



governmentattic.org

"Rummaging in the government's attic"

Description of document: Internal Revenue Service (IRS) Two IBM Reports: Measuring the Indirect Effects of Services and Enforcement on Taxpayer Compliance 2008; and Predicting Taxpayer Behavior 2012

Requested date: 28-August-2024

Release date: 23-September- 2024

Posted date: 14-October-2024

Source of document: FOIA Request
Internal Revenue Service
GLDS Support Services
Stop 211
PO Box 621506
Atlanta, GA 30362-3006
[IRS FOIA Public Access Portal](#)

The governmentattic.org web site ("the site") is a First Amendment free speech web site and is noncommercial and free to the public. The site and materials made available on the site, such as this file, are for reference only. The governmentattic.org web site and its principals have made every effort to make this information as complete and as accurate as possible, however, there may be mistakes and omissions, both typographical and in content. The governmentattic.org web site and its principals shall have neither liability nor responsibility to any person or entity with respect to any loss or damage caused, or alleged to have been caused, directly or indirectly, by the information provided on the governmentattic.org web site or in this file. The public records published on the site were obtained from government agencies using proper legal channels. Each document is identified as to the source. Any concerns about the contents of the site should be directed to the agency originating the document in question. GovernmentAttic.org is not responsible for the contents of documents published on the website.



Department of the Treasury
Internal Revenue Service
Privacy, Governmental Liaison and
Disclosure
GLDS Support Services
Stop 93A
PO Box 621506
Atlanta, GA 30362

Date:
September 23, 2024
Employee name:
A. Mircan
Employee ID number:
1003374216
Telephone number:
904-661-3307
Fax number:
855-205-9335
Case number:
2024-23901

This is a final response to your Freedom of Information Act (FOIA) request we received on August 28, 2024.

You asked for the following:

1. A copy of the complete Final Report dated December 3, 2008, entitled: Measuring the Indirect Effects of Services and Enforcement on Taxpayer Compliance, Produced for the IRS by IBM under Contract TIRNO-08-K-00292.
2. A copy of the complete Final Report dated July 31, 2012, entitled: Predicting Taxpayer Behavior, Produced for the IRS by IBM under Contract TIRNO-09-Z-00021. This report was produced for the Office of Research, Analysis and Statistics.

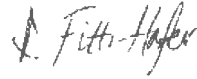
We searched for, and we are enclosing, 285 responsive to your request. This is a full grant of your request.

If you have questions regarding the processing of your FOIA request, please contact the caseworker assigned to your case at the phone number listed at the top of this letter.

If you are not able to resolve any concerns you may have regarding our response with the caseworker, you have the right to seek dispute resolution services by contacting our FOIA Public Liaisons at 312-292-2929. The FOIA Public Liaison is responsible for assisting in reducing delays, increasing transparency, and assisting in the resolution of disputes with respect to the FOIA.

There is no provision for the FOIA Public Liaison to address non-FOIA concerns such as return filing and other tax-related matters or personnel matters. If you need assistance with tax-related issues, you may call the IRS at 800-829-1040.

Sincerely,

A handwritten signature in black ink that reads "D. Fitti-Hafer". The signature is written in a cursive style with a large initial "D".

Deanna J. Fitti-Hafer
Disclosure Manager
Disclosure Office 13

Enclosure:
Responsive Records



Internal Revenue Service

Measuring the Indirect Effects of Services and Enforcement on Taxpayer Compliance

Produced for the IRS under Contract TIRNO-08-K-00292



Final Report

December 3, 2008

Preface

This report is organized into three main sections, which may appeal to different readers:

A. Executive Summary & Introduction

A high-level overview of the entire report, including an understanding of the compelling research value, challenges, and recommendations

Readers seeking an executive understanding of the challenges and opportunities likely to be faced in answering the research question, as well as an overview of the portfolio of research methodologies available should read these sections.

B. Parts I & II: Methodologies and Data Collection

In-depth technical discussion of each of the methodologies, followed by data collection needs and resources to support these methodologies

Readers focused on the planning and implementation of specific projects should read these sections.

C. Parts III & IV: Integrated Plan and Recommendations

The integrated plan with a notional timeline, from conceptual underpinnings to final expected outcomes, including intermediate benefits to the IRS, managed through a portfolio approach

These sections are a guide for those who will be responsible for planning and prioritizing projects within an overall research plan.

Several appendices also provide glossaries and other background material.



Table of Contents

Section	Page
Preface.....	1
Executive Summary.....	1
Introduction.....	6
1.1 Objectives of this Project.....	6
1.2 Project Approach.....	9
1.3 About this Report.....	9
Part I: Proposed Methodologies and Data Requirements.....	11
1 Theoretical Frameworks.....	12
1.1 Theoretical Development.....	12
1.1.1 Value of Theoretical Research.....	12
1.1.2 Existing Theoretical Models.....	12
1.1.3 Scope for New Theoretical Models.....	14
1.2 Segmentation Analysis.....	15
2 Primary Empirical Methods.....	17
2.1 Aggregate Econometric Modeling.....	19
2.1.1 Description/Assumptions/Research Questions Addressed.....	19
2.1.2 Strengths/Weaknesses/Limitations.....	20
2.1.3 Likelihood of Success/Risks.....	21
2.1.4 Expectations (Short-Term and Long-Term).....	21
2.1.5 Dependencies.....	21
2.1.6 Data Needs.....	22
2.2 Microeconomic Modeling – Estimating the Impact on Reporting Compliance.....	22
2.2.1 Description/Assumptions/Research Questions Addressed.....	22
2.2.2 Strengths/Weaknesses/Limitations.....	26
2.2.3 Likelihood of Success/Risks.....	26
2.2.4 Expectations (Short-Term and Long-Term).....	26
2.2.5 Dependencies.....	27
2.2.6 Data Needs.....	27
2.3 Microeconomic Modeling – Estimating the Impact on Filing Compliance.....	27
2.3.1 Description/Assumptions/Research Questions Addressed.....	27
2.3.2 Strengths/Weaknesses/Limitations.....	28
2.3.3 Likelihood of Success/Risks.....	28
2.3.4 Expectations (Short-Term and Long-Term).....	28
2.3.5 Dependencies.....	28
2.3.6 Data Needs.....	29
2.4 Field Experiments and Quasi/Natural Experiments.....	29
2.4.1 Description/Assumptions/Research Questions Addressed.....	29
2.4.2 Strengths/Weaknesses/Limitations.....	30
2.4.3 Likelihood of Success/Risks.....	32
2.4.4 Expectations (Short-Term and Long-Term).....	32
2.4.5 Dependencies.....	32
2.4.6 Data Needs.....	33
2.5 Comparison of Primary Empirical Methodologies.....	34
2.5.1 Speed.....	34
2.5.2 Credibility.....	34



2.5.3	Generalizability	35
2.5.4	Contribution to Understanding.....	36
3	Supporting Empirical Methods	37
3.1	Laboratory Experiments	37
3.1.1	Description/Assumptions/Research Questions Addressed.....	37
3.1.2	Strengths/Weaknesses/Limitations.....	37
3.1.3	Likelihood of Success/Risks	38
3.1.4	Expectations (Short-Term and Long-Term).....	38
3.1.5	Dependencies	38
3.1.6	Data Needs.....	39
3.2	Agent-Based Modeling and Social Network Analysis	39
3.2.1	Description/Assumptions/Research Questions Addressed.....	39
3.2.2	Strengths/Weaknesses/Limitations.....	41
3.2.3	Likelihood of Success/Risks	43
3.2.4	Expectations (Short-Term and Long-Term).....	43
3.2.5	Dependencies	43
3.2.6	Data Needs	44
4	Interrelationships Among Methodologies and Data Needs.....	44
4.1.1	Methodologies that Build a Research Foundation.....	45
4.1.2	Feedback Effects	45
	Part II: Data Collection.....	46
1	Methodology Data Needs	47
1.1	Methodologies with Relatively Little Data Needs.....	47
1.2	Common Data Needs among Methodologies	47
1.3	Specialized Data Needs among Methodologies.....	47
2	Possible Sources to Meet Data Needs	50
2.1	Existing Data Sources.....	50
2.1.1	Tax and Enforcement Data.....	50
2.1.2	Existing Pre-filing / Filing Services Data	50
2.1.3	Consolidating Existing Data: Cross-Channel Services Data Warehouse	54
2.2	Existing Survey Data	55
2.3	Augmenting and Collecting New Data.....	56
2.4	Conducting a Stand-alone Survey	57
2.4.1	Design of Stand-alone Survey.....	58
2.4.2	Extension of this Survey to Provides a Service Panel	60
2.5	Expand Existing IRS Surveys.....	60
2.6	Census Survey	61
2.6.1	Design of Census CPS Survey	61
2.6.2	Census Survey Links to Administrative Data.....	61
2.7	Extension of the National Research Program	62
2.7.1	Purpose of the NRP Extension	62
2.7.2	Links to Administrative Data	64
2.7.3	Data Design.....	64
	Part III: Integrated Plan.....	65
1	Conceptual Underpinnings	66



2	Research Portfolio Management	67
2.1	Managing a Portfolio of Methodologies	67
2.2	Advantages of the “Wide Portfolio” Approach	68
2.3	Portfolio Management Recommendations.....	69
3	Notional Timeline and Staging	71
3.1	Theory (Green on Chart)	73
3.2	Data Collection (Blue on Chart).....	73
3.3	Methodologies (Brown on Chart).....	74
4	Intermediate Outcomes	77
	Part IV: Conclusion and Recommendations.....	79
	References.....	81
	Appendix A: Glossary of Terms	A-1
A.1	Glossary of General Terms	A-1
A.2	Glossary of Methodologies.....	A-2
A.3	Glossary of Data	A-3
	Appendix B: Summary of Current Surveys	B-1
	Appendix C: Review of the Literature	C-1
I.	Introduction.....	C-1
II.	Relevant Literature for Measuring the Indirect Effects of Services	C-1
1.	Theoretical Underpinnings	C-1
2.	Empirical Literature.....	C-6
3.	Deterrence Approaches Used in Environmental Compliance.....	C-10
III.	Bibliography: Papers Selected for In-depth Annotation.....	C-11
	Theoretical.....	C-11
	Econometrics	C-22
	Experimental.....	C-38
	Simulation	C-49
	Other Agencies	C-54
IV.	Bibliography: Full List of Papers Considered.....	C-56
	Appendix D: Point of View Papers.....	D-1
1.	James Alm.....	D-1
2.	Marsha Blumenthal	D-3
3.	Charles Christian.....	D-6
4.	Joe Cordes	D-10
5.	Jon Davis.....	D-12
6.	Don DeLuca.....	D-13
7.	Brenda Dietrich.....	D-15
8.	Jeffrey Dubin.....	D-17
9.	Brian Erard	D-18
10.	Kevin Fletcher.....	D-24
11.	Glenn Galfond.....	D-26
12.	Jim Harvey and Kevin McCrohan.....	D-28
13.	Erich Kirchler.....	D-34
14.	Paul Lehner	D-37
15.	Shelley Metzenbaum.....	D-39



16. Dan Nagin	D-41
17. Alan Plumley.....	D-42
18. John Scholz	D-45
19. Joel Slemrod.....	D-46

List of Figures

Figure 1: Notional Integrated Timeline.....	3
Figure 2: A Continuum of Services	7
Figure 3: Research Design Phases	9
Figure 4: Direct and Indirect Links between Services and Compliance	13
Figure 5: Pre-Filing / Filing Cross-Channel Services Data.....	52
Figure 6: Notional Integrated Timeline.....	72
Figure 7: Microeconomic Modeling Notional Schedule.....	74
Figure 8: Aggregate Econometric Modeling Notional Schedule	74
Figure 9: Field Experiments Notional Schedule	75
Figure 10: Lab Experiment Notional Schedule.....	75
Figure 11: Agent-Based Modeling Notional Schedule	76
Figure 12: The “Slippery Slope” Framework	D-35

List of Tables

Table 1: Comparison of Econometric Analysis and Field Experimentation Approaches	17
Table 2: Ratings of Anticipated Performance of Primary Empirical Methodologies.....	34
Table 3: Merits and Limitations of Supporting Empirical Methodologies	37
Table 4: Summary of Data Needs by Methodological Approach	49
Table 5: Data Desired from Surveys by Method.....	57
Table 6: Intermediate & Long Term Benefits of Methodologies and Data Collection	78

Executive Summary

In 2008, Congress mandated that the Internal Revenue Service (IRS) research the impacts of taxpayer services on voluntary taxpayer compliance. Given the range and scale of IRS service offerings, it is a considerable challenge to measure and monitor their ultimate compliance outcomes. While there is a general consensus among tax administrators that taxpayer services are beneficial to compliance, little more than anecdotal evidence exists to guide judgments about the relative impacts of different services or to make an informed decision on the optimal overall level of service. Furthermore, there is a paucity of evidence on the relative effectiveness of the “carrot” (service activities) and the “stick” (enforcement activities) in promoting voluntary compliance, limiting a tax agency’s ability to strike the right balance between these activities. To enhance understanding of the varied roles that taxpayer services play, the IRS is investing in research on several interrelated areas of taxpayer services, including taxpayer burden, inadvertent taxpayer errors, and taxpayer compliance.

This report focuses specifically on identifying a set of methodologies and supporting data to produce quantitative estimates of the marginal impacts of taxpayer service and enforcement activities on voluntary compliance. These impacts encompass both the “direct effects” on those taxpayers who use an IRS service or who personally experience an enforcement action as well as “indirect effects” on taxpayers who do not. The direct effects of a service include the compliance changes that result when the service helps to clarify a recipient’s filing, reporting, or payment obligations or to reduce the recipient’s compliance burden. The indirect effects of a service include “spillover effects” that result when the delivery of a service to one group of taxpayers influences the transmission of information, attitudes, and perceptions within the general population in ways that ultimately impact the compliance behavior of non-recipients of the service.

Understanding the relative contributions of different service and enforcement activities to tax compliance would be of significant benefit to IRS operations. Specifically, it would enable decisions on resource allocation to be based more squarely on empirical evidence rather than judgments about the probable impacts. The improvements in resource allocation decisions afforded by data-driven estimates should translate into higher levels of voluntary compliance. Moreover, much will be learned about the pathways through which services influence taxpayer behavior, potentially leading to the design of new and more effective taxpayer services that reduce taxpayer burden and improve compliance.

To date, relatively few attempts have been made to measure the influence of IRS services on tax compliance; however, a number of studies have attempted to assess the impact of enforcement, particularly audits, on compliance. To the extent that these enforcement studies serve as a guide, it is reasonable to expect that the measurement of service impacts on compliance will require a concerted effort over a period of years to identify appropriate methodologies; compile and collect necessary data; and implement, interpret, and refine the approaches. Given the technical complexity of the problem, it is also important to have reasonable expectations with respect to the likely outcomes of such an initiative. In particular, it is optimistic to expect that a research effort will result in generally accepted, precise estimates of the impact of all types of service and enforcement activities on all forms of voluntary compliance (filing, reporting, and payment) within several years. More likely, we anticipate that estimates with varying degrees of acceptance and precision will be obtained for a subset of these activities on some forms of

Key Points

- IRS is beginning longrange research to **measure the marginal impact of services on voluntary taxpayer compliance**. This research will further understanding of the mechanisms by which IRS services influence compliance, enabling resource allocation decisions to be based on **empirical evidence and translating into higher voluntary compliance**.
- Choosing a suitable methodological approach for measurement presents a research challenge. Given the nature of data collected on services and compliance, **no single methodology can be guaranteed** to measure the full marginal effects accurately.
- In light of this technical complexity, IRS brought together **academics in the compliance area from around the country to determine the best approach(es)**. This research design plan reflects their recommendations.
- The report recommends that IRS **establish a research portfolio**, composed of **theoretical modeling, econometric analysis and field studies**, augmented with laboratory experiments and social network modeling. The portfolio should be managed through an oversight committee to evaluate success and rebalance the investments.

voluntary compliance, while certain other activities or forms of compliance may go largely unmeasured. Our reasoning is that:

- the estimation of marginal compliance impacts involves very substantial methodological challenges;
- collecting data on all potentially relevant factors for all forms of compliance may be difficult or impossible; and
- alternative estimation methodologies may yield very different results that cannot be reconciled easily or quickly – for instance, different methodologies for evaluating the impact of audits on compliance have yielded widely divergent results.

In addition to generating some estimates of the marginal impact of IRS activities, the research plan will likely yield substantial insights into the nature of taxpayer compliance behavior and the ways that IRS activities impact that behavior. Research in the plan would lead to a deeper understanding of the reasons that taxpayers seek different IRS services; the influence of those services on taxpayer attitudes, perceptions, and motivations; and the ways in which they can have an influence that extends beyond the direct service recipients. This research activity is also likely to generate insights into the reasons that taxpayers use substitute services, such as third-party preparers and software, and how this influences compliance behavior.

To help formulate a research plan, a workshop was held on the *Indirect Effects of Services on Tax Compliance*, with a goal of identifying the most promising methodological approaches and associated data needs for measuring service impacts. The participants included a multidisciplinary group of experts from academia, government and industry with a wealth of knowledge in the areas of tax administration and policy evaluation. The primary research question posed at the workshop was: “What are the direct and indirect impacts of IRS services and enforcement activities on taxpayers’ voluntary compliance?” During the workshop and through a set of point-of-view papers submitted prior to the event, the experts proposed and discussed a variety of alternative estimation methodologies for addressing this question.

Although it was recognized that there were significant challenges and limitations associated with each of the proposed methodologies, the experts ultimately concluded that econometric methods and field experiments have the best potential for generating reasonably reliable quantitative estimates of marginal effects. Several econometric methods were proposed. Some of these take advantage of micro-level data, such as the National Research Program (NRP), while another approach relies on longitudinal data on reporting and filing behavior aggregated at a geographic level such as by state. In the case of field experiments, the experts recommended conducting real world controlled experiments in which the compliance behavior of a treatment group that receives a differential level or quality of a given service is compared to the behavior of a control group that receives the standard level and quality of service.

Three primary methodologies have promise for generating quantitative estimates:

- **Aggregate Econometric approaches** have been applied in past research on direct and indirect effects of enforcement activity, but must overcome some challenging modeling issues to account for services (page 19)
- **Microeconomic approaches** provide insight into individual compliance drivers but may not easily account for social effects or time lags without further theoretical development (page 22).
- **Field Experiments** measure real world compliance impacts in a short amount of time, although the ability to generalize results would depend on the design of the experiment (page 29).

Some methodologies were deemed *supportive*, likely to contribute valuable insights on the relationships between IRS activities and tax compliance. **Laboratory experiments** provide a useful way to test hypotheses in a controlled environment free of confounding factors. However, the external validity of this approach is likely to be limited. **Agent-based models**, though relatively new to the set of analysis and modeling tools, have the potential to generate useful insights into the ways service and enforcement activities impact on compliance behavior. Such models are designed to capture micro-behaviors and variability over time of the actions and decisions of the key players in the tax ecosystem including taxpayers, third parties such as tax professionals and the IRS. Outcomes of such modeling activities have the potential to produce emergent behaviors and to generate insights on social behaviors within the ecosystem. Strictly speaking, it is possible to develop agent-based models that rely solely on assumptions regarding social interactions and behaviors in the absence of relevant data. However, it was generally agreed that such models have greater promise when they are more data-driven, based on detailed individual level data on socio-psychological

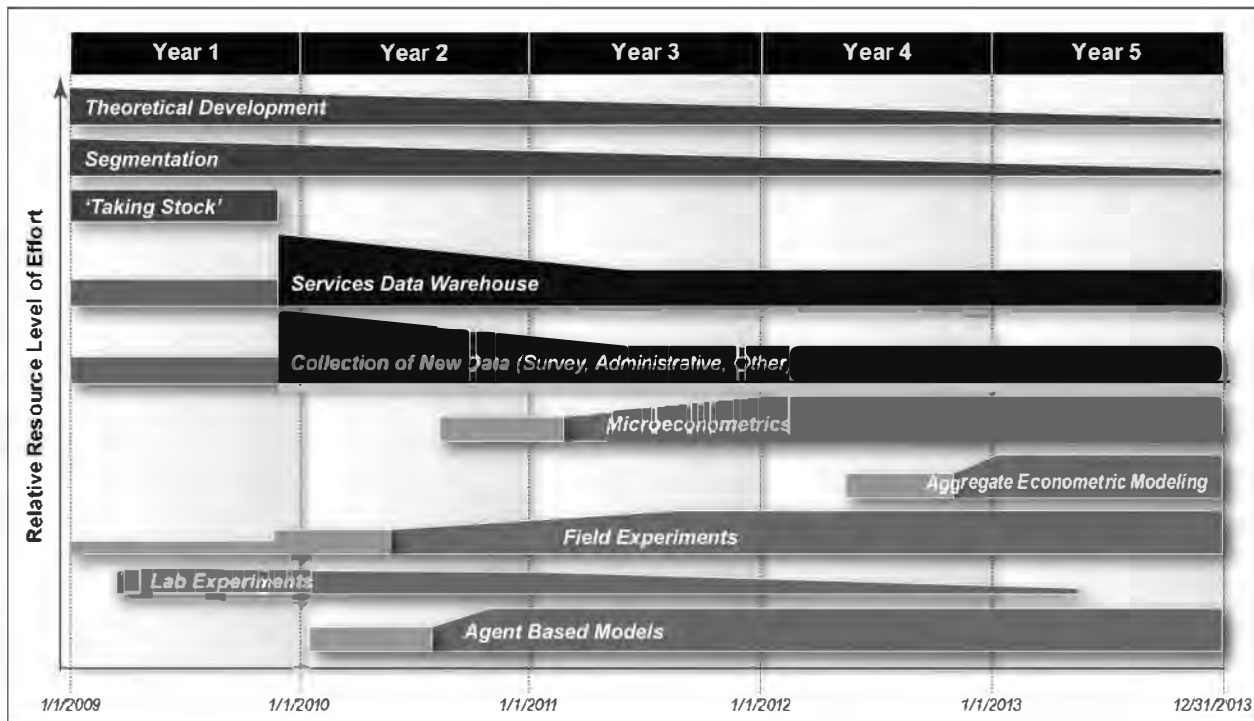
factors, taxpayer networks and compliance behavior. The IRS has some working experience with agent-based models, having already built or leveraged two such models for other initiatives.

There was consensus that the IRS should facilitate the development of theoretical frameworks that incorporate services into models of taxpayer compliance. Such frameworks would generate testable hypotheses, inform the specification of empirical models, and provide a basis for interpreting empirical results. There was also broad consensus across different methodology breakout groups that segmentation analysis would serve as a useful tool for identifying and understanding the sources of heterogeneity among taxpayers with respect to their motivations and compliance behavior.

The proposed methodologies all have significant data requirements, with the exceptions of theoretical models and laboratory experiments. One common data need is service utilization data consolidated in a standardized research format. Also desirable for many methods are data on taxpayer attitudes and perceptions about services, the IRS, and government. At least some of the required information can be leveraged from existing data sources. Establishing a common repository for services data would ensure that the data required by the methodologies would be easily accessible and properly maintained, while facilitating cross-channel research. Existing surveys, conducted by the IRS and other government agencies, could be augmented to acquire behavioral and attitudinal data. Some methods, such as microeconomic methods which require NRP and/or tax return data at the level of the individual taxpayer require more disaggregated data than others. Other methods have additional specialized data needs.

From the experts' identification of methodologies and associated data requirements, a plan was developed for collecting and assembling the required data and implementing the methodologies. The various research threads should be actively managed as part of a portfolio of methods and data collection initiatives. This portfolio should be managed using widely accepted performance management principles and evaluated on a periodic basis, with actions taken to expand or discontinue different research threads depending upon their performance. Based on such an approach, a five-year research plan includes the following notional time schedule for data collection and research activities.

Figure 1: Notional Integrated Timeline



This plan incorporates theoretical development, data collection, and empirical modeling, based on currently available information about the interrelationships among the different methodologies and their associated data requirements. The lighter shading at the beginning of the data collection and methods sections indicate a period of planning which is necessary before implementation and analysis.

- **Theoretical work**, represented in green on the above schedule, is seen as a critical first step in order to enhance understanding prior to the implementation of the various methodologies. Therefore, there is a greater amount of theoretical work done at the beginning. Of particular importance is “taking stock”, an exercise to evaluate the current resources, data, and analysis being performed within the IRS. Assessing current data availability will inform what additional data collection efforts are required.
- **Data collection**, depicted in blue in the chart, include the creation of a services data warehouse and the collection of new data. A Services Data Warehouse would gather and store all service data in a single location for easy accessibility to researchers. Collection of new data would look at ways in which the IRS could gather more data, such as additional surveys. Both of these data efforts are seen as having a planning phase, followed by an increased level of effort during implementation, and a slow ramp down to a maintenance phase.
- **Empirical methods**, colored brown, include both the methodologies that can answer the primary research question and methodologies that can support the primary research. The primary methodologies of aggregate econometrics and microeconometrics cannot begin until enough data has been collected in order to make valid statistical models, which is why these do not begin to ramp up until year 4 and year 3 respectively. Field experiments, another primary methodology, can begin sooner, in year 2, with some planning beforehand in order to design them. Both lab experiments and agent-based modeling (ABM) models are seen as supporting methodologies. Lab experiments can begin early and can be done quickly and inexpensively, but will ramp down as other methodologies begin to yield credible measurements. ABM, while considered a supporting methodology, can be useful in validating the results of other models and enhancing the theoretical understanding, so once it ramps up, it continues at a constant rate.

Built into this schedule are intermediate benefits, as well as the long-term potential outcome of measuring the effects of service on compliance. For instance, a Services Data Warehouse would serve the long-term goal by providing data for econometric modeling, but it would also organize the services data into a comprehensive framework, which could then be used by many researchers in a variety of ways. Theoretical development helps the primary empirical methodologies, while providing the intermediate benefit of aiding the IRS understanding of taxpayer groups, services usage and compliance behavior. Preliminary results from primary methodologies are also expected to improve understanding of taxpayer compliance. This schedule is ultimately notional and should be refined and updated on a regular basis using the aforementioned portfolio management principles.

Partial Project Benefits

- Enhanced understanding of the role of services in compliance, ultimately yielding **quantitative measurements**
- An accessible, **standardized data repository for cross-channel services research**
- Data analysis on the use and distribution of IRS services, including a **services demand model** for more efficient sampling

For a full list, see Table 6 (page 77).

To implement the research plan:

- I. The IRS should invest in a **dynamically managed portfolio of methodologies**.
 - a. The IRS should manage this portfolio through a research oversight committee, with **periodic performance evaluation** and appropriate revisions.
 - b. The IRS should prioritize those methodologies with the best chance of success in estimating the marginal direct and indirect effects of services and enforcement activities.
 - c. The IRS should begin an assessment of all relevant service and compliance data to **catalogue existing knowledge of taxpayer compliance behavior and attitudes**, culminating in a final report.

- II. The IRS should invest in **foundational strategies early on**.
 - a. The IRS should **facilitate the development of theoretical** frameworks to better understand the relationship between services and compliance.
 - b. IRS should distinguish **relevant taxpayer segments from a compliance perspective**.
 - c. IRS should judiciously **test theoretical frameworks in the laboratory environment**.

- III. The IRS should invest in a **strategy that yields desired intermediate outcomes** as well as the final objective, **the measurement of marginal indirect and direct effects**.
 - a. Intermediate benefits include a **better understanding of taxpayer services usage, attitudes and compliance behaviors** before quantitative estimates will be available.
 - b. It is desirable to measure **impacts on intermediate variables** that are likely to improve compliance (such as taxpayer satisfaction) before estimates of the impact on compliance are likely to be available.

- IV. To meet the data needs, the IRS should **leverage existing data and collect new data**.
 - a. The IRS should **consolidate and standardize services data** in a common repository accessible to IRS researchers.
 - b. The IRS should examine existing surveys to determine what data are available on taxpayer attitudes, networks, behaviors and perceptions and consider **augmenting existing surveys or facilitating new ones**.

- V. Once appropriate data and insights are in place, begin investing in the measurement of indirect and direct effects.
 - a. In order to measure the marginal direct and indirect effects of services and enforcement activities, the IRS should **invest in carefully controlled field experiments**.
 - b. Simultaneously, the IRS should **pursue both aggregate econometric methods and microeconomic methods**.
 - c. The IRS should assess existing **social network/ABM models** or develop new ones for the purposes of representing social networks and third parties in compliance.

Introduction

The IRS mission statement emphasizes providing "...America's taxpayers top quality service by helping them understand and meet their tax responsibilities and by applying the tax law with integrity and fairness to all." To achieve this mission, the agency actively manages a wide range of informational and assistive services. The IRS spends over \$500 million on pre-filing taxpayer assistance and education alone, and the effective reach of services is much wider, from refund assistance to answering compliance-related calls and questions.

Given this enormous investment, the outcomes of taxpayer service activities are of great interest to the IRS, Congress, and other agency stakeholders. Understanding such outcomes enables tax administrators to improve voluntary compliance while providing necessary and desirable taxpayer services. In 2005, Congress mandated that the IRS create a unified vision for services that addresses the needs, preferences, and behaviors of taxpayers. This resulting service-wide initiative, the Taxpayer Assistance Blueprint (TAB), launched a five-year plan for the evaluation and improvement of services.¹ The TAB report identified voluntary taxpayer compliance as one of the three major outcomes for services yet has not found a comprehensive measurement approach. This measurement is critical to understanding the full value and return on investment of these services.

The 2008 Budget initiative, *Measuring the Impact of Services on Compliance*, provided a vehicle to address this measurement question within the context of continued research on the relationship between services and compliance. As part of this project, the IRS is researching taxpayer service channels, taxpayer burden, and taxpayer errors. The fourth key component is measuring the direct and indirect effects of IRS services on voluntary taxpayer compliance. In contrast to the direct effect of a service on the compliance behavior of recipients, indirect effects refer to spillover effects on non-recipients. For instance, taxpayers who receive assistance from an IRS help desk may share what they learn with other individuals, thereby indirectly influencing the compliance behavior of those individuals. Alternatively, feedback regarding other taxpayers' direct experiences with the IRS may influence one's own attitudes about the agency, which could ultimately influence one's compliance behavior. This report presents an integrated research plan for investment in a portfolio of research and associated data acquisition activities for measuring the direct and indirect effects of service and enforcement activities on taxpayer compliance.

1.1 Objectives of this Project

The objective of this project is to identify and plan for the implementation of a set of methodologies that are deemed by a group of subject matter experts to hold the most promise for success in generating sound empirical estimates of the impact of services and enforcement activities on the voluntary compliance behavior of taxpayers. Such estimates could be used to guide a range of IRS management decisions including budget decisions for services support, setting priorities for new services development, and addressing other tax administration priorities.

There are two intrinsically related questions regarding services. The first is "How can the IRS measure the effects of its service and enforcement activities on compliance?" The second is "How can the IRS improve these activities (or develop new services) to promote greater compliance?" This report focuses specifically on the former question by examining methodologies that can be used to *measure the effects*.

While the specific focus of this report is on measurements, it is important to note the broader goal of these measurements is ultimately to increase compliance and improve service to taxpayers. Improved knowledge of the relative contributions of different service and enforcement activities to tax compliance should enable adjustments in

¹ TAB is a joint effort of IRS, the IRS Oversight Board, and the National Taxpayer Advocate to develop a five-year plan for taxpayer service, focused on taxpayers who file (or should file) a Form 1040 series tax return. The objectives were to: "1. Establish a credible taxpayer and partner baseline of needs, preferences, and behaviors; 2. Implement a transparent process for making service-related resource and operational decisions; 3. Develop a framework for institutionalizing key research, operational, and assessment activities to holistically manage service delivery; and 4. Utilize both short-term performance and long-term business outcome goals and measures to assess service value." TAB reports are publicly available at <http://www.irs.gov/individuals/article/0,,id=156394,00.html>.



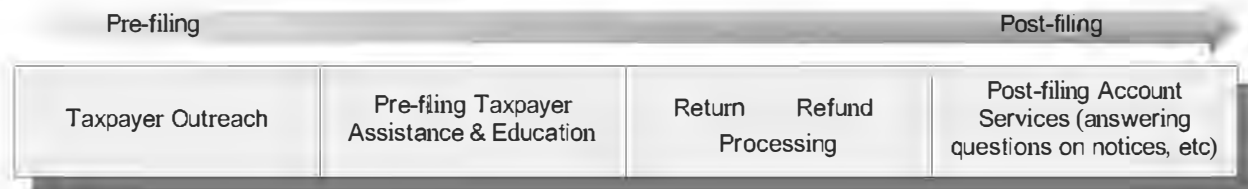
resource allocation that ultimately lead to improvements in the overall compliance impact of IRS activities. Further, through the process of developing and applying methodologies to measure the relative contributions of different activities, it is likely that much will be learned about the pathways through which services influence taxpayer behavior, potentially leading to the design of new and more effective taxpayer services. Finally, activities that are discovered to have little or no impact on taxpayer compliance would be logical targets of research to determine if they can be improved.

The Range of Services

The measurement strategy developed in this report is supportive of the wider IRS community goals of improving taxpayer services and compliance – specifically identifying key benefits of the planned research activities to IRS executive and managerial decision makers. Service improvement is a wide policy domain, incorporating a number of divisions at the IRS, including Wage & Investment (W&I) and W&I Research, Taxpayer Services Project Management Office (TSPMO) and the Taxpayer Advocate Service (TAS). Implementation of the proposed plan for data acquisition and analysis should contribute to many ongoing service initiatives, including efforts to better understand the needs, motivations, and behaviors of service users; assess the best ways to manage and improve services; and jointly model service and enforcement activities.

Generally within the IRS, services are largely the domain of Wage & Investment, the Business Operating Division in the IRS that serves individual taxpayers who do not use a range of “business related” forms such as the Schedule C. This is the scope adopted by TAB. However, we did not want to limit discussion in the design stage by defining a particular range of services by type of interaction, taxpayer type, or channel delivered, for several reasons. First, IRS outreach and education occurs for every major taxpayer segment, including large and mid-sized businesses. Second, all IRS services are duplicated in private sector tax preparation and personal finance services, as well as provided informally or on a volunteer basis. Lastly, “services” occur throughout the full lifecycle of IRS / taxpayer interactions, from taxpayer outreach to post-filing account management. Services and enforcement activities are more appropriately a continuum of interaction (Figure 2: A Continuum of Services).

Figure 2: A Continuum of Services



While no official definition of “services” defined our scope, the majority of recommendations focus around the TAB domain: information provided through walk-in, telephone, and internet venues to individual taxpayers, 1040 filers, self-employed or potentially small businesses. Therefore, the focus of this report is on designing a plan to examine the impact of IRS activities on the compliance behavior of individual taxpayers.

Examples of IRS Services:

- The IRS provides a wide variety of services via the web at irs.gov, such as forms and publications, tax law calculators and guided tax law interviews decision support, refund inquiries, FreeFile preparation links, information on other services, and information on compliance events such as payments, notices, Offers in Compromise, Innocent Spouse, etc.
- The IRS runs numerous **call centers** which answer taxpayers’ pre-filing and post-filing questions. Pre-filing services include refund questions, tax law questions, Ask Taxpayer Advocate, and E-Help, while enforcement-related services include customer account management and tax fraud hotline. When a taxpayer calls in, they hear a voice menu, which may either direct them to automated interactions or connect them with appropriate customer service representatives (CSR). During FY 2007, over 23M calls were received which utilized the automated service option of toll free assistance call center, while over 33M calls were answered by an IRS representative.
- In 2007, there were approximately 7M visits to **Taxpayer Assistance Centers (TAC)**. TAC centers provided a full gamut of services, including providing publications, answering questions and accepting payments. Ongoing pilot studies of electronic self-help resources at the TAC may provide insights into taxpayer service usage.

To the extent possible, the measurement method must distinguish the impacts of IRS services and substitute services, since services received through third parties are rather distinct from those provided by the IRS. The prevalence of paid professional usage varies by group of taxpayer and is driven by a variety of factors, including the complexity of the tax situation and the tax acumen of the taxpayer, while the role of paid preparers in compliance may vary with taxpayer risk preferences.

Defining Compliance

Researchers and administrators have long distinguished among the three general types of compliance: filing, payment and reporting. Underreporting is the focus of current compliance research efforts such as NRP, from which IRS estimates that underreporting accounts for 82.5% of the tax gap. This report also identifies methods for researching how the quality of services influences the incidence of noncompliance by “ghosts” (non-filers and late filers), which the IRS estimates to make up 7.8% of the tax gap. In addition, certain of the proposed methodologies, such as field experiments, may prove useful for investigating service impacts on a variety of compliance issues, including payment compliance. Compliance in the context of corporate or large and mid-sized businesses was not discussed.

Research Considerations

Measuring service impacts on compliance outcomes is a formidable challenge. To achieve the primary research goal of estimating the direct and indirect effects of services and enforcement on compliance, we identified several key issues that need to be addressed:

- Filing, Reporting, and Payment compliance each have different motivations and are likely to be differentially impacted by taxpayer services. Some estimation methods may be able to address multiple types of compliance, yet others will not.
- The research must address the complexity of the relationship between services and tax compliance. Some examples include: Are there threshold effects? Is there a significant lag between the provision of a given service and its impact on compliance? Does this impact attenuate over time? Do different subgroups have very different behaviors?
- Estimating the impact of services invariably requires accounting for the level of enforcement so that interactions and relationships among these activities can be addressed.
- Both the quantity of a service or enforcement activity and its quality (such as accessibility, timeliness, accuracy, taxpayer satisfaction, etc.) may impact compliance.
- The research method will measure both the impact of a change in the level or quality on one activity, holding other activities constant on taxpayers who are direct recipients of the activity (the direct effects) and taxpayers in the general population (the indirect effect).
- Substitute services provided by third-parties are likely to play a major role in tax compliance.
- There may be important interaction effects among enforcement activities, IRS services, and substitute services.
- Accounting for the nature of the relationship between taxpayer perceptions and attitudes about IRS activities and objective measures of these activities is likely to be important when attempting to measure the impact of these activities on compliance.

Some additional challenges include:

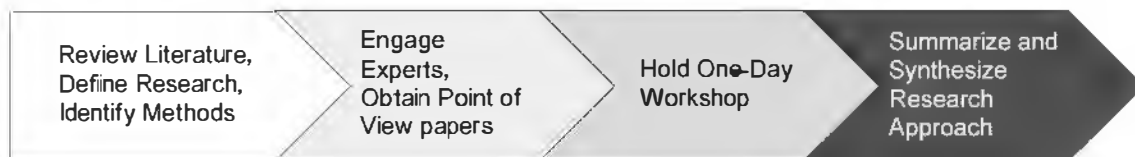
- Pre-filing assistance and education services such as publications, forms, and tax law questions are largely taxpayer-initiated, anonymous inquiries. Therefore, services data collection is limited and procedures vary across service activities. In contrast, taxpayer-specific records are maintained for many post-filing service activities.
- There is a strong belief that services positively impact taxpayers’ psychological impressions of the government, the IRS, and their fellow taxpayers, but there is little hard evidence to confirm these impressions.
- Voluntary taxpayer compliance itself can be only partially observed; even relatively intensive audits are sometimes unable to uncover instances of noncompliance.

1.2 Project Approach

Given these challenges, the IRS sought to engage a multidisciplinary team of subject matter experts from academia, industry, and government to consider the core research objectives discussed above. Allowing for a wide range of expert opinions was necessary to ensure that we considered the full range of methodologies with the most promise of generating sound estimates. To accomplish this, we took the following steps in this project.

- Conducted a Review of the Literature in the field of tax compliance as well as related fields which were believed to be able to provide a unique methodological approach and valuable insight into compliance issues. This literature review provided both an understanding of the current state of the research and a basis from which to select participants for a workshop.
- Recruited a diverse set of experts to participate in a one-day workshop. We asked each participant to provide a “Point of View” statement on the topic that included his or her initial thoughts on how to answer the research questions. The documents were compiled and distributed to all participants before the workshop, and they can be found in the attachment, Workshop Materials.
- Held a one-day workshop, called “Estimating the Indirect Effects of Services on Compliance,” where participants were tasked to identify and evaluate all possible methods, identify the associated data needs, and develop a consensus regarding which of the possible methodologies hold the most promise in addressing the primary research question. Follow-up conference calls gave participants the opportunity to express further opinions.
- Compiled a Research Design plan, described in this report, reflective of the panel’s combined expertise and IRS stakeholder input, paired with a realistic data acquisition plan best suited for implementation.

Figure 3: Research Design Phases



1.3 About this Report

This report presents the recommended methodologies, the additional data needed for their implementation, the steps the IRS needs to take to compile the needed data, and a realistic timetable for short-term and long-term research. Part I describes five methodologies which can aid in estimating the direct and indirect effects of services on compliance, their dependencies and data needs. Part II describes the data sources and data acquisition requirements to implement the methodologies. Part III describes a portfolio approach to accomplishing this research and presents a comprehensive multi-year research plan. Part IV concludes with the full set of research recommendations.

These recommendations represent the contributions, and often the consensus, of a wide range of technical experts. The authors’ goal is to preserve the original ideas and viewpoints expressed in the brainstorming session and point of view papers, while crafting these initial inspirations into viable recommendations. With the impartial, inclusive spirit guiding design of this new and important research initiative, efforts were made to solicit and incorporate further feedback from participants. However, it would be impossible to have this report represent all participants’ perspectives *in toto*. Implementation solutions are solely the suggestion of IBM.



Part I: Proposed Methodologies and Data Requirements

1 Theoretical Frameworks

Two foundational needs emerged as priorities among the workshop participants: theoretical development and segmentation analysis.

Theoretical development was viewed as an urgent need for guiding the development and interpretation of empirical models. It was recognized that theoretical research on the role of tax *enforcement* in tax compliance has been fairly well developed over several decades but that our understanding of the role of *services* is inchoate.

It was further recognized that there exists a substantial degree of heterogeneity in compliance motivations and behavior among the general taxpayer population. Many of the workshop participants emphasized that accounting for such heterogeneity is essential to reaching robust conclusions about the role of taxpayer services in compliance (see Blumethal, Dietrich, Harvey/McCrohan and Kirchler in the Point of View papers in Appendix C: Review of the Literature). Segmentation analysis was proposed as a means of dividing taxpayers into separate groups according to their apparent motivations and behaviors.

In this section, we provide a description of the proposed theoretical development and segmentation analysis activities and their expected contributions to the research project.

1.1 Theoretical Development

In this section, we discuss the value and scope of theoretical research in support of empirical estimation efforts.

1.1.1 Value of Theoretical Research

Theoretical models that link taxpayer services and enforcement activities to compliance behavior can serve as an important guide for empirical studies. They can help to identify and explain causal relationships, make qualitative predictions, and generate testable hypotheses. This helps to clarify which variables and parameters are important to include in an empirical study and whether a given variable should be treated as endogenous or exogenous. Theoretical models also provide a framework for interpreting the results of empirical studies.

Although it is possible to estimate empirical models in the absence of a theoretical foundation, such research often tends to be more exploratory in nature. Under such an approach, one has less confidence that the model is accounting for the most relevant variables, relationships, and parameters; the results are often subject to varying interpretations; and it is more difficult to assess causal implications.

1.1.2 Existing Theoretical Models

As indicated in the literature review in Attachment A, various theories have been posited for the relationship between tax enforcement activities and tax compliance. Under the traditional economic model of tax compliance, the likelihood of enforcement and the severity of the penalty for noncompliance play a key role in discouraging noncompliance. An alternative perspective is provided by certain socio-psychological theories of tax compliance. Under one such theory enforcement activities can, under certain circumstances, “crowd out” internal motivations to comply with the tax laws, thereby resulting in less voluntary compliance. Other theories focus on the interplay between social factors, such as social stigma, and tax enforcement.² These alternative theories argue for a more complex and uncertain relationship between tax enforcement and tax compliance.

At the workshop, it was recognized that much of the general deterrent effect of IRS services was likely to be rooted in social and psychological factors. The compliance decisions of taxpayers generally depend on both power of the

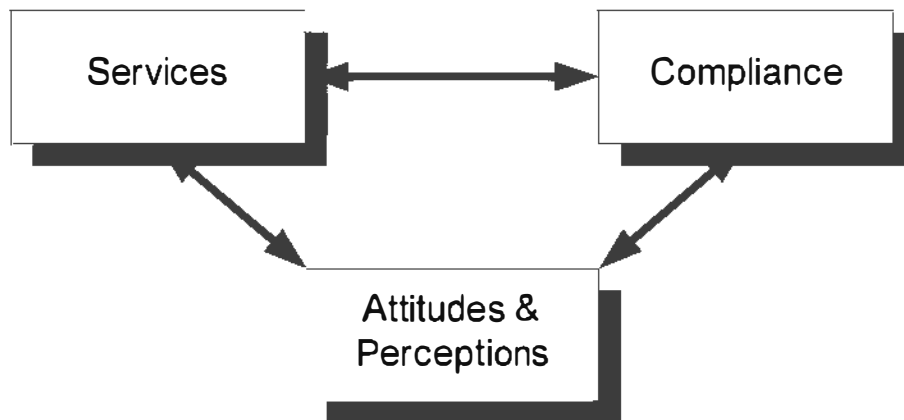
² For details on the traditional model, refer to the discussion of Allingham and Sandmo (1972) and related work provided in the literature review in Appendix C: Review of the Literature. The literature on intrinsic and extrinsic motivations is reviewed in Bruno S. Frey (1997). The papers referenced in the literature review by Gordon (1989) and Myles and Naylor (1996) examine the role of social stigma on tax compliance behavior. In Appendix D: Point of View Papers, Erich Kirchler describes the ‘slippery slope’ theoretical framework.

tax authority to enforce compliance and on the level of taxpayer trust in the tax authority. Trust is influenced by subjective tax knowledge, attitudes towards government in general and the tax authority in particular, personal and social norms, and perceptions of fairness. Such perceptions may be in relation to taxpayer expectations regarding procedural justice, composed of qualities such as service quality, fairness in the application of tax rules and procedures, the level of courtesy in interpersonal communications, and the quality of goods and services provided in exchange for their tax payments. Compliance behavior is also mediated by perceived distributive fairness and normative values, such as reciprocation and duty to pay taxes. Perceptions of fairness, levels of trust, and the strength of norms to comply all may be influenced by direct taxpayer service experiences as well as by information about taxpayer services that is learned through social networks and media communications.

Theoretical modeling of the impact of taxpayer services on tax compliance is in its infancy. As illustrated in *Direct and Indirect Links between Services and Compliance*, taxpayer services play both a direct and an indirect role in influencing compliance. Taxpayer services can directly impact the compliance of recipients by clarifying their filing, reporting, and payment obligations and reducing their compliance burdens. To the extent that direct recipients of taxpayer services share what they have learned with others, the compliance behavior of such individuals also may be affected. More generally, though, taxpayer services may influence taxpayer attitudes and perceptions, which in turn may impact compliance behavior.

Observe that the arrows in Figure 4: *Direct and Indirect Links between Services and Compliance* go in both directions, indicating that compliance behavior can impact one's decisions about using taxpayer services and even one's attitudes and perceptions. Taxpayers who desire to comply but are confused about their obligations may be more prone to using services. In contrast, deliberate misreporters may be reluctant to contact the IRS for assistance. Further, one may sub-consciously develop attitudes about the IRS and its services that help to rationalize one's compliance behavior.³

Figure 4: Direct and Indirect Links between Services and Compliance



There are various socio-psychological theories of tax compliance that capture elements of this relationship. However, they have not focused specifically on the role of taxpayer services. One insight from existing models is that such services may not always promote compliance. For instance, economic theory predicts that a risk-averse taxpayer will tend to be more conservative in his reporting strategy when his true tax liability is uncertain. Therefore, if an informational service, such as a taxpayer help line, serves to reduce the taxpayer's uncertainty, this could actually lead to reduced compliance.⁴

³ Other drivers of compliance are not shown here. For example, excessive burden and complexity may negatively impact compliance.

⁴ For more details, refer to the discussion of Scotchmer and Slemrod (1989) in the literature review in Appendix C: Review of the Literature.

It is important to recognize that the IRS is not the only provider of taxpayer services. Taxpayers also have access to assistance from tax practitioners, tax preparation and tax planning software, third-party publications, and even informal advice from the taxpayer's acquaintances and family members. Existing theoretical research on substitute services has been largely focused on understanding the role of tax practitioners. One of its key insights has been that tax practitioners may play a dual role with respect to tax compliance. According to this view, practitioners will enforce compliance on clear cut tax issues, but will tend to promote noncompliance on more ambiguous issues by taking aggressive tax positions.⁵

1.1.3 Scope for New Theoretical Models

There is a need for additional theoretical research to guide empirical modeling of tax compliance, particularly with regard to the roles of IRS-provided taxpayer services and substitute services in tax compliance. Key issues for investigation include:

- How IRS-provided services impact compliance behavior;
- How substitute services impact compliance behavior;
- The choices among IRS-provided and substitute services; and
- The interactions among taxpayer services and tax enforcement activities.

The objective of this research would be to generate insights into pathways through which services and enforcement activities are likely to impact tax compliance; identify key variables and parameters for empirical analysis; and produce testable hypotheses about taxpayer behavior.

To improve our understanding of indirect effects, it would be desirable to have a unified theoretical framework that specifies the mechanisms through which taxpayer services and enforcement activities impact a clear and measurable set of taxpayer attitudes and perceptions, as well as the mechanisms through which these attitudes and perceptions ultimately impact tax compliance. This specification would explicitly account for relevant social networks and other sources through which information and interactions that influence taxpayer attitudes and perceptions are diffused. Such a framework might provide insights into several key issues, including: (1) the ways in which different segments of the population are likely to be influenced by taxpayer services and enforcement activities; (2) how social networks form and their role in the compliance process; and (3) the dynamics of compliance responses over time. Such a framework would be a useful guide for empirical specifications that might yield valuable intermediate results on the impact of taxpayer services on taxpayer attitudes, perceptions, and levels of satisfaction. Ultimately, it might also contribute to empirical modeling of the extent to which these changes in taxpayer attitudes, perceptions, and levels of satisfaction impact compliance.

Theoretical research would be an ongoing element in the research portfolio, guiding empirical studies. So too could empirical results inspire the revision or refinement of theoretical frameworks. To lay the groundwork for development of theories regarding services, a high level of theoretical development occurs early in the project portfolio. The timing of theoretical research, in conjunction with empirical methods, is discussed more in Part III Section 3 Notional Timeline and Staging (page 71).

⁵ For more details on this issue, refer to the discussion of Klepper, Mazur, and Nagin (1991) in the literature review in Appendix C: Review of the Literature.

1.2 Segmentation Analysis

Segmentation analysis focuses on understanding taxpayer groups and recognizing the differences among those groups that are likely to impact compliance. Through segmentation, measurement strategies reflect the differences in motivations, influences, and behaviors of different segments of the population. This approach involves the analysis of data from surveys, audit studies (such as the NRP), and other sources to identify population segments with different compliance characteristics, learn about factors that influence compliance by segment, and develop insights into how the IRS can enhance compliance by segment.

The main objective of segmentation analysis is to generate insights into the nature and implications of taxpayer heterogeneity that can guide subsequent theoretical and empirical modeling efforts. This methodology can address the following research questions:

- What are the key factors that impact compliance?
- Who is affected by these factors, and by how much?
- Where can the IRS best focus its actions?

A segment represents a group of taxpayers identified by a common viewpoint, likely to respond similarly to changes in the tax environment (such as an improvement in services or a reduction in uncertainty). Various segmentations should be considered. One way of segmenting the taxpayer population is through administrative data on service “market segments,” distinguishing groups of taxpayers by similar use of various IRS services. Another possible segmentation is by risk of noncompliance. Segmentation is generally more complex than simply one variable class (e.g. income), since it delineates enough characteristics to target a meaningful behavioral outcome.

As service utilization and compliance outcomes vary among taxpayers, understanding the unique factors underlying this heterogeneity in behavior is vital to addressing the primary research question regarding the direct and indirect impacts of IRS activities on taxpayer compliance. While segmentation analysis does not empirically measure the impact of the IRS services on compliance, it does provide an important foundation that can be used by all other methods (three of the four workshop breakout groups specifically mentioned the need for better taxpayer population segmentation as a key input to those methodologies). For instance, through an improved understanding of different taxpayer segments, one can tailor econometric specifications to capture the salient characteristics and issues associated with each segment. Such knowledge would also be useful when designing behavioral models that account for heterogeneous behavior among taxpayers, when selecting treatment and control groups for field experiments, and for the creation of behavioral assumptions and “rules” in the development of agent-based models.

The IRS has already utilized segmentation analysis to improve its understanding of the service needs of taxpayers. It has been recognized that taxpayers represent a diverse group with wide variations in age, income, language, education, ability, and tax characteristics. Phase 1 of the IRS Taxpayer Assistance Blueprint (TAB) focused on income and generational segments, while TAB Phase 2 utilized those segments but also identified education as another useful segmentation criterion.⁶ However, experts suggested that existing segmentation is a starting point: refinement includes segmentation based on the data, evaluating whether the data supports segmentation and continual refinement as more data is collected.

In the short term, segmentation analysis will assist researchers in understanding the causes, magnitudes, and potential reasons for noncompliance by population segment. This is likely to contribute to improved theoretical models of what drives the behavior of different groups of taxpayers, and it should provide insight into what data should be collected.

⁶ IRS Taxpayer Assistance Blueprint report (<http://www.irs.gov/pub/irs-pdf/p4579.pdf>). Several earlier studies on compliance segments are discussed in the Point of View paper by Jim Harvey and Kevin McCrohan in Appendix D: Point of View Papers.

Over the longer term, results from segmentation analysis should be useful as a guide for specifying empirical models of the impact of IRS activities on compliance behavior.

A potential challenge for the approach is that some important taxpayer segments may not be easily distinguished by observable or measurable characteristics, such as the social networks to which taxpayers belong. Another risk associated with this methodology is that the compliance response of taxpayers to IRS activities may not vary significantly across different taxpayer segments. However, this seems unlikely as different segments face different issues and will probably respond differently to various services and enforcement actions.

Segmentation analysis would not be dependent upon the results from other methodologies, but would require adequate data on compliance behavior, return characteristics, demographics, attitudes/motivations, beliefs, and opportunities. This information could be collected from surveys and administrative data sources, such as the NRP.

2 Primary Empirical Methods

This section discusses the best methodologies to answer the primary research question of this report: “What are the marginal direct and indirect impacts of IRS services and enforcement activities on taxpayers’ voluntary compliance?” Based on the consensus of the experts at the workshop, only a few methodologies have the potential to provide sound quantitative estimates of marginal impacts; in particular, certain approaches that rely on econometric modeling and field experiments.

Econometrics

In a broad sense, econometrics consists of the development and application of statistical methods and statistical tests for studying economic problems. From this perspective, many of the alternative empirical methods described in this report involve econometric methods, since they typically employ statistical techniques to analyze results and test hypotheses. However, in this section, our focus is on certain “stand-alone” econometric methodologies that were proposed at the workshop for using observational data (as opposed to data from controlled experiments) to measure and understand the impact of taxpayer services and enforcement activities on tax compliance. To examine the impact of these activities on compliance, this report outlines three alternative econometric methodologies:

- aggregate econometric estimation techniques to be applied to longitudinal (panel) data on taxpayer reporting behavior aggregated by geographic area;
- microeconomic estimation techniques to be applied to individual level data from a combination of operational, random, and survey data sources on tax reporting compliance; and
- a microeconomic methodology for investigating the impact of service activities on filing behavior.

Each of these methods seeks to measure the causal association between service and enforcement activities and tax compliance. Since these methods rely on observational data, they must control for numerous potentially confounding factors relating to taxpayer characteristics and incentives that also influence compliance behavior.

Field Experiments

The other primary empirical method that was proposed for answering the primary research question is field experiments. In contrast with the proposed micro and aggregate econometric methodologies, this approach relies on treatment and control groups to investigate the impact of a change in service or enforcement activities on compliance behavior. Under this approach, a researcher evaluates how behavioral responses differ between members of a control group who are offered the usual level and quality of a given service and members of a treatment group who are offered a different level or quality of the service. The behavioral responses of particular interest for the current project are the extent to which taxpayers use the service (take-up) and the level of tax compliance. Ideally, the treatment and control groups that are selected for a field experiment are quite similar, so that in the absence of the treatment, they would both demonstrate very similar behavior. When this is so, one may reasonably infer that a difference in program take-up or compliance among the two groups is attributable to the treatment.

Table 1 illustrates the main differences between the econometric and field experimentation approaches.

Table 1: Comparison of Econometric Analysis and Field Experimentation Approaches

	Econometric Analysis	Field Experimentation
Type of Data	Observational; service and enforcement activities not under the control of the researcher	Experimental; service and enforcement activities largely under the control of the researcher
Method of controlling for confounding factors	Controls statistically (after the fact) using data on those factors	Relies heavily on the similarity between the treatment and control groups, which are designed in advance of the experiment; econometric methods can be used afterwards to address possible dissimilarities

A detailed review of each of the proposed econometric and field experimentation methodologies is provided below. Each of the approaches has some potential to generate reasonably sound estimates of the extent to which service and



enforcement activities impact compliance. However, as currently constructed, none of the methodologies are explicitly linked to a theoretical model of taxpayer compliance behavior, which limits their ability to inform our understanding of why compliance behavior responds to changes in IRS activities. We note that during the workshop and in subsequent follow-up telephone discussions, several experts suggested that empirical research designs be developed for implementing and testing theoretical models of compliance behavior that incorporate social and psychological factors reflected in taxpayer attitudes, perceptions, and motivations, such as those described in Part I Section 1 Theoretical Frameworks (page 12). Possible research designs for this purpose are not included in this section, because no specific designs were proposed. However, the structure of the integrated plan presented in Part III accommodates the development of new empirical methodologies for studying compliance impacts as well as the enhancement of the existing proposed methodologies. As part of the review of the proposed methodologies, we consider their potential for incorporating social and psychological factors. In addition, the data sources recommended for implementation of several of the methodologies, particularly agent-based modeling and segmentation analysis, would facilitate the inclusion of such factors.

2.1 Aggregate Econometric Modeling

2.1.1 Description/Assumptions/Research Questions Addressed

Several past research efforts to investigate the overall impact of tax enforcement activities on tax compliance have applied panel data estimation techniques to a time series of observations on tax reporting and filing behavior aggregated by state. This includes the studies by Dubin, Graetz, and Wilde (1988), Plumley (1996), and Dubin (2007). During the workshop, an extension of this methodology was outlined for investigating the impact of a variety of different taxpayer services on compliance.

A key feature of this approach is that it does not directly attempt to explain compliance behavior. Rather, it employs regression techniques to explain variations across states and over time in such aggregates as reported taxable income and the percentage of households that file tax returns. A variety of control variables are incorporated in the analysis to account for variations in these aggregates that are attributable to geographical and cross-year differences in actual tax reporting and filing requirements. To the extent that such controls are effective, the estimated impact of audit activities on reporting and filing behavior should represent their marginal contribution to compliance.

In the first application of this methodology, Dubin, Graetz and Wilde (1990) (henceforth referred to as DGW) examined differences in filing and reporting activity across all fifty states for the years 1977 to 1986. The authors relied mostly on IRS administrative data contained in the Commissioner's Report for the relevant years to calculate three state level measures of filing and tax liability: (1) returns filed per capita; (2) assessed tax liability per return; and (3) reported tax liability per return. Regression equations were used to explain the variation in these measures across states and over time as a function of variations in audit rates, state income tax rates, per capita income levels, and other socio-economic factors such as age, education, occupation, and the unemployment rate. The specification involving returns filed per capita was employed to estimate the influence of audits on filing behavior. The specification involving reported tax liability per return was used to assess the marginal indirect effect of audits on reporting compliance; i.e., the change in the amount of tax liability reported as a result of a change in the risk of being audited. Finally, the specification involving assessed tax liability (defined as the sum of reported tax liability and any additional taxes assessed during audits) was employed to estimate the marginal combined direct and indirect effect of audits on reporting compliance. The difference between this estimate and the aforementioned estimate of the marginal indirect effect would then represent the marginal direct effect of audits.

In their specification, DGW encountered two important econometric issues. First, they recognized that the state level audit rate was likely to be endogenously determined with state level filing and reporting behavior; all else equal, it seems likely that the IRS would allocate resources so that audit rates would tend to be higher in those states where compliance is relatively poor. Second, they realized that even after controlling for a variety of state-level characteristics, their measures of state-level filing and reporting behavior may be influenced by unmeasured state-specific factors. Under these circumstances, an ordinary least squares regression would yield biased and inconsistent estimates. To address these issues, the authors employed an instrumental variables-generalized least squares estimation procedure. Under this procedure, instrumental variables – variables that are correlated with the audit rate but assumed to be uncorrelated with the error term of the regression model – are used to control for the endogeneity of audit rates, and generalized least squares procedures are employed to account for state level heterogeneity, which is modeled as a time-invariant source of random error, known as a "random effect".

Plumley (1996) extended this approach by using more detailed measures of compliance outcomes, a richer set of explanatory variables to account for variations in reporting and filing requirements, a more accurate measure of audit rates, and some additional enforcement variables relating to information return matching, nonfiler notices, refund offsets, and criminal enforcement activities. Data were pooled for 49 states over the years 1982 to 1991. One filing equation and three reporting equations were estimated. The dependent variables in the latter equations were aggregated state level measures based on (1) total income reported; (2) total offsets to income; and (3) net income (defined as total income minus offsets). In an effort to account for changes over time in reported income amounts that resulted from legal changes in reporting requirements, Plumley experimented with two alternative approaches. One approach was to employ an extended set of explanatory variables to control for changes in legal requirements. The other was to employ constant-law measures of the three income sources.

To address the unobserved sources of heterogeneity across states, Plumley used a somewhat different approach than DGW. In particular, he employed a fixed effects specification under which a distinct time-invariant intercept term was estimated for each state rather than the random-effects specification used in the former study. Like DGW, Plumley also employed an instrumental variables procedure to control for the potential endogeneity of audit rates. However, he used what were arguably improved instruments based on the portion of the total time available to auditors that is devoted to direct examinations.

This comparison of the two studies illustrates the fact that researchers face a variety of different options for how to address the econometric issues of state-level heterogeneity and endogenous regressors when performing such analysis. Decisions regarding these issues are important as they are ultimately likely to influence the estimates that are achieved. Further work is therefore desirable to explore a range of options for how best to address these and other econometric issues when extending the methodology to account for the role of taxpayer services in compliance.

As a first step in extending the methodology to account for the role of taxpayer services, Plumley incorporated additional explanatory variables in his analysis relating to the volumes of telephone and return preparation assistance provided by the IRS. Intuitively, his measures of these services were intended to assess the extent to which filing and reporting behavior varied across states and over time in response to changes in service utilization. A difficulty with this approach, however, is that taxpayers are free to choose whether to use a given service. If unobserved factors that influence the likelihood that taxpayers will use a service (such as confusion about the tax law) are correlated with compliance behavior, one can obtain misleading inferences about causal effects – a problem referred to in the econometric literature as either “self-selection bias” or “sample selection bias”. Further econometric development of the methodology is warranted to address this issue.

A potentially important issue with regard to enforcement and service impacts is that they may occur with a lag. Dubin (2007) updated the earlier DGW study and extended it to account for criminal enforcement and information return matching activities.⁷ Although the quantitative estimates reported in his paper do not appear to account for a lagged behavioral response, Dubin does describe some findings on this issue based on a dynamic econometric specification of reporting behavior. Specifically, he reports having found evidence of a lagged effect for enforcement, with the majority of the compliance adjustment occurring within two to three years of the change in activity levels. The ability to account for lagged responses represents a potentially important advantage of using a panel data approach.

2.1.2 Strengths/Weaknesses/Limitations

The proposed aggregate econometric methodology is meant to directly address the fundamental research issue at the heart of this project; namely, the direct and indirect effects of IRS services on tax compliance. A positive feature of this approach is that it has been used in prior research to examine enforcement effects. Consequently, many of the challenges and limitations of the methodology are reasonably well understood, and researchers have made at least some progress in addressing them. As the approach relies on panel data, the methodology has the advantage that it may be able to account for a lagged response of compliance behavior to changes in IRS activities.

Another advantage of the approach is that much of the data needed to implement the methodology is already available. One key exception is the availability of historical detail on the main IRS service activities. It is generally agreed that a longitudinal database spanning at least 10 years would be necessary to produce meaningful results. Given that it does not appear to be possible to reconstruct the necessary service data for prior years, this means that the data would need to be collected moving forward, so that it will take a number of years before an adequate time series has been collected. In the meantime, immediate efforts should be directed towards data collection and retention and, perhaps, preliminary model development and specification testing.

⁷ The updated paper incorporates some new instrumental variables, including a measure similar to the direct examination time measures used by Plumley (1996) and a variable proposed by Mete (2002), based on political economy considerations.

As currently specified, the methodology is not explicitly linked to a theoretical framework of how services impact compliance. Consequently, while the methodology has the potential to measure the magnitude of the impact of services, it is unlikely to generate insights into the reasons underlying their impact. In principle, it may be possible to adapt the methodology to explore the theoretical connections between services and tax compliance. For instance, variables accounting for differences in social and psychological factors across geographic areas might be incorporated in the specification. However, the aggregate nature of the analysis may impose limitations on the extent to which relevant theoretical issues can be addressed.

One important modeling issue that may be informed by theoretical analysis is the appropriate categories of taxpayers for aggregation. In past applications of the methodology, taxpayers have been aggregated into groups according to their state of residence. However, it is not clear whether this is the best choice for studying service and enforcement effects. To a large extent, the indirect effects of these activities are likely to depend on the nature of information flows and interactions among taxpayers. While it seems intuitive that at least some of these information flows and interactions are likely to occur between taxpayers residing within a common geographic area, such as a state or locality, there is also significant potential for these to occur across borders via media and Internet communications, national networks of professionals and industries, and interactions among family members residing in different locations. Therefore, an analysis that relies on data aggregated by state may not be able to properly account for the spillover effects of service and enforcement activities in one state on taxpayers in other jurisdictions. While theoretical research may provide a guide for the best way to group taxpayers, ultimately the choice of grouping may be constrained by data collection limitations.

Views among the workshop participants varied regarding the potential reliability of the approach. Some researchers felt that the procedure was likely to be quite reliable, but others felt that independent validation would be needed to demonstrate that the results of such an approach were credible. Past applications of the methodology have produced very large and controversial estimates of the general deterrent effect of enforcement.

2.1.3 Likelihood of Success/Risks

If implemented, the likelihood that this approach would provide reliable estimates of the direct and indirect effects of IRS services on compliance is uncertain. In order to be successful, the methodology would need to overcome some very significant econometric challenges, including finding a way to address sample selection bias associated with service utilization, identifying credible and effective instruments to account for the likely endogeneity of enforcement activity, addressing spillover effects and potential lags in behavioral responses, and adequately controlling for potentially confounding effects. While progress has been made on some of these issues, further work is needed. In terms of data needs, implementation of the methodology is contingent on obtaining and retaining accurate and timely information on the use and distribution of IRS services.

2.1.4 Expectations (Short-Term and Long-Term)

In the first few years, the short-term benefits from implementing this methodology include new and important data on the use and distribution of IRS services are collected and analyzed. These early benefits would include descriptive statistics on trends in tax reporting, filing, audit coverage, service utilization and service quality by state and geographic region.

In the longer term, estimates of the direct and indirect marginal impact of taxpayer enforcement and service utilization rates would be obtained. The long-term expectation is that it would take approximately 10 years to realize the benefits of this methodology, since the approach requires sufficient variation in independent variables. Yet progress in refining and adapting the methodology to account for service impacts to produce preliminary estimation results would likely be expected within 5 years.

2.1.5 Dependencies

Preliminary work in identifying important variables that would be used as instruments in controlling for the endogeneity of taxpayer audits could begin immediately, as could efforts to develop an appropriate econometric strategy to address sample selection bias and other challenges of incorporating service utilization variables into the existing framework. On the data side, it would be important in the near future to identify those IRS services where data on service utilization are readily available and where additional data collection efforts should be directed. A

useful starting point would be to review the econometric specifications of Dubin (2007) and Plumley (1996) to ascertain which additional data collection efforts are needed. Results from theoretical models and segmentation analysis would be useful in assessing how best to aggregate taxpayers into groups and possible ways to extend the methodology to generate insights into the underlying reasons that service and enforcement activities impact compliance.

2.1.6 Data Needs

Implementation of the proposed methodology would require a large data collection effort, across many different data sources. Prior research has proved useful in identifying the types of data that would be required in this effort.

A useful framework for elements of the econometric specification is contained in Plumley (1996). He identifies those data elements related to (i) tax policy, such as filing thresholds, marginal tax rates, state income taxes and child exemptions; (ii) Tax burden and opportunity for evasion such as the number of sole proprietorships and average tax burden; (iii) IRS enforcement including audit coverage, number of information returns and conviction rate; (iv) IRS responsiveness (i.e., services) such as telephone assistance and return preparation; and (v) Demographic and socio-economic variables such as marital status, age, births, and personal income.

Additional consideration would be given as to what new types of data on IRS services and utilization of these services would be required. Most likely, data collected for implementing this methodology would combine information from both cross sectional and time series components. That is, the data structure would represent a panel of observations across time and geographical region for each data element identified as being an input into the econometric specification. Data would ultimately be aggregated to a geographic region, such as the state level, for the principal econometric approach but should be collected and retained at the lowest level of aggregation possible.⁸

While comprehensive, the data requirements for implementing this methodology are relatively modest and do not require the use of confidential taxpayer information. Therefore, there is no reason that these data cannot be retained indefinitely.

In addition to IRS administrative data, potentially augmented by survey data to assess service quality and taxpayer attitudes, state level income and other related economic data is required. This could be collected from the Department of Commerce, Bureau of Economic Analysis (BEA). Demographic and socioeconomic data could be obtained from the Bureau of the Census in most cases, or from relevant state agencies.

2.2 Microeconomic Modeling – Estimating the Impact on Reporting Compliance

During the workshop, an outline of a methodology was proposed for explaining the marginal impacts of both service and enforcement activities on tax reporting compliance by individual filers of tax returns. A description of this approach is provided below.

2.2.1 Description/Assumptions/Research Questions Addressed

Under the basic approach proposed at the workshop, one estimates probabilities of enforcement action and service utilization using appropriate data sources and then imputes these measures onto the National Research Program (NRP) data base to support noncompliance analysis. This approach involves four main methodological steps. After describing these steps, we consider various possible extensions of the approach.

Step 1: Estimate Probability of an Operational Audit

The first step is to estimate the likelihood of an operational audit as a function of taxpayer characteristics using a sample of audited and unaudited tax returns. Sample weights would be included to make the returns broadly representative of all audited and unaudited returns in the population. Since audit outcomes are binary (audited or not

⁸ For example, to the extent that IRS administrative data are identified by zip code, this level of aggregation should be maintained, and subsequent aggregation to the state or regional level could be accomplished as needed.

audited), ordinary regression approaches for continuous dependent variables are not appropriate. Instead, one should employ an econometric methodology for qualitative dependent variables. A common choice is a probit specification:

$$A_i^* = \gamma_A' x_{Ai} + u_{Ai},$$

where A_i^* is a latent variable representing the propensity for the return of taxpayer i to be selected for an audit; an audit occurs only if A_i^* is greater than zero.⁹ The term x_{Ai} represents a vector of taxpayer i 's characteristics, γ_A is a vector of coefficients to be estimated, and u_{Ai} represents a standard normal random disturbance term. Such a model can be estimated using the method of maximum likelihood. The resulting parameter estimates can be used to predict, for any given return in the population, the probability of operational audit selection.

Step 2: Estimate Probabilities of Taxpayer Service Utilization

The second step is to estimate a model for the demand for relevant taxpayer services as a function of taxpayer characteristics using data from a representative survey on taxpayer service utilization. Typically, the decision to use a given taxpayer service is a discrete decision; either the taxpayer chooses to use the service, or he does not. Therefore, one approach to estimating the demand for each service would be to estimate a separate probit specification for each distinct service. However, it seems likely that many services are substitutes for one another. In such a case, it might be more appropriate to employ a joint estimation procedure for related services. A suitable statistical model for joint estimation of the likelihood of using different services would need to be developed in this case. The explanatory variables in an econometric model of taxpayer service utilization would include various taxpayer characteristics. They also might include measures of the quality of the services for the geographic area in which the taxpayer resides (accessibility, accuracy, etc.). The resulting parameter estimates could be used to estimate, for any given taxpayer in the population, the probability that a given service would be employed by a taxpayer.

Step 3: Impute Audit and Service Utilization Probabilities to NRP Returns

The third step is to employ the results from the first two steps to impute to each return in the NRP data base, the probability that the return would be selected for an audit and, in the case of each service, the probability that the taxpayer filing the return would employ the service.

Step 4: Estimate Model of Noncompliance Using Augmented NRP Data

Lastly, a model of noncompliance is estimated using the NRP data base, complete with the additional imputed variables. As a starting point, consider the following regression specification:

$$N_i = \beta_A PA_i + \beta_S' PS_i + \beta_N' x_{Ni} + \varepsilon_{Ni},$$

where N_i is a latent variable for the propensity for noncompliance on return i , PA_i represents the predicted probability of an operational audit of return i , PS_i represents the vector of predicted probabilities of service utilization for the various types of services (which may include both IRS-provided and substitute services such as assistance from tax practitioners), x_{Ni} represents a vector of additional tax return characteristics, and ε_{Ni} is a random disturbance term.¹⁰ The measure of noncompliance might be based on the NRP audit adjustment to overall tax liability on the return. Such a model would at least partially capture the marginal general deterrent effect of an audit

⁹ As an alternative, one could employ a logit specification or another appropriate specification for a binary dependent variable. For more details on econometric estimation with qualitative dependent variables, refer to G.S. Maddala (1983).

¹⁰ If desired, one could employ additional imputed variables representing the likelihood of other enforcement actions, such as matching program adjustments or correspondence examinations.

through the coefficient β_A of the audit risk variable PA_i . It would also capture how the likelihood of using various services would impact compliance through the estimated coefficients of vector β_S .

Extensions of the Basic Approach

The basic specification described above might not fully account for the general deterrent effect of taxpayer enforcement and service activities. In the case of enforcement activities, taxpayer perceptions of audit risk may differ from the objective measure of risk used in that specification. In addition, a taxpayer's compliance behavior may be influenced by the extent to which others are being audited for reasons independent of his own audit risk. For instance, a relatively high overall audit rate may reassure the taxpayer that others are being made to pay their fair share, thereby making the taxpayer more willing to comply voluntarily. Similarly, a taxpayer service may impact one's compliance behavior even when one is unlikely to make direct use of the service. Consider, for example, a taxpayer who is unlikely to use a service himself, but who learns about reporting requirements from another taxpayer who does employ the service. The impact of such interpersonal communication on the first taxpayer's reporting behavior is not accounted for by the specification. Similarly, perceptions about the availability and quality of taxpayer services may influence attitudes about compliance, even among taxpayers who are unlikely use the services. Again, the current specification is unable to account for such indirect effects.

To address this issue, one might expand the basic model so that it has a more explicit link to theories of how enforcement and service activities impact compliance. One way to accomplish this objective is to incorporate "social interaction effects". For instance, one might assume that taxpayers are members of groups defined on the basis of their geographic location or their occupation. The behavior of a taxpayer in a given group may be hypothesized to depend on the behavior of other members of the group. To address this possibility, one could include some additional group level regressors in the model, such as the percentages of group members who use various services, the average audit risk within the group, and the average level of noncompliance within the group. Such a specification could more fully account for indirect effects of taxpayer enforcement and service activities on compliance behavior. Consider the above example where a taxpayer who was unlikely to use a given service learned about reporting requirements from another taxpayer who did use it. Presumably, the likelihood of such an interaction would depend on the extent to which other members of the taxpayer's group elected to use the service. By including the percentage of group members using the service as an additional explanatory variable in the specification, such an effect could be accounted for. The level of service usage among the members of the group presumably would also function as an indicator of the quality and availability of services, since such factors would influence the level of service usage. Therefore, to the extent that the quality and availability of services impact tax compliance by influencing taxpayer attitudes, this also would be accounted for in the specification.

The inclusion of explanatory variables describing the average audit risk within the group and the average level of noncompliance within the group would more fully account for the impact of enforcement activity and peer group influences on compliance. The former variable would account for the possibility that one's perceived risk of audit as well as one's willingness to comply may depend on the audit experiences of others within the group. The latter variable would address the possibility that members of the group attempt to conform their own compliance behavior to how other members of the group are behaving. This would allow another pathway for taxpayer services to have an indirect effect on tax compliance. For instance, suppose that some members of a group become more compliant as a result of certain taxpayer services they have received. This would improve average compliance within the group, which in turn might lead to improved compliance by other members as they attempt to conform their behavior to that of the group.

An excellent review of models with social interaction effects is provided by Blume and Durlauf (2005). Such models require careful specification to ensure identification, or in other words, to ensure that unique values can be estimated for each of the structural parameters of the model. Of particular concern is ensuring that the specification can adequately distinguish between individual effects (e.g., own service utilization) and group effects (e.g., average group service utilization) on behavior.

As with the aforementioned aggregate econometric methodology, another challenge for this approach is determining how to divide taxpayers into relevant groups. In the above discussion, we suggested the possibility of using geographic location or occupation as a classification criterion, but relatively little is known about how best to

identify and define the social networks used by different sets of taxpayers for their tax-related decision making. This is an important area for additional research.

Another concern with the above specification is that the formulae for the predicted probabilities of audit and service utilization are a nonlinear function of taxpayer characteristics. If these same characteristics are included as separate explanatory variables in the regression specification, the predicted probabilities might be highly correlated with these variables, a problem known as a high degree of multicollinearity. Under such circumstances, it can be difficult to obtain reasonably precise estimates of the parameters of the model. To better identify the impact of service and enforcement activities on compliance, it would be desirable to include at least one explanatory variable in the specification for each service utilization and enforcement probability that is not included in the regression specification for noncompliance. For a given service activity, an example of such an explanatory variable would be a measure of service accessibility within the taxpayer's geographic area.

An alternative to using imputed probabilities of service utilization would be to employ direct measures of whether the different services were used. Beginning with tax year 2006 returns, NRP examiners have begun asking taxpayers which of the following services they have used:

- Forms or instruction booklets
- Tax assistance center (walk-in site)
- IRS phone representative
- Automated phone system (Teletax)
- E-mail correspondence with the IRS
- Written correspondence with the IRS
- IRS Free File website or software

The taxpayers are also asked an open-ended question about other IRS services they may have used. If a taxpayer reports having contacted the IRS, including via the internet, for assistance, the taxpayer is asked to identify which of the following topics prompted the contact:

- Taxability of income
- Earned Income Credit
- Other tax credits
- Exemptions
- Filing status
- Charitable contributions
- Interest income/expense
- Capital gains/losses
- Deductions
- Pension/IRA/social security benefits
- Small business income/expenses
- State income tax related issues
- Other issues

Given that the taxpayers being surveyed are in the process of being audited, there is some question as to how completely and accurately they will respond to these questions. It therefore would be useful to compare their responses to those of a representative sample of taxpayers from a separate survey to see if they are reasonably consistent. Examples of separate surveys that might be used as a source of comparison include the benchmark Opinion Survey of Taxpayer Resources and Services and the 2006 W&I Market Segment Survey. These surveys pertain to tax year 2005 rather than 2006, but they should still be reasonably informative with regard to the issue.

Assuming that the responses to the NRP questions on service utilization are credible, one could use them as direct measures of service utilization in the regression model for noncompliance in place of the predicted service utilization probabilities. Similarly, the NRP contains details on whether a tax practitioner assisted with the return and, if so, the type of practitioner. Such information could be used to account for usage of certain substitute services. To account for social interaction effects, one could also construct explanatory variables from the NRP of group level utilization of services and average group level audit risk.

A statistical issue associated with including explanatory variables relating to actual service utilization is the problem of sample selection. In particular, individuals are not randomly assigned to use different services but instead make their own decisions. As a result, taxpayers who choose to use a given service may differ in unobserved ways from taxpayers who choose not to use the service. If these unobserved factors are correlated with compliance behavior, an ordinary regression of compliance on service utilization will tend to produce biased estimates of the impact of

services on compliance. There are econometric methods to address these sample selection issues by carefully modeling the decision to employ services.¹¹ As with the multicollinearity issue mentioned previously with regard to social interaction effects, it is desirable when using such methods to have at least one explanatory variable for service utilization that is excluded from the regression specification for noncompliance.

An additional issue with the above regression specification for noncompliance is that a non-trivial number of taxpayers are fully compliant with their tax reporting obligations. This raises a statistical issue known as a “censored dependent variable” problem. Under circumstances where there is a substantial mass of observations for which the dependent variable in a regression take the value zero (in this case representing perfect compliance), ordinary regression results will generally be biased and inconsistent. Typically, censoring is addressed by using somewhat more complex estimation methods, such as tobit estimation.¹² Such methods are feasible, but they are not very robust to specification errors. For instance, one can obtain misleading results if one makes an incorrect assumption about the distribution of the error term in the regression specification. Under a tobit specification, one assumes that the error term is normally distributed. In the case of tax noncompliance, which tends to have a very skewed distribution under which a relatively small number of returns underreport taxes by very large amounts, an alternative distribution such as the lognormal would seem more appropriate. Thus, some care would be needed in developing an appropriate specification to address the censoring issue.

2.2.2 Strengths/Weaknesses/Limitations

Assuming social interaction effects were taken into account, the above approach would have the potential to address the fundamental research question; namely, the marginal deterrent effects of various service and enforcement activities on overall taxpayer reporting compliance.

As noted above, there are several econometric challenges that would need to be addressed, including issues of identification, multicollinearity, censoring, sample selection, and segmentation designed to capture social interaction effects. Finally, in a single cross section of data, it would not be feasible to account for lagged effects. However, it might be possible to address such effects if additional information could be collected about the prior year characteristics of taxpayers.

2.2.3 Likelihood of Success/Risks

If implemented, the likelihood that this approach will provide reliable estimates of the direct and indirect effects of IRS services on compliance is uncertain. As with any complex econometric estimation methodology, there is the potential for model misspecification and, consequently, misleading results. To help to mitigate this risk, the assumptions underlying the specification would need to be carefully tested, and appropriate refinements of the methodology would need to be undertaken. As with all of the methodologies described in this report, the results are likely to be viewed as more credible if they are consistent with results from independent estimation methodologies based on alternative sets of assumptions.

2.2.4 Expectations (Short-Term and Long-Term)

In the course of the first year or two, it should be possible to estimate models of the demand for alternative taxpayer services and the risk of enforcement action. The precise timing depends on the availability of appropriate data. If existing surveys on taxpayer service usage are suitable, estimation of service demand could begin quite soon. Otherwise, it would have to wait until results of an appropriate survey became available. Also in the short term, work could begin on developing the outlined methodology into a full-fledged econometric framework for estimating behavioral responses with the NRP data.

Over the next couple of years, the econometric framework could be applied, tested, and refined using NRP results. Over the longer term, the results could be updated as new NRP studies and surveys become available.

¹¹ For a survey of such methods, refer to Maddala (1983).

¹² Maddala (1983) provides a helpful discussion of censoring issues.

2.2.5 Dependencies

In order to impute the probabilities of service utilization to NRP returns, it would be necessary to first perform an econometric analysis of the demand for taxpayer services using survey data. Similarly, to impute the likelihood of an operational audit to NRP returns, it would be necessary to first perform an econometric analysis of audit risk using a statistically representative sample of audited and unaudited returns. Under the alternative approach, direct measures of service utilization would be obtained from the relevant questions now being asked of NRP taxpayers. Prior to undertaking this approach, though, it would be advisable to examine whether the responses of the NRP taxpayers are consistent with those provided by a statistically representative group of taxpayers on an independent survey. Results from theoretical models and segmentation analysis would be useful in guiding the specification of social interaction effects, including the assignment of social groupings.

2.2.6 Data Needs

This approach would make use of individual level NRP data, augmented by measures of the probability of audit and taxpayer service utilization, as it becomes available. Estimation could be based on a cross-section of taxpayers from a single tax year, or from pooled cross-sections of taxpayers drawn from multiple tax years.¹³ With multiple years of data, there would likely be greater variation in service and enforcement activities, which may allow improved identification of their marginal effects. It would also be desirable to match the NRP data on taxpayers to their prior year returns to aid in extending the framework to account for lagged effects.

In order to impute the probability of audit to NRP returns, the approach would require individual level data on representative samples of audited and unaudited returns from comparable tax years to those in the NRP. This data would include a set of sample weights to make the sample representative of the overall taxpayer population, line item tax return information, and an indicator for whether a return was subjected to an operational audit.

One of the options under this approach involves using measures of actual service utilization among taxpayers selected for the NRP. Such information could be obtained from the questions currently being asked of taxpayers during NRP examinations. An alternative approach to obtaining this information would be to pre-select taxpayers for the NRP and then to include at least a random sample of those pre-selected in an appropriate survey on taxpayer service utilization prior to sending out NRP audit notices.

Instead of relying on information on actual service utilization, a second option under this approach is to impute measures of the likelihood of taxpayers using various services to the NRP. These measures would be based on estimation results from a model of taxpayer service utilization based on an appropriate survey of a representative sample of taxpayers from the overall population. The data should include a set of sample weights to make the sample representative of the overall taxpayer population, line item tax return information and indicators for which types of taxpayer services (including both those provided by the IRS and substitute services provided by third-parties) were used. It would be desirable to augment this data base with statistics on service quality (accessibility, accuracy, utilization rates, etc.), broken down by geographic region. As noted previously, some NRP service utilization surveys were conducted for tax year 2005, which might prove to be useful for this purpose.

2.3 Microeconomic Modeling – Estimating the Impact on Filing Compliance

There was some discussion during the workshop of ways to obtain a representative data base containing micro-level information on nonfilers of individual income tax returns. If such a data base could be obtained, it would be feasible to perform a microeconomic analysis of the impact of services on filing behavior.

2.3.1 Description/Assumptions/Research Questions Addressed

Most of the discussion during the workshop was focused on collection of a suitable data base rather than estimation methodologies. However, one approach that might be used is to perform a qualitative dependent variable analysis of

¹³ Note that the cross-sectional data, even if combined for several years, would not constitute a longitudinal database, because different taxpayers would be audited in each year.

the likelihood of filing a required tax return as a function of measures of taxpayer service quality and other factors. Consider, for example, the following probit specification:

$$F_i^* = \gamma_S' Q_i + \gamma_F' x_{Fi} + u_{Fi},$$

where F_i^* is a latent variable representing the propensity for taxpayer i to file a return; a return is filed only if F_i^* is greater than zero.¹⁴ The term Q_i represents a vector of variables describing the qualities of taxpayer services available to individual i . x_{Fi} represents a vector of individual i 's characteristics (ideally including whether the individual has filed a return in past years), γ_S and γ_F are vectors of coefficients to be estimated, and u_{Fi} represents a standard normal random disturbance term. Measures of service quality might include perceived or actual measures of such attributes as accessibility and accuracy within the individual's geographic location. If available, one might also include some explanatory variables describing various taxpayer attitudes about government, the tax agency, and the use of government resources.

Such a model could be estimated using cross-sectional data from a single tax year, from a time series of cross-sections from multiple years, or from longitudinal data. The results would provide an indication of how an actual or perceived change in service quality would impact filing behavior.

2.3.2 Strengths/Weaknesses/Limitations

This approach has potential to address a fundamental research question - namely, the marginal impact of the qualities of different services, such as accessibility and accuracy, upon the likelihood that an individual will comply with return filing requirements. Application of methodology is contingent on the collection of appropriate data. Although this approach would account for some of the key drivers of individual filing behavior, it may not fully capture the influence of social factors on filing decisions. The feasibility of extending the framework to address social interaction effects would need to be examined. If a single year of cross-sectional data were used in estimation, it would not be feasible to account for lagged effects. However, it might be possible to address such effects if additional information could be collected about the prior year characteristics of the respondents.

2.3.3 Likelihood of Success/Risks

If implemented, the likelihood that this approach will provide reliable estimates of the direct and indirect effects of IRS services on filing compliance is uncertain. An important issue is that the data sample would need to have sufficient variation in measures of service quality to adequately capture their marginal impacts. It would also be important to examine the robustness of the results to alternative specifications. The likelihood of success would be improved if arrangements could be made with Census so that an econometric analysis could be undertaken of relevant Current Population Survey data that has been matched against filed tax returns.

2.3.4 Expectations (Short-Term and Long-Term)

Once suitable data are available, it should be possible to obtain estimates of service impacts on filing behavior within a year. In addition to learning about service impacts, the data base would provide an opportunity to learn about the general characteristics of nonfilers and factors other than service quality that impact one's filing decision.

2.3.5 Dependencies

The nonfiler project should be able to be instituted independently without dependencies on other projects.

¹⁴ As an alternative, one could employ a logit specification or another appropriate specification for a binary dependent variable. For more details on econometric estimation with qualitative dependent variables, refer to G.S. Maddala (1983).

2.3.6 Data Needs

Data would need to be collected for a representative sample of filers and nonfilers of individual tax returns. The data should include information on a broad set of socio-economic characteristics, ideally with detailed information on sources and levels of income and potential tax liability (so that it would be possible to assess whether there is a filing requirement). Other desirable data are indicators for whether the taxpayer filed a return for the current tax year and survey responses to a variety of questions on attitudes and perceptions about taxpayer services, the tax agency, and government in general. Sample weights should be available to make the sample representative of the overall population.

2.4 Field Experiments and Quasi/Natural Experiments

2.4.1 Description/Assumptions/Research Questions Addressed

Field experiments use treatment and control groups to measure the effects of IRS activities. These real world experiments have the potential to provide the most credible estimates of the impact of IRS activities, but they are not always practical to implement, and it is not always possible to fully isolate the effect of IRS activities from other potential influences.

Field experiments create treatment and control groups (groups could differ by factors such as geography or time and would ideally be randomly assigned), vary IRS activities by group, and then observe impacts on compliance. There have been a number of field studies that have focused on the impact of enforcement activities on compliance. In two recent studies (Slemrod, Blumenthal, and Christian, 2001; Hasseldine et al., 2007) the authors designed controlled field experiments to assess the general deterrent effect of tax audits. While there has been substantially less work done using field experiments for services, an interesting recent study (Adelsheim and Zanetti, 2006) examined the impact of various targeted outreach programs that were designed to educate a category of targeted small business owners about their requirements to file particular types of tax information. The outreach programs were delivered by mass media to selected geographic locations, while similar geographic locations outside of the delivery area were used as what is best referred to as a “quasi-control group”.

When using a field experiment to assess the impact of a change in an IRS activity, such as an improvement in service quality on tax compliance, one needs a way to measure the compliance response. A direct approach would be to conduct an audit of each taxpayer at the end of the experiment. The difference in the average audit adjustment amounts between the treatment group (which received the improved level of service) and the control group (which received the standard level of service) could then be used as a measure of the compliance impact. However, auditing is both costly and intrusive, so it may be desirable to employ a less direct method of inference. One alternative is to use the difference in the average reported tax liabilities of the treatment and control groups. Like the difference in average audit adjustment amounts, this statistic would serve as a valid measure of the compliance impact if, outside of the treatment under study, there was no systematic difference in the distribution of factors relevant to tax compliance behavior between the treatment and control groups.

In practice, however, the treatment and control groups may not be sufficiently comparable. For instance, if they are drawn from different geographic areas, they may be subject to different local influences. A difference-in-differences methodology can be used to control for differences between the treatment and control groups that are time-invariant. Under this approach, one subtracts the difference between the average measured outcomes (audit results or reported tax liabilities) for the treatment and control groups in the pre-treatment period from the difference in average measured outcomes in the post-treatment period. If differences in relevant characteristics are observed between the treatment and control groups, a more general regression approach can be employed that includes explanatory variables to account for these observed differences as well as any unmeasured time-invariant differences across groups. Such an approach represents a hybrid between a conventional experimental analysis and an econometric analysis based on observational data.

The kinds of effects that one can measure using a field experiment depend critically on the nature of the design. Under a single factorial design, one can examine the impact of varying a single element of a service (e.g., accessibility) at one time between the treatment and control groups. This yields a measure of the marginal impact of changing that element of the service, holding all other elements of the service constant. Under a multi-factorial design, one can examine the impact of varying two or more elements of the service (e.g., accessibility and accuracy)

at the same time. This can be important when there are potentially important interactions among the elements. So, for instance, a multi-factorial design could account for the possibility that improving the accessibility of an informational service may have a larger impact on compliance when the accuracy of information provided is relatively high than when the accuracy is relatively low. The benefits of a more complex experimental design of this sort come at a price. In particular, more complex designs tend to be more difficult and costly to implement.

Quasi/Natural experiments are closely related to designed field experiments except that they are discovered as happening naturally in the normal course of the IRS's business as opposed to being specifically designed. Discovering a situation in which one taxpayer group was exposed to an IRS activity while, for some reason, another similar taxpayer group was not exposed, could be a rare occurring situation. An example is a recent IRS initiative in which some taxpayers are being sent a new kind of warning letter (CP 2057) that instructs recipients that they may be underreporting, that they should double-check certain portions of their return, and that they should file an amended return if any errors are identified. This differs from the traditional CP 2000 warning letter in that it does not propose a specific amount of additional tax that may be owed. If returns of taxpayers receiving the CP 2057 notice can be matched against returns for a comparable group of taxpayers receiving a CP 2000 notice, it may be possible to estimate the compliance impact of not specifying a proposed adjustment amount on a warning letter.¹⁵

Field experiments could address the primary research question by providing estimates of the impact of selected IRS activities on tax compliance. However it may not be feasible to implement such experiments for all types of service and enforcement activities owing to either design or cost limitations.

2.4.2 Strengths/Weaknesses/Limitations

Field experiments are an appealing option for their real world impact and ability to address the primary research question in a relatively short time period. However, depending on the experiment, the IRS may encounter practical limitations with regard to the extent to which it can control key variables, identify comparable treatment and control groups, address potentially confounding factors, and generalize the findings to other time periods, groups, or activities.

The main strength of field experiments, as articulated by the workshop breakout group on experimentation, is that they measure "real world" impact. When one measures a difference in compliance behavior between the treatment and control groups over a specific period of time, one observes the empirical results of a set of real world factors that were in place over that period. As discussed below, there may be confounding factors. However, such experiments make clear that some factor or factors caused a real world occurrence.

Another appealing feature of field experiments is that they have a reasonable likelihood of producing such estimates (possibly for limited segments of the taxpayer population) within a year or two of the start of their planning. Because experiments could require the creation and implementation of a new or modified service and collection of data over an extended period of time, it is possible that such experiments may take several months to more than a year to complete. Therefore, this method could produce estimates of the impact of services on compliance in a relatively short timeframe, at least as compared to the other primary methods requiring substantial data acquisition such as the econometric methods.¹⁶

Field experiments, however, have technical limitations. Ultimately, some IRS activities may be difficult or impossible to test in field experiments. The testability of the type of service is related to how much "control" is possible with taxpayers. Depending on the service, the IRS may know only that the taxpayers was exposed to the service or had the opportunity to partake of that service, but not that he or she actually did. In such cases, it should still be possible to measure the overall impact of making a given service available on compliance within a

¹⁵ This potential natural experiment was suggested by Charles Christian in his Point of View paper, which is included in Appendix D: Point of View Papers.

¹⁶ As is discussed in the section on econometrics, it would be a period of several years before enough data is available to produce econometric estimates of the direct and indirect effects of services on compliance. In contrast, field experiments are expected to take longer than the supporting methodology of laboratory experiments.

population, but not its direct impact on service recipients. Also, it is not possible to know all of the activities that the set of taxpayers of interest took advantage of in addition using the service that was tested. In such cases, it is difficult to perform a field test.

Additionally, there are two main reasons why it is difficult to generalize from field experiments. First, the impact of a field experiment would depend on the set of conditions and options available to an individual at a point in time. So, for example, decisions whether to use an IRS service would depend on what alternative services are currently available. If the menu of alternative service options changes over time, the impact of a given IRS service on compliance behavior might also change. This makes it difficult to generalize between the measured impact of a taxpayer service at one point in time and its impact in other time periods. Second, taxpayer services may involve a variety of components (personal interaction, information transmission, follow-up activity) and qualities (accessibility, accuracy, helpfulness, etc.). In a given field experiment, either one or a specific set of these components and qualities might be changed. Information obtained from such an experiment may not be very informative as to what the effect would be of changing other elements of the service.

Another concern is that it may be difficult to attribute changes in compliance to IRS activities, as opposed to the effect of other differences between the treatment and control groups. This concern relates to the complexity of isolating effects in the field. In the Adelsheim and Zanetti paper mentioned above, for example, one cannot be sure that the geographic areas chosen for treatment and control are fully comparable (which is why that study was called quasi-controlled). Although the difference-in-differences methodology helps to control for time-invariant differences between taxpayers in the treatment and control areas, one cannot rule out the possibility that factors relevant to the filing decision beyond the treatment (mass media messages) changed in different ways across the two areas. As noted previously, the difference-in-differences methodology can be extended using a regression framework to account for measured differences in relevant factors across areas. However, if one detects the presence of differences in these measured factors, this is likely to raise concerns about possible differences in unmeasured factors that cannot be controlled for in a regression specification.

The selection of comparable treatment and control groups is complicated by concerns over potential spillover effects. As noted in the earlier discussion of the proposed econometric methodologies, the indirect effects of service and enforcement activities on compliance are likely to be driven by information flows and interactions among taxpayers. Therefore, it is desirable to select treatment and control groups such that the potential for information flows and interactions between the two groups is minimized. However, imposing this restriction may have adverse consequences for the comparability of the groups. In practice, comparable groups of taxpayers are likely to have similar interests, which increases the likelihood that they will live in the similar areas, rely on common sources of information, and belong to the same organizations and social networks.

The choice of the unit of observation (individual or group) is a core issue in the research activities related to services. Randomization at the individual level is ideal, but that isn't always possible. Unlike enforcement activities, the IRS does not always know exactly which taxpayers took advantage of a given service. If experiments are relegated to using only aggregated data, the effects of randomization are diluted, thus raising potential difficulties for the analysis.

The workshop participants observed that the implementation of field experiments presents some pragmatic challenges, including:

- Reducing or denying IRS services to taxpayers may raise political/policy issues. As with clinical trials to test the effectiveness of medical treatments, if a researcher is aware that a particular service would be beneficial to a set of taxpayers, it could be considered unfair or even unethical to deny this service to a portion of the taxpaying public.
- Increasing IRS services to taxpayers may involve large implementation costs. If the study under consideration contemplates use of a new or revised service, to be compared to either no service or to the prior "old" service, it could be potentially very costly to build it out. It might require obtaining and configuring new hardware and software and/or the hiring and/or training of staff. For example, in the case of testing alternative web-delivered services, one could conceive of the need to create a complex selection protocol that funnels some web services users to one set of available services, while others are exposed to an alternative set. Designing and implementing such tests would require substantial resources to implement.

- Natural or quasi-natural experiments may be hard to find. However, field experiments are being used by the IRS to evaluate communications to taxpayers. Current research, by W&I Research and Taxpayer Advocate Service, is measuring the effectiveness of alternative IRS notices.¹⁷ The activity of “searching for field experiments” consists of identifying experiments, most likely through detailed discussions with those who manage the delivery of these services. Overall, it seems likely that such a search activity will uncover only a modest number of potentially fruitful natural experiments.

2.4.3 Likelihood of Success/Risks

In terms of providing useful information to the IRS and in providing information that could support other measurement approaches, this method has a reasonably good likelihood of success where it could be applied effectively. This was supported by the working group summary which concluded that “field experiments may be the best way to measure real world effects, but they may not always be practical to implement”.

The core methodological problem in field experimentation is the level of control that the researcher could attain in performing the experiment. The above discussion on weaknesses and limitations provided some examples of cases involving less than perfect control. This lack of control introduces “confounding” factors into the analysis that complicate the interpretation of results. However, it is always provocative when one actually measures a difference in compliance in the presence of some known change. Such a change might represent a response to the treatment used, to some combination of other confounding variables, or to random effects. Such experimentation would clearly spark the increase in knowledge about how taxpayers act and would provide information on the potential magnitudes of change in compliance that might be reasonable. The latter information may, for example, be a very important compliment to estimates of compliance impact that are derived from other sources, such as econometric estimates.

2.4.4 Expectations (Short-Term and Long-Term)

It is reasonable to expect initial estimates of the impact of certain selected services on tax compliance for a subset of the taxpaying population within one or two years of beginning of a planning activity for this area. Of course, it is clear that the results of such work need to be understood in light of the set of weaknesses and limitations mentioned above; however, the relative speed with which they could be accomplished is an important factor to be considered. As the breakout group on experimentation concluded, “In the short-term field experiments could generate causal inferences about the impact of a specific activity that accounts for “real world” complexity.”

The breakout group for experimentation believed that “In the long-term, field experiments have the highest external validity of estimates of impact and could be used to inform predictive models (e.g., agent-based models).” To attain optimal results from field experiments, however, progress must be made on certain data accessibility issues. For example, in the longer term, the IRS might overcome the barriers to obtaining individual level information on taxpayers who partake of services and, therefore, allow for the confidential use of that taxpayer’s individual information in assessing the compliance outcomes.

2.4.5 Dependencies

Since a field experiment is designed around a specific topic, it requires a specific focus. The design of field experiments should be both opportunistic, taking advantage of quasi-natural variation when it exists, and as carefully targeted as possible, focusing on areas of highest policy importance - either the largest service outlays or the largest compliance issues. In order to set up a field experiment with a treatment group and a control group, knowledge about taxpayers and segmentation of taxpayers into groups is a critical prior step. This suggests that, as data on taxpayers’

¹⁷ W&I Research is currently researching ‘notice effectiveness’ by evaluating the behavior of selected taxpayers after receiving an IRS notice. In one instance, the IRS will study the effect of changes to the IRS CP-79 notice, Earned Income Credit Eligibility Requirement. In another instance, IRS notices on the taxable portion of Social Security benefits will be made more effective using taxpayer focus groups. The IRS conducted tests relating to taxpayer correspondence and made subsequent improvements to five notices that reach over 3.8M taxpayers annually.

usage of services improves over time (with the kind of information aggregation and acquisition activities described elsewhere in this document), more targeted and higher quality field experiments could be designed and implemented.

Ideally one would like not only to quantify the marginal impacts of service and enforcement activities on tax compliance, but also to understand the reasons underlying their impact. With this in mind, the breakout group from the workshop that looked at behavioral determinants specifically suggested that hypotheses about the key drivers of taxpayer behavior be tested using field experiments. This group suggested some possible examples. One such example considers a hypothetical compliance problem involving a specific identifiable group of taxpayers. In this case, one could design a new service to alleviate the problem and then implement it in one district, but not in another similar district, both with comparable representation of the relevant taxpayer group. To assess the comparability of the treatment and control groups it would be important to access survey information that summarizes the demographic makeup of the regions to be included in such a study. At the conclusion of the experiment the researchers would look at compliance in the two groups as well as attitudes pre- and post-treatment.

2.4.6 Data Needs

For field experiments, data would be needed to understand the characteristics of the taxpayer populations that would make up the treatment and control groups. Data at the individual level is best, but aggregate data could be used as well. Much of this information might be obtained from segmentation studies. For natural experiments, variations in IRS services (e.g., geographic or temporal) as well as variation in other exogenous variations will need to be recorded so that reasonable analysis can be accomplished.

2.5 Comparison of Primary Empirical Methodologies

The proposed primary empirical methodologies can be compared on a variety of dimensions, including their relative speed, credibility, generalizability, and contribution to understanding of the reasons underlying compliance effects. An overall rating of the anticipated performance of the aggregate econometric, microeconomic, and field experimentation methodologies with respect to each of these dimensions is provided in Table 2, followed by a more detailed comparison of their relative merits with respect to each dimension.

Table 2: Ratings of Anticipated Performance of Primary Empirical Methodologies

	Speed	Credibility	Generalizability	Contribution to Understanding
Aggregate Econometric	Slow	Modest	Fairly Good	Uncertain
Microeconomic	Medium	Modest	Fairly Good	Uncertain
Field Experiments	Fast	Fairly high	Design-dependent	Design-dependent

2.5.1 Speed

It is anticipated that estimates of marginal effects of service and enforcement activities could be obtained most quickly through field experimentation, followed by the proposed microeconomic approaches, followed by the proposed aggregate econometric approach.

Although it may take some time to design and implement large scale field experiments, it is anticipated that some initial field experiments could be implemented and analyzed within a year or two. If some natural experiments could be identified, these might produce results even more quickly, assuming that the relevant data are available for their analysis. Thus, estimates of marginal effects for some services or enforcement activities may be obtained in a relatively short period of time using the field experimentation approach.

It is anticipated that the econometric approaches will require more time before producing estimates of marginal effects owing to data collection considerations. In the case of the proposed microeconomic methodologies, results may be available in three to five years. However, it may take as many as ten years to realize the benefits of the proposed aggregate econometric methodology, although preliminary estimates may be available in five years or so. This long time horizon is due to the need to collect a sufficiently long time series on aggregate service utilization. However, a quicker possibility that might shed light on these impacts is to construct explanatory variables from historical data on the money spent on (or the number of staff years applied to) the various service and enforcement activities.

2.5.2 Credibility

Among the primary empirical methodologies, results from field experiments are likely to be the most credible, although this depends to a large extent on how well they are designed and how effectively they are implemented. One of the main reasons for the relatively high level of credibility associated with the field experimentation approach is the degree of control the researcher has over the analysis. Econometric studies that rely on observational data must contend with challenging issues surrounding joint causality (endogenous responses), because enforcement and service levels tend to be jointly determined with compliance behavior. Results from such studies hinge on the effectiveness of procedures, such as instrumental variables methods, to account for endogenous responses. The complexity of these procedures, combined with uncertainty surrounding their effectiveness, tends to reduce the credibility of the results. In a field experiment, the researcher is able to experimentally manipulate enforcement and service activities, thereby allowing for straightforward causal interpretations.

Another limitation of econometric studies is the need to condition on a variety of factors besides service and enforcement activities that can influence compliance behavior. This further complicates the analysis and raises questions about whether all relevant confounding influences have been identified, properly measured, and appropriately specified in the analysis. In contrast, field experiments rely on the close similarity between the

treatment and control groups to minimize the influence of confounding factors. For the ideal case in which, absent the experimental treatment, only minor random differences in the outcome variable would be observed between the treatment and control groups, any significant difference in outcomes between the two groups when the treatment is present can be reasonably inferred as the causal impact of the treatment. In this ideal case, a simple analysis of the difference in outcomes across groups based on data free of confounding influences, yields clear and credible evidence on the magnitude of the treatment effect. In practice, however, it is often difficult to construct highly comparable treatment and control groups. As discussed previously, it is therefore a fairly common practice to employ statistical procedures such as difference-in-differences estimation or more general forms of regression analysis to control for relevant differences between the two groups. The presence of such differences therefore complicates the analysis and erodes some of the credibility associated with the approach. Effectively, one winds up with a hybrid methodology that falls somewhere between an ideal experimental approach and an econometric study based on observational data.

Each of these methodologies calls for breaking taxpayers into groups, and spillover effects across these groups pose another potential challenge to their credibility. Under the aggregate econometric approach, one examines how variations in service and enforcement activities at the group level relate to variations in reporting and filing behavior. However, the impact of activities experienced by one taxpayer group (such as all taxpayers residing in a particular state) can spill over to taxpayers in other groups (such as taxpayers residing in different states) as a result of inter-group interactions and information flows. At least as currently specified, the methodology is not designed to account for such spillover effects. Theoretical advances may provide insights into how to construct taxpayer groups in a way that reduces the potential for significant spillover effects, although options may be constrained by data collection limitations.

One faces a similar challenge when attempting to define reference groups for social interactions under the proposed microeconomic approaches. As currently specified, these methodologies assume that relevant social interactions occur within the defined reference groups and not across groups.

In the case of field experiments, the concern is that the effects of the treatment (e.g., enhanced service or increased enforcement activity) may spill over to the control group, thereby resulting in an underestimate of its compliance impact. A challenge for this methodology is to identify a control group that is as similar as possible to the treatment group but at the same time is not likely to be subject to spillover effects from the treatment.

It may take several years for a change in a service or enforcement activity to have its full impact on compliance. Therefore, a methodology is likely to be more credible if it can account for lags in behavioral responses. The proposed aggregate econometric methodology employs panel data, which makes it amenable to a dynamic specification that accounts for lagged responses. In the case of the proposed microeconomic methodologies, the potential to account for lagged responses is dependent on the data that is collected. Such responses could not be addressed if only cross-sectional data were available. However, if such data could be augmented with relevant prior year information about taxpayers, it might be possible to account for lagged behavior. Under the field study approach, it may be possible to track outcomes within the treatment and control groups over a period of several years. However, it may be challenging to maintain control over service and enforcement levels for an extended period. In addition, there is a greater potential for confounding influences to have a differential impact on the two groups over a longer time horizon.

2.5.3 Generalizability

Ideally, one would like to be able to generalize from estimates of marginal effects to:

- The overall population of taxpayers;
- Different time periods; and
- Similar programs;

With a field experiment, such generalizations are often difficult to make. For instance, consider an experiment involving a control group of taxpayers from one geographic area and a treatment group from a second similar area. Unless taxpayers and conditions within those areas happen to be representative of taxpayers and conditions within the overall population, the estimated marginal effects may not be indicative of the marginal effects within the overall population. However, if it were feasible to design an experiment in which taxpayers from the general population

were randomly assigned to treatment and control groups, then the results would tend to be representative of the overall population. Generalizing results from a field experiment to a different time period also poses difficulties. Estimates from an experiment conducted over one time period are conditional on the existing values of other influences on the outcome variable beyond the treatment. The impact of the treatment when these other influences take different values (as they are likely to do in other time periods) may be different. Results from field experiments may also be difficult to generalize to another program unless one is confident that the context, conditions, and features of this program are quite similar to the one that has been studied.

In principle, results from econometric studies are more readily generalized to other contexts. Under the proposed microeconomic approaches, estimation is based on random observational data samples from the overall taxpayer population, and under the aggregate econometric approach, estimation is based on aggregated data representing the entire population. Therefore, the estimates of marginal effects from each approach should be representative of that population. Further, these models condition on a set of explanatory variables meant to capture not only the impact of a variety of service and enforcement activities on compliance, but also the impact of other relevant factors. By plugging in the relevant values of these explanatory variables for a different time period into the estimation equations, one can predict the level of tax compliance in that period. However, as with field studies, it is more challenging to generalize to other similar types services, although this may be possible if the model captures the impact of different elements or qualities of the service and these can be measured for the new service. Overall, though, the relative advantage of econometric studies in terms of the generalizability of results should not be overstated. Accurate predictions of compliance outcomes under conditions when circumstances change rely on an adequate specification of the relationship between tax compliance and all relevant influences. As discussed previously, one faces many challenges when attempting to specify such an econometric relationship.

2.5.4 Contribution to Understanding

Each of the methods has the potential to produce quantitative estimates of the marginal impacts of service and enforcement activities on tax compliance. However, their potential for generating insights into the reasons underlying these behavioral responses is uncertain. As currently specified, the proposed aggregate econometric methodology is not linked to an underlying theoretical model of tax compliance behavior, which makes it difficult to interpret the underlying reasons behind any estimated compliance responses. In principle, it may be possible to adapt the methodology to explore the theoretical connections between services and tax compliance. For instance, variables accounting for differences in social and psychological factors across geographic areas might be incorporated in the specification. However, the aggregate nature of the analysis may impose limitations on the extent to which relevant theoretical issues can be addressed.

The proposed microeconomic approach for studying reporting compliance incorporates a specification for social interaction effects that links the behavior of an individual taxpayer to the circumstances and actions of a reference group. Results from theoretical models and segmentation analysis would be useful in guiding the specification of these social interaction effects, including the assignment of social groupings. The incorporation of these effects may generate insights into the extent to which a taxpayer's compliance behavior is driven by perceptions of service and enforcement activity within his reference group and the degree to which taxpayers conform their behavior to that of the group. However, as noted previously, the successful incorporation of social interaction effects into the methodology requires overcoming some challenging econometric issues.

Field experiments also have some potential to generate insights into the reasons that service and enforcement activities impact compliance behavior, although this is contingent on developing appropriate research designs that are linked to an appropriate theoretical framework.

3 Supporting Empirical Methods

In addition to the primary methodologies, the workshop examined several other approaches: laboratory experiments and general social network models - more specifically, agent-based models. Both are ideal tools for augmenting and contributing to our knowledge of taxpayer compliance behavior and validating directional conclusions regarding the relationship between services and compliance. Ultimately, these approaches were judged “supportive” of the primary methods at this time.

A quick overview of the merits and limitations of these methodologies is provided below in Table 3, followed by a more detailed review of each approach.

Table 3: Merits and Limitations of Supporting Empirical Methodologies

	Merits	Limitations
Laboratory Experiments	<ul style="list-style-type: none"> • Highly controlled environment • Quick time frame • Inexpensive 	<ul style="list-style-type: none"> • Cannot represent full range of compliance consequences • Limited external validity • Most appropriate to validate theory
Social Networks / Agent-Based Models	<ul style="list-style-type: none"> • Builds from the “bottom up” to represent individual compliance motivations and behaviors • Models transient and emergent behaviors (no requirements for equilibrium) 	<ul style="list-style-type: none"> • Computationally intensive • Relatively new methodology • May be difficult or impossible to obtain or validate data on complex individual decision drivers

3.1 Laboratory Experiments

3.1.1 Description/Assumptions/Research Questions Addressed

Laboratory experiments measure taxpayer behavior in a lab environment to understand the impact of IRS activities. Implementing laboratory experiments involves the classic tools of controlled experimentation, for example by defining both a treatment group and a control group. A lab experiment observes individual and group behavior in a controlled environment to understand the impact of specific interventions. These methods have become increasingly popular in recent years.

Such experiments are typically fairly inexpensive, relatively easy to implement, especially when using a robust experimental software platform, and very likely to produce informative results. By testing the predictions of theoretical models in a controlled environment, lab experiments could help to validate and improve our understanding of taxpayer behavior. Further, they are capable of producing estimates of both direct and indirect impact of IRS services on compliance. However, because lab experiments cannot precisely duplicate real world conditions, they provide only an approximation of behavioral responses. Nonetheless, these measurements are useful for prioritizing or eliminating factors for further theoretical and empirical research.

3.1.2 Strengths/Weaknesses/Limitations

The strengths of laboratory experiments lie in their ability to estimate compliance changes as the result of an intervention quickly and cheaply in an isolated environment. Laboratory experiments would directly measure changes in compliance as a result of certain types of IRS activities, such as audits and informational services. Because of the ability to control a large number of factors that affect behavior of subjects in the study and to assign subjects to treatment groups randomly, it is possible to conclude that a change in behavior between treatment and control groups is a result of the difference in treatment. Further, since all values in the lab are induced we can obtain exact measures of tax compliance for subjects that are not audited as well as for those that are. Compared to field experiments, laboratory experiments have much more fine-grained control over experiment than a field experiment. The definition of the treatments could be created with minor differences in the characteristics that define the service of interest, creating a very fine-grained set of possible treatments to study. The experimental situation is quite similar to that used for many decades by the medical community to evaluate the benefits of a new medication,

treatment protocol or medical device by use of “clinical trials.” Therefore, the creation of the treatment versus control groups is limited only by the inventiveness and creativity of the research designers

Additionally, laboratory experiments are relatively quick and inexpensive. Because the experimental environment is created in the laboratory and *not* implemented in a real world environment, it is much cheaper to create. Similarly, the experimental timeframe is substantially shortened as compared to having to build out and test a treatment in the field. For instance, one could put participants through many rounds of a laboratory experiment in about an hour’s time, compared to observing real world annual reporting behavior over a period of many years.

The main limitation of laboratory experiments lies precisely in their laboratory setting: it is difficult to ensure the external validity of laboratory experiments or measure the direction of their bias. This challenge is best described by an example. In the real world, being identified by the IRS as noncompliant, or worse, willfully noncompliant could be accompanied by unpleasant audits (requiring time and expenses), payments of penalties and fees, and (again in the worst case) criminal fines or incarceration. However, it is not possible to fully create those real-world consequences in a laboratory experiment. Audits in the lab are relatively quick and the penalties are usually monetized. The laboratory environment does simulate less egregious losses, such as a loss that is limited to something measured in dollars.

Another limitation is that it could be difficult to simulate some types of services in a laboratory setting. To date, laboratory experiments have focused on simple forms of informational services that are relatively straightforward to implement.

3.1.3 Likelihood of Success/Risks

The breakout group for experimentation concluded that laboratory experiments are likely to be successful in providing useful qualitative results. Such results should be of value in prioritizing or eliminating factors for further research, understanding relative differences in the impacts of different types of service, determining the presence or absence of a behavioral response, and informing the development of future theoretical and empirical studies. However, the group was not confident that lab experiments could precisely quantify the real-world impacts of IRS services.

In order to understand whether there is a statistically significant difference in the compliance rate between a treatment group of participants and a control group within a highly controlled environment, a lab experiment could reasonably obtain such information. However, it would be difficult to generalize from the quantitative results of such an experiment to real world behavior. A possible related concept is to consider 3 or more groups, e.g., one control group and two separate treatment groups. Suppose the idea is to understand which of two possible approaches to providing services results in the highest compliance. The lab experiment could find that the control group and the two treatment groups were all significantly different from each other, and that the order of the three in increasing levels of compliance was control group, treatment 1 then treatment 2.

3.1.4 Expectations (Short-Term and Long-Term)

In the short term, lab experiments would generate causal inferences about the impact of a specific activity. Targeted and well-controlled experiments are the hallmark of lab experiments. Therefore, one could expect in the short term, and with a relatively small investment, to begin to compare various services as to the possibility that they affect compliance. Such results should be used primarily to identify possible variables rather than to estimate the numerical effects of such services. This information could be used to design more complex field experiments and to aid theoreticians in formulating econometric and other models for consideration.

In the long term, one could continue to use lab experiments to help inform predictive models (e.g., agent-based model or econometric models), but they would not likely result in confidently answering the primary research question.

3.1.5 Dependencies

Like field experiments, lab experiments are designed around a specific focus and possibly around a specific group of taxpayers. Therefore, they would rely on diagnostic and segmentation analysis to aid in the experimental design. In

addition, some lab experiments may be formulated so as to validate an econometric and agent-based model. In such cases, the experiment design should conform to the objectives of those models.

However, compared to the other methods, lab experiments have minimal dependencies. It is one of the research approaches that can literally be started immediately, limited only by the creativity and ingenuity of the researcher.

3.1.6 Data Needs

Lab experiments would require little prior data except what is necessary to understand taxpayer segments and prioritization of compliance issues through diagnostic analysis. However, they would need to recruit subjects to participate in the study as the treatment and control groups. Most of the data would most likely be generated through the experiment itself at the individual level.

3.2 Agent-Based Modeling and Social Network Analysis

Simulation models are frequently used by policy analysts to provide insights into the behavior of the microeconomic units that make up the larger economy and to help answer questions about how the impact of certain government actions are likely to alter future outcomes. Generally speaking, these models fall into three broad but distinct groups: (1) macroeconomic models, (2) microsimulation models, and (3) computable general equilibrium (CGE) models. Each type of model has its own strengths and weaknesses in answering those questions posed by policymakers. In recent years, a fourth modeling paradigm has emerged – agent-based modeling – that relies on the increased power of computers to address these questions in a common, computationally intensive framework.

Agent-based models are often enhanced by defining an explicit social network structure among the agents. The resulting network topology suggests how agents may interact and how information is shared among agents. As mathematical objects, these social networks are analyzed in much the same way as the physical networks that are so much a part of our everyday life: the Internet, transportation networks and electrical power grids, for example. Research on networks has grown rapidly over the last several years, due mostly to the availability of large, high-quality datasets that facilitate this analysis, lower costs for high capacity computing, and numerical methods for dealing with largescale linear algebra.

With regards to measuring the effects of IRS activities on compliance, agent-based models would simulate taxpayer compliance behavior given a certain type and number of agents and their characteristics as well as information on how these agents make their decisions. Agents can include the taxpayers themselves as well as third parties, such as paid preparers, or even the IRS. In this simulation environment these agents would interact with one another in an assortment of ways resulting in various outcomes. Agent-based models could also take into account taxpayer motivation and attitudes revealed in theoretical development and would utilize results from segmentation analysis when distinguishing agents and their characteristics. These models lend themselves to understanding social interactions by utilizing the social networks in which agents are embedded, and have the capability of including interactions among family and friends to yield the indirect effects of activities.

3.2.1 Description/Assumptions/Research Questions Addressed

Agent-based modeling (ABM) is a relatively recent modeling paradigm that lies at the intersection of computer science, economics and sociology.¹⁸ It is a powerful simulation modeling technique that relies on the power of computers to represent complex, dynamical systems by building them from the “bottom-up” using their constituent components (agents). In ABM, macroscopic outcomes are driven solely by competitive and cooperative interaction among heterogeneous agents.

In an abstract modeling environment, an agent is nothing more than a software component bundled together with data and behavioral methods that interact with these data. As such, ABM lends itself easily to modern computer design principles such as object-oriented programming and parallel computing. Agents communicate with one

¹⁸ Some economists refer to agent-based modeling (ABM) as agent-based computational economics (ACE). An authoritative and up-to-date compendium of papers on the subject is Tesfatsion and Judd (2006).

another, and enabling learning and adaptive behavior on the part of agents (through the ABM software) can greatly increase the usefulness of the modeling technique.

ABM extends conventional simulation approaches through its central focus on modeling micro behaviors and the ability to capture complex system dynamics, which allows for the endogenous interactions among agents and the identification of emergent, global outcomes. In ABM, researchers are often interested in identifying aggregate behaviors that could not have been predicted by using conventional modeling techniques.

Social networks are interesting phenomena in their own right, and much current research is devoted to examining their properties, how they evolve and how they may support certain human behaviors and interactions. In the present context, social networks are viewed as an important and integral part of agent-based modeling. While it is possible to build an agent-based model without an imbedded social network structure, social networks often amplify and enhance the usefulness of ABMs.

3.2.1.1 ABM as a Modeling Framework

In contrast to econometric models, where the modeler is required to assume representative behaviors, ABM allows for diversity of behavior even down to the individual agent. Also, ABM allows for a multiplicity of agents, including taxpayers, third parties such as preparers or tax advisors, legislative entities such as Congress, and administrative entities such as the IRS. In the case of econometric models, the discussion is focused on the form of the model primarily, while with agent-based models, the focus shifts primarily to the data that supports both the decisions that are assumed in the model and the underlying distributions of agents.

Tesfatsion (2006) identifies four main objectives of current ABM research: empirical understanding of stylized, global facts; normative understanding that can lead to the design of better systems; qualitative insight that can lead to better theories; and methodological advancement. As such, ABM can provide a powerful laboratory in support of the main objectives of this project.

With respect to behaviors, there is also a huge data requirement, especially as one attempts to move toward making more credible estimates (not based solely on assumptions about behavior). The basic behavioral information is what agents will do when confronted with a specific situation. These “situations” can be quite complex. For example, if one member of a social network obtains information from an IRS call-in facility that changes his/her compliance behavior (i.e., a direct affect), how shall another member of that social network behave? How many other members obtain information by using the service? How many of them take the information that the first individual obtained and use it to affect their compliance behaviors (without themselves directly using the service)?

The IRS has already begun building several agent simulations. One is the Tax Compliance Simulator (TCS), developed by Kim Bloomquist of IRS Research in NetLogo simulation language. The application of this multi-agent-based model was motivated in part to test the indirect effects of enforcement, initially modeling compliance response to variability in the audit rate (Bloomquist, 2004). Compared to two earlier models, TCS greatly expanded the number of agent attributes, such as opportunity for evasion, produced discrete estimates of direct and indirect effects, and was able to be calibrated to known tax compliance outcomes (Bloomquist, 2006), though the author was also clear to point out that there is no available data to independently validate marginal impacts.

3.2.1.2 Integrating Social Networks Into ABMs

Many complex systems can be represented as a network, a structure that highlights the relationships and dependencies among its constituent components. The mathematical model of a network is a *graph*, an abstract collection of *vertices* (sometimes referred to as nodes) and *edges* (sometimes referred to as arcs). Graph theory provides the intellectual foundation for much of the analysis of real-world networks. When describing a social network in this context, individuals are represented by vertices and when two individuals have a relationship or connection, this relationship is represented as an arc connecting the two.

Modeling social networks can be a valuable modeling tool when examining how information flows among individuals or how infectious diseases may spread within a population. This feature makes social network analysis a natural platform for examining tax compliance outcomes when those outcomes are thought to be influenced by interactions between individuals.

Serious and deliberate study of social networks is a relatively new phenomenon. One reason for this is that accurate data on large, social networks have only recently been available. In the past, information on social networks was often collected through survey instruments and these were often unreliable. Today, however, the Internet and the World Wide Web (WWW) have provided much-needed data on the structure and evolution of real-world social networks, for example the emergence of the social bookmarking web-service called “delicious” (which is currently owned by Yahoo) and related sites and services. This service keeps track of sites that a person bookmarks, allowing for that individual to find other sites of interest, thus encouraging social networking.

The modern study of social networks recognizes that they are dynamic objects that change and evolve over time as new edges and vertices are constantly added or removed. This strand of research views a social network as the ongoing outcome of the endogenous decisions of individuals. Terms like clustering, connectedness and path lengths are used to describe a social network and these characteristics, and how they change over time or in response to events are the focus of intense statistical analysis. In the context of how a social network structure can support agent-based modeling, two important assumptions relate to how agents interact (the *interaction topology*) and how information is transmitted among agents (the *activation regime*).

Research on tax compliance suggests several natural social networks that can be defined over the population of taxpayers:

- Geographical proximity
- Employment relationships
- Occupational similarities
- Investment similarities
- Familial relationships
- Third Party (e.g., paid tax preparers) services
- User groups, such as for financial planning or tax preparation software

An accurate picture of the social network of a particular taxpayer or community of taxpayers may suggest innovative and improved ways to disseminate and deliver a higher quality of service. It may also be possible to study a relatively small community of taxpayers in detail, to derive insights that apply to much larger populations.

For example, the dynamic-network simulation model, CONSTRUCT, developed by Kathleen Carley and colleagues at CMU, is a sophisticated social network representation, which represents belief and knowledge networks (Carley and Maxwell, 2006). This model has been specified for two scenarios directly relevant to the project here. The first represented a tax scheme promoted in a typical Midwestern town, countered by an IRS ad. The second is a simulation of the IRS EITC Certification study done in 2004-5 in Hartford, Connecticut, with the goal of explicitly validating the application of social network simulation to real world experiments.¹⁹ The application was judged successful in representing population response rates to EITC notices. However, the experimentation at present signals some salient challenges in the data requirements for ABMs: namely, challenges in representing critical IRS service interventions (e.g., TAC visits by taxpayers and Service intervention through local partners) and challenges in identifying key taxpayer decision components (in this case, decision on the timing of filing).²⁰

3.2.2 Strengths/Weaknesses/Limitations

The two primary strengths as articulated by the workshop breakout group on this topic were the ability of ABMs:

- To aid in the formulation of theoretical models

¹⁹ In the 2006 paper, Carley cautions that social network model specification is appropriate in a policy situation when good empirical data exists and “...the refinement enables using the simulation to reason about an important outcome or behavior that could not otherwise be reasoned about.” The EITC study had refined data about behavior and participation. However, participation in TACs was not able to be modeled and represents some of the issues that could be expected in many simulations with IRS services data.

²⁰ This experiment is described extensively in the Taxpayer Advocate 2007 Annual Report to Congress.

- To validate, test, or confirm the results of other models

With respect to theoretical model formulation, ABM provides the researcher with substantial flexibility to analyze complex and volatile systems, including the ability to join individual behavior with complex and potentially “non-linear” institutional relationships. In addition, ABM can reflect heterogeneity and social network relationships relatively easily. In the context of the current Indirect Effects Initiative, for example, ABM could examine the likely consequences of moving resources among various services, increasing or decreasing resources available for selected services or building “new” services or interventions into the model. From the perspective of social networks and how they may affect tax compliance, ABM could model the impact of behaviors on the local social system, its evolution and the propagation of information among taxpayers. As a consequence of the range of possible scenarios and the emergent behaviors that often result, the modeler can discover new modeling structures and develop intuition that can aid in the formulation of new models.

ABM is also an excellent platform to validate, test or confirm modeling outcomes from other models. A specific validation concept discussed by the workshop breakout group was the idea of taking estimates of direct or indirect impacts (if available) and using an agent-based model to attempt to reproduce the same impact. The goal would be to vary the assumptions about taxpayer behaviors so that that outcomes matched the produced impacts of the other model, thereby allowing for an independent evaluation of the credibility of the estimates. In addition, and in a limited way, ABM can be used to evaluate the impact of policy changes, how they may evolve dynamically and what population segments may be disproportionately affected. However, ABM would generally not provide specific, detailed estimates of the impact of policy options. Rather, ABM would serve as a guide to suggest what types of macro outcomes would likely occur under alternative assumptions about agent behavior.

The workshop breakout group suggested the following weaknesses of ABM:

- Probably the most mentioned concern for ABM is the challenge of validating behavioral assumptions in the models. ABM relies on rule-based approaches to modeling behavior and eschews theoretical or optimization-based approaches to simulating the actions of agents. As such, ABM frequently assumes agents possess “bounded rationality” and relies on ad hoc rules-of-thumb to drive agent behavior. Thus, there is an ongoing need for validating assumptions coincident with validation and acquisition of supporting data.
- The predictive power of ABM models depend critically on the credibility of the input data. When the decision rules are weakly supported by data, the predictions can only be interpreted as trends or directions but not magnitude.
- As mentioned above, ABM is a relatively new methodology and not well-understood by the general research community.
- ABM is computation-intensive, creating a challenge for scalability and there is a possibility that the outcomes of an ABM might be sensitive to seemingly innocuous changes in the input data that may require extensive sensitivity analysis.
- Also, the cost of maintenance can be high because of the relatively large data input requirements for large models.
- Lastly, there was some concern that these methods might not be easily accessible to policy makers because it is difficult to describe the assumptions and approach and to interpret the outcomes.

A comment by one of the workshop participants focused in on a core difference between ABM and the more standard modeling approaches. Whereas in econometric or other theoretical models, the discussion, concern and effort is focused on the form of the model: functional form, mathematical relationships among the variables and the direction of causality or simultaneity, for example. In an econometric model, it is often the case that structural relationships will be derived from some underlying optimization framework (e.g., utility maximization) and analytical solutions will be known in advance. With respect to agent-based models, the primary discussion and effort is on the data because the model, generally, cannot be solved analytically: the agents themselves provide the solutions to the model. The specific “model” is often quite simple: track the agents as they evolve over time based on the decisions they make when faced with specific choices. However, the underlying data is very complex. It includes information on how many agents of each “type” there are, how one defines choices they have to make, and how each specific agent would behave when faced with a specific choice. We point out that this behavior can be

quite complex, especially when one allows the individual agents to learn and adapt to a dynamically changing environment.²¹

Knowledge about the structure and evolution of social networks could be a valuable tool in modeling tax compliance through a detailed mapping of how information might flow between taxpayers. This analysis could lead to identifying more efficient means to providing services. However, even if a social network structure were known exactly, it is unclear that much could be said about the actual transmittal of information among the group without stronger assumptions. That is, just because two taxpayers are connected in the same social network doesn't mean they communicate and share information about tax compliance or IRS services. But an agent-based model could be used to test a variety of assumptions about whether a social network could play an important role in tax compliance and indicate whether the IRS should promote and use social networks to transmit information to certain taxpayer groups.

3.2.3 Likelihood of Success/Risks

In conjunction with more traditional modeling approaches, ABM can provide valuable insights into taxpayer behavior and how compliance outcomes may be affected by alternative policies. As a supporting, confirming methodology, or a tool for understanding emergent behavior, ABM is a very important and powerful method to be included in a broader portfolio of methods. As a stand-alone modeling platform, however, it is unlikely that ABM can meet the objectives of this project, especially in the short to medium term. Long-term success would require acquisition of much better data about agents' behavior, characteristics and population distribution. As a relatively new technology, the effectiveness would evolve over time.

In addition, modeling a social network structure as part of a well-defined agent-based model would provide a powerful laboratory to test and validate alternative theories about how certain IRS services are likely to affect tax compliance. The results of such a modeling initiative would have to be used in conjunction with other data collection initiatives.

3.2.4 Expectations (Short-Term and Long-Term)

It is unlikely that this methodology would produce estimates of marginal impacts of IRS services in the near term, say over the next three to five years. In the short run, what could be expected is to study via an ABM, a piece of the ecology (e.g., new taxpayers, cities, etc.) develop a model for this microcosm, model it, link it to field work, and potentially link to experiments. Also, short term data acquisition goals of this research could provide important information on segmentation of the various agent populations (taxpayers, third parties, and administrative entities).

In the longer term, it is reasonable to assume that a high fidelity model capable of comparing different strategies or approaches and capable of validating other methodologies could be developed.

3.2.5 Dependencies

There are four areas where significant efforts in data acquisition and analysis would be necessary for ABM research to proceed:

- Sociological research, to clarify the characteristics of different segments of the population
- Surveys or other information to determine taxpayer social networks, how tax information is shared, and how tax attitudes and perceptions are acquired.²²
- The results of econometric models for comparison and mutual confirmation.
- Good information on the demographics and counts of taxpayers in different taxpayer segments.

²¹ Two common approaches to modeling the complex, adaptive behavior of agents that appear in the literature rely on genetic algorithms (GA) and neural networks.

²² For more information on the survey data required, please see Section 1.3 .

Identifying a taxpayer's social network would, in many cases, require a coordinated data collection effort to ensure that existing administrative information is captured and retained. Because social network analysis is a relatively new science, a theoretical foundation for examining taxpayer networks would have to be developed in the early stages of the process. In guiding this development, a useful starting point would be the recent literature on network structure and formation which could be applied, with very little modification, to taxpayer networks.²³

3.2.6 Data Needs

Successful ABMs mimic real-world phenomenon in ways that facilitate exploration, validation, sensitivity analysis and examination of counterfactual policy options. In order to achieve this level of sophistication, ABMs may require large amounts of data. As far as where to obtain these data, multiple sources of information need to be leveraged. These sources include administrative data, surveys, and experimentation (both lab and field).

The administrative data are clearly the most accessible, but may require substantial effort to put into the appropriate form for use in analysis. However, this information is very rich, as it includes simple demographics (for example the number of dependents as shown on individual tax returns) and tax outcome data (for example "reason codes" that are recorded by Revenue Agents at the conclusion of an audit). Survey and experimental data may be the best ways to obtain many of the data elements. Indeed, the survey of service usage—and the predictive formulae that would be derived from such data—that were proposed by the econometrics group—would likely be invaluable for ABM modeling, as well. In addition, the IRS needs a mechanism for storing, and later accessing, this information by creating information archives, data warehouses, or other data structures. Such information would need to be continually refreshed to take advantage of ABM's capability to study the time dimension of behavior.

Demographic data includes information on: geography, education, employment, gender, race, religion, tax type, preparation method, financial status (income, assets) for each taxpayer-agent. Such data defines "population segments." To support additional modeling efforts, information is required on the numbers of each type, how this population is changing over time, how transitions from one status to another are accomplished, and to what subgroups agents belong.

Data to support behavioral modeling would include information on: what services are supplied; the characteristics of those who choose to use (or not use) the services; where agents get information; how this information is processed, and in certain specific situations, how agents respond to service initiatives; and the mechanisms by which information diffuses, or propagates, from one agent to another. Some of the options for surveying behavioral data are further discussed in Section 1.3.

The data needed to construct an accurate social network structure among taxpayers would, in some cases, be impossible to obtain in either the near or long term, primarily because of data confidentiality and privacy rules. However, some social network data could be inferred from readily available sources. For example, if it is believed that a geographical network is relevant for modeling certain types of tax compliance, then zip codes contained on taxpayer returns are all that may be needed. At a finer level of detail, street addresses may suffice. Secondly, employment information contained on W-2s would identify taxpayers working for the same employer and this may be a fruitful avenue for constructing meaningful social networks in some contexts. Similarly, occupational information contained on the 1040, if deemed reliable, could also be used to identify relationships, especially in conjunction with geographical and employment data. Information on paid professionals used by taxpayers is already collection by the IRS and this could, again in conjunction with other information, be used to construct useful network structures.

4 Interrelationships Among Methodologies and Data Needs

The methodologies presented in sections 2 through 4 have some important interrelationships that need to be considered when formulating an integrated plan to prioritize and stage their implementation. Of particular

²³ Two comprehensive references on the theory behind models of network formation and dynamics are Wasserman and Faust (1994) and Watts and Newman (2006).

importance are the contributions of certain methodologies to a research foundation on which other methodologies can build and the feedback effects that are generated in response to the emergence of new empirical findings.

4.1.1 Methodologies that Build a Research Foundation

Segmentation analysis, theoretical modeling, social network analysis, agent-based modeling, and laboratory experimentation are likely to play an important role in building a research foundation for the application of other methodologies.

Segmentation analysis is likely to yield important clues about the nature and implications of heterogeneous behavior of different taxpayer groups. Such information should be of value to researchers in developing theoretical models that explore the implications of services and enforcement activities for different taxpayer segments. As well, it should contribute to the formulation of empirical studies that attempt to account for the distinct motivations and influences acting on different groups.

Theoretical models that link taxpayer services and enforcement activities to compliance behavior are likely to establish a strong foundation for the productive application of empirical methodologies by identifying and explaining potentially important causal relationships, making qualitative predictions, and generating testable hypotheses.

Social network analysis and agent-based modeling can help to refine theoretical predictions and hypotheses by accounting for information diffusion processes and interrelationships among individuals and activities that are analytically intractable.

Laboratory experiments can be used to test theoretical implications in a controlled environment and produce some suggestive results regarding the relative importance of various factors in driving compliance behavior.

The research foundation established by theoretical models and informed by social network analysis, agent-based models, and laboratory experiments should be of important value when specifying empirical models and interpreting empirical results using real world data.

4.1.2 Feedback Effects

Empirical studies based on field experiments and econometric specifications can be used to build on the insights gained from other methodologies to produce initial quantitative estimates of the impact of taxpayer services and enforcement activities on compliance behavior. At the same time, the results of these studies should generate some important feedback effects. As evidence accumulates from such studies on the extent to which various taxpayer segments are distinct, the adequacy of different modeling assumptions, and the validity of theoretical predictions about taxpayer behavior, this should lead to productive new theorizing, improved modeling, and ultimately more reliable empirical results.



Part II: Data Collection

1 Methodology Data Needs

This section describes the data needs of the various methods in light of their commonalities.²⁴ First, we identify methodologies which have little or no data needs, allowing preliminary applications to begin virtually immediately. Secondly, we describe overlapping data needs between methods, for which concerted collection efforts would benefit more than one approach. Lastly, we delineate unique data dependencies, which warrant additional scrutiny for investment.

1.1 Methodologies with Relatively Little Data Needs

Theoretical models and laboratory experiments require relatively little data. However, both may benefit from some summary tabulations relating to taxpayer characteristics, enforcement and service activities. Summary tabulations offer a sense of the potential importance of various issues, the nature of research questions, and provide preliminary segmentation. Insights gained from more thorough segmentation analyses should ultimately contribute to improved theoretical and experimental designs, but preliminary construction of theoretical models and laboratory experiments could begin with little or no preliminary data analysis.

1.2 Common Data Needs among Methodologies

One of the clearest messages emerging from the Indirect Effects Workshop was the need for detailed information about taxpayers' uses, preferences and behaviors with regard to services. At least some level of detail on such information is important for the application of the segmentation, econometric, social network, and ABM methodologies. Not only are data required on how many taxpayers use the service, but "who" such taxpayers are – that is, what are the characteristics of taxpayers who choose to use and not use various services. Since taxpayer-initiated pre-filing/filing services are generally anonymous, such data on taxpayer characteristics is derived from surveys.

In addition to data on services, these methods require information on enforcement activities and taxpayer characteristics. However, some methods require more disaggregated data than others (e.g., microeconomic methods that rely on survey results that are linked to individual income tax returns), while others require data of a longer duration, in some cases in the form of a panel data base (e.g., aggregate econometric methods).

1.3 Specialized Data Needs among Methodologies

Many methods have additional specialized data requirements, including some types of information that, while not strictly required, may lead to improved analysis. The prospective benefits of non-critical data (including both direct benefits for project and collateral benefits) should be weighed against the cost when deciding on how much to invest in their collection.

The proposed aggregate econometric methodology requires panel data aggregated by state or other geographic area on socio-economic factors from sources such as BEA, Census, and the Department of Commerce. In general, such data is publicly available and has been used in existing studies. To constitute an aggregate panel of sufficient duration, such data would need to be compiled for services for a long time period, such as ten years.

The proposed microeconomic models for reporting compliance require NRP data, and micro-level data on tax return characteristics from a random sample of audited and unaudited tax returns. It is desirable, but not necessary, to have additional information on attitudes and perceptions. Under one of the proposed options, a random sample of taxpayers pre-selected for an NRP audit would be included in a survey on taxpayer service utilization, attitudes, and perceptions, thereby allowing a link between the survey responses, tax returns, and NRP audit results.

²⁴ Please see the Data Requirements section of each methodology, which enumerates their data needs in more detail.

For the proposed microeconomic analysis of filing behavior, the data requirements include information on a broad set of socio-economic characteristics, ideally with detailed information on sources and levels of income and potential tax liability (so that it would be possible to assess whether there is a filing requirement) as well as an indicator for whether the taxpayer filed a tax return. Information would also be needed about perceptions of IRS service quality. Although not strictly necessary, it would also be desirable to have taxpayer responses to a variety of questions on attitudes and perceptions about the IRS and about government in general. Such information would make it possible to account for a broader range of potential behavioral determinants of filing behavior.

For field experiments, individual or aggregate level information on taxpayer characteristics and service utilization is needed to select treatment and control groups, depending on the approach taken. In addition, for the subjects of the experiment, results need to be collected on either tax reporting behavior or tax compliance, as well as any variables that might be used to control for differences between treatment and control groups when analyzing the results.

For Social Network, ABM models, other models of Behavioral Determinants, it is desirable to have additional data on taxpayers relating to taxpayer attitudes, perceptions and motivations, social networks and sources of tax knowledge and behavior. ABMs are built upon two key components: good information about the types and number of entities (or agents) in the population of interest and their characteristics; and information on how agents make decisions or choose to behave during the time evolution of the modeling process. This modeling activity is also benefited from knowledge of other third parties, such as paid tax professionals, software developers (who provide tax information via their software), and the IRS as an agent that interacts with the others just mentioned. Some of the information required for ABM to define population segments is already available to the IRS in its administrative records.²⁵ However, an ABM would need additional data to questions such as the following:

- Where do taxpayers get information about how to prepare their taxes?
- If from the IRS which channels do they use?
- If from friends, family and acquaintances, how is this information obtained and processed?
- What are the social networks in which taxpayers live, work, and spend their recreational time?
- How does information obtained in various networks spread?
- How do taxpayers make the choice to seek paid professional help?
- How do taxpayers actually use the information they receive?
- Do they trust the information obtained?
- Do they follow the advice obtained?
- What are the basic underlying motivations of taxpayers?
- Minimizing their taxes
- Making sure they are not audited
- Fulfilling their patriotic duty
- Other factors
- How does the information above affect taxpayer's choices with regard to compliance?
- Why do taxpayers comply?
- Do taxpayers have periods of compliance interspersed with periods of non-compliance? If so, what triggers those changes?
- What fraction of the time is non-compliance based on ignorance of the right thing to do, complexity of the tax system that fosters confusion, or other non-willful behavior?

All these characteristics may vary by taxpayer segments and demographic categories.

The options for obtaining these data are now discussed in the next section, Part II Section 2 Possible Sources to Meet Data Needs (page 50).

²⁵ A detailed discussion of administrative data appears in Part II Section 2.1.2 Existing Pre-filing / Filing Services Data of this document, allowing the focus in this section to be on that information that cannot be obtained from IRS administrative records.

Table 4: Summary of Data Needs by Methodological Approach
 (Requirements are bolded, benefits are italicized)

Method	Summary of Data Needs
Theoretical Development	<ul style="list-style-type: none"> • <i>Benefits from summary tabulations relating to taxpayer characteristics, enforcement and service activities.</i> • <i>Benefits from data regarding attitudes/motivations, beliefs.</i>
Segmentation Analysis	<ul style="list-style-type: none"> • Requires data regarding taxpayer characteristics, demographics. • Requires data regarding enforcement and service activities. • <i>Benefits from data regarding attitudes/motivations, beliefs.</i>
Aggregate Econometric Modeling	<ul style="list-style-type: none"> • Requires information on taxpayer characteristics, enforcement and service activities. • Requires this data in longitudinal historical detail, ideally spanning at least 10 years. • Requires economic data from Department of Commerce, Bureau of Economic Analysis (BEA) estimates, socio-economic data from Bureau of the Census.
Micro Econometric Modeling: Reporting Compliance	<ul style="list-style-type: none"> • Requires information on taxpayer characteristics, enforcement and service activities. • Requires at least one year of existing NRP service utilization questions (begun in 2006) or new service utilization data to estimate a model that predicts the likelihood of taxpayers using different services. • Requires individual level compliance data with representative samples of audited and unaudited returns from comparable tax years, such as NRP.
Micro-econometric Modeling: Filing Compliance	<ul style="list-style-type: none"> • Requires information on taxpayer characteristics and service quality. • Requires data for a representative sample of filers and nonfilers of individual tax returns with a broad set of socio-economic characteristics, including sources and levels of income, potential tax liability, and a return filing indicator • <i>Benefits from data on attitudes and perceptions about the IRS and government in general.</i>
Field Experiments and Quasi/Natural Experiments	<ul style="list-style-type: none"> • Requires information on taxpayer characteristics, enforcement and service activities, at least for the area of focus. • Requires data on variation in services activities. • Requires data on tax reporting behavior or tax compliance of subjects, as well as any variables that might be used to control for differences between treatment and control groups when analyzing the results.
Laboratory Experiments	<ul style="list-style-type: none"> • <i>Benefits from summary tabulations relating to taxpayer characteristics, enforcement and service activities; segmentation.</i>
Social Network and Agent-Based Models	<ul style="list-style-type: none"> • Requires information on taxpayer characteristics, enforcement and service activities. • Requires information / assumptions regarding taxpayers' social networks – either constructed, such as from occupational and employment network data reported on tax returns, or survey responses. • Requires data on how taxpayers share and acquire tax knowledge / attitudes. • <i>Benefits from data on attitudes and perceptions about taxpayer services, the tax agency, and government in general.</i>

2 Possible Sources to Meet Data Needs

Given this rather extensive list of data requirements, the IRS will need to leverage several sources of existing administrative data on enforcement and services, coupled with survey data to understand taxpayer attitudes and motivations. The IRS may also want to expand existing data collection, augment existing surveys and undertake new data collection to get variables that are missing.

First, Section 2.1 Existing Data Sources (page 50) describes existing administrative enforcement and services data. While little new data is required, this report recommends that such data be compiled in a uniform format for research, such as a Services Data Warehouse. This is followed by existing data sources from surveys. Because additional data on taxpayer attitudes and preferences is desirable to further the level of understanding of taxpayers' compliance motivations, we discuss several options to obtain that data. Since survey data on service usage and taxpayer attitudes have not yet been collected in a way that allows for linkage to compliance, we discuss the benefits of such linkage.

2.1 Existing Data Sources

Wherever possible, this project aims to leverage existing data to minimize collection costs. A significant effort has been made to identify all relevant data sources, but a fully exhaustive review would be advisable.

2.1.1 Tax and Enforcement Data

The importance of maintaining tax and enforcement data resources for research purposes is well recognized, resulting in several excellent enforcement research programs and datasets.

Tax return data is available. For research purposes, IRS Statistics of Income branch maintains several time series cross sectional / panel datasets on taxpayers, such as the Continuous Work History Sample (CWHS), a gold standard for data quality. For more specific queries, researchers can pull data from Individual and Business Master File records, the limited tax data retained from each and every return submitted.

Enforcement activity data is also easily available from both the population of tax returns and from research samples.

- National Research Program: Random examination of approximately 15,000 taxpayers annually provides a research sample on population compliance. The NRP includes compliance adjustments by line item for most of those selected. Data is available from selected years in 3 year cycles beginning in 2002. For earlier data, TCMP provides line adjustments from stratified random audits every several years from 1965 to 1988.
- Enforcement Revenue Information System: Designed to support enterprise-wide estimates of "return on investment" for enforcement activities, ERIS tracks assessments made, revenues collected, direct hours reported, and cost incurred by IRS function from 1994 to present.

Enforcement data on individual accounts undergoing current enforcement activity is stored in Integrated Data Retrieval System (IDRS), which serves as the backbone of compliance adjustments. Additionally, for each functional area, data sets exist which interface with IDRS and are of varying accessibility. These are documented in more detail in the attachment, Workshop Materials.

The Compliance Data Warehouse (CDW) links both tax and enforcement data in an accessible research repository of multiple legacy data sources. For example, CDW contains both NRP and ERIS extracts, data from individual and business returns, and audit enforcement data [specifically, Audit Information Management System (AIMS) and Examination Operational Audit Database (EOAD)]. Sources such as CDW and ERIS are more "analyst-friendly" so should be used whenever sufficient.

2.1.2 Existing Pre-filing / Filing Services Data

Currently, the IRS collects a wide variety of operational data on services, used for managerial decision making and process improvement within a given "channel" of services (examples of channels are electronic, telephone, and field office). Quality measurements are sampled for every channel on a regular basis from taxpayers (with the exception of Outreach services, which conduct a Local Partners survey).

There are several impediments preventing the use of existing service data for project research. One is that the length of data retention is very short; these data go back to 2002 at the earliest. Another is that continuous improvement in data collection, as well as methodological refinement of service performance measurements, pose issues of consistency in data representation. Lastly, services data are not currently standardized at the geographic, temporal or topical level.

Primarily, we looked for two dimensions in the data to form the core of our investigation:

- Service utilization (volume, demand)
- Service quality (accuracy, professionalism, taxpayer perceptions)

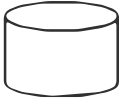



Service quality can be distinguished further as two levels: the level of quality provided and the perception of that level of quality. The first quality measurement is that which the IRS manages from an operational standpoint, as service outputs. The level of taxpayer satisfaction, however, is the ultimate service outcome. The IRS currently employs analysis to measure both the levels and quality of service outputs (such as accessibility, professionalism and accuracy) while ensuring that these are correlated positively with taxpayers' perceptions of the IRS.

While overall levels of service and quality are important in the aggregate, it is beneficial for modeling to break down services to geographic and topical areas. In particular, it is challenging to design Field Experiments, or even to perform aggregate econometric analysis, when variation cannot be captured at a consistent level. A common segmentation approach is to use geographic area. Unlike Taxpayer Assistance Centers and outreach, it does not make sense to apply service quality metrics by geographic area of *the taxpayer served* for phone and web since these services apply to taxpayers nationally. Therefore, the identification of control areas for compliance measurement will need to be addressed.

Each channel collects general volumetric data, as well as characterizing this data by the type and topic of interaction. Pre-Filing / Filing Cross-Channel Services Data represents the available data and potential queries. The first row represents the aggregate service outputs compiled by each channel.



Figure 5: Pre-Filing / Filing Cross-Channel Services Data

SERVICE/ CHANNEL	PHONE	FIELD/ WALK-IN	WEB	OUTREACH
INFORMATION PROVIDED				
Data Elements for Services Provided	JOC Data	QMATIC	WebTrends™ Extract	Form 4913 / 13315 Reports
What quantity of services does IRS provide? What quantity did taxpayers desire?	Demand (incoming calls) and provision (calls answered)	Units of service aggregate	Downloads, views	Taxpayers reached, Returns prepared
What level of services does IRS provide by geographic area?	National routing system	Volume, by TAC area, territory, state	Collection by 'cookies' prohibited	Volume, by SPEC area, territory
In which tax or compliance topic were IRS services provided? Excluding IRS-initiated outreach, what topics were 'in demand' by taxpayers? Which channels are being used most frequently for which topics?	Only through automated call direct (or survey)	Optional 'pop-up' web survey of visitors	By page, which can be categorized	Categorized as Compliance, EITC, E-Services or Other
What was the quality (overall, sampled) of the service provided?	Managerial review of calls (EQRS/NQRS)	Managerial review	Internal Survey of IRS customers for some products	Return prep. accuracy is not currently accessed
Were taxpayers satisfied with the service?	Taxpayer Survey	Cust. Service Comment Cards	Optional 'pop-up' web survey of visitors	Survey of Local Partners
What level of IRS or volunteer resources did it take to provide this service?	Employee time, engaged and un-utilized	Employee time by broad categories – e.g. "BMF return preparation"	Overall total; Not allocated	Volunteer volume but not hours

2.1.2.1 Existing Web Services Data

Since the implementation of IRS.gov in 1996, the IRS has collected web data to guide website development. IRS web data is generally "views and visits" volumes, since extensive web analytics is proscribed by government privacy regulations.²⁶

Data is collected on an aggregate level from individual visits and downloads, specifically the number of visits to the website, the number of pages views and the number of downloads. The IRS.gov site is a managed service through a vendor, and the IRS receives periodic data and reports (through WebTrends™). Information regarding overall

²⁶ For more information on potential web statistics and the inability to collect these through 'cookies,' due to The Privacy Act of 1974 and subsequent OMB and Treasury guidelines, such as The Department of Treasury, (2002), 81-08: Certification Process for the Use of Persistent Cookies on Treasury Web Sites, see Dixon, 2003.

aggregate statistics (hits, page views, visits), while provided since February 2002, significantly changed counting methodology in October 2004.²⁷ Monthly reports are available on the Electronic Tax Administration (ETA) website.

ETA also manages a “pop up” customer satisfaction survey (American Customer Satisfaction Index) that has quality, demographic and open-ended questions.

2.1.2.2 Existing Phone Services Data

The IRS maintains a wide variety of data on phone call assistance. The IRS tracks the volumes of incoming calls from taxpayers, as well as how many of those calls were answered and assisted. They also track call topics: first, as menu selections from the voice interaction and secondly, as a result of call quality sampling. The IRS maintains a plethora of workload information to manage staffing metrics such as employee utilization.

All calls are run through the Joint Operations Center (JOC), which maintains Enterprise Telephone Database (ETD) information online at <http://joc.enterprise.irs.gov>. Current ETD reports are prepopulated, but it is possible to perform ad hoc queries of information. Some data goes back as far as 2002, but generally available from 2004.

Very little identifying information is available on which taxpayers access telephone services, aside from survey data. Phone service is provided and routed nationally through two tiers: Intelligent Contract Manager (ICM), assisted by Voice Response Units (VRUs), routes calls and stores call data; ICM and Aspect data are stored in the Enterprise Telephone Database (ETD). The ETD is the source for the majority of JOC data.

While volume and quality can be tracked at the geographic area of the customer service representative, it cannot be tracked by the geographic area of the incoming (taxpayer) call. Area codes, provided by AT&T to provide an imperfect geographic picture: locations can be aggregated into one originating number, and taxpayers call from cell phones with area codes that no longer reflect their geography. Additionally, phone number data may be problematic to use for privacy considerations. Linking the phone number to other taxpayer information, such as a name, requires the additional assumption that taxpayers are calling from telephones registered in their name. The best data, in this regard, is that obtained from optional customer service surveys. However, because participation in these surveys is optional, the results are unlikely to be representative of the overall population of telephone assistance callers.

Customer service surveys are performed, which provide an average satisfaction level per service (as well as factor analysis to guide high-impact improvements).²⁸ Additionally, calls are monitored for overall program quality. To evaluate quality, a full recording of the call is maintained in a system called Contact Recording for 45 days, during which it may be selected for evaluation by IRS managers. Evaluations are recorded in the quality control database, Embedded Quality Review System (EQRS/NQRS). The IRS plans to add Contact Analytics, a product which will transfer these transcripts to text form.

When answering a tax account question, the taxpayer is generally initiating contact in response to enforcement action. The CSR would assess the taxpayer’s account to provide service, and these actions would be recorded to IDRS.

²⁷ For more information, see the Monthly WebTrends reports, Glossary worksheet, “Data Analysis” and “Considerations for Comparing Data from October 1, 2004 and Beyond to Previous Years” sections. In 2004, the SmartSource Data Collector (SDC) was adopted, causing (a) reported individual IP addresses through session cookies, increasing ‘visits,’ and (b) now excludes redirects from ‘page views.’ The ETA intranet site is <http://win.web.irs.gov/eta/Statistics/default.htm>.

²⁸ IRS quality measures have changed significantly in both methodology and technology over the last few years. In FY2004, IRS began to implement a ‘defects per opportunity’ measurement methodology, while in FY2006 IRS began recording taxpayer calls, which had previously only been able to be monitored in real time (Rosage, 2004).

2.1.2.3 Existing Walk-in / Taxpayer Assistance Center Services Data

Field personnel at Taxpayer Assistance Centers answer a wide variety of questions from tax law to account management, and they record data regarding taxpayer visits in several systems. Visit volumes by area and issue are recorded in QMATIC, which is the most obvious source of service levels for measuring indirect effects.

Centralized Quality Review System (CQRS) is used to perform managerial quality management. Additionally, taxpayers may submit comment cards regarding the service, which reports an overall level of satisfaction.

For post-filing services, resource timesheet recording in a system called FAMIS is assigned to a taxpayer account, and taxpayer information is accessed through an interface to IDRS.

2.1.2.4 Existing Outreach Services Data

Stakeholders Partners Education & Communication (SPEC) leverages IRS partners to perform two main types of services: return preparation and outreach (media and non-media). In contrast to the relatively wide, diffuse impact of outreach efforts, return preparation provides intense assistance to a small number of taxpayers.

Outreach has the potential to inform the compliance behavior of taxpayers as an “audience.” SPEC estimates that 129M taxpayers were assisted through taxpayer education programs in Fiscal Year 2007. However, these estimates are by nature quite imprecise, often based on media circulation figures that may or may not correspond to primary taxpayers. SPEC assesses the aggregate impact of all people who might have received the message and tracks those volumes through a management information system called STARS. This data can be broken down geographically and topically into 4 categories (Compliance, EITC, E-Services or Other).

Return preparation services have a very clear direct impact on the approximately 2.6M taxpayer recipients of preparation services, through Volunteer Income Tax Assistance and Tax Counseling for the Elderly programs. SPEC tracks how many returns are done, and approximately 95% of the returns prepared through partners can be tracked through preparer number captured on Individual Returns Transaction File (IRTF).²⁹ SPEC does not survey taxpayers who had their returns prepared at SPEC Volunteer sites to assess their satisfaction, although they do survey partners on the process.

SPEC’s current data is generally available from October 2004. However, SPEC is in the process of reviewing its output / outcome measurements and restructuring this data system.

2.1.2.5 Existing Publications Services Data

One of the more traditional services performed by the IRS is the printing and distribution of publications in “hard copy.” While IRS Publications designs and prints the forms, IRS Distribution handles form delivery to grocery stores, libraries, post offices etc. Some feedback is provided in the order numbers of distribution sites each year, but otherwise the IRS collects no feedback on consumption of paper products.

2.1.3 Consolidating Existing Data: Cross-Channel Services Data Warehouse

For cross-channel research, it is imperative that such data collection be stored and standardized in a common framework. Across a common unit of measurement, *unit of service provided*, research could be performed to analyze volume and quality fluctuations across channels. For this project, the objective of analysis is measurement of how variability – natural or induced – impacts compliance.

²⁹ Returns prepared at SPEC volunteer return preparation sites can be identified by using the Site Identification Number placed into the Preparer SSN field on a tax return and contained in IRS data files. Sometimes the software used by these sites prevents them from using the SIDN, so EFINs can be used to identify any returns filed from SPEC volunteer return preparation sites. About 85% of volunteer returns are filed electronically. Using the Preparer SSN field and the EFIN together identifies about 95% or more of SPEC volunteer prepared returns.

During the workshop, the IRS Breakout Group recommended establishing a common “Services Data Warehouse,” composed of a data warehouse, interface, and software tools. There are several options for how to implement such a framework; it does not necessarily need to be a separate database. The main options are:

- IRS Research could request extracts from the services data managers and maintain these.
- IRS Research could establish a services database, with a periodic importation routine.
- IRS Research could utilize an existing research database, such as Compliance Data Warehouse (CDW), to store services oriented data.

This common repository would make the services data widely accessible through a variety of analytic software tools and easily refreshed. Importation would include data cleaning and standardization processes. If data were collected by extract, the process would be more manual.

Several elements of standardization are desirable:

- Geographic areas, which are currently captured at various area and territory levels;
- Type and topic of interaction, such as “activity” (ex. “tax law question”) and specific tax topic;
- Periodicity of the data, although monthly and annual levels should provide sufficient granularity for research purposes;
- Elements of quality. Ideally these should be standardized, but directional change is most important.

In order to establish such a repository, the IRS would engage a virtual team representing both *consumers* of services data, such as IRS Research, W&I Research, and potentially all service-wide stakeholders represented on the Services Committee, and *providers*, primarily managers of service channels. The team would examine how services data can be best collected, develop the preferred architecture for database/storage implementation, and produce a set of standards for data quality, backed by appropriate cleaning of all input data sources.

2.2 Existing Survey Data

Some of the data for this project will be obtained through surveys. The IRS already has a substantial amount of information from surveys focused on taxpayer satisfaction with the various IRS services (and in some cases enforcement activities). As discussed in the earlier section, the IRS conducts myriad satisfaction surveys, obtained voluntarily from users of the services, for each channel. However, the IRS also conducts research studies comparing all the channels together and surveying the overall population (rather than simply users). Such research surveys include:

- The TAB Conjoint II Study
- The Opinion Survey or Taxpayer Resources and Services (Benchmark Survey)
- The W&I Market Segment Survey
- The Taxpayer Customer Service and Channel Preference Survey (Oversight Board)

Together, the first three surveys provide an overall picture of service use and taxpayer satisfaction. They tell us how many taxpayers used each IRS channel, for what task, how satisfied they were with the IRS service, and which factors drive taxpayer perceptions of service quality. Importantly, these surveys focus on the taxpayer population at large, rather than known service users.³⁰ Other information about taxpayers includes general demographic data on taxpayer education and income; return preparation choices and characteristics, such as form type and filing status; and taxpayer needs, such as whether the respondent was disabled, had limited English skills, or did not have a computer or internet access. There are some insights on substitute services, and why taxpayers use IRS services or private tax services, such as a paid preparer.

The Market Segment Survey provides 5 years of cross-sectional data on service utilization (2002, 2003, 2006, 2007 and 2008). This random CATI survey of 2,459 taxpayers (30-32% response rate) is aimed at understanding taxpayer

³⁰ Service users are oversampled to obtain a reliable sample size.

service use and satisfaction. The Conjoint study compares taxpayer services models and features, specifically focusing on which factors influence satisfaction. The Conjoint study was done electronically through web or TV with a pre-selected group of market research participants (N = 2,196; response rate was 75%). The Benchmark study, a short mail survey with the largest sample size (over 10,000 responses were received from a sample of 40,000), provided data on perceptions throughout the population of IRS services. Unfortunately, without over-sampling of the populations which use services, the number of service users in the Benchmark study is generally too small.

This survey data could meet the needs of the methodologies in multiple ways. Most importantly, they could meet the needs of theoretical development, segmentation and all methodologies by providing a solid understanding of service utilization. Specifically, these surveys will help to better understand taxpayer groups (segmentation analysis) or to generate a demand model to predict the likelihood of service use (microeconomic analysis). Through the construction of a “services likelihood model,” one can impute a likelihood of service use to an existing data set for which no service data exist. This simulation of services is helpful for microeconomic methods, as well as for efficient stratification of new surveys.

Survey data would be important in determining appropriate control groups for experimental methods. If the aggregate econometric model includes information on service quality, this survey information would provide sufficient measure. The conjoint study uses a specific statistical method to quantify these preferences in ways that can be very useful in defining choices in an ABM. Survey responses which support inference on substitute services (Market Segment and Conjoint surveys) are an important dimension in all methodologies.

Besides services utilization, methodologies are improved with data on taxpayer attitudes. Theoretical development, segmentation and social network / ABM models would benefit from understanding taxpayer attitudes and perceptions, of themselves, of other taxpayers, of the IRS and government in general. Many of the important theoretical drivers of compliance are challenging to observe. For example, few would dispute the hypothesis that taxpayers with high risk aversion and strong beliefs in fairness and civic duties are less likely to be noncompliant. However, risk aversion or attitudes towards government and society, though important, are impossible for the IRS to observe. Through surveys, the IRS could hope to obtain information on general taxpayer attitudes shared in the population. In this area, only one existing survey, the Taxpayer Customer Service and Channel Preference Survey, provides information.

The Taxpayer Customer Service and Channel Preference Survey, commissioned by the IRS Oversight Board, examines attitudinal variables on services, compliance and attitudes towards the IRS and filing and paying taxes. Through identification with common attitudinal postures, such as, “I like to stay up-to-date with changes in the tax laws that might affect me or my family” and “My tax return would be more accurate if I could more easily contact the IRS and get help in filing my return,” the survey segmented taxpayers into six attitudinal profiles. The results also serve as a baseline on public attitudes towards the IRS and perceptions of fairness in tax administration. While this survey is an important source of information, there are several drawbacks to using the survey exclusively. First, the sample size is quite small at 1,000 taxpayers. Secondly, the survey needs to be analyzed for response bias. Lastly, it is not clear to the report authors whether this survey will be on-going or unique.

None of the existing surveys provides information on social networks, how taxpayers acquire and share tax knowledge and attitudes, which is important for theoretical development and ABM models.

In summary, while existing surveys provide an important resource on services utilization, additional information is important to obtain on taxpayer attitude, social networks and social norms, ideally in a stratification design supporting linkage to tax return data. The first step is assessing the full picture about both existing data and modeling by “taking stock” of the IRS surveys in existence now, determining the gaps that need to be closed, and assessing the options for meeting the data requirements. How to acquire the additional data is now presented.

2.3 Augmenting and Collecting New Data

Some of the methods require new survey data. As the section above describes, all methods, with the possible exception of Aggregate Econometric approaches and Laboratory, would be strengthened with data on taxpayer attitudes and perceptions, while ABM in particular requires more data on social networks.



Data Desired from Surveys by Method summarizes the new survey data needs by method.

Table 5: Data Desired from Surveys by Method

Methodology ³¹					
Data Required or Desired	Segmentation	Aggregate Economic	Microeconomic – Reporting	Microeconomic – Filing	ABM ³²
Taxpayer Service Use	Required		Required		Desired
Taxpayer Satisfaction	Required	Desired	Desired	Required	Desired
Taxpayer Attitudes & Perceptions, Norms / Trust	Required		Desired	Desired	Desired
Taxpayer Social Networks	Desired		Desired	Desired	Desired

There are several options of how to obtain this data. The section below suggests four possible approaches:

- Building onto an existing research surveys on services conducted by the IRS
- Commissioning survey data through a national Census survey, CPS
- Extending the NRP with several survey options
- Conducting a stand-alone survey, either to obtain cross-sectional or panel (time series) data

Potential linkage to tax return and compliance data is ideal. There are several options for linking data. First, in IRS surveys, the sample frame could be designed from tax data matched to phone numbers and addresses, rather than randomly selected; that sample pool could be stratified based on the predicted likelihood of noncompliance and of service utilization. In extending the NRP, respondent noncompliance will be known, and respondents could be surveyed in advance, allowing the most direct link to detailed compliance data. In the instance of using Census data, Census may link responses to IRS administrative data to identify nonfilers, for which IRS may be able to obtain analysis results from Census.

Since conducting a stand-alone survey is ideal, in the sense that it could be designed to incorporate almost all desirable traits, we discuss this first. Then, we describe options to augment existing surveys, which might significantly reduce project funding outlays.

2.4 Conducting a Stand-alone Survey

This section discusses the concept of a periodic survey of taxpayers. The primary driver behind this survey is the ability to both segment the population and to use the detailed information obtained to support modeling and estimation of taxpayer metrics. However, much of the thinking behind a survey of the type to be discussed here is driven by the need of creating models of social networks and building agent-based models (ABMs). If this stand-alone survey were extended for panel data, there are unique design considerations discussed in Part II Section 2.4.2 Extension of this Survey to Provides a Service Panel (page 60).

³¹Although theoretical modeling and experimental methods may benefit from segmentation analysis and some summary tabulations of administrative data or survey results, their overall survey requirements are minimal so they have been excluded from this table.

³²Results from ABM are more likely to be informative if driven by a wide variety of data.

2.4.1 Design of Stand-alone Survey

Survey design is critical to the success of this endeavor. While creating a detailed design is not in the scope of this document, it is important to provide some of the key design issues that need to be considered. Careful survey design starts with a clear set of goals and objectives for information acquisition and defines a specific plan for each of the areas discussed in this section. Experience reveals that all of these areas are tightly related to the others, so the design may start with a preliminary plan for each, but then evolve by interacting the needs and requirements of each area of the plan to reach the end. The key areas are:

- Target population
- Sampling frame
- Stratification
- Sample selection
- Data collection instrument
- Privacy Concerns

Simply stated, the **target population** is all taxpayers subject to the U.S. tax law. However, because some individuals who are subject to these laws actually are not required to file (e.g., if that person does not have sufficient income, etc.) and other individuals choose to evade taxes by not filing a tax return, the IRS does not have a totally complete list of this target population. While the first group mentioned is reasonable to ignore in a survey design planning activity, the second is not. Also, there are subgroups within the target population that are more important than others; some important ones are (1) those that are regular or heavy users of taxpayer services and (2) those that do not use services, but are prone to making errors in filing.

Closely related to the question of target population is the **sampling frame**. This is a list from which participants in the survey are selected. With the introduction of phone and web-based methods in survey research, the concept of a sampling frame has undergone some change in the past few years, but the basic concept is important. The core questions are (1) where exactly does the researcher go to select participants in the survey and (2) how well do the participants selected represent the target population? For services, the questions are complicated because the number of taxpayers who use some of the services (e.g., walk-in sites and even telephone call-in services) are relatively small. So doing a simple random sample of taxpayers would “waste” or at least over-sample a lot of cases of those that do not use the services at all.

The last issue leads to the question of **stratification**. This concept is also closely related to the idea of **sample selection**. The design questions here are how should the researcher break up the population into groups for the purposes of either selection or analysis? The statistical issue for stratification is that one should allocate and select the sample so as to minimize the total sampling error in the final estimates. In simple terms, if there is a group that is very homogeneous with respect to the things being measured one does not have to expend a lot of sample on that group; however, if there is a group that is quite variable and could introduce error in the resulting estimates, this group should get relatively more sample allocated. Survey researchers have developed sophisticated methods and tools to define stratification groups and allocate sample among those groups to produce the more useful, least variable, estimates possible.

For the purposes of measuring the effects of services on compliance, an ideal survey design would use existing knowledge of the likelihood of noncompliance, likelihood of services use, or a combination of both to identify the sample.

There are other issues relating to sampling method that are important for the designers to set. A critical one is the “mode” of data collection. This refers to whether the information is collected in person, on the phone, by mail, electronically (by web access), or a combination of these modes. Such decisions also relate to whether to use a simple list (like IRS master file) that has all those who filed tax returns at some time in the past, or whether to use a

more expansive methods like Random Digit Dialing (RDD) methods to seek a wider group. Several of the surveys done by the IRS and mentioned above used RDD methods.³³

Creation of the **sampling instrument or questionnaire** would pull together the core requirements for data collection with the various design decisions that have been discussed, in such a way as to meet the needs of multiple methodologies.

At the minimum, an ideal survey for the Indirect Effects project would include questions described in Part II Section I Methodology Data Needs (page 47):

- Socioeconomic and demographic data
- Taxpayer Service Use: Services, provided both by the IRS and private software and professional services, used in tax preparation
- Taxpayer Satisfaction: Characteristics of those services, experienced and perceived
- Taxpayer Trust, Attitudes, Perception, Norms
- Taxpayers' own attitudes and motivations with regard to taxes, the IRS, government;
- Taxpayers' perceptions of social norms and compliance by others;
- Taxpayers' levels of trust and likelihood of acting on the information and advice they receive
- Taxpayer Social Networks: Social networks inhabited and their influences on tax knowledge and attitudes

An important question in considering the value of the survey is whether it can be linked or imputed to administrative data in order to form the crucial link to compliance. Asking compliance-related questions could bias how a person answers a question. The obvious example is: "Did you comply with the tax law?" While almost all truly compliant taxpayers will answer that question honestly, the noncompliant taxpayers certainly will not—for a wide variety of reasons. Therefore, obtaining the answers to many of the above questions will not be directly from a question asked on a survey, but through linking observed compliance behaviors (administrative data) to attitudes and perceptions (survey results).

While such linkage is desirable, **privacy concerns** are particularly sensitive. Informing participants that their responses will be linked to their return data leads to a selection effect, as fewer noncompliant taxpayers would likely complete the survey. On the other hand, failure to divulge this information raises ethical privacy concerns and may not receive OMB approval. Survey participants must feel comfortable that the IRS will not use results of this study for enforcement rather than research. Fortunately, precedents exist in earlier compliance research. Past surveys have adopted several safeguards for participants' privacy, such as assuring confidentiality or confining survey data to use by independent researchers, not to be shared with the IRS.³⁴ The Benchmark survey, for example, was linked to tax return data, although the percentage of service users was too small to support meaningful inferences and nonresponse analysis was not available.³⁵ Further consideration must be given during survey design in order to both secure the privacy of subjects and to avoid selection effects.

There are other questions that will certainly come up in the design effort. For example the design team needs to decide whether participants will be compensated for their time, even if it is a small incentive fee. Questionnaires differ dramatically by mode and should be matched to the target population of interest. Much research has been done

³³ In such cases, the researchers call a totally random set of people (as long as they have a phone) and "screen" them to see if they fit a set of eligibility requirements. Examples include the Market Segment Survey and Taxpayer Customer Service and Channel Preference (Oversight Board) Survey. That method is not immune from issues such as whether all those in the target population have phones, whether those called will answer the screening questions at all, and, if they answer, would they answer correctly (especially if that person understands that they are not in compliance with the tax law).

³⁴ Examples of the first approach include the Benchmark Survey, which assured respondents that their results would be confidential, while the latter approach was used in compliance analysis by John Scholz and burden survey research by IBM.

³⁵ Sample stratification should address likely noncompliance, likely service use, or a combination of both for survey efficiency.

to create best practices in this area, so bringing that research into the design process will clearly result in a better outcome for the Service.

The above are all complex questions to answer, and depend on the data collection goals of the survey and the resources available to put into the survey process. One of the key projects that will be mentioned in the integrated plan section of this document is to do a focused survey design activity. The key steps and decisions that need to be part of that design activity are those that have been described in this section.

2.4.2 Extension of this Survey to Provides a Service Panel

One of the challenges of obtaining information about taxpayers is that this is a group that is constantly changing over time. While the information obtained from the taxpayer services survey discussed above provides a baseline from which to make decisions and design models, that information could become stale over time. Therefore, the Service should consider the benefits, as well as the costs, of refreshing this information by performing this survey periodically over time. A panel survey is a very important idea to consider here. In a panel survey, one recruits a single set of participants who are surveyed repeatedly, over time, with implementations of essentially the same survey.

In addition to refreshing information, a panel survey allows for measuring very important behaviors that would not be able to be measured in alternative designs. In particular, one can detect changes in behaviors and can look for triggers or drivers of those changes in the data that is collected. Detecting and understanding this type of information from a cross-sectional survey, or even a series of cross-sectional surveys would be very difficult.

A second key question with regard to refreshing data is a statistical one, namely, does a change in the findings of the survey from one version to the next reflect statistical sampling error related to having different subjects in the survey, or does it reflect a “real” change in the population findings? One way to control for, and better understand, the answer to this question is to create a panel survey.

Since a service panel is a generalization of the taxpayer service survey, we delineate only additions for a time series cross sectional (TSCS) data set.

The primary unique challenge for a panel survey is to get a set of participants to be willing to participate over time. The design questions that need to be considered are:

Will there be ongoing compensation required to attract participants?

- How likely will the implementers of the survey be able to find participants between surveys?
- Which questions can be cut back or eliminated because answers do not or are unlikely to change between data collection periods?
- Can there be changes in the data collection mode, for example if one collects the data by phone initially, can updates be obtained annually by email or mail without compromising on quality?
- How does one handle births (new taxpayers entering the population of interest over time) and deaths (taxpayers leaving that population)?

In the panel survey, data needs to be collected periodically over time. The overall set of information required is the same, though sometimes later versions of a panel survey add questions that come to the forefront in later versions of the survey. However, an important issue to consider is whether the subsequent surveys, it is important to collect all of the data. A review of specific data items may allow for some of the information collected from participants to be shortened or eliminated. It is also possible that alternate modes of data collection could be done in subsequent versions of a panel survey.

2.5 Expand Existing IRS Surveys

The IRS already conducts surveys to assess taxpayer usage of IRS services and drivers of taxpayer value in IRS services, serving as excellent sources for data on service usage. It may be possible to augment these surveys with more data on taxpayer attitudes, perceptions and social networks. The disadvantage of these surveys is that they probably could not be linked to taxpayer compliance data, since they are based on random digit dialing with

screening questions. Since response propensity would presumably correlate with compliance, such linkage would be important to defining the relationship with compliance.

In hypothesizing about the behavioral determinants of voluntary compliance, workshop participants suggested a two-tiered approach of correlating services with “intermediate” variables such as trust and attitudes (for example, “duty to pay taxes”), and then correlating these intermediate variables with compliance outcomes. Expanding IRS surveys could identify variation in taxpayer attitudes towards the IRS and government, establishing the relationship between attitudes and services usage. Such surveys would identify hypothesized regional variation in attitudes which could assist in theory building and field experiments. Existing surveys could also be expanded with information of social networks to facilitate building social network and ABM models.

2.6 Census Survey

One option for acquiring information on taxpayer usage of IRS services, satisfaction, attitudes and social networks is to leverage existing data collection efforts by other government agencies. A logical choice would be to rely on the Current Population Survey (CPS), a survey of U.S. households conducted annually by the Bureau of the Census for the Bureau of Labor Statistics. While some of the existing surveys capture service usage, the CPS would provide another source of micro-level data on the use of taxpayer services to estimate the demand for these services. The large sample size of the CPS makes it particularly useful as a way to obtain geographical (e.g., state-level) detail on the use of these services, which would support the aggregate econometrics methodology.

2.6.1 Design of Census CPS Survey

The CPS collects information on households, families and persons with respect to the income, demographics and labor force participation of household members. At present, no information is collected in the CPS on the use of taxpayer services. However, in the past the Census has occasionally supplemented the regular CPS survey instrument with additional questions on special topics (e.g., pension plan participation). The additional resources needed to collect these data is funded, in part, by the agency or organization requesting the information. It may be possible for the IRS to fund additional data collection on the use of IRS services as part of the basic CPS in a cost efficient manner. In addition, this data collection effort could be part of the occasional studies of tax non-filers conducted by the Census.

The CPS, also referred to as the Annual Social and Economic Supplement (ASEC), is a probability sample of the non-institutional population in the U.S and certain members of the Armed Forces. Information on approximately 100,000 households and over 200,000 persons is presently collected across geographical divisions (e.g., state) and certain metropolitan statistical areas (MSAs). Data are organized in a hierarchical structure, with household-, family- and person-level detail. While there is no direct information on filing status or line item tax return details in the public use version of the CPS, Census currently does provide imputed values for tax filing status as well as a variety of tax return variables based on a sophisticated tax model. The survey is conducted annually.

The CPS has been conducted monthly for over 50 years and is the official government source of information on employment and unemployment. It is conducted by the Bureau of the Census for the Bureau of Labor Statistics.

Detailed information on the income, demographics and labor force participation for individuals 15 years of age and older are collected in the CPS. Income variables include: wages and salaries, interest and dividends, business income (or loss) from sole proprietorships, rents and royalties, social security, pensions and unemployment compensation. Also included is information on transfer payments and certain types of non-cash compensation. (Note that information on capital gains and losses is not collected, although imputed values are provided.)

Demographic detail on the CPS includes: age, gender, race, educational attainment, family structure, marital status and health status. Labor force variables include: hours worked, occupation, full- or part-time status, health insurance and pension coverage and hourly earnings.

2.6.2 Census Survey Links to Administrative Data

Census data will generally contain no direct links to administrative data. However, it is relatively straightforward to construct “tax units” from CPS data that mimic the information contained on tax returns. These substitute tax returns

can be used to support the type of demand analysis required. Indeed, imputed values for filing status and various tax return variables are provided on the public use files.

Because the CPS is a probability sample of U.S. households, very little information is collected on very wealthy taxpayers that report a large proportion of taxable income. In addition, to maintain the confidential nature of the data collected, the CPS relies on top-coding of most income items. A careful treatment of high-income taxpayers will be necessary to ensure that the Census data are aligned to IRS administrative data in order for the analysis to be reliable.

There is one case where IRS administrative data are linked directly to CPS data: occasional studies of tax non-filers conducted by the Census where tax returns are exactly matched to survey data. For confidentiality reasons, these “exact match” data are not shared with the IRS. However, it may be possible to acquire selected information (e.g., regression coefficients or summary tabulations) that could support additional analysis of filing compliance. Ideally, special arrangements would be made with the Census for an independent contractor to work with the exact match data to perform analysis along the lines described in Part II Section 2.3 Microeconomic Modeling – Estimating the Impact on Filing Compliance (page 27). However, the only truly essential additional information required for such an analysis would be an indicator for whether each respondent actually filed a tax return. In lieu of giving an independent contractor full access to the exact match data, perhaps this one additional variable could be provided along with the relevant identification code for appending it to the public use CPS file. We note that an imputed variable for filing status (FILESTAT) is already available on the public use file. However, this variable is based on a simulation rather than an actual match against tax returns. As a result, it is not well suited for an analysis of actual filing behavior.

2.7 Extension of the National Research Program

Some alternative microeconomic approaches to estimating the impact of service and enforcement activities on tax compliance were presented in 2.2 Microeconomic Modeling – Estimating the Impact on Reporting Compliance (page 22). To implement any of these approaches, it is necessary to have detailed line item tax return information along with a suitable measure of tax compliance for a reasonably large random sample of taxpayers. Such data are already available from the ongoing National Research Program (NRP). For tax year 2001, this program involved audits of a stratified random sample of approximately 45,000 individual income tax returns. Most of these returns were subjected to a thorough examination, and the line item adjustments by the examiner to reported amounts on the return are recorded in the data base. Recently, a decision was made to conduct an NRP for individual returns on an annual basis, with approximately 15,000 returns randomly selected for audit each year.

2.7.1 Purpose of the NRP Extension

The NRP data will provide much of the necessary information on tax return characteristics and tax compliance behavior for the purposes of the microeconomic analyses. However, service utilization modeling requires additional data sources. Depending on which microeconomic approach is being employed, it also will be necessary to have a means of assessing either which of a set of relevant taxpayer services were actually used by each taxpayer in the NRP or, alternatively, the likelihood that the taxpayer would have used each of the services.

One of the estimation approaches described in 0 Extensions of the Basic Approach (page 24) relies on information on actual service usage by NRP taxpayers. In more recent NRP studies, examiners are now surveying taxpayers on what services were used while asking an open-ended question about other IRS services they may have used. Also, as with the original NRP study for tax year 2001, the more recent studies also contain information about the usage of third-party tax preparation services.

Since NRP taxpayers are undergoing examination, the NRP process may bias their answers to services questions. Comparison to those of a representative sample of taxpayers from a separate survey will determine if NRP and non-NRP subjects are reasonably consistent in their reporting of service utilization. Existing surveys such as the 2006 W&I Market Segment Survey may be used for comparison. If the survey responses are found to be reasonably accurate, one can proceed with the estimation methodology that relies on actual service utilization without any supplementary data on service utilization, although the key challenge for this approach would be adequately accounting for self-selection of services by taxpayers.



If the survey responses are not found to be reasonably accurate, an alternative approach to obtaining details on actual service utilization by NRP taxpayers would be to perform a survey of pre-selected taxpayers prior to informing them that they have been selected for an NRP audit. Under such an approach, one could employ a broader survey that covered topics beyond simply service utilization. For instance, taxpayers also might be asked about their attitudes about the IRS, the quality of its services, views about enforcement, and attitudes about government in general. The additional data collected through this survey approach would make it possible to estimate a richer model of tax compliance that takes into account some of the socio-psychological theories of tax compliance discussed in Part I Section 1.

Theoretical Frameworks (page 12). In this regard, such an approach would seem ideal. However, there are potential drawbacks that deserve careful consideration. Of particular concern is that some survey respondents who were pre-selected for the NRP might (incorrectly) infer that they were chosen for audit as a result of how they answered certain questions on the survey. If this view were to prevail, it could lead to an erosion of trust in the IRS and make it difficult to obtain reliable survey responses from taxpayers in subsequent years. The risk of such an outcome could be mitigated by through the survey design, in particular by surveying both taxpayers that were and were not pre-selected for an NRP audit, and by only selecting a sub-sample of taxpayers from the former group. This sub-sample of pre-selected taxpayers would need to be drawn carefully in order to ensure that it included a sufficient number of taxpayers who used each of the various services being investigated.

An alternative microeconomic estimation approach described in 2.2 Microeconomic Modeling – Estimating the Impact on Reporting Compliance.(page 22) is to perform an analysis of the likelihood that a given taxpayer would use various types of services from a survey of a separate sample of taxpayers. Based on the results of this analysis, one would then impute to each return in the NRP the likelihood that each of the services was used.

2.7.2 Links to Administrative Data

Ultimately, the data on audit risk and service utilization needs to be linked to the NRP data. In some cases, a direct link will be possible. For instance, if the responses to the questions on service usage now being asked during NRP examinations are deemed to be reliable, this information will be directly available on the NRP data base. Similarly, responses to a survey of taxpayers pre-selected for an NRP audit could be readily linked to the NRP data base. In other cases, the probabilities of operational audit and of service utilization would have to be imputed to the NRP using estimated econometric equations derived from the supplementary data.

2.7.3 Data Design

The supplementary data on audit risk would include random samples of audited and unaudited tax returns from the relevant NRP tax year(s). A choice-based sample that included a disproportionate number of audited returns would be ideal for this purpose. The data sample should include a sufficient number of returns from each examination class, and it should include a set of sample weights that make the overall data sample representative of the underlying return population.

An independent survey on taxpayer service utilization should be performed using a representative random sample of taxpayers. The sample should be stratified to ensure that a sufficient number of taxpayers are included who used each of the various services under consideration. Ideally, the sample would include a substantial number of taxpayers who were pre-selected for an NRP audit. The Census-based survey described in Part II 2.6 Census Survey (page 61) provides an example of an appropriate instrument. Alternatively, an existing survey such as the aforementioned benchmark Opinion Survey of Taxpayer Resources and Services or the 2006 W&I Market Segment Survey might be adequate.

At a minimum, the a survey on service utilization should solicit information on a variety of taxpayer and tax return characteristics as well as indicators for what services were used by the taxpayer. Ideally, they should also include information on taxpayer satisfaction, perceptions, and attitudes, as well as the reasons for using the services. As discussed 2.2 Microeconomic Modeling – Estimating the Impact on Reporting Compliance.(page 22), a key challenge for this microeconomic approach would be to avoid problems with multicollinearity between the measures of the likelihood of service utilization and the other explanatory variables in the model. It therefore would be useful to identify one or explanatory variables for service usage (instruments) that do not also belong as separate explanatory variables in the specification of compliance behavior. An example of a possible instrument would be a measure of the accessibility of the service in the taxpayer's geographic area.

To incorporate social interaction effects into the econometric analysis, it would be necessary to be able to have a means of assigning taxpayers to relevant social groups and, for each group, to have relevant data on the utilization and quality of taxpayer services.



Part III: Integrated Plan

1 Conceptual Underpinnings

The first and second part of this report discussed the potential methodologies and data collection respectively. Part III of the report will bring together those methodologies and data collection efforts into a notional plan and will provide recommendations, but before presenting our plan and recommendations, we'll begin with a discussion on the conceptual underpinnings which formed the basis for our recommendations.

For the IRS, the ultimate objective is to provide valuable “intermediate” insights into the relationship between services and compliance while pursuing research with a high likelihood of success in estimating direct and indirect marginal effects of IRS activities on compliance. Meeting this goal must be accomplished with limited resources, acknowledging varied likelihood of technical success, interdependencies among methods, and data needs both common and specialized.

Every method is expected to contribute significant insights and short-term immediate benefits, but the primary methodologies, discussed in Part I of this report, are viewed as having the greatest potential in answering the primary research question. Since every methodology has potential benefits as well as salient limitations and risks of success, in order to have the greatest chance of successfully addressing the primary research question, this plan incorporates a diversified portfolio. This portfolio seeks to integrate a variety of methodologies and data collection efforts as a means of cross-validating results and to mitigating risk.

While an ideal plan may invest in all the methodologies and data collection efforts which show promise, real world constraints may limit investment in all recommended methodologies. In this case, choices between methodologies would then need to be made. Part I Section 4 Interrelationships Among Methodologies and Data Needs (page 44) clearly articulates the dependencies between the methodologies, and Part II Section 1 Methodology Data Needs (page 47) discusses common and special data needs, so that contingency planning may proceed from a well informed position.

Part III, Research Plan, recommends how to manage the portfolio and provides a research plan. Section 2.1 describes our recommended management approach, in which the research strands should be seen as complementary yet competing components of a larger portfolio. Section 3 Notional Timeline and Staging (page 71) shows the notional timeline for the methodologies and data collection and stages these efforts over time. As we recognize the importance of intermediate outcomes on the way to establishing the ultimate goal of quantitative estimates, we discuss then some intermediate outcomes in Section 4 Intermediate Outcomes (page 77).

2 Research Portfolio Management

The culmination of all of the analysis and information contained in this report is the plan, created within the IRS, to invest human resources in working on answering the research questions posed above. This section describes one strategy for reaching those objectives. A more basic message, however, is that the organization needs a strategy and that this strategy should be based on both sound research and management principles. The research principles have been obtained from a variety of experts in the field, as well as deeply knowledgeable staff within the IRS. This section is designed to bring together the technical information and a set of management principles to drive the strategy.

The following set of core management principles is proposed:

- Use organizational priorities to drive decision making
- Plan for and manage risk
- Scope the strategy to be consistent with available resources and time
- Leverage existing resources and capabilities to the maximum extent

While estimating marginal effects is the most important focus of this particular report, the Service must manage many other related goals that are important now or could become important in the future, so the broader range of applications or benefits need to be considered when creating the research plan. Combining the need to meet a broad set of objectives with the additional need to manage risk, it is logical that a diversified set of methods and approaches should be undertaken in what might be called a “portfolio” approach. Portfolios, similar to the concept of a financial investment portfolio, have the benefit that they attempt to reach a broad set of objectives and accommodate a range of contingencies, or risks, which might take place into the future. In this context, the contingencies and risks include such things as the IRS funds expended without payback or the unknown requirement that the Service may want to answer a more diverse set of questions at some time in the future.

A consideration in any research approach is that it must be reasonable in terms of the budget that is likely to be available. Thinking of the organization and research perspective mentioned above, it must be considered from the perspective of whether the Service has the resources available now and into a reasonable future to sustain them.

The final principle relates to organizational synergies. As research activities are going on in various parts of the IRS, for example the services research that has been undertaken within the Wage & Investment Operating Division, it is important to capitalize on common efforts. This principle could relax tight resource constraints.

The sections that follow will continue discussion of the portfolio management approach suggested in this introduction. However, it is important to point out that this is not the only options for organizational priorities for the Service. If the Service was to develop a different set of organizational priorities, the research plan would require revision.

2.1 Managing a Portfolio of Methodologies

The concept of the portfolio approach was introduced in Section 6.1. The basic idea is to support several parallel, but clearly interconnected, research threads at the same time. Those areas cover all of the primary modeling and estimation methods discussed in this document and some of the data collection and aggregation plans. The plan would include all of the following at various levels of investment:

- Aggregation and warehousing of existing IRS administrative data on services
- Primary data collection to support obtaining more information on services
- Theoretical framework development
- Field and laboratory experiments
- Econometrics (both aggregate and micro)
- Simulation methods, in particular agent-based models (ABM)

The input from the workshop participants indicated at least some support for all of these methods, though the field experimentation and econometrics methods had the strongest voices for inclusion. There is also a factor of timing in

the thought process here. Experimentation, in particular laboratory experiments but also to some extent field experiments, can bring results fairly quickly. One could expect results on lab experiments in months, while some results for field experiments could be seen within a year or at most two years. At the other extreme, the proposed aggregate econometric approach, because of the need to build a more robust information base of historical data, could require up to ten years before enough data is in place to support firm estimates. Results from the proposed microeconomic approaches would likely come sooner, perhaps within four years.

The situation with the ABM is less clear. There is some hope to expect information about relative dynamics within the population within one to three years. However, it will be much longer, and possibly decades, before reliable numerical estimates can be obtained from this approach. However, the strength of this approach lies in supporting the other methods, with a test bed to confirm hypotheses, a source of ideas for theoretical development, and a place to understand the transient and emergent behaviors not well-studied with equilibrium-solution models.

In discussing the portfolio support, we'll look at both the advantages and the disadvantages of this approach in the following sections.

2.2 Advantages of the "Wide Portfolio" Approach

The advantages of the portfolio plan are summarized below:

The approach contains several methodologies that have a reasonable likelihood of producing useful estimates of the marginal effects of at least some services on voluntary compliance

Of the various approaches considered at the workshop, the proposed field experiment, microeconomic, and aggregate econometric approaches appear to have the best potential for producing useful quantitative estimates of the marginal effects of services on compliance. The field experimentation method, in particular, is viewed as having a very good chance of producing useful estimates for at least some services, and may be capable of producing such estimates in a relatively short time frame. The potential for the proposed micro and aggregate econometrics methodologies is not as certain, largely because they must overcome some significant technical challenges. They also are likely to require a longer time horizon to produce meaningful estimates, largely because of data collection requirements. In general diversification tends to minimize risk and allow for synergy between methods.

The other methods are likely to improve our understanding of the service-enforcement-compliance relationship

While the other methods might not produce quantitative estimates of the impacts of services on compliance, they are very likely to enhance the general knowledge of taxpayer behavior, provide a fertile ground for idea generation to support theoretical framework development and the generation for hypotheses for experimentation.

Having a range of methods brings more opportunity to obtain intermediate outcomes

Short term successes would bring credibility and support to the program. The recommendations include intermediate benefits from laboratory and field experiments, as well as direct data collection and aggregation activities. Also, having an agent-based modeling platform can allow for "trying out" a new idea quickly in the established platform. The Service has already invested in several ABMs and, therefore, has a start in creating a test bed environment for such tools. For discussion of these activities see papers by Bloomquist (2004) and Carley (2006) in the reference list.

Investing in a wide number of methods is more likely to yield credible quantitative estimates

The likelihood of success, even for well-established primary methods, is uncertain. Additionally, the supporting methodology, ABM does not appear to have the likelihood of producing quantitative estimates in the short run, yet may produce estimates in the future as data driving the model becomes better over time.

Embarking on multiple methods improves the chances of discovering new and better approaches

Any method alone can lead to other ideas or methods that could be beneficial to the goals of the IRS. It follows that if one uses many approaches the chance of finding new ideas improves. Combinations of methods, in particular

econometrics methods with experimentation, could produce stronger estimates than a simple method alone. The portfolio approach allows such synergies to be sought after and leveraged.

The portfolio provides the flexibility to adjust research over time as new information is learned

The last advantage of the portfolio approach is that one always has the opportunity to overcome an ill-advised decision by instituting a periodic review process. This process would have to be informed by tracking progress on each methodology through a performance reporting and review system. In some cases, performance can only be measured by completion of interim milestones. In other cases, it might take more than a year to see research progress, but that should be part of the understanding laid out at the beginning of the plan. If during one of the review cycles, evaluation suggests that one or more of the methods are not performing at the level expected (even if that is because milestones are not being met on time), it is possible to restructure the research plan to allocate the remaining resources differently between portfolio projects in order to achieve a better outcome for the Service.

Disadvantages of the “Wide Portfolio” Approach

Some resources may be spent on methodologies and data sources that turn out to be unproductive

While maintaining a diverse portfolio of approaches is mentioned as a strength above, it is also possible that some of the methods that are selected to be pursued will not work out. In that case, some of the resources will have been spent from which no productive results were obtained. This is, in fact, a fairly common occurrence within research organizations. It would be unduly optimistic to expect all initially promising research paths to turn out to be productive.

It is more complex to manage multiple projects simultaneously requiring both more intense management and broader knowledge of the research areas

More than one project requires stronger management and more communication including cross-project communication and visibility. This fact makes using the aforementioned established management methods more important. Of particular importance are setting clear goals and objectives, defining interim results and reporting milestones, having periodic program and research reviews, and reconsidering the allocation of resources in the research portfolio in light of those reviews.

Resource requirements will be higher

More projects are almost certain to require a higher level of investment than a single project. In particular, because as mentioned above, some of the projects might not produce positive results, some of the resources may be spent without direct payoff.

2.3 Portfolio Management Recommendations

Considering our current understanding of the Service’s needs, we are recommending the portfolio approach. In this method each of the methods that were mentioned in Section 11.1, are recommended to be included in the plan. The exact mix of those methods would require further knowledge of resources and priorities by IRS Research, Analysis and Statistics (RAS). Those deliberations will be informed by much of the information included in this report.

As described in the prior sections, the complexity of the research plan is such that careful management controls are required. Ongoing evaluation and guidance is necessary to periodically readjust the plan, investing more or less in projects as a given approach is successful or falters. It is important to start out with a plan, and we have included a notional plan in the next section. While this plan lays out a projected level of effort into each methodology over the course of the next five years, it is being proposed as a starting point for planning and evaluation by the Service in light of resource availability and other guidance from IRS management.

In other projects of this type, a leading practice is to form an oversight committee to provide both technical and management input and to, in general, oversee the Indirect Effects project. It is one of our recommendations that such a committee be formed in this situation as well. To best use this committee, the direct managers of the portfolio should collect and archive information on progress of each research project included in the portfolio so that this



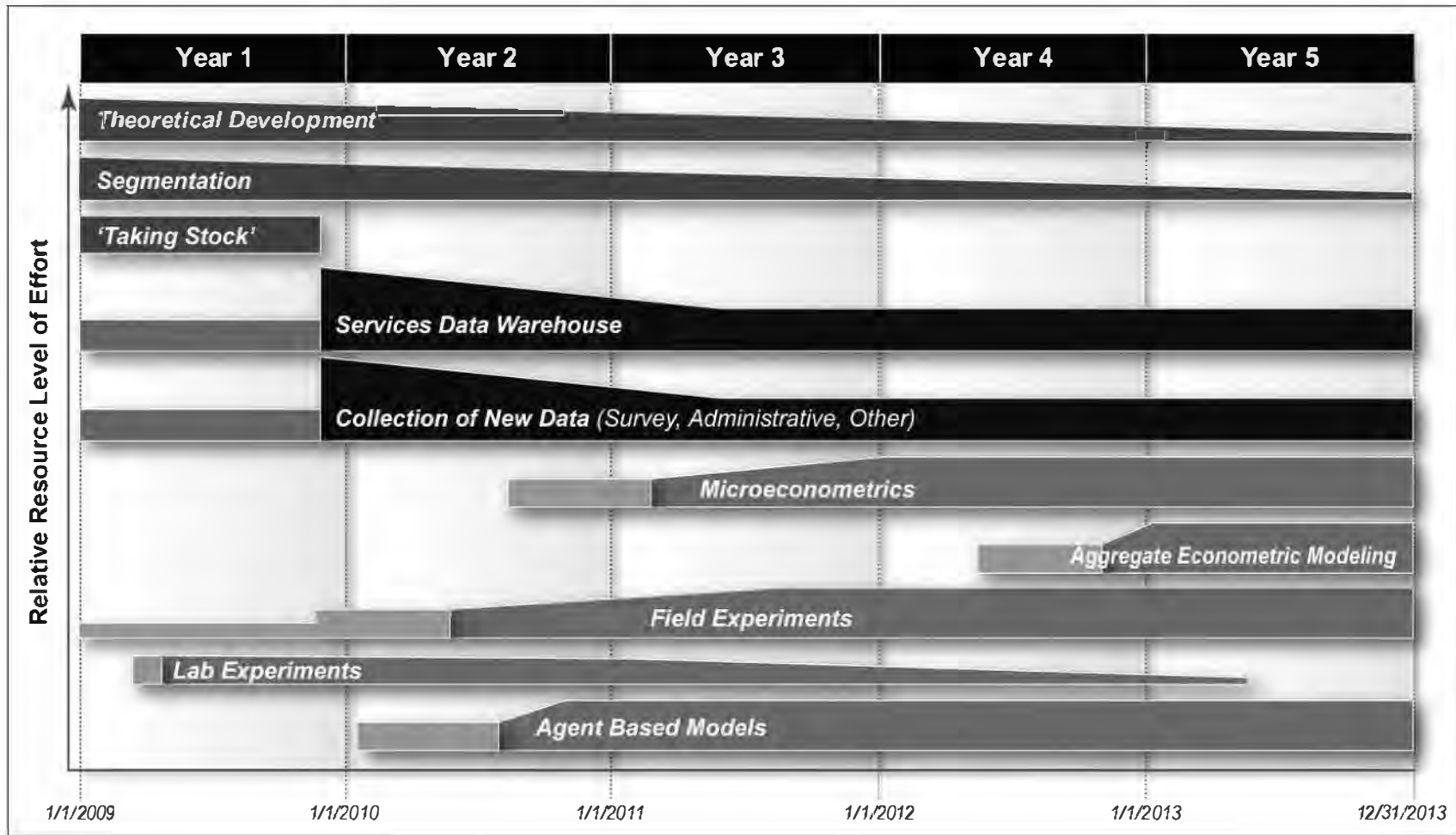
information can be presented to the oversight committee in a structured and useful way. This will allowing the committee to evaluate the status of the data collection and methodologies and to make recommendations back to IRS management whether efforts in a particular area should be increased, decreased, or remain constant.

3 Notional Timeline and Staging

To actualize the vision of an integrated research portfolio, the IRS should develop a scalable research plan with a portfolio of methodologies as described above. This section brings together both research methodologies and data acquisition plans and layers them into a notional schedule. The plan is notional because setting expenditure levels and other specifics of the plan, such as fixing on a set of organizational priorities, were not included in the scope of this project.

The notional research timeline, Notional Integrated Timeline, lays out the various theoretical developments, methodologies, and data collection activities across the span of five years. For ease of explanation, the theoretical developments, methodologies, and data collection activities are shown in different colors (green, brown and blue, respectively). The width of the bar indicates the level of effort required. As the bar increases or decreases, this indicates a corresponding change in the level of effort needed over time. The lighter shading at the beginning of the data collection and methods sections indicate a period of planning which is necessary before implementation and analysis. Below is a discussion of the plan for each individual methodology, data collection, and theoretical development activity as it is laid out across time.

Figure 6: Notional Integrated Timeline



3.1 Theory (Green on Chart)

Theoretical models that link taxpayer services and enforcement activities to compliance behavior can serve as an important guide for empirical studies. The plan therefore calls for a relatively significant early investment in theoretical modeling on the roles of IRS-provided taxpayer services and substitute services in compliance. The objectives of this early research would be to generate insights into pathways through which services and enforcement activities are likely to impact compliance; identify key variables and parameters for empirical analysis; and produce testable hypotheses about taxpayer behavior. Later on in the process, the emergence of empirical results is likely to spur further theoretical developments as some hypotheses are confirmed and others are overturned. Investments in theoretical research should, therefore, begin at a relatively high level and then ramp down to a constant level throughout the rest of the five years.

Segmentation analysis is also likely to play an important role in guiding research. The plan calls for early investments in such analysis to generate insights into sources of taxpayer heterogeneity that can inform development of other methodologies and business decisions. Segmentation analysis would initially involve analysis of existing IRS administrative and survey data, but would also exploit new sources of data as they become available. Segmentation analysis involving compliance data may leverage or extend analysis done for the tax gap, examining nonfiling matches, NRP data, and collection data. Initial segmentation analysis involving taxpayer motivations for using services may involve an examination of return characteristics, demographic information, and service usage patterns from existing services surveys, master file or Statistics of Income (SOI) data sources. Later, it may tap data from new surveys on attitudes/motivations, beliefs, and opportunities for noncompliance.

Since segmentation analysis is used to guide the development of both theoretical and empirical models, this activity should be among those that are done at the beginning of the research plan. The first year begins with the greatest amount of effort in segmentation and decreases throughout the five years. While most of the effort would occur in the beginning, a sustaining level would be desirable to support updating categories as the population changes and tax laws/forms change affecting the prevalence of various compliance issues.

Another important activity for which the plan calls for relatively significant early investment is referred to as “taking stock.” By this we mean a project focused on understanding and enumerating existing sources of IRS data and research in order to ensure that efforts are not duplicated. Within this step, future data needs for research methodologies should also be examined.

3.2 Data Collection (Blue on Chart)

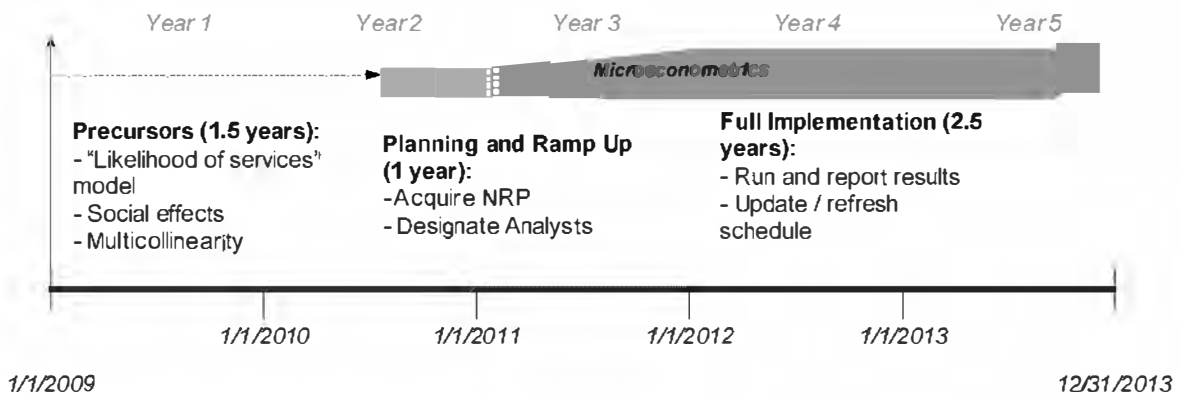
Moving forward on implanting the creation of a Services Data Warehouse would require planning for the first year to establish which data would be imported and the manner in which the data would be stored. This is part of the taking stock process discussed above, but would go into further depth in order to plan the data warehouse. Following the planning stage, the next year or two would be spent implementing the data warehouse. This implementation stage would include determining the right IT platform on which the data would reside importing the appropriate data. This phase is seen to be the most intensive both in terms of personnel as well as equipment cost. After the data warehouse has been completed, less effort would be required in the subsequent years to maintain the data.

While the Services Data Warehouse provides a manner in which to store and easily access existing data, effort must also be placed in collecting new data. This data collection, which includes both surveys and additional administrative data, is currently seen as having a year of planning, followed by increased effort as data collection is underway, and then a maintenance phase in which similar data collection techniques would be done every year and would therefore require less effort than the first time. Collecting new data may come in the form of a new taxpayer survey or it may be possible to piggy back on an already existing survey, allowing for some efficiencies of implementation. In this case, the additional cost of this survey is minimal and the level of effort required by personnel is also relatively small. The relative level of effort would remain about the same as it did during the planning phase and would not increase as the timeline above shows.

3.3 Methodologies (Brown on Chart)

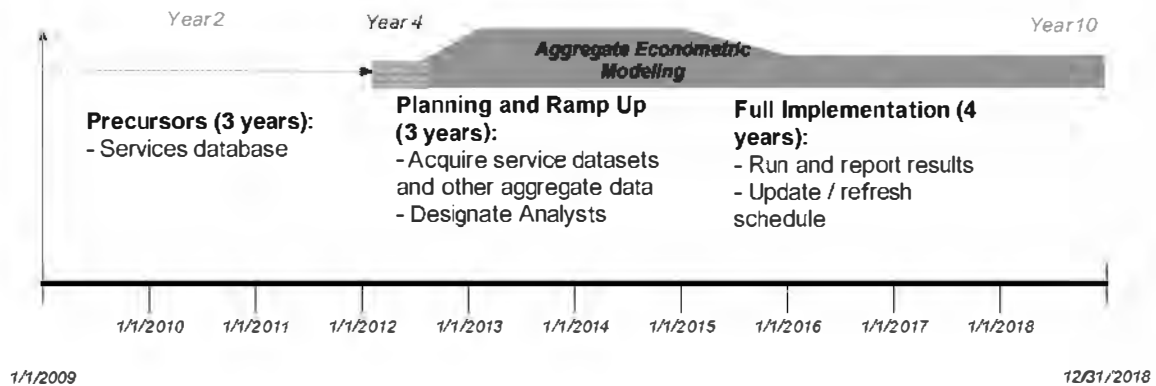
The proposed microeconomic approach for analyzing the compliance impact of service and enforcement activities requires NRP data that has been augmented with either simulated or actual service utilization variables. To incorporate simulated variables for service utilization in the NRP, one must first construct a model of the probability of services usage using data from an appropriate survey. Information on actual service utilization is now being collected during ongoing NRP studies by surveying taxpayers during the audit process. Once this information becomes available, the plan calls for an analysis of the survey responses to assess whether they are reasonably reliable. As a result of the time needed to develop a model for simulating service utilization or obtain and assess NRP data that includes information on actual service utilization, microeconomic approaches do not start until year two and don't ramp up until year three. The theoretical development discussed in the previous section would occur prior to econometric modeling and potentially aid the design of the models.

Figure 7: Microeconomic Modeling Notional Schedule



The proposed aggregate econometrics methodology requires panel data on reporting, filing, service utilization, and enforcement aggregated by state that spans approximately 10 years to produce meaningful results. Reliable historical data on service utilization is not generally available for many services, which means that such data will need to start being collected as the project moves forward. Consequently, planning for this approach starts in the fourth year, later than all the other approaches. While planning will begin around the fourth year, firm results are unlikely to be available until later years. Most of the model building and preliminary specification testing will occur between year four and year eight, after which less effort is required to implement the methodology and update estimates as new data become available.

Figure 8: Aggregate Econometric Modeling Notional Schedule



Implementing field experiments would also require an advance planning activity. In this planning activity, the research team would both search for natural experiments and define a set of designed field experiments that are

desired to be done. The search for natural experiments would include capitalizing on differences in IRS activities across similar taxpayer groups, such as the example provided in the field experiments section where two different warning letters were issued by the IRS regarding underreporting of income. With these natural experiments a researcher could analyze data from an already existing situation. Otherwise field experiments will have to be designed. Which field experiments should be conducted will rely on segmentation and diagnostics to decide which research questions to address and for which taxpayer groups. Having a solid understanding taxpayer segmentation is a critical requirement to be able to define useful treatment and control groups within the field experiments. After this initial planning stage, field experiments would continue through the next few years at a similar level. There may be some field experimentation which could be relatively inexpensive while others may be quite costly and require significant investment. As field experiments have the potential to quantitatively measure the effect of IRS services on compliance, they are important to continue throughout the research plan. Lab experiments would require minimal planning in order to set up. They are early candidates for testing behavioral responses, relatively inexpensive, and less reliant on data. Ideally, lab experiments would be paired with theoretical development to provide a means for testing theoretical hypotheses. In addition to aiding in the evaluation of theoretical models, lab experiments might generate results that would be useful in ABM design. Investments in lab experiments would remain at a constant level for a year or two and then taper off as other methodologies become more prominent.

Figure 9: Field Experiments Notional Schedule

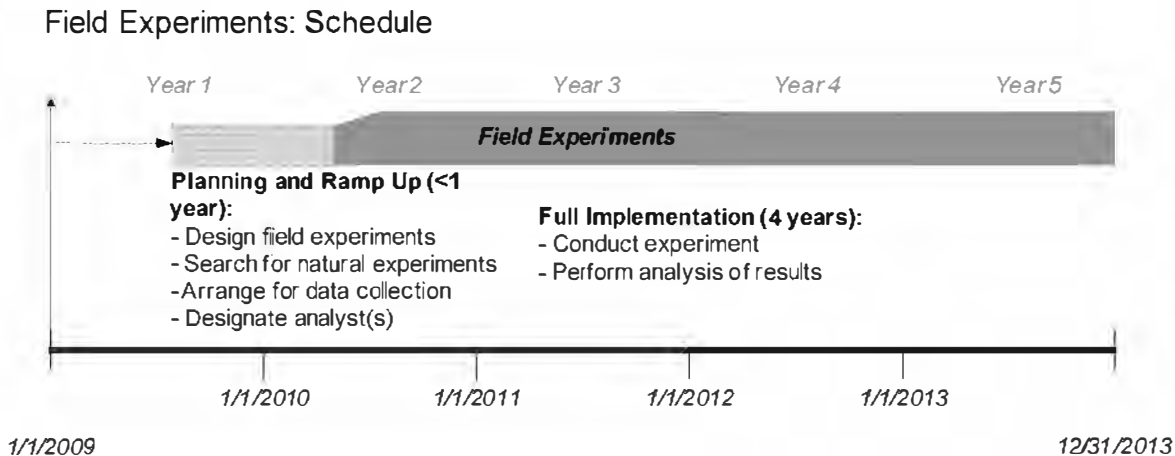
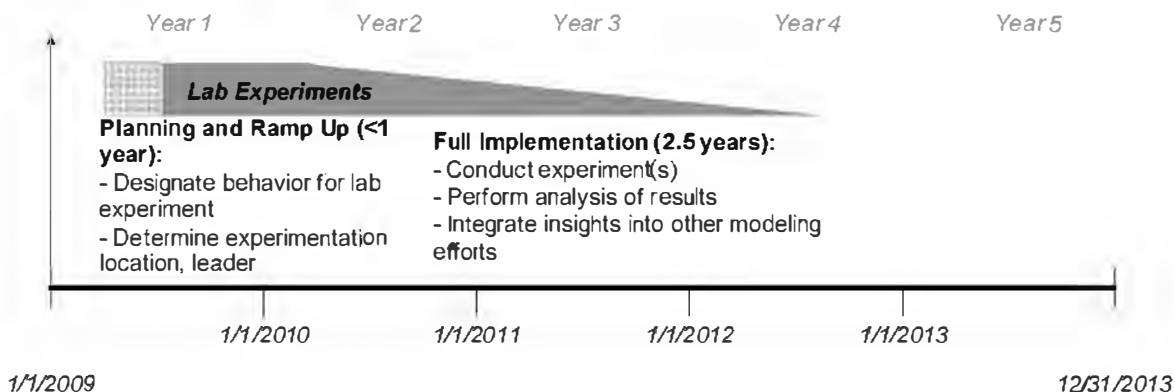


Figure 10: Lab Experiment Notional Schedule

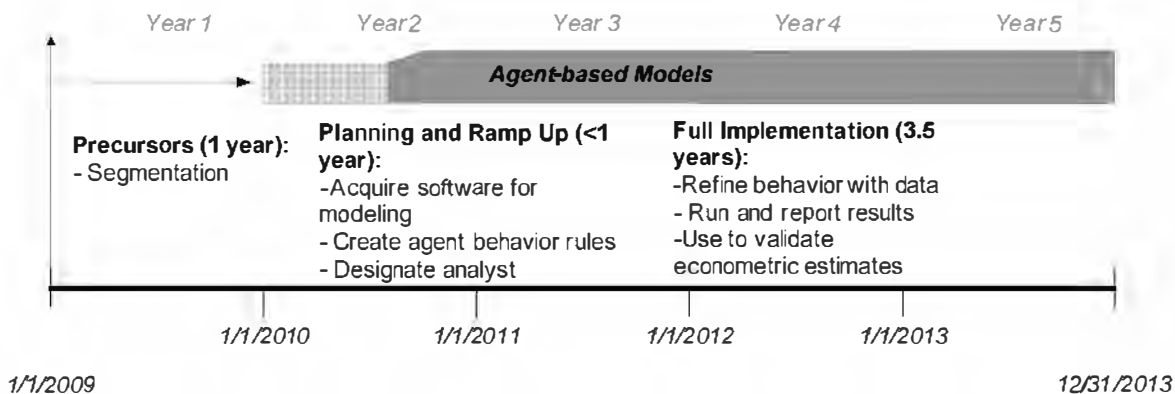


In the case of agent-based modeling, it is generally accepted that data requirements with respect to behavioral assumptions and underlying population demographics are massive. Therefore it is common for such models to begin with assumptions for the behavioral rules. Then, as hard data to support those rules become available they are replaced by behavioral rules and related probabilities based on observation. Because of the size of this challenge, we

are proposing that this planning for this methodology be postponed for about a year to allow the acquisition and organizing of segmentation information. This would allow for a planning activity for the ABM approach to begin in the second year, and model development could start within the second year of the plan. The effort should ramp up as data becomes more available and more credible. ABM is an important companion to econometrics because it focuses on areas where current methods are weakest (e.g. transient behavior, what if analysis, and the discovery of emergent behaviors).

Figure 11: Agent-Based Modeling Notional Schedule

Agent-based Modeling and Social Networks: Schedule





4 Intermediate Outcomes

While the goal of this project is to measure the direct and indirect effects of services on compliance, the primary methodologies potentially capable of doing so may take several years or more to produce estimates. Meanwhile, research methodologies and data collection efforts will provide significant intermediate benefits to the IRS as well. Table 6: Intermediate & Long Term Benefits of Methodologies and Data Collection shows intermediate benefits of the data collection efforts and methodologies discussed in the notional timeline. Most substantially, methodology development will widen understanding on taxpayer compliance behavior, service use, the influence of services on taxpayer attitudes, perceptions, and motivations, the mechanisms through which services influence general behavior, why taxpayers use the IRS or substitute services and how this choices influences compliance behavior.

Table 6: Intermediate & Long Term Benefits of Methodologies and Data Collection

	Immediate Benefit
Data Collection	
Services Data Warehouse	<ul style="list-style-type: none"> • Provide an accessible, standardized repository for services data for research, supporting other services projects • Collect services data to support future research not yet identified
Survey Data	<ul style="list-style-type: none"> • Facilitate segmentation • Create a services demand model for more efficient sampling of services taxpayers in the future • Comparison between NRP and W&I services will allow insights into whether taxpayers undergoing examination appear to have different characteristics
Methodologies	
Theoretical Frameworks	<ul style="list-style-type: none"> • Improve understanding of services and compliance behaviors; the relative mechanisms of “carrots” and “sticks” and the role of intermediate variables
Segmentation	<ul style="list-style-type: none"> • Improve understanding of services and compliance groups and responses
Aggregate Econometric Approaches	<ul style="list-style-type: none"> • Data analysis on the use and distribution of IRS services - for example, descriptive statistics on trends in tax reporting, filing, audit coverage, service utilization and service quality by geographic region or other taxpayer segments
Microeconomic Reporting Compliance	<ul style="list-style-type: none"> • Models of the demand for taxpayer services and the likelihood of audit • Work could begin on developing the outlined methodology into a full-fledged econometric framework for estimating behavioral responses with the NRP data.
Microeconomic Filer Study	<ul style="list-style-type: none"> • Non-filer: estimates of service impacts on filing behavior • Aid understanding about the general characteristics of nonfilers and factors other than service quality that impact one’s filing decision.
Field Experiments	<ul style="list-style-type: none"> • Initial estimates of the impact of certain studied service variation on tax compliance with “real world” results
Lab Experiments	<ul style="list-style-type: none"> • Validation of directional impacts of services and compliance interactions
Agent-Based Models	<ul style="list-style-type: none"> • Test bed in which to examine social networks and the impact of certain assumptions, until more detailed data becomes available • Enhance our theoretical understanding of the tax “ecosystem” in which taxpayers interact with each other, services and substitute services

One of the more tangible benefits will be the data collection in support of services utilization. A “Services Data Warehouse” or similar repository would enhance and organize services data in a comprehensive framework. A services data warehouse would link the currently disconnected data in a meaningful way thus facilitating and encouraging analysis by researchers across the Service.

Services demand analysis, theoretical development and segmentation would also enhance our understanding of services and compliance groups and behaviors. Exploring existing surveys, such as the 2006 W&I Market Survey Report and the new NRP services survey questions, would allow for statistical prediction of services usage and define taxpayer segments along the compliance dimension. While segmentation of taxpayers is a first step in many of the methodologies to measure the effect of service on compliance, it also has an immediate benefit in aiding the understanding of taxpayer groups and some of their basic connections to service (higher income taxpayers are more likely to use paid preparers, lower income are more likely to visit a tax center, younger taxpayers are more likely to use internet, etc.).

Preliminary results from primary methodologies, such as econometrics and field experiments, could improve understanding of taxpayer compliance. While final estimates from these methodologies may take years, preliminary results could be available within a few years and could provide interesting insight for improved theoretical understanding and empirical study design.

While the supporting methodologies may not answer the primary research question, they can help to test directional impacts of compliance strategies. For example, lab experiments could begin to shed some light on impact of services through an experiment where some subjects have access to service and others do not or an experiment in which the quality of service is altered. These experiments would generate preliminary results relatively quickly and inexpensively as to whether the access to and quality of service impacts compliance. Exploiting an ABM test bed environment for services promotes



Part IV: Conclusion and Recommendations

To summarize, this report makes the following recommendations:

- I. The IRS should invest in a **dynamically managed portfolio of methodologies**.
 - a. The IRS should manage this portfolio through a research oversight committee, with **periodic performance evaluation** and appropriate revisions.
 - b. The IRS should prioritize those methodologies with the best chance of success in estimating the marginal direct and indirect effects of services and enforcement activities.
 - c. The IRS should begin an assessment of all relevant service and compliance data to **catalogue existing knowledge of taxpayer compliance behavior and attitudes**, culminating in a final report.

- II. The IRS should invest in **foundational strategies early on**.
 - a. The IRS should **facilitate the development of theoretical** frameworks to better understand the relationship between services and compliance.
 - b. IRS should distinguish **relevant taxpayer segments from a compliance perspective**.
 - c. IRS should judiciously **test theoretical frameworks in the laboratory environment**.

- III. The IRS should invest in a **strategy that yields desired intermediate outcomes** as well as the final objective, **the measurement of marginal indirect and direct effects**.
 - a. Intermediate benefits include a **better understanding of taxpayer services usage, attitudes and compliance behaviors** before quantitative estimates will be available.
 - b. It is desirable to measure **impacts on intermediate variables** that are likely to improve compliance (such as taxpayer satisfaction) before estimates of the impact on compliance are likely to be available.

- IV. To meet the data needs, the IRS should **leverage existing data and collect new data**.
 - a. The IRS should **consolidate and standardize services data** in a common repository accessible to IRS researchers.
 - b. The IRS should examine existing surveys to determine what data are available on taxpayer attitudes, networks, behaviors and perceptions and consider **augmenting existing surveys or facilitating new ones**.

- V. Once appropriate data and insights are in place, begin investing in the measurement of indirect and direct effects.
 - a. In order to measure the marginal direct and indirect effects of services and enforcement activities, **the IRS should invest in carefully controlled field experiments**.
 - b. Simultaneously, the IRS should **pursue both aggregate econometric methods and microeconomic methods**.
 - c. The IRS should assess existing **social network/ABM models** or develop new ones for the purposes of representing social networks and third parties in compliance.

References

- Adelsheim, P.D. and Zanetti, J.L., The effect of targeted outreach on compliance. *IRS Research Conference*, 2006. Washington DC: Publication 1500.
- Allingham, M.G. and Sandmo, A., Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1972. 1: 323-338.
- Blume, L.E. and Durlauf, S.N., Identifying Social Interactions: A Review, *SSRI Working Paper #2005-12*, University of Wisconsin.
- Blumenthal, M., Christian, C., and Slemrod, J., Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, 2001. 54 (1): 125-136.
- Carley, Kathleen M., Simulating EITC Filing Behaviors: Validating Agent-Based Simulation for IRS Analyses: The 2004 Hartford Case Study in Annual Report to Congress, Vol 2: http://www.irs.gov/pub/irs-utl/arc_2007_vol_2.pdf.
- Dixon, Diane M., IRS Seeks to Develop New Web-Based Measurement Indicators for IRS.gov, published for SOI in 2003. Access date: October 10, 2008.
- Dubin, J.A., Criminal investigation enforcement activities and taxpayer noncompliance. *Public Finance Review*, 2007. 35 (4): 500-529.
- Dubin, J.A., Graetz, M.J. and L.L. Wilde, The Effect of Audit Rates on the Federal Individual Income Tax System, 1977-1986. *National Tax Journal*, 1990. 43 (4): 395-409.
- Frey, B., *Not Just for the Money: An Economic Theory of Personal Motivation*, Cheltenham, UK: Edward Elgar, 1997.
- Gordon, J.P.P., Individual morality and reputation costs as deterrents to tax evasion. *European Economic Review*, 1989. 33(4): 797—805.
- Greene, W.H., *Econometric Analysis*. Third Edition, Prentice-Hall, Inc., New York, NY, 1997.
- Hasseldine, J. et al., Carrots, sticks, sole proprietors, and tax accountants. *IRS Research Conference*, 2005. Washington DC: Publication 1500.
- IRS Oversight Board, The 2007 Taxpayer Assistance Blueprint Phase 2. 2007: Department of the Treasury, Internal Revenue Service (<http://www.irs.gov/pub/irs-pdf/p4579.pdf>).
- Klepper, S., Mazur, M. and Nagin, D., Expert intermediaries and legal compliance: The case of tax preparers. *Journal of Law and Economics*, 1991. 34(1): 205—229.
- Maddala, G.S., *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press, 1983.
- Mete, M., Bureaucratic behavior in strategic environments: Politicians, taxpayers, and the IRS. *The Journal of Politics*, 2002. 64(2): p. 384-407.
- Myles, G.D. and R.A. Naylor, A model of tax evasion with group conformity and social customs, *European Journal of Political Economy*, 1996, 12(1): 49—66.
- Plumley, A.H., The Determinants of Individual Income Tax Compliance: Estimating the Impact of Tax Policy, Enforcement and IRS Responsiveness. *Internal Revenue Service. Publication 1916 (Rev. 11-96)*, Washington, DC, 1996.



Rosage, Laura, The Evolution of IRS Telephone Quality Measures. <http://www.irs.gov/pub/irs-soi/04rosasa.pdf>. Accessed October 21, 2008.

Scotchmer, S. and Slemrod, J., Randomness in tax enforcement. *Journal of Public Economics*, 1989. 38 (1): 17—32.

Tesfatsion, L. and Judd K. L., *Handbook of Computational Economics, Volume 2, Agent-Based Computational Economics*, Elsevier: Amsterdam, 2006.

Wasserman, S. and Faust, K., *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994.

Watts, Duncan J. and Newman, Mark, *The Structure and Dynamics of Networks*. Princeton, Princeton University Press, 2006.

Appendix A: Glossary of Terms

A.1 Glossary of General Terms

- 1.) **Services** – Assistance provided by the IRS to taxpayers through various channels including: internet/web, phone centers, field centers, outreach programs, and media/publications. Services can be pre-filing taxpayer initiated services, such as downloading a form or calling an automated Q&A service with inquiries about tax law, or post-filing, in response to an enforcement activity by the IRS. Additionally, the Taxpayer Advocate Service provides advocacy for taxpayers. Appeals and Tax Court activities can be considered “services” available to support taxpayers in obtaining legitimate tax relief from the tax authority; however, these activities also entail an element of enforcement, so they are not strictly considered services.
- 2.) **Substitute Services** – Services providing tax assistance offered by organizations and individuals other than the IRS, known as “partners.” Some common partners include: (1) groups that provide outreach or low (or no) cost return preparation to specific taxpayer groups, such as elderly, low income and of limited English proficiency, including VITA an IRS-sponsored organization; (2) commercial tax preparers, such as Attorneys, Certified Public Accountants, Enrolled Agents, and Enrolled Actuaries, as well as unenrolled tax professionals, such as those offering walk-in services, from H&R Block to less standardized purveyors; and (3) tax preparation software, such as from Intuit or H&R Block, tax newsletters, and various other non-IRS tax-related publications.
- 3.) **Enforcement** – Generally postfiling activities by the IRS that are designed to ensure and compel a particular taxpayer’s compliance with his tax obligations under the law. Examination/audit and matching programs are designed to identify noncompliant taxpayers for enforcement. Collection and criminal investigation may result in additional financial penalties, seizures, indictments and sentencing for noncompliance.
- 4.) **Compliance** – Taxpayers generally have three basic obligations, which define three types of compliance:
 - a. Filing Compliance: filing a required tax return on time
 - b. Reporting Compliance: disclosing all required information on a return, and self-assessing the correct amount of tax
 - c. Payment Compliance: paying on time the amount of tax self-assessed
- 5.) **External Factors** – Factors other than IRS activities that influence compliance behavior. These could include demographics, determinants of income (such as the current state of the economy), characteristics of the tax system beyond IRS control (such as tax complexity and law changes), social perceptions of the government in general and the IRS in particular.
- 6.) **Direct vs. Indirect Effects** – A direct effect is the tax change generated by a direct IRS contact with a particular taxpayer, which could be either a service contact or an enforcement contact. For instance, a taxpayer who calls the IRS toll-free number, gets his question answered, and then files his return accurately may have changed his return based on that contact. Likewise, when a taxpayer is audited, any additional amount of tax assessed would be a direct effect of the audit. Indirect effects are the secondary or subsequent results (sometimes called “spillover effects”) of the service/enforcement activity. Indirect effects are always impacts on *voluntary* compliance—without any direct IRS contact for the tax year in question. An example of this would be when a taxpayer calls the IRS toll-free number, gets his questions answered, tells a friend about this information, and the friend then files her taxes properly. Or a taxpayer who hears about a friend’s audit and then decides to report his taxes more accurately. A subset of indirect effects is subsequent-year effects on a taxpayer who had IRS contact in a prior year.
- 7.) **General vs. Specific Deterrence** – Deterrence is a mechanism that drives behavioral change, generally by providing incentives or disincentives. The effectiveness of a deterrent is thought to depend on people’s perception of the certainty, severity, and celerity (speed) of those incentives and disincentives. Specific deterrence relates to the behavior of an individual who has experienced the incentive or disincentive first-hand, such as the subsequent-year effect on someone who has been audited. General or indirect deterrence

relates to the behavior of the general population—those who have not experienced the incentive or disincentive first-hand, but whose perceptions change based on word-of-mouth reports from those who have had direct contact, official statements, news coverage, etc. (sometimes called the “ripple effect”).

- 8.) **Assurance** – A mechanism that drives behavioral change by reassuring people that relevant systems are being administered fairly. For example, taxpayers in the general population may improve their compliance in response to publicity about the success of criminal investigations by the IRS—not because they fear being caught by similar criminal probes, but because they are assured that justice is being done and scofflaws are not going unpunished.

A.2 Glossary of Methodologies

- 1.) **Theoretical Development** – Theory posits cause-and-effect relationships between variables of interest, such as between services and compliance or through intermediate variables such as trust and satisfaction, and whether those variables are endogenous (derived from model functions and parameters) or exogenous (externally determined). Through theory, testable hypotheses are produced, and theory is revised when empirical studies suggest the theory is misguided.
- 2.) **Segmentation Analysis** – Recognizing that different segments of the population have different reasons for noncompliance, this method uses primarily surveys to gather data to understand (1) population segments with different reasons for compliance, (2) the factors that influence compliance by segment, and (3) how the IRS can influence compliance by segment.
- 3.) **Aggregate Econometric Modeling** – This method uses panel time series cross section data aggregated by geographic region. The dependent variables are observable measures of taxpayer behavior, such as whether the taxpayer filed, how much they reported, and how much they paid. Explanatory variables would include a wide range of IRS variables, such as volume and quality of activities in similar geographic detail and audit risk.
- 4.) **Microeconometric Modeling** – This econometric approach uses individual level data, presuming that service microdata can be merged on an individual level.
- 5.) **Field Experiments** – Field experiments create treatment and control groups (ideally randomly assigned), vary IRS activities by group, and then observe impacts on compliance (groups could differ by factors such as geography or time).
- 6.) **Quasi/Natural Experiments** – Quasi/Natural Experiments are Field Experiments which were not deliberately created, but exploit the variation in the natural world. An example of this would be state-level or service area-level variations in IRS services, effectively creating an experiment where the two regions have different access to a specific service and thus may have different compliance levels. The difficult part of both Field and Natural Experiments is ensuring that both groups are similar in all meaningful aspects, except the IRS activity being altered between them.
- 7.) **Laboratory Experiments** – Laboratory experiments measure the responses of recruited subjects, often university students, in a controlled environment to test hypothesis about compliance behavior.
- 8.) **Agent-Based Modeling** – Agent-based modeling is a micro-level simulation of a system in which individual agents (e.g., taxpayers, 3rd parties or administrative agencies) take actions and interact with other agents based on data-driven rules. This computational model simulates the actions and interactions of individual agents in a network in order to understand the effect on the overall system. The state of the system evolves over time as a result of all of the actions of the full set of agents included.
- 9.) **Social Network Modeling** – Social network modeling views social relationships as a connection of nodes (individual actors within a network) and ties (relationship between the actors). Social networks occur on many levels and play an important role in problem solving. Social network modeling maps the nodes and ties between them to determine the impact of certain events.

A.3 Glossary of Data

- 1.) **Compliance Data Warehouse** is an accessible research repository for multiple legacy data sources. For example, CDW contains both NRP and ERIS extracts, data from individual and business returns, and audit enforcement data. <http://compliance.no.irs.gov/cdw>
- 2.) **Enforcement Revenue Information System (ERIS)** – ERIS tracks assessments made, revenues collected, direct hours reported, and cost incurred as a result of conducting Internal Revenue Service (IRS) enforcement activities. <http://www.irs.gov/privacy/article/0,,id=154931,00.html>
- 3.) **National Research Program (NRP)** – The goal of National Research Program (NRP) is to design and implement a successful strategy to collect data that will be used to measure payment, filing and reporting compliance and to deliver the data to the Business Operation Divisions to meet a wide range of needs including support for the development of strategic plans and improvements in workload identification. The IRS uses the NRP to analyze taxpayer compliance and evaluate the effectiveness of IRS enforcement activities. Data for analysis is available in CDW. <http://www.irs.gov/privacy/article/0,,id=139179,00.html>
- 4.) **Statistics of Income (SOI)** – SOI is the IRS branch responsible for publishing official tax statistics. SOI also performs a number of studies and research projects. The SOI designs a study, determines the sample size and selection criteria, captures the data, assigns weights, and analyzes sample variability. Most SOI operations begin by sampling returns from the master file system compiled by IRS processing centers. They have computers in several of these centers dedicated to statistical processing. They generally compile statistics based on stratified probability samples of tax or information returns, using such classes as size of income, industrial activity, or the like. Data items pulled electronically from the master file are augmented with items captured from the hardcopies of taxpayers' returns. <http://www.irs.gov/taxstats/article/0,,id=120303,00.html>



Appendix B: Summary of Current Surveys

Survey Dimension	W&I Market Segment Survey (2006)	The Taxpayer Customer Service and Channel Preference Survey (Oversight)	Conjoint Study (I, II)	Benchmark Survey of Taxpayer Opinions on IRS Services
Number of Respondents	Cross-sectional sample of 2,459 taxpayers selected through a random-digit dialing process - 30-32% response rate	Cross-sectional random sample of 1,000 taxpayers; over-sampled 101 individuals who had specifically visited an IRS Taxpayer Assistance Center (TAC) in person	75% response rate; Responses were received from 2,196 individual taxpayers, including an over sample of 533 taxpayers who had visited a TAC.	Mailed to 40,000; over 10,000 responded
Date	July 25 through September 11, 2007	March 19 through April 13, 2006	June 23 through July 5, 2006	June through September, 2006
Length of Survey	18 min. for English / 25 min. for Spanish	See instrument / variable	See instrument / variable	10 min.
Selection of respondents	All individual taxpayers - including Schedule C, E, F -who filed a 2006 tax return in 2007 (including some Spanishspeaking taxpayers), regardless of whether or not they received assistance directly from the IRS for tax-related items.	Sampled using random digit dialing and screened so as to reach the individual in the household most responsible for overseeing the preparation of its federal income taxes; must have filed tax year 2004 return filed during calendar year 2005.	Pre-selected respondent pool via vendor	Potential respondents must have been age 18 or over, residing in the United States, and have filed a federal income tax return for tax year 2004 prior to December 31, 2005, using the Form 1040 series of returns.
Modality	CATI - random-digit dialing to U.S. households	CATI - random-digit dialing to U.S. households	Participants submitted their responses online or via a television set top box (for those who don't have computer access).	Mail

Appendix C: Review of the Literature

I. Introduction

To what extent do the specific activities, actions and services that IRS provides impact the voluntary compliance behavior of taxpayers, one of the core metrics of IRS? In 2008, as part of a Congressional Budget Initiative, *Measuring the Impact of Services on Compliance*, IRS began the design stage of a multiyear research project. This project aims to answer the fundamental question, *How can IRS estimate the full marginal effects of services on compliance?* While estimating the impact of services is the core goal, the planning team recognizes that this requires controlling for all other determinants of voluntary compliance, including IRS enforcement activities, as well as non-IRS factors. This project also needs to address the challenge of improving the completeness and accessibility of the data (both internal and external to the IRS) that are needed to answer these questions effectively.

As part of the design phase, this review of the literature is the first step in assessing current knowledge and relevant approaches. In addition to the general goal of updating the knowledge of the literature on the issue of tax compliance, we also have three specific goals for this literature review. A key component of the design phase is the planning and implementation of a workshop in which experts in tax compliance and other types of behavioral modeling will provide IRS with their professional judgment about how best to plan for and operate a research plan. This document is intended to be one of the core inputs to this group. The specific goals of this document are to:

- Provide a reference that will allow a person new to this field to get into the literature quickly;
- Document the lessons learned from prior research in this area; and
- Provide a structure for decision-making and review by members of the workshop.

Section II reviews the compliance literature of the last several decades for key insights. We discuss the relevance of this literature to estimating services impact. We also take a preliminary look at the methods employed here, as potential approaches to be considered by workshop participants. Section III provides a detailed annotated bibliography of over 40 papers. In selecting the bibliography, we began with the core literature on tax compliance estimation. This literature encompasses traditional economic theory, econometric estimation, experimental laboratory studies, observational field studies and surveys, theoretical approaches from policy and law, as well as newer methods such as multi-agent simulation. Section IV provides short descriptions of a wider set of papers encountered during the bibliography compilation.

Integral to the project's approach is an expansive view of relevant areas of technical expertise. One IRS goal is to objectively examine and incorporate insights from a wide variety of fields and technical approaches. The team involved in compiling this bibliography is more familiar with the traditional tax areas and, though we have found some relevant literature from other fields, we believe that more can and will be found outside of the tax literature. Similarly, while most of the research to date has focused on the compliance behavior of individuals, and while estimating individual behavior is the highest priority for this project, we also seek to build on and expand what is known about the compliance of other entities, such as corporations, flow-through businesses, and tax-exempt organizations. In order to incorporate relevant knowledge from a wide variety of areas, we are looking to include experts from other fields in the workshop and we are expecting that those experts will expand our knowledge of the literature. Thus, we expect to expand the sections to include additional insights from fields such as psychology, marketing, environmental policy, and criminal justice. Therefore, this bibliography should be considered to be a living document that will be updated as we acquire new recommendations from participants and through further inquiry.

II. Relevant Literature for Measuring the Indirect Effects of Services

This overview provides a brief summary of past efforts by researchers to identify and estimate the impact of tax service and enforcement activities on tax compliance, as well as related efforts by selected federal agencies to measure the impact of their own service and enforcement activities on their compliance objectives.

1. Theoretical Underpinnings

When attempting to assess the effects of various IRS activities on tax compliance, it is useful to have a theoretical framework for understanding their potential impact on taxpayer behavior. A useful starting point is expected utility

theory, which views taxpayers as rational gamblers who weigh the benefits of successful tax evasion (reduced tax payments) against the potential costs of being caught and punished. To understand this framework, consider the following hypothetical scenario (which is a slight modification of the original formulation of Allingham and Sandmo, 1972) in which a taxpayer with true taxable income y is deciding on how much income to evade E . The individual faces a proportional tax rate t on reported income. In addition, with fixed probability p , he will be audited, in which case he will have to pay his unreported taxes (tE) plus a fine at the rate θ on the unpaid amount. For any given amount of evasion E , the taxpayer would face two possible outcomes. Under the first outcome, the taxpayer is successful in his evasion, in which case he experiences utility in the amount: $U[y(1-t) + tE]$, where $U[\]$ represents the taxpayer's utility function, evaluated at a wealth level equal to legislated after-tax income, $y(1-t)$, plus the amount of tax successfully evaded, tE . Under the second possible outcome, the taxpayer is caught and penalized for cheating, in which case he experiences the lower utility level, $U[y(1-t) - \theta tE]$, evaluated based on his legislated after-tax income less the penalty, θtE . To evaluate his expected utility (EU), the taxpayer computes the weighted average of his utility under these alternative outcomes, using the likelihood of each outcome as the weights:

$$EU = (1-p)U[y(1-t) + tE] + pU[y(1-t) - \theta tE] \quad (1)$$

The taxpayer then chooses to evade by the amount that maximizes his expected utility (i.e., the value of E that maximizes the value of the expression for EU in Equation (1)). Using this framework, it is straightforward to show that tax compliance tends to be higher when the taxpayer faces a greater risk of audit and detection and when the fine rate applied to underreported taxes is relatively severe. Observe that this means the taxpayer will respond to an increase in the risk of audit and punishment, regardless of whether he is ultimately caught or not. This is what is known as a general deterrent or ripple effect. It is this basic prediction that underlies much of the research on tax enforcement.

For given values of the tax, audit, and penalty rate parameters under the above framework, the precise extent of evasion will depend on the form of the utility function. Of particular importance is the attitude of the taxpayer towards risk. The more risk-averse the taxpayer is, other factors held constant, the less he will choose to evade.

In its most basic form, the expected utility model of noncompliance assumes that taxpayers are fully aware of their reporting obligations and all relevant tax and enforcement parameters, such as the marginal tax rate, the probabilities of audit and detection, and the penalty rate that applies to discovered evasion. In such a model, there is essentially no scope for tax agency activities outside of tax enforcement to influence taxpayer compliance. Many researchers have therefore attempted to relax these assumptions while staying within the expected utility framework. For instance, the framework has been extended by Scotchmer and Slemrod (1989) to account for uncertainty among taxpayers about their true tax obligations. Under this extension, it turns out that taxpayers are likely to report a higher tax bill when they are uncertain of how much they owe than when they are certain. Thus, to the extent that taxpayer services (such as telephone and office assistance programs) reduce taxpayers' uncertainty about their tax obligations, the implication is that these services actually may lead to a lower aggregate level of tax compliance. In particular, this framework suggests that while the resolution of uncertainty would reduce the incidence of overreporting errors, this would be more than offset by an increase in the incidence of underreporting violations.

Another extension to the expected utility framework accounts for the possibility that some individuals will elect not to make any report at all to the tax agency; such individuals are known as nonfilers or ghosts in the tax compliance literature. In Erard and Ho (2001) an individual compares his expected utility from not filing a return, taking into account not only the risk of detection and punishment, but also the burden associated with preparing and filing a return. He then compares this result with the expected utility he would achieve if he were to file a return and make an optimal report (based on the utility-maximizing level of evasion obtained in the standard expected utility model of tax compliance). Among these two options, the taxpayer chooses the one that yields the highest expected utility. Observe that under this extension, there is some scope for taxpayer assistance programs to improve filing behavior. In particular, to the extent that the decision not to file is driven by a large perceived burden of preparing and filing a return, taxpayer assistance programs might be able to mitigate the burden sufficiently so that filing becomes an attractive option. Such programs might be particularly important for taxpayers who have not filed returns for many years, who fear that they will have to submit many past due tax returns and make a large payment to square their accounts. Taxpayer filing assistance, combined with procedures to ease these concerns (payment plans, penalty forgiveness, etc.) may be effective in such cases.



In qualitative terms, the expected utility model predicts that taxpayers will tend to be more compliant when the risk of detection and penalty is high. In practice, there is reasonable support for this proposition. Specifically, both third-party information reporting and source withholding are common tools employed by tax agencies to reduce the potential for successful evasion on certain forms of income. According to the expected utility framework, one would therefore expect compliance to be rather high for income sources, such as wages and salaries in the U.S., that are subject to high levels of both third-party information reporting and withholding; somewhat lower for sources, such as interest, that are subject to high levels of information reporting but relatively little source withholding; and substantially lower for sources such as self-employment income, that are subject to low levels of information reporting and withholding. Random audit studies in the U.S. have consistently found such a pattern.

Although the qualitative predictions of expected utility theory about the impact of tax enforcement on tax compliance appear to be borne out in practice, the theory tends to fall short when predicting the magnitude of compliance. Specifically, compliance rates on income items not subject to third-party information reporting or source withholding appear to be much higher than standard expected utility models would predict given the relatively low rates of audit and the low incidence of penalties observed in the U.S. and various other countries. A plausible explanation for the failure of expected utility theory to explain why compliance is as high as it appears to be in many countries is that it focuses too narrowly on financial incentives and the application of rational calculus in individual decisionmaking. To account for a broader range of influences on tax compliance, many economists and psychologists have turned to behavioral economics. The field of behavioral economics attempts to address various cognitive and psychological limitations that individuals encounter when making complex decisions. According to the behavioral economics perspective, such limitations can result in individuals making less than optimal decisions, in some cases contrary to their long run interests. This literature also accounts for the possibility that individuals will sometimes act in an altruistic manner, sacrificing their own self-interests in order to advance the welfare of others.

One strand of behavioral economics research, known as prospect theory (Kahneman and Tversky, 1979), posits that individuals possess different attitudes regarding risks involving gains than they do regarding risks involving losses. In comparison with expected utility theory, prospect theory relies on three distinguishing assumptions regarding how an individual evaluates a risky proposition, each of which has at least some empirical support: (1) utility depends on gains and losses relative to a reference wealth level rather than on final wealth levels; (2) the potential outcomes are evaluated on the basis of decision weights that attach greater significance to low probability outcomes than the objective outcome probabilities; and (3) individuals are subject to “loss aversion”, meaning that they strongly prefer avoiding losses to achieving gains of an equivalent amount. To clarify the differences between the expected utility and prospect theory frameworks, consider our earlier hypothetical scenario in which a taxpayer with true taxable income y is deciding on how much income to evade E . Under prospect theory, the taxpayer would evaluate the alternative outcomes of successful and unsuccessful evasion rather differently than under expected utility theory. Specifically, rather than focusing on his net wealth levels under the two possible outcomes, he would instead focus on the gain or loss relative to a reference level of wealth. To illustrate, suppose that the taxpayer takes his legislated after-tax income $y(1-t)$ as his reference level of wealth. His valuation of the gain under the successful evasion outcome would be $V(tE)$, where the taxpayer’s value function $V(\cdot)$ (which is similar to a utility function) is evaluated at the gain from successful evasion tE . In contrast, the taxpayer’s valuation of the loss $-\theta tE$ associated with the unsuccessful evasion outcome is $V(-\theta tE)$. The taxpayer’s value function is assumed to exhibit loss aversion, meaning that for any amount $z > 0$, $V(z) > -V(-z)$. In other words, the taxpayer values avoiding a loss of z more heavily than making a gain of z . As with expected utility theory, the taxpayer evaluates the overall evasion gamble by taking a weighted average of his valuations of the two alternative outcomes. However, under prospect theory, he weights the outcomes using a function w of the objective probability of each outcome rather than the objective probability, itself:

$$w(1 - p)V(tE) + w(p)V(-\theta tE). \tag{2}$$

Given a relatively low audit probability p , the individual is assumed under prospect theory to place greater weight [$w(p) > p$] on the outcome in which an audit takes place (and lower weight on the outcome where an audit does not take place) than one would under the expected utility framework. In other words, he acts as if the likelihood of an audit is greater than the actual objective probability.

Compared with the results under the expected utility specification, the individual in the prospect theory specification will tend to evade by less, both because he acts as though the risk of audit is higher than the actual risk and because he exhibits an aversion to experiencing losses as a result of being audited. Thus, at least in this example, the

deterrence effects associated with audits and fines obtained under the expected utility framework are magnified under the prospect theory approach.³⁶

Observe that in the above models, the probability of audit is fixed and unrelated to the extent to which a taxpayer is evading. In game-theoretic extensions of such models, the tax agency is treated as a strategic actor who imposes higher audit rates on taxpayers who report suspiciously low levels of taxable income. In this setting, the risk of audit is endogenous, meaning that it is jointly determined along with the level of evasion. As discussed in the review of the empirical literature below, this insight has played an important role in the specification and interpretation of empirical studies of deterrence.

Like expected utility theory, prospect theory assumes that tax compliance is driven only by financial considerations. As such, neither theoretical framework provides much scope for tax agency services (outside of enforcement activities) to have a beneficial impact on tax compliance. However, other strands of behavioral economics research hold more promise in this regard, particularly those that relate to social norms. Social norms are customary rules of behavior that coordinate our interactions with others and act as informal social controls. Individuals typically obey these norms either to avoid feelings of shame associated with social disapproval or because the norms have become internalized as a personal code of conduct (so that their violation would lead to feelings of guilt and remorse). Obeying social norms, especially when they have been internalized, need not be a deliberative process. Rather, one might do so reflexively or unconsciously, much as a matter of habit. In the context of tax compliance, an important social norm is the concept of reciprocity (Cialdini, 1989; Smith, 1992), under which individuals are expected to reciprocate in kind behaviors that are directed towards them.³⁷ Based on this social norm, it has been suggested by various researchers and tax administrators that to the extent taxpayers perceive that they are provided with fair and responsive service, they are more likely to feel positively about the tax agency and reciprocate by complying with their tax obligations (Braithwaite, 2001, 2007, Kirchler et al. 2006, Kornhauser, 2007). Thus, this is one pathway through which taxpayer assistance services may have a beneficial impact on tax compliance. Moreover, to the extent that taxpayers do desire to meet their social responsibilities, taxpayer services can help them to overcome informational and computational barriers that might otherwise prevent them from doing so.

Of course, not all taxpayers require the same type or degree of assistance. Therefore, just as with enforcement, it is important to recognize that many taxpayer services are targeted towards specific groups of taxpayers. Also as with enforcement, though, the compliance impact of these services may extend beyond those who are targeted. Perceptions that the tax agency offers helpful assistance and treats taxpayers fairly may induce reciprocation even among taxpayers who do not seek taxpayer services or have any direct contact with the tax agency. Further, those who have received assistance may share what they have learned with others, thereby helping to educate other taxpayers regarding their responsibilities. Potentially, then, service activities also may be subject to a ripple effect.

Ultimately, the norm of reciprocity is not restricted to relationships between an individual taxpayer and the tax agency; it is also likely to apply to interactions among different taxpayers. For instance, to the extent that some taxpayers are perceived to be cheating on their taxes, other taxpayers may feel less impetus to meet their own tax responsibilities. Thus, the norm of reciprocity also provides an enhanced scope for deterrence to play a role in improving compliance. Through its enforcement activities, the tax agency is able to reassure compliant taxpayers that others are being encouraged to pay their fair share, thereby maintaining incentives for compliant taxpayers to continue to meet their tax obligations.

One strand of behavioral economics research (Gordon, 1989; Myles and Naylor, 1996) focuses on nonmonetary costs, such as reputation or stigma costs, associated with evasion. Typically, such costs are assumed to be high when relatively few taxpayers are cheating and low when relatively many are cheating, so that the utility one achieves

³⁶ A key assumption here is the reference wealth level, because loss aversion tends to make individuals risk averse with respect to earning gains and risk seeking with respect to avoiding losses. If we had assumed that the reference wealth level was pre-tax income (or income after withholding), then taxpayers would perceive any payment of tax (or payment of more than was withheld) as a certain loss situation, and would tend to be risk-seeking in their efforts to avoid such losses. Under such a scenario, one might therefore predict higher levels of evasion for some taxpayers under prospect theory than under expected utility theory.

³⁷ Other potentially important social norms include duty to country and obedience to authority.

from tax evasion is positively associated with the number of others who evade. Under such a framework, there can be a tipping point at which a slight increase in the incidence of evasion triggers an epidemic effect that ultimately leads to a dramatic reduction in both tax compliance and the stigma costs associated with it. By maintaining a sufficient level of enforcement activity, the tax agency in such models can sustain the good equilibrium with relatively high levels of tax compliance. To the extent that the provision of helpful and responsive taxpayer services positively influence the degree of stigma associated with evasion, such services may also have a role in sustaining compliance.

Thus far, we have ignored the question of how taxes are to be used. In practice, tax receipts are employed for a variety of social functions, including the provision of public goods (i.e., goods such as national defense, which impart their benefits on all members of society). In a traditional expected utility model, taxpayers would rationally choose to “free-ride” on these public goods, since in a large economy the taxes paid by a typical individual have only a negligible impact on the overall level of public goods provision. However, one can construct behavioral models that allow for altruism or “warm glow” effects (increased utility from helping others) under which the provision of public goods provides a positive incentive for complying with one’s tax obligations. In this vein, it has been suggested that informational services that highlight the link between tax receipts and the various social programs that they fund may have a positive impact on compliance.

Of course, the tax agency is not the sole provider of taxpayer assistance services. Indeed, approximately half of all taxpayers in the U.S. seek paid tax assistance from third-party tax preparers. Many others receive some forms of informational and processing assistance from tax preparation software, books, significant others, and non-agency websites. Therefore, an improvement in the quantity or quality of tax agency services may lead to a change in relative utilization levels for agency and non-agency services. Relative (as well as absolute) utilization levels also may be influenced by external factors, such as changes in tax complexity resulting from new legislation. The impact of this substitution response on compliance behavior is uncertain. For instance, the theoretical literature on tax compliance suggests that paid tax practitioners play a dual role, helping to enforce compliance with respect to straightforward, unambiguous tax provisions, while assisting taxpayers in exploiting ambiguous provisions to promote aggressive tax positions (Klepper, Mazur, and Nagin, 1991). Thus, shifts in the number of taxpayers (or the kinds of taxpayers) receiving assistance from third party preparers can have important ramifications for overall taxpayer compliance.

In addition to tax assistance services, the IRS engages in a variety of activities to understand, simplify, and report the burdens on taxpayers. While the Service would be quite interested in understanding how such services affect compliance, as far as we know there is no research in this regard. It is possible that because of the size of the projected burden, some taxpayers will choose to be less compliant, and it also possible that a more compliant community of taxpayers results from the successes in lowering compliance burdens.

To summarize, there are 8 main insights from the theoretical literature on tax compliance regarding the relationships among tax compliance, enforcement activities, and taxpayer assistance services:

1. Audits and penalties are likely to have a general deterrent effect on tax evasion.
2. Taxpayers may perceive that, or at least behave as though, the risk of being caught exceeds the objective risk.
3. Enforcement activities are likely to be endogenously determined, in the sense that the tax agency attempts to target its resources towards likely evaders.
4. Some forms of taxpayer assistance services are also likely to be endogenously determined. For instance, tax agencies may determine that some degree of ambiguity in the tax laws is desirable from a compliance standpoint. More generally, the tax agency may target many of its outreach and assistance services towards taxpayers who are deemed to be at risk for certain forms of noncompliance.
5. Taxpayer services are likely to have a direct impact on some forms of compliance by reducing the informational and computational barriers that prevent some taxpayers from filing returns and properly reporting their tax liability.
6. The overall impact of taxpayer services on tax compliance is likely to be influenced by psychological factors, such as social norms.
7. Deterrence activities undertaken by the tax agency may play some role in maintaining social norms to comply.
8. Taxpayers have access to both paid and unpaid forms of tax assistance, including that provided by tax software, outside of the tax agency. Substitution responses across the various types of assistance offerings, in response to

changes that are either internal or external to the tax agency, may have an important impact on the overall level of tax compliance.

2. Empirical Literature

In this section, we provide a brief review of the empirical literature as it relates to the impacts of tax enforcement and taxpayer service on tax compliance. We begin by summarizing the key findings and the extent to which the results of existing studies diverge. We then provide a critical discussion of the main methodological approaches that have been employed.

Key Findings

There have been a substantial number of empirical studies of the general deterrent effect of tax audits, but relatively few of the impact of taxpayer services on tax compliance. With regard to the general deterrent effect of tax audits, results vary widely across studies with some finding only a very modest general deterrent effect (or even in some cases a counter-deterrent effect), and others finding an extraordinarily large effect. As discussed below, much of this variation is attributable to studies that rely on aggregate forms of data. The studies based on individual level data from various sources, including field experiments, laboratory experiments, taxpayer audits, and surveys come to fairly similar conclusions. These latter studies typically find only a fairly modest general deterrent effect (and occasionally even a counter-deterrent effect), even when the perceived risk of audit appears to be quite high.

The relatively small body of empirical research on the impact of taxpayer services on tax compliance comes to different conclusions regarding different forms of taxpayer service. A lesson from these studies is that the impact of services on compliance behavior is likely to depend both on the specific structure of the service offering and on the nature of the intended recipients. As a result, it is difficult to generalize from the results for one particular service offering to other types of service offerings.

General Deterrent Effect of Tax Audits

Researchers have employed a variety of alternative methodological approaches to estimate the general deterrent effect of tax audits, or in other words, the overall increase among taxpayers in their reported tax liabilities in response to an increase in the risk of audit. These approaches can be distinguished by differences in the types of data, the nature of the dependent variable, and the econometric techniques that they employ. As discussed below, each approach is subject to a different set of limitations.

A fundamental problem is that it is very difficult to obtain reliable information on tax noncompliance, which by its very nature is something that taxpayers attempt to conceal. Arguably the best data on noncompliance comes from National Research Program (NRP) and the predecessor Taxpayer Compliance Measurement Program (TCMP). These programs, which have been carried out by the IRS, are large scale audit studies involving randomly selected returns from the overall taxpaying population (particularly individuals). For the most part, tax returns in these studies were subjected to intensive examinations, and detailed results were recorded about both what the taxpayer reported on the return and what the examiner deemed should have been reported. Although even intensive examinations are not always successful in uncovering all noncompliance that is present on a return, such data are potentially quite useful for studying tax compliance behavior.³⁸ In fact, several studies of deterrence have been conducted based on a now very dated version of the TCMP (for tax year 1969), aggregated to the three-digit zip code level. Unfortunately, though, the dependent variable employed in these studies was not an audit-based measure of noncompliance. For instance, two of these studies (Witte and Woodbury, 1985; Dubin and Wilde, 1988) employed a simulated measure of noncompliance based on the computed DIF scores assigned to returns in each zip code area. The DIF score is a statistical index used by the IRS to predict the likelihood that a given return is in need of an adjustment, either up or down. As such, it is a useful tool for the IRS in deciding which returns in the population to audit. However, it presumably provides only a very noisy indication of the magnitude of noncompliance on a tax return. The studies that employed this measure used it as a dependent variable in a

³⁸ Some preliminary research on the deterrence issue based on a combination of NRP data and operational audit data is currently in progress.

regression against the audit rate as well as a set of aggregate socioeconomic control variables. As discussed previously, a lesson from the game-theoretic literature on tax compliance is that the audit rate is likely to be endogenously determined. In Dubin and Wilde (1988) an attempt was made to control for potential endogeneity of the audit rate using instrumental variable techniques. Both studies found only limited evidence for a general deterrent effect, which was relatively small in magnitude.

In the absence of a reliable audit-based measure of tax noncompliance, some researchers have relied on measures of admitted tax evasion based on survey responses (see, for example, Sheffrin and Triest, 1992 and Kinsey, 1992). A difficulty with this approach is that respondents may not be forthcoming about their tax transgressions, so that there is likely to be a substantial errors-in-variables problem. Sheffrin and Triest (1992) attempted to resolve this problem using sophisticated latent variables techniques, but just as with the aforementioned studies that rely on a simulated measure of noncompliance, it is difficult to fully overcome concerns about results that are based on an attempt to explain variations in a poorly measured dependent variable. Typically in survey-based studies, researchers have attempted to explore the impact of attitudes and socio-economic variables on compliance, as well as measures of the perceived risk of audit. However, the causal links between expressed attitudes, perceptions, and compliance behavior are difficult to assess in such studies. On the one hand, the causality may run from attitudes and perceptions to compliance behavior. On the other hand, the causality could run in the opposite direction if respondents are attempting to present a consistent self-image by espousing attitudes and perceptions that serve as an ex-post rationalization for their behavior. In any case, results from such studies have not provided any clear indication of a general deterrent effect of tax audits. Surveys do, however, provide some insight into taxpayer perceptions about audit risk. In particular, they indicate that taxpayers have widely different perceptions of the likelihood of audit, and that these perceptions, on average, exceed the objective audit risk. This suggests that to understand the general deterrent effect of tax audits fully, one needs to understand how audit perceptions are formed and how these perceptions respond to changes in objective risk levels.

As an alternative to attempting to employ a direct measure of noncompliance to study general deterrence effects, several groups of researchers (Dubin, Graetz, and Wilde, 1988; Beron, Tauchen, and Witte, 1992; Plumley, 1996; Dubin, 2004) have relied on aggregate measures of reported taxes, reported income, and/or filing rates as the dependent variables in their analyses. Through regression analysis, they have attempted to explain how the audit rate influences reporting and filing behavior. A major challenge for this approach is how to control for potentially confounding factors. In addition to the audit rate, there are a wide variety of factors that might explain differences in aggregate reporting levels and filing rates across different time periods or geographic locations. Across time periods, there can be significant changes in tax rules, tax rates, and filing thresholds that impact on who should file and how much should be reported. Further, across both time periods and geographic areas, socio-economic and cultural factors may differ in ways that impact on reporting and filing behavior. Changes can also take place in enforcement technologies (such as expansions of document matching or changes in audit selection criteria that can influence taxpayer behavior independently of changes in the audit rate. Indeed, such the IRS has pointed to technological changes such as these in past years to help explain how it has been able to maintain a relatively high degree of tax compliance despite falling overall audit rates. A second challenge is that the IRS presumably tends to devote more audit resources to areas where noncompliance rates are relatively high, in which case audit rates are endogenously determined. To address these concerns, researchers have employed instrumental variables regression specifications with fairly detailed sets of explanatory variables, a method that relies heavily on the judgment of the analyst to locate suitable instrumental variables (variables that are correlated with the audit rate but which have no direct impact on noncompliance) as well as suitable controls for the various potentially confounding factors. Estimation results based on these studies appear to be sensitive to the choice of data set, time period, and model specification. However, with one exception (Beron, Tauchen, and Witte, 1992), such studies have yielded very large estimated general deterrent effects, in some cases indicating that a modest increase in the audit rate could effectively eradicate a very large portion of the overall tax reporting gap (the aggregate difference between the taxes that should have been reported by filers on their returns and what they actual reported).

Two recent studies (Slemrod, Blumenthal, and Christian, 2001; Hasseldine et al., 2007) have employed controlled field experiments to assess the general deterrent effect of tax audits. Like the previously mentioned aggregate data studies, these studies have examined changes in reported amounts of tax liability in response to variations in audit risk. However, because the audit risk has been experimentally manipulated for a random sample of taxpayers, it is not necessary to control for the impact of a large set of potentially confounding factors under this approach. Nor is it necessary to identify and employ valid instrumental variables, because the audit risk has been exogenously assigned.

In Slemrod, Blumenthal, and Christian (2001) a random sample of taxpayers was mailed a notice informing them that the tax returns they soon would be filing would be closely examined (in essence indicating that an audit was certain to occur). In Hasseldine et al. (2007) random samples of self-employed taxpayers were mailed letters with one of several alternative messages ranging from the suggestion of an increased likelihood of audit, to the suggestion of an increased likelihood of audit along with an articulation of potential penalties for noncompliance, to notification that the taxpayer's return has been pre-selected for audit. To assess the compliance impact of these messages, the difference in taxpayer reported amounts for the years just subsequent and prior to the receipt of the messages were compared with the corresponding difference in reported amounts for a control group of taxpayers who did not receive any message. Based on this "difference-in-differences" estimation strategy, both studies found general evidence of somewhat modest but significant deterrent effects in most cases.³⁹ It is striking that even when confronted with a certain audit probability, the deterrent effect was not especially dramatic.

An important issue with regard to these experiments is that taxpayers were provided with direct, and apparently quite credible, information about their audit risk from the tax agency. In more typical situations, tax agencies maintain a substantial degree of secrecy about their enforcement functions, which creates taxpayer uncertainty over their actual audit risk. Based on the aforementioned survey evidence, it appears that taxpayer perceptions of audit risk vary substantially. An important issue for future research is how taxpayer perceptions are formed, and how they respond to changes in the objective audit risk. Presumably, social networking plays a key role here. Of particular importance may be how tax practitioners interact with taxpayers. Presumably tax practitioners who prepare a substantial number of returns and participate in client audits are more knowledgeable about the objective likelihood of an audit, how this risk varies with different reporting strategies, and the potential for sanctions.

Given the considerable challenges associated with obtaining and analyzing useful observational data on tax compliance, some researchers have turned to laboratory experiments to assess how various factors influence compliance behavior (see Blackwell, 2008, for a survey of this approach). An advantage of such experiments is that the researchers can impose considerable control over the laboratory environment to isolate the effects of specific variables on compliance outcomes. Typically, laboratory studies of the deterrent effect of tax audits are designed to mimic the standard expected utility model of tax compliance. Participants take part in multiple rounds, in which they receive income of some sort, are asked to report that income for tax purposes, pay taxes based on the amounts declared, and are subject to potential audits and penalties for noncompliance. At the end of the experiment, they receive a reward based on their net earnings/winnings. A limitation of such experiments is that they are rather artificial. True life compliance decisions often tend to be more subtle and complex than those studied in the laboratory, with higher stakes and a greater sense of gravity and social responsibility. But still, such experiments do capture at least some of the incentives facing actual taxpayers, and do so in a controlled way that allows for more straightforward interpretations of results than typical studies based on observational data. Interestingly, consistent with many individual level observational and field study results, the findings of laboratory experiments generally indicate only modest general deterrent effect of audits. For instance, raising the audit rate from 4 to 6 percent in a typical laboratory experiment (Alm, Jackson, and McKee, 1992) involving a proportional tax rate of 30% and a 100% penalty on unreported taxes results in an improvement in the average compliance rate by approximately 3 percentage points, from 33.2% to 36.5%.

A final approach that has been used to gain experimental insights into tax compliance decisions is agent-based modeling (see Bloomquist, 2006, for a review of this approach). Under this approach, agents (taxpayers) in a simulation model are assigned heterogeneous characteristics that influence their behavior, such as differing attitudes toward risk or differing behavioral strategies (e.g., free riding vs. copycatting). These agents make compliance decisions based on factors such as tax, audit, and penalty rates much as in formal theoretical models of tax compliance. However, interactions among different agents within a social network may influence their subsequent behavior. For instance, copycats may choose to report honestly, because they meet another taxpayer who is honest. Alternatively, an agent may raise his perceptions of audit risk when he meets someone who was recently audited. In this way, researchers can explore potential outcomes based on complex models involving social interactions, group dynamics, and heterogeneous agents that would be intractable using ordinary analytical methods. While agent-based

³⁹ In Slemrod, Blumenthal, and Christian (2001) higher income taxpayers actually demonstrated some evidence a counter-deterrent effect.

modeling cannot currently produce any firm estimates of the general deterrent effect of tax audits, it has potential to suggest a range of potential outcomes and yield new insights into such questions as ways in which individuals may learn about audit risk and how they develop strategies to cope with their compliance decisions.

Impact of Taxpayer Services on Tax Compliance

In contrast to empirical studies of the general deterrent effect of tax enforcement, there have been relatively few empirical studies of the impact of taxpayer services on tax compliance. Overall, though, the methodological approaches that researchers have used to study the impact of taxpayer services are quite similar to those used to study the impact of enforcement.

As with the former topic, a key challenge in analyzing the impact of taxpayer services is obtaining a suitable way to identify changes in tax compliance and to link these changes to specific forms of exposure to taxpayer services. One measure that has been used as a proxy for tax noncompliance is the difference between a national accounts-based measure of income and income reported for tax purposes. Naturally, there are some important conceptual differences between these two measures of income. In addition, national accounts measures of income often rely to some extent on tax return information for certain income sources. Thus, the difference between these two measures of income is hardly a pure measure of tax noncompliance. Nonetheless, researchers have regressed such a measure against a set of explanatory variables in an effort to obtain insights into tax compliance behavior. In one study involving such a proxy noncompliance measure for a group of Swiss Cantons, Frey and Feld (2002) found a negative relationship between the proxy for noncompliance and a survey-based measure of the extent to which a Canton treated taxpayers fairly and respectfully. The latter measure was based on such factors as whether the Canton followed established formal and informal procedures, whether high fines were applied only in the case of more serious and deliberate forms of noncompliance, and whether refunds were granted when a taxpayer was found to have erroneously overpaid his taxes. The implication of the study is that taxpayers will tend to be more compliant in situations where they perceive that they are subject to high levels of procedural fairness. While this result seems plausible, it is difficult to draw any firm conclusions given the rather questionable quality of the dependent variable. The results of the study also suggest a significant counterdeterrent effect of audit activity, once the measure of fair and respectful treatment is included as a regressor for noncompliance. This seems counter-intuitive.

Another approach employed by researchers to assess the impact of taxpayer services on tax compliance has involved laboratory experimentation (Alm and McKee, 2007). Specifically, subjects played a tax compliance game in which there was initially uncertainty regarding the amounts an individual could legitimately claim with respect to a specific credit and a specific deduction on their tax return. In some variants of the game, the subjects had access to “taxpayer services” that helped to resolve this uncertainty. A preliminary comparison of results from the variants of the game with and without these services indicated that the services had a negligible impact on compliance behavior.

A third approach to this research problem involved an observational study of the impact of various targeted outreach programs that were designed to educate various types of small business owners about their requirements to file particular types of information returns (Adelsheim and Zanetti, 2006). The outreach programs were delivered by mass media to selected geographic locations, while similar geographic locations outside of the delivery area were used as a quasi-control group. A difference-in-differences methodology was employed to compare changes in relevant information return filings among the treatment and control group areas after and before the outreach program was delivered. For some outreach programs, filings went up (indicating at least some degree of program success), while in others no significant improvement was observed. A lesson from both this approach and the aforementioned laboratory experiment is that there are many ways to structure taxpayer services, and it is difficult to generalize results from one study regarding a particular service offering to another service offering that differs either in terms of structure or in terms intended recipients.

In one of the controlled field experiments discussed in the previous subsection, an additional experiment was carried out in which randomly selected taxpayers were mailed an offer by the Minnesota Department of Revenue to receive enhanced customer service (see Coleman, 1996, for a discussion of this additional experiment). Specifically, they were provided with a special phone number, through which they could receive tax information, including state and federal tax forms and assistance with both their state and federal tax returns. Ordinarily the Department does not provide federal tax assistance. The taxpayers were also provided with a special information booklet and other useful information, including various departmental contacts. The study did not find any significant impact of this enhanced service offering on overall tax compliance. Interestingly, taxpayers who were offered enhanced telephone services

had a slightly lower rate of take-up for these services (14 percent) than other taxpayers had for ordinary telephone assistance (16 percent).

3. Deterrence Approaches Used in Environmental Compliance⁴⁰

There are many other agencies that have a requirement to monitor, assist, and enforce behaviors of residents and corporations in the U.S. to a specific body of legislation. They include the Environmental Protection Agency (EPA), the Occupational Safety and Health Administration (OSHA), the Mining Safety and Health Administration (MSHA), the Department of Justice (DOJ), Immigration and Customs Enforcement (ICE), and the U.S. Coast Guard (USCG), just to mention a few. While we have included several papers related to the environmental literature, we focus on two recent review articles, by Shelley Metzenbaum and Jay Shimshad, in the bibliography (these papers also have substantial references to the environmental literature). An initial review of these review papers reveals that the basic underlying deterrence model is quite similar to that used in the tax community. Indeed both of these papers contain references to the seminal papers in the tax literature that describe the underlying theoretical framework for compliance.

It is safe to say that the environmental literature is not as old or as mature as that for tax compliance, and the population over which the environmental legislation holds is somewhat less diverse. The tax world includes individuals, corporations, partnerships, tax-exempt entities, etc. The environmental compliance situation is focused on corporations and their responsibilities with respect to a range of environmental legislation, including release of toxic substances, environmental cleanup, etc. However, we believe that there are number of important lessons to be learned from this literature.

First, environmental compliance disciplines have, very early on in their history, focused on the link to social psychological components of compliance, and we believe this is clearly an area in which the tax compliance community can benefit from additional research work. Second, they clearly recognize the importance of timing of the service or enforcement events and the linkage to subsequent environmental outcomes. Therefore, there are number of articles referenced in the reviews by both Metzenbaum and Shimshad that employ time lag variables intended to understand the latencies in the environmental outcomes following a specific stimulus.

Finally, we want to point out that there are many similarities between the two literatures. First, the review papers emphasize the challenges of estimation and needs for better, cleaner and more complete data sources to support the compliance estimation process. The reviews mention the need for statistically reliable surveys and more empirical studies to test theoretical models. There is also an interest in the extent to which environmental activities targeted to specific clients and the dissemination of information (such as publicity surrounding large fines for corporations in a specific area) have “spillover effects” that serve to improve overall compliance within an industry or geographic area. Researchers in this field share the challenge of defining a set of metrics that can be measured accurately and employed to effectively isolate the impact of the intervention under study on compliance from the impact of other potential confounding factors.

While we have included a modest number of studies of compliance issues outside of the tax domain in this bibliography, it is our intention that other papers will be added to later version of this document through the involvement of experts in environmental compliance activities in the workshop that will be held during this design phase of this initiative.

⁴⁰ Ideally, this project plans to investigate parallel policy compliance approaches used by a variety of other fields and governmental compliance areas. At present, only the environmental literature has been reviewed.

III. Bibliography: Papers Selected for In-depth Annotation

Theoretical

1. Alm, J., A perspective on the experimental analysis of taxpayer reporting. *The Accounting Review*, 1991, 66(3): p. 557-593.

Objective:

Alm reviews the application of experimental methods to compliance research, including the appropriate procedures to be followed in experimental analysis, results of previous studies and limitations of experimental methods.

Content:

Alm reviews two studies in depth. Beck et al. (1991) confirms earlier work on the impact of audit and penalty on reported income yet suggests that the effect of an increase in tax rates may increase reported income. Uncertainty influences compliance yet the effects vary based on risk preferences and other experiment parameters (tax, audit, penalty). Collins and Plumley (1991) suggests that audit schemes affect compliance levels, with random audits resulting in the lowest compliance levels.

Conclusions:

Alm concludes that experimental economics provides information on individual behavior in the absence of reliable data. He notes several new factors which seem to affect compliance; in particular, experimental economics suggests that taxpayers overweight the probability of detection and value public goods, while social norms and perceptions of equity affect taxpayers' compliance levels.

2. Andreoni, J., Erard, B., and Feinstein, J., Tax compliance. *Journal of Economic Literature*, 1998. 36(2): p. 818-860. (<http://www.jonathanfeinstein.com/>)⁴¹

Objective:

This paper provides an excellent review of the literature and important research issues and questions relating to tax compliance.

Content:

The paper begins with a description of tax compliance in general, with definitions and basic facts. The beginning section also talks about the core data sources, many of which are managed and disseminated by the IRS. The paper looks at both theoretical and empirical research papers. The theoretical methods discussed include the initial work of Allingham and Sandmo using a basic expected utility model and some of its generalizations. A second major area of theoretical work reviewed in this paper relates to the situation in which the taxpayer and the tax authority interact, using either principle agent approaches or game-theoretic models.

The second half of the paper focuses on empirical research on tax compliance. The review examines survey research approaches, laboratory experiments, as well as field experimentation and the challenges of fitting data to mathematical or statistical models. The authors report the impact of this empirical work on a wide range of tax compliance inquiries, generally based around key variables such as tax rates, audit rates, demographic variation, use of penalties, subjective beliefs of taxpayers, and the influence of tax practitioners. Lastly, the authors explore the theoretical contributions regarding psychological, moral, and social influences on behavior of taxpayers. These include: moral rules and sentiments; perceptions of equity and fairness; citizen satisfaction and governmental legitimacy.

Conclusions:

The authors conclude that the progress of the last 25 years of tax compliance research reassures them that “tax compliance and enforcement systems are amenable to systematic research.” However, they conclude that “a greater synthesis of theory with empirical research” is needed. A second major finding is that “more work needs to be done exploring the diverse psychological, moral and social influences on compliance behavior.” A third conclusion is that “greater attention should be paid to the dynamic and complex institutional framework of tax compliance.” Finally, they believe that “there is a need for more empirical and institutional research within jurisdictions outside of the U.S.”

⁴¹ The access data for all web documents is considered to be the date of this deliverable – July 18, 2008.

3. Kornhauser, M.E., Normative and cognitive aspects of tax compliance: Literature review and recommendations for the IRS regarding individual taxpayers. *Annual Report to Congress*, 2007. 2. (http://www.irs.gov/pub/irs-utl/arc_2007_vol_2.pdf)

Objective:

This study was commissioned by the National Taxpayer Advocate to better understand the role of “tax morale” in compliance through behavioral research. The paper discusses norms impacting tax compliance, cognitive processes that influence individual taxpayer behavior and the frequently ambiguous role of demographic variables (gender, age, education, marital status, religion and income).

Content:

Based on a review of the behavioral literature, Kornhauser concludes that the IRS “...can activate compliance in a variety of ways including education, properly framing communications, fair procedures and a regulatory framework that incorporates current and future findings of tax morale research in its operations and dealings with taxpayers.” The author outlines the following dynamics that should be considered in compliance strategy: first, the importance of “framing” in behavioral response (for example, taxpayers will be risk averse with respect to gains); second, the importance of cognitive “rules of thumb.” These are integrally related to workviews and values, such as reciprocity and cooperation vs. independence and self-reliance. The success of IRS outreach and media will be determined by how these tax authority postures are perceived by the taxpayer. Additionally, Kornhauser cautions that traditional enforcement-based strategies may encourage perverse taxpayer outcomes. For example, publicizing the tax gap may undermine morale by diminishing taxpayer perceptions of reciprocity, while enforcement efforts that are rude or arbitrary may undermine perceptions of procedural justice and legitimacy.

Conclusions:

This report recommends that IRS establish a permanent organizational unit to perform and manage behavioral science of compliance, as well as implement educational and media programs on compliance. The author advocates a “tax morale” / “responsive regulation” approach to improving compliance, incorporating attention to psychological and cognitive factors compliance motivators.

4. Bikhchandani, S., Hirshleifer, D., and Welch, I., Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 1998. 12(3): p. 151-170. (<http://pages.stern.nyu.edu/~wgreene/entertainmentandmedia/Fads.pdf>)

Objective:

This article argues that learning by observing the past decisions of others can help explain some otherwise puzzling phenomena about human behavior.

Content:

The authors consider two scenarios. In the *observable actions* scenario, individuals can observe the actions but not the signals of their predecessors. They compare this to a benchmark *observable signals* scenario in which individuals can observe both the actions and signals of predecessors.

Conclusions:

These authors posit that “learning from past decisions of others” can explain patterns of convergent behavior and fluctuations, which traditional economic models fail to explain. Consistent with the “pervasive but fragile herd behavior [implied by informational cascades theory],” this study finds that informational cascades are triggered by a small amount of information. Under informational cascades, the system spontaneously fluctuates until it reaches a precarious resting point in which behavior is sensitive to small shocks.

5. Caballe, J. and Panades, J., Cost uncertainty and taxpayer compliance. *International Tax and Public Finance*, 2005. 12: p. 239-263.

Objective:

This paper examines the effects of multiple sources of uncertainty in a game-theoretic context with strategic interaction between auditors and taxpayers. In addition to the typical uncertainty associated with taxpayers' income (the tax authority does not know true taxpayer income), the model introduces two sources of uncertainty: the cost of audit to the taxpayer is not known by the auditor, and the cost of the audit and audit strategy is not known to the taxpayer.

Content:

The authors use a mathematical model with derivation of the sequential equilibrium.

Conclusions:

The authors find that the addition of variation to explain uncertainty increases tax compliance. The effects of the uncertainty about the audit cost faced by the tax authority are generally ambiguous. Their main results can be summarized directly (in their own words):

- Larger variance of the distribution of taxpayers' cost results in more average income reported, less average audit effort, more net revenue for the government, and more expected welfare for the taxpayers.
- Larger variance of the income distribution results in less average income reported, more average audit effort, less net revenue for the government, and less expected welfare for the taxpayers.
- Larger average audit cost borne by the tax agency typically results in less average income reported and less net revenue for the government.
- The relation between the average expected tax rate and the true income could be nonmonotonic. Therefore, the tax system could be locally progressive on some range of income levels and locally regressive on another range.

6. Davis, J.S., Hecht, G., and Perkins, J.D., Social behaviors, enforcement, and tax compliance dynamics. *The Accounting Review*, 2003. 78(1): p. 39-69.

Objective:

This paper analyzes the effect of social norms and enforcement on the dynamics of taxpayer compliance.

Content:

The authors develop two complementary theoretical models to evaluate compliance behavior through the movement between classes of compliant and noncompliant taxpayers. The first model is a differential equations, aggregate compliance model based on an epidemiological approach, where tax evasion behavior is infectious at a given rate r . The authors model the flow between evader and honest classes based on two key dynamics: enforcement (multidirectional) and social norms (from evader to honest). The second model is an individual level agent-based model, which validates the aggregate model results.

Conclusions:

The authors' analysis suggests that the efficacy of enforcement depends on the preexisting levels of compliance in a population. The authors describe nonlinear impacts of enforcement on compliance: "Compliant populations are insensitive to changes in enforcement policies until enforcement becomes sufficiently lax, when they observe a sudden shift to high levels of noncompliance in equilibrium. In contrast, relatively noncompliant populations respond to increased enforcement by gradually increasing compliance. Then, when enforcement becomes sufficiently harsh, they find a sudden shift in equilibrium to very high levels of compliance." Thus, after epidemiological shifts, their models predict that returning to the previous enforcement policy will not cause the population to return to its previous state. Results provide a possible explanation for "why taxpayer compliance varies across time and across geographic regions, even under similar enforcement regimes."

7. Plumley, A.H. A framework for optimal tax administration. *IRS Research Conference*. 2007. Washington DC: Publication 1500. (<http://www.irs.gov/pub/irs-soi/07resconf.pdf>)

Objective:

This paper attempts to provide a framework for how a tax administration agency might address resource allocations. Specifically, what are the tax objectives, how can resources be applied most efficiently to meet them, and what are the short and long term planning steps that can get the agency closer to these goals?

Content:

This paper uses a benefit-cost framework to lay out a series of principles and steps that the IRS could decide to take to improve tax administration.

Conclusions:

Plumley provides 15 principals (in his own words) which he believes critical to improved tax administration:

- the stakes are huge
- optimal tax administration can be modeled as a constrained optimization problem, with allowance for of-model judgments when necessary
- model features that cannot (yet) be quantified empirically or theoretically can in some cases be quantified (or at least plausibly bounded) by consensus assumptions
- the optimal solution to the constrained optimization problem is the allocation of resources that equalizes the marginal benefit/cost (“bang for the buck”) across all opportunities to use those resources
- whether allocating the budget across organizations and programs, selecting discretionary workload within programs, or selecting discretionary issues to pursue on a given case, the same ultimate objective should be applied
- other (more common) approaches to tax administration almost certainly do not lead to optimal decisions, will often lead to internally inconsistent decisions, and should be avoided.
- the current IRS mission statement is not specific enough to identify, by itself, the optimal way to administer the tax laws.
- the ultimate objective of the IRS is to maximize the weighted net benefits (weighted benefits minus weighted costs), subject to the relevant constraints.
- if we take all of the right benefits and costs into account, then it is not optimal to pursue noncompliance that is not cost-effective at the margin.
- reduction of taxpayer compliance costs (burden) is not an independent objective.
- customer satisfaction and employee satisfaction (as measured by surveys) are also means to achieve our ultimate objective; they are not ends that compete with that objective.
- senior IRS decision makers need to develop consensus on the components and form of the ultimate objective
- IRS needs to take a long-term view of compiling the right data and developing the estimates necessary to model optimal resource allocation, etc
- IRS needs to estimate marginal direct and indirect benefit/cost for each activity (both enforcement and service activities) as a function of resource levels

8. Raskolnikov, A., Crime and punishment in taxation: Deceit, deterrence, and the self-adjusting penalty. *Columbia Law Review*, 2006. 106: p. 569-642. (<http://www.columbialawreview.org/pdf/Raskonikov.pdf>)

Objective:

This article examines a weakness in the application of audit sanctions and “proposes a new type of penalty that would strengthen tax enforcement while improving efficiency.”

Content:

Raskolnikov’s reasoning is as follows: “Economic analysis of deterrence suggests that rational taxpayers choose avoidance and evasion strategies based on expected rather than nominal sanctions. This paper argues that many taxpayers do just that. Because the probability of detection varies dramatically among different items on a tax return while nominal penalties do not take the likelihood of detection into account, expected penalties for inconspicuous noncompliance are particularly low. Adjusting existing penalties will not solve the problem because what is (and is not) inconspicuous depends on a given return and, therefore, is not susceptible to the type of generalization on which the current penalties rely. This Article offers a novel solution. Because taxpayers often hide aggressive subtractions (such as deductions, credits, and losses) by mixing them with legitimate subtractions of the same type, this paper proposes to set the new penalty to equal a fraction of the legitimate subtraction reported on the same line of a tax return that contains the illegitimate one. With this penalty in place, the harder it is for the government to find an aggressive transaction, the higher is the statutory sanction if the transaction is detected. As a result, the inefficient incentives to hide noncompliance are diminished and deterrence is improved.”

Conclusions:

This article has identified a considerable weakness in the existing enforcement regime, one which “induces wasteful behavior, increases the welfare cost of noncompliance, and lowers overall deterrence.” The author concludes that through a gradient penalty based on the probability of detection, the tax authority can counter taxpayer incentives to hide aggressive transactions. Moreover, Raskolnikov concludes that taxpayer ability to deceive the government is disparate between taxpayers who have few opportunities to reduce their tax liabilities (for example, earn income subject to withholding or information reporting and take a standard deduction) and the taxpayers with significant opportunity to evade: taxpayers who receive income not subject to any monitoring regime, take numerous deductions that are unlikely to be examined, and engage the services of creative, aggressive tax professionals. The proposed penalty aims to diminish the socially undesirable incentives to conceal tax by strengthening the sanction regime.



9. Snow, A. and Jr, R.S.W., Ambiguity about audit probability, tax compliance, and taxpayer welfare. *Economic Inquiry*, 2005. 43(4): p. 865-871.

Objective:

This paper introduces ambiguity preferences that allow taxpayer welfare to depend nonlinearly on the probability of audit. Taxpayer uncertainty about the probability of audit can be systematically biased from the true probability, depending on ambiguity preference.

Content:

In constructing a measure of ambiguity preferences, this paper relies on reported data that a large proportion of the population (70.79%) are ambiguity averse, slightly more are utility maximizers (26.16%), while a small number (4.5%) are ambiguity lovers. This paper uses theoretical estimation of a utility function.

Conclusions:

The authors state that for the ambiguity averse portion of the population, "...a policy of deliberately fostering uncertainty about the probability of audit reduces tax evasion." For the small number of ambiguity lovers, such a policy will have perverse effects. The tax authority has no practical means of differentiating taxpayers based on ambiguity preferences, yet should recognize that the presence of a small number of ambiguity lovers may complicate a policy of fostering deliberate audit ambiguity.

10. Torgler, B., Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *Journal of Economic Surveys*, 2002. 16(5): p. 657-683.

Objective:

This paper looks at the experimental findings of research concerning tax morale and tax compliance, focusing on personal income. After a brief discussion of the topic of deterrence, the main focus is on the social and institutional factors.

Content:

This paper is a literature review. It discusses the traditional tax compliance literature, such as the threat of detection and punishment and the level of tax rates, but extends the discussion to a review of the literature regarding social factors as well as institutions.

Conclusions:

Examining most tax compliance models, Torgler finds that the observed level of tax compliance is higher than predicted. Torgler concludes that "Experiments that analyzed the effect of deterrence determinants have given mixed insights into the changes in tax compliance in response to different policies. However, despite the mixed results, the findings tend to suggest that a higher audit rate leads to more compliance and that tax compliance is an increasing function of income and a decreasing function of the tax rate." This survey reports results on the impact of social and institutional factors while holding the probability of penalty, penalty rate and taxpayer's risk aversion constant. Torgler also concludes that, "experiments which consider the interaction between subjects indicate that moral constraints work as a disincentive to evade taxes."

11. Turk, A., et al. Charitable contributions in a voluntary compliance income tax system: Itemized deductions versus matching subsidies. *The IRS Research Bulletin*. 2007. Washington D.C.: Publication 1500. (<http://www.irs.gov/pub/irs-soi/07resconf.pdf>)

Objective:

Because charitable contributions are not subject to information reporting, there may be alternative policies to improve compliance (as well as increasing equity and decreasing taxpayer burden). In the experimental design in this paper, taxpayers instead “receive utility from total contributions to the public good and from a private good that is based on the amount of the individual’s charitable contribution.” The impacts of matching, instead of allowing deductions, are considered from the burden, administration, and compliance perspectives.

Content:

The authors describe this experiment, “This paper explores optimal subsidized contribution behavior in the context of three cases. In the first, an individual’s utility depends only on his or her consumption of a private good and of a public good toward which he or she may contribute. Contributions are subsidized by either a rebate or a match. In the second case, contributions produce a private benefit (e.g., a “warm glow”), as well as the public good, and, again, contributions are subsidized either by a rebate or a match. Finally, we alter the second case by introducing required reporting of contributions, accompanied by random Government audits and penalties for overstatements. This paper proposes a laboratory experiment designed to test several of the hypotheses raised by the preceding theoretical model. The experiment consists of several stages, each including elements designed to test specific hypotheses. In the first stage of the experiment, subjects participate in an activity in which they can voluntarily contribute to a public good without any subsidy. In the second stage, they will examine the difference between the contribution-matching and rebate systems for Government support of charities. Two subsequent variations on the experiment will measure the price elasticity of donations and test whether the value of the subsidy rate influences donor behavior. Finally, they will introduce noncompliance in reporting charitable contributions and examine changes in donor behavior when the subsidy is based on reported, rather than actual, contributions.”

Conclusions:

The authors report, “Experimental research has shown that the matching subsidies can actually increase charitable contributions. This result is inconsistent with most simple economic models. However, a minor departure from these models can predict behavior that is consistent with the experimental research.”

Econometrics

12. Alm, J., Bahl, R., and Murray, M.N., Audit selection and income tax underreporting in the tax compliance game. *Journal of Development Economics*, 1993. 42: p. 1-33.

Objective:

Using data from Jamaica, this paper estimates the determinants of individual audit selection and individual tax evasion behavior in a way that treats the tax agency and the taxpayer as strategic players in the compliance game (as cannot be done with aggregate econometric analysis).

Content:

The authors use two data sets from Jamaica, which include direct, independently derived estimates of individual income tax evasion based on audited income tax returns for self employed 1980-82, and detailed individual tax return data from the population from which the audits were selected. The data sets, therefore, provide information on both audit selection factors and the amount of underreporting found. They (1) estimate audit selection based on taxpayer reported line items, (2) estimate likelihood on underreporting based on income, marginal tax rate, benefits to which payroll contributions entitle the taxpayer, and various socioeconomic variables, and (3) estimate level of income underreporting and level of tax underreporting. The authors use a three stage estimation process, referencing Heckman and Erard: bivariate probit analysis is applied to the pooled data sets to estimate the factors determining return selection for audit and likelihood of noncompliance. Selectivity bias controls are employed. However, results are generally robust with two alternative specifications. Lastly, the authors also simulated the impacts of fundamental tax reforms.

Conclusions:

The authors conclude that higher income and marginal tax rates increase underreporting. Increased payroll benefits (social program benefits, such as social security for example) decrease underreporting, as does wage income (i.e. withholding). Once taxpayers report, there is no impact of family size or capital allowances on the level of underreporting. Game theoretic approaches are relevant to tax compliance. In this instance, they allow the taxpayer behavior to impact selection and the agency's behavior to be determined based on the taxpayer's behavior. Additionally, governmental fiscal structure has a major impact on tax compliance behavior.

13. Bergman, M. and Nevarez, A., Do audits enhance compliance? An empirical assessment of vat enforcement. *National Tax Journal*, 2006. 59(4): p. 817-832. (<http://ntj.tax.org/> – available only to members of the National Tax Association)

Objective:

This paper analyzes the efficacy of audits in improving individual compliance with the VAT in Chile and Argentina.

Content:

In this paper, the authors employ OLS regression, where the dependent variables are the percentage change in the tax report pre- and post-audit. These are regressed on assessments, size, region, increase in debits to control for tax payers due to natural increase in sales, negative previous balance (similar to NOLs, credit for VAT from prior return period), and No-Compliance (Chile, previously audited without any change). The data is complex, since Argentina and Chile differ in the level of data available, as well as the behavior and enforcement approaches.

Conclusions:

The authors find that net payments are lower among the audited group than the control group. Taxpayers in Chile and Argentina “compensate” for the cost of an audit through subsequent noncompliance. However, previously compliant taxpayers in Chile (as indicated by previous “no change” audit) “pay considerably more in payments post-audit than other sanctioned taxpayers.” This suggests that in an environment of relatively high noncompliance, the effects of audit depend on the taxpayer: “cheaters” increase cheating, while more compliant taxpayers increase compliance.

14. Bloomquist, K.M. Trends as changes in variance: The case of tax noncompliance. *IRS Research Conference*. 2003. Washington DC: Publication 1500.

Objective:

Within the context of a purported rise in global tax evasion, this paper aims to explain the phenomenon, at least in part, “as a shift in the distribution of taxpayer income away from more visible to less visible sources.”

Content:

The authors use OLS econometric regression to analyze the impact of the shadow economy and social inequality on tax reporting. The independent variables are Income inequality and the size of the shadow economy while the dependent variable is the total 1990 tax revenues as a percentage of GDP. This paper uses TCMP data, SOI data, and Gini Coefficients from a World Bank study (as a measure of income inequality).

Conclusions:

The author demonstrates a correlation between income inequality and size of the shadow economy for a sample of developed and developing nations. Bloomquist concludes, “Taxpayers simply may be enjoying greater success at evasion due to a decline in transaction visibility. U.S. taxpayer data show that nonmatchable income has grown from 8.7 percent of reported AGI in 1980 to 18.4 percent in 2000. Applying constant misreporting rates for matchable and nonmatchable income from the 1988 TCMP study, taxpayer underreporting is estimated to have increased from 3.6 percent to 5.6 percent of reported AGI between 1980 and 2000.” Preliminary evidence is found for the hypothesized relationship between income inequality and a measure of tax noncompliance for a sample of 23 developed and developing nations. This finding supports the view that a widening variation in taxpayer incomes, and the associated decline in transaction visibility, could be contributing to the presumed growth in taxpayer noncompliance.”

15. Das-Gupta, A., Ghosh, S., and Mookherjee, D., Tax administration reform and taxpayer compliance in India. *International Tax and Public Finance*, 2004. 11(5): p. 575-600.
(<http://people.bu.edu/dilipm/publications/DasguptaGhMookITPF.pdf>)

Objective:

This paper develop a model of taxpayer compliance decisions within the context of Indian assessment units comprising self-employed taxpayers, where taxpayers are grouped into different assessment units based on the incomes disclosed.

Content:

The authors utilized econometric regression (based on Allingham and Sandmo model with self-selection modifications) of aggregate tax returns with respect to personnel assigned and enforcements efforts of tax administration personnel (who perform exams, etc with virtually no functional specialization). The following variables are regressed on prepaid tax revenues: scrutiny / summary assessment probabilities, indicators of work load (support staff), penalties, and appeal success. Information was obtained from data on 49 tax units in 3 major cities handling small business and professional (self-employed) taxpayers for 2 successive years (89-90, 90-91) including workload and characteristics of the assessment unit, revenues, assessment and follow-up.

Conclusions:

Das-Gupta et al. find significant, positive impact on compliance from increased penalty and enforcement levels. Other results relate specifically to Indian tax administration. Results provide evidence that taxpayers strategically self-selected into assessment units via reporting income levels, while spillover effects between units can be seen from variations in support staff or enforcement effort. Policy consideration should be given to random assignment rules for assessment rather than the current unit assignments to prevent strategic underreporting incentive, workforce expansion could increase revenues, reallocating support staff to most productive locations, etc.

16. Dubin, J.A. Criminal investigation enforcement activities and taxpayer noncompliance. *The IRS Research Bulletin*. 2004. Washington DC: Publication 1500. (<http://www.irs.treas.gov/pub/irs-soi/04dubin.pdf>)

Objective:

This study extended the results from a previous paper by this author, and his co-authors, (Dubin, Graetz and Wilde, 1990) to seek to measure the impact of Criminal Investigation (CI) on compliance as well as to extend the results from an earlier paper to a larger span of time.

Content:

This paper included all of the variables from the previous DGW papers but added several related to CI. The dependent variables include Assessed Liability per Return (ALR), reported tax per return (RTR), and Returns per Capital (RCAP). Predictor variables included several measure of audit rates (to show different activity by Revenue Agents, Tax Agents and correspondence audits). The paper employed instrumental variables to avoid the concern that compliance metrics were correlated with the error term of fitted equations. The papers expanded on the number of socioeconomic variables included in the study and, of course, added the variables for CI enforcement factors for both tax evasion and money laundering cases. The author also included information on media reports of indictments to determine the impact of publicity on compliance. A unique issue for this paper was the focus on recommendations for prosecution in addition to the actually number of indictments. The data was obtained from the Annual Report of the Commissioner of the Internal Revenue for the years 1977-1986. The data were generally recorded and used at the state level and by district. The tax collections and examination variables were obtained from the Annual Report of the Commissioner of the Internal Revenue, IRS Data Book, and IRS Statistics of Income Bulletin. The primary focus methodologically of this paper was through an empirical, econometric analysis of the relationships between the compliance measures and enforcement and demographic variables. The author fits more and more complex equations with new factors added at each stage. The author's analysis is quite complete and substantial. After fitting model equations, the author performs what he describes as "simulations." The author essentially suggests using the model equations with proposed (new) levels for some of the enforcement variables to study the potential impact (i.e., what happens if we double the number of audits?).

Conclusions:

The author concludes that CI activities have a measurable effect on voluntary compliance. He finds that the mix of sentenced cases (tax versus money laundering) is not a significant determinant of tax compliance. Finally, he claims that media attention shows some weak evidence of increasing compliance. The author also revisited some of the analysis performed in the DGW paper of 1990 to see the impact of a larger time frame of analysis. In some cases, the findings in the earlier paper were confirmed, and in other cases the findings changed.



17. Dubin, J.A., Graetz, M.J., and Wilde, L.L., Are we a nation of tax cheaters? New econometric evidence on tax compliance. *The American Economic Review*, 1987. 77(2): p. 240-245.

Objective:

This paper reviews the theoretical as well as the empirical work on tax compliance and uses an instrumental variables econometric approach to empirically examine the effects of audits and socioeconomic factors on compliance.

Content:

The data was obtained from The Annual Report of the Commissioner and socioeconomic and demographic variables. The data was analyzed using econometrics and instrumental variables. The dependent variable is the percentage return per audit for individual returns (additional tax and penalties from audit / total collections per million audits, in 1972 dollars). The independent variables include lagged values of audit rate (exams per 100 returns filed), percent adult population with HS education, percent of population over 45 yrs, per capita income and its square, percent of work force employed in manufacturing, and time trend. The instruments include budget per return, percent of individual returns filed.

Conclusions:

Dubin et al. find that:

- The audit rate is endogenous.
- “There is a deterrent effect associated with increases in audit rate, but in equilibrium it is dominated by the IRS’s incentive to audit according to expected yield”
- “Compliance increase with per capita income at a decreasing rate”
- “Significant negative time trend in the audit rate and compliance, surprisingly not found when individual collections are examined.”
- Dubin, Graetz and Wilde also use an instrumental variables approach to conclude that audits are endogenous and have a positive significant deterrent effect. More research is required into the time trends within audit and compliance data.

18. Dubin, J.A., Graetz, M.J., and Wilde, L.L., The effect of audit rates on the federal individual income tax, 1977-1986. *National Tax Journal*, 1990. 43(4): p. 395-409.

Objective:

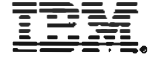
This paper attempts to investigate empirically the role of audits in federal revenue collection, including “spillover effects.” (i.e. indirect effects). Their principal innovation is estimating directly taxes due.

Content:

This paper utilizes data from The Annual Report of the Commissioner (collections, number of returns, refunds, exams, taxes and penalties, budget) generally at the state (IRS district) level, and well as SOI total tax liability. The authors use econometric with instrumental variables and corrective transformations. The dependent variables included (1) reported individual income tax / number of returns filed and (2) filing rate as measured by returns per capita. These are constructed using total tax returns filed, number of individual income tax returns filed, number of income tax returns examined, additional tax and penalty recommended after exam, and costs incurred by IRS (Commissioners Report); tax liability (SOI). The independent variables include an estimated impact of change in enforcement strategies, the number of information returns (other than W2s) / number of individual returns filed, and demographics such as percent with HS education, percent labor force in manufacturing and services industries, number of farms / population, audit coverage, unemployment rate, income, and average state income tax.

Conclusions:

The authors find that audit rate is endogenous and has a positive significant effect for reported tax. Variables related to the tax base were significant in all 3 equations: unemployment negatively and income positively. Higher education and manufacturing percentages result in higher filing. Audit rate has a positive effect, with important policy implications. IRS should reach out to nonfiler populations where they can be identified (i.e. communities with lower education levels, etc) and should simplify the return filing process.



19. Erard, B. and Ho, C.-C., Searching for ghosts: Who are the nonfilers and how much tax do they owe?
Journal of Public Economics, 2001. 81: p. 25-50.

Objective:

This paper examines compliance with income tax filing requirements, the characteristics of nonfilers and the factors driving their decision not to file a tax return, as well as measures their unpaid tax liability.

Content:

The authors begin by considering a standard model of taxpayer reporting behavior and then extend the model to account for nonfiling as a strategic option. They develop an econometric framework for analyzing the decision whether to comply with one's income tax filing requirement. The data used for filers of 1988 federal income tax returns is based on a 25 percent random subsample of the IRS TCMP Phase III Survey. The data on potential nonfilers is from the collection-based segment of the IRS TCMP Phase IX Nonfiler Survey for tax year 1988.

Conclusions:

The results of this study indicate that only 57 percent of the potential nonfiler population could be located through an intensive search. However, locatable nonfilers apparently account for a disproportionate share of all unpaid taxes.

20. Hasseldine, J., et al. Carrots, sticks, sole proprietors, and tax accountants. *IRS Research Conference*. 2005. Washington DC: Publication 1500. (<http://www.irs.gov/pub/irs-soi/05hasse ldine.pdf>)

Objective:

This paper analyzes the impact of tax authority notices on the sole proprietor population, based on a UK study in 2000.

Content:

This paper analyzes the change in tax reporting by 7,307 UK sole proprietors in response to a controlled field experiment, where taxpayers received one of five treatments: (1) a simple offer of assistance, (2) an appeal to citizenship, (3) a threat of audit, (4) a threat of audit which articulated penalties for noncompliance and (5) a notification of preselection for audit. Specifically, these sole proprietors report income below a particular threshold of income allowing for simplified reporting in the UK. Note: This experiment is most similar in approach to the US Minnesota experiment (Blumenthal, Christian).

The data is obtained from UK Inland Revenue for sample. Years 1999, 2000 are used for sample selection while 2000-2001 are analyzed for change in behavior.

Authors employ two types of statistical analysis of the field results. Chi square tests were used to analyze the percentage of the sole props reporting income below the threshold, broken into those who used a paid preparer or self-prepared. Difference-in-difference approach for overall treatment effects between the 5 groups and control group. The variables of this analysis include the percentage of the sole props reporting income below the threshold and the effects on the two fundamental components of reporting which are considered separately: income (“turnover”) and net profit (i.e. deductions).

Conclusions:

The authors showed that the percent sole props reporting income below the threshold decreased in all treatment groups. However, the “enablement” (offer of services) letter had the least effect. The sanction-based approaches all resulted in statistically significant increases in both turnover and net profit, of which preselection for audit was most effective, followed by audit with iteration of penalty, followed by simple notification of an increase in audits to be conducted. The penalty letter was most effective for taxpayers who used a paid preparer (suggesting that penalties hint at wrong-doing on the paid preparers part, more so than audit). Additionally, paid preparer treatment of deductions proves interesting: net profits may “have been more malleable than turnover (income), and therefore the treatments were more effective.” While any notification from the IRS appears to have some impact in this context, sanction-based approaches have the most impact on sole proprietors who prepare their own taxes. Notification of tax authority services had very little impact across the board.

21. Kamdar, N., Information reporting and tax compliance: An investigation using individual TCMP data.
Atlantic Economic Journal, 1995. 23(4): p. 278-292.

Objective:

This paper examines the determinants of tax compliance, particularly looking at the influence of third-party information reporting on compliance.

Content:

The authors use econometric regression to estimate drivers of tax compliance, using micro data from the 1971 examination cycle of the Tax Compliance and measurement program (TCMP). The independent variables are third-party information reporting and tax rate changes.

Conclusions:

The author states that the results indicate, as expected, that noncompliance is smaller on returns that do not include a Schedule D or Schedule E. This study also suggests that noncompliance increases as the marginal tax rate increases. Furthermore, the paper finds that taxpayer behavior is affected by noncompliance opportunities. One constraint on noncompliance is the existence of third-party information reporting requirements.

22. Kirchler, E., Niemirowski, A., and Wearing, A., Shared subjective views, intent to cooperate and tax compliance: Similarities between Australian taxpayers and tax officers. *Journal of Economic Psychology*, 2006. 27: p. 502-517.

Objective:

This study investigates how shared beliefs between taxpayers (both compliant and non-compliant) and tax officers in Australia affect tax compliance. The principal mechanism through which tax compliant behavior is affected is willingness to cooperate. The study builds on psychological theories of social interaction and “tax psychology” that emphasize the quality of interaction between taxpayers and the tax authority.

Content:

Taxpayer and tax officer attitudes and beliefs are measured across several dimensions including: tax law complexity, tax mentality, ethics, social norms and notions of horizontal and vertical equity. The authors hypothesize that shared beliefs across these dimensions will improve cooperation and voluntary tax compliance. A sample of 292 compliant taxpayers, 547 non-compliant taxpayers and 358 tax officers in the Australian Tax Office (ATO) were sent questionnaires prior to tax filing. The sample was chosen from a group of fully-compliant taxpayers in 1995 who were observed over the next five years. Taxpayers from this group were deemed non-compliant if they engaged in non-compliant behavior (e.g., under-reporting of income) in at least two of the five years. The sample frame was comprised of active, non-business taxpayers between the ages of 21 and 65. Taxpayers and ATO tax officers were asked their opinions on tax-related phenomenon. Responses were recorded on a Likert scale ranging from 1 (strongly agree) to 7 (strongly disagree) and similarity measures across each group were constructed. Mean responses were tested and multivariate analysis of variance were performed to isolate effects.

Conclusions:

The authors conclude that the three population groups (compliant taxpayers, non-compliant taxpayers and tax officers) differed significantly across all dimensions except notions of fairness as measured by horizontal and vertical equity. Willingness to cooperate is shown to be significantly related to reported compliance behavior and shared beliefs and evaluations between taxpayers and tax officers facilitate this willingness.

Tax laws that are understood by taxpayers and readily administered by tax officers can lead to similar beliefs and evaluations which is found to be a prerequisite for perceived procedural fairness: “...enabling, treating taxpayers fairly and reasonably, explaining rules and decisions, and providing reliable information and solutions to questions, will lead to improving the reputation of tax administrators and to perceived procedural fairness which may lead to an increased willingness to comply with the spirit of the law.”

23. Martínez-Vazquez, J. and Rider, M., Multiple modes of tax evasion: Theory and evidence. *National Tax Journal*, 2005. 58(1): p. 51-76. (<http://ntj.tax.org/>— Search for Martínez)

Objective:

In response to increased enforcement for one tax concept / line item, taxpayers may correspondingly shift noncompliance to another line item, using “multiple modes” (e.g. line items) of tax evasion. The paper develops a theoretical model of two modes of tax evasion.

Content:

The authors use simultaneous equations estimating income and deductions reporting compliance, two stage Tobit. Variables include marginal tax rate in each mode, total pre-tax income, proxies for probability of detection for each mode, demographic information. The dependent variables are the log of the income gap and the log of the deductions / adjustments gap. The independent variables are the log of total income, the log of the last dollar tax price, the proxies for objective audit probabilities (Wage share, interest and dividend share, itemized deductions share, and items without third party information reporting as dummy variables) and demographics (marital status, number of dependents, age and age squared). The data comes from TCMP 1985 stratified random sample of 49,162 returns, excluding EITC and returns where audit-adjusted filing status differs from the original reported filing status, returns with negative AGI,

Conclusions:

The results indicate that the elasticity of deduction gap respective to income is .8 (a ten percent increase in deductions reporting compliance results in a decrease in income reporting compliance of .8%), while the elasticity for income with respect to deductions is .1. While there is substitution between the two modes of evasion, increasing income compliance is less elastic than increasing deduction compliance. Authors also find that income reporting compliance decreases with an increase in the marginal tax rate (elasticity of -3.92). Sole proprietors, farm and rental income are positive and significant predictors of noncompliance, yet partnership income is not (consistent with Slemrod Minnesota study; these authors suggest that collusion is required among partners). Increasing enforcement in one mode will increase compliance in that target area partially offset by decreasing compliance in another mode.

24. Mete, M., Bureaucratic behavior in strategic environments: Politicians, taxpayers, and the IRS. *The Journal of Politics*, 2002. 64(2): p. 384-407.

Objective:

From the perspective of bureaucratic behavior, the authors model interactions between an enforcement agency, taxpayers, and elected officials, who serve as political principals. "Specifically, the study focuses on the joint analysis of IRS enforcement decisions and taxpayer compliance behavior in order to fully explore the extent of political controls over the agency. The primary objective in this study is not necessarily to distinguish between competing explanations, but rather to establish that the political context does affect compliance as well as enforcement. The goal is to encourage a more integrated study of the political system as a whole, including studies designed to test these alternative explanations of citizen's responses to elected officials." Mete suggests that earlier compliance work fails to incorporate Crusoe fallacy,

Content:

Statistical analysis utilizing correlation evaluates the relationship between audit rates and partisan changes in national politics on taxpayer compliance. This model estimates the effect of audits on compliance through simultaneous equations using three-stage least squares (3SLS), controlling for the reciprocal impact of audits on compliance. The independent variables are IRS audit levels and partisan changes in national politics, and the dependent variables are audit rate and noncompliance rate. The data comes from audit data compiled by Susan Long of TRAC from the IRS audit management information system for the 1965-980 period and combined it with IRS data on returns filed over the same period, as well as the Congressional Almanac, the Book of the States and IRS Taxpayer Compliance Measurement Program.

Conclusions:

Mete finds strong endogeneity among all audit classes. The effect of audit rates on noncompliance are significant for middle- and high-income returns as well as for low-income business, but not for middle- and high-income business returns. The political responsiveness variables support the hypothesized impact of politics on both taxpayer and IRS behavior from the president and Senate Appropriations Committee. Mete finds, "Noncompliance is significantly lower during Democratic presidential administrations. The fact that only presidents have this effect, and that the effect diminishes with income, appears most consistent with the bounded rationality hypothesis: taxpayers in low information settings evaluate the signals they receive from the political environment and change their behavior according to the perceived preferences of the president." Overall, the results suggest that there is a very strong reciprocal relationship between IRS audit levels and taxpayer compliance levels across states and over time and that taxpayer compliance behavior is also influenced by partisan changes in national politics.

25. Plumley, A.H., The determinants of individual income tax compliance: Estimating the impacts of tax policy, enforcement, and IRS responsiveness, I.R. Service, Editor. 1996, *Publication 1916* (Rev. 11-96): Washington DC. (<http://www.irs.gov/pub/irs-soi/pub1916b.pdf>)

Objective:

Focusing on indirect behavioral response, the paper seeks to develop estimates of the compliance impact of important tax policy parameters—most notably marginal tax rates. “This paper provides the first empirical basis for choosing the best combination of major IRS activities to improve voluntary filing and reporting compliance.”

Content:

Plumley’s econometric model consists of four compliance equations (FilingRate, IncomePct, OffsetsPct, and NetIncomePct), and one first-stage AuditRate equation. The independent variables include a rich set of explanatory variables grouped in the following categories: Tax Policy, Burden/Opportunity, IRS Enforcement, IRS Responsiveness, and Demographics/Economics. The dependent variables include filing compliance and reporting compliance (numerators are total income reported, total offsets reported, and net income reported while denominators use exogenous surrogates for reporting and filing obligations).

For filing compliance, returns filed comes from SOI and IMF while number of returns required is estimated from the CPS. In measuring reporting compliance, income and income offsets are from SOI data, while personal income is estimated from BEA data. Census, HHS and BLS sources are used for a variety of independent variables on tax policy and demographic data. IRS enforcement data variables for TDI nonfilers, refund offsets and Criminal Investigation convictions were received from ATTR, AIMS and CI tax conviction systems; service provision variables for IRS Responsiveness included toll free calls, returns prepared, correspondence and education from RMIS paper reports.

Plumley reports, “The data collected for this study form one of the most comprehensive datasets ever compiled of potential determinants of voluntary filing and reporting compliance, and took over eight years to assemble. Some IRS and external data were available as representative samples of individuals; certain IRS variables were available at the district level; and certain external data were available at the state level. All variables were aggregated to the state level and were compiled for a ten-year period: 1982-1991.”

Conclusions:

Importantly for this study, Plumley reports findings for IRS Responsiveness, which measures the provision of IRS services including return preparation and telephone calls. IRS return preparation assistance had a positive and significant impact on filing compliance and income reporting compliance, though little impact on offset reporting. Calls to the IRS had a weakly significant negative impact on income reporting, which Plumley highlighted as a candidate for further study. The other service variables studied – speed in refunds, volume of Taxpayer Services (TPS) correspondence and TPS educational efforts – provided no evidence of compliance effects.

Major drivers of the decision to file include burden, third party information documents and nonfiler notices, as does the filing threshold and demographic and economic conditions (“Of the two policy variables included, only the filing threshold (the sum of one’s standard deduction and personal exemptions) significantly influences the FilingRate and is strongly negative. All of the included burden variables significantly affect filing compliance—at least at the 10 percent level of significance. Both the matching of third-party information documents (IRPDocRate) and the issuance of TDI nonfiler notices ($\ln(\text{TDI}+1)$) provide a fairly strong deterrent against nonfiling. Refund offsets seem to have a negative impact on filing compliance, but it is too weak to be considered significant. Five of the six demographic and economic control variables included have significant and intuitive effects on the FilingRate”).

Plumley finds a “strongly significant, and positive deterrent effect of audits on the general population,” which he estimates to be eleven times as large as the direct effect on tax revenue. Criminal convictions also illustrated a deterrent effect with income and offset reporting. He concludes that more research into indirect effects of enforcement is desirable. His other findings regarding reporting include that the marginal tax rate is significant for low income taxpayers; above this level of income, Plumley hypothesized that marginal tax rates increases are offset by underreporting substitution. Burden has a weak negative impact, suggesting that some offsets are not claimed because taxpayers “view it as too burdensome.” Income reporting appears to be influenced by “high opportunity” variables (e.g. sole proprietors), but offset reporting does not. Plumley found no evidence of paid preparers



increasing underreporting – in fact, the impact of preparers on income reporting was strongly negative, perhaps because paid preparers leveraging business expenses offset income.

Plumley concludes, “this study provides the first quantitative basis for allocating IRS resources optimally across its various enforcement and non-enforcement activities. Perhaps it can also lead to a method for evaluating the effectiveness of those activities, since it controls for a wide variety of other determinants of voluntary compliance as well.”

26. Ritsema, C.M., Thomas, D.W., and Ferrier, G.D. Economic and behavioral determinants of tax compliance: Evidence from the 1997 Arkansas tax penalty amnesty program. *IRS Research Conference*. 2003. Washington DC: Publication 1500. (<http://www.irs.gov/pub/irs-soi/ritsema.pdf>)

Objective:

This paper covers the motivations of taxpayers voluntarily participating in the 1997 Arkansas Tax Penalty Amnesty Program, in both the noncompliance and the amnesty decision making. The two research questions asked were, what are the effects of taxpayer demographic and behavioral characteristics on the total amount of tax noncompliance, and what are the effects of taxpayer demographic and behavioral characteristics on the level of tax noncompliance relative to income?

Content:

The authors divided participants into intentional and non-intentional evaders. The same two questions are investigated for the segmented sample. The independent variables is regressed against the total tax due and tax due as a percentage of income to determine the affect of the tax variables, the demographic variables and the behavioral variables on the amount and level of noncompliance. The dependent variables are measures of noncompliance, including tax due (tax due is the amount of tax due as reported on the tax return) and tax due as a percent of reported income (tax due/total Arkansas income). The independent variables include tax-related variables (Total Arkansas income and scaled total income, opportunity to evade, prior state contact, knowledge of tax due, multiple amnesty returns filed, filing status), demographic variables (Age and education), behavioral variables (excuses for noncompliance - lack of money, unfairness, complexity, ignorance), behavioral variables (reasons for amnesty participation - family and friends, fear of being caught, morality, avoidance of penalties). The data for this study came from tax forms filed during the amnesty period, federal tax data-sharing list, and a survey of amnesty participants.

Conclusions:

The authors find:

- “For unintentional evaders it appears that tax due is a simple linear function of income, while for the overall sample and for the neutral group tax due is an increasing function of income (i.e., the coefficient on income is positive and statistically significant). However, for intentional evaders tax due is a decreasing function of income (i.e., the coefficient on income is negative and statistically significant).
- “As expected, the opportunity to evade is also positively related to tax due. Only for intentional evaders was prior state contact positively related to tax owed.
- “Whether a taxpayer filed for amnesty for more than one tax year is negative and significant only for the intentionally noncompliant group.
- “Only in the group of intentional evaders do we find that age is related to tax due. Education is inversely related to the amount of tax reported only for the complete sample and for the neutral group.”

In terms of practical tax policy advice, the authors suggest that moral appeals did not appear to be effective. Broader publicity regarding tax payment programs may assist unintentional evaders. The authors conclude that “taxpayer motivations may differ across groups and that all taxpayers will not respond to the same motivating factors. Tax administrators should use multiple enforcement measures to appeal to a broader spectrum of taxpayers.”

Experimental

27. Adelsheim, P.D. and Zanetti, J.L. The effect of targeted outreach on compliance. *IRS Research Conference*. 2006. Washington DC: Publication 1500.

Objective:

Test the hypothesis that targeted, industry and issuespecific educational outreach programs is effective in increasing voluntary compliance and changing taxpayer behavior in a positive way.

Content:

Compliance variables varied for the 2 major types of training. One training focused on tip reporting compliance in the restaurant industry, and the other focused on several of the income reporting forms compliance in the construction industry. For tip income the compliance measure was the proportion of taxpayers reporting tip wages on Form 941. For the construction industry the compliance measure was the proportion of business taxpayers in that industry reporting on the appropriate forms. Compliance with forms was obtained in the test versus of the “control” cities. “Test” cities and control cities (matched to the test cities using business metrics) were matched for an observational study. The training protocols were done in the test cities and the compliance metrics were computed in both the test and control cities, and compared.

Conclusions:

For the tip compliance training in the restaurant industry, the authors found that their hypothesis of an increase in compliance immediately following the training program was upheld. However, for the studies of construction industry training, the results were not conclusive.

28. Alm, J., Jackson, B.R., and McKee, M., Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal*, 1992. 45(1): p. 107-114. (<http://ntj.tax.org/> – search for Alm)

Objective:

This study estimates the individual responses to various policy changes under controlled laboratory conditions.

Content:

The experiment measures level of income reporting compliance with variation in parameters: tax, penalty, audit rate, and the presence of a public good. Subjects receive income (randomly generated, initial endowment of tokens), pay taxes (at an unannounced rate) on income voluntarily reported, face a penalty of audit (random, up to 20% of population – 1 of 5 students), which results in a fine (multiple of unreported income). Actual amount of subject participation fee depends on performance over 3 rounds.

Conclusions:

This study observes that higher tax rates decrease reporting, higher income increases reporting, and higher fines have a negligible impact. The elasticity of audit is approximately 0.17. The presence of a public good is weakly negative, yet the payoff to a public good is positive and highly significant. Authors conclude that, given the relatively small audit elasticity, there may be limited benefits from increased audit. Compliance may be increased when taxpayers perceive benefits of a public good yet unaffected by increase in fines / penalties. Therefore, government should consider alternative policy instruments (public good awareness) to enforcement, which may be ineffective.

29. Alm, J., Jackson, B.R., and Mckee, M. The effects of communication among taxpayers on compliance. *The IRS Research Bulletin*. 2004. Washington DC: Publication 1500. (<http://www.irs.gov/pub/irs-soi/04jacnta.pdf>)

Objective:

This study examines the roles of information dissemination and taxpayer communication on voluntary compliance.

Content:

The dependent variable is the compliance rate or amount of declared on income. The independent variables are accumulated wealth (over the rounds), probability of audit and a dummy variable for each type of treatment. Compliance of student subjects is tracked under three conditions: baseline compliance when individual knows whether she was audited only, compliance when “official information” on the number of audits is announced (mirroring government publication of audit rate), and when taxpayers are allowed to send each other unofficial messages regarding own audit results and truthfulness. The compliance was regressed using Tobit maximum likelihood methods and panel estimation methods with fixed effects.

Conclusions:

The authors find that preliminary results indicate that in all cases, “official announcement” of the audit rate does not improve compliance, yet unofficial information does increase the compliance rate. Preliminary conclusions suggest that publication of the audit rate by the tax authority may not improve compliance, while communication among taxpayers may increase compliance.

30. Alm, J., Jones, M., and Mckee, M. Taxpayer services and tax compliance. *IRS Research Conference*. 2007. Washington DC: Publication 1500. (<http://www.irs.gov/pub/irs-soi/07resconfalm.pdf>)

Objective:

The goal of this study to understand, and, if possible, quantify the relationship between critical IRS services, in this case information resources supplied, to the voluntary compliance of taxpayers.

Content:

Subjects faced a complex protocol for preparing a tax return (and many options) and were provided with different levels of information on interpretation of tax rules . Within this experiment, the tax authority could either provide the subject information regarding the interpretation of the rules or not. In instances where information is made available to participants, subjects were able to click on a button on the screen and the true levels of deduction and foreign tax credit are revealed before they file their taxes. Where no information is given to subjects, they learn the true values of these uncertain variables (true level of deduction and FTC) only after they have filed their taxes. Data was then collected to determine which information levels resulted in the most accurate tax return preparation.

Results and Conclusions:

The authors conclude that the preliminary results indicate that increasing enforcement effort increases all three types of compliance. However, providing uncertainty-reducing information has no significant impact on compliance. While the results of this work are not yet definitive, it is also clear that much work needs to be done on services.

31. Alm, J. and McKee, M., Tax compliance as a coordination game. *Journal of Economic Behavior & Organization*, 2004. 54: p. 297–312. (<http://unjobs.org/authors/james-alm>)

Objective:

The authors investigate compliance when audits are conducted on the basis of deviation of reported return items (mirroring DIF), with and without tolerance levels. Can taxpayers coordinate compliance behavior with “cheap talk” (mirroring tax advice) to achieve zero-compliance equilibrium, and if so can the threat of additional random audit overcome coordination?

Content:

The experiment measures level of income reporting compliance with variation in audit method (by deviation, by deviation with tolerance, and deviation in combination with audit) and in communication by players (allowed to discuss strategies). The authors place this within a game theoretic (coordination games without dominant strategy) framework within experimental economics.

Conclusions:

The authors found that subjects are typically unable to coordinate on the payoff dominant Nash equilibrium (test compliance rate of 55%) and coordination is further impeded by a cutoff (tolerance, 75%). Pre-play communication reduces the compliance rate to virtually zero, which is largely mitigated by the effect of a conditional random audit. This experiment suggests that a deviation-based audit strategy is often able to achieve high levels of compliance while the addition of a conditional random audit would achieve greater compliance efficiently. The following scheme would be suggested, “...conduct DIF audits first, then conduct conditional random audits if the DIF audits do not exhaust the audit budget, and also conduct a small subset of purely random audits to construct the DIF audit rule (and also generate some taxpayer uncertainty.”



32. Blumenthal, M., Christian, C., and Slemrod, J., Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, 2001. 54(1): p. 125-136. (<http://ntj.tax.org/> - search for Blumenthal)

Objective:

This paper explores one part of a 1994 Minnesota Department of Revenue field experiment designed to study the effectiveness of alternative enforcement strategies. (See also Slemrod, J. and Christian, M.B.C.,)

Content:

This study measures compliance by the change in reported income and in taxes paid. They investigate the change in the dependent variables as a percentage of federal Total Positive Income and the proportion of participants reporting larger values of the dependent variables in 1994 compared to 1993. The data comes from federal taxable income as reported on the Minnesota individual income tax form. Two letters containing different normative appeals were sent to two large groups of taxpayers; a control group received no letter. The impact of the letters on voluntary compliance is measured by comparing the change in reported income and in taxes paid for treated versus control taxpayers (a difference-in-difference approach).

Conclusions:

The authors concluded that neither of the normative appeals in the letters made a significant overall impact on reported income or tax liability. Although it is impossible to generalize from these findings to make conclusions about any and all normative appeals, no matter how constituted, they find no evidence that inexpensive, mail-based appeals will significantly increase tax compliance. Communications of a different sort, delivered in a different way, or with greater frequency might still produce a compliance effect.

33. Bobek, D.D., Roberts, R.W., and Sweeney, J.T., The social norms of tax compliance: Evidence from Australia, Singapore, and the United States. *Journal of Business Ethics*, 2007. 74: p. 49-64.

Objective:

This paper tests the supposition that differences in compliance across countries are the result of differences in social norms.

Content:

Student subjects from Australia, Singapore, and the U.S. were given a tax compliance dilemma and asked to determine which action they would take in response to it. They also answered questions regarding each of the four types of social norms. The independent variables are social norms (descriptive, injunctive, subjective and personal) and personal norms, while the dependent variable is compliance intention.

Conclusions:

The authors found that social norms were consistently important in explaining compliance in all three countries, but the different types of social norms had differing influences compliance. Personal/subjective norms were found to be four times as important as any other variable.

34. Hite, P.A., Hasseldine, J., and Fatemi, D.J. Tax rate preferences: Understanding the effects of perceived and actual current tax assessments. *IRS Research Conference*. 2007. Washington DC: Publication 1500. (<http://www.irs.gov/pub/irs-soi/07resconfhite.pdf>)

Objective:

This study looks at preferences for progressivity and examines how status quo tax liabilities, both real and perceived, affect them. It looks at how access to actual and hypothetical but comparable rates affect preferences toward negative taxes. This experiment tests for a perceived status quo effect and for an anchoring effect.

Content:

Subjects are given status quo data and another set of subjects were given hypothetical set of tax assessments as an anchor while some subjects were asked to report the tax assessments that they believe current tax law requires. Pairwise comparisons are made in order to evaluate the impact of the status quo manipulation versus the control group, the anchor manipulation versus the control group, and the status quo group versus the anchor group. The independent variables include STATUS QUO tax assessments and ANCHOR (set of hypothetical tax liabilities). The dependent variables include progressivity as measured by the Suits Index (provides a parsimonious measure capturing the overall steepness of a set of tax assessments to the respective set of taxpayer incomes), the tax assessment preference for a low-income bracket, the average tax assigned to a household in the highest bracket, and the range of tax rates that each respondent assess, and the relative tax burden for each income group.

Conclusions:

The results indicate that informed subjects tend to assess lower overall taxes and steeper progressivity as measured by the Suits Index and relative tax burden. Even so, while status quo tax assessments lead to lower tax preferences compared to subjects without status quo information, those tax assessments are still significantly higher than the actual status quo information. The implication is that average Federal income tax liabilities under current law may be much lower than most taxpayers realize, since media coverage highlights marginal tax rates, not average tax rates.

35. Laury, S. and Wallace, S. Does confidentiality affect tax compliance? *IRS Research Conference*. 2005. Washington DC: Publication 1500.
(<http://www.irs.gov/pub/irs-soi/05laury.pdf>)

Objective:

This paper looks at whether perceived breaches in confidentiality or a weakening in the ability to keep taxpayer data confidential can affect taxpayer compliance. How does tax compliance respond to changes in the level of confidentiality of taxpayer information?

Content:

The experimental design follows that used by Cummings, Martinez-Vazquez, and McKee (2001) and mimics the basic tax reporting decision faced by most individuals. At the beginning, subjects are given an income level and must decide how much to report. If income is reported, it is taxed at the preannounced rate. Otherwise, it is not taxed unless the subject is audited and will incur an additional penalty. The authors use two treatments, full and partial confidentiality. All decisions are kept private for subjects in the full confidentiality treatment, while some subjects in the partial confidentiality treatment are randomly chosen to view the reported tax information of other subjects. The independent variable is confidentiality and the dependent variable is tax compliance.

Conclusions:

The authors found that the most frequently observed decisions are reporting no income or reporting income fully. The most observed outcome in the full confidentiality treatment was reporting no income at all with the second most observed outcome reporting full income. Overall, the authors reported that the level of reported income was higher in the partial confidentiality sessions when subjects know that there is a twenty five percent chance that others will view their reported income. The results suggest that compliance is affected by confidentiality. The authors believe that the experimental methodology used in this paper may be refined to shed some light on the issue of confidentiality as it relates to tax compliance, in an age when people are growing ever more concerned about their privacy.

36. Mckee, J.A.M., Audit certainty, audit productivity, and taxpayer compliance. *National Tax Journal*, 2006. 59(4): p. 801-816. (<http://ntj.tax.org/> – available only to members of the National Tax Association)

Objective:

This paper looks at the effect of advance information on audit probability and productivity on compliance responses.

Content:

Subjects are told whether or not they will be audited prior to filling out their returns as well as their productivity (fraction of unreported income discovered) of the audit. The independent variable is information letting participants know that they will or will not be audited and the dependent variable is compliance.

Conclusions:

This paper finds that taxpayer subjects who know they are to be audited increase their compliance when so informed, and they are also responsive to increases in audit probability. Announcement increases compliance of those told they will be audited, but reduces compliance of those knowing they will not be audited leading to a negative overall net effect as compliance falls.

37. Slemrod, J. and Christian, M.B.C., Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 2001. 79: p. 455-483.

Objective:

Examine the impact of increase probability of audit on compliance.

Content:

As part of the Minnesota 1995 field experiment, 1724 taxpayers were randomly selected to receive a letter informing them that their returns would be “closely examined.” A difference-in-difference econometric approach in multivariate regression, was used to estimate the partial effect of return characteristics.

The data comes from income tax return data from 1993 and 1994, from the state (Minnesota Dept of Revenue) and federal level, for the treatment group and the control group. Total returns numbered 22,368,

Change in reported tax income 93-94 was regressed on income / opportunity segments:

- Income strata (3 groups of AGI: <10K, 10-100K, >100K)
- Variables reflecting opportunity to evade: Minnesota estimated tax, Schedule C (trade or business income) or F (farm income)

Results were further broken down by Schedule A-F, ES, marginal tax rate, preparers, age and marital status for taxpayers with income < \$100K.

Conclusions:

The authors show that the treatment effect varies based on the level of income and opportunity to evade, as measured by their source of income (schedules filed). Among taxpayers with a high opportunity to evade (approx 3% of population), low income taxpayers had a treatment effect of 145.3% increase in reported liability, while middle income taxpayers had a 12.1% increase. For taxpayers with a low opportunity to evade, the differences for low- and middle-income groups are not statistically significant (between 2-4% increase). Return characteristics indicating a positive treatment effect were: Schedule A (itemized deductions), Schedule B (interest and dividends), Schedule F (farm), Schedule ES (estimated tax payments), while Schedule E was negative. Married filing joint and 65+ are associated with lower changes (greater compliance). High-income taxpayers (3% of pop, 30% of liability) have high variability and interestingly, reduced their liability in light of the notice, indicating a different set of tax beliefs. Increased probability of an audit, reflected in a threatening notice, increases positive compliance improvement for low and middle income taxpayers who have items providing the opportunity to evade. The overall improvement is small (2% of total expected tax), and only a small piece of the estimated tax gap.

Simulation

38. Bloomquist, K.M. Multi-agent-based simulation of the deterrent effects of taxpayer audits. *Annual Conference of the National Tax Association*. 2004. Minneapolis, MN. (<http://ftp.irs.gov/pub/irs-soi/04blonta.pdf>)

Objective:

This paper describes the development of an agent-based computational model of income tax evasion.

Content:

An agent-based computational model of income tax evasion, known as the Tax Compliance Simulator (TCS), is developed. The agent-based computational model makes the assumption that individuals behave as utility maximizers but also exhibit heterogeneous behavior, receive income from sources with different levels of “visibility,” and are subject to peer influences. While the model does not claim to accurately portray taxpayer behavior, TCS combines both theoretical models and known empirical relationships into a decision support tool that analysts can use to investigate the possible implications of alternative compliance strategies.

Conclusions:

This paper showed how multi-agent-based simulation (MABS) models provide a way to incorporate several important characteristics of taxpayers and tax systems typically excluded from analytical models of tax evasion, and it described the TCS as a prototype for MABS models. Such issues as taxpayer behavioral heterogeneity, variation in the visibility of income from multiple sources, and group influences on individuals can be handled easily using tools of agent-based simulation. This paper reviewed two previous MABS models of income tax evasion and presented a hypothetical case study showing the model’s capability to estimate deterrence resulting from different assumptions concerning taxpayers’ perceptions of audit probability, the proportion of income visible to tax authorities, and social network size. The results of the simulation suggest a significant portion of audit-based deterrence on compliance could come from indirect effects through group influences.

39. Bloomquist, K.M., A comparison of agent-based models of income tax evasion. *Social Science Computer Review*, 2006. 24: p. 411. (<http://ssc.sagepub.com/cgi/content/abstract/24/4/411>)

Objective:

This paper reviews three current, agent-based simulation models (ABMs) of income tax evasion. The models comprise the current state-of-the-art in ABM. Agent-based approaches to modeling tax evasion can offer insights into taxpayer behavior in situations where formal mathematical models cannot be solved.

Models are compared and contrasted across a number of dimensions and similarities and differences in modeling approaches are highlighted. Contributions of these models and how they lead to a better understanding of tax evasion are discussed.

Content:

Each of the models reviewed in this paper focuses on the audit rate as a principal driver of compliance behavior. The level and intensity of the enforcement activity is allowed to vary across simulations and results are compared. Taxpayers are assumed to possess varying levels of honesty which will effect compliance behavior. The proportion of “honest” taxpayers is allowed to vary across simulations in two of the models. In ABM, a large number of heterogeneous software agents interact and follow relatively simple behavioral rules regarding income tax evasion. The “data”, per se, are agentspecific and relate to, among other things, attitudes towards risk, the degree and timing of agent interactions, honesty and memory. In each of the agent-based models discussed in this paper, simulation experiments are conducted where several key parameters are allowed to vary and enforcement outcomes are compared. Each of the models are similar with respect to how agent interactions occur and how this affects tax evasion, the nature of taxpayer audits and how these audits detect (or fail to detect) income, the time interval over which the experiments are conducted and the size of the model (e.g., number of agents).

The models differ with respect to the decision to evade taxes, how each model is validated, the deterrent effects of enforcement activity and the complexity of the taxpayer agents. Only one of the models is designed to explicitly calculate the indirect effects of enforcement activity. Here, indirect effects include the increased compliance observed in subsequent time periods following an enforcement action (e.g., audit) or knowledge that an enforcement action has taken place.

Conclusions:

The author identifies four criteria for evaluating ABM models of tax evasion: (1) outcome validity; (2) process validity; (3) internal validity; and (4) reliability. Of these four, process validity is considered the most important for widespread acceptance of ABM for tax evasion.

40. Carley, K.M. and Maxwell, D.T. Understanding taxpayer behavior and assessing potential IRS interventions using multiagent dynamic-network simulation. *IRS Research Conference*. 2006. Washington DC: Publication 1 500. (<http://www.irs.gov/pub/irs-soi/06carley.pdf>)

Objective:

This paper looks at the effects of interactions with social circles, tax experts, and various types of potential IRS interventions influence taxpayer behavior. Specifically they focus on the combination of multiagent simulations and dynamic social network analysis.

Content:

Carley and Maxwell suggest that, “Dynamic-network simulation systems can be used to examine how networks, a set of nodes and relations, evolve and change over time and the repercussions of those changes for individual behavior.” In this subset of agent-based simulation, “the agents exist in a multidimensional or “metanetwork” space that changes as they interact.” To initialize and tune internal processes, real world data is used as an anchor and validation of these systems. In this paper, the authors present a high-level description of a specific dynamic-network model and discuss a virtual experiment for assessing the impact of an intervention.

Conclusions:

The authors conclude that dynamic-network simulations have significant potential for representing the decisions and behaviors of the taxpaying population and can be useful in evaluating different intervention strategies.

41. Cronkite, R.C., Social psychological simulations: An alternative to experiments? *Social Psychology Quarterly*, 1980. 43(2): p. 199-216.

Objective:

This paper critiques attempts by researchers to substitute simulations for actual experiments in social psychology. The rationale for relying on simulations in these situations is reviewed and the likelihood of making faulty inferences from both simulations and experimental data are investigated.

Content:

The author surveys a broad range of social psychological research where the dependent variables, generally, reflect attitudes or beliefs. Comparisons are then made with respect to actual experimental data versus data observed in simulations. Reasons for relying on simulation rather than actual experiments include a desire on the part of the researcher to avoid deception that may be necessary in an experimental setting; convenience and efficiency in data collection; and hypothesis testing. In the studies critiqued in this paper, data are collected from experiments and simulations (e.g., role playing). The principal concern is whether the data collected under both methods are comparable and provide reliable estimates from which inferences can be made. Simulation approaches to measuring social psychological phenomenon are “as-if” exercises rather than actual experiments and the question arises as to how accurate are the inferences derived from these simulations. Several problems in this approach are identified including: imperfect measurement under both approaches, misspecification that may arise in either (or both) models, and indistinguishable theories that lead to identification problems in interpreting results. Comparing variables across simulation and experiments require very strong assumptions which may not be met in practice.

Conclusions:

This paper cautions researchers who wish to use simulation as a substitute for actual, real world experiments because of the high probability that incorrect inferences will be made. The author suggests a useful set of guidelines for researchers making inferences from data collected under each regime. These guidelines include: (1) the use of multiple indicators to correct and control for identification; (2) evaluation of the hypothesized model using various goodness-of-fit measures; and (3) comparing parameter estimates from both simulated and experimental outcomes. The author suggests full-information, maximum likelihood (FIML) methods to carry out this research.

42. Wildt, A.R. and Winer, R.S., Modeling and estimation in changing market environments. *The Journal of Business*, 1983. 56(3): p. 365-388.

Objectives:

This paper discusses variable-parameter models as related to reported marketing applications and problems.

Content:

The authors discuss two major issues which are: “(1) Is the presently available variable parameter methodology capable of delivering these accurate estimates? (2) If it is, what is the most effective implementation strategy, that is, for any given situation what estimation procedure or method or what sequence of procedures or methods is best?”

Conclusions:

The authors foresee these two issues to be the next step in the research and that this will utilize a combination of analytical and simulation approaches in conjunction with a variety of known time-varying and constant coefficient processes and will be directed toward the evaluation of existing estimation methods and testing procedures.

Other Agencies

43. Metzenbaum, S., et al., Compliance and deterrence research project: Measuring compliance assistance outcomes. *State of Science and Practice White Paper*, 2007.
(<http://www.epa.gov/compliance/resources/reports/compliance/research/mcao-whitepaper.pdf>)

Objectives:

This paper looks at measuring outcomes from California area efforts of the Environmental Protection Agency, specifically responding to the following questions posed by EPA:

- “1. Is OECA’s current Government Performance and Results Act (GPRA) objective the most effective and compelling measure of compliance assistance performance? If not, what are potentially better measures and why?”
2. According to the relevant compliance literature, when do regulated entities seek compliance assistance, how and when does compliance assistance drive behavior or motivate change, and how can these outcomes be measured?”

The author also considers the questions in the context of OECA policy decisions and how “OECA use information on the effectiveness of compliance assistance strategically to select and target future compliance and enforcement priorities and activities.”

Content:

Assessment of EPA’s Current National GPRA Measures Based on the Following Criteria:

- A. Comprehensibility and Coherence
- B. Motivation Value
- C. Diagnostic Value
- D. Replication Value
- E. Feasibility of Measuring Accurately and Credibly
- F. Attribution

Conclusions:

This paper concludes that in the context of these key criteria, the EPA’s current GPRA CA measures are conceptually comprehensible and coherent, and align with the agency’s mission and strategic goals. Based on the strong findings in the cognitive psychology literature about the motivational power of specific, challenging targets, the authors suggest that the EPA’s current national CA targets could be more specific and challenging. The authors also recommend the use of measures in assessing what works and what doesn’t work. The review of the compliance assistance literature, especially reports on regional and STAG grant-funded compliance assistance and measurement projects, suggest beneficial CA work across the country, reporting noteworthy improvements in compliance levels, EMP, or pollution.

44. Shimshack, J.P., **Monitoring, enforcement, & environmental compliance: Understanding specific & general deterrence.** *State-of-Science White Paper*, 2007.
(<http://www.epa.gov/compliance/resources/reports/compliance/research/meec-whitepaper.pdf>)

Objective:

This report reviews the policy-relevant environmental compliance literature. It focuses specifically on measuring the general and specific deterrence effects of environmental monitoring and enforcement.

Content:

This paper reviews the literature on factors that influence environmental compliance, explores the deterrence impacts of compliance monitoring and enforcement in detail, and synthesizes research methodologies and presents frameworks for deterrence assessment.

Conclusions:

The major findings include: “(1) Regulation, monitoring, and enforcement have historically been, and remain, critical determinants of environmental behavior. (2) Environmental monitoring and enforcement activities generate substantial specific deterrence. (3) Environmental monitoring and enforcement activities generate substantial general deterrence. (4) Environmental monitoring and enforcement activities may generate significant emissions reductions, even for sector/contaminant combinations where compliance is typically high. (5) Enforcement and monitoring responses are heterogeneous across sources. (6) Quantitative databases and statistical methods exist for deterrence measurement. (7) Qualitative survey methods exist for deterrence measurement.” Key recommendations include: “(1) OECA should consider closely replicating statistical database analyses for measuring the specific and general deterrence effects of monitoring and enforcement. (2) OECA should consider closely replicating qualitative survey analyses for measuring the specific and general deterrence effects of monitoring and enforcement. (3) OECA should consider expanding existing qualitative survey analyses for measuring the specific and general deterrence effects of monitoring and enforcement. (4) For both quantitative and qualitative methods, studies based upon a universal metric are not advised. (5) Specific compliance rates and deterrence metrics should be statistically valid.”

IV. Bibliography: Full List of Papers Considered

Since many times the authors' own words provided the most succinct summary of their objectives and research, some of these short references are simply abstracts or excerpts. Papers with longer descriptions in the previous section are bolded.

- 1. Adelsheim, P.D. and Zanetti, J.L. The effect of targeted outreach on compliance. *IRS Research Conference*. 2006. Washington DC: Publication 1500.**

This paper studies the effect of targeted outreach on compliance and seeks to determine whether such a targeted approach is effective. None of their tests provided evidence of the effectiveness targeted outreach. Their diverse results indicate that perhaps the issue matters, perhaps the different mode of delivery matters, or perhaps industries simply have different characteristics that make them more or less susceptible to outreach.

- 2. Alexander, C. and Feinstein, J.S., A microeconomic analysis of income tax evasion. *Massachusetts Institute of Technology*, mimeo, 1987.**

This paper presents a microeconomic analysis of income tax evasion and proposes three new econometric methods for analyzing evasion data.

- 3. Allingham, M.G. and Sandmo, A., Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1972. 1: p. 323-338.**

This paper starts by considering a simple static model where the decision to evade is the only one with which the individual is concerned, and after a detailed study of this simple case they proceed with an analysis of the dynamic case where the individual has to make a sequence of tax declaration decisions. This paper is generally considered the quintessential origination point for tax evasion literature using a utility function approach.

- 4. Alm, J., A perspective on the experimental analysis of taxpayer reporting. *The Accounting Review*, 1991. 66(3): p. 557-593.**

Alm reviews the application of experimental methods to compliance research, including the appropriate procedures to be followed in experimental analysis, results of previous studies and limitations of experimental methods.

- 5. Alm, J., Bahl, R., and Murray, M.N., Audit section and income tax underreporting in the tax compliance game. *Journal of Development Economics*, 1993. 42: p. 1-33.**

This paper uses data from Jamaica to estimate the determinants of individual audit selection and individual tax evasion behavior in a way that treats the tax agency and the taxpayer as strategic players in the compliance game (as cannot be done with aggregate econometric analysis). They conclude that higher income and marginal tax rates increase underreporting.

- 6. Alm, J., Jackson, B.R., and Mckee, M., Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal*, 1992. 45(1): p. 107-114.**

This paper conducts laboratory experiments to estimate the individual responses to various policy changes. It concludes that higher tax rates decrease reporting compliance and higher income increases reporting compliance, yet higher penalties have a negligible impact.

- 7. Alm, J., Jackson, B.R., and Mckee, M. The effects of communication among taxpayers on compliance. *The IRS Research Bulletin*. 2004. Washington DC: Publication 1500.**

The purpose of this study is to examine the roles of information dissemination and taxpayer communication on voluntary compliance. In particular, they examine three types of communication about audit frequency and audit results using laboratory experiments in which the audit setting and communication opportunities are controlled.

- 8. Alm, J., Jones, M., and Mckee, M. Taxpayer services and tax compliance. *IRS Research Conference, 2007. Washington DC: Publication 1500.***

This paper utilizes laboratory experiments as a means of testing the range of possible information programs in terms of their effectiveness in enhancing tax compliance, and of comparing the compliance impacts of these “service” programs to the impacts of increased “enforcement” efforts.

- 9. Alm, J. and Mckee, M., Tax compliance as a coordination game. *Journal of Economic Behavior & Organization, 2004. 54: p. 297–312.***

This paper uses laboratory experiments to investigate compliance behavior when returns are selected for audit based upon the deviation of each individual’s tax report from the average report of all other taxpayers. The experimental results indicate that individuals find it difficult to coordinate on the zero-compliance equilibrium. Communication facilitates noncompliance in a DIF-like (deviation-based) audit strategy, yet this noncompliance can be reduced through the introduction of a secondary random audit approach.

- 10. Alm, J. and Torgler, B., Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology, 2006. 27: p. 224–246.***

This paper estimates the determinants of an individual’s intrinsic willingness to pay taxes, what is sometimes termed “tax morale,” using information from the World Values Survey for a wide range of countries over several years of data. The results indicate that individuals in the United States have the highest tax morale across all countries examined, followed by Austria and Switzerland. It is also found that there exists a strong negative correlation between the size of shadow economy and the degree of tax morale in those countries.

- 11. Anderson, L.R. and Stafford, S.L., Does crime pay? A classroom demonstration of monitoring and enforcement. *Southern Economic Journal, 2006. 72(4): p. 1016-1025.***

This article presents a classroom game in which students choose whether or not to comply with pollution regulations. By varying the level of monitoring and fines for noncompliance, the game shows students how the probability and severity of enforcement affects incentives for compliance.

- 12. Andreoni, J., Erard, B., and Feinstein, J., Tax compliance. *Journal of Economic Literature, 1998. 36(2): p. 818-860.***

This paper reviews the literature on tax compliance and describes the major theoretical and empirical findings in the recent tax compliance literature, focusing solely on personal income tax compliance.

- 13. Arcand, J.-L. and Graziosi, G.R., Tax compliance and rank dependent expected utility. *The Geneva Risk and Insurance Review, 2005. 30: p. 57-69.***

Formulating the classic Allingham and Sandmo [1972] tax compliance problem under Rank Dependent Expected Utility (RDEU) provides a simple explanation for the “excess” level of full compliance observed in empirical studies, which standard Expected Utility (EU) theory is unable to explain. They show that the threshold audit probability or penalty rate at which full compliance becomes optimal for the decision maker are significantly lower under RDEU axiomatics than in the EU case, and that the optimal level of underreporting is lower under RDEU.

- 14. Bauman, C.C., Luna, D., and Peracchio, L.A. Improving tax compliance of bilingual taxpayers with effective consumer communication. *IRS Research Conference, 2005. Washington DC: Publication 1500.***

The purpose of this study is to examine how bilingual taxpayers process technical tax information by comparing the accuracy of their answers on short exercises when only written rules are provided as compared to condensed written rules accompanied by figures and pictures. The study is grounded in consumer marketing research that finds pictorial cues can improve bilingual individuals’ understanding of a message. Comprehension of tax concepts by Spanish-speaking subjects was marginally improved by the addition of a picture / diagram, but only when paired with Spanish-speaking text.

- 15. Beckera, D., Kessler, D., and McClellan, M., Detecting Medicare abuse. *Journal of Health Economics*, 2005. 24: p. 189-210.**

This paper identifies which types of patients and hospitals have abusive Medicare billings that are responsive to law enforcement. They show that increased enforcement leads certain types of types of patients and hospitals to have lower billings, without adverse consequences for patients' health outcomes.

- 16. Bergman, M. and Nevarez, A., Do audits enhance compliance? An empirical assessment of vat enforcement. *National Tax Journal*, 2006. 59(4): p. 817-832.**

This paper analyzes for the first time the VAT tax return information and enforcement data to assess the impact of audits on subsequent compliance of taxpayers in Argentina and Chile. The evidence shows that audits have the undesired effect of furthering noncompliance behavior among cheaters, but a more positive effect among those prone to compliance.

- 17. Beron, K.J., H.V. Tauchen, and A.D. Witte, 1992, The effect of audits and socioeconomic variables on compliance, in *Why people pay taxes: Tax Compliance and enforcement*, ed. J. Slemrod, Ann Arbor: University of Michigan Press, 67—89.**

This essay examines the impact of audits and a variety of socio-economic factors on tax compliance. A simultaneous equations regression model is presented that includes separate specifications for taxpayers' reported income, taxpayers' reported tax liability, and the probability of an audit. This model explicitly accounts for the possibility that the audit rate is endogenously determined with the levels of reported income and tax liability. The model is applied to U.S. federal income tax return data for a random sample of taxpayers in 1969, aggregated to the three-digit zip code level. The data is augmented with a rich set of socio-economic variables. The results of the analysis indicate a modest, but significant, deterrent effect of audits for some groups of taxpayers, but no significant deterrent effect for other groups.

- 18. Bikhchandani, S., Hirshleifer, D., and Welch, I., Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 1998. 12(3): p. 151-170.**

This article argues that learning by observing the past decisions of others can help explain some otherwise puzzling phenomena about human behavior. They consider two scenarios. In the observable actions scenario, individuals can observe the actions but not the signals of their predecessors. They compare this to a benchmark observable signals scenario in which individuals can observe both the actions and signals of predecessors. Although other factors—such as network externalities and preference interactions—can lock in an inefficient behavior, the informational cascades theory differs in that it implies pervasive but fragile herd behavior. This occurs because cascades are triggered by a small amount of information. Under informational cascades, the system spontaneously fluctuates until it reaches a precarious resting point in which behavior is sensitive to small shocks.

- 19. Blackwell, C., 2008, A meta-analysis of tax compliance experiments, in *Tax Compliance and Evasion*, ed. J. Alm, J. Martinez-Vazquez, and B. Torgler, Routledge, forthcoming.**

Since 1978, economists, psychologists, sociologists and accountants have used laboratory experiments to investigate the determinants of tax compliance. In this paper the author attempts to synthesize this literature in a meta-analysis to draw conclusions regarding the determinants of tax compliance. Specifically, the author examines the impacts of traditional economic determinants of tax compliance: the tax rate, the penalty rate, and the probability of audit. In addition the author examines the effect of a public good "return" to taxes paid. The author finds strong evidence that increasing the penalty rate, the probability of audit, and the marginal per capita return to the public good lead to higher compliance, but finds no statistically significant effect of the tax rate on compliance.

- 20. Bloomquist, K.M. Trends as changes in variance: The case of tax noncompliance. *IRS Research Conference*. 2003. Washington DC: Publication 1500.**

The purpose of this paper is to show that the presumed rise in evasion activity may be due, at least in part, to a shift in the distribution of taxpayer income away from more visible to less visible sources. They present some preliminary

findings that show a correlation between income inequality and size of the shadow economy for a sample of developed and developing nations.

21. Bloomquist, K.M. Multi-agent-based simulation of the deterrent effects of taxpayer audits. *Annual Conference of the National Tax Association*. 2004. Minneapolis, MN.

This paper describes the development of an agent-based computational model of income tax evasion that assumes individuals behave as utility maximizers but who also exhibit heterogeneous behavior, receive income from sources with different levels of “visibility,” and whose behavior is subject to peer influences. This paper also described the development of the Tax Compliance Simulator (TCS), a prototype MABS model developed by IRS to estimate the deterrent effects of tax compliance alternatives.

22. Bloomquist, K.M., A comparison of agent-based models of income tax evasion. *Social Science Computer Review*, 2006. 24: p. 411.

This article compares three multiagent-based simulation (MABS) models of individual income tax evasion. It gives a brief description of each of the three models and highlights the similarities and differences among them. They also give a discussion of the contributions these models make to literature on MABS modeling and, more broadly, to the field of computational social science.

23. Bloomquist, K.M., Albert, M.F., and Edgerton, R.L. Evaluating preparation accuracy of tax practitioners: A bootstrap approach. *IRS Research Conference*, 2007. Washington DC: Publication 1500.

This paper presents summary measures of return preparation accuracy using population data from the IRS Office of Research’s Compliance Data Warehouse (CDW), and also develops a computational method for identifying tax practitioners who have a high percentage of returns with one or more errors.

24. Blumenthal, M., Christian, C., and Slemrod, J., Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, 2001. 54(1): p. 125-136.

This paper explores one part of a 1994 Minnesota Department of Revenue field experiment designed to study the effectiveness of alternative enforcement strategies. They find that neither of the normative appeals in the letters sent made a significant overall impact on reported income or tax liability.

25. Board, IRS Oversight, *The 2007 Taxpayer Assistance Blueprint Phase 2*. 2007: Department of the Treasury, Internal Revenue Service (<http://www.irs.gov/pub/irs-pdf/p4579.pdf>)

This report provided Congress a comprehensive overview of strategic planning for IRS services. Outcome measures, services initiatives, and existing knowledge of service usage are all extensively reported. Updates can be found in subsequent annual reports to Congress, of which 2008 is the first. See the Taxpayer Assistance Blueprint webpage for more information.

26. Bobek, D.D., Roberts, R.W., and Sweeney, J.T., The social norms of tax compliance: Evidence from Australia, Singapore, and the United States. *Journal of Business Ethics*, 2007. 74: p. 49-64.

This study examines the role of social norms on tax compliance in three different countries, Australia, Singapore, and the U.S. Factor analysis of the social norm questions identified three distinct social norm constructs. Two of these factors were significant in explaining tax compliance behavior. The first and most influential factor was taxpayers’ own personal moral beliefs, along with the beliefs of those close to them. The second significant factor represented societal views of proper behavior. They conclude that social norms help to explain tax compliance intentions and why tax compliance rates are higher than would be predicted by strictly economic models.

27. Bonabeau, E., Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 2002. 99(3): p. 7280-7287.

The author argues that agent-based modeling, with a rising number of applications in the last few years, provides a powerful approach to real-world business problems. After the basic principles of agent-based simulation are briefly introduced, its four areas of application are discussed by using real-world applications: flow simulation,

organizational simulation, market simulation, and diffusion simulation. For each category, one or several business applications are described and analyzed.

28. Braithwaite, V., Responsive regulation and taxation: Introduction. *Law & Policy*, 2007. 29(1): p. 3-10.

The implementation of responsive regulation in taxation means influencing the community's commitment to pay tax through respectful treatment, through attending to resistance and reforming faulty processes, through fairly directed and fully explained disapproval of non-compliant behavior, through preparedness to administer sanctions, and capacity to follow through to escalate regulatory intervention in the face of continuing noncompliance. Responsive regulation and regulatory formalism are pitted against each other in this issue on responsive regulation and taxation. Normative and explanatory arguments in favor of responsive regulation are explored by data collected in taxation contexts; and institutional obstacles are identified that limit effective implementation.

29. Braithwaite, V. and Braithwaite, J., An evolving compliance model for tax enforcement, in *Crimes of privilege*, N. Shover and J.P. Wright, Editors. 2001, Oxford University Press: New York p. 405-419.

This paper describes the application of the responsive regulation approach into the Australian Taxation Office (ATO) compliance model. This "pyramid based" policy approach begins with broad-based persuasion policy tactics and escalates to sanction. The application of this approach is considered in three compliance contexts: large corporations, high wealth individuals and in the cash economy. ATO corporate strategies include targeting known, aggressive tax advisors while building relationships within the financial planning community; Advance Pricing Arrangement (APA) and risk leveraging. ATO's high wealth approach consisted of increased surveillance and one high profile enforcement, while the cash economy approach built relationships in cash-based industries and educated industry members on tax responsibilities.

30. Braithwaite, V., Murphy, K., and Reinhart, M., Taxation threat, motivational postures, and responsive regulation. *Law & Policy*, 2007. 29(1): p. 137-158.

The article argues that the most effective regulatory outcome is achieved when the regulatory process can dampen the "taking control" and "feeling oppressed" sensibilities, and strengthen the "thinking morally" sensibility. Responsive regulation is an approach that encourages tax authorities to read motivational postures, understand the sensibilities that shape them, and tailor a regulatory intervention accordingly.

31. Caballe, J. and Panades, J., Cost uncertainty and taxpayer compliance. *International Tax and Public Finance*, 2005. 12: p. 239-263.

This paper provides a unified framework to analyze the effects of all sources of uncertainty (including the typical uncertainty associated with taxpayers' incomes and heterogeneous attitudes towards tax compliance) in a model of tax compliance with strategic interaction between auditors and taxpayers. They show that more variance in the distribution of the taxpayers' private cost of evading raises both tax compliance and the ex-ante welfare of taxpayers. The effects of the uncertainty about the audit cost faced by the tax authority are generally ambiguous.

32. Carley, K.M. and Maxwell, D.T. Understanding taxpayer behavior and assessing potential IRS interventions using multiagent dynamic-network simulation. *IRS Research Conference*, 2006. Washington DC: Publication 1500.

This paper describes how emerging research in the computational social sciences, specifically the combination of multiagent simulations and dynamic social network analysis could assist the Service in better understanding taxpayer behavior, as well as how taxpayer behavior changes in response to their interactions with others in their social and family circles, perceived tax experts, and various types of potential IRS interventions. Dynamic-network simulations can be used to effectively and systematically evaluate the relative efficacy of different intervention strategies. Dynamic-network simulations (Carley, 2003) are multiagent simulation systems in which the agents are enabled and constrained by their positions in dynamic metanetworks that include both social and knowledge networks.

33. Chamley, C.P., *Rational herds: Economic models of social learning*, 2004: Cambridge University Press.

This book looks at social learning which involves individuals learning from the behavior of others and may lead to spectacular outcomes such as herding, fads, frenzies, crashes, and booms. Providing a synthesis of the theoretical literature of the last ten years, the author demonstrates how these pathologies may occur in a society of rational individuals.

34. Chang, J.-J. and Lai, C.-C., *Collaborative tax evasion and social norms: Why deterrence does not work*. *Oxford Economic Papers*, 2004. 56(2): p. 344-368.

This paper makes a preliminary attempt to model a phenomenon of collaborative tax evasion between a seller and his customer and incorporates the social norm into such collusive tax-evading activities. It is found that, due to the existence of the social norm, more prevalent collaborative tax evasion at the status quo tends to intensify the extent of the tax evasion itself. It is also found that the snowballing effect stemming from social norm has a decisive influence not only in determining the conformity of consumers to the tax code, but also the deterrent effect of the authority's tax enforcement.

35. Christian, Charles. *The deterrent effects of audits on noncompliance by sole proprietors*. *IRS Research Conference*, 1992. Washington DC.

This paper summarizes prior empirical research on the effects of audits on noncompliance and presents the results of a test for audit rate effects from the TCMP survey of individual returns. The results are consistent with a deterrent effect: returns from zip codes with higher audit rates are more compliant than others even after controlling for differences in a variety of economic, sociodemographic, and tax administration factors thought to be associated with compliance.

36. Cialdini, R.B., 1989, *Social motivations to comply: Norms, values, and principles, in Taxpayer compliance, vol. 2, social science perspectives*, ed. J.A. Roth and J.T. Scholz, Philadelphia: University of Pennsylvania Press, 220—227.

This study examines the role of social cues that may affect an individual's willingness to comply with the demands of others. The author identifies procedures commonly used by compliance professionals—individuals in such fields as advertising, negotiation, recruitment, and sales, whose success depends on obtaining others' compliance. He then relates these procedures to several general principles of behavior. Among those most pertinent to tax compliance are commitment/consistency and reciprocity. The first of these relates to the desire of individuals to be consistent within their attitudes, beliefs, words, and deeds. The latter relates to the norm to respond in kind to positive behaviors that are directed towards oneself. The author draws out some potential applications of these principles to the problem of tax compliance.

37. Clotfelter, C.T., *Tax evasion and tax rates: An analysis of individual returns*. *The Review of Economics and Statistics*, 1983. 65(3): p. 363-373.

This paper investigates the relationship between marginal tax rates and tax evasion. The basic contribution of this paper is to provide econometric evidence suggesting that tax evasion is sensitive to marginal tax rates.

38. Coleman, S., 1996, *The Minnesota income tax compliance experiment: State tax results*, MPRA Paper No. 4827, University Library of Munich, Germany, April.

This report describes the Minnesota Income Tax Compliance Experiment conducted by the Minnesota Department of Revenue in 1995. The experiment tested alternate strategies to improve voluntary compliance with the state income tax. These strategies included: increased examination and auditing of tax returns with prior notice to taxpayers, enhanced services to taxpayers, information messages in letters sent to taxpayers, and a new M-1 tax form. About 47,000 taxpayers participated in the experiment. They were randomly selected for the alternative compliance strategies. The primary measures used to evaluate compliance strategies were the change in reported income and the change in state taxes paid from tax year 1993 to tax year 1994. Changes in taxpayer groups subject to a compliance strategy were compared to changes in similar groups of taxpayers who were unaffected by the experiment (control groups).

**39. Copeland, V. Research into the impact of audit on compliance. *IRS Research Conference, 2005.*
Washington DC: Publication 1500.**

This paper looks at the influence of audits on compliance. It finds that use of a control group shows promise as a method for establishing the extent of changes in audited taxpayer behaviors. Through examining the proportion of late payment penalties applied, audit appears to improve compliance. Examination of tax assessed did not suggest that tax assessed rose to a greater extent for auditees than for the control group.

40. Cox, D., National Tax Journal (Pre-1986); Sep 1984; 37, A.I.G., and 283, P., Raising revenue in the underground economy. *National Tax Journal, 1984. 37(3): p. 283-288.*

This paper looks at the relationship between marginal tax rates and compliance and finds no evidence of a relationship between them. The least compliant taxpayers are those at both ends of the income distribution within a given tax rate, the highest and lowest taxpayers. This early study concludes that further investigation should classify by more taxpayer characteristics than income.

41. Cronkite, R.C., Social psychological simulations: An alternative to experiments? *Social Psychology Quarterly, 1980. 43(2): p. 199-216.*

This paper critically reviews previous research on simulations and shows how faulty inference may have arisen. It then demonstrates that some of these logical flaws are due to underidentification of the hypothesized models in simulations and experiments, and can be alleviated through the use of multiple indicators. However, the logic of bridging the gap between simulations and experiments still remains problematic and the rationale for performing simulations as substitutes for experiments is discussed in light of the issues raised in this paper.

42. Das-Gupta, A., Ghosh, S., and Mookherjee, D., Tax administration reform and taxpayer compliance in India. *International Tax and Public Finance, 2004. 11(5): p. 575-600.*

This paper evaluates effects on tax compliance of simple reforms in personnel policy in the Indian income tax administration. The results imply significant compliance gains would accrue from expanded staff employment and changes in assignment procedures for staff and taxpayers.

43. Davis, J.S., Hecht, G., and Perkins, J.D., Social behaviors, enforcement, and tax compliance dynamics. *The Accounting Review, 2003. 78(1): p. 39-69.*

This paper analyzes the effect of social norms and enforcement on the dynamics of taxpayer compliance. The analysis suggests that the effect on compliance of changing enforcement levels depends on whether the taxpayer population is initially compliant or noncompliant.

44. Dhami, S. and Al-Nowaihi, A., Why do people pay taxes? Prospect theory versus expected utility theory. *Journal of Economic Behavior & Organization, 2007. 64: p. 171-192.*

Expected utility theory (EUT) is unable to explain observed high levels of compliance. This paper posits that cumulative prospect theory (CPT), which measures utility as gains and losses relative to some reference point, provides a much more satisfactory account of tax evasion, while providing independent confirmation of prospect theory. Several principals from prospect theory are (1) individuals are risk seeking with respect to gains but risk averse with regard to losses, (2) under uncertainty, individuals will overweight small probabilities and overweight large ones, and (3) that decisions are influenced by the context in which they are presented. The authors construct a CPT utility function where the probability of detection varies with the amount evaded and stigma costs of evasion are introduced. Using parameters from experimental data, the authors conclude that their model predicts an observable magnitude of tax evasion,

45. Dodge, J.M. and Soled, J.A., Debunking the basis myth under the income tax *Indiana Law Journal, 2001.*

Under the current income tax regime, taxpayers often lack the acumen and requisite records and information to fulfill their tax basis reporting obligations, the rules themselves are unwieldy and complicated, and the IRS is unable to fulfill its compliance mission insofar as basis and gain monitoring is concerned. This paper argues that most

plausible solution to this problem would be for the IRS to design an electronic filing system website that would assign an asset (and tax-payer) identification number to an asset adequately described by the broker.

46. Dubin, J.A. Criminal investigation enforcement activities and taxpayer noncompliance. *The IRS Research Bulletin*. 2004. Washington DC: Publication 1500.

The purpose of this paper is to answer several basic questions. First, does Criminal Investigation (CI) have a measurable effect on voluntary compliance, which includes both civil and criminal tax laws? Second, if CI does have a measurable effect on voluntary compliance, what mix of CI investigations has the greatest influence on voluntary compliance? Third, does media attention and publicity on CI investigations increase the compliance effect? Fourth, do convictions that result in prison sentences affect compliance differently from cases that result in probation? This paper finds that CI activities have a measurable effect on voluntary compliance, the mix of sentenced cases (tax and money laundering) is not a significant determinant of tax compliance, media attention shows some weak evidence of increasing compliance, and incarceration and probation (rather than fines) have the most influence on taxpayers.

47. Dubin, J.A., et al., The demand for tax return preparation services. *The Review of Economics and Statistics*, 1992. 74(1): p. 75-82.

This paper analyzes taxpayer choices of return preparation services, distinguishing between two types of nonpaid preparers, six types of paid third parties, and self-preparation. They find significant differences in the factors which explain the demand for paid third parties who are and are not able to represent clients before the IRS. Among these factors are increases in IRS audit rates and the frequency of IRS penalties.

48. Dubin, J.A., Graetz, M.J., and Wilde, L.L., Are we a nation of tax cheaters? New econometric evidence on tax compliance. *The American Economic Review*, 1987. 77(2): p. 240-245.

This paper reviews the theoretical as well as the empirical work on tax compliance and uses an instrumental variables approach to empirically examine the effects of audits and socioeconomic factors on compliance.

49. Dubin, J.A., Graetz, M.J., and Wilde, L.L., The effect of audit rates on the federal individual income tax, 1977-1986. *National Tax Journal*, 1990. 43(4): p. 395-409.

This paper investigate empirically the role of audits in federal revenue collection, including "spillover effects." (i.e. indirect effects). Their principal innovation is estimating taxes due directly. It is found that audit rate is endogenous and has a positive significant effect for reported tax.

50. Dubin, J.A. and Wilde, L.L., An empirical analysis of federal income tax auditing and compliance. *National Tax Journal*, 1988. 41(1): p. 61-74.

This paper provides empirical evidence on the relationship between compliance with the Federal Income Tax and auditing by the IRS. They find support for an economic approach to tax compliance that incorporates the IRS as a strategic actor. Moreover, after allowing for the simultaneous determination of audit rates and compliance levers, they find significant deterrent effects of auditing on noncompliance.

51. Elffers, H., Weigel, R.H., and Hessing, D.J., The consequences of different strategies for measuring tax evasion behavior. *Journal of Economic Psychology*, 1987. 8:p. 311-337.

This paper discusses tax evasion and measurement strategies. It examined various norms and attitudes and their correlation with tax evasion, both reported and actual evasion.

52. Engel, E.M.R.A., Galetovic, A., and Raddatz, C.E., A note on enforcement spending and vat revenues. *The Review of Economics and Statistics*, 2001. 83(2): p. 384-387.

This paper shows how tax revenues can be used instead of evasion data to estimate the impact of changes in enforcement spending. Applying the method to Chilean data, we find that \$1 (USD) of additional enforcement spending increases VAT revenues by \$31. Moreover, current levels of spending could increase by 40% and still be within sample values. Hence, a 10% increase in spending could reduce evasion from its current rate of 23% to 20%.

53. Erard, B. and Ho, C.-C., Searching for ghosts: Who are the nonfilers and how much tax do they owe? *Journal of Public Economics*, 2001. 81: p. 25-50.

This paper looks at individuals who fail to comply with their income tax filing requirements. It extends the standard model of tax evasion to account for their existence and examines empirical significance and policy relevance of the extension using a unique data set containing detailed tax and audit information for both filers and nonfilers of U.S. federal income tax returns.

54. Feinstein, J.S. Statistical analysis of compliance using the NRP data: Detection controlled models. *IRS Research Conference*. 2004. Washington DC.

This presentation discusses the development of statistical methodology – based on detection controlled modeling -- for using the National Research Program (NRP) database to estimate individual income tax reporting noncompliance and the associated “tax gap.”

55. Feinstein, J.S., An econometric analysis of income tax evasion and its detection. *Journal of Economics*, 1991. 22(1): p. 14-35.

This article presents an econometric analysis of income tax evasion and its detection based on individual-level data. The empirical analysis explores the effects of income, the marginal tax rate, and various socioeconomic characteristics on filer evasion behavior, and it assess the variability in detection rates among IRS examiners.

56. Feld, L.P. and Frey, B.S., Trust breeds trust: How taxpayers are treated. *Economics of Governance*, 2002. 3: p. 87-99.

In this paper, empirical evidence based on a survey of tax authorities of the 26 Swiss states (cantons) is presented, indicating that the differences in the treatment of taxpayers by tax authorities can be explained by differences in political participation rights as well.

57. Fortin, B., Lacroix, G., and Villeval, M.-C., Tax evasion and social interactions. *Journal of Public Economics*, 2007. 91: p. 2089-2112.

The paper extends the standard tax evasion model by allowing for social interactions. In an experiment, participants must decide how much income to report given individual and group tax rates and audit probabilities, and given a feedback on the other members' reporting behavior. A unique social equilibrium exists when the model satisfies coherency conditions. The results provide evidence of fairness effects but reject social conformity.

58. Frey, B.S. and L.P. Feld, 2002, “Deterrence and morale in taxation: An empirical analysis”, CESIFO Working Paper No. 760, Category 1: Public Finance, August.

The standard model of tax evasion based on subjective expected utility maximization does not perform particularly well in econometric analyses: it predicts too little evasion and produces unsatisfactory econometric parameter estimates. The model is extended by looking at how the tax authority deals with the taxpayers. Based on econometric estimates, it is shown that taxpayers' tax morale is raised when the tax officials treat them with respect. In contrast, when tax officials solely rely on deterrence, taxpayers tend to respond by actively trying to avoid taxation.

59. Fugazza, M. and Jacques, J.-F., Labor market institutions, taxation and the underground economy. *Journal of Public Economics*, 2003. 88: p. 395-418.

The paper aims at qualifying the links between labor market institutions, taxation, tax monitoring, and underground economic activity. It supports the view that policies aiming at increasing individuals benefits of participating in the regular sector are more desirable than a deterrence policy.

- 60. Gerxhani, K. and Schram, A., Tax evasion and income source: A comparative experimental study. *Journal of Economic Psychology*, 2006. 27: p. 402-422.**

This paper compares tax evasive behavior in a country in transition from communism to that in a developed economy by running an experiment across distinct social groups in Albania and the Netherlands. Aside from the tax compliance decision, subjects choose a source of income, where one type enables subsequent tax evasion. It is shown that they take the possibility of evasion into account when deciding on the income source. In addition, Dutch subjects evade more than the Albanians. This paper concludes that the different levels of tax evasion outside of the laboratory in the two types of countries can be attributed to distinct formal tax institutions.

- 61. Giles, D.E.A. and Caragata, P.J., The learning path of the hidden economy: The tax burden and tax evasion in New Zealand. *Applied Economics*, 2001. 33(14): p. 1857-1868.**

This paper considers the “learning curve” relationship between the aggregate tax rate and the relative size of the hidden economy in New Zealand. This study finds that about half of the hidden activity in New Zealand is a learned response to changing opportunities and constraints in fiscal policy, but this amount varies over the business cycle.

- 62. Gordon, J.P.P., Individual morality and reputation costs as deterrents to tax evasion, *European Economic Review*, 1989. 33(4): 797—805.**

Empirical evidence suggests that some taxpayers never evade, evasion increases with the tax rate, and evasion decisions are interdependent. These findings are not easily explained by the existing theory, which models evasion as an independently-made amoral gamble. This paper uses non-pecuniary costs of evasion to reconcile theory with evidence. Once individuals are assumed to differ by an honesty characteristic, taxpayers dichotomize into groups, the more honest of which responds to a tax change as the evidence suggests. A further extension is to introduce an endogenous reputation cost. This suggests a positive relationship between the number of evaders and the tax rate.

- 63. Gray, L.N. and Tallman, I., Theories of choice: Contingent reward and punishment applications. *Social Psychology Quarterly*, 1987. 50(1): p. 16-23.**

This paper suggests that conventional ideas of “rational” choice based on game theoretic or economic principles need to be closely examined and new principles of behavioral choice developed.

- 64. Hasseldine, J., P. Hite, S. James, and M. Toumi, Persuasive communications: tax compliance enforcement strategies for sole proprietors, *Contemporary Accounting Research*, 2007, 24(1): 171—194.**

The paper reports on a controlled field experiment to analyze the effect on actual tax reporting behaviors of a large sample of sole proprietors in the United Kingdom who had previously reported a turnover (sales) level just low enough to qualify for a simplified method of reporting that lowered their compliance burden. Changes in turnover and net profit over a 2-year period are examined after the taxpayer receives one of five alternative treatments, including a simple offer of assistance (a small carrot), a citizenship approach, a threat of audit (a moderate stick), a threat of an audit and possible penalties (a larger stick), and, finally, notification that the return to be filed by the taxpayer has been preselected for audit (a very large stick). The results indicate a significant overall treatment effect for all five treatments.

- 65. Hasseldine, J., et al. Carrots, sticks, sole proprietors, and tax accountants. *IRS Research Conference*. 2005. Washington DC: Publication 1500.**

The paper uses a controlled field experiment to analyze the effect of communication on actual tax reporting behaviors of sole proprietors. It reveals that, even though the taxpayers in the five treatment groups received only one letter from the Internal Revenue during the 2001 tax return filing season,⁶ all five approaches were effective at significantly increasing the proportion of sole proprietors reporting above the £15,000 turnover threshold for simplified tax reporting. Furthermore, the sanction-based approaches appear to have induced taxpayers to report increased turnovers and net profits (relative to the control group).

- 66. Hasseldine, J. and Li, Z., More tax evasion research required in new millennium. *Crime, Law and Social Change*, 1999. 31(2): p. 91-104.**

This article highlights the pervasive phenomenon of tax evasion and discusses the different theoretical models and research approaches that have been used to study the problem in the last two decades.

- 67. Heijden, H.E.P.V.D. and Hezemans, M., Explaining regulatory non-compliance: A survey study of rule transgression for two Dutch instrumental laws, applying the randomized response method. *Journal of Quantitative Criminology*, 2003. 19(4): p. 409-439.**

Within a rational choice framework, secondary data analysis of a survey study on compliance with two Dutch regulatory laws is carried out. In this paper the authors make use of the Table-of-Eleven (T11), which is a Fishbein-Azjen-style explanatory model that explains behavior in attitudinal cost-benefit terms. It has been developed especially for use in the regulatory compliance context. Using adapted logistic regression analysis, they show that self-reported compliance, measured by means of a randomized response procedure, can be explained in terms of benefits of non-compliance, social norms and deterrence, while knowledge and general norm-conformity have no role to play. The impact of various contributing factors turns out to be rather different in size for the two laws.

- 68. Hessing, D.J., Effers, H., and Weigel, R.H., Exploring the limits of self-reports and reasoned action: An investigation of the psychology of tax evasion behavior. *Journal of Personality and Social Psychology*, 54(3): p.405-413.**

This study examined attitudes and norms in relation to tax evasion behavior. The outcomes of this study "suggest that the explanatory power of the theory of reasoned action may not extend to the domain of socially proscribed behaviors where self-presentation concerns are likely to prompt both misrepresentations of past behavior and reports of attitudes and perceived norms consistent with those misrepresentations."

- 69. Hite, P.A., Hasseldine, J., and Fatemi, D.J. Tax rate preferences: Understanding the effects of perceived and actual current tax assessments. *IRS Research Conference*. 2007. Washington DC: Publication 1500.**

This study examines how status quo tax liabilities, both real and perceived, affect preferences for progressivity. It looks at how access to actual and hypothetical but comparable rates affect preferences toward negative taxes.

- 70. Holland, K. and Rasey, H. Lab research on customer preferences and the relationship between service and compliance. *IRS Research Conference*. 2007. Washington DC: Publication 1500.**

This study looked at the impact of providing groups of participants with different types of service while they completed hypothetical tax situations.

- 71. Hunter, W.J. and Nelson, M.A., An IRS production function. *National Tax Journal*, 1996. 49(1): p. 105.**

In this paper, the effectiveness of the IRS is assessed by estimating a production function for the years 1955-90. They provide estimates of the contribution of various factor inputs to output, including labor and capital at both the field and national office level. Output is measured by the value of additional individual income taxes and penalties assessed. They determine that several technological innovations introduced by the IRS in recent years have produced positive effects on production. However, they also find that the IRS can produce additional tax revenues at no additional cost if it were to alter its mix of inputs.

- 72. Judson, D.H., Gray, L.N., and Duran-Aydintug, C., Predicting unique behavioral choices: Direct numeric estimation of rewards and costs and the satisfaction-balance decision model. *Social Psychology Quarterly*, 1994. 57(2): p. 140-149.**

This paper uses a satisfaction-balance model, attempting to predict a signal, unique behavioral choice. The results strongly support the applicability of the satisfaction-balance model in situations of nonrepetitive behavioral choice, with the use of direct numeric estimation.

73. Kahan, D.M., Social influence, social meaning, and deterrence. *Virginia Law Review*, 1997. 83: p. 349-395.

This paper suggests that individuals decide to commit crimes based in part on their perception of the values, beliefs, and behavior of other individuals. The social influence conception reveals the utility of certain law-enforcement strategies - including order-maintenance policies and gang-loitering laws - that have little effect on the price of crime but that nonetheless discourage it through suppressing signals of criminality.

74. Kahneman, D., and A. Tversky, Prospect theory: An analysis of decisions under risk, *Econometrica*, 1979. 47: 313—327.

This very widely cited paper presents a critique of expected utility theory as a descriptive model of decision-making under risk and lays out the foundations for an alternative paradigm known as prospect theory. According to prospect theory, individuals tend to underweigh outcomes that are merely probable in comparison with outcomes that are obtained with certainty, with the result that individuals tend to be risk-averse when confronted with choices involving sure gains and risk-seeking when confronted with choices involving sure losses. Further, it is argued that individuals attach value to gains and losses rather than to final assets (as they do according to expected utility theory). Finally, individuals are assumed to apply decision weights rather than objective probabilities when evaluating risky outcomes, with the effect that they tend to place more weight on low probability events in their decision-making than would be accorded by a mere consideration of objective probabilities.

75. Kamdar, N., Information reporting and tax compliance: An investigation using individual TCMP data. *Atlantic Economic Journal*, 1995. 23(4): p. 278-292.

This paper uses micro data from individual tax returns audited during the 1971 cycle of the IRS TCMP to reexamine the determinants of tax compliance. The results show third-party information reporting to be an effective deterrent to noncompliance, but cast doubt on the presumption that lower marginal tax rates led to greater compliance.

76. Klepper, S., M. Mazur, and D. Nagin, Expert intermediaries and legal compliance: The case of tax preparers, *Journal of Law and Economics*, 1991, 34(1): 205—229.

In this study, the authors propose that tax practitioners play dual roles with respect to tax compliance. Specifically, it is argued that they serve as enforcers of the tax law in unambiguous situations and as exploiters of tax law in ambiguous situations. To test their theory, the authors perform an econometric analysis of the adjustments that were made to a large sample of individual income tax returns that were randomly subjected to tax audits; these adjustments were aggregated at the line item level by mode of tax preparation. The results are consistent with their theory that tax practitioners will act to discourage non-compliance on legally unambiguous income sources but encourage compliance on ambiguous sources.

77. Kinsey, K.A., 1992, Deterrence and alienation effects of IRS enforcement, in *Why people pay taxes: Tax compliance and enforcement*, ed. J. Slemrod, Ann Arbor: University of Michigan Press, 259—285.

This study employs survey data to explore two potentially opposing effects of enforcement techniques on tax compliance: (1) they force people to comply with the law; and (2) they negatively affect willingness to comply voluntarily. The author's regression results indicate that respondents who report having paid additional taxes as a result of an enforcement contact perceive a greater likelihood of detection for noncompliance and are more likely to report an intention to comply on tax future returns. On the other hand, those who report having avoided detection of noncompliance during enforcement contacts perceive a lower certainty of detection and are more likely to report an intention to be noncompliant on future tax returns. Further, results based on vicarious contacts from other taxpayers indicate little support for a general deterrence effect. In particular, hearing vicarious reports of bad behavior by the tax agency is found to lower both the perceived fairness of the tax laws and the personal willingness to comply.

78. Kirchler, E. and Maciejovsky, B., Tax compliance within the context of gain and loss situations, expected and current asset position, and profession. *Journal of Economic Psychology*, 2001. 22: p. 173-194.

This study investigates self-reported tax evasion within the context of taxpayers' previous expected tax payments at filing time (payment or refund), their asset position (expected versus current asset position), and their tax category (self-employed versus business entrepreneurs). The results indicate that for the self-employed in the study an

unexpected surprise payment leads to low tax compliance, whereas an unexpected surprise refund leads to high tax compliance. Thus, their self-reported tax compliance can be best described by the current asset position.

79. Kirchler, E., Niemiowski, A., and Wearing, A., Shared subjective views, intent to cooperate and tax compliance: Similarities between Australian taxpayers and tax officers. *Journal of Economic Psychology*, 2006. 27: p. 502-517.

This study investigates how taxation is represented in the minds of Australian taxpayers and tax officers and whether shared beliefs and evaluations are related to taxpayer compliance. The ATO reported non-compliance was found to be related to respondents' willingness to cooperate (i.e., filing timely and correctly) which, in turn, was related to respondents' beliefs and evaluation of tax issues and their shared views on these issues. Willingness to cooperate appears to function as a mediator between tax compliance and taxpayer's beliefs and evaluations as well as shared tax officials' and taxpayers' views.

80. Kornhauser, M.E., Normative and cognitive aspects of tax compliance: Literature review and recommendations for the IRS regarding individual taxpayers. *Annual Report to Congress*, 2007. 2.

This study was commissioned by the Taxpayer Advocate to better understand the role of "tax morale" in compliance through behavioral research. The paper discusses norms impacting tax compliance, cognitive processes that influence individual taxpayer behavior and the frequently ambiguous role of demographic variables, gender, age, education, marital status, religion and income.

81. Kotler, P. and Schultz, R.L., Marketing simulations: Review and prospects. *The Journal of Business*, 1970. 43(3): p. 237-295.

This paper reviews the uses of simulations in marketing, distinguishing and describing four types of simulations. These include simulation as behavioral modeling, simulation as a way of introducing and handling uncertainty, simulation as a computational technique for measuring parametric sensitivity, and simulation as a heuristic technique for finding an approximately optimal solution.

82. Kuran, T. and Sunstein, C.R., Availability cascades and risk regulation. *Stanford Law Review*, 1999. 51(4): p. 683-768.

This paper analyzes availability cascades and suggests reforms to alleviate their potential hazards. The authors' proposals include new governmental structures designed to give civil servants better insulation against mass demands for regulatory change and an easily accessible scientific database to reduce people's dependence on popular (mis)perceptions.

83. Kydland, F.E. and Prescott, E.C., The computational experiment: An econometric tool. *The Journal of Economic Perspectives*, 1996. 10(1): p. 69-85.

Kydland and Prescott describe the application of "computational experiments," a technique frequently used in the physical sciences, to generate macroeconomic estimates (in lieu controlled experiments with national economies). In such an experiment, the analyst constructs a model economy and records the time paths of subjects' economic behavior. The authors specify the steps in designing a computational experiment to address some well posed quantitative question. They emphasize that the computational experiment is an econometric tool used in the task of deriving the quantitative implications of theory.

84. Laury, S. and Wallace, S. Does confidentiality affect tax compliance? *IRS Research Conference*. 2005. Washington DC: Publication 1500.

This paper looks at whether perceived breaches in confidentiality or a weakening in the ability to keep taxpayer data confidential can affect taxpayer compliance. The experiments and the results are preliminary but are suggestive of an impact of confidentiality on compliance.

85. Lowenstein, R., Exuberance is rational, in *New York Times*. 2001.

This newspaper article discusses the work of Richard Thaler and his contributions to behavioral economics.

86. Luan, F., Graco, W., and Norrie, M. Instance-based classifiers for tax agent modeling. *IRS Research Conference*. 2006. Washington DC: Publication 1500.

The Australian Taxation Office (ATO) is responsible for identifying high-risk tax agents who are engaging in unacceptable practice. The methods described in this paper were aimed at identifying high-risk agents. In this paper, the authors report some results from modeling tax agent behavior using a distance-from-the-centroid (DFC) method with assistance from a genetic algorithm (GA).

87. Martinez-Vazquez, J. and Rider, M., Multiple modes of tax evasion: Theory and evidence. *National Tax Journal*, 2005. 58(1): p. 51-76.

This paper examines the theoretical and empirical implications of accounting for multiple modes of tax evasion. They find that increasing the probability of detection in a given mode has an ambiguous effect on compliance in the target mode as well as the untargeted mode. Furthermore, they find that increased enforcement effort has a positive effect on compliance in the targeted mode, a negative effect in the untargeted mode, and a positive overall effect on tax compliance.

88. Masken, K. Longitudinal study of EITC claimants. *IRS Research Conference*. 2006. Washington DC: Publication 1500.

While it is known that there is significant turnover in EITC claimants from one year to the next, the reasons for this are not well understood. In order to better understand why taxpayers move in and out of the EITC population, the Office of Research is conducting a longitudinal study of tax returns filed for Tax Years 1996 through 2004. In addition to tracking taxpayers who claimed EITC in at least 1 of the last 9 years, the study will also track the children claimed in the last 4 years (due to data problems, it is not possible at this time to track the children for all 9 years). This paper presents some of the data issues encountered and a preliminary analysis of taxpayer patterns during the study period. It also looks at the pattern of children claimed as qualifying children for the shorter time period.

89. McKee, J.A.M., Audit certainty, audit productivity, and taxpayer compliance. *National Tax Journal*, 2006. 59(4): p. 801-816.

This paper uses experimental methods to examine individual compliance responses to advance information on audit probability and productivity. They find that announcement increases compliance of those told they will be audited, but reduces compliance of those knowing they will not be audited with the net effect being that overall compliance falls.

90. Mete, M., Bureaucratic behavior in strategic environments: Politicians, taxpayers, and the IRS. *The Journal of Politics*, 2002. 64(2): p. 384-407.

This article extends the empirical literature on bureaucratic behavior by modeling the strategic interactions between an enforcement agency, its clients, and its political principals. Specifically, the study focuses on the joint analysis of IRS enforcement decisions and taxpayer compliance behavior in order to fully explore the extent of political controls over the agency. The results suggest that there is a very strong reciprocal relationship between IRS audit levels and taxpayer compliance levels across states and over time and that taxpayer compliance behavior is also influenced by partisan changes in national politics.

91. Metzenbaum, S., et al., Compliance and deterrence research project: Measuring compliance assistance outcomes. *State of Science and Practice White Paper*, 2007.

This paper seeks to inform Environmental Protection Agency's efforts to measuring outcomes from CA efforts.

92. Morris, T. and Lonsdale, M. Translating the compliance model into practical reality. *The IRS Research Bulletin*, 2004. Washington DC: Publication 1500.

This paper puts forth a model which will allow for improvement of long-term voluntary compliance and create an environment that promotes compliance. The compliance model pyramid has two different components, the customer's attitudes and behaviors and the response to encourage compliance.

93. Myles, G.D and R.A. Naylor, A model of tax evasion with group conformity and social customs, *European Journal of Political Economy*, 1996, 12(1): 49—66.

A model of tax evasion is investigated in which a “social custom utility” is derived when taxes are paid honestly and there is a conformity payoff from adhering to the standard pattern of social behavior. A taxpayer faces a choice between whether to evade or not and, if evasion is chosen, a straightforward choice with risk over the level of evasion. It is shown that both equilibria - with no evasion and with all taxpayers choosing to evade - may exist. The no-evasion equilibrium may be destroyed by a small change in the tax rate leading to epidemics of evasion after seemingly insignificant tax changes.

94. Orcutt, G.H., Simulation of economic systems. *The American Economic Review*, 1960. 50(5): p. 893-907.

This paper describes various uses of simulation, models and model building with focus on various means by which complex, large-scale systems, such as economies, can be simulated. It also describes a concrete example of simulation of a demographic model of the United States.

95. Park, C.-G. and Hyunb, J.K., Examining the determinants of tax compliance by experimental data: A case of Korea. *Journal of Policy Modeling*, 2003. 25: p. 673-684.

This paper examines the determinants of tax compliance using experimental data for Korea. The empirical results can be summarized as follows. First, taxpayers have the same degree of compliance regardless of their income levels, which can be interpreted as an evidence of the utility function with constant relative risk aversion. Second, both the tax audit and the penalty rate are important deterrents from tax evasion, although the penalty rate is more effective. Third, taxpayers have a strong tendency for a free-ride, which means less compliance for financing their government. Fourth, tax education is one of the effective tools to induce taxpayers to comply more.

96. Plumley, A.H., The determinants of individual income tax compliance: Estimating the impacts of tax policy, enforcement, and IRS responsiveness, IRS. 1996, *Publication 1916* (Rev. 11-96): Washington DC.

This paper presents an econometric analysis of the impact of a wide variety of potential determinants of voluntary compliance with individual income tax filing and reporting obligations. Based on perhaps the richest dataset yet compiled (by state and year, from 1982 through 1991), including data on taxpayer behavior, IRS actions, and other factors, the analysis finds significant compliance effects attributable to many tax policy and tax administration parameters, including: audits; the matching of third-party information documents; the issuance of targeted nonfiler notices; criminal tax convictions; marginal tax rates; the burden associated with completing the myriad tax forms and schedules; and the preparation of returns by the IRS Taxpayer Service function.

97. Plumley, A.H. A framework for optimal tax administration. in IRS Research Conference. 2007. Washington DC: *Publication 1500*.

Congress asked the Internal Revenue Service (IRS) several important questions: What is the right-sized IRS budget to increase voluntary compliance by 1 percent? By 5 percent? How would the rate of voluntary tax compliance be affected by sustained growth in the IRS budget of \$500 million over each of 5 years? What would be the optimal use of these resources in providing improved services, stronger enforcement, and enhanced information technology? Unfortunately, the answer to each of these questions is that we do not currently know. Therefore, this paper attempts to provide a framework for how a tax administration agency might address these, and similar, questions. This approach uses a benefit-cost framework that is very familiar to economists and many policymakers. In the context of this framework, the author lays out a series of principles and steps that the IRS could decide to take to improve tax administration. These steps may not all be feasible, but the issues they are aimed at will need to be addressed if the IRS is to make significant strides toward improved tax administration.

98. Posner, E.A., Law and social norms: The case of tax compliance. *Virginia Law Review*, November, 2000.

This paper discusses the topic of the law of tax compliance, drawing an important distinction between the signaling model and other models of social norms. The test would focus on mainstream communities; in communities where, for historical or other reasons, tax compliance never emerged as a signal, there would be no relationship between

discount rate and tax compliance. Tax compliance is likely to become a signal in mainstream communities only if it is plausible to associate tax compliance with being a good type

99. Raskolnikov, A., Crime and punishment in taxation: Deceit, deterrence, and the self-adjusting penalty. *Columbia Law Review*, 2006. 106: p. 569-642.

Avoidance and evasion continue to frustrate the government's efforts to collect much-needed tax revenues. This Article articulates one of the reasons for this lack of success and proposes a new type of penalty that would strengthen tax enforcement while improving efficiency.

100. Reinganum, J.F. and Wilde, L.L., Income tax compliance in a principal-agent framework. *Journal of Public Economics*, 1985. 26: p. 1-18.

This paper compares an alternative audit policy to the standard random audit policy. They focus on an audit cutoff policy, in which an agent triggers an audit if reported income is too low, and is not audited if reporting income is sufficiently high. They find that random audit rules are weakly dominated by audit cutoff rules. Given lump-sum taxes and fines, audit cutoff rules are the least-cost policies which induce truthful reporting of income

101. Reinganum, J.F. and Wilde, L.L., Equilibrium enforcement and compliance in the presence of tax practitioners. *Journal of Law, Economics, and Organization*, 1991. 7(1): p. 163-181.

This paper focuses on tax practitioners as providers of services and their potential for lowering the costs to taxpayers of filing returns and facing the risk of detection.

102. Ritsema, C.M., Thomas, D.W., and Ferrier, G.D. Economic and behavioral determinants of tax compliance: Evidence from the 1997 Arkansas tax penalty amnesty program. *IRS Research Conference*, 2003. Washington DC: Publication 1500.

This paper asked taxpayers, who came forth during the 1997 Arkansas Tax Penalty Amnesty Program, why they didn't pay taxes when they should have and what motivated them to come forward during the amnesty. It confirmed previous research, finding that filing status, income, and the opportunity to evade are positively related to the tax owed. They also found that the lack of available funds is a primary factor related to their failure to report and pay taxes when due, but only for taxpayers who owe a higher percent of their income in tax. They also discovered that a sense of morality is not reported as a motivating factor for their participation in the amnesty. Importantly, it showed that taxpayers are not all alike in their motivation for failure to report or pay taxes and therefore enforcement techniques are likely to be received by taxpayers differently.

103. Roine, J., The political economics of not paying taxes. *Public Choice*, 2006. 126: p. 107-134.

This paper considers redistributive as well as political consequences of tax avoidance. When the avoidance possibilities are limited, the classical conflict between rich and poor is sustained. If the tax avoidance technology is more effective, however, the equilibrium outcome can change to a situation where a coalition of poor and the very richest favor a higher tax rate. Comparing the model predictions with data on income inequality and evidence of avoidance activity, it comes surprisingly close to actual observations.

104. Sandmo, A., The theory of tax evasion: A retrospective view. *National Tax Journal*, 2005. 58(4): p. 643-663.

The paper gives an overview of some main themes in the theory of tax evasion, starting from Allingham and Sandmo (1972). It reviews the comparative statistics of the original model of individual behavior where the tax evasion decision is analogous to portfolio choice, and its extensions to incorporate socially conscious behavior, participation in the black labor market and tax evasion by firms. It also discusses the analysis of tax incidence and the problems involved in moving from individual to aggregate analysis. Finally, it reviews the issues that arise in formulating models of optimal taxation in the presence of tax evasion.

105.Scholz, J.T. and Lubell, M., Adaptive political attitudes: Duty, trust, and fear as monitors of tax policy. *American Journal of Political Science*, 1998. 42(3): p. 903-920.

This paper estimates the impact of individual tax changes created by the Tax Reform Act of 1986 on attitudes of upper-income taxpayers, using tax returns and two waves of survey data. Trust, duty, and fear increase significantly when taxes decrease, and decrease when taxes increase.

106.Scholz, J.T. and Lubell, M., Trust and taxpaying: Testing the heuristic approach to collective action. *American Journal of Political Science*, 1998. 42(2): p. 389-417.

This paper tests whether trust in government and in other citizens increases compliance over and above the levels expected from an internalized sense of duty to obey laws and the fear of getting caught by enforcement agencies like the IRS. They find that both dimensions of trust significantly increase the likelihood of tax compliance, even after controlling for duty, fear, selection bias, and potential endogeneity effects.

107.Scotchmer, S. and J. Siemrod, Randomness in tax enforcement, *Journal of Public Economics*, 1989, 38(1): 17—32.

One of the consequences of complexity is that it makes it difficult for taxpayers to understand their tax obligations. In this paper, the authors employ an expected utility framework to examine the effect of randomness with respect to how much taxable income an auditor would assess if a tax return were examined. They find that when there is tax evasion, increased randomness generally leads to higher reported income and more revenue. When reducing randomness is costly, optimality requires some randomness in assessed taxable income. Even if reducing randomness is costless, taxpayers may prefer some randomness when the increased revenue can be rebated, so that the government's revenue stays fixed.

108.Sell, J., Gender, strategies, and contributions to public goods. *Psychology Quarterly*, 1997. 60(3): p. 252-265.

This study examines factors affecting cooperation in public goods settings involving relatively small groups, particularly focusing on gender composition of the group and the history of the group's level of cooperation. All game theory predictions are supported. Further, for the public goods setting examined, expectation states theories provide more accurate models of group members' behavior than social identity theories.

109.Sheffrin, S.M. and R.K. Triest, 1992, Can brute deterrence backfire? Perceptions and attitudes in taxpayer compliance, in *Why people pay taxes: Tax compliance and enforcement*, ed. J. Slemrod, Ann Arbor: University of Michigan Press, 193—218.

This essay provides new estimates of the effects of attitudes, perceptions of honesty, and perceived deterrence on tax compliance based on a national survey of 1,444 U.S. taxpayers. To account for errors and inconsistencies among their self-reported measures of tax compliance behavior and taxpayer attitudes, the authors employ a latent variables econometric methodology. The results indicate that attitudes are important determinants of tax compliance. In particular, having a negative attitude toward the tax system or perceiving other taxpayers as dishonest is positively associated with noncompliance. The results also indicate a role for deterrence. Specifically, increases in the perceived probability of detection are found to decrease the likelihood of noncompliance. Surprisingly, though, personal knowledge of someone who has had difficulties with the IRS decreases the perceived probability of detection. This raises the possibility that an increase in enforcement rates could (by lowering perceptions of the probability of detection) have the perverse effect of increasing the level of noncompliance in future years.

110.Shimshack, J.P., Monitoring, enforcement, & environmental compliance: Understanding specific & general deterrence. *State-of-Science White Paper*, 2007.

This report reviews the policy-relevant environmental compliance literature, with emphasis on the growing empirical literature measuring the specific and general deterrence effects of environmental monitoring and enforcement.

111. Slemrod, J., Why people pay taxes: Tax compliance and enforcement. 1992: University of Michigan Press.

This volume contains twelve essays on tax compliance and enforcement from a conference held at the School of Business Administration of the University of Michigan on December 7—8, 1990. The essays investigate the determinants of tax compliance from a variety of different perspectives. The resulting interchange of views brings out many new insights into the causes and consequences of tax compliance.

112. Slemrod, J. and Christian, M.B.C., Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 2001. 79: p. 455-483.

In 1995 a group of 1724 randomly selected Minnesota taxpayers was informed by letter that the returns they were about to file would be “closely examined.” Compared to a control group that did not receive this letter, low and middle-income taxpayers in the treatment group on average increased tax payments compared to the previous year, which the authors interpret as indicating the presence of noncompliance. The effect was much stronger for those with more opportunity to evade; in fact, the difference in differences is not statistically significant for those who do not have self-employment or farm income, and do not pay estimated tax. Surprisingly, however, the reported tax liability of the high income treatment group fell sharply relative to the control group.

113. Smith, K.W., 1992, Reciprocity and fairness: Positive incentives for tax compliance, in *Why people pay taxes: Tax compliance and enforcement*, ed. J. Slemrod, Ann Arbor: University of Michigan Press, 223—250.

This study provides an analysis of the impact of positive incentives on tax compliance based on a national survey of approximately 1,500 U.S. taxpayers. The author’s regression results indicate that responsive service and procedural fairness tend to increase citizens’ normative commitment to tax compliance. The results further indicate that normative commitment to compliance can be positively reinforced by the effective detection and punishment of noncompliance.

114. Snow, A. and Jr, R.S.W., Ambiguity about audit probability, tax compliance, and taxpayer welfare. *Economic Inquiry*, 2005. 43(4): p. 865-871.

This paper shows that an increase in uncertainty about the probability of being audited (ambiguity) increases tax compliance for ambiguity-averse taxpayers but reduces compliance for ambiguity lovers.

115. Stafford, M.C., et al., Modeling the deterrent effects of punishment. *Social Psychology Quarterly*, 1986. 49(4): p. 338-347.

This paper argues that a satisfaction balance model, a social psychological theory representing a modification of the matching equation and positing interactive effects of certainty and severity, fits both nonexperimental and experimental data well and substantially better than an additive model. The implications of the satisfaction balance model for a theory of deterrence are discussed.

116. Swanson, G.E., Doing things together: Some basic forms of agency and structure in collective action and some explanations. *Social Psychology Quarterly*, 1992. 55(2): p. 94-117.

Studies of the growth and continuous regeneration of groups suggest answers to recent questions about the relations between “agency” and “structure.” Several of these studies are reviewed within a framework built on G.H. Mead’s social behaviorism and on research concerning organizational “growth” and collective behavior. The explanations based on this framework are contrasted with alternative approaches.

117. Tauchen, H.V., Witte, A.D., and Beron, K.J., Tax compliance: An investigation using individual taxpayer compliance measurement program (TCMP) data. *Journal of Quantitative Criminology*, 1993. 9(2): p. 177-202.

This paper analyzes the tax compliance behavior of U.S. taxpayers by using a 1979 data set that combines information from a random sample of individual tax returns, each of which has been thoroughly audited, IRS administrative records, and sociodemographic data from the Census. They find evidence that both audits and tax

code provisions affect compliance. However, the effects are significant for only low- and high-income groups. The results for audits also suggest that the "ripple" or general deterrent effect of audits may be substantially larger than the direct revenue yield of audits for high-income taxpayers.

118. Torgler, B., Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *Journal of Economic Surveys*, 2002. 16(5): p. 657-683.

This paper looks at the experimental findings of research concerning tax morale and tax compliance, focusing on personal income. After briefly discusses the topic of deterrence, the main focus is on the social and institutional factors.

119. Turk, A., et al. Charitable contributions in a voluntary compliance income tax system: Itemized deductions versus matching subsidies. *The IRS Research Bulletin*. 2007. Washington D.C.: Publication 1500.

This paper explores the tax policy and tax administration implications of itemizing deductions versus matching subsidies of charitable contributions. It was found that subsidizing charitable contributions as an itemized deduction creates varying subsidy rates, opportunities for misreporting, and a considerable amount of burden on individual taxpayers. Subsidizing contributions via a match can disconnect the subsidy rate and the marginal tax rate and can change the opportunities and the incentive for misreporting. In addition, moving the reporting responsibility to the charitable organizations clearly reduces the burden placed on individual taxpayers.

120. Wenzel, M., The impact of outcome orientation and justice concerns on tax compliance: The role of taxpayers' identity. *Journal of Applied Psychology*, 2002. 87(4): p. 629-645.

Previous research has yielded inconsistent evidence for the impact of justice perceptions on tax compliance. This article suggests a more differentiated view on the basis of 2 congenial theories of procedural and distributive justice. The group-value model and a categorization approach argue that taxpayers are more concerned about justice and less about personal outcomes when they identify strongly with the inclusive category within which procedures and distributions apply. Regression analyses of survey data from 2,040 Australian citizens showed that 2 forms of tax compliance (pay-income reporting and tax minimization) were determined by self-interest variables. For 2 other forms (nonpay income and deductions), inclusive identification had an additional effect and moderated the effects of self-interest and justice variables as predicted.

121. Wenzel, M., A letter from the tax office: Compliance effects of informational and interpersonal justice. *Social Justice Research*, 2006. 19(3).

This study investigates the effects on behavioral compliance of reminder letters adopting principles of informational and interpersonal fairness compared with a standard reminder notice. The two fairness letters yielded a significantly greater compliance rate than the control letter.

122. Wildt, A.R. and Winer, R.S., Modeling and estimation in changing market environments. *The Journal of Business*, 1983. 56(3): p. 365-388.

In this paper various potentially useful variable-parameter models are discussed, reported marketing applications are examined, and issues pertaining to the application of this relatively new methodology to marketing problems are investigated.

123. Wilson, J.A., et al. Impact of taxpayer representation on the outcome of earned income credit audits. *IRS Research Conference*, 2007. Washington DC: Publication 1500.

This paper examines if taxpayers with representation in EIC audits are more likely to be determined eligible for EIC (and to have a higher no change rate) than taxpayers without representation in EIC audits.²⁵ It also determines if taxpayers with representation in EIC audits retain a greater proportion (measuring the proportion retained will help guard against the bias of one group claiming more EIC than the other) of the EIC originally claimed than taxpayers without representation in EIC audits.



124. Witte, A.D. and Woodbury, D.F., The effect of tax laws and tax administration on tax compliance: The case of the U.S. individual income tax. *National Tax Journal*, 1985, 38(1): p. 1-13.

This paper develops and estimates an economic model of tax compliance which specifically incorporates important aspects of the U.S. Federal income tax law. The model indicates that IRS compliance activities, taxpayer opportunities for non-compliance, and taxpayer attitudes all have significant effects on compliance.

Appendix D: Point of View Papers

1. James Alm

Using Laboratory Experiments to Investigate the Impact of Services on Voluntary Tax Compliance

A Rationale for Laboratory Experiments on Voluntary Tax Compliance

An important trend in tax administration policies in recent years is the recognition that the traditional “enforcement” paradigm of tax administration, in which taxpayers are viewed and treated as potential criminals and the emphasis is exclusively on repression of illegal behavior through frequent audits and stiff penalties, is incomplete. A revised “service” paradigm recognizes the role of enforcement, but also emphasizes the role of the tax administration as a facilitator and a provider of services to taxpayer-citizens. Indeed, many recent tax administration reforms around the world have also embraced this alternative paradigm with some success. However, while such “kinder, friendlier” provisions may improve the image of the tax authority, their actual effects on tax compliance have not, to our knowledge, been quantified.

One possible approach to measuring the effects of services on voluntary tax compliance is through the use of experimental economics. The use of laboratory experiments in economics began in the early 1960s with the work of Nobel laureate Vernon Smith. Growth in its applications came with the establishment of a well-defined framework for experimental work, and it is now widely accepted as a methodological approach in the analysis of theory and policy.

Tax compliance is an area that seems especially amenable to laboratory experiments, given limitations in both theoretical and empirical approaches. Theoretical models are not able to incorporate fully, appropriately, or tractably many factors deemed relevant to the individual compliance decision, such as services. In a sense, theory is both too complex and too simple. It is only in the simpler models that clearcut analytical results can be generated on the compliance impact of basic policy parameters; when more complex dimensions of individual behavior are introduced, results generally become ambiguous. Paradoxically, the theoretical models of individual choice are also too simple. There are numerous factors that affect the reporting decisions of individuals, but theoretical models are capable of including only a few.

Empirical studies of tax compliance are also plagued by problems, especially the absence of reliable information on individual compliance decisions. It is difficult to measure – and to measure accurately – something that by its very nature people want to conceal. It is also difficult to control in econometric work for the many factors that may affect the compliance decision, and to identify their separate effects on compliance.

In contrast, laboratory methods allow many factors suggested by theory to be introduced in experimental settings. Also, experiments generate precise data on individual compliance decisions, which allow econometric estimation of individual responses in ways that are not possible with field data. For example, penalty rates do not vary exogenously across individuals, and so it has not been possible to estimate reliably individual responses to changes in penalty rates. Independent variation in audit rates across individuals has also proven difficult to generate in natural settings. Examining with field data the impact of other factors often deemed important in compliance decisions (e.g., services) is even more problematic. Indeed, laboratory methods have examined a wide range of factors in the compliance decision, factors that have not proven amenable to either theoretical analyses or empirical analyses with field data, such as public good provision, audit selection rules, social processes, and institutions. There are some obvious limitations of experimental methods, as discussed later. However, given the weaknesses of other methodologies, as well as the strengths of experimental methods, there are compelling reasons for the use of experiments.

The Design of Compliance and Service Experiments

Experimental economics involves the creation of a real microeconomic system in the laboratory, one that parallels the naturally occurring world that is the subject of investigation. The essence of such a system is control over the environment, the institutions, and the preferences that subjects face. Of these, control over preferences is crucial. Such control can be achieved by using a reward structure that induces prescribed monetary value on actions; that is,

the investigator specifies a system by which subjects are paid, so that the amount earned by a subject is determined exclusively by the skill of the subject in figuring out the prescribed rules of the game.

In the area of taxpayer compliance, the basic design of most compliance experiments is similar. Human subjects in a controlled laboratory are told that they should feel free to make as much income as possible. At the beginning of each round of the experiment, each subject is given (or earns) income, and must decide how much income to report. Taxes are paid at some rate on all reported, but not on underreported, income. Underreporting is discovered with some probability, and the subject must then pay a fine on unpaid taxes. This process is repeated for a given number of rounds. At the completion of the experiment, each subject is paid an amount that depends on his or her performance during the experiment. Most experiments are now fully computerized.

A wide range of compliance issues has been examined using laboratory methods. Of particular relevance here is the possibility of examining the direct and indirect effects of services on voluntary tax compliance. Indeed, in some preliminary work, I and my colleagues have examined this issue. To do so, we complicate the compliance decision of subjects, and then provide services that allow subjects to compute more easily their tax liabilities. By comparing the compliance response of subjects to these service programs relative to more traditional enforcement methods (e.g., penalties and audits), we are able to determine the relative effectiveness of the alternative paradigms in generating greater taxpayer compliance.

Specifically, we complicate the compliance decision in two main respects. First, an individual's earned income is attributed to both "domestic" and "foreign" sources, and in filing their tax forms, the participants are allowed to claim a tax credit for foreign taxes paid (on the foreign income). Second, the participants are allowed to claim a deduction from total income in determining taxable income. Both the exact level of the credit for foreign taxes paid and the level of the allowed deduction may be uncertain to the taxpayer at the time of filing. This uncertainty is implemented via mean-preserving spreads (with a uniform distribution) on the tax credit and on the deduction. The subjects are always informed of the means of the allowed deduction and credit and the ranges for each, but in some sessions this uncertainty is resolved by the provision of "services" from the "tax authority".

The tax authority can either provide the information regarding the interpretation of the rules or not. In all settings the subjects face (initial) uncertainty regarding the allowed deduction and the allowed foreign tax credit, being told only the mean values and the spreads of the deduction and the credit; the level of this uncertainty is held constant throughout all treatments. When information is made available, the subjects are able to click on a button on the computer screen, and the true levels of deduction and foreign tax credit are revealed *before* they file their taxes. This mimics the ability to call a tax agency information help line and obtain accurate information at zero cost. (In subsequent research we can impose a cost on obtaining this information.) If no information is provided, then the subjects learn the true values of these uncertain variables only *after* they have filed their taxes.

Our experimental treatments are changes in the provision by the tax authority of information that reduces the credit and deduction uncertainty versus changes in the level of enforcement effort.

Our preliminary results indicate that increasing enforcement effort increases all three types of compliance. Perhaps surprisingly, providing uncertainty-reducing information has no significant impact on compliance. Again, these results are preliminary.

Information is only one mechanism by which the tax agency can increase its service provided to taxpayers. We are currently designing treatments to investigate the effects of these other mechanisms as well. Thus, even if the aggregate effect of information provision is found to be zero with larger sample sizes and broader parameter spaces, there are many other avenues by which the tax agency may be able to enhance compliance through the adoption of a service paradigm.

A Cautionary Note on Experimental Methods

There are many reasons for caution in the use and interpretation of experimental economics. A common criticism of experimental economics is that the student subjects typically used may not be representative of taxpayers. However, there is now much evidence that the experimental responses of students are seldom different than the responses of other subject pools. There is also no reason to believe that the cognitive processes of students are different from those of "real" people. Another common criticism is that it is not possible to control for many relevant factors in the

laboratory. However, if one cannot control for such factors in the laboratory where the experimenter establishes the institutions, the rules, and the reward structure, then one cannot hope to control for these factors in the naturally occurring world.

Of more legitimate concern, the results may well be sensitive to the specific experimental design, so that replication is crucial. It is also possible that subjects may modify their behavior simply because they know that they are participating in an experiment. Most importantly, there is a certain artificiality in any laboratory setting. A decision to report, say, three tokens of income in a tax compliance experiment is clearly different from a decision to report actual income on an annual tax return, even if the laboratory incentives are salient. In particular, the laboratory setting cannot capture a catastrophic loss such as jail, and it cannot capture the social stigma that some surveys suggest is an important factor in taxpayer reporting.

In short, one must use the results from laboratory experiments with some care. However, such use depends largely upon the purpose of the experiment. Experiments can be classified into three broad categories that depend upon the dialogue in which they are meant to participate. “Speaking to Theorists” includes those experiments designed to test well-articulated theories. “Searching for Facts” involves experiments that examine the effects of variables about which existing theory has little to say. “Whispering in the Ears of Princes” identifies those experiments motivated by specific policy issues. To date, most experiments in tax compliance have fallen into the first two categories. This now seems to be changing somewhat.

2. Marsha Blumenthal

In laying out my views about the various methodologies for studying tax compliance, I begin with three general comments, followed by a set of more specific ones. Along the way, I weave in my thoughts about data and appropriate indices of compliance.

In general, I see theoretical work primarily as the preferred tool for developing empirically testable hypotheses. If those hypotheses are not confirmed (e.g., actual levels of taxpayer compliance exceed the predictions of basic expected utility-maximization models), then one returns to the theoretical drawing board, amending assumptions, adding or subtracting situational features – or beginning again from scratch – and repeating. We seem to be some distance down this road for understanding the direct impact of enforcement activities; considerably less is known either about their indirect effects or about either the direct or indirect impacts of services to taxpayers. The potential contribution of computational models (ABM) in capturing (indirect) social network effects is intriguing, though the response functions programmed for participating agents would seem capable of determining whatever outcome the modeler desires.⁴²

As a second general comment, I am in favor, to the extent possible, of using micro rather than macro indices of compliance. That is, in order to estimate how an IRS activity influences individual taxpayers’ behavior, look at the individual responses of those taxpayers. Macro indices of compliance (e.g., the difference between National Income in the national accounts and all income reported on tax returns) are too crude to reliably and validly reflect any correlation with an IRS activity, even if it were possible to construct an index with and without that activity.

My third general comment concerns measuring the impact of an IRS activity, some number of tax-years after it occurs. Such “lagged” effects are considered an element of the “indirect” influence of IRS activities. As is true when studying the contemporaneous impact of IRS activities, one also needs to be careful in a multi-year context to control for the presence of any confounding events. The difficulty of identifying likely confounders and controlling for them is harder the longer is the time frame. To give an example of an admittedly somewhat crude control, in conducting the Minnesota Tax Experiment, researchers removed from the panel each taxpayer whose marital status changed significantly between the two years under analysis. Over a longer period, may other events might matter. For example, what about job changes? Residential moves? Changes in health status? Births or adoptions? To the

⁴² That is, if I model agents as behaving more compliantly when a friend has a child (because education is then a more important public good), and I populate the model with some agents who beget children, it seems circular if the resulting hypothesis is that compliance increases with the fertility rate in one’s social network. Perhaps I don’t understand agent-based models well enough to get past this sticking point.



extent that such events (some of which cannot be observed) affect treateds in an experiment differentially than controls, they will bias measures of experimental impact. While multi-year lagged effects may be important, and while it may yet become possible to develop methods of estimating them free of confounding influences, my preference is to direct the research agenda first toward understanding contemporaneous influences. The balance of my remarks reflects that view.

More specifically, the research questions posed reach across a wide range of IRS activities: from audits to correspondence to congressional testimony to press releases to walk-in centers to websites. Given this diversity, a single empirical methodological approach is unlikely to be optimal. Rather, I think that methodological approaches should be matched up with particular types of like activities. One dimension for partitioning the set of activities could be the usual carrot vs. stick (enforcement vs. service) distinction. But that distinction alone could not capture the fact that some activities initiated by the IRS and targeted to particular taxpayers will almost certainly be responded to by those taxpayers (e.g., audits) while other IRS-initiated activities, aimed indiscriminately (anonymously) at all taxpayers, might be ignored by many (e.g., a press release or a website). Surely measuring the impact of the latter requires a different approach than measuring the impact of the former. Along yet another dimension, an IRS contact with one taxpayer may spin off an external (or thirdparty) influence on a taxpayer who was not a party to that contact. Identifying the externally-affected taxpayers (e.g., the corresponding social network) is a step not needed in order to measure the impact of an audit on the audited taxpayer. Toward considering these differences more systematically, I propose a two-by-three partition of the IRS-taxpayer activity space, as in the table below. In what follows, I discuss my take on how best to measure the marginal impact of activities in each of the three rows.

IRS ACTIVITIES⁴³

	Service	Enforcement
IRS-initiated, targeted to particular taxpayers, response required	A	B
IRS-offered to all taxpayers; no response required	C	D
Third-party experiences; no response required	E	F

A and B activities

Most of the extant literature explores cell B, perhaps because there are so few examples of A activities. In my view, the most productive methodology for measuring the effects of these activities on compliance is to conduct a short-term experiment, first randomly assigning a suitably large number of taxpayers to “treated” and “control” groups, administering an enforcement or service to the treateds, and then calculating the (after-before) changes in some index of each group’s compliance behavior (“dif-i-dif”). That index must be available both for treateds and controls and both before and after the treatment. Good candidates would include reported taxable income and reported tax. Alternatively, if the returns in the after-treatment year for both treateds and controls were audited, the difference in detected evasion would be a measure of the enforcement/service impact. Access to taxpayer data (which seems available for B but perhaps not for A), properly sanitized so as to preserve confidentiality is, of course, the key prerequisite here. It would also be important to make sure, for enforcement studies, that taxpayer-IRS interactions of a service nature are distributed similarly across the treated and control groups. Symmetrically, for service studies, the rates of audit should be similar.

C and D activities

⁴³ Examples of specific activities include letters to advise taxpayers of their eligibility for a deduction/credit (somewhat of a stretch: at one time, for the EITC, A), audits (B), press release regarding provision of a public good (C), help centers (C), website(C), press release reporting a decrease in taxpayer evasion (D), a friend’s recommendation of an IRS website (E), and a sibling sharing news of her assessed audit penalty (F).

Measures of the marginal impact on compliance of this set of activities must address/consider/control for taxpayers' decisions about whether to take-up the offered activities. Here the current data collection system, as I understand it, is a roadblock. Focusing on services (C), the tension between wanting taxpayers to feel comfortably anonymous when they visit a website or call with a question conflicts with the research objectives of understanding what distinguishes such taxpayers from those who do not take-up services, much less of measuring the impact of service take-up on subsequent compliance behavior. Toward better understanding the taxpayer characteristic piece, it would seem possible to collect some non-threatening demographic data in the course of the Customer Satisfaction Surveys already being employed (perhaps the surveys already do so?). Because I doubt that the link between activity take-up and subsequent compliance will be enabled and made available to empirical researchers anytime soon, I think the second-best solution is to study that link with laboratory experiments. For example, assuming that one stimulus driving take-up is the presence of some ambiguity or uncertainty, laboratories would be good places to explore whether a taxpayer service that reduces uncertainty is associated with more or less compliance, and by what magnitude.⁴⁴

E and F activities

Here existing databases provide virtually no clues about the social networks connecting taxpayers, the information transmitted through them, or their impact on compliance behavior. Concentrating again on services (E) (though the following can be easily extended to enforcement, F), one could infer a network from the subscription lists of religious, lifestyle, or occupational publications (probably available via marketing organizations), though of course one could not ensure that an article entitled "I reduced my tax bill by visiting irs.gov" would be read by every subscriber to *Plumber's Monthly*, nor that everyone who read the article would visit the IRS website. Still, if both of those hurdles were surmounted, and if it were possible to link subscription lists to tax records (imagine the hue and cry about that!) then, in theory, one could track and measure subsequent compliance behavior. An intriguing alternative would be to use agent-based models (ABM). Though I know little about them, perhaps those project participants who are experts could explain how this method can offer a computational laboratory in which researchers could study how networks form (attract/repel new members), how they facilitate information transmission and how they support/encourage changes in member behavior.

⁴⁴ I think it is worth noting, from the literature, that some research suggests that reducing uncertainty breeds noncompliance (Scotchmer and Siemrod, 1989) while other work suggests that reductions in uncertainty lead to lower filing burdens and then, perhaps, to greater (filing) compliance (Erard and Ho, 2001).

3. Charles Christian

IBM has asked participants in the upcoming IRS “indirect effects” panel discussion to provide a one to three page “point of view” on research methods that might be used to estimate the separate marginal effects of IRS taxpayer service and enforcement activities on voluntary income tax compliance. I assume that compliance includes accurate reporting as well as timely filing and fully paying tax obligations. I also assume that all types of taxpayers are to be considered including individuals as well as corporations. The request specifically asks that recommendations consider 1) the expected term, 2) the strengths, weaknesses, and plausibility, 3) the potential for quantifying voluntary compliance, and 4) the data needs for alternative research methods.

In preparation for the panel discussion IBM distributed a recent literature review: “Indirect Effects Project: Review Of Compliance Literature” (IRS 2008). It provides an overview of theoretical, econometric (archival-empirical), experimental (both laboratory and field), and simulation-based research on estimating the indirect effects of IRS activities on voluntary income tax compliance .

I argue that future compliance research should employ all four basic methods because each has somewhat different strengths and weaknesses, although I question the feasibility of field-based experimental methods for studying the effects of IRS enforcement activities on individuals.

Theoretical Methods and Simulations

My personal view is that both short-term and long-term theoretical research and simulations would continue to be useful in understanding the indirect effects of IRS activities on all forms of compliance and for all types of taxpayers. An important advantage of both is that they can address the role of factors that are difficult or costly to observe, and both have the potential to yield useful insights into the complex interactions between the IRS and taxpayers in a game theoretic framework.

Theoretical research is particularly useful in guiding the specification of future empirical models for both econometric and experimental research. A disadvantage of theoretical research is that it is less likely to produce estimates of the magnitude of voluntary compliance relative to other methods. Simulations on the other hand are well suited to quantifying indirect compliance effects, and they have the advantage of producing estimates in the short-term.

Econometric Methods

Plumley’s tour de force research, *The Determinants of Individual Income Tax Compliance* (IRS 1996), describes the challenges inherent in collecting and analyzing archival data that were produced by disparate organizations within and outside of the IRS – data that most often are highly aggregated and data that were not created with compliance research in mind.⁴⁵

Just as daunting when employing econometric methods is specifying a model that captures the complex relationships between IRS activities and compliance. I suspect that many IRS activities are interactive in their effect, and all agree that IRS activities are not independent of taxpayer behavior. IRS enforcement in particular is clearly dependent on variation in taxpayer compliance, so it is very difficult to estimate the effect of enforcement on compliance. Although most designs, including Plumley (1996), attempt to address the endogenous nature of IRS activities, it is very difficult to assess the effectiveness of the instruments and the complex specifications used to mitigate endogeneity and simultaneity bias.

⁴⁵ Plumley argues that state-level aggregate data are preferred for compliance research because compliance is an “inherently aggregate phenomena,” but he does acknowledge that individual-level data may provide insight into the “mechanisms” of taxpayer behavior (IRS 1996, p. 42). I look forward to a discussion of the nature of the indirect effects of IRS activities (individual or aggregate phenomena) and the appropriate level of analysis.

Most also agree that the unobservable socioeconomic factors associated with voluntary compliance raise serious concerns for omitted variables bias. Fixed-effects models estimated with panel data may mitigate omitted variables bias, but only to the extent the unobserved factors are constant over time.

Taking these factors into account, it is still my opinion that long-term econometric research holds the greatest promise for quantifying the indirect effects of IRS activities on voluntary compliance, albeit with a multi-year lag. I believe it is feasible notwithstanding the formidable data collection task given the availability of the IRS National Research Program data on individual reporting compliance and the promise of data in the future on other entities.

Experimental Methods

While one could argue that randomized-controlled *field-based* experiments offer the best hope for isolating the true effect of IRS activities on voluntary compliance, the IBM literature review reflects the rarity of the design in practice. Only one paper appeared to address federal income tax compliance (Adelsheim and Zanetti 2006), and the analysis was at the city level.

Statutory protection of taxpayer confidentiality and political sensitivity pose possibly insurmountable obstacles to comprehensive field research on the effects of IRS activities, especially enforcement activities, on voluntary compliance by individuals. The Congressional opposition to the Taxpayer Compliance Measurement Program comes to mind. However, the Minnesota legislature did support field experiments testing for the effects of taxpayer service and Department of Revenue enforcement activities on voluntary compliance, so it is not without precedent. The IRS has successfully conducted national randomized-controlled field experiments on the effects of taxpayer service activities on individuals filing returns with a balance due (IRS 2004).

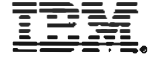
A number of randomized-controlled *laboratory* experiments have revealed important insights into behavioral responses to experimental manipulations related to tax agency activities. In my opinion, the promise of laboratory experiments lies in guiding the specification of econometric models much as does theoretical research, and this could be accomplished in the short-term with little or no data needs. I think it unlikely that laboratory experiments will improve estimates of the magnitude of the indirect effects of IRS activities on voluntary compliance.

Conclusion

Finally, I would like to comment on the potential to exploit “natural” experiments. In concept these are no different than using econometric methods to analyze archival data because they lack randomized controls. However, changes in IRS activities offer opportunities to observe taxpayer responses and make inferences about indirect effects (assuming effective econometric controls).

For example, according to the attached *Wall Street Journal* article (Dale 2008) the IRS is planning to mail notices this fall to a wide group of taxpayers informing them that they may not be reporting enough income on their tax returns. It seems feasible to compare the year-over changes in reporting by the recipients relative to a quasi-control group and make inferences on the effect of the notices.

Of course we could learn much more about the indirect effects of these new notices if the CP 2057 program managers would collaborate with the Office of Research in the selection of taxpayers to receive the notices. I will leave the feasibility assessment of that strategy to others.



References

Adelsheim, P.D., and J.L. Zanetti (2006). "The Effect of Targeted Outreach on Compliance." IRS Research Conference. Washington DC: Publication 1500.

Dale, Arden (2008). "IRS to Ramp Up Warning Letters." *Wall Street Journal*, August 21, 2008, p. D3.

Internal Revenue Service (1996). The Determinants of Individual Income Tax Compliance: Estimating The Impacts of Tax Policy, Enforcement, and IRS Responsiveness. Publication 1916 (Rev. 11-96).

Internal Revenue Service (2004). "Experimental Tests of Remedial Actions to Reduce Insufficient Prepayments: Effectiveness of 2003 Letters." Wage and Investment Division Research Project 5-04-06-2-058N.

Internal Revenue Service (2008). "Indirect Effects Project: Review Of Compliance Literature." Contract TIRNO08-K-00292 (7-08).

IRS to Ramp Up Warning Letters, By ARDEN DALE, Wall Street Journal, August 21, 2008, Page D3

This fall, more taxpayers won't be able to say they weren't given fair warning.

The Internal Revenue Service is ramping up efforts to mail out notices in October to a wider group of people, noting that they may not be reporting enough income on their tax returns.

The warning letter differs from one the IRS has been sending for years, which suggests specific changes to income, payments, credits, deductions or other parts of the return.

That letter, known as CP 2000, proposes an adjustment to the return and may ask the taxpayer to pay additional tax. Nearly 3.5 million such letters were sent for 2007, and will continue to be sent out.

The new letter, called a CP 2057, tells recipients they may be underreporting, and instructs them to double-check parts of the return and file an amended return Form 1040X if in error. It doesn't, however, specify the amount of additional tax that is owed.

The push is part of the IRS's desire to narrow the so-called tax gap, or the amount of uncollected taxes. Some tax experts view the new notices as helpful reminders, while others say they will rattle many taxpayers. "If you get one of these things, you should respond," says Robert E. McKenzie, a partner at Arnstein & Lehr, a law firm in Chicago.

The new warnings are "a major rollout by the IRS, using a totally different form than in the past," says Michael O'Keefe, a manager of tax research at H&R Block.

A CP 2057 would most likely show income omitted or an expense such as mortgage interest. It might note the discrepancy if the taxpayer reported \$10 in interest income but information documents showed an interest income of \$200.

It asks the taxpayer to work with employers and others to correct errors in W2s, 1099s, K1s and other documents. Finally, it says the IRS will scrutinize the following year's return to see that similar problems don't crop up. Roughly 31,000 of the CP 2057 letters will go out in the fall, and the IRS will expand their use depending on how effective they are in getting people to correct their own returns.

The new notice is automatically generated using an IRS system of computer-matching returns with 1099, W2, K1 and other documents -- a system the IRS also uses for the notices it currently sends.

The IRS hopes that the automated nature of the notices will allow it to run the program at low cost. It is designating only a tiny portion of its staff to work on it, with the aim that taxpayers will be able to resolve underreporting issues without having to correspond extensively with the IRS.

But some tax experts question whether the program will be effective. Robert Kerr, senior director of government relations at the National Association of Enrolled Agents, wondered how the IRS is going to field all the calls from taxpayers with concerns about the notices.

The warning letters mark a significant shift in the IRS's strategy for recovering taxes, according to Thomas P. Ochsenschlager, vice president of taxation at the American Institute of Certified Public Accountants. By giving taxpayers lead time to correct an error without actually being penalized, it "avoids the 'gotcha!' situation that happens so often," particularly for small-business owners who file Schedule Cs. The IRS has identified Schedule Cs as being particularly rife with potential errors.

Write to Arden Dale at arden.dale@dowjones.com

URL for this article: <http://online.wsj.com/article/SB121928519997559171.html>

4. Joe Cordes

Given the paucity of available data available both from academic research and from IRS data estimating the marginal impact of IRS services on compliance will require both creative use of existing research and new research to produce better estimates. My comments below represent thoughts on how such a research agenda might be structured.

1. Defining What is to be Measured

The aim of the proposed project is to measure the full marginal effect of IRS services on compliance. This, in turn, presumes that there is an agreed-upon measure or set of measures of taxpayer compliance.

- The 1990 study by Dubin, Graetz, and Wilde focuses on filing compliance, reporting compliance, and assessed liability as measures. Plumley (1996) focuses on filing and reporting compliance. Audit probability (e.g. as measured by DIF scores or similar indexes) might be another measure. Are these the main outcomes that one would like to measure? Are there other dimensions of compliance to be considered?
- What concretely does it mean to capture the marginal impact of IRS customer services on such measures, as distinct from other IRS actions?
- What issues are involved in making the measure(s) operational, not only for research purposes, but also for the purposes of “real-time” tracking by the IRS as part of its on-going operations? Put a bit differently, knowing what one would like to measure could, over time, affect IRS efforts to obtain and maintain administrative records that can then be used for performance measurement and assessment.
- Are there “intermediate outcome measures” (such as taxpayer satisfaction) that (a) are relatively easily measured, while at the same time (b) can be linked conceptually and empirically to final compliance outcomes so that one can infer final impacts from changes in intermediate measures (based on empirical estimates of the link between intermediate measures and final outcomes)?
- What specific types of taxpayer services are the “object of inquiry” and how would one measure changes in the levels of such services? Which of these measures can be easily linked to the ongoing need for the IRS to measure and track its services?
- At the risk of interjecting a “four-letter word” should any effort be made to identify possible links between compliance measures developed for purposes of this research and the development of PART (Program Assessment Rating Tool) measures currently required of federal agencies?

2. Strategies for Measuring Effects

Once one has defined what is to be measured, the next issue is to devise empirically valid ways of determining whether (a) there is a causal link between varying levels of taxpayer services and (b) one or more compliance measures, and (b) determining the quantitative size of such effects.

- **Field studies** (e.g. Blumenthal and Slemrod, Slemrod and Christian, Adelsheim and Zanetti) have many attractive attributes. It would be great if data from existing field studies could shed some light on the question at hand, however, these studies were designed to answer other questions so the ability to extrapolate from the results of these studies may be limited. The general structure of a new field study specifically designed to estimate the relationship between IRS services and compliance could be similar to the earlier Slemrod and Christian study: a group of taxpayers would be randomly selected and offered a set of compliance services by the IRS; another set of taxpayers would not receive such services and the compliance outcomes of the two groups would be compared. Within the field study, one might consider offering differing levels or mixes of IRS services.
- **Laboratory experiments** (e.g. Alm and McKee, Alm, Jackson and McKee) have also been undertaken of compliance behavior. Such experiments would seem easier to set up than a field experiment, although the “realism” of the results from such exercises would seem to be less than from an actual field experiment. In addition, one would need to devise ways in the laboratory of “simulating” IRS services.” It may be easier to create laboratory equivalents of audit rates than of compliance services.
- **Nonexperimental studies** Plumley’s 1996 study is an example of an effort to estimate the impact on compliance of various IRS activities using data from existing sources. Although there are “IRS responsiveness variables in this study, these variables are defined in terms of IRS phone contacts presumably because this

dimension of service is relatively easy to measure. The advantage of this approach is that it relies on data that are available from other sources, and can be linked to operational data from the IRS assuming that such data are gathered and maintained to provide the requisite cross-sectional and time variation in the variables. A downside would be the usual concerns that are voiced about non-experimental studies of this sort, which one would attempt to address with the usual econometric techniques.

- **Natural experiments.** One can attempt to take opportunities to estimate parameters of interest presented by “natural experiments” arising either from exogenous variation in the availability of IRS services over time, or across IRS regions, or perhaps variation in taxpayer compliance programs run by the states. The question is whether such natural experiments exist.
- **Formal Benefit Cost Analysis of Compliance Services.** One might also attempt to undertake a formal benefit-cost analysis of investments in taxpayer compliance services in the manner in which such analyses are frequently undertaken by other federal agencies as part of the regulatory process. Such work could build on Plumley (2007). The goal of such an analysis would be to (a) develop the appropriate social accounting framework for the benefits and costs of the various IRS service activities, (b) identify measures of the relevant benefits and costs, (c) identify ways of quantifying these benefits and costs by extrapolating from estimates from existing research, and (d) attempt to come to a bottom-line estimate of total benefits and costs. Such a study could proceed independently of and in tandem with efforts to obtain “better estimates” of the impact of IRS services. My own experience in doing benefit-cost analyses with existing (though flawed and incomplete) data and estimates is that such exercises help clarify what one is trying to measure in a manner that complements research to improve measurement of outcomes; and often suggests ways of arriving at reasonable estimates of outcomes that one might initially have thought were unavailable and/or difficult to measure.

5. Jon Davis

1. One size does not fit all for this research problem. The IRS provides a wide range of services. Each is targeted to different taxpayer constituencies. The effect of enforcement efforts is not uniform across the population. For most offered services, taxpayers self-select as users. Similarly, there are different notions of non-compliance (filing, reporting, and payment) identified and the reasons underlying each of these are orthogonal in important respects. The issue is further complicated by the presence of third-party tax service providers, who also elect to use various IRS services in their work and who compete with the IRS as service and information providers in some cases. As a result of the above complexities, using a single methodology is not the best solution to a full understanding of the indirect effects of services and enforcement activities on voluntary compliance.

2. Because of the complexities noted in (1), a contingency approach to theoretical development and data analysis should be a central strategy for the research. That is, one aspect of the research program should be to independently focus on specific services and their impact on specific forms of noncompliance. Archival analysis and related theoretical work would likely be the most important methodologies for this part of the research program, with experimental methods taking a supporting role. Econometric analysis could also be used to estimate quantitative effects (although I'm not an econometrician so I won't opine as to how this might be done).

3. Understanding the relation between specific forms of compliance and specific current service and enforcement activities (and related magnitude estimations) may not provide the full picture and may provide misleading conclusions because the dynamics of taxpayer behavior in the aggregate may be related in complex ways to the portfolio of services and enforcement activities provided by the IRS. Analysis of the relationships is further complicated by heterogeneity in the taxpaying population (differences in motivations for noncompliance and differences in forms of noncompliance), interaction within this population regarding perceptions and values, and competing sources of information (IRS, third party service providers, and other taxpayers). Agent-based models (ABMs) could help deal with these difficulties (and especially provide insights into dynamics of taxpayer behavior in response to changes in the service/enforcement portfolio). The construction of ABMs could be informed and validated by insights provided from the research suggested in (2). Recent innovations in representing realistic social networks in agent-based environments also provide some substantial promise for construction of tax compliance ABMs (if such network complexities are necessary to create a valid model, which is an open question). Some ABM predictions may be amenable to subsequent testing through further archival and experimental analysis to provide a more comprehensive understanding of the relation between services, enforcement and compliance. ABMs in this context would aid in understanding qualitative relationships between compliance and indirect effects (especially dynamics) but not in quantitative estimation. It should be noted, however, that ABM is still a fairly young method and it does have some potentially serious limitations. Seemingly innocuous assumptions may have unknown or unexpected impacts that are not yet fully understood, solution concepts are weak and there is, to date, little formal work on replicability using different models. Some of these limitations can be overcome with sufficient care.

4. The evaluation of the impact of services on voluntary compliance in (2) should include analysis of quality of service measures (including taxpayer perceptions). Following on (3) above, the overall taxpayer perception of the tax agency and the tax system (formed in reaction to personal experience with specific services enforcement activities and through the experiences of others via social networks and other mechanisms) may be more useful in evaluating indirect effects than measures of specific services.

5. The role of third party service providers and the effect of their use of IRS services seems to me to be an important (but complex) piece of the overall picture.

6. Regarding data, in the short-term, it would be helpful to collect information on the characteristics of taxpayers that utilize various classes of IRS services if available. In some cases (e.g., SPEC), it may be straightforward to collect data on taxpayer/user characteristics. Longer term, it would be useful to collect additional data during NRP audits if this is possible. Ideally, more than the standard demographics, income levels etc., would be collected, including variables that could be related to compliance and indirect (ripple) effects (from a sociological perspective).

6. Don DeLuca

Measurement models are inextricable from the nature of data collection. Developing a research design to measure and model the separate marginal impacts of taxpayer services and enforcement activities on voluntary taxpayer compliance requires assembling or creating detailed data about taxpayer behavior. Research that seeks to determine taxpayer behavior from the results of tax filings, tax audits, etc. are searching for impacts or influences “revealed” by the data. In comparison, research that seeks to understand taxpayer values, attitudes, perceptions, and experiences as they relate to compliance behavior is looking for impacts or influences “perceived” by the taxpayer.

Measurement models used in the “revealed” impact research design employ analysis of secondary data sources (i.e., IRS administrative or other collected data). By and large these data are not collected for the purpose of constructing measures of the marginal impact of IRS services on voluntary tax compliance, and hence, dependent (and endogenous) variables are *ex post facto* constructs. Yet the direct connection of the results of this research to taxpayer compliance behavior outcomes is compelling. Needless to say, the research problem at hand would benefit greatly from the careful development of more comprehensive compliance databases that capture indicators appropriate for measuring and modeling the marginal impacts of taxpayer services.

By contrast, measurement models used in “perceived” impact research designs employ primary data collection (i.e., mostly surveys, but I am also including laboratory experiments and qualitative research in this category). Developing measures in these research designs are directly constructed through questionnaires, moderators’ guides, lab experimental protocols, etc. Yet the advantage gained by direct measurement is restrained by measurement challenges derived from the fact that the dependent variable involves asking threatening questions about behavior (i.e., noncompliance). Threatening questions are intrinsically more difficult to ask than nonthreatening questions. Many survey respondents are reluctant to give full and honest answers to questions they may consider threatening (i.e., illegal), such as questions on tax noncompliance. As the questions become more threatening, research shows that substantial response bias should be expected regardless of the survey techniques or question wordings used.

More accurate measurements of threatening topics from the survey research literature include use of open, long questions with familiar words, use of key informants, deliberate loading of the questions, the use of an appropriate time frame for socially desirable questions, asking about last behavior rather than usual behavior, asking about whether behavior has ever been done before asking about current behavior, embedding the topic to reduce perceived importance, use of anonymous forms such as card sorting and randomized response. Additional questions are useful at the end of the survey questionnaire to help reveal the respondent’s view of the perceived threat about the topic of tax noncompliance. Also, any possible validation of administrative records is helpful, even if only available at the aggregate level.

Results from carefully crafted laboratory experiments provide excellent insights into taxpayer behavior. But these designs are not easily generalized to the population of all taxpayers, and some designs have been criticized as artificial representations, not likely to reflect true taxpayer decisions and behavior beyond the laboratory. One extremely useful outcome of the results of these studies is to inform the measurement of operational variables for quantitative studies. And yet, in order to predict the *marginal* impact of IRS services and enforcement on voluntary compliance implies an experimental research design (using experimental/control groups or operationalizing control variables) such as those implemented in laboratory research designs.

The data collected for analyses, whether from existing administrative records or from new primary research, should be a large dataset and of sufficient high quality to generalize to the larger population of taxpayers so that tests for significant differences between and among population subgroups is allowed. For survey research designs, careful questionnaire construction, random sample selection, and data collection protocols designed to achieve the maximum possible response rate are minimal requirements for obtaining high quality results. Better yet, a hybrid design that combines taxpayer administrative records to survey responses (through sample selection) would combine the best of revealed and perceived measurement methodologies referenced above.

So to summarize, it is my point of view that improving measurement and data collection aspects of research designs will go a long way towards advancing knowledge about the marginal impacts of IRS services and enforcement activities on voluntary compliance. Short term advances could derive from laboratory and other qualitative experimental designs. Medium term gains from hypotheses testing using quantitative surveys and analyses of



administrative records would advance both theory and empirical evidence. However, I believe the most promising methodology for long term gains would be achieved best from a hybrid approach that combines large-scale quantitative surveys with administrative records linked at the micro level.

7. Brenda Dietrich

- **What methodology do you recommend? Is this an approach for the short term and long term (since short-term, limited studies may be useful for refining the design of longer-term studies and data collection)?**

The IRS may wish to consider structuring the research plan as a portfolio of different research approaches and to monitor the performance and productivity of each of those research directions through time to determine continued support. In defining this portfolio, the IRS should consider both activities aimed at exploiting available predictive information about taxpayer responses, and experiments aimed at exploring taxpayer responses to improve predictions. Further, the IRS should use available data to segment the taxpayer population to improve the value of predictive models, and should monitor and refine the segments as additional data becomes available.

The IRS should consider constrained resource allocation and optimization as one of its primary research directions. Research resources to be considered include but are not limited to funds and skilled personnel. Data is also a resource but fortunately, unlike money and labor hours, it is a resource that once acquired can be used in multiple activities without being consumed. The predictive value of data does tend to degrade over time, so some financial and labor resources need to be expended to either refresh or revalidate aging data sets. Additional resources to be considered are those that actually deliver the services to the taxpayers, such as staff at call centers or help centers, informational mailings, and websites.

Several dimensions across which to segment the taxpayer population should be considered, including geography, filing method (e.g., hand written, computer aided/mailed, online, preparer), income source (e.g., wages, investments, farm/business, self employed) and demographic information (age, primary language, income level, etc). The response to an IRS services may differ significantly across the range defined by each dimension. Ideally, for each defined taxpayer segment (defined for example by a geography, a filing method, an income source category, and a demographic category (e.g. suburban NYC, computer aided/mailed, >90% wages, 35-65, English, 100-250K)) and each potential service or action (800 number, web site, mailing, walk-in help center, informational letter, etc) a response coefficient, indicating the additional tax paid (or additional probability of compliance) can be estimated from the available data. The predictive value of these estimates depends on the quality of the data, the integrity of the segmentation of the ranges, and having selected the right set of dimensions (features). Quality of data can be improved by targeted experiments, which can either validate estimates, or provide insight for refining the features and/or segmentations. However, experiments, analysis of data from experiments, and other IRS service activity all require both funding from the research budget and time of skilled research staff.

- **Likely strengths and weaknesses of alternative methodologies, including an explicit assessment of the plausibility of the assumptions made in each**

The literature reviewed provided few examples of taxpayer segmentation, and in most cases the segments seem to have been generated prior to the collection of data, rather than derived from the data. In numerous other examples of predictive modeling (e.g. insurance underwriting, airline loyalty, business-to-business propensity to buy analysis), the segments (or clusters) representing populations exhibiting similar behavior, are derived from the data. Within each segment a response curve (or in linear cases, a coefficient) is computed from the data. Available data will need to be examined to determine whether the data is sufficient to support segmentation.

- **Approaches to quantifying voluntary compliance (filing, reporting, and payment compliance are all included, but could conceivably be studied separately)**

Voluntary compliance should be quantified by taxpayer segment, once appropriate segments are identified. Change in compliance in response to actions should be estimated for each applicable segment. Return (new tax collected) per responding individual should also be estimated for each action and each segment. Additionally the required resource for each action (or each "unit" of each action) for each segment should be estimated. Time effects should be modeled (e.g. does action need to be repeated in each year, or does effect carry over to future years with some decay). Availability of each resource should be estimated. A resource allocation model in which the decision variables represent the amount of each action to take with each segment can be formulated. An objective function, such as total new tax collected, or some weighted function of new tax collected, can be formulated. Additional policy constraints setting minimum and maximum levels of each activity for each population can be specified.

Mathematical optimization methods (e.g. linear programming) can be used to determine the “optimal” levels of activities. Single year models and multiple year models can be formulated. Sensitivity of the optimal solution to specific data elements (e.g. response rates) can be determined. Additionally, allocation of resources to experiments designed to collect additional data can be incorporated into the optimization model.

▪ **Current data availability and future desired data**

With recent advances in computational power and data analysis technology, it may be possible to “mine” data sets that represent a significant portion of the taxpayer population. Such mining could be used to determine segments, to detect year to year changes and long term trends, and to identify anomalies. It could also be used to compare against other sources of demographic data as permitted by policy (e.g. labor statistics, census). Privacy issues will need to be addressed before such a large scale mining effort is launched.

The Statistics of Income CWS file (around .1% (120,000) of taxpayer population) has been used to analyze impact on collection. The characteristics of this file should be examined to understand whether it provides a distribution of data points that is sufficient for prediction of taxpayer responses. Perhaps more importantly, some consideration should be given to creation of a data set that can be used to gain insight into non taxpayers, perhaps through tracking of some subset of newly issued taxpayer identification numbers.

8. Jeffrey Dubin

I propose to update and extend the empirical approach of Dubin, Graetz, and Wilde. This method was successfully used to measure direct and indirect effects of audits. It was most recently extended in work that I did for the IRS Criminal Investigations Division in two completed projects. See “Criminal Investigation Enforcement Activities and Taxpayer Noncompliance, in Public Finance Review, Volume 35, June 2007. My approach to measuring the marginal impacts of services on compliance extends the DGW analysis by (i) updating the econometric model and (ii) by including explanatory factors that have been identified by the IRS for service center and web taxpayer contact as well as for taxpayer assistance and return preparation.

The empirical approach used in this article follows Dubin, Graetz, and Wilde (hereafter, DGW; 1987, 1990). The DGW method can determine specific and general deterrence effects of IRS service activities, as well as the effects of audit rates on taxpayer compliance. Although the general deterrence effects provided by audits have been widely acknowledged, the IRS has never reported the “spillover benefits” of audits. Spillover benefits are the increase in collections from taxpayers, whether or not they are audited, who report more taxes in response to an increased likelihood of an audit. DGW’s principal innovation was to directly estimate taxes due rather than first attempting to construct a noncompliance measure and then extrapolating from noncompliance to revenue. The empirical analysis in DGW was based on two econometric models that were both estimated using a state-level, time-series, cross-section data set for the years 1977-1987, with various measures of compliance specified as a function of audit rates and socioeconomic factors.

The DGW empirical analysis was based on three models that were estimated using a state-level, time-series cross-section. One model specified reported taxes per return filed as a function of audit rates and a variety of socioeconomic control variables. The second model used assessed taxes per return (combining reported taxes with additional taxes and penalties) as the dependent variable. The final model specified returns filed per capita as a function of the same variables. DGW selected explanatory variables for this system of equations based on two considerations: the size of the tax base and tax-payers’ compliance behavior. Clearly, some factors affect both the tax base and taxpayer compliance. In general, this effect may lead to ambiguous predictions for the a priori signs of some regression coefficients. Taxpayers confront three options: to file a return and report honestly, to file a return and underreport taxes, or not to file a return. Deterrence theory maintains that factors that either reduce the benefit or increase the costs of filing a return and underreporting taxes will increase the likelihood of selecting one of the other options. Variables related to compliance behavior either reflect opportunities to evade (e.g., more educated or savvy taxpayers) or the enforcement activity of the IRS (e.g., examination and criminal prosecution). With respect to variables that relate to the tax base, any change that increases the tax base (e.g., shifting taxpayers above minimum reporting requirements or into higher tax brackets) will increase reported taxes and the number of returns filed. In addition to updating the DGW models, a specific objective of this study is to augment the list of compliance factors to measure the nature and extent of service center activities.

With respect to civil audit examination, a measure such as the audit rate may be significant to a potential tax evader because it measures the probability that the taxpayer will be subjected to an audit. In the current setting, the natural analogue to the audit rate is the rate at which taxpayers have contact with IRS service activities. The IRS tracks both web based and phone based service contacts. Data has been collected by the IRS with respect to the number of website visits, page views, and customer satisfaction. In addition, the IRS tracks the volume of telephone calls, call accessibility, call quality and satisfaction. While the data exist in somewhat aggregate form, phone number information and web visit data may allow a breakdown of web and phone activity by geographic region and time period. The DGW analysis combines state-level and time-series information to estimate econometric models using time-series cross-section methods. The aggregate information present for call center and web based contacts may be problematic for micro analysis but is well suited to the DGW approach. Additionally, the IRS tracks detailed information on Taxpayer Assistance Center contacts including volume of visits by area and various quality control measures. The IRS also tracks return preparation services and taxpayer outreach contacts. Again aggregate information measured over time could be assembled to fit into the DGW econometric modeling framework. There are several good candidates for explanatory factors that could affect specific and general deterrence of IRS services and these variables can be tested and measured in the DGW framework with immediate calculation of marginal impacts and cost and benefits of IRS provided services.

9. Brian Erard

1. Deterrent Effect of Tax Audits

Although there have been a number of empirical studies of the deterrent effect of tax audits using a variety of alternative approaches, I think it is fair to say that we do not yet have a firm understanding of the extent to which audits pose a deterrent to tax noncompliance.

1.1 General Deterrent Effect

Of particular interest is the general deterrent effect of tax audits, or in other words, the impact of the perceived risk of an audit on tax compliance. To improve our understanding of this issue, one potentially useful approach is to apply regression techniques to a combination of NRP and operational tax data. I have done some preliminary work along these lines with colleagues from the IRS (Ed Emblom and Chih-Chin Ho). The basic idea is that one uses a choice-based sample including returns that were and were not subjected to an operational audit to estimate the objective risk of an audit as a function of tax return characteristics within an audit class. Next, one uses the resulting prediction formula to estimate the risk of audit for each randomly selected return from that audit class that is included in the NRP data base. Finally, one performs an econometric analysis of the magnitude of noncompliance (e.g., the adjustment made to reported tax by the NRP examiner) as a function of the predicted risk of audit and other taxpayer characteristics. This provides a measure of the responsiveness of noncompliance to a change in the risk of audit (holding other factors constant), or in other words, the marginal general deterrent effect of an audit. If multiple years of NRP data were available (a time series of cross-sections), one could expand this approach to account for changes in audit risk over time as well as across individuals. A challenge for this approach is that noncompliance is imperfectly measured. In principle, this can be addressed using detection controlled estimation techniques, but this adds considerable complexity to the analysis. In addition, data bases such as the NRP have only limited details on taxpayer characteristics, largely based on what is reported on the tax return. In principle, though, additional information about the taxpayer could be collected at the time of the audit if it was deemed important for the analysis.

A second approach to studying general deterrence is to employ an updated aggregate time-series cross-sectional analysis along the lines of the studies by Dubin, Graetz, and Wilde and Plumley. Under this approach, one estimates regression equations that relate aggregate measures of reported income or reported taxes (as well as the filing rate) over distinct geographic areas, such as IRS districts, to the audit rate for that area as well as a variety of control variables. The marginal predicted changes in the dependent variables to a change in the audit rate, holding other factors constant, are then used to assess general deterrence effects. My chief concern with this approach is that variations in aggregate filing and reporting for tax purposes across areas and over time are influenced by a wide variety of factors (socio-economic variables, tax laws, state and local enforcement conditions, etc.) beyond the federal audit coverage rate. Moreover, much of this variation has nothing to do with tax noncompliance. Properly controlling for all relevant factors poses a significant challenge for the analysis; if some factors are omitted and happen to be correlated with federal audit activity, the results of the analysis are likely to be misleading. As noted in the review of the compliance literature provided for the conference, applications based on this approach have generally produced very large deterrence estimates, both in absolute terms and in relation to other estimation approaches. Before large estimates such as these can be relied upon for IRS resource allocation decisions, it is important to have confidence that they really reflect what they are intended to represent (i.e. the impact of audit activity) and not the impact of confounding factors.

A third potentially useful approach to examining the general deterrence issue is to perform a field experiment. The Minnesota study of Slemrod, Blumenthal, and Christian provides an example of such an approach. This study provides a fascinating test of the standard economic theory of tax compliance, under which taxpayers are fully aware of the risk of an audit. However, in practice, taxpayers normally are not directly informed of the risk of an audit by the tax agency. So, it might be interesting to examine how an unannounced change in audit risk impacts upon compliance behavior over time. One possibility would be to ramp up the audit rate in one or more locations, while holding the audit rate constant or ramping it down in one or more comparable locations that are geographically separated (to avoid contamination effects). One could then perform a difference-in-differences (DID) analysis of reporting behavior over time in the separate locations.

More complex experiments are also possible. For instance, audit rates in one area might be increased during the first year and then decreased the following year. In a second area, they might be decreased during the first year and then increased the following year. In a third area, they might be maintained at an increased level for two years, while in a fourth area, they may be maintained at a reduced level for two years. Finally, a fifth area would serve as a control with audit coverage levels held constant over the two years. This would provide a more powerful test of deterrence by providing more contrasts that could be tested. In addition, it would allow a somewhat enhanced analysis of the dynamics of the deterrence process.

Ideally, one would also perform surveys over the test period to investigate the extent to which taxpayer (and perhaps tax practitioner) perceptions of the risk of audit changed over time in the different areas. This would help to clarify whether observed changes in reporting behavior are linked to changes in taxpayer perceptions of audit risk. Clearly, a field experiment of this sort would present some significant administrative challenges, but it would provide a somewhat cleaner test of deterrence than studies based on observational data. In observational studies of deterrence, one needs to be mindful that audit rates are likely to be endogenous (i.e., to depend on reported magnitudes of income and taxes). Finding suitable instruments to convincingly control for endogeneity is a significant challenge for such studies. In a field experiment, audit coverage rates could be experimentally manipulated, which would eliminate the need to employ instrumental variables procedures. A challenge that would remain for a field experiment would be verifying the comparability (and possibly adjusting for any lack of comparability) among the taxpayers from the various geographic areas that are subjected to different rates of audit coverage.

1.2 Perceptions of Audit Risk

As alluded to in the preceding discussion, perceptions of audit risk tend to differ from the objective risk and are likely to respond to changes in objective risk with a lag. Relatively little is known about how perceptions of audit risk are formed and transmitted among taxpayers. For instance, suppose individuals tend to form their perceptions of audit risk primarily from national news reports and from published statistics on nationwide audit coverage rates. In this case, tests of deterrence based on differences in objective audit rates across tax districts would presumably have little power. Alternatively, suppose that perceptions of audit risk were found to be based on objective audit coverage rates of individuals from the same occupational group, business activity, or economic class. In such a case, it would seem useful to disaggregate audit coverage rates by occupation, industry, or economic status when studying deterrence rather than to work with overall audit coverage rates. As a third possibility, suppose that perceived audit risk depends largely on one's knowledge of the enforcement experiences of one's family, neighbors, and acquaintances. In such a case, one might want to perform an analysis of how subsequent reporting and filing behavior among different "networks" of taxpayers is affected when one or more members of the network are subjected to an enforcement action.

To improve our understanding of perceptions of audit risks, one approach would be to perform periodic surveys so that one would obtain either a time series of cross-sections or a panel data base on audit perceptions. In addition to inquiring about audit perceptions, the surveys would also inquire about relevant taxpayer characteristics and attitudes. In this way, one could generate reasonable estimates of objective audit risk, explore what factors seem to drive audit perceptions, and investigate how these perceptions respond to changes in objective risks. It would probably make sense to develop a comparable survey for tax practitioners. Tax practitioners are employed by a large percentage of individual taxpayers, so their perceptions of audit risk are also of importance. One might use the results of such surveys to calibrate and/or test agent-based models of information transmission. Such models might help to generate new insights into how audit perceptions evolve, potentially leading to improved strategies for conducting audits and measuring their impact.

1.3 Specific Deterrent Effect

The impact of audits on a taxpayer's subsequent compliance behavior (i.e., specific deterrence) is another area where available evidence is limited. Some time ago, I performed a study of this issue using a combination of operational and random audit data. In particular, the 1982 TCMP data base included some taxpayers who, by chance, had been subjected to an operational audit two years prior to the random TCMP audit (of their 1980 tax return). I estimated an econometric model where noncompliance on the taxpayer's 1982 tax return was specified as a function of whether the taxpayer's 1980 tax return had been audited as well as a variety of other control variables. A separate equation accounted for the likelihood that a return filed for the 1980 tax year would have been audited. This equation was estimated using a sample of audited and unaudited returns from that year. In an effort to account for

sample selection bias, I allowed for a correlation between the errors of the noncompliance and audit equations. The specification for the likelihood of a tax year 1980 audit was based on only a few explanatory variables, and ultimately, it proved difficult to obtain a convincing estimate of the selection bias term. However, it might be worth attempting an improved specification using more recent data. Perhaps the NRP includes a sufficient number of returns that were subjected to a prior audit, in which case it could be employed along with some operational data relating to the prior audit year. Alternatively, when designing the sample for future random audit studies, one could oversample returns that had been recently audited.

For confirmatory purposes, one might also estimate an alternative to a sample selection model. For instance, one might use a statistical matching procedure. Specifically, each taxpayer who received a prior year audit could be matched against a taxpayer who, in the prior year, was observationally similar, but was not audited. The average within-match difference in noncompliance in the random audit year could then be computed and compared with the estimate of specific deterrence based on the sample selection model. This comparison would provide a sense of how sensitive the results are to the method one uses to account for differences among taxpayers selected for an operational audit and taxpayers who are not selected.

A randomized field experiment would provide an alternative method for assessing the specific deterrent effect of an audit. As an example of such an experiment, suppose that one had a list of taxpayers within an audit class whose returns had been selected by IRS classifiers for examination. In this example, a randomly selected portion of the selected returns would be subjected to operational audits in that year, while the remainder would not be audited. Then, in a subsequent tax year, all of the taxpayers in the original sample would be subjected to a thorough audit of their returns for that year. A comparison of the subsequent year audit results for those who had and had not been audited in the earlier year would then provide an estimate of the specific deterrent effect of an audit. As an alternative strategy, one could perform a difference-in-differences analysis of *reported* taxes filed by the two groups using the year of the operational audit and one or more subsequent tax years. The advantage of the latter approach is that one would not need to perform detailed audits of all taxpayers. On the other hand, results based on such audits might be more informative and convincing.

As an alternative to a randomized field experiment, one might uncover certain “natural experiments” where, as a result resource limitations, only a fraction of the returns that are classified for an audit wind up being audited in some locations each year. If those classified returns that are ultimately audited are considered to be reasonably comparable to those classified returns that are not audited, one could employ a difference-in-differences analysis of reporting in the current and subsequent year(s) among the two groups of returns to see whether there appears to be a compliance impact of the audits.

1.4 A New Data Source on Verification and Compliance

In addition to data that would be obtained from field experiments, surveys, or future random audit studies, I think it is worth considering whether something akin to the Canadian Processing Review program should be adopted. This program involves various credit and deduction items on the tax return that cannot be verified from information already available to the tax agency. There are two elements of the program. The first element involves the selection of a random sample of returns each year for verification of selected return items. Typically, a selected taxpayer will receive a request to submit receipts pertaining to one or more credits and deductions claimed on the return. If the receipts are submitted and match the amount claimed, no action is taken. Otherwise, an appropriate adjustment is made and any additional taxes are assessed. A data base records the results for each return in the sample. With the aid of these random verification results, criteria are developed for targeting returns that are potentially noncompliant with respect to one or more of credits or deductions. Then, under the second element of the program, these targeting criteria are applied to select additional returns for compliance verification.

I believe such a program may have a number of potential benefits. First, from a compliance perspective, this is a relatively low-cost way of verifying and improving compliance among taxpayers who, in many cases, would not be productive to audit in a face-to-face setting. Second, from a research perspective, this would provide a valuable data source for understanding deterrence. Specifically, it would make it possible to examine to what extent compliance on various credit and deduction items improves as a result of both random and targeted verification efforts.

2. Compliance Impact of Taxpayer Services

Since relatively little formal research has been conducted on the impact of taxpayer services on tax compliance, it would be useful to begin by developing some profiles of taxpayers who use the different types of services that are available, if this has not already been done. This would include details on the types of questions or problems the taxpayers are trying to address, their geographic location, and to the extent that the information can be collected, other more detailed characteristics.

2.1 Direct Impact on Service Recipients

For certain taxpayer services, eligibility rules might be helpful in identifying the impact of tax services on taxpayer compliance. Examples include the IRS-sponsored Tax Counseling for the Elderly (TCE) program, which is restricted to taxpayers who are age 60 and older, and the IRS-sponsored Voluntary Income Tax Assistance (VITA) program, which is restricted to taxpayers whose income is \$42,000 or below. Each of these programs provides free tax assistance to eligible participants. For such programs, one might employ a regression discontinuity approach to estimate the impact of the services on compliance. Under this approach, taxpayers just above and just below the age or income threshold are treated as participants in a random experiment where one group has been assigned the “treatment” of being offered the free service, while the other group has been assigned as a control. Separate estimation procedures are employed for each group to generate local estimates of noncompliance. The average difference between the estimates for the treatment and control groups serve as an estimate of the impact of the availability of the free service on taxpayer compliance. One can then divide this estimate by the program participation rate to arrive at an estimate of the average impact on those taxpayers actually receiving the service. The key assumption under this approach is that noncompliance is a continuous function of the running variable (i.e., age or income in our examples; this is also sometimes called the “forcing variable”) in the neighborhood of the threshold, both for individuals receiving the free service and for individuals not receiving the service. Although this is probably a reasonable assumption, it is not a totally innocuous one. For instance, the assumption would be violated if individuals just above the income threshold (or just below the age threshold) tended to understate their income (or overstate their age) in order to qualify for free assistance. This approach could be applied using NRP data. Alternatively, one could use a random sample of (unaudited) tax returns from the general population if one wanted to perform the analysis using a measure of reported taxable income as the outcome variable rather than a measure of noncompliance. In this case, the assumption would be that reported taxable income is a continuous function of the running variable. The results of a regression discontinuity analysis would provide an estimate of the impact of the service on individuals who are close to the eligibility threshold. While such results certainly would be suggestive, there is no guarantee that the results would extend to individuals far away from the threshold.

For the many taxpayer services that are available to the general taxpayer population, it is likely to be more challenging to estimate their impact on tax compliance. First of all, one would like to have access to detailed information about taxpayers who did and did not receive such services. However, for services such as telephone inquiries about the tax laws or IRS web site visits, it seems unlikely that identifying information about the taxpayers is collected. At the micro-level, an alternative approach is to employ a survey to learn about the services used by different taxpayers. The chief difficulty with this approach is that it would be difficult to link the taxpayers’ responses to a reasonable measure of their tax compliance behavior. Although various techniques have been employed in surveys to inquire about tax compliance, it is notoriously difficult to obtain reliable self-reports on this subject. Even at a more aggregated level, it is not obvious how to obtain details about tax reporting or tax compliance behavior for groups of individuals who have received certain specific types of taxpayer assistance.

One possibility would be to develop a data base that linked a survey on service utilization with the tax returns of respondents. There are no doubt some important privacy issues here, but presumably a match between a survey and tax records could be undertaken by a third-party, independent of the IRS – much in the same way that “exact match” data has been obtained for the CPS. I believe such a data base would be an important resource for understanding the compliance impact of services. Ideally the survey would be conducted periodically, perhaps including a longitudinal component.

2.2 Indirect Effects

Even among taxpayers who do not avail themselves of any IRS-provided services (beyond the basic ones provided to all taxpayers, such as processing of refunds), compliance may respond positively to perceptions that the IRS

provides timely, accessible, reliable, and responsive services. The magnitude of this response is likely to depend on psychological factors. In addition, taxpayers who receive IRS assistance may share what they learn with other taxpayers, thereby magnifying the initial compliance impact.

One approach to measuring such indirect effects is to conduct a field experiment similar to the one proposed above for examining the general deterrent effect of tax audits. For example, one might provide improved quality service (such as improved telephone accessibility) in one geographic area while maintaining standard quality service in a comparable area. One would then perform a difference-in-differences analysis of reported taxes in the two areas to assess the impact of improved quality service. Ideally, such an experiment would be supplemented with a survey to examine the extent to which taxpayer perceptions of service quality in the two areas are consistent with the objective levels of quality.

As an alternative to a field experiment, a natural experiment may be possible involving a location where, due to technical or resource limitations, service quality was lower during a tax season. In such a case, one might perform a difference-in-differences analysis of reporting behavior in this and some comparable (unaffected) locations during the affected tax season and other seasons.

Another alternative approach would be to extend the methodology employed by Dubin, Graetz, and Wilde and Plumley to include aggregated measures of service utilization and/or service quality as additional explanatory variables. In addition to the challenge of controlling for confounding factors that was mentioned in the preceding discussion of this method, the introduction of taxpayer service utilization variables would raise additional concerns about potential sample selection bias. In particular, taxpayers are not randomly assigned to receive services, but instead voluntarily choose whether to receive them. It might be difficult to control for the impact of self-selection of services, particularly with aggregated data.

Each of the above approaches treat the mechanism through which services impact on compliance as essentially a “black box”. An alternative approach to this problem would be to attempt to understand what goes on inside of this box. As mentioned previously, the magnitude of indirect effects of services on compliance is likely to depend to a large extent on psychological factors. In particular, a change in the quality, quantity, or type of service offerings is likely to have an influence on taxpayer attitudes and perceptions, which in turn, may impact on their compliance behavior. Therefore, a starting point for learning about this process would be to investigate the extent to which various taxpayer services impact on taxpayer attitudes and perceptions, both among those who use the services and within the general population. It would also be useful to investigate how taxpayer perceptions of service quality (timeliness, accessibility, reliability, responsiveness, etc.) compare to objective measures compiled by the IRS. As with audit risk, it is presumably perceptions more than objective realities that drive behavior. Such issues would be amenable to measurement using a survey approach.

If there are significant indirect effects, one would hope that such an approach would be able to find a positive relationship between taxpayer attitudes and perceived service quality as well as a positive association between changes in perceived service quality and changes in objective measures of service quality (over time or across geographic areas), perhaps with a lagged effect. Such findings would serve as useful “intermediate outcomes” that document a potential pathway through which services can have an indirect effect on compliance. The next step would be to attempt to link taxpayer attitudes and perceptions to actual compliance behavior. There have been some preliminary efforts by past researchers to explore this relationship, but more much more work is needed to obtain convincing estimates. A key issue is whether taxpayers express their true attitudes on a survey or if they instead provide responses that are intended to present themselves in a better light. Further, there is the question whether taxpayers adjust their attitudes after making their tax reports in an attempt to rationalize their behavior. If so, then attitudes are determined endogenously with noncompliance.

2.3 Substitute Services

Clearly, the IRS does not have a monopoly on taxpayer services. Many taxpayers rely on third-party tax practitioners for tax assistance. Others rely on various types of third-party software, third-party tax publications, or Internet sources for tax assistance.



I think it would be useful to update and improve some of the past research that has been done on the impact of tax practitioners on tax compliance; two examples are the study of dual preparer roles undertaken by Klepper, Mazur, and Nagin and the sample selection methodology I used in some of my own past research on the topic.

In addition, it would be useful to perform an analysis of how utilization rates for substitute services respond to changes in the quantity and quality of IRS service offerings, and how the choice among IRS and substitute services is influenced by factors such as tax complexity.

3.0 Payment Compliance

A relatively recent third-party service that has become available to taxpayers is the opportunity to pay taxes due using a credit card. It would be interesting to examine the impact of this service on the overall level of payment compliance.

10. Kevin Fletcher

Introduction

Since 2005/06, HMRC has had a Public Service Agreement to reduce tax losses⁴⁶. For Value-Added Tax and Excise Duties, these have been expressed as targets to reduce the tax gap⁴⁷. With such targets, it is important for HMRC to understand the wider behavioral effects of its activities to ensure optimal allocation of resources to enforcement activity in aggregate, across heads of duty and to particular kinds of operations.

This need is reinforced by important changes to the policy and operational environment in which tax authorities operate. Important recent policy changes include the introduction of more stringent disclosure rules on promoters of tax avoidance schemes, as well as increased potential for retrospective legislation to counter avoidance activity.

In addition to these changes in the policy environment, there have been marked shifts in the way in which HMRC engages with customers. The new ways of working with Large Business implemented following a review by Sir David Varney envisages a more collaborative approach to managing risks among large businesses, involving more work with customers before returns are submitted and a greater focus on the highest priority risks. Further, lighter touch interventions are increasingly used to engage with the whole range of taxpayers, particularly through Targeted, Education, Enabling, and Leverage Activity. Finally, understanding the behavioral effects of enforcement activity is relevant in assessing the effectiveness of specific campaigns such as the introduction of the Offshore Disclosure Facility⁴⁸.

Evaluating deterrent effects

The basic problem in measuring the marginal deterrent effects of compliance activity is that there is seldom sufficient random variation in audit risk faced by [otherwise equivalent] taxpayers to enable estimation of the effect of such variation on the tax gap or, even more modestly, voluntarily declared tax liabilities. Ideally, we would want to identify two identical groups of the population and introduce random variation to the audit rates for each population group with sufficient intensity so that after this random variation has been introduced, a difference-in-difference approach can be utilized to determine the effects of the change in audit risk on marginal income and tax declarations and, thereby deterrent effects. Given the practical difficulty of doing this on such large scale, we often resort to indirect or inferential approaches.

- Social Network Analysis

In recent years, we expanded our ability to link taxpayers to their social networks, as well as to augment their reported information with third-party information. Social networks that could be constructed may include linkages across common company boards of directors, family networks, or tax representatives. In principle, we are able to exploit variation in audit intensity across these social networks to begin to get an understanding of the indirect effects of audit activity targeted at one taxpayer on his or her social network.

As a first pass at this, we may compare changes over time in the declared income and tax liability of members of a social network who have not been subject to an audit if someone else in the network has been audited. For instance, we may compare the change over time in declared income by individuals where someone in their family was subject to a tax audit compared to the change in reported income/tax in the population where they did not have a family member subject to an audit. Of course, this does rely on the strong assumption that the two groups are otherwise identical. A more sophisticated form of this could involve trying to match families in the base year to identify a control and treatment group that would be otherwise identical. This could be attempted across a variety of alternative social networks permitted by the data. The key to success here is going to be the ability within the social network analysis to match the networks appropriately, ensure sufficiently rich information on the taxpayers and their

⁴⁶ Details can be found at <http://www.hmrc.gov.uk/psa/psa2005-2008.htm>.

⁴⁷ The direct tax target was also to reduce tax losses, though was based mainly on measures of extra tax directly recovered each year.

⁴⁸ See box 2.6 in of <http://www.hmrc.gov.uk/budget2008/protecting-tax-revenues.pdf>.

network peers and to identify sufficient numbers of identical networks [one will always have the risk that there is some factor that is not observed that is actually determining the decision to audit and the change in reported income/tax].

- Illicit activity and criminal attack

In respect of efforts to reduce the size of illicit markets in excise goods – tobacco, alcohol products and carbon fuels in particular – the focus requires a further step in translating disruption of activity in the illegal supply of such products to changes in the volume of legitimate consumption. More precisely, we would want to focus efforts in understanding whether changes in enforcement presence in a particular area of illicit supply translates into higher sales in the legitimate market. To understand the localized deterrent effects of enforcement activity, it is possible to consider interrupted time series analysis on sales in the legitimate trade when HMRC activity changes in a particular area, especially for commodities that are relatively insensitive to changes in prices, e.g., tobacco. In principle, successful disruption to the local illegitimate supply should see a statistically significant increase in sales by legitimate retailers in the presence of behavioral shifts. Attempts to do this require large amounts of retail sales data, covering a particular geographic area over sufficiently long periods of time to generate a suitably robust explanation of movements in retail sales against which the effects of revenue authority activity could be detected. Clearly a number of analytical questions arise in having sufficient data and sample sizes to isolate the effects of the change in activity to a change in sales in a particular area, as well as the issue of generalizing local results to the population [especially, as is likely to be the case, if the area that was selected for activity was selected on the basis of expectations of the scale of non-compliance].

- Tracking attitudes

A further way to indirectly track the effects of tax authority activity on deterrent effects is to track attitudes in surveys of taxpayers. We could introduce a set of questions into general household surveys to ask questions related to the perceived levels of non-compliance, the acceptability of such non-compliance as well as the perceived risks of being caught and the consequences of being caught. This would enable the tax authority to infer changes in general attitudes toward compliance with tax obligations, though will remain quite blunt with respect to whether attitude [or changes therein] is actually converted to changes in compliance rates or whether specific policies are more or less successful.

11. Glenn Galfond

I must admit that I am skeptical that the IRS can accurately measure the total impact of its services on voluntary compliance. By “accurately measure” I mean produce estimates of known precision without relying on unproven assumptions. For example, in an ideal research world we could accurately measure the direct impact of IRS services on compliance by creating an experiment where taxpayers are randomly divided into two groups, one of which is denied the benefits of IRS services, and then auditing both groups’ tax returns (or noting when no return was filed) to measure differences in voluntary compliance between the two groups. However, what is ideal in the research world is impossible in the real world.

Since the real world limits our research options, it is tempting to fill gaps by relying on assumptions—for example, that taxpayer behavior follows models from utility theory, behavioral economics or prospect theory. While such models provide a useful framework for understanding behavior, I believe that the IRS should be skeptical of approaches that rely too heavily on unproven assumptions. Instead, the IRS should rely primarily on empirical data collected through surveys and experiments. I have four recommendations.

1. Narrow the Focus

There are many factors that affect the impact of IRS services on compliance. The IRS has multiple service channels, each with different direct and indirect effects. If the IRS didn’t offer these services, some people would find substitute services (e.g., the advice of a paid tax preparer or a friend) and others would go without help. Compliance itself has many dimensions, including filing, reporting, and payment compliance. Rather than trying to estimate the impact of IRS services on compliance considering all of these factors, I recommend the IRS begin by conducting focused surveys and experiments to identify the factors that are the most important to study in more detail.

One example of how the IRS could narrow the scope of the problem is a past study conducted by Price Waterhouse for the IRS. There are other examples in the literature, but I mention this one because I was involved, so I know it best. That study was a laboratory test of the impact of Taxpayer Assistance on compliance. In the study, taxpayers with demographic characteristics similar to users of IRS Taxpayer Assistance were recruited to fill out hypothetical tax returns typical of such users. Some of the taxpayers were randomly assigned to study groups that could use IRS Taxpayer Assistance, and some were assigned to a control group that had no assistance. The compliance of both groups on the hypothetical returns was compared to determine the impact Taxpayer Assistance on compliance in the test. While the impact observed in the test may not be an accurate measurement of the impact of Taxpayer Assistance in the real world, it does help focus future research. For example, if such a test found little or no impact on compliance, further research on the impact of Taxpayer Assistance would be given a lower priority or eliminated from consideration entirely. By conducting surveys and experiments on various components of the problem, the IRS can make more informed decisions about which factors (service channel, direct vs. indirect effect, type of compliance, etc.) to study in more detail.

2. Extend the National Research Program (NRP) to Address Some of the Effects of IRS Services

The NRP provides the IRS’s best measurement of the accuracy of filed tax returns. Extending the NRP through follow-up surveys that gather information on the direct and indirect use of IRS services by NRP taxpayers would enable the IRS to correlate the use of IRS services with the accuracy of filed returns. This approach would not address the impact of substitution (e.g., that some taxpayers might use a paid tax preparer if IRS services were unavailable) or the impact of IRS services on compliance with filing requirements (since NRP only studies filed returns). In addition, observed differences in the compliance of users and non-users of IRS services may in part be due to differences between the populations, and thus cannot entirely be attributed to IRS services. Nonetheless, such a study would address very important components of the problem.

3. Better Leverage Limited Research Funding by Coordinating this Research with Other Research Activities

The prior recommendation related to the NRP is one example of how the IRS can cost-effectively extend existing research to address multiple objectives. The IRS should look for other ways to coordinate its research activities across initiatives in order to make research more cost-effective. For example, existing surveys of customer satisfaction with IRS services or surveys related to taxpayer burden could easily be extended to ask questions related to the impact of IRS services on compliance. In addition, it is likely that research studies focused on the impact of



IRS services could be extended to address other areas such as customer satisfaction and burden. I cannot suggest specific examples without more knowledge of the IRS's research agenda, but I expect many exist.

4. Evaluate Benefits Beyond Compliance (if not already being done)

Improved compliance is not the only benefit of IRS services. For example, IRS services reduce taxpayer burden and improve taxpayer satisfaction. If the IRS is not already doing so, it should perform research to quantify all of the major benefits of IRS services to more fully inform decisions on which services to provide. This quantification can best be achieved through surveys and experiments.

12. Jim Harvey and Kevin McCrohan

The research question for the IRS is how to measure and model the impact of taxpayer services and enforcement activities on the voluntary compliance of taxpayers—both those who have interacted directly with the IRS and those in the general population who have not. We believe there is a measure of taxpayer compliance due to enforcement as well as one for service. An intriguing aspect of voluntary compliance is that the overwhelming number of studies does not explain why compliance is as high as it appears to be in the US. One reason for this could be the treatment of taxpayers as a homogeneous population or as rational decision makers. We would suggest that the market for voluntary compliance is heterogeneous – with significant subpopulations that when addressed can raise overall US compliance with tax obligations. For example, the annual news reports of tax filer hysteria hours prior to the filing deadline provide some insight into the attitudes, demographics and behaviors of one of these significant groups.

Market Heterogeneity

Marketers have long known that heterogeneity exists in most markets and that even for the most similar appearing products, expensive wrist watches or soft drinks for example; there are real differences among consumer segments. In the \$3,000 and above wrist watch market Rolex and Baume and Mercier consumers tend to have different perceptions of themselves, view different media, and have somewhat different behavioral and socio-economic profiles. It would seem that if consumers do not behave in a homogeneous manner than there should not be an expectation that all taxpayers would.

In 1982 the IRS commissioned Yankelovich and White to conduct a survey of taxpayer attitudes. The results of their survey identified five taxpayer segments that included those wishing to pay but confused by the regulations to those dubbed “strategic non-compliers,” taxpayers that enjoyed jousting with the IRS (Skelly 1984).

Studies conducted on the IRS Informal Supplier Studies data sets by on the nature of the consumers in the underground economy pointed segmentation among consumers of informal service. McCrohan, Smith, and Adams (1991) suggested that more fortunate consumers benefit from its existence. Lemieux, Fortin and Frechette (1994) reported that while the earnings in the underground sector were concentrated among workers with low income in the regular sector, purchases from providers in the underground sector were concentrated among people with high incomes in the regular sector. Following their paper McCrohan and Sugrue (1998) reexamined the IRS Informal Supplier Studies data sets and identified four consumer and supplier segments in informal markets.

Participation in informal markets was defined as the ratio of all expenditures (both formal and informal for the fourteen categories of goods and services studied) to informal expenditures for the same set of purchases. This ratio was found to range from a low of eight percent among older and less educated consumers to a high of twenty-three percent among the most upscale consumers. Of interest for market segmentation among taxpayers was that consumers who demonstrated the highest ratios of expenditures with informal suppliers were found to be characterized by having the highest income per person, highest education, and work the most hours. It is in this segment that the most barter activity was found.

These findings support the earlier work by Clinard (1952) and Tanzi and Shone (1993) that questioned the involvement of consumers with suppliers that were operating out of the taxation or measurement stream. It also supports the work of Lemieux, Fortin and Frechette (1994) which found that purchases in the underground economy were concentrated among higher income populations.

Key Principles for Increased Voluntary Compliance

A comparison of approaches taken to increase voluntary compliance within both the philanthropic and tax compliance literature postulated by Harvey and McCrohan (1988a) found that Americans were more likely to act voluntarily when:

- Individuals believe the operations of the organization are efficient
- They think the social, political and economic goals of the organization are legitimate
- Their norms of voluntary compliance are reinforced and rewarded
- The benefit to the individual and family is clear
- Training is available for the desired behavior

- Relevant subgroups are identified
 - Economic well-being
 - Higher education
 - Older
- Communications are appropriate

These were found to raise compliance for philanthropic organizations (Harvey and McCrohan 1988b). In comparing voluntary compliance between the taxation and philanthropy literature there were numerous similarities with the exception of economic well-being and higher education. However, in the case of higher education and income being linked to higher charity giving and higher tax noncompliance the seeming contradiction can be explained by sacrifice, donation in relation to the income. In this case percent goes down as income and education goes up with the highest “sacrifice” (percent) levels found among lower income and education groups.

A Program for Tax Compliance Behavior

In developing a program for the IRS, a valuable perspective is to view tax compliance as a social issue behavior. Such behaviors are individual actions connected to matters of overall public well being, including early childhood immunization, smoking secession and responsible alcohol use. These behaviors are related to social issues because the outcomes of choices are connected to significant social costs that are borne by the public. For example, the current estimate of Gross tax gap was estimated to be \$345 billion in 2001. After enforcement efforts and late payments, this amount was reduced to a net tax gap of approximately \$290 billion (IRS 2007).

A field that provides insight into behavior management with significant social costs is Social Marketing - a discipline that uses traditional marketing methods for changing behaviors for social benefit, public healthcare and safety. Typical Social Marketing investigations include seatbelt use, smoking secession, immunization, responsible sex, unhealthy diet, blood pressure monitoring and excessive drinking. Each has high public costs stemming from individual noncompliance.

An example of the effectiveness of Social Marketing in changing social issue behaviors is the development of the Road Crew initiative (<http://www.roadcrewonline.org>) in Wisconsin. During the five year startup of this program (2002 – 2007), over 85,000 rides for inebriated drivers were provided that avoided an estimated 140 alcohol related crashes and six alcohol related fatalities. The estimated cost savings of the program to the state is \$31 million.

Background for the success of the Road Crew success is found in Rothschild’s (1999) discussion of the interrelated roles of law, education and marketing in managing social issue behaviors as they apply to targets that may or may not have any motivation, opportunity, and/or ability to cooperate but that nevertheless have been selected by management.

In that paper Rothschild defines

- Education as “messages of any type that attempt to inform and/or persuade a target to behave voluntarily in a particular manner but do not provide, on their own, direct and/or immediate reward or punishment,”
- Marketing as “attempts to manage behavior by offering reinforcing incentives and/or consequences in an environment that invites voluntary exchange” and
- Law as “the use of coercion to achieve behavior in a non-voluntary manner or to threaten with punishment for noncompliance or inappropriate behavior.”

Rothschild argues that the key drivers for understanding the relative effectiveness of Education, Marketing and Law are based on the character of three constructs of the individual:

- Self Interest – Change is more likely to change when the individual can discern immediate self interest in the behavior and if there is no perceived benefit, it is less likely to occur. If the behavior change is perceived to be minor, it is more likely to occur, whereas major change is less likely to occur.
- Exchange – The specificity and timing of the payback in the transaction. With social issue behaviors the payback is often vague, uncertain, and in the distant future.

- Power and Competition – The level of apathy and involvement individuals perceive about the behavior. Competition stems from the choices individuals have such as not filing taxes or report misrepresentation. In a free choice society, many laws are not followed if the target cannot discern the reward in doing so.

Rothschild offers insight into the best use of Education, Marketing and Law by examining whether (a) the individual is prone, resistant or unable to accept the advocated behavior and (b) their level of motivation, their opportunity and their ability to comply.

In that work Rothschild concludes that

- The best role for Education is for individuals who are prone to behave (are motivated - positive self interest, have an opportunity – there is a mechanism at hand, and have the ability – skill and proficiency),
- The clearest application of Marketing is for individuals who are unable to behave (are motivated, have the ability, but no opportunity), and
- The most appropriate place for Law is when individuals are resistant to behave (have the opportunity, have the ability, but no motivation).
- His analysis suggests that understanding individual self-interest, the opportunity to comply and their skills in complying will lead to appropriate application of Education, Marketing and Law.

Elements of self interest include perceptions of tax law fairness and whether taxes advance society, the likelihood of an audit, and the size of the cost and benefit of compliance. Opportunity can be viewed as the ease of availability of tax forms, assistance in answering questions, software for self-prepared filing, and availability of commercial preparation services. Compliance skills include ease of interpretation of forms, record-keeping discipline, and functional literacy (reading, comprehension and calculation).

The Role of Service Quality in Citizen Satisfaction

Customer satisfaction is a key metric watched by most organizations (<http://www.theacsi.org/index.php>). Recent news reports focused on the high customer satisfaction ratings of Google and Apple Computers (http://www.theacsi.org/index.php?option=com_content&task=view&id=15&Itemid=188).

Overall American Customer Satisfaction Index (ACSI) scores 1999 - 2007 for Federal Agencies are found at http://www.theacsi.org/index.php?option=com_content&task=view&id=123&Itemid=134. Among the 2007 findings for Federal Agencies were significant differences between and within agencies. The Pension Benefit Guaranty Corporation (Retirees) was reported to have high satisfaction among Retirees (88) but fared less well for other constituencies, such as Premium Filers.

Large & Midsize Business Corporate tax filers rated the Internal Revenue Service low (52) but satisfaction with the IRS for Individual electronic tax filers was higher than seven other Federal Agencies with different constituencies: NASA (Data & Information Systems); Pension Benefit Guarantee Corporation (Premium filers); Corporation for National and Community Service (Grantees); Army Corps of Engineers, Defense (Recreational visitors); Pension Benefit Guarantee Corporation (Premium filers); Consular Affairs, State (Recent passport applicants/renewals); and Federal Aviation Administration, Transportation (Commercial pilots and Aviation mechanics).

Drivers of Customer Satisfaction

While customer satisfaction is a useful outcome tracking measure, the drivers of this metric are not defined and may vary industry to industry. Valuable insight into the causes of customer satisfaction for the IRS is the literature of service quality (Parasuraman, et al 1985). This team of researchers developed the most widely accepted scale for measuring service quality, SERVQUAL.

Since higher service quality is linked to greater customer satisfaction, the components of service quality provide good insight into managing satisfaction with service. The dimensions of service quality are widely documented. A convenient summary of service quality is found at <http://ils.unc.edu/daniel/131/servicequality.html> and is reproduced below. *



Quality Dimension	Samples of questions to ask
Tangibles: Appearance of physical facilities, equipment, personnel, printed and visual materials	<ul style="list-style-type: none"> • Are facilities attractive? • Are staff dressed appropriately? • Are written materials easy to understand? • Does technology look modern?
Reliability: Ability to perform promised service dependably and accurately	<ul style="list-style-type: none"> • If a response is promised in a certain time, does it happen? • Are exact specifications of client followed? • Are statements or reports free of error? • Is service performed right the first time? • Is level of service same at all times of day and for all members of staff?
Responsiveness: Willingness to help customers to provide prompt service	<ul style="list-style-type: none"> • When there is a problem, does organization respond to it quickly? • Are staff willing to answer client questions? • Are specific times for service accomplishments given to client? • Are public situations treated with care and seriousness?
Competence: Possession of required skill and knowledge to perform service	<ul style="list-style-type: none"> • Can staff provide service without fumbling around? • Are materials provided appropriate and up to date? • Can staff use the technology quickly and skillfully? • Does staff appear to know what they are doing?
Courtesy: Politeness, respect, consideration and friendliness of contact personnel	<ul style="list-style-type: none"> • Does staff member have a pleasant demeanor? • Does staff refrain from acting busy or being rude when clients ask questions? • Are those who answer the telephone considerate and polite? • Do staff observe consideration of the property and values of clients?
Credibility: Trustworthiness, believability, honesty of the service provider	<ul style="list-style-type: none"> • Does service organization have a good reputation? • Do staff members refrain from pressuring the client? • Are responses given accurate and consistent with other reliable sources? • Does the organization guarantee its services?
Security: Freedom from danger, risk, or doubt	<ul style="list-style-type: none"> • Is it safe to enter the premises and to use the equipment? • Are documents and other information provided for the client held securely? • Are use records of clients safe from unauthorized use? • Can client be confident that service provided was done correctly?
Access: Approachability and ease of contact.	<ul style="list-style-type: none"> • How easy is it to talk to knowledgeable staff member when client has a problem? • Is it easy to reach the appropriate staff person <ul style="list-style-type: none"> – in person? – by telephone? – by email? • Are service access points conveniently located?
Communication: Listening to customers and acknowledging their comments; Keeping customers informed in a language they can understand.	<ul style="list-style-type: none"> • When client contacts service point, will staff person listen to their problem and demonstrate understanding and concern? • Can staff explain clearly the various options available to a particular query? • Do staff avoid using technical jargon when speaking with clients? • Does staff member call if a scheduled appointment will be missed?
Understanding the Customer: Making the effort to know customers and their needs.	<ul style="list-style-type: none"> • Does someone on staff recognize each regular client and address them by name? • Do staff try to determine what client's specific objectives are? • Is level of service and cost of service consistent with what client requires and can afford? • Are service providers flexible enough to accommodate to client's schedule?

*Adapted from SERVQUAL, an instrument for measuring quality service developed by Zeithaml, et al, 1990.

Other convenient sources for insight into service quality are provided by

<http://www.slideshare.net/bob.hogg/services-marketing-the-servqual-model/>

http://www.12manage.com/methods_zeithaml_servqual.html

<http://www.businessadvantagcuk.biz/SERVQUAL.htm>

<http://www.istheory.yorku.ca/SERVQUAL.htm>

and for government and taxation, see

<http://www.waset.org/pwaset/v28/v28-56.pdf>

<http://www.acis2007.usq.edu.au/assets/papers/117.pdf> (to launch this link must be copied and pasted into a new browser)

<http://www.emeraldinsight.com/Insight/viewContentItem.do;jsessionid=9E5CA9A9947C6E3E65BEE6D16981D03D?contentType=Article&contentId=867721>

<http://www.ejeg.com/volume-5/vol5-iss2/Connolly.pdf>

Conclusion

In this piece we have summarized previous works that collectively offer insight into developing a model of elevated tax compliance. Key elements of these views include

- Viewing noncompliance as a heterogeneous phenomena that can benefit from developing targeted service and enforcement strategies,
- Identifying sample differences in demographic, attitudinal and lifestyles that represent high probability target for IRS program development,
- The value of viewing tax compliance as a social issue behavior and the role of Social Marketing - a discipline that uses traditional marketing methods for changing behaviors for social benefit, public healthcare and safety,
- A brief summary of the complimentary roles of Education, Marketing and Law in developing a program of tax compliance,
- The role of service quality in customer satisfaction,
- The drivers of service quality that cause customer satisfaction to vary, and
- Samples of factors to monitor in managing service quality.



References

- Clinard, Marshall B (1952), *The Black Market*, Rinehart & Company, Inc.: New York.
- Harvey, James W. & McCrohan, Kevin F. (1988a). Is There a Better Way of Improving Compliance with Taxation: Insights from the Philanthropic Literature. *Journal of Public Policy and Marketing*, (7), 138-151.
- Harvey, James W. & McCrohan, Kevin F. (1988b). Voluntary Compliance and the Effectiveness of Public and Nonprofit Institutions: American Philanthropy and Taxation. *Journal of Economic Psychology*, 9, 369-386.
- IRS, Reducing the Federal Tax Gap: A Report on Improving Voluntary Compliance. Internal Revenue Service, U.S. Department of the Treasury, August 2, 2007, http://www.irs.gov/pub/irs-news/tax_gap_report_final_080207_linked.pdf.
- Lemieux, Thomas, Bernard Fortin, and Pierre Frechette (1994), "The Effect of Taxes on Labour Supply in the Underground Economy," *The American Economic Review*, 84 (March), 231-254.
- McCrohan, Kevin F. and Timothy Sugrue (1998), "An Empirical Analysis of Informal Markets Participation," *Social Science Quarterly*, 79, 1, (March) 212-226.
- McCrohan, Kevin F., James D. Smith, and Terry K. Adams (1991), "Consumer Purchases in Informal Markets: Estimates for the 1980s, Prospects for the 1990s," *Journal of Retailing*, 67 (Spring), 22-50.
- Parasuraman, A., V. A. Zeithaml and Leonard L. Berry, "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing* 49 (Fall 1985): 41 – 50).
- Rothschild, Michael L. (1999), Carrots, Sticks, and Promises: A Conceptual Framework for the Management of Public Health and Social Issue Behaviors *Journal of Marketing*, Vol. 63 (October), 24-37.
- Skelly, Florence (1984), 1984 General Purpose Taxpayer Opinion Survey, Stanford, CT: Yankelovich, Skelly & White, Inc.
- Tanzi, Vito and Parthasarathi Shome (1993), "A Primer on Tax Evasion," *IMF Staff Papers*, 40, (December), pp. 807-828.
- Zeithaml, Valarie, A. Parasuraman, and Leonard Berry (1990), *Delivering Quality Service; Balancing Customer Perceptions and Expectations*, Free Press, 1990.

13. Erich Kirchler

Methods for estimating the impact of services

1. Enforced versus voluntary tax compliance: the “slippery slope” framework

The “slippery slope” framework (Kirchler, Hoelzl & Wahl, 2008) starts from the idea that the tax climate in a society can vary on a continuum between an antagonistic climate and a synergistic climate. In an antagonistic climate, taxpayers and tax authorities work against each other (“cops and robbers” attitude; high social distance between authorities and taxpayers); in a synergistic climate, they work together (“service and client” attitude, close social distance).

The framework distinguishes between **enforced compliance** and **voluntary compliance**, and proceeds with the idea to think about tax compliance along two major dimensions: the (coercive) **power of tax authorities** to enforce compliance and **trust in tax authorities**. These dimensions and their interactions jointly influence the type and level of tax compliance.

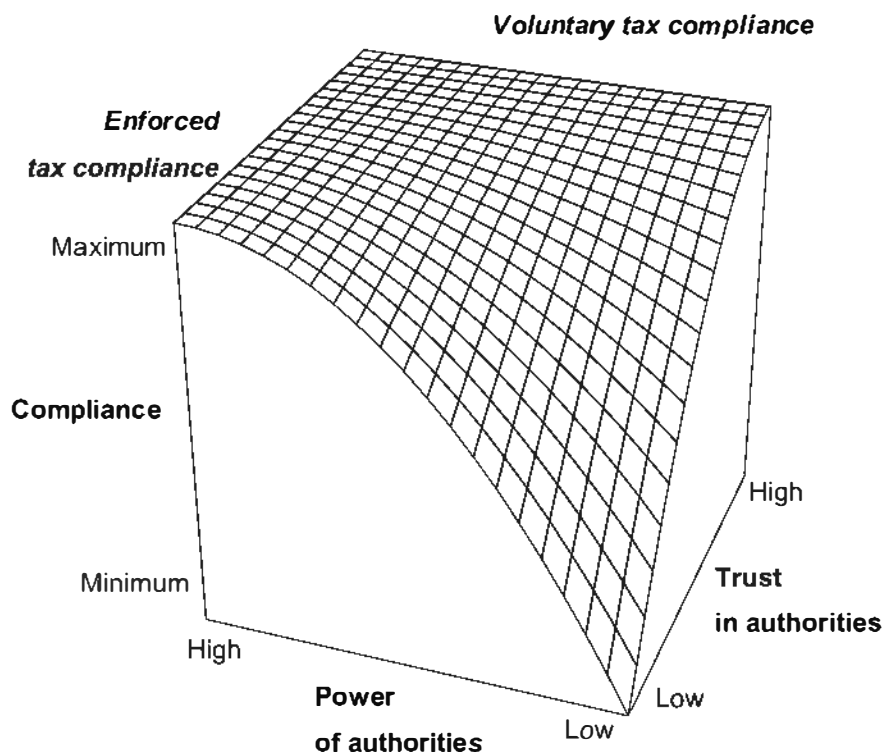
Two different ways of how authorities could gain cooperation from the public are distinguished: the first way claims that the threat of **audits and punishment** can enforce compliance (economic approach). The second way claims that perceived competence in managing problems can activate citizens to aid the authorities and to feel obliged to adhere to decisions, policies, and rules (psychological approach). Trust in authorities depends on the following variables:

- (a) subjective tax **knowledge**;
- (b) **attitudes** towards tax authorities and the government;
- (c) personal, social, and national **norms**;
- (d) perceived **fairness**:
 - distributive fairness (horizontal, vertical, exchange fairness);
 - procedural fairness (fairness of procedures, i.e. neutrality, transparency, etc., and interactional justice with two aspects of interpersonal treatment: i) interpersonal justice, reflecting the degree to which people are treated with politeness, dignity, and respect by authorities. ii) informational justice, focusing on the explanations provided to people that convey information about why procedures were used in a certain way or why outcomes were distributed in a certain fashion);
 - retributive / restorative fairness (restorative justice focuses on crime as an act against another individual or community, and the victim should receive some type of restitution from the offender).

Perceived fairness is connected to the trust dimension because a just treatment of taxpayers (i.e., distributive, procedural fairness, retributive fairness) helps to build and maintain trust. Retributive justice is connected to the power dimension as well, because it depends also on detecting and fining wrongdoers. In turn, an inconsiderate exertion of power that is perceived as intrusive can reduce trust.

The responsive regulation approach (Braithwaite, 2007) fits well with the current framework. It proposes regulatory rules and suggests that the authorities should act responding to the beliefs and attitudes of the taxpayers. These are captured in the concept of “motivational postures”, defined as “an interconnected set of beliefs and attitudes that are consciously held and openly shared with others”. Motivational postures are commitment, capitulation, resistance, disengagement, and game playing.

Figure 12: The “Slippery Slope” Framework



2. Methods to investigate the impact of services

Before selecting methods to investigate the impact of services, a theoretical framework of determinants of tax compliance should be developed or an existing framework might be used, e.g., the “slippery slope” framework. This is necessary to define dependent and independent variables as well as situational circumstances which might have moderator effects in cause-effect relationships.

From a psychological perspective, the following questions appear relevant for the development and selection of research methods:

- (a) **Service criteria:** What are the variables on which services should have an impact? Enforced compliance, voluntary compliance, perceived power of authorities, trust in authorities, perceived procedural justice, satisfaction with tax authorities and tax system, etc.
- (b) **Definition of service:** How is service defined? Information and support provided to taxpayers and tax practitioners by tax officials, internet, brochures, etc. How are taxpayers treated at tax offices and during audits; provision of public goods.

After clarifying these questions, the following methods can be used:

- (a) **Surveys:** What are tax officials’ and taxpayers’ expectations about services and perceptions of actual services?
 - Survey techniques: Tax officials and taxpayers should respond to questionnaires focusing on their knowledge about services, expectations, evaluation and satisfaction with services.

- Tax officials and taxpayers should not only indicate their own position but also put themselves in the position of the other party and indicate how much knowledge etc the others are thought to have. This procedure allows to measuring the discrepancy of perceptions of interaction parties.
 - For existing services, it seems relevant to examine knowledge of these services, but also their interpretation by the taxpayers. Interpretation might differ according to motivational postures.
 - Besides measuring knowledge, etc. about services, also perceptions of social norms, fairness, and compliance data should be collected in order to assess the impact of service perception on compliance. According to the “slippery slope” framework, voluntary and enforced compliance, trust and power need to be measured. In a synergistic climate, perceived service quality should have an impact of perceived procedural fairness, on trust and voluntary compliance. With regard to motivational postures, commitment should prevail. In an antagonistic climate the correlations should be low; if power of authorities is high, compliance should be enforced rather than voluntary; capitulation as a motivational posture should prevail.
 - The perspective of tax preparers could be helpful to understand the prevalent motives of taxpayers and their view of services.
- (b) Systematic **collection of complaints** by taxpayers and content analysis.
- (c) **Focus groups** (tax officials and taxpayers) discussing quality of service and necessity of improvement.
- (d) **“Mystery shopping”** (tool ranging from simple questionnaires to complete audio and video recordings, used by market research companies to measure quality of retail service or to gather specific information about products and services. Mystery shoppers posing as normal customers perform specific tasks - such as purchasing a product, asking questions, registering complaints or behaving in a certain way - and then provide detailed reports or feedback about service.
- (e) **Quasi-experimental approaches** using comparable districts where a new service is introduced in one district and the current service is continued in the other district, and potential changes in trust etc. are measured.
- (f) Experimental methods to investigate specific theoretical cause-effect relations.

To conclude, a stringent theoretical framework should be employed to organize and to integrate methods and results.

References

- Braithwaite, V. (2007). Responsive regulation and taxation: Introduction. Law & Policy, 29(1), 3-11. (Special issue on responsive regulation and taxation).*
- Kirchler, E., Hölzl, E. & Wahl. (2008). Enforced versus voluntary compliance: the “slippery slope” framework. Journal of Economic Psychology, 29(2), 210-225.*



14. Paul Lehner

To begin with let me note that it is a delight to see such serious efforts to investigate this hard measurement problem. In many of my other endeavors I'm often struck by the extent to which "experience and expertise" are relied upon in the absence, and often in contradiction to, hard scientific data on "what works". I strongly support the general tenor of this effort.

My comments below are limited to a few thoughts on how to approach this measurement problem a little differently.

From a policy maker's perspective, determining tax policy is in part a forecasting problem. While the use of statistical regression models is an empirically-justified practice for many forecasting problems, it is worth noting that there is an emerging literature on predicting decisions in conflict situations. The following table, which summarizes data from a couple of studies, shows the accuracy of different "methods" for predicting the decisions that were made in various historical conflict situations (where the context was redacted, but details verified with individuals knowledgeable of those cases). Forecasts based on game theory, expert judgment, "briefly briefed" undergraduates all did no better than chance. By contrast "briefly briefed" undergraduates who engaged in some role playing did markedly better. (Check www.conflictforecasting.com for these and related studies.)

	Chance	Novices	Experts (domain)	Experts (game theory)	Role playing w/ novices
Artist Protest	17%	5%	10%	6%	29%
Distribution Channel	33%	10%	38%	23%	75%
55% Pay Plan	25%	13%	18%	29%	60%
Telco Takeover	25%	20%	0%	0%	40%
Personal Grievance	25%	22%	31%	43%	60%
Zenith Investment	33%	33%	36%	22%	59%
Water Dispute	33%	48%	50%	75%	90%
Nurses Dispute	33%	66%	73%	50%	82%
Averages (unweighted)	28%	27%	32%	31%	62%
Total number forecasts		523	106	101	105

Results such as these indicate that there may be innovative methods that can reasonably forecast individual decisions in specific situations, but that good research is needed to discover the typically counterintuitive methods that do work.

Another approach that is relevant is the research on models for predicting individual judgments. There is a substantial literature that indicates that individual judgments' including expert judgments (e.g. medical diagnosis), can be reliably predicted with simple linear models containing a small number of factors. Although intuition and introspection suggests that our judgments are complex and nuanced, our behavior consistently suggests otherwise. To build these predictive models researchers routinely use a method where the people make hypothetical judgments against descriptions of real or hypothetical cases and then use that data to build a predictive model for each individual's judgments.

It might be possible to build predictive models of likely tax compliance decisions from research with individuals projecting "what they would do" in various hypothetical cases. If there is sufficient diversity in the subject population, then it might also be possible to build profiles of the principal drivers of tax compliance decisions within different subpopulations – if indeed there is commonality among different taxpayers.



Both of the approaches described above emphasize the development of methods that predict individual taxpayer decisions in specific settings. Whatever the method, the accuracy of these forecasts should be tested against a corpus of real cases drawn from IRS archives and random audits.

Overall I'm suggesting a three stage research program. First, engage the research community to develop alternative methods for forecasting individual taxpayer decisions in specific contexts. Second, develop an extensive corpus of individual taxpayer decision cases that could be used to test the predictive accuracy of these methods. Third once a diversity of vetted methods are developed then those methods could be used to support forecasting of the likely impact of various changes in tax policy, including those related to tax services. And finally, the predicted impact of changes in tax policy could again be tested against new cases in the corpus of taxpayer decisions.

15. Shelley Metzenbaum

- The first challenge for measurement is clarifying the objective of service/compliance assistance (CA). Possible objectives include:
 - Reduced percentage and number of non-filers;
 - Fewer intentional errors;
 - Fewer unintentional errors;
 - Increased awareness of compliance obligations;
 - Increased commitment/sense of obligation to comply;
 - Increased understanding of specific compliance obligations;
 - Reduced burden of compliance
 - Actual
 - Perceived

Other objectives may also exist. To design effective service/CA programs and choose useful measurement methods, it is essential to identify clearly what the service objectives are.

- Measurement systems can have distinct purposes, as well. The same measurement can sometimes serve multiple purposes, but different purposes often require differences in measurement methods. Possible purposes include:
 - Diagnostic or operational measurement that management uses to find:
 - Programs that work worth replicating;
 - Programs that don't work and need attention;
 - Relative risk (consequence, frequency) of problems to inform priority-setting.
 - Reporting progress and problems to Congress and the public
 - Evaluation to find factor that affect outcomes and assess their relative import (coefficient of independent variables)

To choose useful measurement methods, it is essential to think about and design for all the relevant measurement uses. In most situations, it is useful to serve all three measurement uses.

- Congress would like to know that compliance is high and improving (reporting progress using outcome measures.) It may also want to know how much compliance assistance affects compliance relative to enforcement and other possible influencers (evaluation using outcome measures as the dependent variable and possible causal factors as the independent variables), although that is not a useful way to think about resource allocation or running compliance assurance programs since compliance assistance and enforcement work best when used together as part of a graduated continuum of responses.
- To increase compliance levels and reduce compliance filing errors, the IRS needs more frequent feedback to find what most likely is working and what probably is not.
- To use the data it gathers that way, the IRS needs tagged data that can be sliced and diced to look for patterns, similarities, differences, and anomalies.
- Individuals and organizations have different attitudes, awareness, and understanding about compliance obligations. They are therefore likely to respond differently to compliance assistance. CA programs may need to be tailored to specific audiences, and measured separately.
- Measuring all aspects of CA at one time is unlikely to be affordable, so decisions need to be made about taxpayer segments and specific objectives before figuring out how to measure impacts.
- Controlled field studies are a likely to be practical way to assess CA program impacts
 - Consider grants to states to participate in controlled studies
 - For on-line CA, try randomized trials a la Monster.com, as described in Ayres *Supercrunchers*
 - Are there measurable variations for any of the outcomes of interest that differ by third-party assisters, serving as *de facto* controlled experiments?
- Surveys are likely to be helpful for assessing errors, commitment, understanding, and perceived burden, including which questions/compliance obligations are well understood and which are not.
- Analysis of patterns, variations, etc. of questions and service demand, e.g., looking at sections of forms prompting most questions and errors, types of assistance requested, timing of request, likely to be very useful for designing better CA programs.



- If some questions/compliance obligations have less accurate compliance than others, target CA programs to fix those problems and measure change in understanding and compliance for specific obligations.
- Understanding market segments will be helpful. Analyses of existing data bases to look for variations and similarities by type of filer should be useful.
- Do audits ask about assistance? Would that be possible, and tag records accordingly?
- Avoid temptation to report enforcement and assistance impacts separately under GPRA or PART. Will cause or sustain artificial separation in practice.
- Finally, let me recommend a movie. It's not at your local theater, but at the Spy Museum. Just a short clip from World War II urging citizens to do their duty and pay their taxes early. It will stretch your thinking about what CA or IRS service might include – working with Hollywood to build the public's sense of duty to pay taxes. Wow!



16. Dan Nagin

I have not done research on tax compliance for nearly 20 years. Thus, having been away from the field for so long I do not have a good basis for making useful recommendations for tax compliance research particularly as it relates to services. However, in reading over the literature review prepared by IRS staff, one research hypothesis stood out in my mind as suggesting a potentially effective tactic for improving tax compliance in a relatively painless way both in terms of economic cost and public sentiment. That hypothesis involved the question of whether being in a balance due versus refund status affects compliance. Kahneman and Tversky's prospect theory predicts that due to loss aversion the balance due condition will produce more tax noncompliance. This hypothesis dates back to when I was doing compliance research but was never tested. Seems it remains untested (but at least Kahneman has a Nobel Prize under his belt for prospect theory among other things.)

I recommend that a randomized experiment be conducted to test whether (1) absent a statutory change requiring increased withholding, taxpayer withholding levels can be influenced by a message describing the cash flow advantages of not having to pay a large balance due on April 15 and (2) whether balance due status affects reporting behavior. As an aside this experiment also provides a nice opportunity to apply recently developed methods designed to measure treatment effects among compliers. In this setting a complier is any person whether in treatment or control who has a refund due. In designing this experiment it would be useful to block taxpayers by return characteristics in prior filings. I say this because the response to having a balance due will likely depend upon a taxpayer's opportunities for noncompliance. Thus, I think it important at a minimum to block on whether the taxpayer had previously filed a long or short form. Further blocking on long form characteristics would probably also be desirable (e.g., reports business income or not).

17. Alan Plumley

The Problem

I would prefer to have a comprehensive and accurate model of taxpayer behavior (incorporating all of the factors that influence that behavior), as well as years of data on each of those factors for each taxpayer, or at least for each taxpayer in a representative sample. Under those conditions, quantifying the impact that changes in IRS actions have on taxpayer compliance would be relatively straightforward. Unfortunately, I do not believe that we will ever have comprehensive data, let alone a comprehensive model—especially at the micro level. Surely, the number of potential determinants of taxpayer behavior is virtually endless, most of them are subjective perceptions and motivations that the taxpayers themselves do not fully comprehend (and could not explain correctly even if they wanted to), the interactions between these factors are extremely complex, and the weights that each taxpayer places on them may be unique to him. The observable drivers of those determinants (such as IRS service and enforcement actions) are surely indirect influences; they help to shape the ultimate determinants of taxpayer behavior, which are generally not observable.

The complexity of that network of factors and the inevitable lack of data on ultimate perceptions and motivations, however, do not mean that we cannot successfully quantify the independent impact of most IRS activities. Rather, they mean that we need to be quite modest about what we say the mechanisms are that link IRS actions to taxpayer behavior. Indeed, we're probably better off not assuming much at all about the nature of those mechanisms. For example, a single IRS activity such as an audit could affect some taxpayers through a deterrence mechanism (if it changes their perception of detection and penalty), others through an assurance mechanism (if it assures them that scofflaws are getting punished, or that IRS audits actually help taxpayers to understand the complicated tax law), others through a burden mechanism (if it is merely a drain on their time, money, and patience), and it could affect others not at all. Understanding the ultimate internal drivers of compliance behavior and the mechanisms that trigger these drivers would be extremely helpful to the IRS, but complete knowledge of them is not necessary for the purpose of quantifying the link between IRS actions and taxpayer behavior.

My Recommendation

I believe that the most promising approach to estimating the form and magnitude of those relationships is standard regression analysis applied to aggregate panel data. I and others have successfully used this approach to estimate the impact of several IRS activities. The main criticism of these efforts has been that the estimated results are sometimes implausibly large.

That may be true, but if it is, it could easily be due to significant limitations in the data that could potentially be overcome. I believe that not modeling one or more specific theories as to the mechanisms that link IRS actions to taxpayer behavior is actually a strength of this approach, not a weakness. Microanalyses that have attempted to model potential compliance mechanisms have inevitably oversimplified those mechanisms (often using aggregate variables anyway), and have not recognized that many divergent mechanisms are undoubtedly at play simultaneously. Moreover, from the perspective of IRS resource allocation decisions, taxpayer compliance is an aggregate phenomenon. That is, we are not interested in the impact of Joe's audit on Joe's nextdoor neighbor; we are interested instead in the impact of groups of similar audits on multiplied thousands of other taxpayers in the aggregate. Applying the results of micro models to the population in the aggregate is no better than applying the results of aggregate models to individuals; if we want to apply our results to the population as a whole, then we're better off estimating the relationships in the aggregate to begin with.

The unit of observation in such an analysis would be some subset of the population (e.g., all taxpayers who reside in a certain geographic area, such as a state) and a specific tax (or fiscal) year. It would be helpful to divide the population into many subsets, but if the group is too small (e.g., even 3-digit ZIP code areas), then the implicit assumption that each unit of observation is independent from the others would almost certainly be violated significantly. Starting with a common definition of the unit of observation would allow aggregate data to be compiled on the same basis across all IRS activities each year. (This is mostly possible for enforcement activities already, but it would probably take significant work to do this for service activities.) The panel nature of such an analysis would make it possible to account for trends, lags, and attenuations in the effects, as well as for geographic variations in both the dependent and independent variables.

One of the key decisions in any attempt to quantify the impact of the IRS on compliance behavior is **how to measure compliance behavior**. The problem with measures of compliance or noncompliance in this context is that they depend on knowing (or, more typically, estimating) true taxpayer obligations. The basic measure of taxpayer compliance is what taxpayers do: returns filed, tax (or income) reported, and tax paid. However, those observed behaviors will vary across units of observation primarily because the corresponding true obligations also vary. Perhaps the best way to control for those variations in true obligations is to divide observed behavior by the true obligation, thus expressing the compliance measure (dependent variable) as a compliance rate. However, the true obligation often cannot be estimated well (such as for reporting compliance). In that case, it may be best to divide by a major determinant or predictor of the true obligation, and then control for other determinants of true obligations by introducing additional explanatory variables. Finally, I do not think that we should define our compliance measures in this analysis in terms of tax reported or paid. That is because some of the independent variables (e.g., marginal tax rates?) that are likely to have an indirect impact on compliance behavior also have a direct role in the tax calculation itself, making it hard to separate those effects and interpret the results. Therefore, I have used and recommend using the amount of income reported and the amount of offsets to income claimed in lieu of the amount of tax reported. Even then, it would be necessary to control for changes in tax law over time that affect what taxpayers should report.

Alternatives

The next most promising approach, in my view, is **field experiments**. I'm not optimistic that this approach could produce estimates of the marginal impact for many IRS activities, however. I'm concerned about two problems. First, I expect that in this case it will be extremely difficult to establish a control group that is identical to the test group(s) in all respects other than the test. One way to do that is to design both groups as large random samples from the population, presumably having the same mix of relevant characteristics. However, the larger each group is, the more likely that the test will influence the control group (we are very interested in the "ripple" from those contacted to those who are not, after all). The smaller (and further apart geographically) the groups are, the less likely they will be similar to each other. Moreover, I think that the problem of "left-out variable bias" is a greater risk in field experiments than it is in econometrics. For example, let's assume that an important determinant of tax compliance is religion, and that it was not specifically controlled for in either an econometric analysis or a field experiment. Since religious backgrounds and preferences tend to vary a lot geographically, it's quite possible that our control and test groups in the field experiment would not be identical in this regard. Therefore, some of the difference between their compliance behaviors would be attributable to their different religious makeups, but we would attribute all of it to the test. The same mistake could happen in an econometric analysis, but since religion is bound to be uncorrelated with the variables of interest (i.e., IRS service and enforcement activities), our estimates for those variables are not likely to be biased.

My other concern about field experiments is that they are not likely to be a practical way to monitor trends. Regularly repeating the same tests such that the results would be comparable would likely introduce lots of logistical problems. However, field experiments will likely be very helpful for evaluating new services or enforcement interventions.

Perhaps an even better alternative for this project would be what would amount to **a blend of field experiments and econometrics**. We would control for all relevant factors statistically, but we would introduce more variation in the IRS variables than would normally be the case. That variation could be random, but it would probably be better for endogenous variables (e.g., audits) if the variation were accomplished through a different allocation of the budget across geographic areas (preferably defined the same as our units of observation) and over time, without changing the method for selecting returns. One of the problems with many established IRS programs is that the level of output tends to be fairly stable over time, making it difficult to detect marginal effects. Intentionally introducing more variation than normal could solve that problem.

I would expect **lab experiments** to be even less useful than field experiments in producing estimates of indirect compliance effects. Even if the subjects were real taxpayers (instead of students), it is virtually impossible to reproduce real-world tax compliance incentives and disincentives in the lab, making it impossible to project the results to the population. Moreover, I'm even concerned that most experiments on tax compliance behavior may have little internal validity, as well, because the subjects are very likely to adjust their behavior, knowing that they're being watched in an experiment (the Hawthorne effect). Lab experiments may have value, however, in helping to demonstrate the possibility of an effect, which could later be tested econometrically. They might also help to identify important non-IRS determinants that should be controlled for in the econometrics.



It may be possible to learn some useful things from taxpayer **surveys**, but I'm skeptical that taxpayers can be completely truthful about their compliance behavior—even if they understood their own behavior well enough and could articulate it. It's much more valuable to focus on what taxpayers do than on what they say, particularly about their tax compliance behavior. Having said that, surveys could be helpful in identifying some trends and some relevant explanatory variables.

I am also skeptical that **simulation models** can produce accurate estimates of tax compliance behavior. Like microeconomic models, they depend on modeling the mechanisms that drive taxpayer behavior, and I believe that these mechanisms are too complex, too varied, and too based on unobservable factors to be practical (though I'm willing to be convinced otherwise). Ultimately, the accuracy of these models (as with all of the other methods) depends heavily on the assumptions that are made in them—both explicitly and implicitly. I need to be convinced that the assumptions underlying a simulation model are stronger than the assumptions underlying an econometric approach.

18. John Scholz

I. Theory:

Analyzing the impact of taxpayer services on compliance is best done in the broader context of how taxpayers find out about strategies to reduce their tax liabilities, and requires some understanding of what impact the various sources of information have on compliance behavior.

For example, a study of the 1986 TRA found that the tax compliance orientation of individual who provided information about TRA to a taxpayer had a significant impact on the respondent's own orientation toward compliance. Surprisingly, those most hurt by the TRA did not necessarily seek advice from those most likely to encourage noncompliance, but if they did by chance encounter noncompliant advisors, they would become more noncompliant. Thus IRS services that could provide information as an alternative in the most noncompliant areas may reduce the spread of noncompliant attitudes.

II. Research design issues:

Matching survey and longitudinal tax return data. This allows the integration of questions of social context and of usage of IRS services with tax reporting behavior before and after such usage. Increases in reporting of relevant income/deduction items by those exposed to the service compared with those not exposed (controlling for potential selection effects and historical individual income trends) can provide an indicator of increased compliance. The potential is good, but the results from the 1986TRA study suggest that significant differences are difficult to detect because of the considerable variability of year to year tax reporting. Thus homogeneous groups and appropriate sizes of samples would have to be carefully determined for studies linking the use of taxpayer services with increases in compliance.

Matching field experiments and tax return data,

This provides a powerful potential to test the effects of particular services, but requires some means of following up the reporting behavior of those receiving the service to determine impact of the service on reporting (at least on the change in reporting after receiving the service.)

III. Focus of research:

There is a trade-off between generalizable, large-scale studies that could quantify impacts over a broad set of tax returns and smaller studies that produce less generalizable studies but can better focus on specific issues and specific aspects of the process.

The CID of the IRS did several interesting studies in the 1970s that focused on a specific location and population (catfish farms in Louisiana, waitresses in Nevada). They researched the problem, did several targeted audits with considerable publicity, and compared income reporting in the sector before and after to show major changes in compliance. The Australian Tax Office did a similar series of enforcement/service/research projects for the more noncompliant small business categories. They used enforcement to find out the major areas of underreporting in troublesome industries, then developed an information service package with associations representing the industry, and followed up with tax return studies to examine changes in reporting patterns after the program. Unfortunately, the knowledge acquired from individual projects was not (to my knowledge) used to provide a broader picture of the impact of taxpayer services on compliance. But this model could be very useful to provide more context-rich information in conjunction with more broad-based field experiment and survey-based research.

19. Joel Slemrod

Statement of the Problem:

The Service would like to measure and model the separate marginal impacts (as a function of the level of IRS effort) of a wide range of taxpayer services and enforcement activities on the voluntary compliance of taxpayers—both those who have interacted directly with the IRS and those in the general population who have not.

- These impacts (what we call “indirect effects”) are behavioral responses, which are different from the “direct effect” of IRS activities, which is the additional revenue generated directly from enforcement contacts.
- Enforcement contacts are not anonymous, while many service contacts are anonymous. Furthermore, many service contacts are initiated by the taxpayer, while enforcement contacts generally are not.
- What methodology do you recommend? Is this an approach for the short term and long term (since short-term, limited studies may be useful for refining the design of longer-term studies and data collection)?
- Likely strengths and weaknesses of alternative methodologies, including an explicit assessment of the plausibility of the assumptions made in each.
- Approaches to quantifying voluntary compliance (filing, reporting, and payment compliance are all included, but could conceivably be studied separately).
- Current data availability and future desired data

My Two Cents:

I recommend that the behavioral response to possible policy initiatives be analyzed by a series of randomized field trials: pilot programs (or carveouts of broadly implemented programs) with a randomized treatment group as well as a statistically identical control group.

An example may help illustrate the proposed methodology. Say we are interested in knowing the impact of a “surge” in audits of self-employment (Schedule C) income.

Limit the surge, say, to returns filed from a set of IRS service centers. Publicize it, and implement it as advertised, emphasizing the geographic limitation of the treatment. We note that randomization by area, not by person, is subject to cross-area differences biasing results, so many treatment areas must be selected and difference-in-difference results emphasized.

To measure the impact on behavior one needs access to data from a random sample of returns of the treated group, before and after the treatment, and of a control group, before and after. Ideally that data would include information from returns as filed, and also from a sample of intensively audited returns.

Data from a baseline control group should be maintained. This would consist of a stratified random (i.e., based on the last SSN digits) sample of tax returns as filed, plus information on subsequent enforcement actions. It would be stratified to emphasize those categories of tax returns more likely to feature noncompliance issues, and therefore more likely to have significant behavioral responses to policy changes. It would link individual tax returns to the returns of pass-through entities of which the taxpayer is a principal. It would have imbedded within it a panel, as with the Continuous Work History Sample. A random subset of these returns (and associated pass-through entities) would be subject to a standardized but intensive audit. (These samples would have to be separate from the other SOI samples, because assessing noncompliance through an audit arguably affects the future behavior of the audited party, in part but not necessarily limited to the fact that being audited might affect the taxpayer’s subjective probability of future audit. This particularly matters for the embedded panel.)

In order to achieve a satisfactorily large sample size, for some field trials the data for the treatment group might have to be collected separately from the baseline data described above for each pilot policy. Depending on the nature of the treatment, the baseline control group might have to be enlarged, as well.

The effect of a given policy would be measured by the difference-in-difference effect of policies on two sorts of data: (i) what is reported on tax returns (which requires no audit), and (ii) the results from random audits of the treatment and control groups. With an appropriately designed treatment, the change in reporting captures what the

analysis needs to know. The information collected from audits will enrich one's understanding of the nature of the response, but will add considerable expense to the project. The effect on non-filing would be measured by differences in the number of returns filed, or the number of returns reporting Schedule C income.

Depending on the policy at issue, the randomization could be within a geographical area. For example, callers to a service center could be randomly assigned to different response protocols. This sort of procedure, though, raised difficult issues of tracking any behavioral response to the treatment. Follow-up surveys may be useful.

Publicity about the program raises the issue of selection bias because service users are, for the most part, "volunteers" and some may be attracted or dissuaded to make a service contact because of the program publicity. Random assignment does not eliminate selection bias, but ensures that it is the same between the treatment and control groups. If we can assume the policy effect is a constant across individuals, no program impact bias arises.

Like all randomized field trials, some practical and statistical questions arise, for example:

- How to learn why a program works or does not, so as to generalize the findings to other policy initiatives?
- How important is information dispersal and social interaction to the program impact?
- What is the distribution of program impact across taxpayer characteristics?
- How important is the unavoidable limited duration of field trials?

I look forward to discussing these issues.



Department of the Treasury
Office of Research, Analysis, and Statistics
Internal Revenue Service

Predicting Taxpayer Behavior

Produced for IRS under Contract TIRNO-09-Z-00021



Final Report
July 31, 2012

Table of Contents

Section	Page
Executive Summary.....	2
1 Introduction.....	8
2 Econometric Methodology.....	9
2.1 Similarities to Foundational Studies.....	9
2.1.1 Accounting for Unobserved Heterogeneity	10
2.1.2 Controlling for Changes in Taxpayer Reporting and Filing Requirements	11
2.1.3 Endogeneity of Audit Rate.....	13
2.2 Refinements and Extensions	14
2.2.1 Forecasting and Evaluating Performance.....	14
2.2.2 Dynamic Models.....	16
2.2.3 Reporting of Specific Income Sources.....	17
2.2.4 More Extensive Specification Analysis	18
2.2.5 Alternative Time Series Forecasting Methods.....	18
3 Main Findings and Lessons Learned	18
3.1 Reporting of Constant Law Total Income.....	19
3.1.1 Trend Effects vs. Two-Way Fixed Effects.....	19
3.1.2 Trend Effects vs. Random Effects	26
3.1.3 Endogenous Explanatory Variables and Instrument Choices	30
3.1.4 Choice of Functional Form	32
3.1.5 Dynamic Models.....	32
3.1.6 Alternative Forecasting Methods.....	34
3.2 Reporting of Specific income Sources	34
3.2.1 Net Nonfarm Self-Employment Earnings.....	35
3.2.2 Gross Pensions and IRAs.....	39
3.3 Filing Compliance.....	44
3.3.1 Estimation Results for Filing Rate.....	46
3.3.2 Forecasting Performance	51
4 Conclusions and Recommendations.....	53
4.1 Further Specification Analysis and Testing	54
4.2 Alternative Data Sources and Software Programs	55
4.3 Future Applications of the Methodology	56
4.4 Potential for Microeconomic Approaches.....	57
5 References	58
Appendix A: Referenced Project Reports	60
A.1 Task Order 1 Final Report.....	60
A.2 Research Focus Areas 1 (Updated Data) and 2 (Endogeneity) Report.....	60
A.3 Research Focus Area 4 (Filing Compliance) Report.....	61
A.4 Research Focus Area 5 (Time Series Methods) Report	62
A.5 Research Focus Area 6 (Line Item Predictions) Report	62
A.6 Research Focus Area 7 (Fixed Effects Alternatives) Report.....	63
A.7 Research Focus Area 8 (Dynamic Effects) Report.....	64

Executive Summary

This report summarizes extensive econometric research undertaken from September 2009 through July 2012 to evaluate alternative methodologies for predicting aggregate taxpayer reporting and filing behavior by state, gather and compile a research database to implement these methodologies, and report on their performance. The diverse econometric techniques discussed here represent unique and significant contributions to the tax compliance literature, while serving as the foundation for future studies that IRS may conduct. An important focus of the current study includes forecasting future taxpayer filing and reporting behavior at the level it would have been in the absence of some intervention (such as an increase in the quality or quantity of an existing taxpayer service or the introduction of a new service).

The data base for our research (described separately in the *Data User's Guide*) includes longitudinal information on households, taxes, tax administration, tax policy, and socio-economic factors, all of which has been aggregated to the state level. We consider alternative panel data models of the form:

$$Y_{it} = \alpha_i + \gamma_t + \beta'_A A_{it} + \beta'_O O_{it} + \varepsilon_{it}, \quad (1)$$

where Y represents a measure of filing or reporting behavior, A represents a set of IRS activities (such as enforcement and/or service activities), and O represents a set of other relevant measured determinants of taxpayer behavior. The subscripts “ i ” and “ t ” represent individual states and years, respectively, reflecting our objective of explaining the variation in reporting behavior across both states and time. In the above specification, the parameters β_A and β_O represent coefficients to be estimated. The term ε_{it} is an error term that is meant to capture the net impact of unobserved factors across states and over time on state-level taxpayer behavior. Finally, the terms α_i and γ_t represent possible sources of state-specific and year-specific heterogeneity. More specifically, α_i represents unobserved time-invariant differences across states and that drive inter-state differences in taxpayer behavior, while γ_t represents unobserved state-invariant differences across years that drive inter-temporal differences in behavior.

This specification presents a series of considerations. Our approach follows the earlier panel data studies by Dubin et al. (1986), Dubin (2007), and especially Plumley (1996) in several respects:

- Following Plumley, we restrict our analysis to required returns, relying on a “constant law” definition of total reported income, and employing a measure of income excluded from taxation as an explanatory variable.
- We account for unobserved heterogeneity among taxpayers and over time. Like Plumley, we tend to favor the fixed effects

Builds on Prior Research Foundations

- Following Plumley (1996), we use a state and year specification for required returns under constant law
- Fixed effects serves as the basis of specification, although we explore the performance of random effects
- A one-way fixed effects model is complemented with a trend term to address the limitations of fixed effects for forecasting

approach as it yields consistent estimates under a wider range of circumstances than the random effects approach.¹ However, we perform some comparisons with the random effects approach to see how sensitive the findings are to the choice of method. We also employ a Hausman test to investigate the validity of a random effects specification.

- We control for year-specific shocks (such as political events or relevant tax law changes) or behavioral trends that are national in scope. When we focus on prediction, the two-way fixed effects model, which we would otherwise favor, has no direct means of generating estimates of the year-specific effects over the forecast period; for this approach, it is therefore necessary to employ an ad hoc method to estimate those effects. In addition to estimating two-way fixed effects models, we explore one-way (state-level) fixed effects models that use trend terms and dummy variables for selected time intervals as an alternative – a “trend effects” model. The disadvantage of this specification is that it may not account as completely for year-specific events that influence national taxpayer filing and reporting behavior.

An extensive panel of behavioral determinants

- Over 750 unique variables from over 20 data sources
- Imputations of pensions, social security, self-employment earnings, taxpayer burden, and combined state-federal marginal tax rates
- Coverage of many variables from 1991-2009; some for an even longer period.

This study ploughs new ground. We present an extensive analysis of potential explanatory variables, alternative time periods, and choices regarding functional forms. Our base specifications follow Plumley in modelling the dependent variable and many of the regressors as ratio variables. However, to address concerns that models containing ratio variables sometimes yield unreliable results, we also test some specifications that do not rely as heavily on ratios (such as specifications in which the natural log of reported income is regressed against the natural log of personal income and other explanatory variables rather than using the ratio of reported income to personal income as the dependent variable).

In carrying out our analysis, we are aided by new data and improved measures of some key variables for the analysis. This includes better criteria for evaluating whether taxpayers have a legal filing obligation; improved measures of pensions and IRAs, social security income, and self-employment earnings; new measures of combined state and federal marginal tax rates; and some alternative measures of taxpayer burden that are motivated by IRS survey research studies.

Focused on prediction, we faced a limitation of modeling the time-specific heterogeneity term using fixed effects: the value of the fixed effect is not known for years outside of the estimation sample, which makes forecasting difficult. We have explored several approaches to account for

¹ In the case of dynamic models discussed later in Section 2.2, neither the random nor the fixed effects approaches are generally appropriate, at least when the panel data base covers a relatively small number of years. For these models, we explore some alternative instrumental variables methods of estimation.

this limitation. In general, we employ trend terms rather than yearly fixed effects in much of our analysis. However, a comparison of our results based on our longer panel analyses indicates that certain parameter estimates (notably, the audit rate and the marginal tax rate coefficients) are sensitive to whether yearly fixed effects or trend terms are employed. To investigate the impact of this choice on predictive performance, we have developed an econometric approach to forecasting with yearly fixed effects. Under this approach, we predict the value of the fixed effects for years outside of the sample period based on the estimated sample period fixed effects. Our approach uses time series techniques to model the fixed effects.

To evaluate the forecasting performance of alternative models, we have developed a “step-ahead” forecasting methodology for evaluating out-of-sample performance. Under this approach, we begin by estimating our specifications using all but the last one to several years of data in our sample. We then employ the estimation results to predict the reporting (or filing) behavior for each state in each of the excluded years based on the observed values of the explanatory variables in those years. In some cases, we have also assessed out-of-sample predictive performance using a “leave-one-out” prediction methodology. Under this approach, one estimates the panel data specification using all years in the data sample except the first and then uses the results to predict the value of reported income in each state in the left out year. One then repeats the exercise, this time leaving out the second rather than the first year and predicting the value of reported income in each state in the second year. The process continues until out-of-sample predictions have been made for each state in every year of the sample.

New contributions

- Focus on prediction for new time periods, using a one-way fixed effects model with trend and a variety of time series techniques
- Dynamic (lagged) effects to express the impact of past behavior and prior enforcement and service experiences
- Examination of the potential for panel data models to predict both overall income reporting and the reporting of specific income sources
- New econometric evidence on the drivers of filing compliance

For both our step-ahead and leave-one-out forecasting approaches, we focus on two alternative measures of out-of-sample predictive performance. The first is the mean absolute deviation of the out-of-sample prediction of reported income in each state and time period from the true value of reported income. The second is the root mean-squared error (i.e., the square root of the average squared deviation of the out-of-sample prediction from the actual value). Both of these measures are normalized by dividing them by the average value of reported income over all states and time periods. We refer to the first measure as the “absolute deviation as a percentage of income”. The second measure is known in the statistics literature as the “coefficient of variation of root mean-squared error”.

In addition to evaluating the predictive performance of alternative panel data econometric models of taxpayer reporting behavior, we have also experimented with some alternative time series approaches to forecasting that are rooted in autoregressive integrated moving average (ARIMA) processes. We have found that some of these approaches work reasonably well and may serve as attractive options in applications where relatively few explanatory variables for taxpayer filing or reporting behavior are present.

Often the impact of changes in public policy or economic factors on individual behavior takes place gradually over time rather than all at once. While the models presented in the foundational studies are static in nature, we have extended our analysis to include some dynamic models of behavior. These dynamic models account for the influence of past taxpayer behavior as well as current and lagged values of other causal factors. Our dynamic panel data models account more generally for temporal factors when predicting how taxpayers will adjust their filing and reporting behavior in response to changing circumstances. However, the usual two-way fixed effects, trend effects, and random effects estimators are biased when the list of explanatory variables includes a lagged dependent variable. This bias becomes negligible as the number of time periods in estimation becomes sufficiently large. For our experiments with up to 17 years, it is uncertain whether there are a sufficient number of periods to safely ignore the bias. Therefore we compare results from the standard fixed effects estimation with those from a fixed effects estimator (Kiviet, 1995; Bruno, 2005) that corrects for the bias associated with dynamic estimation. We also explore the performance of some alternative estimation methods proposed by Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991). These alternative methods produce consistent parameter estimates even when the time dimension of the panel data base is small.

In addition to examining overall income reporting, we examine reporting of two specific income sources: net nonfarm self-employment earnings, restricted to sole proprietors with positive net earnings, and an imputed measure of pension and IRA income.

The main findings of this research program are as follows:

- Overall, it appears to be more difficult to forecast taxpayer reporting for a specific income source, such as net nonfarm self-employment income or pensions and IRAs, than for overall income. In the case of nonfarm self-employment income, a challenge is the lack of a high quality independent measure of the incidence and magnitude of this income source within the overall population against which to compare net earnings reported on tax returns. In the case of pensions and IRAs, a challenge is the relatively high degree of voluntary compliance in the reporting of this income item. Many of the observed variations in the reporting of this income source over time and across states are likely due to random factors rather than changes in compliance behavior.
- Longer panels provide more degrees of freedom for estimation and prediction; however, there are also some drawbacks. For instance, if there have been structural changes in taxpayer reporting or filing activities over the estimation period, earlier time periods in the sample may not have much predictive content for recent years. As well, shorter panels permit the inclusion of potentially important explanatory variables that are not available over a longer time span. An advantage of a sufficiently long panel is that the bias associated with dynamic fixed effects estimation approaches tends to be relatively small. However, alternative dynamic estimation techniques such as Arellano-Bond and Anderson-Hsiao produce consistent parameter estimates in shorter panels.
- For interpreting the role of certain key factors in overall reporting behavior (such as the audit rate and federal-state marginal tax rate), a two-way fixed effects model seems to be preferable; trend effects and random effects models sometimes produce counter-intuitive

predictions regarding the marginal effects of such factors. However, it is challenging to predict fixed year effects outside of the estimation sample under the two-way fixed effects approach, which limits its usefulness for forecasting future taxpayer behavior.

- The best choice of model for forecasting purposes is somewhat case dependent. Among static models, the trend effects model tends to perform reasonably well in reasonably long panels. However, the exhaustion of degrees of freedom in estimating state level fixed effects can hamper its performance in shorter panels. In shorter panels, a random effects model tends to perform reasonably well. This is interesting, because a Hausman test consistently rejects random effects in favor of fixed effects in our models of overall taxpayer reporting behavior. In the case of dynamic models, trend effects estimation or bias-corrected trend effects estimation sometimes works quite well, at least in reasonably long panels. However, in other cases, the forecasting performance of the Anderson-Hsiao or Arellano-Bond approaches is superior.
- The choice of functional form in panel data specifications can be very important for forecast quality. While a ratio specification for the dependent variable performed reasonably well in our models of overall taxpayer reporting behavior (constant law total income as a share of total personal income) and filing behavior (the filing rate), a levels specification (natural log of pensions and IRAs reported) performed much better in our model of pension and IRA reporting.
- In applications where limited explanatory variables are available, an ARIMA or ARIMAX forecasting procedure represents a viable alternative to panel data econometric methods.
- Estimates of forecast performance can be sensitive to the choice of forecast period. In many cases, we found that our models performed relatively poorly in predicting taxpayer reporting and filing outcomes in tax year 2007 based on estimation sample that ended in tax year 2006. Taxpayer reporting and filing behavior for tax year 2007 was atypical as a result of behavioral responses to the Economic Stimulus Act of 2008. This illustrates that forecasting models tend to perform better over relatively calm periods that are not characterized by sudden changes in the policy or economic conditions. It also provides an illustration of how these prediction methods can be applied to estimate the impact of an intervention (even if in a “natural” experiment).

Our recommendations encompass the following four topic areas: (1) further specification analysis and testing; (2) alternative data sources and software programs; (3) future applications of the methodology; (4) potential for micro-econometric approaches.

In the area of further specification analysis and testing, we propose further analysis of the potential endogeneity of certain explanatory variables (including measures of IRS enforcement and services, the marginal tax, and the filing rate) and the suitability of alternative instruments. We also suggest a more extensive analysis of filing compliance that considers some new and alternative explanatory variables, different estimation periods, and the drivers of late filing behavior. In addition, we propose a more comprehensive analysis of the impact of the Economic Stimulus Act of 2008 on filing compliance.

With regard to alternative data sources, we propose consideration of the American Community Survey (ACS) as a future panel data source for income and demographic variables. The principal advantage of the ACS is its very large sample size. Compared to our current data source (Current Population Survey Annual Social and Economic Supplement, or CPS ASEC), which annually surveys about 100,000 household addresses, the ACS surveys several million addresses each year. This large sample size dramatically reduces the sampling variation in the state level measures of variables. Currently, nationally representative micro-level ACS survey findings are available from 2005 (tax year 2004) forward. With regard to the software programs used to compile the current data (particularly those used to process the CPS ASEC data) we suggest development of more user-friendly programs to facilitate future refinements and updates.

Based on our research findings, state level panel data econometric methods do not seem to be well suited for directly predicting the effects of taxpayer services on tax compliance. Panel data on IRS and third-party services provided to taxpayers are limited, and the existing measures are potentially endogenous (since taxpayers who elect to use these services are likely to be different in important ways from taxpayers who do not). In addition, attempts to include service measures in panel data specifications have often yielded statistically insignificant or counter-intuitive estimates of their impact on taxpayer behavior. Panel data econometric methods show more promise as a tool for predicting what future taxpayer behavior would be in the absence of an intervention. Consider, for example, a field experiment in which a treatment group (perhaps those taxpayers in a selected geographic region) is provided with a service intervention, while a control group (perhaps taxpayers in a similar but distinct geographic region) is not offered the intervention. Panel data econometric techniques may prove useful in accounting for inevitable differences between the control group and the treatment group that would cause treatment group behavior to differ from control group behavior even in the absence of the intervention. Research on ways to integrate panel data and field experiment techniques seems warranted.

Another area for future research is microeconomic applications. We suggest conducting a micro-level study of the determinants of taxpayer demand for IRS and third-party services. Such a study would require a cross-sectional or panel data base containing information on various services employed by taxpayers as well as a variety of tax and socio-economic factors. We are less sanguine regarding the prospects of directly estimating the impact of taxpayer services on taxpayer compliance using individual level data. Our concerns include the lack of reliable instruments to control for the self-selection of service options by taxpayers and the difficulties associated with attempting to account for indirect effects of services on taxpayers who do not directly use such services. Rather, we suggest that some field experiments be conducted to learn how services impact taxpayer filing and reporting behavior. Depending on the experiment, micro-level or state level panel data econometric techniques may prove useful in controlling for differences in control and treatment groups that would potentially impact inferences from the experiment.



1 Introduction

In September 2009, IRS engaged IBM (Task Order 1 under Contract TIRNO-09-Z-00021) to design a preliminary methodology for predicting aggregate taxpayer reporting behavior by state, gather and compile a research database to implement this methodology, and report on its performance. A final report on this work was completed on October 21, 2010. Subsequent to this initial task order, IRS engaged IBM under two additional task orders (4 and 6) to conduct more extensive research in this area and to refine and expand the supporting research data base. Table 1 summarizes the research focus areas that directed this additional work.

Table 1: Research Focus Areas Under Task Orders 4 and 6

Research Focus Area	Objective
1	Repeat the analysis conducted under <i>Task Order 1</i> using a revised and updated data base
2	Explore the suitability and performance of alternative instrumental variables
3	Impute measures of social security income, pensions and IRAs, and self-employment earnings for use in developing filing rate measures
4	Perform a panel data econometric analysis of taxpayer filing behavior
5	Explore alternatives to panel data econometric methods for predicting taxpayer reporting behavior
6	Perform a panel data econometric analysis of taxpayer reporting behavior with respect to a selected income source
7	Evaluate the relative forecasting performance of one-way and two-way fixed effects panel data models
8	Extend the panel data econometric framework to examine the role of dynamic factors

IBM has submitted reports describing the work completed within each of these research focus areas. In this report, we synthesize the findings of these individual research focus area reports and the Task Order 1 final report; present new results on taxpayer filing behavior; identify the key lessons learned about modeling and predicting state level taxpayer and filing behavior; and provide recommendations for future research. As a companion to this report, the *Data User's Guide* describes the Analysis Dataset that has been developed to support this and other research.

The remainder of this report is organized as follows. Section 2 provides an overview of the econometric methodology used to analyze state level filing and reporting behavior among U.S. federal taxpayers. One encounters many of the same econometric issues whether one is modeling filing or reporting behavior among taxpayers. We follow the foundational studies by Dubin et al. (1990), Plumley (1996), and Dubin (2007) in our approach to several of these issues. We begin in Section 2 by detailing the similarities between our own econometric approach and those employed in these earlier studies. We then provide a description of the refinements and extensions we have employed to make our study of taxpayer filing and reporting behavior more comprehensive than previous research on these topics. Section 3 describes the main findings and lessons learned from our analysis. The discussion is divided into subsections covering the distinct taxpayer behaviors that have been modeled: overall income reporting, filing behavior, and reporting of selected income sources (net self-employment income and gross pensions and IRAs.) Selected results from the earlier research focus area reports and the Task Order 1 final report are used to highlight these findings. Links to the full reports are provided in Appendix A. In addition, recent extensions to the analyses of taxpayer filing behavior and reporting of gross pensions and IRAs are discussed. Section 4 provides recommendations for future research. Issues where further analysis seems warranted are highlighted, and suggestions are made regarding potential future applications of the methodology. Also included is a discussion of potential challenges and options for a micro-level data analysis of taxpayer reporting and filing behavior.

2 Econometric Methodology

This section lays out the econometric methodology we have employed to examine both taxpayer reporting and filing behavior. Our approach builds on the research foundation set by Dubin et al. (1990), Dubin (2007), and especially Plumley (1996). In our analyses of both reporting and filing decisions, we follow these earlier studies when addressing issues of unobserved heterogeneity across states and over time, changes in tax rules, and the potential endogeneity of the audit rate. These similarities of our econometric methodology with the foundational studies are explored in Section 2.1. However, our approach goes beyond these studies in a variety of directions. In Section 2.2, we discuss how we have refined and extended the econometric framework to more fully explore the potential for predicting taxpayer reporting and filing behavior.

2.1 Similarities to Foundational Studies

In developing our panel data econometric models of reporting and filing behavior, we were faced with many of the same issues encountered by the authors of the foundational studies in this area, including how best to: (1) account for unobserved heterogeneity among taxpayers and over time; (2) control for changes in taxpayer reporting and filing requirements; and (3) address the potential endogeneity of the audit rate.

In discussing these issues, we will refer to the following panel data specification:

$$Y_{it} = \alpha_i + \gamma_t + \beta_A' A_{it} + \beta_O' O_{it} + \varepsilon_{it}, \quad (1)$$

where Y represents a measure of filing or reporting behavior, A represents a set of IRS activities (such as enforcement and/or service activities), and O represents a set of other relevant measured determinants of taxpayer behavior. The subscripts “ i ” and “ t ” represent individual states and years, respectively, reflecting our objective of explaining the variation in reporting behavior across both states and time. In the above specification, the parameters β_A and β_O represent coefficients to be estimated. The term ε_{it} is an error term that is meant to capture the net impact of unobserved factors across states and over time on state-level taxpayer behavior. Finally, the terms α_i and γ_t represent possible sources of state-specific and year-specific heterogeneity. More specifically, α_i represents unobserved time-invariant differences across states and that drive inter-state differences in taxpayer behavior, while γ_t represents unobserved state-invariant differences across years that drive inter-temporal differences in behavior.

2.1.1 Accounting for Unobserved Heterogeneity

Our panel data base, which is described in a companion document to this report, tracks state level filing and reporting behavior over a substantial period of years. This presents an opportunity to control for unobserved state-specific factors (e.g., cultural issues) that remain constant over time when analyzing taxpayer behavior. In Equation 1 above, these factors are reflected in the α_i terms. Two common approaches to accounting for unobserved heterogeneity are referred to as “fixed effects” and “random effects” in the econometrics literature. A good review of these approaches is provided in Hsiao (2003). Under the fixed effects approach, one treats the α_i terms as state-specific constant terms to be estimated like other regression coefficients in the model. One approach to estimating this “one-way fixed effects” model is to include dummy variables for individual states as additional explanatory variables in the model. This is sometimes referred to as the “least squares dummy variable” approach.

Under the random effects approach, one treats the value of α_i in each state as a random draw from a probability distribution, thereby making α_i an element of the overall disturbance term in the model. Owing to the presence of these state-specific error components, the overall disturbance term is correlated over time. Typically, generalized least squares estimation is employed to account for the non-spherical nature of the disturbance.

An advantage of the fixed effects specification is that it produces consistent estimates of the parameters of the model even when the unobserved state-specific factors are correlated with one or more of the explanatory variables in the model. However, if these factors are not correlated with any of the explanatory variables, the random effects specification produces more efficient (precise) estimates; intuitively, the random effects specification exhausts fewer degrees of freedom, because it is not necessary to estimate the value of α_i (“nuisance parameter”) for each state as one does with the fixed effects specification. The fixed effects specification also yields only conditional predictions in the sense that it is limited to predicting observations that come from units for which a fixed effect has been estimated. However, as Plumley points out, since the units in our case are states and essentially all states are included in our analysis (the exception is Alaska, which is excluded because of compatibility issues resulting from the need for all

recipients of Alaska Permanent Fund Dividends to file federal tax returns; also Maryland and DC have been combined together), this is not a meaningful limitation for our application. Like Plumley, we tend to favor the fixed effects approach as it yields consistent estimates under a wider range of circumstances than the random effects approach.² However, we perform some comparisons with the random effects approach to see how sensitive the findings are to the choice of method. We also employ a Hausman test to investigate the validity of a random effects specification.

The panel data base also provides an opportunity to control for year-specific shocks (such as political events or relevant tax law changes) or behavioral trends that are national in scope. In Equation 1 above, these factors are represented by the γ_i terms. Dubin and his colleagues account for such factors in only a limited way; for instance, Dubin et al. (1990) include a dummy variable equal to one for the period 1980-1986 to account for changes in IRS enforcement strategies and tax law changes that were likely to have affected taxpayer behavior in all states. Plumley accounts for year-specific heterogeneity in a more general way: he employs a “two-way fixed effects” model that includes both state-specific and year-specific fixed effects.

In the current study, our focus is on using our models to forecast future taxpayer filing and reporting behavior. A challenge with employing a two-way fixed effects model is that the value of the year-specific effects over the forecast period will generally be unknown. In Section 2.2, we present a methodology we have developed to address this challenge. In addition to estimating two-way fixed effects models, we explore one-way (state-level) fixed effects models that use trend terms and dummy variables for selected time intervals as an alternative. In this report, we refer to a one-way fixed effect model that includes a trend term as a “trend effects” model. An advantage of trend effects approach is that it does not require an ad hoc method to forecast year-specific effects. The disadvantage is that it may not account as completely for year-specific events that influence national taxpayer filing and reporting behavior.

2.1.2 Controlling for Changes in Taxpayer Reporting and Filing Requirements

Federal tax laws and regulations are subject to frequent changes. Changes in the filing threshold and the taxability of certain income sources (such as social security income) have a direct impact on the number of filers in a state who are required to file a federal individual income tax return. In addition, many households file a tax return even in the absence of a legal filing obligation in order to receive a refund or claim a refundable tax credit. Therefore, changes in withholding requirements or regulations governing refundable credits also impact the number of returns that are filed. Among those who do file, the overall amount of income reported will depend on the extent to which different income sources are taxable. Over time, the rules governing the extent to which income sources such as dividends, capital gains, unemployment compensation, and social security benefits are taxable have varied. For a given level of income, the amount of taxes

² In the case of dynamic models discussed later in Section 2.2, neither the random nor the fixed effects approaches are generally appropriate, at least when the panel data base covers a relatively small number of years. For these models, we explore some alternative instrumental variables methods of estimation.

reported by filers will depend directly on tax rates, exemption amounts, depreciation rules, and other offset provisions – all of which are subject to substantial variation over time.

Dubin et al. (1990) include a dummy variable equal to one for the period 1980-1986 to account for changes in IRS enforcement strategies and tax law changes that were likely to have affected taxpayer behavior in all states. Dubin (2007) controls for the average combined state and federal marginal tax rate. Beyond these limited measures, neither of the studies controls for the direct impact of changes in tax rules on taxpayer behavior over time. This confounds the interpretation of the estimated regression coefficients in the models. For instance, the estimated coefficient on the audit rate may reflect the impact of changes in enforcement on compliance behavior. However, if trends and interstate variations in audit activity happen to be correlated with changes in tax requirements that have not been accounted for, the coefficient may also be picking up the impact of these latter changes on taxpayer activity. Thus, it is difficult to assess the extent to which a given explanatory variable is associated with noncompliance and the extent to which it is associated with legitimate changes in filing or reporting behavior as a direct response to changes in the tax law.

Plumley (1996) takes several steps to control for changes in the tax laws. First, he restricts his estimation sample to taxpayers with a legal filing obligation. Thus, individuals who file just to obtain a refund or to claim a refundable credit are excluded from the analysis. When analyzing filing behavior, the restriction of the estimation sample to taxpayers with a legal filing obligation is a natural one for studying filing compliance; those without a legal filing obligation do not violate the law when they elect not to file. When analyzing reporting behavior, this restriction eliminates reporting variations that are attributable to changes in the level of filing activity by those without a legal filing obligation. Second, the dependent variable in Plumley's reporting equation is based on a measure of overall reported income rather than reported tax; the dependent variable in the analyses by Dubin et al. (1990) and Dubin (1997) are based on reported tax (or reported tax plus additional audit assessments). Unlike the amount of tax, the overall amount income that must be reported by filers for tax purposes is not directly related to changes in such tax provisions as the tax rate schedule, personal exemption amounts, and tax credits.³ Although Plumley experiments with some alternative definitions of overall income, we focus on his constant law measure of total income. Plumley recognizes that certain income sources, including dividends, capital gains, unemployment compensation, and social security benefits, have not always been fully reportable for tax purposes. To control for this, he excludes all but the first of these income sources from his measure of reported total income, thereby creating a constant law measure. In the case of dividends, taxpayers were able to exclude the first \$100 (\$200 in the case of married joint filers) of dividends from taxation through tax year 1986. The exclusion was disallowed under the Tax Reform Act of 1986 for subsequent tax years. As a constant law adjustment, Plumley continues to apply the dividend exclusion when computing his measure of total income in subsequent tax years.

³ Changes in such provisions may change incentives regarding labor supply, investments, and compliance behavior, thereby *indirectly* influencing the amount of income reported. However, changes in such provisions do not *directly* impact reporting in the way that, say, a change in the rate of exclusion on social security income would impact reporting of this income source.

In his reporting equation, Plumley employs (100 times) the ratio of his constant law measure of total reported income to total personal income (as measured by the Bureau of Economic Analysis or BEA) as the dependent variable. To control for differences between the tax concept of total income and the BEA concept of total personal income, he includes a measure of personal income excluded from taxation as an explanatory variable in his analysis. This measure accounts for all personal income received by households that are below the filing threshold as well as certain non-taxable sources of personal income received by those who are above the filing threshold (including veterans' benefits, workers' compensation, and supplemental security income).⁴

Through these steps, Plumley is able to control for many of the changes in filing rates and income reporting rates that result directly from changes in tax requirements, such as a change in the definition of the tax base. This helps to disentangle legitimate and appropriate taxpayer responses to tax law changes from changes in compliance behavior. We follow Plumley in restricting our analysis to required returns, relying on a constant law definition of total reported income, and employing a measure of income excluded from taxation as an explanatory variable.⁵

2.1.3 Endogeneity of Audit Rate

Each of the foundational studies recognizes that the audit rate is likely to be an endogenous regressor; specifically, the audit rate and the level of noncompliance in a given state will tend to be jointly determined since the IRS strategically targets its enforcement resources across jurisdictions in accordance with compliance levels. To address this issue, each of the studies employs an instrumental variables estimation procedure. A good introductory review of instrumental variables methods in ordinary least squares regression models (which are suitable for fixed effects panel data models) is provided in Kmenta (1997). Instrumental variables approaches for random effects panel data models is covered in Baltagi (2008) and Balestra and Varadharajan-Krishnakumar (1987). Plumley identified some flaws with the original instruments chosen by Dubin et al. and employed some arguably superior instruments in his own study based on direct examination time (i.e., time spent by auditors directly on tax examination activities rather than other administrative functions). We employ the instrumental variables proposed by Plumley in our models. We also experiment with some alternative instruments.

⁴ Since our econometric analysis focuses on individuals with a legal filing obligation, the measure of excluded income in our sample is the one that applies to individuals above the filing threshold. However, the measure in the research data base is extended to account for those below the filing threshold for potential future studies that examine households without a filing requirement.

⁵ Since much of our recent analysis focuses on the period beginning in tax year 1991, it would be feasible to fully include unemployment insurance benefits and dividends (both of which became fully taxable as of tax year 1987). The exclusion on long-term capital gains also was eliminated as of 1987. However, the realization of capital gains is discretionary and sensitive to tax law changes and economic conditions. As well, the maximum tax rate on capital gains has varied over time. Thus, it is difficult to discern whether significant shifts in reported capital gains from one year to the next are attributable to legal discretionary decisions or noncompliant reporting. In the case of social security income, the rules for determining the share of social security earnings that is taxable (which had been in place since tax year 1984) changed in tax year 1994.

2.2 Refinements and Extensions

Our research has gone beyond the foundational studies by Plumley and Dubin et al. in several important ways. First, the primary focus of these studies has been interpreting the parameters of the panel data model and simulating the effect of a change in enforcement on overall taxpayer reporting and filing behavior. In contrast, the focus of the current project is using state level panel data models to predict what taxpayer filing and reporting behavior would have been in the absence of some intervention (such as an increase in the quality or quantity of an existing taxpayer service or the introduction of a new service). Consequently, we have developed methods for forecasting taxpayer behavior out of sample along with criteria to evaluate the performance of those forecasts. Second, the models presented in the foundational studies are static in nature, meaning that filing and reporting behavior is modeled as a contemporaneous function of causal factors. We have extended our analysis to include some dynamic models of behavior. These dynamic models account for the influence of past taxpayer behavior as well as current and lagged values of other causal factors. Third, the foundational models of taxpayer reporting behavior focused on overall income or tax reporting. In addition to examining overall income reporting, we examine reporting of two specific income sources: net nonfarm self-employment earnings and pensions and IRAs. Fourth, we have conducted a more extensive analysis of potential explanatory variables, alternative time periods, and functional forms than the earlier studies. Fifth, we have examined the potential for forecasting taxpayer reporting behavior based on some alternative time series techniques. Finally, as discussed in companion report on the research data base, we have developed improved measures of some key variables for the analysis. This includes better criteria for evaluating whether taxpayers have a legal filing obligation; improved measures of pensions and IRAs, social security income, and self-employment earnings; new measures of combined state and federal marginal tax rates; and some alternative measures of taxpayer burden that are motivated by IRS survey research studies. We discuss each of these refinements and extensions below.

2.2.1 Forecasting and Evaluating Performance

The primary focus of the studies by Plumley and Dubin and his co-authors was interpreting the parameters of the panel data model and simulating the effect of a change in enforcement rates. In contrast, the focus of the current project is using the model to predict what taxpayer reporting behavior would have been in the absence of some intervention (such as an increase in the quality or quantity of an existing taxpayer service or the introduction of a new service). We have therefore investigated ways to develop and evaluate out-of-sample forecasts using our panel data results.

Our primary approach for generating out-of-sample predictions is a “step-ahead” forecasting methodology. Under this approach, we begin by estimating our specifications using all but the last one to several years of data in our sample. We then employ the estimation results to predict the reporting (or filing) behavior for each state in each of the excluded years based on the observed values of the explanatory variables in those years. In some cases, we have also assessed out-of-sample predictive performance using a “leave-one-out” prediction methodology. Under this approach, one estimates the panel data specification using all years in the data sample except the first and then uses the results to predict the value of reported income in each state in the left out year. One then repeats the exercise, this time leaving out the second rather than the first year

and predicting the value of reported income in each state in the second year. The process continues until out-of-sample predictions have been made for each state in every year of the sample.

For both our step-ahead and leave-one-out forecasting approaches, we focus on two alternative measures of out-of-sample predictive performance. The first is the mean absolute deviation of the out-of-sample prediction of reported income in each state and time period from the true value of reported income. The second is the root mean-squared error (i.e., the square root of the average squared deviation of the out-of-sample prediction from the actual value). Both of these measures are normalized by dividing them by the average value of reported income over all states and time periods. We refer to the first measure as the “absolute deviation as a percentage of income”. The second measure is known in the statistics literature as the “coefficient of variation of root mean-squared error”.

As discussed above in Section 2.1, a limitation of modeling the time-specific heterogeneity term using fixed effects is that the value of the fixed effect would not be known for years outside of the sample period, which makes forecasting difficult. We therefore employ trend terms rather than yearly fixed effects in much of our analysis. However, a comparison of our results based on our longer panel analyses indicates that certain parameter estimates (notably, the audit rate and the marginal tax rate coefficients) are sensitive to whether yearly fixed effects or trend terms are employed. To investigate the impact of this choice on predictive performance, we have developed an econometric approach to forecasting with yearly fixed effects. Under this approach, we predict the value of the fixed effects for years outside of the sample period based on the estimated sample period fixed effects. Our approach uses time series techniques to model the fixed effects. The methodology is described in the combined report on Research Focus Areas 1 and 2.

Another complication of our analysis for prediction purposes is the presence of endogenous explanatory variables. Consider the panel specification presented previously as Equation 1:

$$Y_{it} = \alpha_i + \gamma_i + \beta_A A_{it} + \beta_O O_{it} + \varepsilon_{it},$$

where the variable A represents the audit rate – an endogenous explanatory variable. We can consistently estimate the parameters of this model using an instrumental variables approach. After applying instrumental variables estimation using our two-way fixed effects methodology, one can substitute the predicted values of the coefficients in for the actual values and attempt to predict Y as:

$$\hat{Y}_{it} = \hat{\alpha}_i + \hat{\gamma}_i + \hat{\beta}_A A_{it} + \hat{\beta}_O O_{it}. \quad (2)$$

When forecasting years outside the sample, our forecasts of the time specific fixed effect $\hat{\gamma}_i$ can be applied. In general, however, this will not be a consistent predictor of Y , because the error term ε will be correlated with A . Consequently, the expectation of $(Y - \hat{Y})$ given A will (asymptotically) converge to the value:

$$E(\varepsilon | A), \quad (3)$$

the value of the conditional expectation of the error term given the audit rate A . Since ε and A are correlated, this expectation will not be equal to zero. To address this issue, we employ a two-stage approach to prediction motivated by the Wu-Hausman specification test for endogeneity.⁶ In the first stage, we regress the audit rate against all of the explanatory variables of the model as well as the instruments (just as in the first stage of two-stage least squares estimation). We obtain the residual (u) from this regression. In the second stage, we estimate the following regression specification:

$$Y_{it} = \alpha_i + \gamma_t + \beta_A A_{it} + \beta_O' O_{it} + \lambda u_{it} + \varepsilon_{it} . \quad (4)$$

Under the Wu-Hausman test, one performs a t-test of whether the coefficient λ is equal to zero. The intuition for this test is that this extra term involving the residual u accounts for the correlation between ε and A , so that if $\lambda = 0$, there is no correlation and, hence, A is exogenous. Although we have not seen this specification used in the econometric literature for purposes of forecasting, it can also serve this function. In particular, this extra term involving the residual u directly accounts for the conditional expectation of ε given A that was left out of the above prediction formula and was the source of inconsistent estimation. An analogous approach is used when forecasting based on a one-way fixed effects model with a trend term (“trend effects” specification).

For our random effects specification, it is not straightforward to produce forecasts using a Wu-Hausman specification. As an alternative, we have experimented with generating forecasts using the predicted value of the audit rate \hat{A}_{it} in place of the actual audit rate, where the predicted value is obtained from a regression of the audit rate against the other explanatory variables of the model O_{it} as well as the instruments. An advantage of using the predicted audit rate over the actual audit rate in the forecast is that it will not be correlated with the error term of the model.

2.2.2 Dynamic Models

Often the impact of changes in public policy or economic factors on individual behavior takes place gradually over time rather than all at once. For instance, one might anticipate that an upward shift in the audit rate would lead to a gradual improvement in tax reporting compliance over time as taxpayers learn about and react to the increased incidence of audits. Thus, current behavior may depend not only on the contemporaneous values of causal factors, but also on past values. Our static models of taxpayer reporting and filing behavior have addressed dynamic influences in only a very limited way by incorporating a time trend or fixed year effects into alternative panel data specifications. These approaches have accounted for the impact of changes over time in unmeasured factors (such as social norms or attitudes) on taxpayer filing and reporting behavior. Our dynamic panel data models account more generally for temporal factors when predicting how taxpayers will adjust their filing and reporting behavior in response to changing circumstances. The Research Focus Area 8 report provides a detailed discussion of our

⁶ This is based on specification tests developed by Wu (1974) and Hausman (1978) .



approach to the estimation of dynamic panel data models. Some good introductory discussions on dynamic panel data models are provided in Hsiao (2003), Roodman (2008), and Baltagi (2008).

Consider the following extension of the static panel data model presented in Equation 1:

$$Y_{it} = \alpha_i + \gamma_t + \delta Y_{i,t-1} + \beta' Z_{it} + \varepsilon_{it}, \quad (5)$$

where Z_{it} includes the contemporaneous values of the audit rate (A_{it}) and other explanatory variables (O_{it}) as well as lagged values of selected explanatory variables. In this autoregressive distributed lag (ADL) specification, the current value of the dependent variable (a measure of filing or reporting behavior) depends on past taxpayer behavior as well as current and lagged values of other factors. As discussed in the Research Focus Area 8 report, the usual two-way fixed effects, trend effects, and random effects estimators are biased when the list of explanatory variables includes a lagged dependent variable. However, this bias becomes negligible as the number of time periods in estimation becomes sufficiently large. In this report, we experiment with data samples that extend up to 17 years for each state. Whether this is a sufficient number of periods to safely ignore the bias is uncertain. Therefore we compare results from the standard fixed effects estimation with those from a fixed effects estimator (Kiviet, 1995; Bruno, 2005) that corrects for the bias associated with dynamic estimation.⁷ We also explore the performance of some alternative estimation methods proposed by Anderson and Hsiao (1981, 1982) and Arellano and Bond (1991). These alternative methods produce consistent parameter estimates even when the time dimension of the panel data base is small.

2.2.3 Reporting of Specific Income Sources

The foundational models of taxpayer reporting behavior by Plumley and Dubin et al. focused on overall income or tax reporting. In addition to examining overall income reporting, we examine reporting of two specific income sources: net nonfarm self-employment earnings and gross pensions and IRAs. For the former income source, we restrict our attention to sole proprietors with positive net earnings. We do so because the CPS ASEC substantially undercounts the number of sole proprietors with negative earnings. The CPS ASEC is our data source for an independent measure of self-employment earnings that should be reported.

As discussed in the companion report on the research data base, the CPS ASEC understates the number of earners and the gross amount of earnings from pensions and IRAs in the overall population. We therefore impute additional pension and IRA earnings to the data base using prediction equations derived from an econometric analysis of IRS Form 1099-R information documents.

⁷ Under both approaches, we employ an instrumental variable for the current value of the audit rate.

2.2.4 More Extensive Specification Analysis

We have undertaken a more extensive specification analysis than the earlier foundational studies. We have investigated a broader set of explanatory variables, including some new measures of IRS services. We have also experimented with some alternative time periods of different durations. As well, we have investigated alternative functional forms. Our base specifications follow Plumley in their reliance on a dependent variable in ratio form. For instance, the dependent variable in the overall income reporting equation is defined 100 times as the ratio of constant law total income reported on tax returns in a state to personal income. Many of the explanatory variables are also initially specified in ratio form. As summarized by Wiseman (2009), however, the use of ratio measures in regression analysis is controversial, and there is a growing literature demonstrating that such measures can sometimes lead to spurious and inconsistent findings. We have therefore estimated some alternative specifications that do not rely as heavily on ratios. For instance, we have investigated specifications in which the natural log of reported income is regressed against the natural log of personal income and other explanatory variables rather than using the ratio of reported income to personal income as the dependent variable as is done in Plumley's study. We have also estimated specifications in which many of the ratio explanatory variables have been replaced by variables that separately account for their numerators and denominators. We compare the forecasting performance among the ratio and levels variable specifications.

2.2.5 Alternative Time Series Forecasting Methods

The primary focus of our research has been forecasting state level taxpayer reporting and filing behavior. While panel data econometric techniques represent one viable approach for carrying out these predictions, alternative methodologies do exist for generating forecasts. In the Research Focus Area 5 report, we explore some time series forecasting alternatives and investigate their relative forecasting performance. Among the alternatives considered are: linear exponential smoothing, pre-specified low-order autoregressive integrated moving average (ARIMA) processes, flexible ARIMA and trend-stationary autoregressive moving average (ARMA) processes customized to fit individual state time series, and a low-order ARIMA specification that incorporates exogenous explanatory variables (ARIMAX). Some useful references on these techniques include Bowerman and O'Connell (1993), Box, Jenkins, and Reinsel (1994), Chatfield (2004), and Hamilton (1994).

3 Main Findings and Lessons Learned

In this section we highlight the main findings of our research and summarize the lessons learned. We begin by discussing results for our models of overall income reporting by federal taxpayers. Next we describe findings for the reporting of specific income sources, including net self-employment earnings and gross pensions and IRAs. We then cover results for models of filing behavior.



3.1 Reporting of Constant Law Total Income

The Task Order 1 Final Report presented preliminary results from some alternative panel data models of taxpayer reporting of constant law total income. An updated analysis using improved data was provided in the combined report on Research Focus Areas 1 and 2 and, more recently, in the Research Focus Area 7 report. A final update that included results from some alternative dynamic panel data models was given in the Research Focus Area 8 report. Overall, the more recent reports based on improved data corroborated and extended the findings of the earlier reports. Below we highlight the main findings and lessons learned from these analyses. As noted in Section 2, all of our specifications relate to required returns; taxpayers with no legal filing obligation who file simply to receive a refund or to claim a credit are excluded from the analysis.

3.1.1 Trend Effects vs. Two-Way Fixed Effects

We have estimated two-way fixed effects and trend effects specifications of taxpayer reporting behavior. As discussed in Section 2, the two-way fixed effects specification controls for both unobserved time-invariant state level characteristics (such as cultural factors) and unobserved year-specific national factors (such as political events or tax law changes) that influence reporting behavior. The trend effects specification also controls for unobserved time-invariant state level characteristics. However, the fixed year-specific effects estimated under the two-way approach are replaced by a trend term. The advantage of the trend effects specification is that it does not require an ad hoc methodology for predicting future year-specific effects when forecasting taxpayer behavior. The disadvantage is that it may not control as effectively for unobserved year-specific factors.

In the Research Focus Area 7 report, we estimated trend effects and two-way fixed effects specifications using both a long panel covering the tax year period from 1983-2008 and a short panel covering the period 2002-2008. The short panel allowed us to incorporate some additional explanatory variables that were not available prior to tax year 2002, including state-level measures of call attempts to the IRS help line and non-campus staff hours devoted to mission critical taxpayer service activities. The variables employed in these specifications are summarized below in Table 2. In many of our specifications, the reporting concept used as the dependent variable in the analysis is expressed as a ratio of constant law total income reported on federal income tax returns in the state to total personal income. The denominator may be thought of as a rough proxy of how much income should have been reported on required returns. To account for some of the imperfections in the denominator as a measure of what should have been reported, the share of personal income that is excluded from federal income taxation (exclincomepct) is included as an explanatory variable.

Table 2: Variable Definitions

Variable	Definition
incomepcta	Dependent variable defined as 100 times the ratio of constant law measure of total income reported on federal income tax returns to the BEA measure of personal income
exclincomepct	The ratio of income excluded from federal income taxation (veterans' benefits, workers' compensation, supplemental security income on required returns as well as the total personal income of households without a filing requirement) to total personal income
filingrate	Number of required returns filed divided by number of required returns in population
fthresholdpct	Aggregate value of filing threshold among all required returns as a % of personal income
c_marg_95_fixed	Dollar-weighted average combined federal-state marginal tax rate based on a fixed tax year 1995 distribution of income
depamountpct	Total value of the dependent exemption from required returns as a percentage of personal income
lnaud	Natural log of one plus the audit rate of individual returns by agents and auditors/TCO in fiscal year.
det_pct	Percentage of examiners' overall time that was devoted to direct examination activities (used as an instrumental variable).
lndetlag	Natural log of the lagged value of the average direct examination time of revenue agents per individual tax return audited (used as an instrumental variable).
callattemptpct	Total call attempts as a percentage of the population
annsrvrscap	State level measure of the average number of non-campus staff hours in mission critical service occupations per potential return
paidprep	Percentage of required returns completed by a paid preparer
soleprops	Percentage of potential returns having non-farm proprietorship income
soleproptfs	Variable soleprops multiplied by the percentage of non-farm employment in the trade, finance & service sectors
singles	Number of single or head of household returns as a percentage of the number of potential returns
under30	Number of returns with primary taxpayer under age 30 as a percentage of the number of potential returns
over64	Number of returns with a primary taxpayer over age 64 as a percentage of the number of potential returns
pcbirths	Number born per thousand of population
unemplrate	Unemployment rate
collegepct	Percentage of potential returns with a college-educated primary filer



Variable	Definition
malepct	Percentage of single and head of household potential returns belonging to males
homeownerpct	Percentage of potential returns belonging to homeowners
popdensity	Population density
trend	Tax year minus 1981

The filing rate (percentage of required returns that are actually filed) is included to account for the role of filing compliance on taxpayer reporting behavior. All else equal, a decline in filing compliance would result in fewer returns being filed, and therefore a smaller share of aggregate personal income being reported on tax returns.

Several variables relating to tax policy are included. The filing threshold percentage (fthresholdpct) controls for changes in the value of the standard deduction and personal exemption. For a given set of tax rates, a higher filing threshold implies lower taxes, which could influence the reporting of income. The impact of such a change on taxpayer reporting behavior is uncertain. According to the standard economic model of taxation (Allingham and Sandmo, 1972), a higher level of after tax income would induce more risk taking and, hence, lower compliance. On the other hand, low income taxpayers who continue to have a filing requirement may be able to report more of their income without incurring a tax liability, in which case their compliance may improve. Another policy variable is the combined federal-state marginal tax rate. The impact of an increase in this rate on taxpayer reporting behavior is also uncertain. A higher marginal tax rate induces a negative income effect, which would tend to discourage risk taking and, hence, improve compliance. At the same time, high tax rates may induce taxpayers to seek additional ways to legally reduce their tax liabilities. The federal-state marginal tax rate will tend to be endogenously determined, because the rate is graduated and depends upon the amounts of income actually reported by taxpayers within a state. To address this issue in a simple manner, we employ a marginal tax rate measure based on a fixed national distribution of income. This proxy measure does not suffer from the endogeneity problem and seems to perform as well as a more complex instrumental variables procedure that jointly treats the audit rate and the marginal tax rate as endogenous. Further discussion of the potential endogeneity of the marginal tax rate as well as of the filing rate is provided below in Section 3.1. A third policy variable is the value of dependent exemptions as a share of total personal income (depaamountpct). The potential impact of this variable on taxpayer reporting behavior is uncertain for the same reasons that the impact of the filing threshold percentage is uncertain.

A subset of the regressors in Table 2 relate to administrative policy. The natural log of the audit rate (lnaud) is one such variable. All else equal, an increase in the audit rate is expected to



increase compliance. As noted in Section 2, this variable is potentially endogenous since the IRS will tend to invest more audit resources in areas where noncompliance is more prevalent. We account for this by applying an instrumental variables estimation procedure. The variables `det_pct` and `lagdet` are instruments for the audit rate that are used in this procedure. The ratio of call attempts to the population (`callattemptspct`) and average number of IRS non-campus staff hours in mission-critical services per potential return are included to account for the role of IRS services in taxpayer reporting behavior. The share of returns prepared by a paid tax practitioner (`paidprep`) is included to control for the role of third-party services on taxpayer reporting behavior. Refer to Erard (1993) for an analysis of how tax practitioners influence tax compliance.

The remaining variables listed in Table 2 refer to socio-economic factors that may influence reporting behavior.

The main findings for the long and short panels are discussed below. Further details are provided in the Research Focus Area 7 report.

3.1.1.1 Long Panel Results

The long panel estimation results for the trend effects and two-way fixed effects specifications are compared in Table 3.

Table 3: Long Panel Estimation Results, Tax Years 1993-2008

	Trend Effects	Two-way Fixed Effects
<code>lnaud</code>	-5.441** (3.06)	10.467** (3.14)
<code>c_marg_95_fixed</code>	12.278 (0.93)	-107.717*** (4.39)
<code>filingrate</code>	0.381*** (10.24)	0.307*** (8.67)
<code>fthresholdpct</code>	1.440*** (7.53)	1.691*** (8.55)
<code>depamountpct</code>	0.577 (1.22)	0.396 (0.89)
<code>soleprops</code>	-0.045 (0.14)	0.127 (0.41)
<code>soleproptfs</code>	-0.001 (0.14)	-0.004 (0.76)
<code>paidprep</code>	-0.022 (0.55)	0.039 (1.10)
<code>singles</code>	0.111 (1.61)	0.139* (2.20)
<code>under30</code>	-0.040 (0.68)	0.077 (1.38)
<code>over64</code>	-0.056 (0.83)	-0.061 (0.96)



	Trend Effects	Two-way Fixed Effects
pcbirths	0.427* (2.39)	0.677*** (3.36)
exclincomepct	-0.830*** (3.35)	-0.588* (2.44)
unemplrate	-0.688*** (6.38)	-0.629*** (5.46)
collesepct	0.142*** (3.33)	0.027 (0.64)
malepct	0.096 (1.28)	0.030 (0.43)
homeownerpct	-0.174** (2.94)	-0.026 (0.49)
popdensity	0.010 (1.16)	-0.006 (0.75)
trend	-0.065 (0.82)	
N	784	784

Absolute Value of t-Statistics in Parentheses

* p<0.05, ** p<0.01, *** p<0.001

As indicated in the table, the parameter estimates for most explanatory variables are qualitatively fairly similar across the two specifications. The key exceptions are the estimated coefficients of the audit rate and marginal tax rate. The trend effects model counter-intuitively predicts that audit rates are negatively associated with compliance. It also predicts a positive, but insignificant, relationship between the marginal tax rate and compliance. Some theoretical models predict that the marginal tax rate is, in fact, positively associated with compliance. For instance, the traditional economic deterrence model predicts that, when penalties are proportional to understated taxes, an increase in the marginal tax rate will have a negative income effect, thereby reducing the willingness of the taxpayer to gamble by understating taxes. However, many policy analysts believe that high marginal tax rates are responsible for relatively high levels of tax evasion. Consistent with this latter perspective, the two-way fixed effect model predicts a negative and statistically significant relationship between marginal tax rates and tax compliance. The two-way fixed effect model also predicts a more intuitive positive and statistically significant relationship between the audit rate and tax compliance.

In the combined report on Research Focus Areas 1 and 2, a measure of criminal sentences resulting from IRS criminal prosecutions as a share of the state population was included as an explanatory variable for taxpayer reporting behavior. The estimated coefficient was positive, indicating that the criminal sentence rate is positively associated with income reporting by taxpayers. However, the coefficient was not statistically significant. This explanatory variable is not included in the above specification, because the measure is not available for tax years beyond 2005.

Overall, the two-way fixed effects model for this period accords better with prior expectations regarding the impact of audits and tax rates on taxpayer reporting behavior. However, the focus of our analysis is on forecasting performance. Our step-ahead methodology for evaluating



forecast performance was described earlier in Section 2. Table 4 summarizes the relative performance of our trend effects and two-way fixed effects specifications based on that methodology. Recall that for the two-way fixed effects model, we employ a time series approach to predict the value of the fixed year effects in the forecast years. A double exponential smoothing model was used to predict these effects for the results presented in the table. Both specifications perform reasonably well in predicting reported income out of sample. However, the Research Focus Area 7 report indicates that the forecasting performance of the two-way model is somewhat sensitive to the choice of approach for forecasting the yearly effects. From this standpoint, the trend effects specification is somewhat more reliable.

Another interesting finding in Table 4 is that the forecast performance tends to be worse in tax year 2007 than in tax year 2008 when estimation samples ending in either 2005 or 2006 have been employed. Normally, one would expect forecast performance to drop off as one predicts further into the future. A plausible explanation for this finding is that tax year 2007 was an unusual year. In particular, the Economic Stimulus Act of 2008 introduced a tax rebate of up to \$1,200 for households that filed a tax year 2007 return. This stimulus payment helped to promote an unusually high number of filers (both with and without a legal filing obligation) in that year (143 million compared to 138.4 million in tax year 2006). Restricting attention to timely filers with a legal filing obligation, there were 117 million in tax year 2007 compared to only 112.9 million in the prior year. Although we have not investigated the characteristics of the additional timely required filers in tax year 2007, intuition would suggest that they tended to have relatively modest incomes and, in many cases, were only filing to receive the stimulus payment. In other words, these new filers were unlikely to be typical of existing filers in prior years. A lesson from this finding is that econometric models implicitly assume that the future will be similar to the past. Consequently, they will tend to perform more effectively when generating forecasts for a “normal” year than for an “unusual” year.

Table 4: Performance in Forecasting Aggregate State Level Reported Income: Long Panel Specifications

Last Sample Year	Forecast Year	Trend Effects		Two-way FE	
		% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE
2005	2006	1.748	2.760	1.790	2.674
	2007	2.381	3.508	2.333	3.691
	2008	1.901	3.161	2.306	4.064
2006	2007	2.263	3.311	2.231	3.559
	2008	1.838	3.031	1.953	3.470
2007	2008	1.726	3.098	3.466	5.479

3.1.1.2 Short Panel Results

The short panel estimation results for the trend effects and two-way fixed effects specifications are compared in Table 5. For this time period, the estimated signs of the audit rate and marginal tax rate coefficients are consistent across specifications. However, the coefficients are in most



cases imprecisely estimated. The short panel specifications include two explanatory variables related to taxpayer service: callattemptpct and annsrvhrscap. Neither specification finds these variables to be statistically significant predictors of taxpayer reporting behavior.

Table 5: Short Panel Estimation Results, Tax Years 2002-2008

	Trend Effects	Two-way Fixed Effects
lnaud	16.321 (1.71)	19.814 (1.37)
c marg 95 fixed	-8.726 (0.38)	-179.044*** (3.35)
filingrate	-0.115 (1.43)	-0.034 (0.40)
fthresholdpct	-0.146 (0.45)	0.487 (1.16)
depamountpct	0.786 (0.93)	0.325 (0.37)
soleprops	-1.184 (1.83)	-1.444* (2.06)
soleproptfs	0.019 (1.75)	0.024* (2.11)
paidprep	-0.111 (1.88)	-0.048 (0.77)
singles	-0.032 (0.34)	0.010 (0.10)
under30	0.015 (0.16)	0.058 (0.60)
over64	-0.052 (0.53)	-0.076 (0.78)
pcbirths	0.337 (0.98)	0.739 (1.61)
exclincomepct	0.745* (1.98)	0.448 (1.14)
unemplrate	-0.676*** (3.70)	-0.611** (3.20)
collegepct	0.047 (0.76)	0.025 (0.36)
malepct	-0.038 (0.27)	-0.043 (0.31)
homeownerpct	0.002 (0.03)	0.031 (0.39)
popdensity	-0.115*** (3.84)	-0.113*** (3.87)
callattemptpct	0.001 (0.03)	-0.013 (0.23)
annsrvhrscap	-1.661 (0.06)	3.006 (0.11)
trend	0.019 (0.10)	
N	343	343

Absolute Value of t-Statistics in Parentheses

* p<0.05, ** p<0.01, *** p<0.001



A comparison of the results in Tables 3 and 5 indicates fewer significant predictors of taxpayer reporting behavior in the short panel; variables `filingrate`, `fthresoldpct`, `pcbirths`, and `exclincpct` lose their significance in the shorter timeframe. This may be a reflection of fewer degrees of freedom. Alternatively, it could be an indication of possible structural change. Interestingly, population density is found to be negatively associated with taxpayer reporting behavior in the short panel, but insignificant in the long panel.

Table 6 summarizes the relative step-ahead forecasting performances of our trend effects and two-way fixed effects specifications. Predictions of the fixed year effects over the forecast period are based on a double exponential smoothing model. The trend effects model performs well in predicting reported income in tax years 2006 and 2007. However, the performance degrades when predicting tax year 2008. The two-way fixed effects model also tends to perform poorly in that year. This may be attributable to the relatively small number of years in the estimation sample under the short panel. One would expect lower quality estimates of the state level fixed effects as the number of years available to estimate those effects becomes small. When the last year of the estimation sample is 2005, the trend effects model outperforms the two-way fixed effects model. However, this is not the case when the last year of the estimation sample is 2006 or 2007. An important finding in the Research Focus Area 7 report is that the forecasting performance of the two-way fixed effects model is highly sensitive to the choice of prediction method for the year-specific fixed effects; some seemingly reasonable choices yield very poor forecasts. Given this finding, the trend effects approach seems preferable overall.

Table 6: Performance in Forecasting Aggregate State Level Reported Income: Short Panel Specifications

Last Sample Year	Forecast Year	Trend Effects		Two-way FE	
		% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE
2005	2006	2.008	3.468	3.096	5.415
	2007	2.031	3.244	3.352	5.870
	2008	4.636	7.256	7.344	11.60
2006	2007	1.527	2.221	1.527	2.221
	2008	4.328	6.337	3.260	4.872
2007	2008	5.190	7.421	4.139	5.872

3.1.2 Trend Effects vs. Random Effects

As discussed in Section 2, we favor a fixed effects over a random effects specification of taxpayer reporting behavior, primarily because fixed effects estimation yields consistent estimates under more general conditions. However, we have explored random effects estimation as an alternative. A Hausman test has been used to investigate whether the assumptions of the random effects model are satisfied in our context. Based on this test, the random effects specification has been consistently rejected in favor of the fixed effects approach. Of course, while this rejection of the random effects model indicates potentially biased parameter estimates



from that approach, it does not necessarily indicate that the approach would perform poorly in forecasting taxpayer behavior. We therefore also perform a comparison of the relative out-of-sample forecast performance of our fixed and random effects specifications. In this section, we focus on a comparison between our trend effects model and a random effects specification. We highlight the main findings on the relative performance of these models below. Results are provided separately for a long and a short panel. Further details on the long panel analysis are provided in the Research Focus Area 8 report, while additional details on the short panel analysis are provided in the Research Focus Area 7 report. Again, instrumental variables estimation is employed for the audit rate using the same instrument concepts used by Plumley (1996).

3.1.2.1 Long Panel Results

Our most up-to-date results comparing trend and random effects specifications over a long panel period are discussed in the Research Focus Area 8 report. The tax year period covered by this analysis extends from 1983-2009. The estimation results for these two specifications are summarized in Table 7. In this analysis, variable `c_marg` (federal-state marginal tax rate based on a fixed 1984 distribution of income) is used in place of variable `c_marg_95_fixed` (federal-state marginal tax rate based on a fixed 1995 distribution of income). Overall the trend effects and random effects parameter estimates are rather similar. Both models produce counter-intuitive estimates of the coefficients of the audit rate and marginal tax rate variables.

Table 7: Long Panel Estimation Results, Tax Years 1993-2009

	Trend Effects	Random Effects
<code>lnaud</code>	-9.849*** (5.39)	-5.775** (3.26)
<code>c_marg</code>	16.889 (1.15)	4.130 (0.45)
<code>filingrate</code>	0.447*** (11.13)	0.383*** (10.18)
<code>fthresholdpct</code>	2.097*** (8.66)	1.636*** (7.95)
<code>depamountpct</code>	-0.028 (0.06)	-0.119 (0.28)
<code>soleprops</code>	-0.286 (0.99)	0.049 (0.24)
<code>soleproptfs</code>	0.003 (0.64)	-0.003 (0.97)
<code>paidprep</code>	-0.042 (1.16)	-0.089** (2.95)
<code>singles</code>	0.116 (1.78)	0.081 (1.26)
<code>under30</code>	-0.051 (0.93)	-0.046 (0.87)
<code>over64</code>	-0.088 (1.37)	-0.182** (2.82)
<code>pcbirths</code>	0.254	0.084



	Trend Effects	Random Effects
	(1.58)	(0.67)
exclincomepct	-0.449*	-0.751***
	(1.96)	(3.44)
unemplrate	-0.845***	-0.778***
	(11.89)	(11.49)
collegepct	0.137***	0.074*
	(3.35)	(2.18)
malepct	0.077	0.060
	(1.07)	(0.82)
homeownerpct	-0.148**	-0.082
	(2.68)	(1.58)
popdensity	0.006	0.007***
	(0.83)	(5.68)
trend	-0.102	
	(1.54)	
N	833	833

Absolute Value of t-Statistics in Parentheses

* p<0.05, ** p<0.01, *** p<0.001

Table 8 summarizes the relative step-ahead forecasting performances of our trend effects and random effects models. In the case of the random effects specification, we develop our forecasts after substituting the predicted value of the audit rate in place of the true rate. Both methods perform reasonably well, although the trend effects model performs slightly better overall. Again we observe that predictive performance tends to be somewhat better in 2008 than in 2007, presumably owing to the effects of the economic stimulus on filing activity.

Table 8: Performance in Forecasting Aggregate State Level Reported Income: Long Panel Specifications

Last Sample Year	Forecast Year	Trend Effects		Random Effects	
		% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE
2006	2007	2.438	4.235	2.369	4.154
	2008	1.839	2.975	2.056	3.607
	2009	2.567	4.568	3.055	5.812
2007	2008	1.964	3.369	1.891	3.304
	2009	2.807	5.690	2.987	5.628
2008	2009	2.328	4.331	2.788	5.255

3.1.2.2 Short Panel Results

Our most recent estimates of trend and random effects models based on a short panel are presented in the Research Focus Area 7 report. The tax year period covered by this analysis extends from 2002-2008. The estimation results for these two models are summarized in Table 9 below. For this time period, the estimated coefficient of the audit rate (and also of the marginal



tax rate in the case of the trend effects model) is more intuitive. However, the values are not very precisely estimated. Overall, the coefficient estimates differ somewhat more across models for this time period. For instance, the models disagree on the sign of the effect of population density. In addition, the random effects model produces a counter-intuitive and statistically significant negative relationship between IRS activity in mission critical services and taxpayer reporting behavior. The estimated coefficient of this service measure is much smaller and statistically insignificant under the trend effects framework.

Table 9: Short Panel Estimation Results, Tax Years 2002-2008

	Trend Effects	Random Effects
lnaud	16.321	7.395
	(1.71)	(0.64)
c_marg_95_fixed	-8.726	3.383
	(0.38)	(0.29)
filingrate	-0.115	0.047
	(1.43)	(0.75)
fthresholdpct	-0.146	0.299
	(0.45)	(1.27)
depamountpct	0.786	-0.253
	(0.93)	(0.36)
soleprops	-1.184	-0.126
	(1.83)	(0.31)
soleproptfs	0.019	0.000
	(1.75)	(0.06)
paidprep	-0.111	-0.146**
	(1.88)	(3.01)
singles	-0.032	-0.018
	(0.34)	(0.20)
under30	0.015	0.069
	(0.16)	(0.84)
over64	-0.052	-0.192*
	(0.53)	(2.05)
pcbirths	0.337	-0.116
	(0.98)	(0.56)
exclincomepct	0.745*	0.043
	(1.98)	(0.13)
unemplrate	-0.676***	-0.630***
	(3.70)	(3.29)
collegpct	0.047	0.080
	(0.76)	(1.48)
malepct	-0.038	0.147
	(0.27)	(1.11)
homeownerpct	0.002	0.147*
	(0.03)	(2.25)
popdensity	-0.115***	0.006**
	(3.84)	(3.18)
callattemptpct	0.001	-0.021
	(0.03)	(0.80)
annsrvhrcap	-1.661	-49.643*



	Trend Effects	Random Effects
	(0.06)	(2.10)
trend	0.019	
	(0.10)	
N	343	343

Absolute Value of t-Statistics in Parentheses

* p<0.05, ** p<0.01, *** p<0.001

Table 10 summarizes the relative step-ahead forecasting performances of our trend effects and random effects models. In the case of the random effects specification, we develop our forecasts after substituting the predicted value of the audit rate in place of the true rate. Overall, the performances are reasonably comparable for predicting reported income in tax years 2006 and 2007. However, the trend effects model predicts relatively poorly in tax year 2008. This may be a reflection of the relatively small number of degrees of freedom available in this short panel for estimating the state level fixed effects. For this short sample period, the random effects model seems preferable from a forecasting perspective. In fact, a comparison of the results in Tables 8 and 9 indicates that the random effects model for the short panel actually tends to outperform the random effects model for the long panel. This may be an indication of some structural changes over time in the relationship between taxpayer reporting behavior and its determinants; such changes would make data from earlier years in a long panel less helpful in forecasting reporting activity in the current environment. Moreover, tax year 2007 was an atypical year as a result of the stimulus. In fact, tax year 2008 was also rather atypical as some taxpayers were eligible for the recovery rebate credit in that year. Indeed the number of timely filers of required returns declined by only 0.9 million (from 117.0 to 116.1) between tax years 2007 and 2008. The following year, the number dropped to 113.2, which was only 0.3 million higher than the number of timely required filers in tax year 2006 (112.9). The forecasting performance of the long panel specifications may have been more impressive if a more typical set of years had been used for evaluation.

Table 10: Performance in Forecasting Aggregate State Level Reported Income: Short Panel Specifications

Last Sample Year	Forecast Year	Trend Effects		Random Effects	
		% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE
2005	2006	2.008	3.468	1.748	2.973
	2007	2.031	3.244	2.049	3.264
	2008	4.636	7.256	2.588	3.688
2006	2007	1.527	2.221	1.900	2.868
	2008	4.328	6.337	2.020	2.836
2007	2008	5.190	7.421	1.658	2.494

3.1.3 Endogenous Explanatory Variables and Instrument Choices

In each of our models, we specify instruments for the potentially endogenous audit rate variable (lnaud). The federal-state marginal tax rate is also potentially endogenous; owing to the graduated rate structure, the marginal tax rate is sensitive to the overall amount of income

reported on tax returns. In most specifications, we have employed a federal-state marginal tax rate measure based on a fixed distribution of income as an explanatory variable. This variable is closely related to the actual federal-state marginal rate, but it is exogenously determined. A third potentially endogenous variable is the filing rate. In an economic model of tax compliance, one would expect the decision whether to file an individual income tax return to be assessed jointly with the decision of how much to report if one does file an individual income tax return. In the combined report on Research Focus Areas 1 and 2, we undertake some tests of endogeneity for these three variables and investigate the suitability of various instruments for inclusion in an instrumental variables approach to estimation.

We begin by conducting a Wu-Hausman test of the endogeneity of the audit rate variable ($Inaud$) in our two-way fixed effects model. Our initial choice of instruments are the measures proposed by Plumley (2006): $Indetlag$ and det_pct . As discussed in Section 2, this test is based on the following regression:

$$Y_{it} = \alpha_i + \gamma_t + \beta_A A_{it} + \beta_O O_{it} + \lambda u_{it} + \varepsilon_{it}, \quad (6)$$

where u_{it} is the residual from a separate regression of the audit rate variable (A_{it}) against the other explanatory variables in Equation 6 (O_{it} and the fixed effects α_i and γ_t) as well as the instruments. One performs a t-test for the statistical significance of coefficient λ in this specification. If this coefficient is statistically significant, one rejects the null hypothesis that A_{it} is exogenous in favor of the alternative hypothesis that it is, in fact, endogenous. On the basis of the test, we conclude that the audit rate is endogenous in our model. An analogous test leads to the same finding for our trend effects model. Further testing confirms that the instruments $Indetlag$ and det_pct are suitable for the two-way fixed effects model.⁸ Somewhat surprisingly, the results indicate that variable $Indetlag$ may not be a suitable instrument for the trend effects model. Results of some preliminary experimentation indicate that the trend effects model yields qualitatively quite similar results when variable $Indetlag$ is excluded from the instrument set. In our analysis we also investigate several alternative instruments for the audit rate that are based on political economy considerations. A similar set of instruments was employed by Dubin (2007) in some of his specifications. Our preliminary findings suggest that our group of political economy variables does not appear to be suitable as an instrument set. Further investigation seems warranted to identify the most appropriate instrument set for the audit rate. Based on the existing findings, variable det_pct seems like a generally appropriate instrument. This instrument alone would suffice for models with a single audit rate measure. However, in extensions of the model to account for additional audit rate measures (such as separate measures for tax auditors and revenue agents), additional instruments would be required. As discussed below, we have estimated some dynamic specifications of taxpayer compliance. Under the Anderson-Hsiao and Arellano-Bond approaches to estimation of such specifications, one employs instruments based on lagged values of the endogenous explanatory variables. One might analogously consider

⁸ To be suitable, an instrument must be significantly correlated with the endogenous explanatory variable (after accounting for the correlation between the endogenous regressor and the other regressors in the model) and it must be uncorrelated with the error term of the model.

employing lagged values of the audit rate as instruments in future static models of taxpayer reporting behavior. One might also experiment with lagged values of variable `det_pct` as additional instruments.

When we extend the analysis to allow for the potential endogeneity of the federal-state marginal tax rate and the filing rate as well as the audit rate, the test results suggest that at least some subset of these three variables is endogenous in both the two-way fixed effects and trend effects models. Our preliminary analysis indicates that the federal-state marginal tax rate based on a fixed income distribution is a suitable instrument for the marginal tax rate, while the ratio of required to potential returns is a suitable instrument for the filing rate; potential returns are returns that would have a filing requirement if the filing threshold were zero. Further research seems warranted to confirm the endogeneity of this group of three variables and to explore the implications of jointly employing instruments for each of the endogenous members of the group.

3.1.4 Choice of Functional Form

As discussed in Section 2, evidence from the econometrics literature suggests that regression models that specify dependent and explanatory variables in ratio form sometimes yield misleading or spurious results. This is of some concern, because the dependent variable in our main specifications is constructed as a ratio of two income measures, while the set of explanatory variables also includes several variables in ratio form. Appendix E of the Task 1 Final Report presents estimation and forecasting results based on some alternative functional forms. This includes specifications in which the dependent variable is the natural log of constant law total income reported rather than the ratio of constant law total income reported to total personal income. In one variant, the natural log of total personal income is included as an additional explanatory variable and the other regressors are the same as in the previous specification. In the other variant, many of the ratio variables are eliminated. In their place are separate measures of the numerators and denominators of these ratios. The estimation results are qualitatively similar to the results for the original ratio dependent variable specification, and the forecasting performance is also fairly similar. Thus, at least for our models of overall reporting behavior, our specification based on ratio variables seems to be adequate.

3.1.5 Dynamic Models

Often the impact of changes in public policy or economic factors on individual behavior takes place gradually over time rather than all at once. For instance, one might anticipate that an upward shift in the audit rate would lead to a gradual improvement in reporting compliance over time as taxpayers learn about and react to the increased incidence of audits. Thus, current behavior may depend not only on the contemporaneous values of causal factors, but also on past values. The static models of taxpayer reporting behavior discussed so far are not sufficiently general to capture this adjustment process. In the Research Focus Area 8 report, we extend our estimation framework to account for graduated behavioral responses. In one extension, lagged values of selected regressors, including the audit rate, the marginal tax rate, and the unemployment rate, are incorporated as additional explanatory variables in the static long panel model described above in Section 3.1.2.1. Alternative specifications involving up to two lags of



each of these variables are examined. The results indicate that the coefficients of at least some of the lagged variables are statistically significant. The forecasting performance of the extended trend effects specification containing lags of selected explanatory variables is compared that of with the original static trend effects specification with no lagged regressors. Interestingly, the original static trend effects specification performs best when tax year 2006 serves as the last sample year. However, the specifications that include lagged regressors (particularly, the one based on the current and one lagged value of the three selected explanatory variables) tend to perform better when the tax year 2007 or tax year 2008 serves as the last sample year. Recall that tax year 2007 was an unusual year owing to the economic stimulus. This finding may be an indication that lagged values of the selected explanatory variables assist the model in identifying that the impact of the stimulus was temporary in nature.

In another extension, some auto-distributed lag specifications of taxpayer reporting behavior are estimated. These specifications included the lag of the dependent variable as an explanatory variable. As discussed in Section 2, the trend effects estimation of a lagged dependent variable model yields biased parameter estimates. However, this bias becomes negligible as the time dimension of the panel data base becomes sufficiently large. The analysis in the Research Focus Area 8 report involves long panels of up to a 14 year duration. Whether this duration is sufficient to negate the bias associated with trend effects estimation is uncertain. To investigate this question, trend effects method is compared with some alternative estimation procedures, including bias-corrected trend effects estimation (Kiviet, 1995; Bruno, 2005), Anderson-Hsiao estimation (Anderson and Hsiao, 1981 and 1982), and Arellano-Bond estimation (Arellano and Bond, 1991). Interestingly, ordinary trend effects estimation yields the best forecasting performance. Further, once the lagged value of the dependent variable is included as a regressor, lagged values of the audit rate, the marginal tax rate, and the unemployment rate no longer serve to improve forecasting performance. A comparison of the forecasting performance of the trend effects specification that includes a lagged dependent variable with that of the original static trend effects specification is presented in Table 11. At least for a reasonably long panel, it appears that a simple extension of the trend effects approach to include a lagged value of the dependent variable can yield a significant improvement in forecasting performance.

Table 11: Performance in Forecasting Aggregate State Level Reported Income With and Without a Lagged Dependent Variable: Long Panel Specifications

Last Sample Year	Forecast Year	Original Static Trend Effects Specification		Specification With Lagged Dependent Variable	
		% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE
2006	2007	2.438	4.235	2.302	4.356
	2008	1.839	2.975	1.911	3.011
	2009	2.567	4.568	2.296	4.534
2007	2008	1.964	3.369	1.417	2.160
	2009	2.807	5.690	2.060	4.215
2008	2009	2.328	4.331	1.911	4.116

3.1.6 Alternative Forecasting Methods

The Research Focus Area 7 Report reviews some alternative approaches to forecasting what taxpayer reporting behavior would be in the absence of an intervention. The alternatives fall within the autoregressive integrated moving average (ARIMA) framework. The specific approaches include: linear exponential smoothing, pre-specified low-order autoregressive integrated moving average (ARIMA) processes, flexible ARIMA and trend-stationary autoregressive moving average (ARMA) processes customized to fit individual state time series, and a low-order ARIMA specification that incorporates exogenous explanatory variables (ARIMAX). A limitation of these alternatives is that they rely solely on observed variations over time in state-level behavior. Available time series for taxpayer reporting behavior are relatively short and are punctuated by periods of significant legislative and administrative changes in tax filing and reporting requirements. The strength of these approaches is that they are capable of flexibly capturing the dynamics of behavior, in terms of both the impact of past reporting decisions and the lingering influence of external shocks on future behavior.

Overall, we find that a simple low-order moving average process in first differences –performs relatively well among competing time series approaches, including a set of customized state-specific ARIMA and trend-stationary ARMA processes. When this model is augmented by explanatory variables reflecting the level of personal income and the period subsequent to the passage of the Tax Reform Act of 2006, the performance improves and becomes more robust to changes in the size of the estimation sample. Overall, this ARIMAX specification performs only slightly less well than our earlier panel data econometric specifications. The ARIMAX approach could be useful in cases where a simple methodology is needed to forecast what behavior would have been in the absence of an intervention when a full panel data base is not available.

3.2 Reporting of Specific Income Sources

In addition to our modeling of taxpayer reporting behavior with respect to constant law total income, we have undertaken preliminary analyses of reporting with respect to two specific income sources: net nonfarm self-employment earnings and gross pensions and IRAs. These income sources were selected through discussions with IRS Office of Research. Under Task Order 1, we elected to model reporting behavior with respect to Schedule C (nonfarm self-employment) net income. Based on prior tax gap studies, this income source has consistently been an important source of noncompliance. As well, it is a source that is rather commonly reported on tax returns nationwide. In addition, the CPS ASEC provides us with an independent measure of net self-employment earnings, albeit an imperfect one.

Ultimately, we found that our preliminary specifications of Schedule C income reporting tended to predict rather poorly. Under Task Order 6, we elected to select a second income source for analysis. Based on our experience with Schedule C net income, we sought an income source for which a reasonably high quality independent measure was available. In addition, it was deemed important for the source to be fairly commonly reported among taxpayers in all states. Further, we felt that income concept should be associated with a non-trivial degree of noncompliance. Ultimately, it would be desirable to apply panel data estimation techniques to aid in an



evaluation of programs to improve compliance. Therefore, it seems appropriate to focus on a line item where noncompliance poses a significant problem. Based on these considerations, we ultimately selected gross pensions and IRAs. Our reasoning was that we had a good independent measure of pension earnings (the imputed measure described in the Research Focus Area 3 Documentation Report); pension earnings are subject to a modest but significant level of misreporting on tax returns; and pensions are a significant income source in all states. We selected gross earnings rather than taxable earnings primarily because our imputed measure refers to gross earnings and it would be difficult to construct taxable earnings from our gross earnings measure. Table 12 lists the other potential line item income concepts that were considered along with the main reason that they were not selected.

Table 12: Rationales for Not Selecting Alternative Line Item Income Concepts

Variable	Rationale
Wages and Salaries	Very low rate of misreporting
Interest	Very low rate of misreporting
Dividends	Lack of high quality independent measure
State and Local Refunds	Lack of high quality independent measure; difficult to model state-level requirements; low rate of misreporting
Social Security Income	Taxable share of social security earnings difficult to assess; gross amount not always reported on tax return
Unemployment Insurance	Very low rate of misreporting; lack of high quality independent measure
Alimony Received	Very low rate of misreporting; very small share of overall state level income
Schedule F (Farm Self-Employment) Income	Small share of overall state level income in many states; often negative (net losses)
Rents and Royalties	Lack of high quality independent measure – especially for losses
Estates and Trusts	Lack of high quality independent measure – especially for losses
Other Income	Lack of high quality independent measure; many instances of negative income

3.2.1 Net Nonfarm Self-Employment Earnings

The Task Order 1 Final Report presents the results of some preliminary panel data specifications of Schedule C (nonfarm sole proprietor) net income reporting. The dependent variable in these specifications is the Schedule C reporting ratio, which is defined as 100 times the ratio of aggregate state-level Schedule C net income reported on required returns to a CPS-based measure of aggregate state-level Schedule C net income earned by sole proprietors with a filing requirement. The denominator of this expression serves as a rough proxy for the amount of



Schedule C net income that should have been reported on required tax returns, while the numerator represents the amount that actually was reported.

The preliminary panel data specifications include one-way and two-way fixed effects models of the Schedule C reporting ratio. The explanatory variables include many of the same regressors used in our analysis of constant law total income reporting, including a measure of the audit rate, the federal-state marginal tax rate, and various socio-economic factors. However, there are relatively few sole proprietors without a federal income tax filing requirement, because all taxpayers with at least \$433 in net self-employment earnings are required to file a return. We therefore do not include measures such as the excluded income percentage, the filing rate, and the filing threshold percentage, which were all included in our analysis of constant law total income reporting to account for income that was legally excluded from tax returns. However, we do include some new variables that represent the percentages of sole proprietors in various industries, including agriculture, forestry, and fishing; construction; finance; professional business services; education and health services; and leisure and hospitality services. We also include CPS-based measures of total income and the ratio of Schedule C income to total income as controls. We experiment with two alternative measures of the audit rate. The first is the average overall state audit rate for individual returns that was used in our analysis of total income reporting. This variable is available for our full sample period covering tax years 1991-2007. The second is a measure of the average state audit rate within Schedule C (business) examination classes. This variable is only available for the tax year period from 1998-2007. As in our models of constant law total income reporting, we perform an instrumental variables analysis to control for the potential endogeneity of the audit rate using measures of direct examination time as instruments. Table 13 describes the additional explanatory variables used in the analysis of nonfarm sole proprietor net income reporting.

Table 13: Definitions of Additional Variables Used In Schedule C Net Income Reporting Analysis

Variable	Definition
schcpct	Dependent variable measured as 100 times the ratio of Sch. C net income actually reported on required returns to a CPS measure of Sch. C net income on all required returns
lnbusaudr	Natural log of one plus the average audit rate within Schedule C examination classes
Intotinc	Natural log of total income on required returns
ctotincpct	Sch. C net income on required returns as a percentage of total income
lnburden	Natural log of average taxpayer burden (in dollars) based on the Individual Master File population
agforfishpct	Sole proprietors in agriculture, forestry, fishing, hunting as a percentage of all nonfarm sole proprietors



Variable	Definition
constrpct	Sole proprietors in construction as percentage of all nonfarm sole proprietors
finanpct	Sole proprietors in finance as percentage of all nonfarm sole proprietors
probusservpct	Sole proprietors in professional business services as percentage of all nonfarm sole proprietors
edhlthservpct	Sole proprietors in education and health services as percentage of all nonfarm sole proprietors
leishospervpct	Sole proprietors in leisure and hospitality services as percentage of all nonfarm sole proprietors
trendsq	The square of the value of (tax year-1981)

The full estimation results for all panel data specifications of Schedule C net income reporting behavior are provided in the Task Order 1 Final Report. In Table 14, we focus on two alternative trend effects specifications. The first relies on the overall audit rate and is based on tax years 1991-2007; the second relies on the average audit rate within Schedule C examination classes and is based on tax years 1998-2007. The results are qualitatively fairly similar. For each of the alternative audit rate measures, the estimated coefficient is positive, which is indicative of a deterrence effect. However, the coefficients are imprecisely estimated. Both specifications yield a negative estimated coefficient on the marginal tax rate, suggesting that reporting is negatively impacted by high tax rates. Again, however, the coefficient is imprecisely estimated. The overall level of income as well as the share of sole proprietor income in total income show a significant negative association with Schedule C net income reporting. Reporting is also negatively associated with the unemployment rate in both specifications. Reporting does not show a significant relationship with most industry categories. All else equal, however, sole proprietors engaged in agriculture, forestry, and fishing tend to report a higher share of their income than those engaged in other industries.

Table 14: Trend Effects Specifications of Schedule C Net Income Reporting

	Overall audit rate	Sch. C exam class audit rate
lnaudnw	13.681 (1.53)	
lnbusaudr		10.514 (0.86)
c marg 95 fixed	-21.406 (0.26)	-170.049 (1.72)
lntotinc	-59.436	-64.176



	(6.52)**	(5.48)**
ctotincpct	-13.729	-13.314
	(32.57)**	(24.67)**
depamountpct	0.470	-4.623
	(0.17)	(1.18)
lnburden	4.154	6.111
	(0.77)	(1.08)
paidprep	0.253	-0.295
	(1.16)	(0.99)
singles	-0.695	-0.192
	(2.27)*	(0.53)
under30	0.723	0.569
	(2.16)*	(1.37)
over64	0.099	0.932
	(0.25)	(1.98)*
pcbirths	-1.719	-2.094
	(1.44)	(1.13)
unemplrate	-1.852	-3.698
	(2.82)**	(4.06)**
agforfishpct	0.243	0.275
	(3.05)**	(2.75)**
constrpct	-0.125	0.099
	(1.17)	(0.81)
finanpct	0.050	0.020
	(0.43)	(0.14)
probusservpct	0.123	0.180
	(1.02)	(1.40)
edhlthservpct	0.056	-0.013
	(0.47)	(0.10)
leishospervpct	0.211	-0.022
	(1.47)	(0.13)
trend	-0.657	0.466
	(0.55)	(0.08)
trendlsq	0.165	0.095
	(3.45)**	(0.45)
Constant	1,622.147	1,779.237
	(7.21)**	(5.75)**
N	833	490

Absolute Value of t-Statistics in Parentheses

* p<0.05, ** p<0.01

Unfortunately, the predictive performance of these models is rather poor. For instance, based on the leave-one-out methodology described in Section 2, the average deviation as a percentage of income reported was 9.7 percent for the longer panel analysis using the overall audit rate and 13 percent for the shorter panel analysis using the average Schedule C exam class audit rate. In contrast, the value of this statistic ranges from 2 to 2.8 percent for our specifications of constant law total income reporting. One possible explanation for this finding is that the CPS ASEC measure of net self-employment earnings is a weak proxy for actual earnings. The average value of the our dependent variable (the Schedule C reporting ratio) across all years and states used in estimation is approximately 80 percent, which is high relative to NRP-based measures of voluntary compliance with respect to Schedule C reporting. Second, the CPS ASEC seems to



very substantially undercount the number of returns with a Schedule C filing requirement. For instance, our CPS ASEC tabulations indicate a population of only 11.4 million required 1040 filers with sole-proprietorship income in tax year 2007. In contrast, the Statistics of Income (SOI) tabulations indicate that a much larger number of taxpayers with a filing requirement – 26.6 million – actually completed a Schedule C for that year. Indeed, these findings led us to impute additional net self-employment earnings to the CPS ASEC under Task Order 4. As described in the Research Focus Area 3 report, we based these imputations on an econometric analysis of self-employment income reported on federal individual income tax returns. Since our imputed measure is based on findings from actual tax returns, it is not a fully independent measure of self-employment earnings. As such, it is not very well suited for inclusion as an explanatory variable in a panel data model intended for use in forecasting future taxpayer reporting behavior.

3.2.2 Gross Pensions and IRAs

The Research Focus Area 6 report presents findings from a preliminary analysis of taxpayer reporting of gross pensions and IRAs. We felt that one of the strengths of our analysis was that we had a good independent measure of this income source based on our imputation work under Research Focus Area 3. As with our analysis of constant law total income reporting behavior, our models for gross pension and IRA reporting include specifications in which the dependent variable is in ratio form (ratio of gross earnings actually reported on required returns to our imputed measure of gross earnings on required returns) as well as specifications in which it is in levels form (natural log of gross earnings actually reported on required returns). As discussed above in Section 3.1, we found that the ratio dependent variable specification for constant law total income performed at least as well at forecasting as the level specification. In the case of gross pensions and IRAs, however, the level specification turned out to be far superior. Recently we have extended our analysis of gross pension and IRA reporting to include some additional explanatory variables that relate to the shares of taxpayers who are retired, working, and both retired and working. Table 15 summarizes the variables used in our level specifications. In many cases, these have been constructed by taking the natural log of the numerator or the denominator of explanatory variables used in our ratio specification for constant law total income defined in Table 2.

Table 15: Definitions of Variables Used in Level Specifications

Variable	Definition
lnpenrep	Natural log of gross pension and IRA earnings reported on required returns
lnaud	Natural log of one plus the audit rate of individual returns by agents and auditors/TCO in fiscal year.
det_pct	Percentage of examiners' overall time that was devoted to direct examination activities (used as an instrumental variable).
lndetlag	Natural log of the lagged value of the average direct examination time of revenue agents per individual tax return audited (used as an



Variable	Definition
	instrumental variable).
lnpenimp	Natural log of imputed gross pension and IRA earnings on required returns
c_marg	Federal-state marginal tax rate based on a fixed 1984 distribution of income
lnreturn	Natural log of number of timely filed required returns
lnfthresh	Natural log of aggregate filing threshold among required returns
lndepamt	Natural log of aggregate value of dependent exemption amount among required returns
lnsoleprop	Natural log of number of potential returns with income from a sole proprietorship
lntrfs	Natural log of number of nonfarm sole proprietors in the trade, finance and service industries
lnprep	Natural log of number of required returns completed by a paid preparer
lnsingles	Natural log of number of single or head of household returns
lnunder30	Natural log of number of returns with primary taxpayer under age 30
lnover64	Natural log of number of returns with a primary taxpayer over age 64
pcbirths	Number born per thousand of population
lnexclinc	Natural log of income excluded from federal income taxation (veterans' benefits, workers' compensation, supplemental security income) and as a percentage of total income
unemplrate	Unemployment rate
lncollege	Natural log of number of potential returns with a college-educated primary filer
lnmale	Natural log of number of single and head of household potential returns belonging to males
lnhomeowner	Natural log of number of potential returns belonging to homeowners
popdensity	Population density
trend	Tax year minus 1981

The results of estimation based on our level dependent variable specification are summarized in Table 16. Three separate specifications are employed: trend effects; trend effects with a lagged dependent variable; and trend effects with a lagged dependent variable as well as lags of selected explanatory variables, including the audit rate, the imputed value of gross pensions and IRAs, the federal-state marginal tax rate, and the unemployment rate. The lagged explanatory variables in



the table are assigned the prefix “L.” for ease of identification. In all three specifications, instrumental variables estimation is applied to account for the potential endogeneity of the audit rate using variables `det_pct` and `Indetlag` as instruments. The estimation period for the specification without lagged variables spans tax years 1993-2009. The remaining specifications span one less year (1994-2009) as a result of the lagged dependent variable. The coefficient of the audit rate is positive but statistically insignificant. Surprisingly, the current value of the imputed gross pension and IRA measure is not a statistically significant predictor of reported gross pensions and IRAs, despite being highly correlated with the latter variable. Apparently, our imputed measure loses its explanatory power once the role of other variables, such as the natural log of the filing threshold, is taken into account. When the lagged value of the imputed gross pension and IRA measure is included as an explanatory variable in the dynamic specification, its coefficient is negative and statistically significant. An increase in the lagged value of the imputed measure (`L.lnpenimp`), holding the lagged level of actual reporting (`L.lnpenrep`) constant, would signify a potential increase in pension and IRA underreporting by taxpayers in the prior year. The negative coefficient suggests that this increase in underreporting in the prior period would tend to persist in the current period.

The natural log of the aggregate filing threshold (`lnfthresh`) has a negative and statistically significant relationship with taxpayer reporting behavior. As this threshold increases, fewer taxpayers are required to file returns. The negative coefficient indicates that those who continue to have a filing requirement will tend to report less gross pension and IRA earnings relative to our imputed measure (`lnpenimp`). This seems to suggest that those who are near the filing threshold tend to be more compliant in their reporting of pension income than those who are well above the threshold. The results also indicate that sole proprietors operating in the trade, financial, and service industries; taxpayers who use paid preparers; younger taxpayers; single males; and homeowners all tend to be relatively compliant in their reporting of gross pensions and IRAs. On the other hand, retired taxpayers who continue to hold a job tend to be somewhat less compliant.

Table 16: Estimation Results for Specifications of Models in Levels Form (LnPenRep)

	Trend Effects	Trend Effects with Lagged Dependent Variable	Trend Effects with Lagged Dependent and Selected Explanatory Variables
	b/t	b/t	b/t
<code>L.lnpenrep</code>		0.287*** (8.59)	0.297*** (8.94)
<code>Lnaud</code>	0.182 (1.39)	0.211 (1.61)	0.119 (0.26)
<code>L.lnaud</code>			0.117 (0.34)
<code>lnpenimp</code>	0.008 (0.18)	-0.013 (0.29)	-0.069 (1.48)
<code>L.lnpenimp</code>			-0.090* (2.09)
<code>c_marg</code>	0.257	0.052	4.303*



	Trend Effects	Trend Effects with Lagged Dependent Variable	Trend Effects with Lagged Dependent and Selected Explanatory Variables
	(0.20)	(0.04)	(2.39)
L.c marg			-5.157***
			(3.36)
lnreturn	0.019	0.344	0.230
	(0.07)	(1.18)	(0.76)
lnfthresh	-1.450***	-1.095***	-0.963***
	(5.76)	(4.58)	(3.81)
lndepamt	-0.163	-0.168	-0.176
	(1.46)	(1.58)	(1.59)
lnsoleprop	-0.099	-0.036	-0.111
	(1.83)	(0.67)	(1.76)
lntrfs	1.670***	1.361***	1.475***
	(4.63)	(3.66)	(3.48)
lnprep	0.496***	0.304*	0.350*
	(3.31)	(2.14)	(2.47)
lnsingles	-0.387	-0.227	-0.261
	(1.38)	(0.85)	(0.99)
lnunder30	0.342**	0.297**	0.256*
	(2.94)	(2.72)	(2.35)
lnover64	0.079	0.007	0.038
	(0.68)	(0.06)	(0.34)
pcbirths	0.025*	0.004	0.001
	(1.98)	(0.33)	(0.10)
lnexcline	-0.115	-0.126	-0.135
	(1.58)	(1.81)	(1.90)
unemplrate	-0.006	-0.011	-0.012
	(0.80)	(1.57)	(1.37)
L.unemplrate			0.015
			(1.32)
lncollege	-0.019	-0.029	0.027
	(0.11)	(0.18)	(0.16)
lnmale	0.506***	0.468***	0.510***
	(3.46)	(3.35)	(3.65)
lnhomeowner	0.737***	0.746***	0.712**
	(3.35)	(3.41)	(3.24)
popdensity	-0.001*	-0.001	-0.001
	(2.37)	(1.91)	(1.60)
pctwork	0.001	0.000	0.001
	(0.09)	(0.04)	(0.12)
pctret	0.004	0.006	0.008
	(0.40)	(0.69)	(0.91)
pctworkret	-0.030*	-0.035*	-0.036**
	(2.06)	(2.45)	(2.59)
trend	0.120***	0.087***	0.084***
	(14.83)	(10.17)	(8.73)



	Trend Effects	Trend Effects with Lagged Dependent Variable	Trend Effects with Lagged Dependent and Selected Explanatory Variables
Constant	34.269***	23.818***	23.301***
	(6.88)	(4.82)	(4.64)
N	833	784	784

Absolute Value of t-Statistics in Parentheses

* p<0.05, ** p<0.01, *** p<0.001

Table 17 summarizes the forecasting performance of our three alternative level dependent variable specifications. The results indicate that the inclusion of a lagged dependent variable tends to improve forecasting performance. Performance is further improved when lagged values of selected explanatory variables are also included in the specification. Although our specifications based on a level dependent variable substantially outperform those based on a ratio specification, even the former specifications are not especially accurate in their forecasts.

Table 17: Forecasting Performance for Specifications of Models in Levels Form

Last Sample Year	Forecast Year	Static		Lagged Dependent Variable		Lagged Dependent and Selected Regressors	
		% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE
2006	2007	8.83	13.59	8.68	13.74	7.69	12.76
	2008	7.32	11.20	5.81	8.65	5.34	8.00
	2009	13.99	21.60	10.07	15.89	8.90	14.39
2007	2008	6.23	9.44	6.04	8.94	5.59	8.17
	2009	11.50	17.54	8.90	14.55	8.10	13.99
2008	2009	9.84	15.70	8.35	13.60	7.70	13.16

In the case of our models of Schedule C net income reporting, we attributed a poor forecast performance to the lack of a good independent measure of net self-employment income. In the case of gross pension and IRA earnings, we do have a good independent measure of earnings. Unfortunately, though, this measure does not appear to be strongly correlated with reporting behavior after accounting for other explanatory factors. The relatively weak forecast performance in the case of gross pension and IRA reporting may be attributable to a fairly low “signal to noise” ratio. In particular, changes in state level pension reporting that are attributable to random factors (noise) may swamp changes that are attributable to compliance trends. As well, although our imputed measure of pension income is likely to be reasonably accurate at the national level, it may be subject to nontrivial measurement errors at state level. This is because the CPS ASEC data used as the base for our imputations has a relatively small sample size at the state level. In addition, our econometric imputation approach does not directly account for the

distribution of IRS Form 1099-R pension and IRA records across states. Such measurement errors could adversely impact forecast performance.

3.3 Filing Compliance

In addition to exploring the potential of panel data econometric techniques to forecast federal taxpayer reporting behavior at the state level, we have undertaken an analysis of their usefulness in explaining and predicting filing compliance. The results of a preliminary investigation of this issue were discussed in the Research Focus Area 4 report. Recently, we have extended this analysis to evaluate the explanatory power of some additional regressors and to compare the forecasting performance of some static and dynamic specifications. In this section, we focus on these more recent findings.

We consider two alternative specifications of the dependent variable for our analysis. The first is a measure of the filing rate, defined as 100 times the ratio of the number of timely filed required returns to the total number of required returns. The filing rate is a key measure of filing compliance that is tracked over time by the IRS, and it is comparable to the dependent variable used in the Plumley (1996) analysis. The second specification of the dependent variable is the natural log of the number of timely filers. Models using this version of the dependent variable include the number of required returns as an explanatory variable.

The explanatory variables in our analysis of filing compliance include many of the same regressors used in our analysis of constant law total income reporting, such as measures of the filing threshold percentage, and various socio-economic factors. However, certain explanatory variables for total income reporting are less relevant for the filing decision and have therefore been excluded. These include the combined federal-state marginal tax rate, the audit rate, the percentage of filed returns prepared by a paid tax practitioner, the birth rate, the ratio of the value of the dependent exemption to total personal income, and the share of total personal income that is excluded from taxation. At the same time, we include some new explanatory variables that seem potentially relevant to filing compliance. These include the natural log of the taxpayer delinquency investigation (TDI) rate, a dummy variable for state tax amnesty, measures of real average personal income and the rate of growth rate in real personal income, the natural log of taxpayer burden, and proxy measures of earned income tax credit (EITC) eligibility constructed from the CPS ASEC data. Taxpayer delinquency investigations are a potential deterrent to filing noncompliance. State level tax amnesties may tend to improve filing compliance, at least in the short run. Average real personal income serves as a proxy for taxes owed as well as taxpayer prosperity. The growth rate of average real personal income serves as a business cycle measure. A relatively high burden of preparing and filing a return may serve to discourage filing compliance. Finally, households that are eligible for the EITC may be relatively more prone to comply with their filing requirement.⁹

⁹ Recall that the analysis focuses on taxpayers with a legal filing obligation; taxpayers who file solely to claim the EITC are excluded from the analysis.



We have also experimented with some explanatory variables that are specific to taxpayers whose third-party matchable income is only modestly above the filing threshold. Such taxpayers are relatively less likely to face enforcement if they do not file a return, and they may be more sensitive to a variety of factors when deciding whether to file a return compared to the general population. These explanatory variables are calculated over the set of returns that would have a tax liability of less than \$200 based solely on specified third-party matchable sources of income (specifically, interest, dividends, alimony, pensions and IRAs, social security, and unemployment compensation), the relevant standard deduction, personal exemptions, and the tax rate in the lowest bracket for the return. Observe that wages are not included among the third-party matchable income sources under this criterion; it is assumed that the amount of tax withheld by employers on wage income fully offsets any tax liability. This criterion for being close to the filing threshold is one of four alternative criteria discussed in Section 4 of the *Data User's Guide*. The additional variables employed in our analysis of the filing decision are defined in Table 18.

Table 18: Definitions of Additional Variables Used In Filing Compliance Analysis

Variable	Definition
filingrate	Dependent variable measured as 100 times the ratio of timely filed required returns to total required returns*
Innumfilers	Natural log of the number of timely filers
Inrcquired	Natural log of the number of required returns
Intdirate	Natural log of 100 times the ratio of taxpayer delinquency investigation notices to the number of potential returns
amnesty5	Indicator of a state tax amnesty in the past 5 years
avgpi	Average real (2005) personal income
avgpigrowth	Annual growth rate for real average personal income
Inavpi	Natural log of avgpi
Inavburden	Natural log of average real (2007) taxpayer burden
pctcitclignok	Percentage of required returns with no qualifying children that are eligible for the EITC
pcteitcligk	Percentage of required returns with qualifying children that are eligible for the EITC
lowriskpctr2	Percentage of all required returns with third-party matchable income close to filing threshold
Inavburdenlr2	Natural log of average taxpayer burden among required returns with



Variable	Definition
	third-party matchable income close to filing threshold
pctund30lr2	Share of required returns with third-party matchable income close to filing threshold that have a primary taxpayer under age 30
pctovr64lr2	Share of required returns with third-party matchable income close to filing threshold that have a primary taxpayer over age 64
pctmarlr2	Share of required returns with third-party matchable income close to filing threshold that have married joint filing status
pcthscolllr2	Share of required returns with third-party matchable income close to filing threshold that have a primary taxpayer with at least a high school diploma
pctnonmatlr2	Share of required returns with third-party non-matchable income close to filing threshold that have income from sources subject to little or no third-party information reporting, including our CPS ASEC measures of farm income, rental income, and imputed nonfarm self-employment income

*This variable represents a new dependent variable for this analysis; however, it previously served as an explanatory variable in the analysis of taxpayer reporting behavior.

3.3.1 Estimation Results for Filing Rate

We have estimated both trend effects and two-way fixed effects models of the filing rate. This includes both static specifications and dynamic specifications. The dynamic specifications include the lagged value of the filing rate as an explanatory variable as well as the lagged value of the natural log of the TDI rate. We estimate the dynamic models using two alternative approaches. Under the first approach, we employ our standard trend effects or two-way fixed effects estimation method using a specification that includes the relevant lagged variables. As discussed in Section 2, this approach yields biased estimates; however, the magnitude of the bias decreases with the number of time periods in estimation. Our dynamic models are estimated over periods of 15 (tax years 1995-2009) to 18 years (tax years 1992-2009), depending on whether the TDI rate is included as a regressor (the TDI rate is only available for tax years 2003-2009) and on the choice of estimation method. Whether these periods are sufficiently long to render the bias negligible is uncertain. As an alternative, we employ Arellano-Bond estimation, which is a consistent estimation procedure even in short panels. For each of our models, we begin by including a full set of potential explanatory variables for the filing rate. We then employ F-tests to eliminate variables that are not statistically significant predictors. In this section, we report the final specifications after insignificant regressors have been eliminated.



3.3.1.1 Static Model Estimation Results for Filing Rate

The estimation results for our static models of the filing rate are presented in Table 19. Overall, the estimated coefficients of the common explanatory variables are similar across the two specifications. The coefficient of the filing threshold percentage variable is negative, suggesting that households with third-party matchable income that is fairly close to the filing threshold tend to be relatively more likely to file than those with higher matchable income amounts. This finding is reinforced by the negative estimated coefficient on average real personal income (avgpi). It seems plausible that many taxpayers with relatively low income are inclined to file to receive a refund, while those with higher levels of income are sometimes deterred from filing by the prospect of having significant taxes due. Single individuals tend to be relatively less likely to comply with their filing obligation. Consistent with the finding of Erard and Ho (1991), the burden of preparing and filing a return tends to serve as a deterrent to filing for those near the filing threshold, although the effect is not precisely estimated in the two-way fixed effects model. Interestingly, among the group of taxpayers with third-party matchable income close to the filing threshold, those who have additional income from sources subject to little or no third-party information reporting actually tend to be more prone to file a return. Of course, the overall income (matchable and unmatchable) of such taxpayers is further over the filing threshold than the overall income of those in the group with no sources of unmatchable income. Consequently, the duty to file a return may be more apparent to those with unmatchable income.

Table 19: Static Model Estimation Results

	Two-Way Fixed Effects	Trend Effects
	b/t	b/t
fthresholdpct	-3.407*** (16.09)	-2.980*** (14.77)
soleprops	-0.368 (1.54)	
soleproptfs	0.699 (1.77)	
singles	-0.491*** (8.77)	-0.781*** (10.97)
under30		0.103 (1.71)
avgpi	-0.147** (3.02)	-0.122* (2.43)
avgpigrowth		0.071* (2.32)
lowriskpctlr2	0.166** (2.98)	-0.189*** (3.69)
lnavburdenlr2	-5.049 (1.49)	-14.672*** (5.83)
pctund30lr2	-0.072 (1.75)	-0.115* (2.23)
pctmarlr2		-0.154** (3.02)



	Two-Way Fixed Effects	Trend Effects
pctnonmatlr2	0.104* (2.18)	0.151*** (4.51)
trend		-0.128* (2.48)
Constant	205.531*** (9.50)	297.327*** (16.40)
N	931	931

Absolute Value of t-Statistics in Parentheses
 * p<0.05, ** p<0.01, *** p<0.001

3.3.1.2 Dynamic Model Estimation Results for Filing Rate

Table 20 presents the results from our dynamic specifications of filing compliance, which include lagged values of the filing rate and the taxpayer delinquency investigation rate as explanatory variables. Ignoring for a moment the lagged explanatory variables, the estimated coefficients in the dynamic specifications are qualitatively similar to the static results in Table 19. All else equal, states with a high filing rate in one year tend to have a high filing rate in the subsequent year. The results for the TDI rate are somewhat difficult to interpret. In the case of the Arellano-Bond specification that includes year dummies, the results indicate that delinquency investigations are a deterrent to filing noncompliance. This seems intuitive. However, in the trend effects specification as well as the Arellano-Bond specification that includes a trend term, the current TDI rate is positively associated with the filing rate, but the lagged TDI rate is negatively associated with the filing rate, leaving the overall impact uncertain.¹⁰ As we found earlier in our analysis of reporting behavior, a trend term may not account as fully for year-specific factors as year dummies, which can lead to counter-intuitive estimates of the impact of enforcement on taxpayer behavior.

Table 20: Dynamic Model Estimation Results

	Two-Way Fixed Effects	Trend Effects	Arellano-Bond with Year Dummies	Arellano-Bond with Trend Term
	b/t	b/t	b/t	b/t
L.filingrate	0.246*** (8.85)	0.265*** (9.24)	0.223*** (4.47)	0.256*** (5.10)
Intdi		0.827*** (4.31)	0.413* (2.31)	1.000*** (4.99)
L.Intdi		-1.100*** (5.28)		-1.062*** (6.87)

¹⁰ The TDI rate is potentially endogenous, because the rate is likely to be greater in states with more filing noncompliance. Our dynamic specifications address the potential endogeneity by using lagged values of the TDI rate as instruments. However, the static specifications do not correct for potential endogeneity.



	Two-Way Fixed Effects	Trend Effects	Arellano-Bond with Year Dummies	Arellano-Bond with Trend Term
fhresholdpct	-3.079*** (15.05)	-3.086*** (15.18)	-4.025*** (9.79)	-3.319*** (9.55)
singles	-0.445*** (8.54)	-0.536*** (9.49)	-0.502*** (7.23)	-0.546*** (8.24)
under30		0.172** (3.10)	0.118* (2.17)	0.275*** (4.48)
over64		0.155* (2.04)	0.090 (1.47)	0.254*** (3.49)
avgpi	-0.233*** (5.09)	-0.244*** (4.79)	-0.225* (2.41)	-0.294** (3.05)
avgpigrowth	0.146*** (4.35)	0.214*** (7.15)		0.205*** (3.67)
lowriskpctr2	0.161** (3.11)	0.002 (0.04)	0.186** (2.64)	0.027 (0.55)
lnavburdenlr2		-14.903*** (6.48)	-5.436 (1.60)	-16.693*** (8.68)
pctund30lr2	-0.092* (2.46)	-0.096 (1.93)	-0.115* (2.48)	-0.188*** (3.86)
pctovr64lr2		-0.142* (2.27)		-0.206* (2.44)
pcthscollr2		0.069 (1.77)		
pctnonmatlr2		0.130*** (4.16)	0.086 (1.94)	0.141*** (4.40)
trend		0.033 (0.60)		0.115 (1.50)
N	882	784	784	735

Absolute Value of t-Statistics in Parentheses

* p<0.05, ** p<0.01, *** p<0.001

3.3.1.3 Results for Levels Specifications

As an alternative to employing the filing rate as the dependent variable, we have also estimated some models that use the natural log of the number of timely filers as the dependent variable. In these models, we include the natural log of the total number of required returns as an explanatory variable. We have estimated both static and dynamic trend effects and two-way fixed effects specifications. Our dynamic specifications employ the Arellano-Bond estimation approach. The estimation results are presented in Table 21.

Table 21: Estimation Results for Levels Specifications

	Static Two-Way Fixed Effects	Static Trend Effects	Arellano-Bond with Year Dummies	Arellano-Bond with Trend Term
	b/t	b/t	b/t	b/t
L.Infilers			0.457*** (6.57)	0.566*** (10.01)
Intdi				0.008*** (4.22)
L.Intdi				-0.005*** (4.16)
lnrequired	0.881*** (54.22)	0.849*** (49.24)	0.389*** (6.70)	0.290*** (5.50)
fthresholdpct	-0.039*** (13.82)	-0.035*** (13.84)	-0.013** (3.10)	-0.007* (2.11)
soleprops				0.007** (2.60)
soleproptfs				-0.012* (2.35)
singles	-0.005*** (8.37)	-0.006*** (11.51)	-0.002** (2.80)	-0.002*** (3.78)
under30		0.002** (3.02)		0.002*** (4.10)
over64		0.002* (2.21)		
lnavpi	-0.231*** (5.40)	-0.214*** (5.04)		
unemplrate		-0.002* (2.04)		-0.006*** (6.41)
amnesty5				0.004 (1.90)
pcteitceligk				-0.002* (2.36)
lowriskpctr2	0.002*** (3.38)		0.001* (2.09)	
lnavburdenlr2		-0.115*** (4.41)		-0.076*** (4.02)
pctund30lr2		-0.001** (2.69)		
pctovr64lr2		-0.001* (1.97)		
pctnonmatlr2	0.001* (2.32)	0.001** (3.25)		
trend		0.002*** (3.82)		0.002** (3.13)
Constant	-3.581*** (13.44)	-2.475*** (8.28)		



	Static Two-Way Fixed Effects	Static Trend Effects	Arellano-Bond with Year Dummies	Arellano-Bond with Trend Term
N	931	931	833	735

Absolute Value of t-Statistics in Parentheses

* p<0.05, ** p<0.01, *** p<0.001

Overall, the estimation results for the specifications of the natural log of the number of timely filers are reasonably comparable to those of the filing rate. As expected, the overall number of required returns is positively and strongly associated with the filing rate in Table 21.

3.3.2 Forecasting Performance

Table 22 summarizes the step-ahead performance in forecasting the number of timely filed required returns using our alternative static and dynamic models of the filing rate. The two-way fixed effects specifications are excluded, because it is difficult to predict the value of the year effects outside of the estimation period. Overall, the dynamic model outperforms the static specification, especially when Arellano-Bond estimation is employed.

Table 22: Forecasting Performance of Filing Rate Specifications

Last Sample Year	Forecast Year	Static Model		Dynamic Model			
		Trend Effects		Trend Effects		Arellano-Bond with Trend Term	
		% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE
2006	2007	4.66	7.88	3.96	6.67	3.76	6.35
	2008	3.59	6.35	2.41	4.45	2.11	3.82
	2009	3.97	7.90	3.02	6.42	2.66	5.70
2007	2008	2.31	4.16	1.28	2.35	1.16	1.80
	2009	2.70	5.60	1.91	4.31	1.68	3.65
2008	2009	2.38	4.98	1.86	4.18	1.70	3.74

The forecast accuracy for tax year 2007 is relatively weak across all specifications when the last sample year used in estimation is tax year 2006. This is presumably due to the aforementioned spike in filing activity in tax year 2007 as a result of the Economic Stimulus Act of 2008. The Act may be thought of as a national level “intervention” or a “natural experiment”. In Table 23, some preliminary estimates of the impact of this intervention on filing compliance are obtained by comparing the actual national filing rates in tax years 2007, 2008, and 2009 to the forecasted rates based on the three alternative models. The estimated tax year 2007 improvement in filing compliance in response to the stimulus benefit is large, ranging from 3.6 to 4.4 percentage points. Although the filing rate declines over the following two years, it remains substantially higher than predicted on the basis of historical data. This finding is consistent with the



conclusion of Erard and Ho (2001) that there is substantial persistence in filing behavior over time.

Table 23: Estimated Impact of Economic Stimulus Act of 2008 on Filing Compliance, Filing Rate Specifications

Tax Year	Actual Filing Rate	Filing Rate Forecast			Preliminary Estimate of Impact of Economic Stimulus Act of 2008		
		Static Trend Effects	Dynamic Trend Effects	Arellano-Bond with Trend	Static Trend Effects	Dynamic Trend Effects	Arellano-Bond with Trend
2007	95.7	91.3	91.9	92.1	4.4	3.8	3.6
2008	94.4	91.1	92.2	92.5	3.3	2.2	1.9
2009	92.8	89.2	90.0	90.4	3.6	2.8	2.4

Table 24 summarizes the step-ahead performance in forecasting the number of timely filed required returns using our alternative static and dynamic models of the natural log of the number of timely filed returns. The two-way fixed effects specifications are again excluded, because it is difficult to predict the value of the year effects outside of the estimation period. Once again, the dynamic model outperforms the static specification. Comparing the dynamic model results in Tables 22 and 24, neither the filing rate specification nor the filing levels specification appears to be consistently superior in its forecasting performance.

Table 24: Forecasting Performance of Filing Levels Specifications

Last Sample Year	Forecast Year	Static Trend Effects		Dynamic Arellano-Bond with Trend Term	
		% Abs Dev	CV of RMSE	% Abs Dev	CV of RMSE
2006	2007	4.02	6.72	2.83	4.35
	2008	3.29	5.85	1.70	2.99
	2009	4.36	8.24	3.16	5.64
2007	2008	2.17	4.11	0.82	1.36
	2009	3.29	6.51	2.34	4.11
2008	2009	2.63	5.37	2.07	3.72



Like our filing rate specifications, our filing levels specifications tend to under-predict the extent of filing in tax years 2007-2009. In Table 25 we present the forecast errors for the filing rate as preliminary estimates of the impact of the Economic Stimulus Act of 2008 on filing compliance. Overall, these estimates tend to be lower than those shown in Table 23 for our filing rate specifications, but they continue to imply a substantial and persistent impact.

Table 25: Estimated Impact of Economic Stimulus Act of 2008 on Filing Compliance, Filing Levels Specifications

Tax Year	Actual Filing Rate	Filing Rate Forecast		Preliminary Estimate of Impact of Economic Stimulus Act of 2008	
		Static Trend Effects	Arellano-Bond with Trend	Static Trend Effects	Arellano-Bond with Trend
2007	95.7	91.9	93.0	3.8	2.7
2008	94.4	91.3	92.9	3.0	1.5
2009	92.8	88.8	89.9	4.0	2.9

4 Conclusions and Recommendations

The main findings of this research are as follows:

- Overall, it appears to be more difficult to forecast taxpayer reporting for a specific income source, such as net nonfarm self-employment income or pensions and IRAs, than for overall income. In the case of nonfarm self-employment income, a challenge is the lack of a high quality independent measure of the incidence and magnitude of this income source within the overall population against which to compare net earnings reported on tax returns. In the case of pensions and IRAs, a challenge is the relatively high degree of voluntary compliance in the reporting of this income item. Many of the observed variations in the reporting of this income source over time and across states are likely due to random factors rather than changes in compliance behavior.
- Longer panels provide more degrees of freedom for estimation and prediction; however, there are also some drawbacks. For instance, if there have been structural changes in taxpayer reporting or filing activities over the estimation period, earlier time periods in the sample may not have much predictive value for recent years. As well, shorter panels permit the inclusion of potentially important explanatory variables that are not available over a longer time span. An advantage of a sufficiently long panel is that the bias associated with dynamic fixed effects estimation approaches tends to be relatively small. However, alternative dynamic estimation techniques such as Arellano-Bond and Anderson-Hsiao produce consistent parameter estimates in shorter panels.

- For interpreting the role of certain key factors in overall reporting behavior (such as the audit rate and federal-state marginal tax rate), a two-way fixed effects model seems to be preferable; trend effects and random effects models sometimes produce counter-intuitive predictions regarding the marginal effects of such factors. However, it is challenging to predict fixed year effects outside of the estimation sample under the two-way fixed effects approach, which limits its usefulness for forecasting future taxpayer behavior.
- The best choice of model for forecasting purposes is somewhat case dependent. Among static models, the trend effects model tends to perform reasonably well in reasonably long panels. However, the exhaustion of degrees of freedom in estimating state level fixed effects can hamper its performance in shorter panels. In shorter panels, a random effects model tends to perform reasonably well. This is interesting, because a Hausman test consistently rejects random effects in favor of fixed effects in our models of overall taxpayer reporting behavior. In the case of dynamic models, trend effects estimation or bias-corrected trend effects estimation sometimes works quite well, at least in reasonably long panels. However, in other cases, the forecasting performance of the Anderson-Hsiao or Arellano-Bond approaches is superior.
- The choice of functional form in panel data specifications can be very important for forecast quality. While a ratio specification for the dependent variable performed reasonably well in our models of overall taxpayer reporting behavior (constant law total income as a share of total personal income) and filing behavior (the filing rate), a levels specification (natural log of pensions and IRAs reported) performed much better in our model of pension and IRA reporting.
- In applications where limited explanatory variables are available, an ARIMA or ARIMAX forecasting procedure represents a viable alternative to panel data econometric methods.
- Estimates of forecast performance can be sensitive to the choice of forecast period. In many cases, we found that our models performed relatively poorly in predicting taxpayer reporting and filing outcomes in tax year 2007 based on estimation sample that ended in tax year 2006. Taxpayer reporting and filing behavior for tax year 2007 was atypical as a result of behavioral responses to the Economic Stimulus Act of 2008. This illustrates that forecasting models tend to perform better over relatively calm periods that are not characterized by sudden changes in the policy or economic conditions. It also provides an illustration of how these prediction methods can be applied to estimate the impact of an intervention (even if in a “natural” experiment).

We offer recommendations encompassing the following four topic areas: (1) further specification analysis and testing; (2) alternative data sources and software programs; (3) future applications of the methodology; (4) potential for micro-econometric approaches.

4.1 Further Specification Analysis and Testing

One issue that has not been fully resolved is the endogeneity of certain explanatory variables and the best instruments to employ for them. As discussed in Section 3.1.3, there is strong evidence



that the audit rate is endogenous in our models of taxpayer reporting behavior. One of the direct examination time variables (det_pct) seems to be a consistently good instrument for the audit rate across specifications. However, the suitability of the other variable (detlag) is more questionable. It would be desirable to explore some specifications that include separate audit rate measures for examinations by tax examiners and revenue agents. However, multiple instruments would be required in this case. The evidence suggests that the filing rate and/or the federal-state marginal tax rates are also endogenous explanatory variables in our reporting equations. We have identified some instruments for these variables that seem suitable. It would be useful to estimate some specifications of taxpayer reporting that simultaneously treat the audit rate, the filing rate, and the federal-state marginal tax rate as endogenous regressors and to work out a suitable methodology for forecasting future reporting behavior in the presence of multiple instruments. The endogeneity of certain other variables should also be resolved. For instance, the TDI rate is a potentially endogenous variable in our models of filing behavior. As well, taxpayer services are potentially endogenous regressors in models of reporting or filing behavior. Unfortunately, it is not clear that there are good instruments available for these variables. One possibility that warrants examination would be to use lagged values of these variables as instruments.

There is also scope for more panel data analysis of taxpayer filing compliance. It would be useful to estimate some shorter panels using some additional explanatory variables relating to service (taxpayer help line and IRS staff hours in mission critical service activities) and enforcement (document matching and refund offsets) that are not available for the full sample period. A second extension would be to test alternative definitions of taxpayer burden as well as the explanatory variables relating to taxpayers who are close to the filing threshold. A third extension would be to examine the determinants of late filing by taxpayers with a filing requirement. Finally, it would be desirable to undertake a more comprehensive evaluation of the impact of the Economic Stimulus Act of 2008 on taxpayer filing compliance.

In the case of overall taxpayer reporting behavior, it would be interesting to refine the ratio dependent variable to make it a better proxy for tax compliance. In particular, instead of using the excluded income percentage as an explanatory variable, one could subtract excluded income from personal income in the denominator of the reporting ratio measure that is currently being used as the dependent variable. Another extension would be to use a broader overall income measure than the current constant law concept. For the tax year period 1991-2009, the definition of income subject to taxation has changed very little; the main factor to account for is changes in the rules (beginning in tax year 1994) for the share of social security income that is subject to taxation.

4.2 Alternative Data Sources and Software Programs

Two important sources of demographic and income information in our data base are the CPS ASEC and S●I. One difficulty with these data sources is that the sample size for some states (especially smaller ones) is rather modest. As a consequence, our measures of state level taxpayer characteristics tend to be subject to significant sampling error. This can result in counter-intuitive values for certain variables, such as estimated filing rates in excess of one in certain states and time periods. Overall, the modest sample sizes add “noise” to the analysis, which may adversely impact forecasting performance. As noted earlier, we adjust the weights on

the SOI data so that the number of required returns in each state equals the number of required returns actually filed (per the Master File).

An alternative to the CPS ASEC for demographic and income information is the American Community Survey (ACS). This data source has a much larger sample size (several million addresses compared to about 100,000 in the CPS ASEC). It does not have as much detail on income sources, but it does contain key measures including wages, interest, dividends, net rental income, self-employment earnings, social security benefits, SSI, retirement income, and other income. In addition to having less sampling error at the state level, the ACS provides the potential to perform a more disaggregated panel data analysis, for instance by metropolitan area. This may prove useful when attempting to use panel data techniques to account for differences in a control and treatment group from different metropolitan areas in a field experiment. Unfortunately, however, the first year of full implementation of the survey was 2005. Thus, there is currently only a limited number of years of data for use in a panel data analysis. However, the ACS appears to have promise as an excellent data source to build from as more years of data are added over time.

The processing of CPS ASEC data to create return unit measures of variables is currently managed using SAS software programs. These programs are effective, but they are not user-friendly. RAS users are likely to find the programs somewhat difficult to modify and refine; in addition, it is somewhat cumbersome to employ the software to compile additional years of CPS ASEC data. It may be worthwhile to develop a more user-friendly software application to support further development of the data base.

4.3 Future Applications of the Methodology

State level panel data econometric methods do not seem to be well suited for directly predicting the effects of taxpayer services on tax compliance. Panel data on IRS and third-party services provided to taxpayers are limited, and the existing measures are potentially endogenous (since taxpayers who elect to use these services are likely to be different in important ways from taxpayer who do not). In addition, attempts to include service measures in panel data specifications have often yielded statistically insignificant or counter-intuitive estimates of their impact on taxpayer behavior. Panel data econometric methods show more promise as a tool for predicting what future taxpayer behavior would be in the absence of an intervention. Consider, for example, a field experiment in which a treatment group (perhaps taxpayers in a selected geographic region) is provided with a service intervention, while a control group (perhaps taxpayers in a similar but distinct geographic region) is not offered the intervention. Panel data econometric techniques may prove useful in accounting for inevitable differences between the control group and the treatment group that would cause treatment group behavior to differ from control group behavior even in the absence of the intervention. Research on ways to integrate panel data and field experiment techniques seems warranted.

4.4 Potential for Microeconomic Approaches

Under this project, we have investigated the potential for understanding and predicting taxpayer reporting and filing behavior at the state level using panel data econometric techniques. An alternative approach would be to apply microeconomic methods in an attempt to understand and predict individual taxpayer reporting and filing behavior. If the focus is on understanding the role of taxpayer services in tax compliance, a first step in this direction would be to undertake a micro-level analysis of the demand for IRS and third-party services. This would require a representative data sample containing information on various services employed by taxpayers and potential determinants of service usage. Some elements of taxpayer service usage are measured by the IRS (such as third-party tax preparation and calls to the help-line for account-related activities). However, taxpayer usage of many services is not captured at the individual level. Thus, a survey would seem necessary to obtain information on the full range of service offerings used by different taxpayers. Ideally, this survey would be linked to tax records to permit a broader understanding of taxpayer characteristics. While a cross-sectional data source would provide useful insights into taxpayer service demand, even greater insights into service usage would be obtained from a longitudinal study (including dynamic issues such as repeat usage of services and changes in demand resulting from changes in circumstances or tax policy).

If the micro-level data source also included tax return information, one could attempt to estimate the impact of various services on tax reporting behavior. However, this would be a significant challenge owing to the endogeneity of service demand. It is likely to be difficult to construct good instruments for service usage that meet the requirement of being independent of taxpayer reporting behavior. With a longitudinal data source, one might consider employing lagged measures of service usage as instruments. However, these instruments would be valid only if the lagged measures did not belong as explanatory variables in the first place. These lagged measures might be important explanatory variables if the receipt of taxpayer services only impacts taxpayer behavior gradually over a period of years. A further consideration is that one taxpayer's use of services may influence the behavior of other taxpayers who do not directly use that service. For instance, the user may share information learned from the service experience with other taxpayers, who may then change their own reporting or filing behavior. Alternatively, knowledge about services offered by the IRS may impact general taxpayer perceptions of how fair and responsive the organization is, which may ultimately feed into taxpayer reporting and filing decisions. It would be difficult to capture these indirect effects in a microeconomic analysis.

Overall, microeconomic techniques seem well suited for understanding what drives the demand for IRS and third-party taxpayer services. However, it is likely to be difficult to apply such techniques to learn directly about the role of taxpayer services in tax compliance. Field experiments (and "natural experiments") seem like a more promising approach to measuring the impact of services on compliance. Depending on the experiment, micro-level or state level panel data econometric techniques may prove useful in controlling for differences in control and treatment groups that would potentially impact inferences from the experiment.

5 References

- Allingham, M.G. and A. Sandmo (1972) "Income Tax Evasion: A Theoretical Analysis", *Journal of Public Economics* 1, 323-338.
- Alm, J. and M. Yunis (2009) "Spatiality and Persistence in U.S. Individual Income Tax Compliance", *National Tax Journal* 62(1), 101-124.
- Anderson, T.W. and C. Hsiao (1981) "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association* 76(375), 598-606.
- Anderson, T.W. and C. Hsiao (1982) "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics* 18, 47-82.
- Arellano, M., and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies* 58, 277-297.
- Baltagi, B. H. (2008) *Econometric Analysis of Panel Data*, 4th Ed., New York: Wiley.
- Balestra, P., and J. Varadharajan-Krishnakumar (1987) "Full Information Estimation of a System of Simultaneous Equations with Error Component Structure," *Econometric Theory* 3, 223-246.
- Bruno, G.S.F. (2005) "Estimation and Inference in Dynamic Unbalanced Panel Data Models with a Small Number of Individuals," CESPRI Working Paper Number 165, Università Bocconi-CESPRI, Milan.
- Bowerman, B. L., and R. T. O'Connell (1993) *Forecasting and Time Series: An Applied Approach*, 3rd Ed., Pacific Grove, CA: Duxbury/Thomson Learning.
- Box, G. E. P., G.M. Jenkins, and G. C. Reinsel (1994) *Time Series Analysis: Forecasting and Control*, 3rd Ed., Englewood Cliffs, NJ: Prentice-Hall.
- Chatfield, C. (2004) *The Analysis Of Time Series: An Introduction*, 6th Ed., Boca Raton, FL: Chapman & Hall/CRC.
- Dubin, J.A. (2007) "Criminal Investigation Enforcement Activities and Taxpayer Noncompliance", *Public Finance Review* 35, 500-529.
- Dubin, J.A., Graetz, M.J., and Wilde, L.L. (1990) "The Effect of Audit Rates on the Federal Individual Income Tax, 1977-1986", *National Tax Journal* 43(4), 395-409.
- Elliott, G., T. J. Rothenberg, and J. H. Stock (1996) "Efficient Tests for an Autoregressive Unit Root", *Econometrica* 64, 813-836.
- Erard, B. (1993) "Taxation with Representation: An Analysis of the Role of Tax Practitioners in Tax Compliance", *Journal of Public Economics* 52(2), 163-197.

- Erard, B. and Ho, C.-C. (2001) “Searching for Ghosts: Who Are the Nonfilers and How Much Tax Do They owe?”, *Journal of Public Economics* 81(1), 25-50.
- Hamilton, J.D. (1994) *Time Series Analysis*. Princeton: Princeton University Press.
- Kiviet, J. F. (1995) “On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models”, *Journal of Econometrics* 68, 53–78.
- Kmenta, J. (1997) *Elements of Econometrics*, 2nd Ed., Ann Arbor, MI: University of Michigan Press.
- Hausman, J. A. (1978) “Specification Tests in Econometrics”, *Econometrica* 46, 1251–1271.
- Hsiao, C. (2003) *Analysis of Panel Data*, 2nd Ed., Cambridge: Cambridge University Press.
- Plumley, A.H. (1996) “The Determinants of Individual Income Tax Compliance: Estimating the Impact of Tax Policy, Enforcement, and IRS Responsiveness”, *Internal Revenue Service, Publication 1916 (Rev. 11-96)*, Washington, DC.
- Roodman, D. (2008) “How To Do Xtabond2: An Introduction to Difference And System Estimation in Stata”, Center For Global Development Working Paper Number 103, December.
- Wiseman, R.M. (2009) “On the Use and Misuse of Ratios in Strategic Management Research”, in *Research Methodology in Strategy and Management*, Volume 5, Ed. by D.D. Berg and D.J. Ketchen, Jr., pp. 75-110, Bingley, U.K.: Emerald Group Publishing, Ltd.
- Wu, D.-M. (1974) “Alternative Tests of Independence Between Stochastic Regressors and Disturbances: Finite Sample Results”, *Econometrica* 42, 529–546.

Appendix A: Referenced Project Reports

A.1 Task Order 1 Final Report

This report discusses all aspects of the project's initial phase. It begins with the compilation of the database and a summary of the literature. Econometric specifications of state-level income reporting compliance update the earlier research by Plumley (1996) to include a variety of sensitivity tests. The preliminary econometric analysis explores several approaches: fixed versus random effects; specifications with ratio dependent and explanatory variables versus alternative functional forms; short versus long panels; and year dummies versus trend terms. A novel methodology is developed for forecasting and evaluation with two-way (individual unit and year) fixed effects panel data econometric models that employ instrumental variables. In addition to specifications of overall income reporting, we explore specifications of net income reporting by sole proprietorships.

The results confirm that it is possible to develop reasonably good forecasts of what overall state level income reporting behavior would be in the absence of a major innovation, such as a significant change in service level or quality. Most of the results prove to be reasonably robust against alternative modeling approaches. However, the substitution of trend terms in place of year dummies in the fixed effects model yields conflicting results with regard to the impact of audit rates and marginal tax rates, indicating that the yearly fixed effects are able to pick up certain state-invariant changes over time that trend terms are unable to capture.

Our analysis of the impact of some new explanatory variables produces some mixed results. For IRS hours of service work, the results are counter-intuitive, suggesting that greater service hours result in lower overall income reporting, while call attempts show no significant relationship with income reporting. Tax complexity and taxpayer burden are, perhaps surprisingly, positively associated with the degree of overall income reporting. Generally we obtain a positive and sometimes statistically significant estimate of the impact of audit rates on taxpayer reporting behavior in models where individual year effects are estimated. Significant relationships are found for a variety of new explanatory variables in overall income reporting, including measures of college attendance, homeownership, male-headed households, population density, and the state level Gini coefficient (a measure of income inequality).



PredictingBehav TO1
- Final Report - 200d

A.2 Research Focus Areas 1 (Updated Data) and 2 (Endogeneity) Report

In the Focus Area 1 study, the main panel data specifications considered in the Task Order 1 Final Report are re-estimated using the updated state level longitudinal research data base. For the most part, the revised estimates are qualitatively similar to the original findings in the case of estimates for the tax year 1996 through 2005 period. However, there are a few notable differences which bring the trend effects results more in line with the two-way fixed effects

results for this time period. The trend effects model for the tax year 1996 through 2007 time period performs reasonably well in terms of out-of-sample predictive performance, while the two-way fixed effect model performs poorly at times (likely attributable to difficulties in predicting year effects outside of the estimation period).

In the Focus Area 2 study, the endogeneity of certain explanatory variables and the suitability of various instruments are assessed. The preliminary findings indicate that the audit rate is endogenous. Two instruments are explored: the direct examination time percentage (*det_pct*) and the natural log of the lagged value of the average number of direct examination hours (*ln_{detlag}*). In the case of the trend effects model, *det_pct* appears to be a suitable instrument while *ln_{detlag}* appears to be invalid, at least for the sample period being tested. In the case of the two-way fixed effects model, both instruments seem suitable. Many of the potential instruments based on political economy considerations also appear unsuitable. We also investigate whether the marginal tax rate and/or the filing rate are endogenous. The preliminary findings indicated that one or both of these explanatory variables is in fact endogenous.



PredictingBehav TO4
- Focus Area 1-2 Red

A.3 Research Focus Area 4 (Filing Compliance) Report

This report presents a preliminary panel data econometric analysis of *filing* compliance, using several different models (two-way fixed effects and trend effects), variable specifications, and time periods. The findings consistently point to a positive association between filing and the rate of IRS/VITA/TCE tax return preparation assistance, the share of the potential filer population over age 64, and the rate of growth in personal income. Additionally, filing compliance is negatively associated with the magnitude of the filing threshold in relation to personal income, the share of the potential filer population that is single, and the share of the potential filer population that is self-employed in certain industry sectors. The filing rate is not significantly associated with the incidence of state amnesties or refund offsets.

The results are less consistent with regard to the roles of taxpayer delinquency investigations, third-party information document matching, the unemployment rate, and average state personal income. The differences in estimation results across time periods may be indicative of a true change in behavioral relationships over time or reflect recent improvements in the quality of some of the measured variables.



PredictingBehav TO4
- Focus Area 4 Predic

A.4 Research Focus Area 5 (Time Series Methods) Report

This report explores the predictive performance of time series approaches in estimating *reporting* behavior. The report begins with a description of the autoregressive integrated moving average (ARIMA) process, a general class of models for forecasting time series. A limitation of these alternatives is that they rely solely on observed variations over time in state-level behavior. Available time series for taxpayer reporting behavior are relatively short and are punctuated by periods of significant legislative and administrative changes in tax filing and reporting requirements. The strength of these approaches is that they are capable of flexibly capturing the dynamics of behavior, in terms of both the impact of past reporting decisions and the lingering influence of external shocks on future behavior. In contrast, the static panel data models estimated in earlier reports exploit both annual and state variation in taxpayer reporting behavior but fail to directly account for dynamic factors.

We report on the performance of several ARIMA approaches: linear exponential smoothing, pre-specified low-order autoregressive integrated moving average (ARIMA) processes, flexible ARIMA and trend-stationary autoregressive moving average (ARMA) processes customized to fit individual state time series, and a low-order ARIMA specification that incorporates exogenous explanatory variables (ARIMAX). Overall, we find that a simple low-order moving average process in first differences – ARIMA(0,1,1) – performs relatively well among the competing time series approaches. In addition, we find that the ARIMAX approach is potentially useful in cases where a simple methodology is needed to forecast what behavior would have been in the absence of an intervention when a full panel data base is not available.



PredictingBehav TO4
- Focus Area 5 Altern

A.5 Research Focus Area 6 (Line Item Predictions) Report

This report further explores the potential for predicting state level federal individual income tax reporting behavior for a specific income source, complementing the earlier analysis of net self-employment income reporting in A.1 Task Order 1 Final Report. The main objective is to examine whether it is possible to predict state level gross pension and IRA income with reasonable precision using panel data techniques comparable to those employed for overall income reporting. Good candidate income sources for panel data analysis are those associated with a non-trivial degree of noncompliance, for which a suitable independent measure of the income source exists.

A variety of alternative static and dynamic panel data specifications are explored. In most of our models, we specify our dependent variable and many of our explanatory variables as ratios. However, we do consider an alternative specification in which the dependent variable and most explanatory variables are in level form. Unlike previous studies of overall income reporting, the ratio specification involving gross pension earnings yields a rather weak forecasting performance across a range of static and dynamic panel data models. A substantially stronger forecasting performance is achieved using a levels specification. This suggests that it is important to

experiment with alternative functional forms when attempting to forecast income reporting at the line item level. A second finding is that, although our imputed ASEC pension measure is very highly correlated with the Form 1040 reported measure, it had essentially no explanatory power after controlling for state level effects, a linear time trend, and other factors. We had anticipated that our imputed measure would play an important role in producing forecasts of reasonable quality. Apparently, the presence of a good independent income measure is neither necessary nor sufficient for producing reasonable panel data forecasts.



PredictingBehav TO4
- Focus Area 6 Line It

A.6 Research Focus Area 7 (Fixed Effects Alternatives) Report

The challenge with employing year effects in models used for making forward predictions is that one does not know what value the year effects will take for years outside of the estimation sample, such as future years for which one would like to predict taxpayer reporting behavior. We address this challenge by developing a preliminary forecasting methodology for predicting out-of-sample values for the year effects based on a time series analysis of estimated year effects in A.1 Task Order 1 Final Report. Further work in A.2 Research Focus Areas 1 and 2 Report compares the relative forecasting performance of our two-way fixed effects and trend effects approaches. Relative to the trend effects models, the two-way fixed effects models perform less well in forecasting one and two years ahead yet outperform the trend effects model in forecasting three and four years ahead. In this report, we refine our methodology for predicting state level reporting behavior based on a two-way fixed effects model and evaluated its relative performance to other panel data forecasting approaches; specifically:

1. A one-way “trend effects” model that treats unobserved state-specific factors as fixed effects and also controls for a constant linear time trend;
2. A one-way random effects model that treats unobserved state-specific factors as a random variable; and
3. A two-way random effects model that treats both unobserved state-specific factors and unobserved year-specific factors as random variables.

We investigate relative performance using both a short panel spanning a modest number of tax years and a long panel spanning up to 15 tax years.

Overall, the two-way fixed effects model seems preferable for point estimation, both because it yields a more intuitive prediction of the effect of audit rates on reporting behavior than the other models and because both theoretical arguments and empirical findings (i.e., the Hausman test) favor a fixed effects specification over a random effects specification. However, the two-way fixed effects model does not seem as well suited as some of the alternatives for forecasting state level taxpayer behavior in the absence of an intervention. In particular, the trend effects model demonstrates a relatively strong performance, at least in reasonably long panels. In the case of a short panel, the one-way random effects model appears to be a reasonably good alternative.



PredictingBehav TO4
- Focus Area 7 Fixed

A.7 Research Focus Area 8 (Dynamic Effects) Report

Often the impact of changes in public policy or economic factors on individual behavior takes place gradually over time rather than all at once. This report investigates whether forecasting performance can be improved by accounting for dynamic effects, positing a relationship between current taxpayer reporting behavior and both current and past values of selected explanatory variables - the audit rate, the marginal tax rate, and unemployment rate. We then extend our analysis to allow a link between current reporting activity and prior reporting levels. We estimate various ADL specifications using alternative approaches, including standard fixed effects with a linear trend (trend effects), bias-corrected trend effects, Anderson-Hsiao instrumental variables, and Arellano-Bond GMM.

Our results demonstrate that dynamic panel data methods may improve panel data forecasting performance. Particularly promising is a relatively simple autoregressive distributed lag specification in which current reporting behavior is modeled as a function of past behavior as well as other explanatory factors. For the time windows we have employed (14 to 17 years), the standard fixed effect model outperforms both the Anderson-Hsiao and Arellano Bond methodologies in terms of forecasting performance. This suggests that the bias associated with fixed effects estimation in the presence of a lagged dependent variable is modest for time windows of this length. Our experiments do not indicate any forecasting benefit associated with the inclusion lagged values of selected independent variables (such as the audit rate, the marginal tax rate, and the unemployment rate) in trend effects estimation when a lagged dependent variable is present.



PredictingBehav TO4
- Focus Area 8 Dynar