



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: title page and table of contents only

Requested date: 06-August-2024

Release date: 27-August-2024

Posted date: 16- September-2024

Source of document: FOIA Request
Internal Revenue Service
GLDS Support Services
Stop 211
PO Box 621506
Atlanta, GA 30362-3006
Fax: 877-807-9215
[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:

August 27, 2024

Employee name:

Anamaria B. Mircan

Employee ID number:

1003374216

Telephone number:

904-661-3307

Fax number:

855-205-9335

Case number:

2024-22452

This is a final response to your Freedom of Information Act (FOIA) request we received on August 6, 2024.

You asked for copy of reports produced by International Business Machines for the IRS under contract TIRNO08K00292 and TIRNO09Z0002. We searched for, and we are enclosing, 22 pages 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, appearing to read 'Deanna J. Fitti-Hafer'.

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



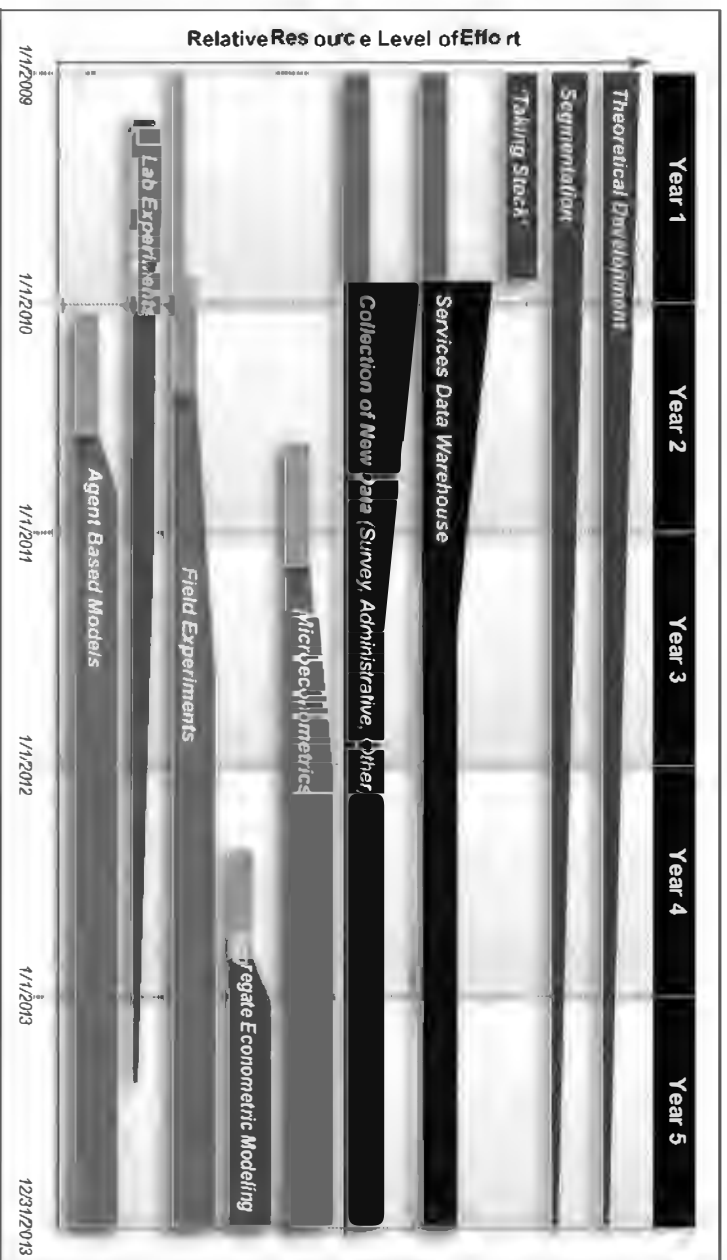
factors, taxpayer networks and compliance behavior. The IRS has some working experience with agent-based models, having already built or leveraged two such models for other initiatives.

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:

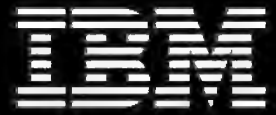
- I. The IRS should invest in a **dynamically managed portfolio of methodologies**.
 - a. The IRS should manage this portfolio through a research oversight committee, with **periodic performance evaluation** and appropriate revisions.
 - b. The IRS should prioritize those methodologies with the best 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**.

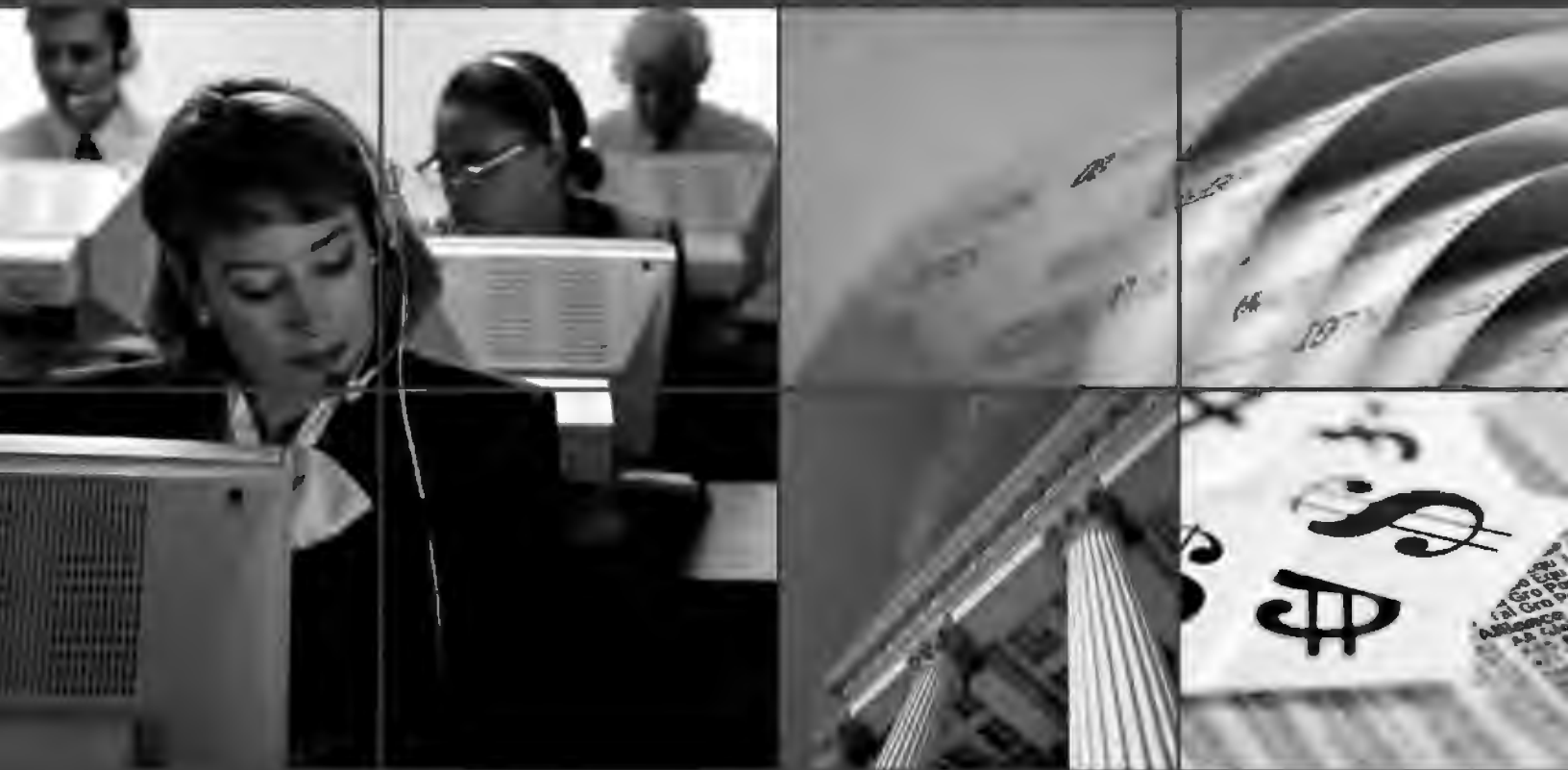
- V. Once appropriate data and insights are in place, begin investing in the measurement of indirect and direct effects.
 - a. In order to measure the marginal direct and indirect effects of services and enforcement activities, the IRS should **invest in carefully controlled field experiments**.
 - b. 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.



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. TwoWay 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.