Challenges in Developing a Small Business Tax payer Burden Model

Authors:
Don DeLuca
Arnold Greenland
Audrey Kindlon
Michael Stavrianos

I. Introduction

In 1998, the Internal Revenue Service (IRS) contracted with IBM Business Consulting Services (IBM) to develop an improved methodology for measuring and modeling the compliance burden imposed by the federal tax system. At that time, the IRS estimated that taxpayers spent approximately six billion hours each year on tax compliance activities, such as tax planning, record keeping, and form completion.

The IRS's plan to measure and model tax compliance burden is an ambitious one—particularly when one considers the complex network of burden created by the tax system. Each type of tax (e.g., income tax, employment tax, excise tax) imposes a unique set of compliance requirements, which can generate burden in the pre-filing, filing, or post-filing time period. Moreover, compliance burden can mean different things to different subgroups of taxpayers (e.g., individuals, businesses, tax exempt entities). To identify administrative practices that can reduce compliance burden, the IRS must understand how each group is affected.

In light of the complexity and scope of this issue, the IRS is building the capacity to measure and model compliance burden in incremental steps, starting with two segments of individual taxpayers—Wage and Investment (W&I) and Self-Employed (SE). In March 2002, IBM delivered the first burden simulation model, addressing the federal income tax compliance burden of W&I taxpayers. The second model, addressing analogous burdens for SE taxpayer burden, was completed and integrated with the W&I model in January 2003.

IRS's Office of Research now seeks to complement these prior studies of individual taxpayer burden with a study of compliance burden among Small Business (SB) taxpayers. Like the previous studies, this new study focuses on compliance burden incurred in the pre-filing and filing time periods. Unlike the previous studies, the scope of this new research extends beyond federal income tax compliance burden to include compliance burdens associated with employment taxes and excise taxes.

II. Challenges in Developing a Small Business Taxpayer Burden Model

The task of developing a model of tax compliance burden among SB taxpayers involves a number of challenges. Many of these challenges are now familiar, as they were encountered during both of the previous studies of individual taxpayer burden. Having successfully completed these studies, we have a firm understanding of the challenges associated with each of the core model development tasks:

Design Tasks
- Operationalize the Definition of Burden
- Establish Model Objectives and Data Requirements
- Evaluate Data Collection Options
- Develop an Operational Model Design
- Develop an Implementation Plan

Implementation Tasks
- Collect Data for Use in Model Development
- Estimate Statistical Relationships / Develop Simulation Algorithms
- Develop a Simulation Model

In addition to these known challenges, we face several new challenges that are specific to this study of SB taxpayer compliance burden. First, the scope of the SB study—which includes federal income, employment, and excise tax burden—is substantially broader than the scope of previous studies, which were limited to federal income tax compliance burden. Second, we anticipate that SB taxpayers experience a greater level of compliance burden and greater diversity in compliance activities, than do their individual counterparts.

This remainder of this section provides an overview of the eight tasks listed above, with particular emphasis on the challenges associated with designing and implementing a model of SB taxpayer compliance burden.

II.1 Operationalize the Definition of Burden

The first task in the design phase of the SB burden project is to refine and sharpen the definition of taxpayer compliance burden that was developed as part of the W&I and SE studies. Establishing a precise and measurable definition of burden is a critical first step towards the development of an SB burden model, as it provides clarity with respect to the primary focus of the model.

To determine whether changes were needed in the definition of burden that was used with individual taxpayers, we revisited each of the five key elements of our definition and asked whether they could accommodate the diverse compliance issues surrounding SB taxpayers. Specifically, we sought to answer the following questions:

- What is the conceptual definition of total taxpayer burden?
- What is the population of interest for this study of small business burden?
- What filing requirements and activities are within the scope of tax compliance burden?
- What is the proper time period for measuring burden?
- What are the appropriate units of measurement for small business burden?

We relied on three primary sources of information to answer these questions. First, we conducted a series of meetings and in-depth-interviews with stakeholders and subject matter experts. Second, we reviewed recent literature, including studies of tax compliance burden in the US, UK, Canada, Asia, Australia, and New Zealand. Third, we analyzed public-use data obtained from IRS, the Small Business Administration, and the Census Bureau.

As we refined our definition of burden for this study of SB taxpayers, we encountered many complex decisions. On a number of issues, the guidance provided by stakeholders or other information sources did not point to any clear resolution.[1] To help us sort through this ambiguity and move towards an operational definition of SB compliance burden, we established five guiding principles. We used these guiding principles (listed below) to assess how well each proposed definition supported the fundamental objectives of the project:

- **Alignment with Business / Analytic Objectives –** Does the proposed definition support stakeholder business / analytic objectives?
- **Internal and External Integrability –** Does the proposed definition support comprehensive measurement of Small Business burden and external integration with measures of total taxpayer burden?
- **Definitional Consistency –** Is the proposed definition consistent with definitional decisions from IRS's prior studies of tax compliance burden?
- **Practicality and Measurability –** Can the proposed definition be measured accurately and practicably through primary and secondary data collection efforts?
- **Expert Consensus –** Do subject matter experts generally agree with the proposed definition?

II.2 Establish Model Objectives and Data Requirements

The second task of the study is to establish a more detailed set of "functional objectives" for the SB burden model and describe the data that would be needed to support this functionality. This is a critical task in the design of our SB burden model, as it defines the capabilities of the model and draws boundaries around the information the model can provide. Most functional objectives can be thought of as answers to the following three questions:

- **What simulation levers are needed?** Simulation levers are model inputs that can be adjusted by model users to simulate a change in one or more factors that influence taxpayer burden. A number of factors are likely to influence SB taxpayer burden, such as taxpayer demographics (e.g., industry, size of business, organizational structure), tax complexity, and compliance methods. Understanding which of these factors are of particular interest to IRS will help us design a model(s) that provides an appropriate set of simulation levers.

---

[1] For example, stakeholders proposed a diverse array of metrics to define a small business, including assets, employees, business structure (corporation, partnership, sole proprietorship, etc.), receipts, type of ownership, accounting method, etc.

- **What are the outcomes of interest?** Beyond the total level of burden, what outcomes should the model be able to forecast? Examples might include specific components of burden (e.g., record keeping, tax planning) or business decisions that relate to tax compliance (e.g., which professional services to use, which accounting or depreciation method to use, how to structure one's business).

- **What are the opportunities for model integration?** The compliance burden created by the federal tax system can be allocated across several dimensions, including tax types, taxpayer types, and filing periods. In designing the SB burden model(s), careful consideration must be given to the desired degree of model integration across these dimensions—as well as integration of burden estimates with other analytic tools or processes (e.g., tax revenue estimates, IRS resource allocation).

We relied on three primary sources of information to identify and prioritize functional objectives. First, we conducted a series of meetings and in-depth-interviews with stakeholders and subject matter experts. Second, we compiled a database of recent burden reduction initiatives and other proposals that could influence the level or distribution of taxpayer compliance burden. Third, we examined the characteristics of the SB taxpayer population and gave a higher priority to functional objectives that were relevant for a large proportion of SB taxpayers.

II.3 Evaluate Data Collection Options

During Task 3, we will evaluate options for obtaining data to support the functional objectives of the model. The challenge inherent in this task will be to understand whether the information we have deemed as necessary for the model's functionality (as defined during Task 2) is practical to collect from taxpayers or obtain through IRS administrative data. We will examine and recommend a strategy for data collection as well as identify and evaluate potential administrative data sources and evaluate primary data collection options. During this task, we are addressing two major questions:

- What criteria should be used to evaluate IRS administrative data sources?
- What evaluation criteria should be used for primary data?

To evaluate what IRS administrative data is available and useful, we have conducted several meetings with IRS data experts to obtain information on potential administrative data sources. To identify potential sources of IRS administrative data, we are evaluating selected data files (SOI, CRIS, and MACS) based on the following criteria: (1) ability to obtain a large/representative sample of taxpayers, (2) quality and detail of data, (3) access to data in a timely manner, (4) currency of data and ability to update data over time, and (5) ability to integrate samples across tax types.

As for the primary data collection, we will draw on our experience of collecting data from the previous taxpayer burden studies. However, in light of the many differences between individual taxpayers (who file 1040-series forms) and SB taxpayers (who file a multitude of forms), it is important to re-evaluate our options and limitations with respect to collecting primary data from Small Business taxpayers. We are conducting focus groups with both Small Business taxpayers and tax preparer who work with Small Business taxpayer to best understand our these options and limitations. Qualitative research helps us better understand the ability of potential respondents to provide the information required by the model and will help to inform the decisions we will make with regards to the method of collecting data, our questionnaire design and sampling design.

II.4 Develop Operational Model Design

During Task 4 we will develop an operational design for the burden model. This operational design will describe the model architecture in sufficient detail as to illustrate how the model will be built, how it will operate, and what information it will be able to provide. In particular, the operational design will address the following questions:

- What are the functional components of the proposed model, and how do these functional components work together to predict taxpayer compliance burden?
- What is the proposed architecture for each functional component and for the integrated model? How will these functional components be developed?

- What data will be used to drive the model? Will different data be used for model development and operation?
- What simulation levers are available in the proposed model? What scenarios can the model handle? What scenarios can the model not handle, and why?
- How will reform scenarios be implemented in the proposed model? Will there be a graphical user interface? What simulation levers will this interface provide?
- What resources and expertise will be needed to develop, maintain, and operate the proposed model? What is the proposed strategy for model deployment (i.e., web-access, remote use on stand-alone machines, centralized use at IRS headquarters)?

The central challenge of this task is to develop an operational design that balances several competing objectives and constraints. At the most basic level, we need to weigh the objective of model functionality against the constraint of data availability. Juxtaposing these two pieces of information should reveal an outer limit of desired functionality that can be supported by the available data.

IRS may, however, want to step back from this outer limit of functionality, for a variety of reasons. Increasing model functionality typically results in a more complex model, with higher costs of development, operation, and maintenance. Complex models also tend to be less accessible to a wide audience, resulting in more centralized use by a smaller group of dedicated analysts.

II.5   Develop Implementation Plan

The design phase of the SB Taxpayer Compliance Burden Study will culminate with a detailed work plan for the implementation phase. The specific elements of the implementation plan will depend on information gathered during the design phase, but the work plan is likely to include the following tasks:

- **Data Collection Tasks**
  - Design, develop, and pretest data collection instruments
  - Prepare an OMB clearance package
  - Design and select samples of taxpayers and/or tax professionals
  - Collect primary and secondary data
- **Data File Development Tasks**
  - Validate primary and secondary data sources (e.g., check for internal consistency)
  - Clean primary and secondary data sources, as needed
  - Integrate data files
  - Document data files
- **Data Analysis and Model Estimation Tasks**
  - Conduct descriptive analysis
  - Conduct explanatory analysis and estimate models
  - Evaluate findings and discuss with Working Group
- **Software Development and Testing Tasks**
  - Develop production data file
  - Develop software tools to support simulation modeling
  - Conduct unit tests and system tests to verify that model is operating correctly
  - Conduct simulation tests to verify that model produces reasonableness results
  - Prepare documentation for project and model
- **Model Deployment Tasks**
  - Support IRS efforts to deploy the model to end users and conduct user training
  - Support IRS efforts to advertise the model to key stakeholders

Several challenges surround the development of this implementation plan. Foremost, the plan must lead to development of a model that will support the project's core objectives: (1) measuring the level of SB taxpayer burden, and (2) helping IRS understand how burden changes in response to changes in tax policy, tax system administration, and taxpayer characteristics. Of equal importance is that the implementation plan be achievable within the proposed time frame and budget—and that it be integrable with prior and future studies of taxpayer compliance burden.

II.6 Collect Data for Use in Model Development

Collecting data on SB taxpayer burden is a complex and exacting task. Information must be collected from a representative group of taxpayers, the information must be accurate, and it must support estimation of the burden model (i.e., uncovering statistical relationships between burden and other factors). Key tasks related to data collection include determining the mode of data collection, developing the sampling plan, writing and pre-testing the questionnaire(s), collecting data, editing and cleaning data, and preparing data for estimation.

When determining the mode of data collection the following considerations must be balanced including the desired response rate, budget available, scheduling, and type of information needed. For the W&I and SE taxpayer studies we implemented a mixed mode data collection methodology, employing both telephone and self-administered mail questionnaires.

When developing the sampling plan, the most significant challenge will be to ensure that the sample sizes are large enough to: a) provide statistically significant estimates at the national level, and b) enable the detection of significant differences between the taxpayer groups contemplated to be important for the burden model.

The primary purpose of the data collection instrument is to collect comprehensive and accurate measures of small business taxpayer burden. Therefore, much of the instrument will be devoted to measuring the specific activities completed by each respondent, and the time and money spent on these activities. Supporting questions will investigate the methods that taxpayers use to comply with the tax system and other factors that may relate to taxpayer burden. After the instrument is completed, it must be pre-tested to ensure that respondents are interpreting questions consistently and that each question elicits the desired information.

Post data collection techniques include cleaning and editing the data, weighting the data, and linking the data collected in the survey to the tax returns. In weighting the data, various techniques can be used to calibrate the survey results to known totals derived from IRS administrative data files.

II.7 Estimate Statistical Relationships / Develop Simulation Algorithms

Of all the tasks involved in building a model of compliance burden, perhaps the most challenging is to develop the algorithms that simulate taxpayer behavior and burden. A number of factors contribute to the complexity of this task—there are several outcomes of interest, each outcome is influenced by multiple drivers, and some of these drivers are unobservable or difficult to quantify. The algorithms must reflect enough of these factors to provide accurate forecasts of taxpayer behavior, but not so many that the model becomes intractable—and collectively, the factors included in each algorithm must provide model users with simulation levers that accommodate a wide variety of what-if scenarios.

In both studies of compliance burden among individual taxpayers, the model estimation task focused on the development of three sets of algorithms:

- **Tax Engine.** An enhanced tax calculator that evaluates tax rules and taxpayer characteristics to determine filing requirements and filing outcomes. Inputs to the tax engine include tax parameters and structures, taxpayer characteristics, and macroeconomic growth rates; outputs include the specific forms, schedules, worksheets, and line items completed by the taxpayer, as well as the dollar amounts reported on most lines.

- **Decision Module.** A model that simulates taxpayer decisions and behavior with respect to compliance methods and optional filing outcomes. Inputs to the decision module include taxpayer characteristics and preferences, as well as filing outcomes predicted in the tax engine; outputs include the taxpayers decisions with respect to preparation method, submission method, and selected non-mandatory filing outcomes (e.g., participation in credits, choice of primary tax form).

- **Burden Module.** A simulation model that predicts the level of compliance burden. Inputs include taxpayer characteristics, filing outcomes, and compliance methods; outputs include the level of time and money burden for selected activity categories (e.g., record keeping, tax planning, form completion).

The SB population creates unique challenges that impact each of these model components. In the Tax Engine, limited detail in the tax data reported by SB taxpayers creates challenges with respect to the simulation of tax liability and filing outcomes. In the Decision Module, the existence of several compliance decisions that are specific to the SB study (e.g., accounting method, tax deposit method, use of payroll services) creates additional modelling challenges. In the Burden Module, ambiguity surrounding the distinction between tax compliance burdens and general business management burdens creates challenges in terms of the allocation of time and money burdens.

II.8 Develop a Simulation Model

The final task in each study of compliance burden is to develop a simulation model (i.e., a software tool) that: (1) applies the estimated relationships to a nationally representative sample of taxpayers, (2) gives users access to simulation levers that support a wide range of "what-if" scenarios (e.g., tax parameters, forecast parameters), and (3) generates summary reports that describe the level and distribution of burden under different scenarios.

The primary challenge surrounding this task is to ensure that the simulation model is easy to use, yet still provides all of the functionality demanded by model users. An additional challenge that is unique to this study is the challenge of coordinating the SB simulation model with the existing Individual Taxpayer Burden Model, to provide a consistent look and feel to the user interface, and integrated estimates of time and money burden.