The IRS Research Bulletin

Proceedings of the 2021 IRS / TPC Research Conference

Research, Applied Analytics & Statistics

Publication 1500 (Rev. 5-2022) Catalog Number 11546J Department of the Treasury **Internal Revenue Service** www.irs.gov

Papers given at the

11th Annual Joint Research Conference on Tax Administration

Cosponsored by the IRS and the Urban-Brookings Tax Policy Center

Held Virtually
June 24, 2021

Compiled and edited by Alan Plumley*
Research, Applied Analytics, and Statistics, Internal Revenue Service

*Prepared under the direction of Barry W. Johnson, Acting IRS Chief Research and Analytics Officer

Foreword

This edition of the IRS Research Bulletin (Publication 1500) features selected papers from the IRS-Tax Policy Center (TPC) Research Conference held virtually on June 24, 2021. Conference presenters and attendees included researchers from many areas of the IRS, officials from other government agencies, and academic and private sector experts on tax policy, tax administration, and tax compliance. Many people participated in this, our second fully virtual conference. Videos of the presentations are archived on the Tax Policy Center website to enable additional participation. Attendees participated in the discussions by submitting questions via e-mail as the sessions proceeded.

The conference began with welcoming remarks by Eric Toder, Co-Director of the Tax Policy Center, and by Barry Johnson, the Acting IRS Chief Research and Analytics Officer, who introduced IRS Commissioner Charles Rettig, who gave some opening remarks. The remainder of the conference included sessions on improving individual taxpayer compliance, the impacts of variations in tax administration processes, developments in technology and analytics, and enhancing taxpayers' customer experience. The keynote speaker was Mark Mazur, the Acting Assistant Secretary for Tax Policy, who offered his insights on current tax issues.

We trust that this volume will enable IRS executives, managers, employees, stakeholders, and tax administrators elsewhere to stay abreast of the latest trends and research findings affecting tax administration. We anticipate that the research featured here will stimulate improved tax administration, additional helpful research, and even greater cooperation among tax administration researchers worldwide.

Acknowledgments

This IRS-TPC Research Conference was the result of preparation over a number of months by many people. The conference program was assembled by a committee representing research organizations throughout the IRS. Members of the program committee included: Alan Plumley, Brett Collins, Patrick Martin, Christine Oehlert, and Melanie Patrick (Research, Applied Analytics, and Statistics); Terry Ashley (Taxpayer Advocate); Javier Framiñan (Wage and Investment Division); John Dattoma and Larson Silbaugh (Small Business and Self-Employed); Connor Kitchings (Tax Exempt and Government Entities); and Rob McClelland (Tax Policy Center). In addition, Ivy Hunter and Ann Cleven from the Tax Policy Center oversaw numerous details to ensure that the conference ran smoothly.

This volume was prepared by Camille Swick and Lisa Smith (layout and graphics) and Beth Kilss and Sarah Swisher (editors), all of the IRS Statistics of Income Division. The authors of the papers are responsible for their content, and views expressed in these papers do not necessarily represent the views of the Department of the Treasury or the Internal Revenue Service.

We appreciate the contributions of everyone who helped make this conference a success.

Melanie Krause
IRS Chief Data and Analytics Officer

11th Annual IRS-TPC Joint Research Conference on Tax Administration

Contents

1

Improving Individual Taxpayer Compliance

Badgley ◆ Furlong ◆ Lykke ◆ Nicholl ◆ Plumley

Lin ◆ Patel ◆ Yuskavage

Audit Contagion? Investigating the General Indirect Effect of Audits Through Tax Preparer Networks

Ellen Badgley, Kyle Furlong, Lucia Lykke, and Leigh Nicholl (MITRE Corporation), and Alan Plumley (IRS, Research, Applied Analytics, and Statistics)

Introduction

The tax enforcement literature posits that there are two types of indirect effects of audits: taxpayers who experience an audit might change their own future reporting behavior (the "specific" indirect effect), and taxpayers who are **not** audited may still be affected by others' audits via a network or "spillover" effect (the "general" indirect effect). Much of the literature and research conducted on taxpayer compliance behavior has rested on the assumption that tax agencies' enforcement activities—particularly audits—encourage tax compliance by deterring tax evasion or, conversely, by assuring that the tax system is fair and just. Much research has been done to test whether and how a taxpayer's experience of enforcement threat or activity (e.g., a visit from an IRS officer or an audit) will affect that taxpayer's future probability of compliance (Slemrod (2016)) or, more generally, as the specific indirect effect. More difficult to identify operationally is the general indirect effect, the phenomenon in which taxpayers who are not themselves audited are affected by enforcement activities applied to others in their network.

Unlike in the case of the specific effect, where the unit of observation is clearly the individual taxpayer or tax entity, the challenge with the general effect is identifying the network or context within which the effect of audits does indeed "spill over." Researchers have approached this problem at the aggregate level, finding that higher audit rates at the state level are associated with more taxes remitted and higher compliance rates, with a magnitude significantly greater than the revenue collected from the audits themselves (Ali *et al.* (2001); Dubin *et al.* (1990); Plumley (1996)). However, this research does not give us insight into why and how audits in the general population translate into more tax dollars remitted. Although this knowledge regarding how the general effect propagates is not necessary for quantifying the overall general effect, it could be helpful for informing, for example, campaigns to encourage compliance, given that simulation research has shown that information to promote compliance has the strongest effect among dense networks (Andrei *et al.* (2014)).

In this study, we investigate the effect of audits on tax reporting within tax preparer networks. Tax preparers are known to affect taxpayer reporting and compliance (e.g., Erard (1992); Klepper *et al.* (1991)), and a recent study has shown that revenue officer enforcement of employment tax remittance among some firms induces increased employment tax payments among other firms related to an enforced firm via a shared tax preparer (Boning *et al.* (2020)). Given this foundation, we hypothesize that sharing a tax preparer with one or more audited taxpayers may be a route by which the general effect propagates among individuals, as measured by total tax reported. In this paper, we focus specifically on one type of correspondence audit that addresses Schedule C business expenses. This is an area that is ripe for noncompliance, due to the lack of third-party reporting documents on sole proprietorship income. Prior research on the specific indirect effect has shown that indeed, individual taxpayer audits affect sole proprietorship taxpayers' subsequent income reporting far more than wage income reporting (DeBacker *et al.* (2015); Kleven *et al.* (2011)).

However, to draw conclusions about the general indirect effect in tax preparer networks, we must first establish whether and how taxpayers interact with their tax preparers over time. It is the nature of audits that they happen after the preparation and filing of the return, and there may be a time lag of one to three years between when a return is filed and when a correspondence audit is opened. Thus, our first research objective in this paper is to understand the longevity of taxpayers' relationships with tax preparers. This is necessary to establish whether it is reasonable to imagine that tax preparers are a nexus of audit information given the typical time lags of audit openings.

Second, we estimate longitudinal mixed-effects linear models to test whether the total tax reporting differs over time between two groups of individual taxpayers: taxpayers who share a tax preparer with taxpayers subject to a Schedule C correspondence audit, and taxpayers whose tax preparer did not have any Schedule C correspondence audits in their network.

There are several important implications of this study. First, empirically measuring the indirect effect of audits is of paramount importance for the IRS's ability to allocate audit program resources on the basis of *total* revenue, which includes indirect revenue along with the revenue collected directly by the audits. This is the focus of a Government Accountability Office (GAO) report from 2012, and to date, measurement of the indirect effect at a level that can be applied to resource allocation decision-making is still evolving (Nicholl *et al.* (2020)). In this study, we develop an approach to evaluate one facet of the general indirect effect—tax preparer networks—in a way that could be operationalized for resource allocation decision-making if reproduced across multiple types of enforcement activities. We also use operational business rules to define our study population, meaning that all taxpayers in our study possess the characteristics that could trigger a Schedule C correspondence audit, thus closely aligning our analysis with real IRS audit operations. Second, we provide a framework for better understanding how the general indirect effect might move through tax preparer networks by explicitly analyzing the relationship between taxpayers and tax preparers over time and incorporating this relationship into our model specification.

Literature Review

Evidence for General Indirect Effect

Due to the challenge of evaluating, let alone identifying, the general indirect effect of enforcement activities on tax compliance and reporting, research in this area is still in its relative infancy. However, at a conceptual level, the idea that enforcement activities cause ripple effects that go beyond the initially treated individual is common in various fields, such as behavioral economics (Alm *et al.* (2017)). Early studies on the general indirect effect of tax enforcement actions on compliance largely explored the relationship between the audit rate of a specific jurisdiction (e.g., state, municipality, etc.) and the overall compliance of individuals residing in that jurisdiction. Using this approach, both Alm *et al.* (1992) and Plumley (1996) observed strong positive effects of the audit rate within a state on reporting compliance in that state. These studies do not explicitly specify the mechanism through which the statewide audit rate impacts tax reporting compliance; however, the underlying assumption is that audit rates act as proxies for the probability of noncompliance detection for taxpayers (Ali *et al.* (2001)) and that geographic proximity results in communication between these residents (Alstadsæter and Jacob (2017)) and may help to control for attitudes about tax compliance and cultural beliefs or norms, such as trust in government (Beers *et al.* (2012)).

Studies using individual-level data seek to identify precise mechanisms underlying the general indirect effect, yet these findings often rely on simulation-based methods from which it may be harder to show causal relationships compared to more traditional econometric approaches. Building on results from tax experiments conducted by Choo *et al.* (2013) regarding the influence of social norms on tax compliance, Andrei *et al.* (2014) showed in an agent-based simulation that targeting enforcement activities at highly connected individuals could result in higher system-wide tax compliance. Moreover, Bloomquist's extensive work (2012) on applying agent-based modeling in the tax compliance domain provides further evidence for a positive, and nonnegligible, general indirect effect (Bloomquist and Koehler (2015)). Although simulation studies help build intuition into the general indirect effect, it is important to note that calibration of these models can be challenging, and external validity of the results may be contingent on the underlying assumptions driving the behaviors of individual agents (Hokamp *et al.* (2018)).

Given that earlier studies implicitly assume that taxpayers communicate about tax compliance habits and are aware of one another's reporting behaviors, it follows that general indirect effects might be better observed at a more local level or between more clearly connected entities. To this end, various studies on Austrian

TV licensing fees have observed strong spillover effects of enforcement activities (Rincke and Traxler (2011); Fellner *et al.* (2013)) at the ZIP code/municipality level and between households up to 500 meters away (Drago *et al.* (2020)). Beyond the Austrian TV fee setting, Lediga *et al.* (2019) observe a 1-percent increase in tax payments among South African firms that are less than 500 meters from an audited firm. Findings in the United States on general indirect effects are mixed, however, with enforcement activities not having a spillover effect on filing compliance among Detroit neighbors (Meiselman (2018)) nor on reporting compliance among accounting firms that share a ZIP code (Boning *et al.* (2020)). Boning *et al.* (2020) do, however, observe that U.S. firms that share a tax preparer with another firm that is visited by a revenue officer remit 2 percent more employment taxes in the following quarter, which provides evidence for a general indirect effect in the United States and motivates the focus of the present study.

Tax Preparers and Taxpayer Behavior

Central to our theory in this paper is the assumption that tax preparers exert influence over the reporting and compliance of their clients, and we contribute to prior literature on the general indirect effect by investigating the longevity of tax preparer-taxpayer relationships over time. Previous research finds that there is nuance in the relationship between using a tax preparer and taxpayer compliance: although tax preparers increase compliance on legally straightforward tax matters, they decrease taxpayer compliance where there is ambiguity in tax law (Erard (1992); Klepper *et al.* (1991)). Additionally, taxpayers with complex returns who hire tax preparers are less motivated by legal compliance (Fleischman and Stephenson (2012)). These findings are particularly relevant to the study at hand, where we focus on a taxpayer population reporting sole proprietorship income or losses on the Schedule C; this type of reporting is not subject to third-party documentation, and therefore is more prone to noncompliance (Slemrod (2019)) and may fall into that "gray area" where tax preparers might not discourage noncompliance effectively. Additional factors, such as competition between paid preparers as the result of regional supply, may moderate the preparer-compliance association, with greater competition among preparers driving up noncompliance as preparers presumably seek to decrease clients' tax payments as much as possible (Batta *et al.* (2019)).

Other studies have found additional evidence that tax preparers have influence over their clients' reporting. For example, when tax preparers educate clients on the Earned Income Tax Credit (EITC), they drive taxpayer take-up of the credit (Chetty *et al.* (2013); Kopczuk and Pop-Eleches (2007)).

Additionally, taxpayers have individual characteristics and preferences that influence whether they use a tax preparer, and if so, what attributes they seek in a tax preparer. For example, return complexity and the opportunity cost of the taxpayer's time both increase the likelihood that a taxpayer will use a preparer (Klepper *et al.* (1991)). Furthermore, taxpayers have different viewpoints on how aggressive or risk-averse a tax preparer should be in minimizing tax payments. For example, some taxpayers prefer their preparers to be "creative" and "aggressive" in pursuing tax breaks, whereas others prefer a "no fuss, low risk" preparer (Murphy (2004); Sakurai and Braithwaite (2003)). This individual preference, while impossible to control for in observational tax return data, undoubtedly mediates the relationship between tax preparer use and taxpayer compliance.

Most closely relevant to our study is the recent finding that tax preparers serve as a conduit for enforcement risk information among accounting firms. Boning *et al.* (2020), in a field experiment of accounting firms, find evidence of an indirect effect of revenue officer visits on quarterly employment tax remittance among firms that share a tax preparer, particularly in the first three quarters after the visit. This study found that over the course of a year, each visit resulted in the visited firm paying, on average, an additional \$10,233; the visits also led to an additional \$12,258 from firms sharing a tax preparer with the visited firms. This finding provides motivation for continuing to study tax preparer networks as the context for general indirect effects; we extend the prior research by testing this hypothesis among individual taxpayers who report sole proprietorship gains/losses, using observational correspondence audit data.

Operational Context for This Study: Schedule C Correspondence Audits

Correspondence audits are one of many types of audits conducted by the IRS's Small Business/Self-Employed (SBSE) division (Rettig (2016)). In a typical correspondence audit, SBSE identifies narrowly defined individual taxpayer populations for examination, and corresponds with the taxpayer by mail to examine specific line items. As such, a correspondence audit is not a comprehensive examination of the entire tax return.

In this study, we focus on one type of correspondence audit that looks at Schedule C line items. This is an expansion of our previous work using operational data for this type of Schedule C audit, where we previously showed a specific indirect effect on audited taxpayers' total tax reporting in subsequent years. Specifically, we found that taxpayers subject to a Schedule C audit reported an average of $4,479 dollars in total tax more than audit-eligible taxpayers in the five years post-audit (Nicholl *et al.* (2020)).

Key to our analysis in the present study and in our prior work is the incorporation of operational audit selection criteria into sample selection and model specification. This mitigates the selection bias inherent in using operational data, given that operational audits are not selected at random (nor should they be). To identify narrowly defined populations of eligible taxpayers for correspondence audits, SBSE applies what Rettig (2016) calls "user-developed criteria" to identify eligible taxpayers. We define our not-audited "control" group on the basis of eligibility criteria for the audit type, to ensure operational similarity between the audited and control groups in our analysis. We then use the operational prioritization metrics as control variables in our analysis, which serve as a proxy measure to predict audit value.

One important implication of using these operational criteria with a particular category of correspondence audit is that indirect effects can be used by the IRS for future resource allocation decision-making. That is, different types of audits have different "bang for the buck" and the overall audit value is the combination of direct and indirect revenue (GAO (2012); Nicholl *et al.* (2020)). To most accurately account for general indirect effects, the IRS would ideally generate estimates of tax preparer network effects on revenue across many different types of audits and incorporate this information into resource allocation. However, to do this, analyses generated in research such as this and our prior work (Nicholl *et al.* (2020)) **must** align to IRS enforcement activities and processes; without this alignment, indirect effects estimation will remain a purely academic pursuit.

Research Questions

In this study, we address the following research questions; for the purposes of these research questions and the subsequent findings, we refer to taxpayers as having a "baseline year": the first year between Tax Years (TYs) 2006 and 2012 that the taxpayer was eligible for a Schedule C correspondence audit.

1. What are the dynamics of taxpayer-tax preparer relationships in the years leading up to and following audit eligibility? That is, what proportion of taxpayers remain with their "baseline" year tax preparer over time, and to what extent are we able to observe empirical proxies for this behavioral phenomenon in the data?
2. How does total tax reported by audit-eligible taxpayers who shared a tax preparer with an audited taxpayer for Tax Years 2006 through 2012 vary over time compared to the tax reporting of similarly audit-eligible taxpayers whose tax preparer did *not* have any audited clients? More simply put, do we observe a general indirect effect among tax preparer networks subject to a Schedule C Correspondence audit?

Data and Methods

Data

We use tax return and audit record data for primary Taxpayer Identification Numbers from the IRS's Compliance Data Warehouse (CDW) for TYs 2004–2018. Recall that the "baseline" year for a given taxpayer is the TY from 2006 to 2012 in which that taxpayer entered the sample, either in the treated or untreated group. Taxpayers enter the sample because they were eligible for this audit type, and either a) were in a preparer network where at least one client had an audited return for that TY (treatment group), or b) used one of the sample of preparers without any audited clients for that audit type for that TY (control group) and fell within the random sampling criteria.

Population Definitions

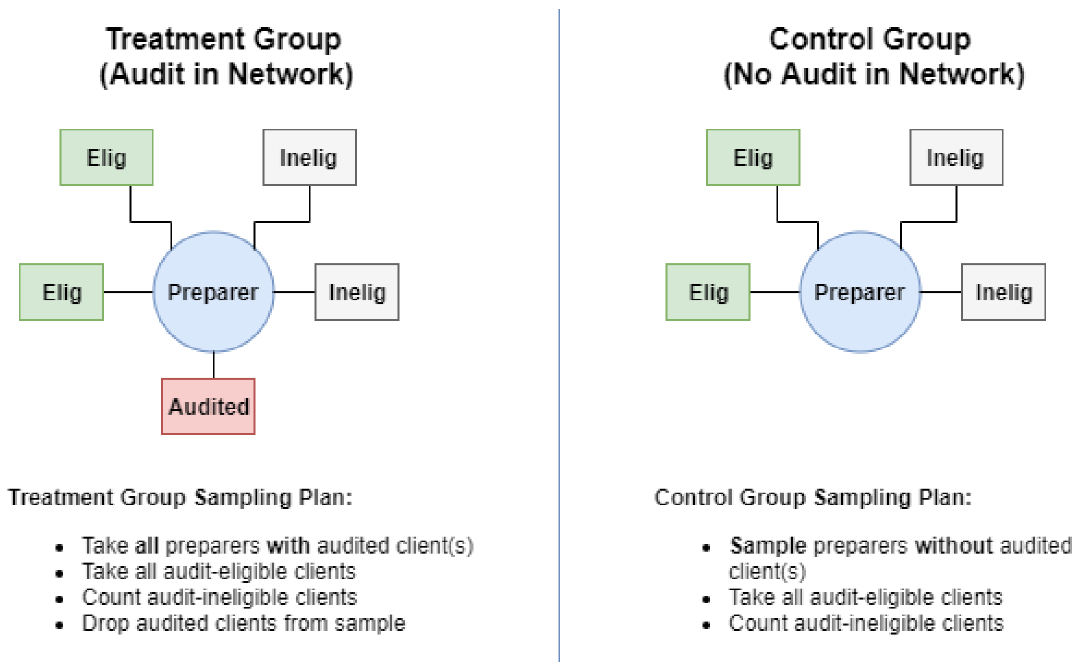
Our goal is to understand whether taxpayers who were connected to an audit **via their tax preparer network** differed in their total tax reporting compared to taxpayers with **no** audits in their tax preparer network. As such, to define our population, we gathered information on tax preparers' networks and individual taxpayer audit eligibility.[1]

The definitions for treated and untreated taxpayers, and their corresponding tax preparers, are provided in Table 1, and illustrated in Figure 1.

TABLE 1. Description of treatment and control groups for preparer networks

	Treatment Group	Control Group
Tax Preparers (PTINs)	All preparers who had at least one client audited in the Schedule C category of correspondence audit in the TYs 2006–2012 time period (baseline year).	A random sample of 10,000 preparers who did not have any clients audited in the Schedule C category of correspondence audit in the baseline year but had at least one client eligible for the audit.
Individual Taxpayers (TINs) (Preparer Clients)	All individual taxpayers who used a preparer in the Treatment PTINs group and were eligible for audit but not audited.	All individual taxpayers who used a preparer in the Control PTINs group and were eligible for audit, but not audited.

FIGURE 1. Sampling approach and definitions for treatment vs. control groups



[1] We require taxpayer populations to be audit eligible in their baseline TY; however, we do not require that they maintain this audit eligibility in subsequent TYs.

In cases in which a taxpayer entered the analytical sample multiple times (due to being in an audited network for multiple years and/or entering the sample in multiple treatment/control groups), we developed de-duplication rules to keep only one taxpayer observation.[2]

To evaluate taxpayer behavior and reporting patterns before and after the baseline year, we define our queries to draw taxpayer return data from two years prior to the baseline year through seven years after the baseline year. The only exceptions to this time period are taxpayers who entered the sample in the baseline TY2012, whose TY2019 data we omit from the analysis since it was not yet available at the time of our data extraction. Preparers with greater than 600 clients or fewer than 10 clients are removed from the sample to ensure realistic network size.[3] It is worth highlighting that unlike many studies on audit effectiveness in which the treatment group is subject to an audit—and therefore, care must be taken to ensure that the control group appropriately resembles the audited treatment group to mitigate sample bias—**neither group in our study is subject to the audit in question**. This choice in specification mitigates some of the challenges of designing a control sample comparable to the treatment sample, but also draws sharper focus on how tax preparers with similar characteristics work with diverse taxpayers for different periods of time.

Dependent Variable

Total Tax. Our dependent variable is total tax as reported on Form 1040, since the change in tax paid over time most closely represents the "return on investment" that the IRS reaps from any observable general indirect effect that results from the audit. Total tax, along with all other variables measured in dollars, is adjusted for inflation to 2018 U.S. Dollars (USD).[4] Following deflation, all monetary variables are Winsorized at the 99th percentile per norms from the literature to account for erroneous outliers that may arise due to IRS data entry or taxpayer calculation errors (DeBacker *et al.* (2015)). Because total tax is strongly right skewed, we fit our analysis models using the natural logarithm of (total tax plus one dollar) to account for cases in which the taxpayer has reported zero total tax. If an indirect effect is present, we would expect total tax reporting to increase for the treatment group relative to the control group.

Independent Variables

Treatment-Time Interaction. The primary variables of interest are treatment status and its interaction with time, specified as Tax Years since the baseline year. Treatment status is a time-invariant variable for each taxpayer, as they can be considered only as "treated" or "control" in our sample. We define the baseline year as Year 0, and we fit years after baseline as a categorical variable rather than a continuous, numeric variable, such that its slope is not constrained to be linear. This allows for any potential attenuation in indirect effect to be captured.

Control Variables. A variety of control variables were assessed with the intent to account for possible changes in taxpayer characteristics over time, including financial situation, living situation, and family structure. For all models, we control for *Total Positive Income* (TPI), Winsorized and adjusted to reflect 2018 U.S. dollars.[5] We treat *Filing Status* (FS) as a binary variable, with 1 being Married Filing Jointly and the reference level being other filing statuses collapsed into one category (Single, Married Filing Separately, Widow/er, Head of Household). A binary wage indicator is derived on the basis of the presence of any nonwage income reported on Form 1040 (*AnyWages)*. TPI, FS, and any wages are treated as time-varying covariates. We also fit *Tax Year* of the return as a categorical variable with possible values Tax Years 2006–2018. We control for

---

[2] Our de-duplication rules are:

1) For any taxpayers whom our queries returned multiple times because they were captured as "eligible" multiple times and were not in an audited preparer network in TYs 2006–2012, we declare the most recent eligibility year as the "baseline" year.

2) For any taxpayers whom our queries returned multiple times because they were in an audited network multiple times, we declare the first record as the "baseline" year.

3) For any taxpayers whom our queries returned as being in the control group in one or more years and in the treated group in one or more years, we declare the earliest (or only) treatment group record as the "baseline" year and consider them solely in the "treatment" group.

[3] The 10 to 600 client range was selected under guidance from subject-matter experts within the IRS Return Preparer Office.

[4] Inflation adjustment was based on the Consumer Price Index (CPI): value in 2018 USD = (CPI in 2018)/(CPI in the TY of interest) * value in TY of interest.

[5] Total Positive Income is defined as the sum of all income without subtracting losses or deductions.

Priority in the baseline year, a variable representing the metric used operationally to rank and select returns for Schedule C correspondence audit. This is also Winsorized, being a monetary variable and measured in 2018 USD with the interpretation that higher priority is more likely to be audited. We derive a time-invariant binary variable for taxpayers who were subject to another audit (*OtherAudit*) at any point after their baseline year as a potential surrogate for taxpayer behavior. Finally, we derive the total number of audit-eligible taxpayers for each preparer in the baseline year (*TotalEligible*) to control for network-level preferences that may contribute to taxpayers' behavior with respect to preparer choice, which in turn could sharpen or dull the hypothesized general indirect effect.

Methods

To evaluate the presence of a general indirect effect, it is critical that we first establish the mechanism through which we believe the audit will affect taxpayers in the treatment group. Specifically, our underlying assumption is that tax preparers learn of clients' Schedule C audits. We hypothesize that tax preparers carry knowledge of these audits with them and use it when assisting other clients in the following years when filing taxes. Complicating this mechanism, however, is the nature of the time delay in audits. Correspondence audits typically begin one to three years after filing and are completed two to four years after filing. Therefore, we assume that taxpayers need to work with their tax preparer for at least one to three years in order to observe an indirect effect. We first assess the degree to which taxpayers remain with their tax preparer over the period of our study and then determine how these dynamics may affect the interpretation of our results.

Network Visualization

We use data visualization to build intuition about the mechanism in the general indirect effect of a Schedule C correspondence audit. Specifically, we build and present this complex and inter-related information as Sankey diagrams to show how taxpayers remain, or part ways, with their baseline year tax preparer in the following years. We construct a separate visualization for each treatment group to identify any differences in taxpayer-tax preparer longevity in our sample. Moreover, to visualize a larger set of taxpayer behaviors as they relate to tax preparer longevity, we assign taxpayers into the following groups for each year based on who, if any, they use as their tax preparer:

1. **Taxpayers with their Baseline Preparer** are taxpayers who use the same tax preparer as for their baseline year for a given year.
2. **Taxpayers without a Preparer** are taxpayers who do not use a preparer in the given year despite using a tax preparer in, at least, the baseline year.
3. **Taxpayers with a New Preparer** are taxpayers who are not using their baseline year preparer but are using a new tax preparer in the given year with whom they have never worked in the past.
4. **Taxpayers with a Prior Preparer** are taxpayers who are not using their baseline year preparer but are using a tax preparer in the given year with whom they have some history of working.

We use Sankey diagrams to analyze the flow of taxpayers between these categories for two years prior to the baseline year and up to seven years after the baseline year to fully study the range of longevity trends present in our sample. To construct diagrams, we used R version 3.5.3 with the package *networkD3* (Gandrud (2017)).

Model

We fit a linear mixed model to assess the relationship between treatment status and the outcome of interest over time. Linear mixed effects models are a form of linear regression allowing for repeated measurements on subjects and in which within-subject correlation is captured and accounted for in the standard errors (Moulton (1986), in Bell and Jones (2015)). A random effect (γ_{0i}) is included for each taxpayer and their preparers, nesting taxpayers within preparers. This allows observations to have their own "baseline" intercept for the dependent variable. A mixed effects model specification also has the advantage of allowing both time-varying and time-invariant predictor and outcome variables (Bell and Jones (2015)), unlike fixed-effects-only models. In each model, we interact the audit variable with the number of years since the audit to investigate whether the indirect effect varies over time. The model specifications are provided in equation (1) for the i^{th} taxpayer, p^{th} preparer, and j^{th} return (years after baseline). Analyses were conducted with R version 3.5.3, using the package *nlme* (Pinheiro (2020)).

In equation (1), $\ln(\text{total tax} + 1)_{ipj}$ denotes the natural logarithm of total tax in U.S. dollars plus one dollar, adjusted for inflation, for each individual i using preparer p at year j. γ_{0i} denotes a random effect on individual i and γ_{0p} denotes a random effect on tax preparer p.

$$\begin{aligned} (1) \qquad \ln(\text{total tax} + 1)_{ipj} &= \beta_0 + \gamma_{0i} + \gamma_{0p} \\ &+ \beta_1 baseline\ priority_i + \beta_2 treated_i + \beta_{3-9} year\ after\ baseline_{ij} \\ &+ \beta_{10-16} treated_i * year\ after\ baseline_{ij} + \beta_{17} FS_{ij} + \beta_{13} TY_{ij} + \beta_{14} TPI_{ij} \\ &+ \beta_{15} AnyWages_{ij} + \beta_{16} TotalEligible_p + \beta_{17} OtherAudit_i + \epsilon_{ipj} \end{aligned}$$

Results

Descriptive Statistics

Table 2 displays summary statistics for the treatment and control taxpayers in the baseline year.

TABLE 2. Taxpayer baseline characteristics

Categorical variables are described with N(%) and continuous variables with mean (SD)

Variable	Control Group (N = 51,823 taxpayers)	Treatment Group (N = 73,189 taxpayers)	Total (N = 125,012 taxpayers)	*p*-value
Married Filing Jointly	33,830 (65.28 %)	36,330 (49.64 %)	70,160 (56.12 %)	< 0.0001[1]
Total Positive Income (2018 USD)	$136,807.44 ($252,136.67)	$136,407.34 ($256,068.16)	$136,573.20 ($254,444.82)	< 0.0001[2]
Priority (2018 USD)	$18,697.29 ($284,291.68)	$21,647.65 ($494,126.69)	$20,424.60 ($420,060.35)	< 0.0001[2]
Any F1040 Wage Income	46,227 (89.2 %)	67,469 (92.18 %)	113,696 (90.95 %)	< 0.0001[1]
Other Audit	389 (0.8%)	1,095 (1.5%)	1,484 (1.2%)	< 0.0001[1]

[1] Chi-Squared Test

[2] Kruskal-Wallis Test

There is evidence that the five covariates of interest are different between the control group and treatment group at baseline ($p < 0.0001$ for all). Taxpayers in the control group are more likely to have a filing status of Married Filing Jointly than those in treated networks. Although the group difference is statistically significant, the average TPI in both groups is substantively similar: $136,807 for untreated networks compared to $136,407. However, taxpayers in the treated network have higher audit priority of approximately $2,950. This is unexpected, given that both groups were defined using operational audit filters, but neither group contains taxpayers who were themselves audited. It should be noted that both TPI and priority are heavily right skewed and therefore have high variance.

Table 3 shows summary statistics for taxpayer network size, presence of audit-eligible taxpayers, and presence of audited taxpayers in the tax preparer treatment and control groups for the baseline year.

TABLE 3. Preparer baseline characteristics

Both variables are displayed as mean (SD)

Variable	Control Group (N = 34,350 preparers)	Treatment Group (N = 22,410 preparers)	Total (N = 56,760 preparers)	*p*-value
Number of Clients	212.63 (143.49)	268.04 (147.88)	234.51 (147.74)	< 0.0001[2]
Number of Audit-Eligible Clients	1.51 (1.37)	3.31 (6.13)	2.22 (4.09)	< 0.0001[2]
Number of Audited Clients	NA	1.34 (3.08)	NA	NA

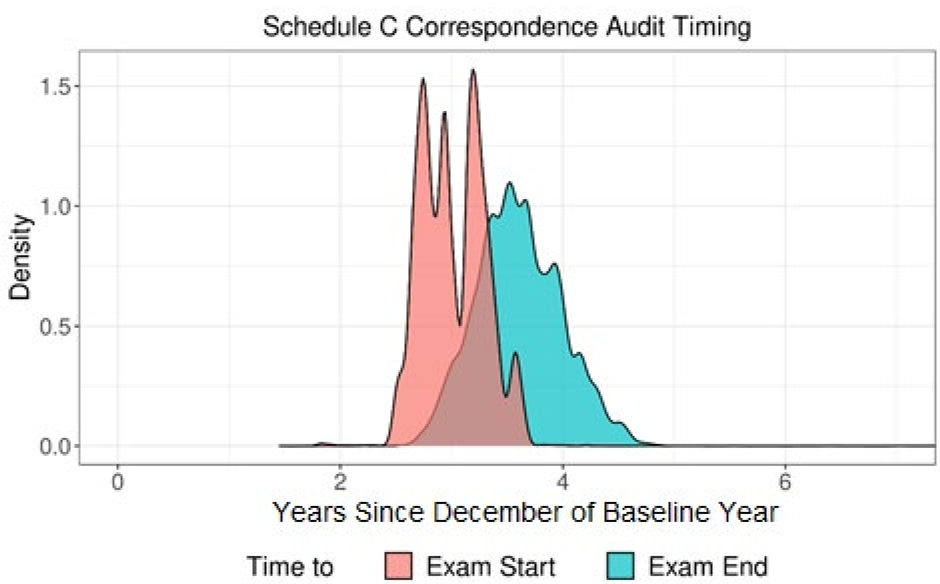
[2] Kruskal-Wallis Test

The average network size is 235 clients per tax preparer, with treated networks slightly larger (268 vs. 213, p < 0.0001). Treated networks have on average 3.3 audit-eligible clients while control networks have 1.5. Treated networks have an average of 1.34 audits per network.

Timing of Audits

Figure 2 summarizes the timing of Schedule C correspondence audits (start and end) for the audits in the treatment tax preparer group. Note that the audited clients who contribute to Figure 2 are *not* in our analysis dataset and that some preparers may have more than one audited client in their baseline TY; per Table 3, the average number of audited clients in a preparer treatment network was 1.34 (SD 3.08).

FIGURE 2. Schedule C correspondence audit timing for tax preparer networks and connected taxpayer in the treatment group

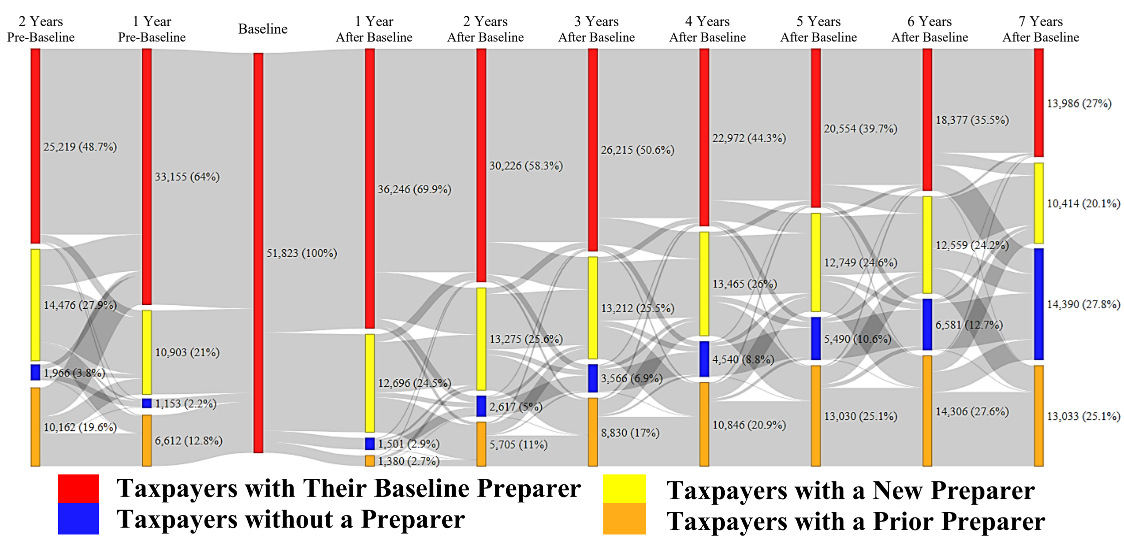


We assume that audited taxpayers are notified that their return is under examination at the time of the exam start date. Additionally, we assume that audited taxpayers immediately convey this information to their baseline year tax preparers who were responsible for the return under audit. While audited taxpayer reach-back to their baseline tax preparer is not guaranteed, many preparers offer audit protection services for their clients, which would incentivize this transfer of information on the part of the taxpayer (Stephenson *et al.* (2017)). As seen in Figure 2, most taxpayers are notified of their audit approximately two to three years after the December of the TY for which they filed the audited return. Additionally, almost all audited taxpayers, and by extension their baseline year tax preparers, are notified within four years of the audited TY with a small fraction (0.2 percent) receiving notification prior to two years. Given these stipulations, in our study we would not expect a behavioral response because of the audit to manifest before two years following the TY of the audited return.

Taxpayer-Tax Preparer Sankey Diagrams

Sankey diagrams illustrating the progression of taxpayers' relationship with their baseline year tax preparer are depicted in Figures 3 and 4 for control network taxpayers and treatment network taxpayers, respectively. As indicated in the legend underneath each of the diagrams, each "node" in the diagram is color-coded to represent whether a taxpayer is with their baseline tax preparer, a new tax preparer, or no tax preparer. Nodes are proportionately sized within each year, and labels provide the raw number of taxpayers at each node and their corresponding proportion.

FIGURE 3. Control group taxpayers: Baseline year tax preparer longevity

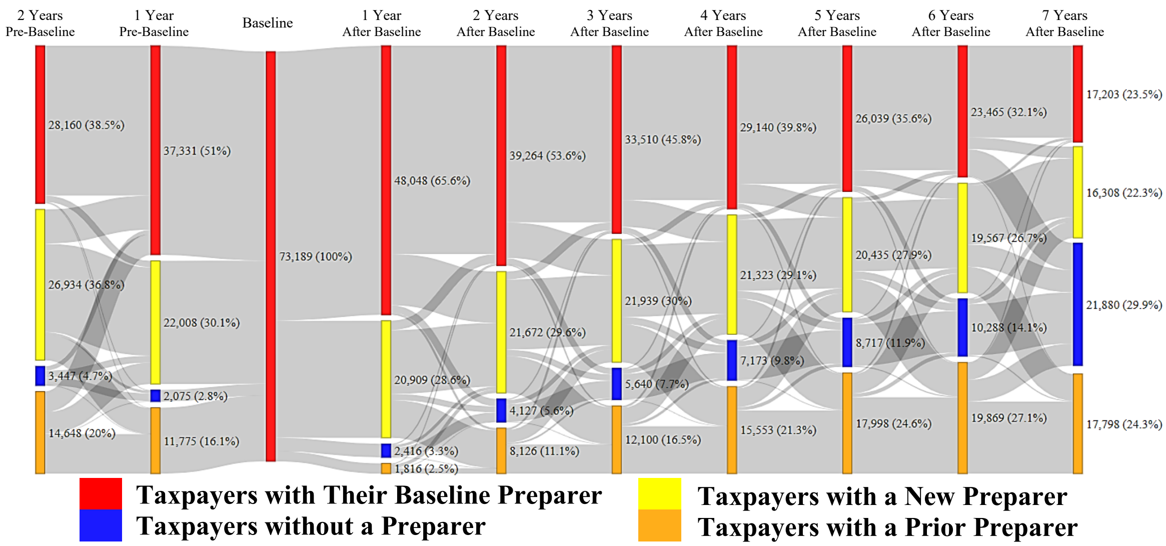


At first glace, Sankey diagrams may appear visually cluttered and challenging to read; however, we begin to see interesting patterns of taxpayer-tax preparer behavior once we break down the flows of the diagram into smaller pieces. Beginning with the third band from the left, we see a single red node—indicating that in the baseline year, 100 percent of the control group taxpayers were (by definition) using their baseline year tax preparer. Moving either forward or backward in time, however, we see a stark dropoff, with roughly 30 percent taxpayers not using their baseline year preparer. The majority of these taxpayers in the years immediately surrounding the baseline flow from using their baseline preparer (red nodes) to a new preparer entirely (yellow nodes). Applying this approach to analyzing the entire diagram, we observe the following general trends among control network taxpayers:

1. In every year after the baseline, we see a steady decrease in taxpayers using their baseline year preparer.

2. Similarly, taxpayer-tax preparer relationships appear to exhibit some degree of "stickiness." That is, taxpayers who use their baseline tax preparer for at least two years are likely to continue to use that tax preparer in subsequent years.
3. At least 75 percent of taxpayers in the control group use a tax preparer for the entirety of the study period. By seven years after the baseline, we see approximately 25 percent of taxpayers not reporting use of a tax preparer.[6]

FIGURE 4. Treatment group taxpayer-baseline year tax preparer longevity over time



Comparing the patterns exhibited among the treatment group in Figure 4, we observe that taxpayers in the treatment group follow a similar trend as the control group (Figure 3) with some slight differences at the margins. Notably, while taxpayers in our treatment group still use tax preparers at higher rates than the general public (United States Internal Revenue Office (2020)), they do seem to display a less "sticky" relationship with their baseline year preparer. That is, taxpayers in the treatment group are uniformly less likely to remain with their baseline year preparer in the following years, resulting in greater proportions of them turning to new preparers (yellow), relative to the control group. For example, four years after the baseline year, 44.3 percent of control group taxpayers continue using use their baseline year preparer whereas this is true for only 39.8 percent of treatment group taxpayers. Notably, taxpayers in the treatment group do not use prior preparers (orange) substantially different from the control group.

Model Results: Two-Level Treatment Group

Table 4 shows the estimated coefficients from the linear mixed effects model. Estimates reflect a multiplicative increase on total tax reporting, as compared to the control group's total tax reporting at baseline.

[6] We note that the proportion of our sample that uses a tax preparer for seven or more years is larger than the proportion of the general taxpayer population that uses tax preparers each year, which is approximately 53 percent in TY18 (SOI Tax Stats—Tax Stats-at-a-Glance, available at https://www.irs.gov/statistics/soi-tax-stats-tax-stats-at-a-glance). That our population would have a higher rate of tax preparer use makes sense, however, given that we focus on a Schedule C/sole proprietorship population, and taxpayers with more complex returns are more likely to use tax preparers (Klepper et al. (1991)).

TABLE 4. Exponentiated coefficients and 95 percent confidence intervals for effect on total tax

Variable	Estimate	95% CI	*p*-value
Married Filing Jointly	1.214	(1.192, 1.237)	< 0.0001
TPI	1	(1, 1)	< 0.0001
Baseline Priority	1	(1, 1)	< 0.0001
Any Wage Income	1.936	(1.897, 1.976)	< 0.0001
Any Other Audits in Study Period	0.679	(0.651, 0.709)	< 0.0001
TY2006	Reference	Reference	Reference
TY2007	0.992	(0.941, 1.046)	0.7654
TY2008	0.895	(0.848, 0.944)	< 0.0001
TY2009	0.833	(0.789, 0.880)	< 0.0001
TY2010	0.898	(0.848, 0.950)	0.0002
TY2011	0.924	(0.870, 0.981)	0.0099
TY2012	1.051	(0.987, 1.120)	0.121
TY2013	1.101	(1.028, 1.179)	0.0062
TY2014	1.179	(1.095, 1.269)	< 0.0001
TY2015	1.224	(1.131, 1.325)	< 0.0001
TY2016	1.261	(1.159, 1.373)	< 0.0001
TY2017	1.324	(1.209, 1.450)	< 0.0001
TY2018	1.141	(1.035, 1.259)	0.0083
Treatment	0.706	(0.681, 0.732)	< 0.0001
Year 0	Reference	Reference	Reference
Year 1	0.915	(0.890, 0.941)	< 0.0001
Year 2	0.949	(0.920, 0.978)	0.0008
Year 3	0.950	(0.918, 0.984)	0.0043
Year 4	0.945	(0.908, 0.984)	0.0055
Year 5	0.926	(0.886, 0.969)	0.0009
Year 6	0.91	(0.865, 0.958)	0.0003
Year 7	0.889	(0.839, 0.941)	0.0001
Number of Eligible Clients	0.972	(0.970, 0.973)	< 0.0001
Treatment*Year 1	1.126	(1.087, 1.166)	< 0.0001
Treatment*Year 2	1.215	(1.173, 1.259)	< 0.0001
Treatment*Year 3	1.32	(1.274, 1.367)	< 0.0001
Treatment*Year 4	1.438	(1.387, 1.490)	< 0.0001
Treatment*Year 5	1.497	(1.444, 1.551)	< 0.0001
Treatment*Year 6	1.536	(1.481, 1.593)	< 0.0001
Treatment*Year 7	1.597	(1.537, 1.660)	< 0.0001

There is sufficient evidence to suggest a difference in total tax reporting between groups for the baseline year: on average, the treated taxpayers remit 29 percent less tax than that of the control taxpayers (95 percent confidence interval (CI) 27 - 32, $p < 0.0001$), while holding the control variables constant. There is evidence to suggest that taxpayers whose filing status is Married Filing Jointly remit 21 percent higher tax than those of other filing statuses (95 percent CI 19 - 24, $p < 0.0001$). Taxpayers with any wage income reported on their F1040 report nearly twice as much total tax in a given year as do taxpayers without wage income (95 percent CI 1.90 - 1.98, $p < 0.0001$). Not all TYs have a statistically significant effect on total tax reporting compared to the reference TY2006. As the number of audit-eligible taxpayers in the tax preparers' networks increase, on average, the total tax reported decreases by 3 percent (95 percent CI 3 - 3, $p < 0.0001$).

FIGURE 5. Estimated percent change in total tax using treatment status, year, and treatment status-year interaction effects.

Shading represents 95% confidence intervals

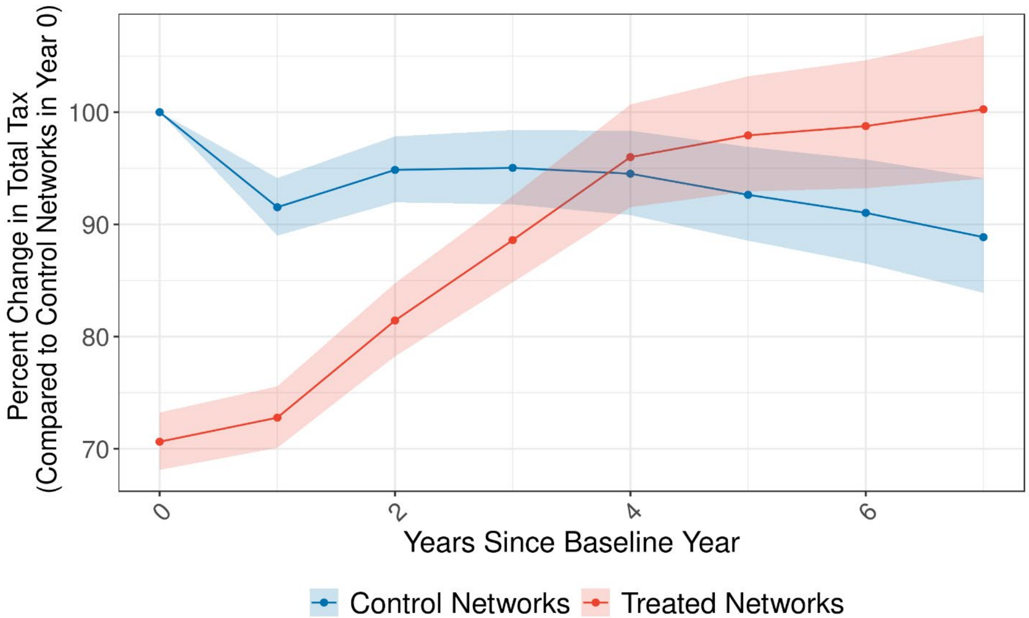


Figure 5 displays the linear contrasts of treatment status, years since baseline, and treatment-year interaction effects. The two groups have different tax reporting between the baseline year and year 1, with the treatment taxpayers greatly increasing their total tax while the control taxpayers' tax reporting decreases slightly. The treatment group continues to increase its total tax reporting over all seven study years, with years 1–4 having the steepest increase. From years 1 to 2, the control group reports increasing tax, and plateaus in years 2 to 3, followed by a slow decrease in total tax reporting.

Model Results by Treatment Group Taxpayer-Tax Preparer Longevity

Given our understanding of the mechanism under question, we would not expect to observe a general indirect effect prior to audit start, which occurs roughly two to three years after the baseline year (Figure 2). However, the Sankey diagrams presented in Figures 3 and 4 for our control and treatment groups, respectively, indicated that half of the taxpayers in our sample did not continue with their baseline year preparer in subsequent years. Approximately 50 percent of taxpayers still worked with their baseline preparer three years after the baseline. Therefore, as a sensitivity analysis on both our model results and understanding of the mechanism, we update our original model from equation (1) to redefine our treatment variable from a traditional two-level indicator to the following specification, which accounts for longevity of the taxpayer-tax preparer relationship:

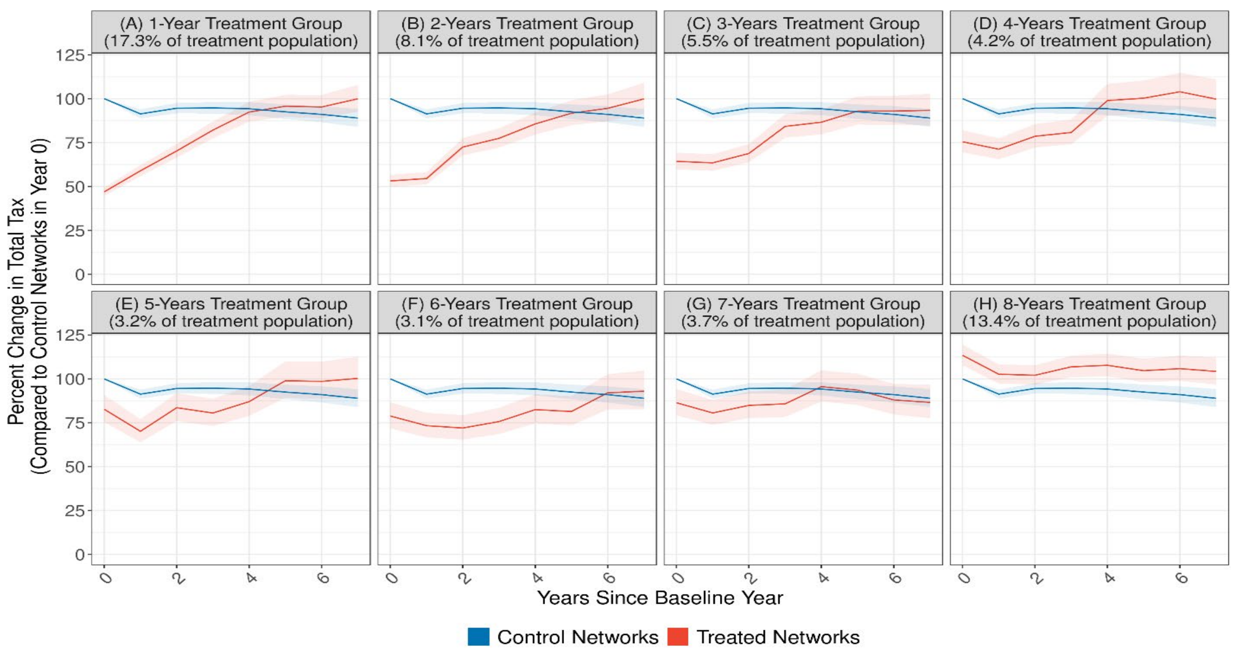
$$treated_i = \begin{cases} 0, & \text{if taxpayer } i \text{ is in the control group;} \\ \sum_{t=0}^{7} I(BaselinePreparer_i \;\; == TaxPreparer_{it}), & \text{if taxpayer } i \text{ is in the treatment group} \end{cases}$$

More simply stated, we redefine treatment group as the count of the number of years that a treatment group taxpayer used their baseline year preparer. While this specification does not require taxpayers to remain with their baseline preparer in consecutive years, our earlier Sankey analysis indicated that the pattern of leaving and then returning to one's baseline tax preparer is relatively rare among our population, particularly in the later years. Therefore, most taxpayers in treatment group *t* are with their baseline preparer for *t* consecutive years, including the baseline year.

As with years after baseline, we fit our updated treatment specification as a categorical variable rather than a continuous, numeric variable to allow for any nonlinear relationships between taxpayer-tax preparer longevity and our dependent variable and to capture any potential differences in attenuation across longevity groups. Figure 6 displays the linear contrasts of treatment, years since baseline, and treatment-year interaction effects for each treatment group. Figure 6 also displays the relative size of each treatment group. Estimated coefficients for all variables may be referenced in Table 5 in the Appendix.

FIGURE 6. Estimated percent change in total tax using treatment status, year, and treatment status-year interaction effects. Model results are broken down by number of years a taxpayer used their baseline year preparer.

Shading represents 95% confidence intervals



Compared to our two-level model from equation (1), there remains sufficient evidence to suggest a difference in total tax reporting for the baseline year although this difference varies significantly across treatment groups. The largest difference in total tax reporting between the treatment and control groups occurs among taxpayers in Plot (A) of Figure 6 who, on average, remit 53 percent less tax than the control group (95 percent CI: 49 - 55, $p < 0.0001$). The difference in total tax reporting between the treatment and control groups is smallest when we examine treatment taxpayers who have the longest relationships with their preparers (six to seven years). Among taxpayers in Plot (G) and Plot (H) of Figure 6, there is no statistically significant difference between tax reporting for the treatment and control groups in years 2 or 3 after baseline. In year 4, there is weak evidence to suggest a positive effect on total tax reporting for the clients in Plot (G) ($p = 0.041$) and stronger evidence of different treatment group tax reporting among those in Plot (H) ($p = 0.0007$). However, the impact of controlling for the duration of the taxpayer-tax preparer relationship cannot be understated as these results indicate a clear "flattening" of the slope as we move toward longer-standing relationships.

Under this new treatment design, we immediately observe that the relationship between having an audited taxpayer in the tax preparer network and total tax reporting over time varies by the longevity of the taxpayer-tax preparer relationship. Although the treatment groups demonstrate an upward slope on total tax compared to the control groups, the steepness of this trajectory varies, and indeed, the taxpayers who are with their tax preparers for the most years seem to exhibit a smaller indirect effect compared to taxpayers who are with their preparers for fewer years. This trend differs slightly among the clients in Plot (H) of Figure 6, but even among these taxpayers there appears to be an increase in total tax reporting in years 3-4 as compared to the control group. Overall, however, this trajectory appears to be counterintuitive and demonstrates the importance of investigating taxpayer-tax preparer longevity before drawing strong conclusions.

Discussion

This study is motivated by several factors. First, research finds that tax preparers have an influence on their clients' compliance (Erard (1992); Klepper *et al.* (1991)); therefore, tax preparers are of key interest when taxpayer reporting behavior and potential interventions to address the tax gap are considered. Second, tax preparers may be conduits of enforcement risk information, making them of even greater interest when taxpayer compliance is considered. Recent research has shown that tax preparer networks are a channel for information between firms that had revenue officer visits and firms that did not themselves receive a visit, resulting in the latter remitting additional tax dollars, anyway (Boning *et al.* (2020)). Finally, although the general indirect result of an audit is difficult to define and even more difficult to measure, doing so could be of great value for the IRS.

Simulation studies have shown that auditing highly connected taxpayers results in overall higher compliance in the population, for example (Andrei *et al.* (2014)). Understanding how the general indirect effect propagates through networks could increase the IRS's ability to strategically allocate audits toward high-value networks. Additionally, understanding how to observe and measure the general indirect effect can enhance the IRS's ability to fully account for enforcement revenue, and even better, to maximize overall revenue through better resource allocation (GAO (2012)). In this study, we pull these threads together to investigate whether and how total tax reporting varies between taxpayers whose tax preparers have an audited client and taxpayers whose tax preparers' networks are "clean" of audits.

A major contribution of this research is that we explicitly address an inherent assumption in any study of network effects in tax enforcement: in order to reasonably expect an indirect effect to manifest, there must be some mechanism by which the network audit influences another taxpayer. We tease this out by examining the longevity of the taxpayer-tax preparer relationship. In this way, we add to the literature on taxpayer behavior and the use of tax preparers, as well as the indirect effects literature.

First, regarding taxpayer-tax preparer relationship longevity, we find that a significant proportion of the taxpayers in our sample—about 30 percent—are with their baseline tax preparer for only one year. This has major impact on the ability to draw conclusions about the indirect effect among tax preparer networks, given that audits do not typically start until at least one year after the taxpayer files. Additionally, we observe slight differences in longevity between the treatment and control groups: control group taxpayers, who are in "clean" tax preparer networks with no audits, more often stay with their baseline taxpayers in subsequent years. It is possible that tax preparers with audits in their network have characteristics that make taxpayers less likely to stay with them over time, though our methods do not allow us to test this hypothesis.

Leveraging alternative methodologies, such as matching strategies or exploiting variation in audit notification timing within the treatment group, may be viable approaches to *control* for these differences. Matching methods used as nonparametric pre-processing can be used to better balance characteristics of the treatment and control groups and reduce model dependence (Ho *et al.* (2007)). Specifically, matching on characteristics of tax preparer networks may help balance the inherent differences between networks that are audited more often than others. Various algorithms can be used for matching, including propensity scores (Beer *et al.* (2015)), Malahanobis distance (Beer *et al.* (2019)), and coarsened exact matching, with theory suggesting that the latter two are superior to propensity scores by approximating fully blocked randomized experiments (King and Nielsen (2019)). However, it should be noted that such methods would not inherently help to *identify* which behavioral factors drive baseline differences between more and less frequently audited networks. Understanding these factors, therefore, remains a critical—if challenging—area of future work under our current methodology and identification process.

Second, our findings are suggestive of a general indirect effect through tax preparer networks but are counterintuitive in some ways. Consistent with Boning *et al.* (2020), we see taxpayers in "treated" networks—in our case, in networks where another client of the tax preparer was audited—increased their total tax reported in the years subsequent to an estimated audit notification in the network. However, when we disaggregated our findings by longevity of taxpayer-tax preparer relationship, we found that the effect was *less* strong for taxpayers who exhibited the longest duration with the tax preparer. This is counterintuitive because one would expect that taxpayers who use the same tax preparer over multiple years might communicate more with that tax preparer and therefore be more influenced by the in-network audit. However, it may also be the case that taxpayers who frequently change tax preparers have different characteristics from those with more stable relationships. Our key takeaway from these results is that future studies of the general indirect effect of audits within tax preparer networks should continue to consider the duration of taxpayer-tax preparer relationships as a moderating factor, meaning that the association between having an audit in-network may vary by the longevity of the taxpayer-tax-preparer relationship.

In this study, we cannot draw conclusions about why taxpayers may move on from their tax preparers. It could be that they are not happy with the tax liability they incur after working with a particular tax preparer, or that there is a mismatch between taxpayer and tax preparer tolerance for aggressiveness in pursuing tax breaks (Murphy (2004); Sakurai and Braithwaite (2003)). That is, some tax preparers may be incentivized to recommend riskier reporting strategies in order to minimize a client's tax liability and, accordingly, increase the likelihood that the client will employ the preparer in subsequent filing years. Conversely, preparers may be sensitive to the fact that they could directly face penalties, in monetary or reputational terms, if found to be consistently noncompliant by enforcement officers or previous clients. These attitudinal attributes, which we cannot observe from taxpayer return data, reflect the possibility of conflicting interests between taxpayers and their preparer and could influence both relationship longevity and taxpayer reporting. Future research should continue to tease out the effects tax preparers have on taxpayers by better understanding why taxpayers move between tax preparers—is it driven by tax preparer competence/aggressiveness, taxpayer characteristics/risk tolerance, or both? Understanding the dynamics of this relationship may help us better understand whether and how tax preparers serve as a nexus of enforcement risk information.

Our findings also raise questions about determining selection bias versus causality in the taxpayer-tax preparer relationship. That is, do certain types of taxpayers—e.g., taxpayers who report higher business expenses and are therefore more likely to be audited—select their tax preparers and then get audited, or do certain tax preparers *encourage* higher business expenses reporting that then causes their clients to be more likely to be audited? This question stands out given our finding that taxpayers in the treated network have higher audit priority (business expenses) than those in out control group of $2,950, despite neither group being audited. We cannot determine if this difference is because taxpayers with higher expenses gravitate toward certain tax preparers who happen to have more audits in their network, or because tax preparers with more audits in their network caused the unaudited taxpayers to report higher expenses.

Limitations

We applied operational eligibility criteria to construct both our treatment group and our control group. In doing so, we operate under the assumption that the category of audit we analyze here has been relatively stable over time, especially regarding the types of line items examined in the audits. Still, it is possible that the current selection filters did not apply to all historic Tax Years: we are informed of current selection criteria (e.g., those used for TY2018), but these filters may not necessarily apply to TYs 2006–2012, and we do not have knowledge of the eligibility criteria used in historic years for this type of audit. Similarly, formulation of the prioritization variables may have changed over time, but, without easy access to this knowledge, we must assume that the current prioritization applies to TYs 2006–2012. In addition to these assumptions about the consistency of operational eligibility criteria, we also note that not all taxpayers have a complete set of returns after the baseline year; we assume this absence is Missing at Random (MAR).

Additionally, audited taxpayers have varying notification times, even for audits of returns from the same TY, and results must be interpreted with the knowledge that not every taxpayer is aware of their audit by the time they prepare their tax return for a subsequent TY; therefore, there is variation in when the clients' preparers may be notified. While we control for the presence of other audits (e.g., not a Schedule C Correspondence audit) at the individual level in our sample, these other audits may result in other general indirect effects, which are not directly accounted for in our current model specification. Our current sampling strategy also omits all noneligible Schedule C Correspondence audit taxpayers, though some of these individuals are undoubtedly selected for other types of audits which may, or may not, produce compliance spillover effects through the tax preparer network.

Moreover, the proposed mechanism through which the general indirect effect propagates relies on assumptions which, if deemed unreasonable, may jeopardize our results. Although many tax preparers offer audit assistance and representation to their clients, audited taxpayers are not required to notify their tax preparers about whether their return is under audit. It is plausible that taxpayers do not communicate with preparers from previous years due to logistical constraints, such as geographic barriers or prohibitive financial burden, or behavioral constraints, such as a loss of trust in the tax preparer who failed to protect the taxpayer from an audit. Audited taxpayers may also prefer to reach out to their current TY tax preparer for audit assistance, which, if widespread, would fundamentally alter our understanding of the mechanism as presented in this study and shift the treatment population to the preparer network of the audited taxpayer at the time of notification. It is also possible that information regarding the audit in a preparer's network may spread between taxpayers—thereby shortcutting our theorized pathway of audit knowledge spreading from audited client to preparer to unaudited clients and potentially driving behavioral differences in taxpayers' propensity to work with that preparer in the future. Without qualitative research of taxpayer-tax preparer relationships over time, these types of insights remain speculative, and it remains challenging to discern which of these study designs most appropriately models reality.

Future Research

In addition to continuing to incorporate our insight that taxpayer-tax preparer relationship duration matters for estimating the general indirect effect of audits, future research should seek to quantify the general indirect effect in dollar values. This is crucial for using the general indirect effect as part of a measure of *total* revenue, including indirect revenue. In our prior work, we demonstrated translating exponentiated coefficients from a linear mixed model into dollar values representing the average yearly *specific* indirect effect for taxpayers after being audited (Nicholl *et al.* (2020)).

The other necessary activity to move from an academic study of the indirect effect to estimates that can be used for operational resource allocation is replication across many different types of enforcement activities. If we know the total value of audits—including their indirect effect—for many types of audits, it allows decision-makers to better allocate resources to maximize total audit value. Additionally, different types of audits select different types of taxpayers (e.g., taxpayers who report Schedule C gains/losses, taxpayers who report Schedule A deductions, etc.), and these distinct audit candidate populations have different underlying characteristics that may make them respond differently to audits. For example, in National Research Program data, sole proprietorship income is more likely to be affected by an audit compared to wage income (DeBacker *et al.* (2015)). In operational correspondence audits, sole proprietorship taxpayers who are examined for Schedule C expenses exhibit a larger specific indirect effect compared to taxpayers who are examined for Schedule E gains/losses, who exhibit no statistically significant specific indirect effect (Nicholl *et al.* (2020)). Future research should repeat the approach we lay out in this study across multiple types of enforcement activity as the next step toward determining whether incorporating tax preparer network effects into resource allocation decisions would help the IRS maximize total revenue and encourage greater taxpayer compliance.

Our analysis underscores the importance of understanding drivers of taxpayer-tax preparer relationship longevity. Our counterintuitive results and baseline differences between treatment and control groups suggest that our model suffers from omitted variable bias, which we posit is likely due to preparer selection mechanisms we are not measuring. Future research should investigate why taxpayers remain with or leave tax preparers over time; specifically, whether taxpayer-tax preparer longevity is driven by taxpayers' awareness of audit incidence among the tax preparer's clients or by taxpayer demographic factors, such as age or total positive income. These factors may explain differences in baseline preparer longevity between the treatment and control groups observed in our study.

Finally, future research should consider alternative definitions for the "baseline" year. Here, we define the baseline year as the year the taxpayer was eligible for a Schedule C correspondence audit, and we define whether eligible taxpayers were in the "treatment" group by detecting audits in their tax preparer network for that same year (e.g., an unaudited taxpayer was audit-eligible in TY2008 and their tax preparer had an in-network audit for another TY2008 return). However, we know that taxpayers shift tax preparers frequently, and there is a lag between the TY of an audited return and when the audit opens. Therefore, we suggest defining the taxpayer population on the basis of the taxpayer's audit eligibility and employment of a tax preparer in the one to two years after the tax preparer had an audit in their network. This definition would better align the timing of a taxpayer's interaction with their preparer to when awareness of the audit would propagate through the network via their tax preparer, which could better represent and capture the general indirect effect.

References

Ali, Mukhtar M., H. Wayne Cecil, and James A. Knoblett (2001). "The Effects of Tax Rates and Enforcement Policies on Taxpayer Compliance: A Study of Self-Employed Taxpayers." *Atlantic Economic Journal* 29(2): 186–202. https://doi.org/10.1007/BF02299137.

Alm, James, Betty R. Jackson, and Michael McKee (1992). "Estimating the Determinants of Taxpayer Compliance with Experimental Data." *National Tax Journal* 45(1): 107–114.

Alm, James, Kim M. Bloomquist, and Michael McKee (2017). "When You Know Your Neighbour Pays Taxes: Information, Peer Effects and Tax Compliance." *Fiscal Studies* 38: 587–613.

Alstadsæter, Annette, and Martin Jacob (2017). "Who Participates in Tax Avoidance? Evidence from Swedish Microdata." *Applied Economics* 49(28): 2779–2796. https://doi.org/10.1080/00036846.2016.1248285.

Andrei, Amanda L., Kevin Comer, and Matthew Koehler (2014). "An Agent-based Model of Network Effects on Tax Compliance and Evasion." *Journal of Economic Psychology* 40(C): 119–133.

Batta, George, Andrew Finley, and Joshua Rosett (2019). "The Effect of Paid Preparer Competition on Individual Tax Avoidance." *Hawaii Accounting Research Conference 2020* 19 (Taxation). http://hdl.handle.net/10125/64898.

Beer, Sebastian, Mathias Kasper, Erich Kirchler, and Brian Erard (2015). "Audit Impact Study." *TAS Research and Related Studies* 2.

Beer, S., M. Kasper, E. Kirchler, and B. Erard (2019). Do Audits *Deter or Provoke Future Tax Noncompliance? Evidence on Self-employed Taxpayers.* IMF Working Paper No. 19/223. doi: https://doi.org/10.5089/9781513515373.001.

Beers, Tom, Eric LoPresti, and Eric San Juan (2012). "Factors Influencing Voluntary Compliance by Small Businesses: Preliminary Survey Results." *National Taxpayer Advocate Annual Report to Congress* 2.

Bell, Andrew, and Kelvyn Jones (2015). "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3(1): 133–153. doi:10.1017/psrm.2014.7

Bloomquist, Kim M (2012). "Agent-based Simulation of Tax Reporting Compliance." PhD diss. https://hdl.handle.net/1920/7927.

Bloomquist, Kim M., and Matt Koehler (2015). "A Large-scale Agent-based Model of Taxpayer Reporting Compliance." *Journal of Artificial Societies and Social Simulation* 18(2): 20.

Boning, William C., John Guyton, Ronald Hodge, and Joel Slemrod (2020). "Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms." *Journal of Public Economics* 190(C), Elsevier.

Chetty, Raj, John N. Friedman, and Emmanuel Saez (2013). "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103(7): 2683–2721.

Choo, Lawrence, Miguel A. Fonseca, and Gareth D. Myles (2013). *Lab Experiment to Investigate Tax Compliance: Audit Strategies and Messaging.* HM Customs and Revenue Research Report 308. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/342680/Audit_Strategies_and_Messaging-HMRCResearchReport308.pdf.

DeBacker, Jason, Bradley T. Heim, Anh Tran, and Alexander Yuskavage (2015). "Once Bitten, Twice Shy? The Lasting Impact of IRS Audits on Individual Tax Reporting—Semantic Scholar." *Journal of Financial Economics* 117(1).

Drago, Francesco, Friederike Mengele, and Christian Traxler (2020). "Compliance Behavior in Networks: Evidence from a Field Experiment." *American Economic Journal: Applied Economics* 12(2): 96–133. DOI: 10.1257/app.20170690.

Dubin, Jeffrey A., Michael J. Graetz, and Louis L. Wilde (1990). "The Effect of Audit Rates on the Federal Individual Income Tax, 1977-1986." *National Tax Journal* 43(4): 395–409.

Erard, Brian (1992). "The Influence of Tax Audits on Reporting Behavior." In *Why People Pay Taxes: Compliance and Enforcement*, edited by Joel Slemrod.

Fellner, G., R. Sausgruber, and C. Traxler (2013). "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information." *Journal of the European Economic Association* 11(3): 634–660. https://doi.org/10.1111/jeea.12013.

Fleischman, Gary, and Teresa Stephenson (2012). "Client Variables Associated with Four Key Determinants of Demand for Tax Preparer Services: An Exploratory Study." *Accounting Horizons* 26(3).

Gandrud, Christopher (2017). Package "networkD3." Available at tps://cran.r-project.org/web/packages/networkD3/networkD3.pdf.

Government Accountability Office (GAO) (2012). *IRS Could Significantly Increase Revenues by Better Targeting Enforcement Resources.* Report GAO-13–151.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15: 199–236. https://j.mp/2oD70EE.

Hokamp, Sascha, László Gulyás, Matthew Koehler, and Sanith Wijesinghe (2018). "Agent-based modeling and tax evasion: theory and application." *Agent-Based Modeling of Tax Evasion: Theoretical Aspects and Computational Simulations*, Chichester, UK: Wiley and Sons: 1–36.

Internal Revenue Office (2020). *SOI Tax Stats—Tax Stats-at-a-Glance.* https://www.irs.gov/statistics/soi-tax-stats-tax-stats-at-a-glance.

King, Gary, and Richard Nielsen (2019). "Why Propensity Scores Should Not Be Used for Matching." *Political Analysis* 27(4): 435–454. https://j.mp/2ovYGsW.

Klepper, S., M. Mazur, and D. Nagin (1991). "Expert Intermediaries and Legal Compliance: The Case of Tax Preparers." *The Journal of Law and Economics* 34(1): 205–229.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez (2011). "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica* 79(3): 651–692. https://doi.org/10.3982/ECTA9113.

Kopczuk, Wojciech, and Cristian Pop-Eleches (2007). "Electronic filing, tax preparers and participation in the Earned Income Tax Credit." *Journal of Public Economics* 91(7-8): 1351–1367.

Lediga, Collen, Nadine Riedel, and Kristina Strohmaier (2019). *Spatial Tax Enforcement Spillovers: Evidence from South Africa.* Working Paper. https://www.semanticscholar.org/paper/Spatial-Tax-Enforcement-Spillovers-%3A-Evidence-from-Lediga-Riedel/ab31ddc6723a68b532bfe522c3ac8af81c727873.

Meiselman, Ben (2018). "Ghostbusting in Detroit: Evidence on Nonfilers from a Controlled Field Experiment." *Journal of Public Economics* 158(C): 180–193.

Moulton, Brent (1986). "Random group effects and the precision of regression estimates." *Journal of Econometrics* 32(3): 385–397. https://econpapers.repec.org/article/eeeeconom/v_3a32_3ay_3a1986_3ai_3a3_3ap_3a385-397.htm.

Murphy, Kristina (2004). "Aggressive tax planning: Differentiating those playing the game from those who don't." *Journal of Economic Psychology* 25(3): 307–329.

Nicholl, Leigh C., Lucia Lykke, Max McGill, and Alan Plumley (2020). "The Specific Indirect Effect of Correspondence Audits: Moving from Research to Operational Application." *2020 IRS-TPC Joint Research Conference on Tax Administration* (Virtual).

Pinheiro, José (2020). Package "nlme." Available at https://cran.r-project.org/web/packages/nlme/nlme.pdf.

Plumley, Alan (1996). *The Determinants of Individual Income Tax Compliance.* Department of the Treasury, Internal Revenue Service, Publication 1916 (Rev. 11–96).

Rettig, Charles P (2016). "IRS Audit Selection and Classification Processes." *J. Tax Prac. & Proc.* 18: 17.

Rincke, Johannes, and Christian Traxler (2011). "Enforcement Spillovers." *The Review of Economics and Statistics* 93(4): 1224–1234. http://tiny.cc/uriffz.

Sakurai, Yuka, and Valerie Braithwaite (2003). "Taxpayers' perceptions of practitioners: Finding one who is effective and does the right thing?" *Journal of Business Ethics* 46(4): 375–387.

Slemrod, Joel (2016). *Tax Compliance and Enforcement: New Research and Its Policy Implications.* Ross School of Business Working Paper 1302. University of Michigan.

Slemrod, Joel (2019). "Tax Compliance and Enforcement." *Journal of Economic Literature* 57(4): 904–954.

Stephenson, T., G. Fleischman, G. and M. Peterson (2017). "Demand for Tax-Preparation Services: An Exploratory Examination of Client Versus Tax-Preparer Expectation Gaps." *Advances in Taxation* 24, Bingley, UK: Emerald Publishing Limited, 199–231. https://doi.org/10.1108/S1058-749720170000024005.

Appendix

TABLE 5. Estimated coefficients from the linear mixed effects model on total tax with a nine-level treatment group. Estimates reflect multiplicative effect on total tax

Variable	Estimate	85% CI	*p*-value
Married Filing Jointly	1.201	(1.179, 1.223)	< 0.0001
TPI	1	(1, 1)	< 0.0001
Baseline Priority	1	(1, 1)	< 0.0001
Any Wage Income	1.938	(1.899, 1.978)	< 0.0001
Any Other Audits in Study period	0.679	(0.650, 0.708)	< 0.0001
TY2007	1	(0.949, 1.054)	0.991
TY2008	0.909	(0.862, 0.958)	0.0004
TY2009	0.845	(0.800, 0.892)	< 0.0001
TY2010	0.916	(0.866, 0.968)	0.0021
TY2011	0.946	(0.891, 1.004)	0.0681
TY2012	1.069	(1.003, 1.139)	0.0398
TY2013	1.118	(1.044, 1.198)	0.0013
TY2014	1.197	(1.112, 1.289)	< 0.0001
TY2015	1.241	(1.147, 1.343)	< 0.0001
TY2016	1.278	(1.175, 1.391)	< 0.0001
TY2017	1.339	(1.223, 1.465)	< 0.0001
TY2018	1.154	(1.046, 1.272)	0.0041
Treatment 1	0.469	(0.446, 0.494)	< 0.0001
Treatment 2	0.532	(0.499, 0.567)	< 0.0001
Treatment 3	0.643	(0.597, 0.692)	< 0.0001
Treatment 4	0.754	(0.694, 0.819)	< 0.0001
Treatment 5	0.827	(0.754, 0.907)	0.0001
Treatment 6	0.789	(0.717, 0.867)	< 0.0001
Treatment 7	0.864	(0.792, 0.943)	0.001
Treatment 8	1.134	(1.077, 1.195)	< 0.0001
Year 1	0.913	(0.888, 0.939)	< 0.0001
Year 2	0.945	(0.917, 0.975)	0.0004
Year 3	0.947	(0.915, 0.981)	0.0024
Year 4	0.942	(0.906, 0.981)	0.0034
Year 5	0.925	(0.884, 0.967)	0.0007
Year 6	0.91	(0.865, 0.958)	0.0003
Year 7	0.889	(0.840, 0.942)	0.0001
Number of Eligible Clients	0.974	(0.972, 0.975)	< 0.0001
Treatment 1*Year 1	1.375	(1.308, 1.445)	< 0.0001
Treatment 2*Year 1	1.123	(1.052, 1.198)	0.0005
Treatment 3*Year 1	1.08	(1.000, 1.167)	0.0487
Treatment 4*Year 1	1.033	(0.947, 1.127)	0.4577
Treatment 5*Year 1	0.929	(0.842, 1.024)	0.139

TABLE 5. Estimated coefficients from the linear mixed effects model on total tax with a nine-level treatment group. Estimates reflect multiplicative effect on total tax—Continued

Variable	Estimate	85% CI	*p*-value
Treatment 6*Year 1	1.019	(0.922, 1.126)	0.7137
Treatment 7*Year 1	1.021	(0.932, 1.119)	0.6529
Treatment 8*Year 1	0.991	(0.939, 1.045)	0.7374
Treatment 1*Year 2	1.583	(1.505, 1.664)	< 0.0001
Treatment 2*Year 2	1.441	(1.348, 1.541)	< 0.0001
Treatment 3*Year 2	1.131	(1.048, 1.222)	0.0016
Treatment 4*Year 2	1.102	(1.010, 1.202)	0.0289
Treatment 5*Year 2	1.069	(0.969, 1.179)	0.181
Treatment 6*Year 2	0.966	(0.874, 1.068)	0.5022
Treatment 7*Year 2	1.039	(0.948, 1.139)	0.4111
Treatment 8*Year 2	0.952	(0.902, 1.004)	0.069
Treatment 1*Year 3	1.847	(1.755, 1.942)	< 0.0001
Treatment 2*Year 3	1.537	(1.437, 1.644)	< 0.0001
Treatment 3*Year 3	1.382	(1.277, 1.496)	< 0.0001
Treatment 4*Year 3	1.13	(1.035, 1.233)	0.0061
Treatment 5*Year 3	1.028	(0.932, 1.134)	0.5785
Treatment 6*Year 3	1.013	(0.917, 1.120)	0.7959
Treatment 7*Year 3	1.048	(0.956, 1.148)	0.3217
Treatment 8*Year 3	0.994	(0.943, 1.049)	0.8357
Treatment 1*Year 4	2.088	(1.984, 2.198)	< 0.0001
Treatment 2*Year 4	1.709	(1.596, 1.829)	< 0.0001
Treatment 3*Year 4	1.429	(1.319, 1.548)	< 0.0001
Treatment 4*Year 4	1.391	(1.271, 1.523)	< 0.0001
Treatment 5*Year 4	1.117	(1.012, 1.232)	0.0275
Treatment 6*Year 4	1.11	(1.004, 1.227)	0.0409
Treatment 7*Year 4	1.173	(1.070, 1.286)	0.0007
Treatment 8*Year 4	1.008	(0.955, 1.063)	0.7763
Treatment 1*Year 5	2.205	(2.095, 2.321)	< 0.0001
Treatment 2*Year 5	1.866	(1.742, 1.998)	< 0.0001
Treatment 3*Year 5	1.562	(1.441, 1.694)	< 0.0001
Treatment 4*Year 5	1.437	(1.312, 1.575)	< 0.0001
Treatment 5*Year 5	1.294	(1.169, 1.432)	< 0.0001
Treatment 6*Year 5	1.117	(1.010, 1.235)	0.0311
Treatment 7*Year 5	1.171	(1.068, 1.284)	0.0008
Treatment 8*Year 5	0.997	(0.946, 1.052)	0.9266
Treatment 1*Year 6	2.23	(2.117, 2.348)	< 0.0001
Treatment 2*Year 6	1.952	(1.822, 2.092)	< 0.0001
Treatment 3*Year 6	1.588	(1.463, 1.722)	< 0.0001
Treatment 4*Year 6	1.513	(1.380, 1.660)	< 0.0001
Treatment 5*Year 6	1.309	(1.182, 1.451)	< 0.0001

TABLE 5. Estimated coefficients from the linear mixed effects model on total tax with a nine-level treatment group. Estimates reflect multiplicative effect on total tax—Continued

Variable	Estimate	85% CI	*p*-value
Treatment 6*Year 6	1.282	(1.155, 1.424)	< 0.0001
Treatment 7*Year 6	1.119	(1.020, 1.227)	0.0171
Treatment 8*Year 6	1.025	(0.971, 1.081)	0.3695
Treatment 1*Year 7	2.394	(2.266, 2.530)	< 0.0001
Treatment 2*Year 7	2.112	(1.964, 2.272)	< 0.0001
Treatment 3*Year 7	1.632	(1.497, 1.779)	< 0.0001
Treatment 4*Year 7	1.487	(1.347, 1.640)	< 0.0001
Treatment 5*Year 7	1.363	(1.221, 1.522)	< 0.0001
Treatment 6*Year 7	1.326	(1.185, 1.484)	< 0.0001
Treatment 7*Year 7	1.127	(1.018, 1.247)	0.0213
Treatment 8*Year 7	1.033	(0.976, 1.094)	0.2566

EITC Noncompliance: Examining the Roles of the Dynamics of EITC Claims and Paid Preparer Use

Emily Y. Lin, Ankur Patel, and Alexander Yuskavage (Office of Tax Analysis, U.S. Department of the Treasury)[1]

I. Introduction

This paper examines the role taxpayers' experience with prior claims plays in explaining their current noncompliance with the Earned Income Tax Credit (EITC). The EITC is the country's largest benefit program for low- and moderate-income workers; in 2019, over 25 million families received $63 billion in federal EITC. Of the many policy issues related to the credit, taxpayer noncompliance and its causes have been under intense scrutiny by policymakers. To reduce erroneous claims, Congress and the IRS have made frequent changes and adjustments to EITC statutes, regulations, and tax administration strategies. Despite these efforts, the EITC error rate has remained in the range from 23 to 27 percent since 2002 (U.S. Department of the Treasury (various years)).

The complexity of eligibility rules for claiming the EITC with qualifying children is well documented in the literature.[2] These complex rules can make it difficult for taxpayers to navigate through a web of criteria to determine their qualification with certainty. In addition, the IRS lacks third-party information to readily verify several of the eligibility rules, including a claimed child's residency test and the taxpayer's filing status, and therefore cannot confirm the accuracy of some of the claims it receives without an audit. For example, a universal, up-to-date database on a child's living arrangement is not yet available and there is no third-party information report on a taxpayer's self-employment income or marital status. Consequently, the taxpayer's—or her preparer's, if applicable—knowledge about the EITC rules and her familiarity with the administration of the credit by the tax authority can be related to the likelihood that the taxpayer commits an error, either inadvertent or intentional, on her EITC claim.

In this paper, we consider two ways in which a claimant's current compliance is affected by her prior experience with the credit. The first is through the taxpayer's past claiming of the EITC, while the second is through the taxpayer's experience with using a paid return preparer. Without empirical analysis, we cannot be certain if either of these factors has a positive or negative effect on claim accuracy. With respect to claim experience, new claimants can be expected to have a higher error rate than do experienced claimants because they may be less familiar with the eligibility rules. Moreover, as the IRS conducts education and outreach programs for potential EITC recipients, taxpayers who have never claimed the credit are less likely to be aware of these programs or know how to seek appropriate service. In fact, the high turnover rate among EITC recipients is cited by the Treasury Department as a barrier to reducing EITC errors because it creates challenges for the IRS to target compliance activities.[3]

However, it is possible that, once taxpayers accumulate experience in claiming the credit, they know how to take advantage of the gaps in the IRS verification systems and thereby make erroneous claims accordingly. Moreover, if a taxpayer's claim error has never been corrected by an enforcement function, she may not know that there is an error on her claim or she may believe that the error will not be caught by the tax authority.

[1] The authors would like to thank Jacklyn Pi and Navodhya Samarakoon for excellent research assistance, and Edith Brashares, David Splinter, Tatiana Homonoff, Robert McClelland, and the participants in the 2021 IRS/TPC Joint Research Conference for helpful feedback. The findings and views expressed in this paper are those of the authors and do not necessarily reflect those of the Department of the Treasury.

[2] Leibel *et al.* (2020), provides a thorough review of the EITC rules, especially those with respect to claiming a qualifying child, enacted before 2020.

[3] See p. 253 of the Agency Financial Report (U.S. Department of the Treasury (2020)) for a list of factors the Treasury Department identifies as potential barriers to reducing improper payments for refundable tax credits.

In these cases, experienced claimants can make more EITC errors; that is, the taxpayer's past noncompliance persists over time.

Similarly, the effect of using a paid preparer is also uncertain ex ante. As a preparer assists taxpayers in meeting their obligations to comply with the tax law, the preparer also attempts to help taxpayers maximize credits and minimize tax liability. Given the wide variation in the qualification of return preparers, unethical or incompetent preparers can contribute to EITC noncompliance by making either inadvertent or intentional errors on the claim. For example, a paid preparer may direct her client to take a risky position on certain aspects of the claim, knowing that the tax authority likely has priorities set on other areas of EITC noncompliance. It is likewise unclear if we should expect prior use of a paid preparer to correlate positively or negatively with the taxpayer's current compliance level. Past experience with paid preparers can either signal a type of taxpayer who is wary of preparing her own return and thus has little first-hand experience with tax preparation or indicate that the taxpayer can reference professional tax preparation decisions from prior years, compliant or not, when preparing her own return. In addition, it should be noted that the attribution of responsibility with respect to a claim error is not always unambiguous. Book (2007) describes many scenarios in which erroneous EITC claims result from neglect or intentional noncompliance by the taxpayer, the paid preparer, or both.

To better understand how taxpayer experience affects EITC claims, we present trends in publicly released EITC error rates on the basis of audit data from Tax Years 1999 through 2016 and show that EITC error rates have moved little during the past 2 decades. In contrast to the stable error rates are the dynamic patterns exhibited in EITC claims and paid preparer use—the two factors which we identify that could drive EITC compliance outcomes. We find that, in each year, the population of EITC claimants with children is composed of 20 to 25 percent of taxpayers who did not claim the credit in the previous year, i.e., those who need to acquire or update their knowledge about the credit. In addition, 20 years ago, approximately three quarters of EITC claimants with children used a return preparer to help them file the return but only about 55 percent did so in 2018.

We use data from a 2-percent panel of EITC claimants with children to track EITC claiming and preparation methods from 2010 and 2018 for each taxpayer. We discover a wide range of claiming behavior; some taxpayers claimed the credit for only a few years while some claimed the EITC for the entire 9-year period. We also observe variation in claim preparation methods, with high continuation rates among those who self-prepare their returns and a persistent transition into the self-preparation method among those who previously sought assistance from unregulated preparers.

More importantly, in the regression analysis, we find that taxpayer experience matters—in both positive and negative ways. Specifically, whether experienced claimants are more or less likely to make claim errors relative to new claimants in the current year depends on whether a claim they made in the past contained possible compliance issues. Having prior compliance issues increases the probability of noncompliance in the current year, whereas having prior claims without compliance issues reduces the probability of current noncompliance. In addition, return preparation methods and types of preparers used are also positively or negatively associated with EITC noncompliance—depending on the types of EITC errors, types of paid preparers, and whether a prior claim contained possible compliance issues. For taxpayers with prior claims that contained possible compliance issues, returns prepared by unregulated preparers and self-prepared returns are more likely to have errors associated with EITC qualification and less likely to have other general errors that affect the EITC amount. However, the effect of prior preparation methods is generally small compared to the effect of the taxpayer's own claiming experience.

The paper is organized as follows: Section II provides a review of previous studies. Section III presents background on publicly released EITC error rates and describes the indicators of likely EITC noncompliance we construct for the paper. It also shows the correlation between the indicators of likely noncompliance and actual noncompliance detected in audits. Section IV shows trends in the portion of all EITC claimants with children who newly claim the credit in a year and the mix of return preparation methods. Section V contains analysis and results, which consist of three parts: (1) connecting a taxpayer's new claim status and return preparation method to the noncompliance indicators in the cross-section data, (2) leveraging panel data to measure EITC claim dynamics and transitions across preparation methods, and (3) examining the correlation of current noncompliance with prior claims for the credit and their associated preparation methods. Section VI concludes and discusses policy implications.

II. Previous Studies

Using data from a stratified random sample of tax audits, an IRS study (2014) finds that, of all types of EITC claim errors, qualifying child errors, misreporting of self-employment income, and filing status errors account for the most EITC overclaim dollars for Tax Years 2006 through 2008. With respect to the errors associated with qualifying children, the most common mistakes take place when the child claimed does not live with the taxpayers for the required length of time during the year, i.e., more than 6 months, or does not have the required relationship with the taxpayer. Other qualifying child errors that account for smaller shares of total overclaim dollars include failures to meet the age test, the Social Security number (SSN) test, the joint-filing test, and the child disability test as well as failures by the claimant to meet the tiebreaker rules when the child is eligible to be claimed by more than one taxpayer. Further analyzing child-related EITC errors, Leibel *et al.* (2020), shows that the vast majority of children claimed with qualifying child errors had an eligible familial relationship with their claimants and a small portion had lived with the claimants during part of the year.

Because a complete child-related database covering all of these eligibility rules does not exist, the IRS cannot readily verify the accuracy of self-certified eligibility when processing EITC claims. To assist in enforcing the EITC rules, the IRS leverages administrative databases from other agencies or local governments to help detect potential noncompliance with the credit. For example, one of the IRS's computer-scoring systems incorporates the SSN application records available at the Social Security Administration as well as state data on child support orders and cases maintained by the Department of Health and Human Services, among other databases, to help identify potential claim errors (GAO, 2016). In the empirical analysis, our paper uses the claim error indicator derived from this scoring system as one of the noncompliance measures. In addition, Hotz and Scholz (2008), Maag *et al.* (2015), Pergamit *et al.* (2015), and Guyton *et al.* (2017) provide examples and assessments of the extent to which various state-level administrative data or third-party information reports can be used to improve the administration of the EITC.

In the face of the complex eligibility rules, taxpayers, especially those with children, may seek professional help in preparing EITC claims. Early research studying the effects of paid return preparers on tax compliance, e.g., Klepper and Nagin (1989) and Klepper *et al.* (1991), generally finds mixed results, suggesting dual roles played by paid preparers as enforcers of compliance and enablers of noncompliance. Other studies, such as Erard (1993) and DeBacker *et al.* (2021), find that paid tax return preparers tend to increase noncompliance relative to self-prepared returns. Lin (2020) shows that the legislative and administrative changes concerning the paid return preparer industry during 2009 through 2012 moderately increased compliance of EITC claims prepared by paid preparers relative to self-prepared claims.

In addition, a number of studies exist that examine the compliance effect by type of preparer. Langetieg *et al.* (2013) use tax administrative data from the years transitioning into the requirement of preparer taxpayer identification numbers (PTINs) and finds that returns prepared by PTIN holders, consisting of credentialed preparers and otherwise early adopters of PTINs, contain fewer mismatches with third-party information than do returns prepared by non-PTIN holders. In terms of preparer qualifications, GAO (2006) and Treasury Inspector General for Tax Administration (2008) reveal that returns prepared by unlicensed and unenrolled preparers observed in their limited sample studies contain significant errors. Studying EITC claim errors, IRS (2014) discovers a wide range of dollar overclaim rates across different types of paid preparers. Specifically, unenrolled tax preparers have a much higher overclaim rate than do credentialed preparers and preparers from a national tax preparation firm. DeBacker *et al.* (2021) find that, while paid preparers increase income-reporting noncompliance, voluntary return preparers decrease noncompliance among low-income taxpayers.

Other than seeking professional help, taxpayers can learn about the credit through many channels. Tax knowledge can be diffused through a neighborhood effect. For example, Chetty *et al.* (2012) show substantial heterogeneity across 3-digit ZIP codes in the portion of self-employed EITC claimants who report income at the first kink of the EITC schedule, a behavior found to be associated with misreporting of income by self-employed individuals to erroneously maximize the credit (Saez, (2010); Chetty *et al.* (2012)). In addition, when preparing EITC claims with a tax preparation software package, a taxpayer can gain knowledge about how the credit works by gauging the relationship between the resulting refund and the answers she provides to the software. Lin (2020) documents the dramatic increase in the share of software-aided, self-prepared EITC claims between 2004 and 2016—from 20 to 40 percent of all claims. Mortenson and Whitten (2020) not only find evidence of bunching at EITC kink points but find that this bunching is more pronounced for self-employed taxpayers who prepare their own returns and file electronically. Our paper is the first in the field, to our knowledge, to document the transitions across return preparation methods as well as across types of preparers among EITC claimants.

Furthermore, a taxpayer's claiming experience can affect her subsequent compliance with the credit. Once taxpayers claim the credit or are identified by the IRS as potential claimants, they will be targeted by IRS outreach efforts for education and updates on the credit. They may also receive notices or letters from the IRS about their potential EITC noncompliance or for a participation reminder. They could even change their behavior because of the experience of being subject to an EITC enforcement action, such as an audit. Using a panel of taxpayers who at any point received the EITC from 2000 and 2006, Ackerman *et al*. (2009), finds that, while the EITC claiming population each year has a high turnover rate, 70 percent of the sampled taxpayers received the credit for more than 1 year and 11 percent received the credit for all of the years. Using a panel of taxpayers from 1989 to 2006, Dowd and Horowitz (2011) also find a combination of high turnover and persistent EITC claiming. While most households claiming the EITC are found to stop after a couple of years, many of those same households resume claiming again after a few more years.

A few studies, including Gemmell and Ratto (2012), DeBacker *et al.* (2018a), and Beer *et al.* (2019), discover heterogeneity in taxpayer compliance subsequent to an audit. In particular, whether taxpayers improve compliance after an audit depends on whether they received a tax assessment or, more generally, their perceived enforcement risk. Specific to the EITC, DeBacker *et al.* (2018b) find improved compliance and a reduced probability of claiming among taxpayers subsequent to an audit. Similarly, Guyton *et al.* (2018) show that, in the years following a correspondence audit, taxpayers—especially those whose claims had been flagged as noncompliant—became less likely to claim the EITC.

III. EITC Errors

a. Trend in Error Rates

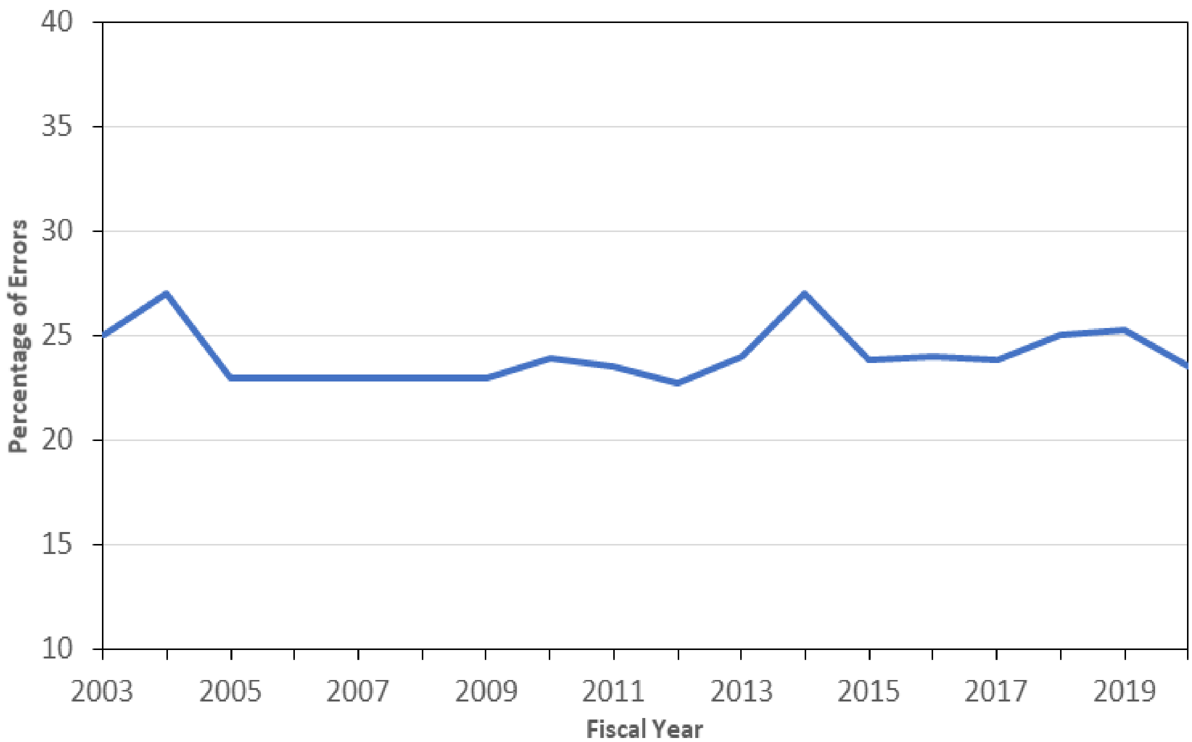
The Department of the Treasury has reported EITC erroneous payments under the requirements of the Improper Payment Information Act of 2002 (IPIA), as amended, since 2002. Under Office of Management and Budget (OMB) guidance, outlays of refundable tax credits are considered "payments" within the scope of the IPIA, which, in general, applies to federal spending programs. An improper payment, as defined, is any payment that should not have been made because it was made either in an incorrect amount or to an ineligible recipient. Furthermore, a payment program is susceptible to significant improper payments—and thus subject to IPIA requirements—when its estimated improper payments exceed (1) both 1.5 percent of program outlays and $10 million of all payments or (2) $100 million at any percent of program outlays. These are considerably low thresholds for the EITC given the large scale of the EITC program and the self-certification nature of the tax system. For programs identified as susceptible to significant improper payments, the agencies are required to estimate the amount of the improper payments and report to Congress the causes of and the steps taken to reduce the improper payments.

Outside the IPIA requirements, the IRS reports EITC errors in the context of the tax gap, which for the EITC includes estimates of the associated tax gap and the misreporting percentage. The EITC improper payment estimates and the EITC tax gap estimates differ in their reporting frequency[4] and estimation methodology to meet respective needs or requirements. However, both estimates are based on the same underlying source of audit data from a stratified random sample of individual tax returns conducted by the IRS, i.e., the Form 1040 Study of the IRS's National Research Program (NRP).

[4] Prior to 2010, the EITC improper payment rate was updated periodically along with the tax gap estimates. There has been annual reporting of the EITC improper payment rate since 2010, while the reporting of the tax gap estimates remains periodic.

We present in Figure 1 the annual estimates of EITC error rates reported in the Department of the Treasury's annual Agency Finance Report (AFR) under the IPIA requirements.[5] Between Fiscal Year 2002 and Fiscal Year 2020, the estimated EITC error rates generally fluctuated from 23 percent to 27 percent. Note that, due to the time needed to conduct tax audits and produce data, there is an average of a 4-year lag between the tax year for which the NRP data are used and the year for which the estimated error rate is reported. As shown in Table A1, the estimated error rates for the reporting period 2002–2020 in Figure 1 are based on data from Tax Years 1999 through 2016. During this period, a couple of notable expansions in the program occurred, including the EITC marriage penalty relief throughout the 2000s and the provision of a larger credit for larger families that was enacted in 2009.[6] There were also a number of legislative or major tax administrative changes aimed at simplifying the EITC rules or improving EITC compliance. Leibel *et al.* (2020) and Lin (2020) document these changes. Other factors, including labor market conditions as well as the level of IRS enforcement and taxpayer service activities, can also impact EITC compliance. For example, the IRS budget (in constant dollars) increased by about 18 percent during the 2000s and then dropped by 20 percent from its peak reached in 2010 (IRS Data Book (various years)). However, despite these changes, there has not been an apparent movement or trend in the estimated EITC error rates.

FIGURE 1. EITC Error Rate by Year



SOURCE: Department of the Treasury Annual Financial Report, 2003–2020.

b. Indicators of Likely EITC Noncompliance

The annual NRP audits of a statistical sample of individual income tax returns provide data to produce reliable estimates of EITC error rates. Specifically, this sample design allows the estimate of the annual improper payment rate to achieve a margin of error of plus or minus 3 percentage points within the 95 percent confidence

[5] When referring to the reported estimates under the IPIA requirements, this paper uses EITC "error" rates and EITC "improper payment" rates interchangeably. The term "error" is more appropriate because the estimates are based on the entire credit claims, not just the outlays portion or the concept of "payments" focused by the statute.

[6] The American Recovery and Reinvestment Act of 2009 (ARRA) created a new category for the EITC that provided a larger credit for families with three or more children. Prior to the ARRA, the credit was limited to two children.

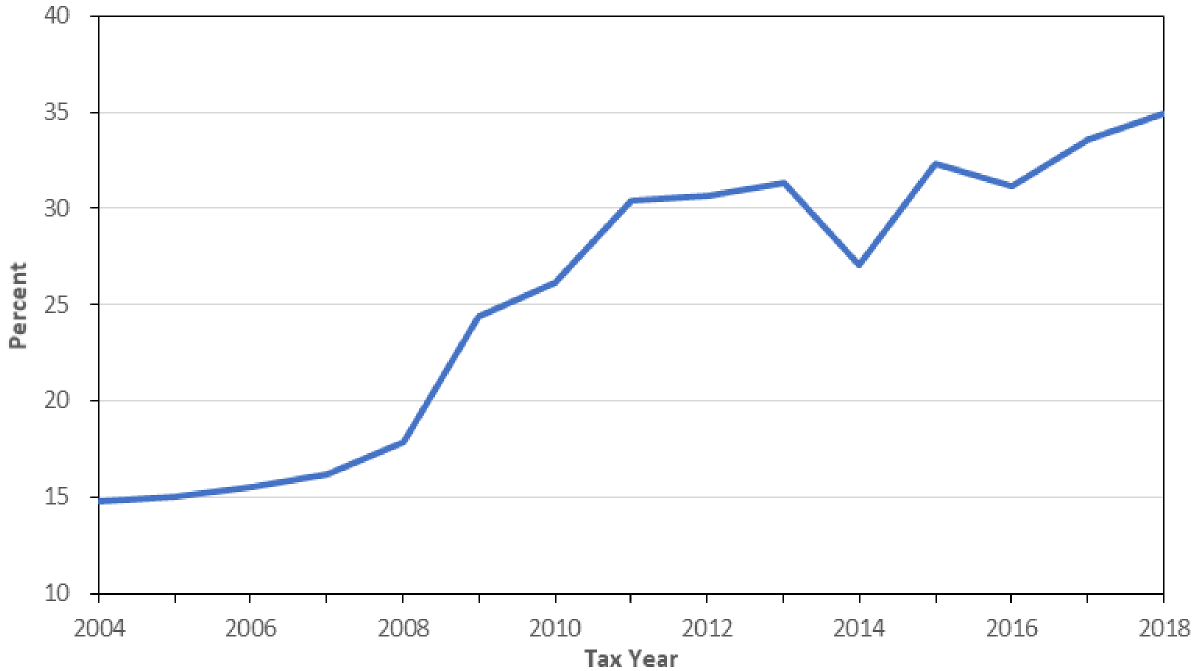
level, pursuant to OMB guidance. In addition, precision on granular estimates regarding the nature of EITC errors can be enhanced by combining multiple years of EITC claims in the NRP. The other advantage of NRP data is that, for each line on the tax return, we have both the amount reported by the taxpayer and the amount that an IRS auditor determines to be the true amount. This line-item adjustment provides a direct measurement regarding the types and extent of tax noncompliance for each taxpayer in the sample. However, the size of the sample, coupled with the method of sample selection, means that taxpayers are unlikely to experience this particular type of audit multiple times. As a result, the data do not allow researchers to follow a taxpayer's compliance behavior over time.

In order to track compliance trends for individual taxpayers, we instead rely on two proxy measures which are available for the entire population of EITC returns: the Discriminant Function System (DIF) score and the Dependent Database (DDb) indicator. The DIF is one of the computer-scoring systems developed by the IRS research staff to rank taxpayer compliance risks to facilitate audit selection. Returns are classified into different categories based on return characteristics including the amount of total positive income reported, the presence of business income, the presence of an EITC claim, etc. Each of these categories is then defined by an "activity code." The examination activity codes for EITC returns are 270 and 271. From the annual population file of federal individual income tax returns processed by the IRS, we construct indicator variables for each EITC claim with qualifying children that is within either the top 20 percent or top 10 percent of DIF scores within that activity code. While having a high DIF score is not a direct measure of noncompliance, the flagged claims should be more likely to be erroneous as they are determined by the IRS algorithm to have a high noncompliance risk relative to 90 or 80 percent of claims.

The DDb is an IRS database that applies multiple decision rules—rules that are connected to the eligibility criteria for child-related tax benefits—to tax returns claiming the EITC and other child-related tax benefits (GAO (2016). The DDb data are available for Tax Years 2004 and after. Figure 2 shows the share of returns claiming the EITC with qualifying children that are identified as breaking any of the rules in the DDb system each year. The share of claims with a DDb rule violation rose from 15 percent in 2004 to 35 percent in 2018. In addition, the ratio jumped in 2009 and 2011, and dipped in 2014. Neither the rising trend nor the large swings were seen in the improper payment rates in Figure 1. Because DDb is a rule-based system, we conclude that significant changes in the portion of returns that break DDb rules can result from updates of the decision rules. For example, an increase in the number of rules included in the system when the IRS obtains, or is newly authorized to use, a reliable database can cause the share of claimants breaking any DDb rule to rise.

Neither the DIF score nor a DDb violation is itself a direct indicator of noncompliance. Therefore, it is necessary to validate that these are in fact predictors of EITC noncompliance. To do so, we look at the relationship between these proxy indicators and the actual results of NRP audits from 2007 to 2013. We focus on two outcomes for NRP audits: (1) the "true" (post-audit) EITC amount conditional on the reported EITC claim and (2) the "true" amount of total tax before refundable credits conditional on the reported tax before credits. The first outcome is a direct measure of EITC and child-related noncompliance while the second one is a measure of general noncompliance which affects tax liability but not EITC qualification. Because the DDb flag is based on issues related to taxpayer and qualifying child eligibility, we expect the DDb flag to strongly predict noncompliance with the EITC while being a weaker predictor of general noncompliance that affects tax before refundable credits. In contrast, because the construction and use of the DIF score is intentionally opaque, it is difficult to say how we expect the DIF score flag to differ in the two regressions of audit adjustments. However, it should be broadly predictive of taxpayer noncompliance in both regressions.

FIGURE 2. Share of EITC Claims with a Violation of Any Dependent Database Rule by Year



SOURCE: Individual Return Transaction File, 2004–2018.

Table 1 shows the results, which validate the noncompliance indicators and give insights into the types of errors associated with each indicator. In the EITC regressions, all three flags have negative coefficients, meaning that the flags correlate with having a downward post-audit adjustment relative to the amount of the EITC claim. The adjustment is the largest in size for the DDb indicator (over $630), followed by DIF10 ($140) and DIF20 ($50). These results are consistent with the hypothesis that the indicators, especially the DDb flag, are predictive of noncompliance with the qualification rules for claiming a dependent child. For the tax liability regressions, all three flags still have the expected sign, which is positive in this case, but the DDb indicator is much less important ($65) than the DIF indicators ($436 and $285, respectively, for DIF10 and DIF20). These results suggest that, compared to the DDb flag, the DIF scores pick up errors associated with other tax elements that result in an upward adjustment to the claimant's tax liability before the application of the EITC.

TABLE 1. Correlation Between Indicators of Likely Noncompliance and Actual Noncompliance, Tax Years 2007–2013

Item	Post-Audit EITC		Post-Audit Total Tax	
	Unweighted	Weighted	Unweighted	Weighted
Reported EITC	0.741***	0.742***		
	(0.010)	(0.000)		
Reported Total Tax			1.616***	1.589***
			(0.062)	(0.001)
DDb Flag	-636.585***	-638.624***	65.223	70.550***
	(38.729)	(0.323)	(71.525)	(0.532)
DIF10 Flag	-136.653***	-191.596***	435.760***	570.928***
	(42.036)	(0.441)	(163.250)	(1.607)
DIF20 Flag	-47.481	-45.933***	284.588**	320.068***
	(33.942)	(0.338)	(126.740)	(1.137)
Year 2007 Dummy	-6.589	19.674***	-333.060	-224.742***
	(40.557)	(0.379)	(231.132)	(1.649)
Year 2008 Dummy	9.162	1.570***	-200.796	-102.766***
	(39.903)	(0.379)	(242.819)	(1.753)
Year 2009 Dummy	84.750**	56.081***	-349.986	-217.522***
	(41.306)	(0.395)	(230.594)	(1.672)
Year 2010 Dummy	-13.015	-29.644***	-263.621	-149.543***
	(43.401)	(0.395)	(230.745)	(1.645)
Year 2011 Dummy	99.039**	104.134***	-268.242	-166.927***
	43.860)	(0.397)	(232.460)	(1.656)
Year 2012 Dummy	95.929**	93.588***	-259.856	-165.753***
	(44.458)	(0.407)	(228.930)	(1.640)
Constant	117.679***	102.356***	401.459*	306.716***
	(35.765)	(0.332)	(235.819)	(1.671)
R-squared	0.445	0.415	0.126	0.148
N	11,091	136,919,373	11,091	136,919,373

SOURCE: NRP 1040 Study, 2007–2013.

Robust standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

IV. Trends in New Claimant Shares and Return Preparation Methods

In this section, we first document the trend in the portion of all claimants of the EITC with qualifying children in a year who did not claim in the previous year, i.e., those whom we define as new claimants. Over the period from 2003 to 2018, there was a persistent turnover rate among the EITC claiming population, which generally ranged between 20 to 25 percent. In addition, the share of new claimants was lower after 2009 than the share in earlier years. Specifically, Table 2 shows that the portion of new claimants stayed stable around 25 percent from 2003 to 2008; it then spiked, reaching 27.4 percent in 2009, before gradually declining to 20 percent in more recent years. Hence, in each year, at least one-fifth of the claiming population is replaced by those who did not receive the credit in the previous year, or those who may need to acquire or update their knowledge about the credit.

TABLE 2. Shares of New Claimants and Return Preparation Methods, All Returns Claiming EITC with Qualifying Children, by Year

Year	Number of Claimants (Million)	New Claim (Did Not Claim in Year t-1)	Self-Prepared	Self-Prepared and E-filed	Use of Any Preparer	Paid Preparer Types					
						CPA	Enrolled Agent	State-Regulated	VITA or TCE	Other Credential	Unknown or No Credential
2003	17.71	24%	25%	14%	75%						
2004	17.82	25%	25%	16%	75%						
2005	18.04	25%	24%	17%	76%						
2006	18.24	25%	25%	19%	75%						
2007	18.79	25%	26%	21%	74%						
2008	19.13	25%	28%	24%	72%						
2009	20.8	27%	31%	27%	69%						
2010	20.86	23%	33%	30%	67%	5%	6%	4%	2%	1%	50%
2011	20.99	23%	35%	33%	65%	5%	6%	4%	2%	1%	48%
2012	20.9	22%	38%	36%	62%	5%	5%	4%	2%	1%	46%
2013	21.19	23%	40%	38%	60%	5%	5%	4%	2%	1%	44%
2014	20.88	21%	41%	40%	59%	5%	5%	4%	2%	1%	43%
2015	20.66	21%	43%	42%	57%	4%	5%	3%	2%	1%	42%
2016	20.2	20%	44%	42%	56%	4%	5%	3%	2%	1%	42%
2017	19.72	20%	44%	42%	56%	4%	4%	3%	2%	1%	42%
2018	19.16	20%	45%	43%	55%	4%	4%	3%	2%	1%	41%

SOURCE: Individual Return Transaction File, 2003–2018.

In contrast to the relatively stable EITC turnover rate, the period between 2003 and 2018 saw transformational changes in return preparation methods used by EITC claimants with children. About 75 percent of the taxpayers who claimed the EITC for qualifying children used a preparer in 2003, but only 55 percent did so in 2018, a drop of 20 percentage points over the period. Persistent reductions in the share of claimants using a preparer occurred throughout years 2007 to 2015, with the year-to-year decline ranging from 1 percentage point (2007) to over 2 percentage points (2008 through 2012). Meanwhile, there was rapid growth in the portion of all claimants who filed software-aided returns without using a preparer (i.e., the percent of self-prepared e-filed). The share of software-aided, self-prepared returns more than tripled from 14 percent to 43 percent, a rise of 30 percentage points, over the period. Accordingly, the share of paper-filed EITC returns claiming children decreased by 10 percentage points, from 11 percent in 2003 to below 2 percent in 2018.

Data on Preparer Identification Numbers (PTINs), available beginning in Tax Year 2010, shed light on the types of preparers used by EITC claimants with children—e.g., paid preparers versus those free of cost, or credentialed preparers versus unregulated preparers. Preparers who do not report any credential or license to the IRS, whom we broadly refer to as unregulated preparers, assisted in preparing the vast majority of all claims prepared by any types of paid preparers. The category of unregulated preparers includes a wide range of backgrounds and qualifications. It can include anyone, from those with no training at all to those who have substantial experience or those who work at a tax preparation chain.[7] Although the share of unregulated-preparer

[7] In theory, it can also include tax preparers who do have relevant credentials but choose not to report them to the IRS. We have no data on the extent, if any, to which this occurs.

claims dropped from 50 percent to 41 percent of the total EITC claims with children, this category remained the dominant type across all categories of preparer-assisted claims.

In contrast, each type of the credentialed, or otherwise state-regulated, preparer constituted a modest or negligible portion of all EITC claims claiming children. Of all types of credentialed preparers, enrolled agents (EAs) and certified public accounts (CPAs) prepared the most claims, followed by state-regulated preparers and those with all other types of credentials—attorneys, enrolled actuaries, and enrolled retirement plan agents—combined. Between 2010 and 2018, the share of all claims prepared by any credentialed preparer dropped from 15 percent to 12 percent, roughly in the same proportion as the decline in the claims prepared by unregulated preparers. Hence, there has been little change to the relative market representation between these two types of paid preparers. Finally, over the period, the IRS's Volunteer Income Tax Assistance (VITA) and Tax Counseling for the Elderly (TCE) programs, which offer free tax return preparation to qualified lower-income individuals, regularly prepared about 2 percent of all EITC claims claiming children.[8]

V. Analysis and Results

In this section, we conduct analysis using a 2 percent sample of taxpayers claiming the EITC with qualifying children between 2010 and 2018. We begin with the 2-percent repeated cross-section samples to provide summary statistics on the relationship between the two proxy variables of taxpayer knowledge—new claimant status and return preparation method—and the two indicators of likely EITC noncompliance—DDb and DIF score flags[9]—described in the previous sections. We then follow claimants in the 2010 sample every year after that regardless of EITC claiming for a panel data study. We use data from this 2-percent panel of taxpayers to track EITC claiming and preparation methods over time for each taxpayer. We investigate how these experiences are associated with the taxpayer's subsequent EITC noncompliance using the 2-percent panel data.

a. Connecting the New Claim Status and Return Preparation Method to EITC Noncompliance—Cross-Section Data Analysis

Table 3 shows year-by-year noncompliance outcomes by new claimant status and return preparation method. Specifically, using repeated cross-section samples, we regress each of the two likely noncompliance flags against the indicator of new claimant status interacted with the year dummy variables (columns (1) and (3)), or against the indicator of a particular type of tax preparer interacted with the year dummy variables (columns (2) and (4)). The omitted group in columns (2) and (4) is the self-preparation category.

[8] A fourth category of tax preparer which presumably appears in our data is the so-called "ghost preparer"—a preparer who is paid to prepare a return but does not provide a signature or PTIN on the form. Because of the nature of these returns, we cannot identify them in our data.

[9] We do not run the analysis for DIF20 because DIF10 is shown to be a stronger predictor for tax and EITC adjustments than is DIF20.

TABLE 3. Linear Probability Model of EITC Noncompliance, by New Claimant Status and Return Preparation Method

Tax Year	Violating Any DDb Rule				DIF Score in Top Decile			
	New Claimant (1)	Preparation Methods, Self-Preparation Omitted (2)			New Claimant (3)	Preparation Methods, Self-Preparation Omitted (4)		
		Credentialed Preparer	Unregulated Preparer	VITA/TCE		Credentialed Preparer	Unregulated Preparer	VITA/TCE
2010	0.043***	-0.099***	-0.006***	-0.152***	0.014***	0.051***	-0.009***	-0.028***
	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)
2011	0.077***	-0.063***	0.039***	-0.098***	0.014***	0.048***	-0.008***	-0.025***
	(0.002)	(0.002)	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)	(0.003)
2012	0.076***	-0.066***	0.039***	-0.098***	0.015***	0.049***	-0.006***	-0.028***
	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)
2013	0.088***	-0.062***	0.047***	-0.088***	0.012***	0.044***	-0.007***	-0.034***
	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)
2014	0.033***	-0.092***	(0.002)	-0.107***	0.015***	0.049***	-0.003***	-0.033***
	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.003)
2015	0.082***	-0.053***	0.058***	-0.071***	0.017***	0.051***	0.001	-0.028***
	(0.002)	(0.002)	(0.001)	(0.005)	(0.001)	(0.002)	(0.001)	(0.003)
2016	0.056***	-0.069***	0.044***	-0.070***	0.022***	0.051***	0.002**	-0.028***
	(0.002)	(0.002)	(0.001)	(0.005)	(0.001)	(0.002)	(0.001)	(0.003)
2017	0.090***	-0.048***	0.073***	-0.053***	0.019***	0.047***	0.003***	-0.035***
	(0.002)	(0.002)	(0.001)	(0.005)	(0.001)	(0.002)	(0.001)	(0.003)
2018	0.088***	-0.038***	0.089***	-0.058***	-0.029***	-0.012***	0.026***	-0.064***
	(0.002)	(0.002)	(0.001)	(0.006)	(0.001)	(0.001)	(0.001)	(0.002)

Robust standard errors are in parentheses. ** p<0.05, *** p<0.01.

NOTES: The sample includes 2 percent of all taxpayers claiming the EITC with qualifying children each year 2010-2018. Regressions also include a constant term. N=3,694,729 for all years.

Columns (1) and (3) suggest that being a new claimant is positively related to the probability of violating any DDb rule and, in all but one year, positively related to the probability of having a high DIF score. Specifically, new claimants are more likely to break DDb rules by 4 to 9 percentage points, and more likely to have a high DIF score by about 1 to 2 percentage points, than are those with prior claiming experiences. Note that this analysis does not consider the effects of the variables that might be simultaneously correlated to both tax filing decisions and individual compliance behavior. We will explore the effects of past claim behavior and outcomes controlling for individual and neighborhood characteristics in the regression analysis in subsection c.

With respect to return preparation methods, the most consistent result shown in Table 3 is the positive correlation between using the VITA or TCE service and EITC compliance. Not only are returns filed using this preparation method less likely to have a noncompliance flag than are self-prepared returns, but they are also less likely to have a noncompliance flag relative to returns prepared by either type of paid preparer. This result is consistent with other work, such as the finding in Mortensen and Whitten (2020), which finds significantly less clustering at EITC kink points in returns prepared through IRS-associated programs such as VITA services.[10]

For paid preparers, the results are more complex. Returns prepared by a credentialed paid preparer have a much lower probability of breaking DDb rules, by 4 to 10 percentage points depending on the year, than do self-prepared returns, but they are in general more likely to have a high DIF score than do self-prepared returns by about 5 percentage points. In contrast, returns prepared by an unregulated paid preparer, in most years, are much more likely to break DDb rules than are self-prepared returns, making them the bottom category in DDb compliance rankings across the four return preparation methods. However, these returns had a slightly lower, or about equal, probability of having a DIF score in the top decile relative to self-prepared returns before 2016.

Results in Table 3 suggest that different types of preparers may have different impacts on varying types of EITC noncompliance, relative to self-prepared returns. However, considering that many factors that contribute to the decision of using a specific type of preparer may be correlated with individual compliance behavior, we cannot readily establish a causal relationship on the basis of these results.

b. Dynamics of EITC Claims and Return Preparation Methods—Panel Data Analysis

In any given year, an EITC claimant could enter into one of the following three states in the subsequent year for filing and EITC statuses: not filing, filing and not claiming the EITC, and continuing the EITC claim.[11] If she stops claiming the credit in a year, for the year that follows, she could resume the claim or continue to be a nonclaimant. To better understand these dynamics, Table 4 documents these between-year transitions for the panel of taxpayers who claimed the EITC with qualifying children in 2010. All of the probabilities in the table are conditional on the previous-year status of the taxpayer.

[10] It is unclear how much of this effect to attribute to the behavior of the tax preparation services themselves and how much to self-selection on the part of clients willing to seek help at an IRS-associated program. See Goldin *et al.* (2021) for a recent study that finds difficulty in inducing low-income taxpayers to accept free tax assistance while simultaneously finding positive outcomes for those who do choose to participate.

[11] The first two states are consolidated into the category of "stop EITC claim." To simplify the analysis, we include taxpayers claiming the EITC without qualifying children in the category of having an EITC claim after 2010.

TABLE 4. Transition of EITC Status 2011-2018 Among Claimants in 2010

Tax Year	Transition Rates				Sum of Previous Columns	Making EITC Claim = (1) + (3)
	EITC Claim in Year *t-1*		No EITC Claim or Not Filing in Year *t-1*			
	Claim in *t* (1)	Stop Claim in *t* (2)	Resume Claim in *t* (3)	Continue No Claim in *t* (4)	(5)	(6)
2010						1
2011	0.779	0.221			1	0.779
2012	0.634	0.145	0.057	0.164	1	0.691
2013	0.569	0.121	0.063	0.246	1	0.633
2014	0.517	0.116	0.060	0.308	1	0.577
2015	0.469	0.108	0.059	0.364	1	0.528
2016	0.425	0.103	0.056	0.416	1	0.481
2017	0.385	0.096	0.054	0.466	1	0.439
2018	0.349	0.090	0.052	0.510	1	0.401

NOTES: The sample includes a panel of 2 percent of EITC claimants with qualifying children in 2010.
N=420,780.

Of the claimants in 2010, about 22 percent stopped making an EITC claim in 2011, because they either became a nonfiler or filed a return without an EITC claim. This ratio is consistent with the turnover rate documented in the previous section. Most of these taxpayers continued to be nonclaimants in 2012, as indicated by the rate of 16 percent in column (4). In fact, the taxpayers in the analysis appear to consist of a portion of claimants who exited shortly and stayed out of the claim status, a portion who made persistent claims, and the rest who transitioned in and out of the claim status. First, the exit rates were relatively high during the first few years subsequent to 2010 (column (2)). Each year, an increasing portion of the initial claimants stopped making claims for at least 2 consecutive years (column (4)). However, at the end of the 9-year period, there was still a nontrivial portion, about 40 percent, of the initial claimants who claimed the EITC, including 35 percent who did so for at least 2 consecutive years (column (1)) and 5 percent who resumed claiming in 2018 (column (3)). Additionally, a stable share of the initial claimants, about 5 to 6 percent, stopped and then resumed their claims each year.[12]

Another way to evaluate the dynamics of EITC claims is to assess the distribution of the number of claims made by taxpayers over a period. Table 5 presents the frequencies with which an EITC claim was made by the 2010 panel of claimants during the subsequent 8 years. About 30 percent of taxpayers made 3 or fewer EITC claims and yet 34 percent made 8 or 9 claims over the 9-year period. These disparate claiming behaviors across taxpayers indicate that the claiming population each year likely comprises a wide range of experiences with the credit. In the next section, we leverage this variation in prior experience across taxpayers to examine the effect of claiming experience. Although our results are not directly comparable to Dowd and Horowitz (2011), we find our results to be broadly consistent with theirs in the pattern of re-entry after short periods of exit and the overall distribution of claim tenures.

[12] These claiming patterns are affected by macroeconomic conditions and therefore may not apply in other decades.

TABLE 5. Number of EITC Claims 2010–2018 Among Claimants in 2010

Number of Claims	Frequency	Percentage (%)	Cumulative Percentage (%)
1	42,319	10.1	10.1
2	42,565	10.1	20.2
3	40,439	9.6	29.8
4	39,570	9.4	39.2
5	38,193	9.1	48.3
6	37,567	8.9	57.2
7	37,788	9.0	66.2
8	43,485	10.3	76.5
9	98,854	23.5	100

NOTES: The sample includes a panel of 2 percent of EITC claimants with qualifying children in 2010.
N=420,780.

We also analyze the dynamics of return preparation methods by looking at the transition probabilities for use of paid preparers within the panel of taxpayers. Following the initial year in 2010, we document the preparation methods used by these taxpayers from 2011 through 2018 when they claimed the EITC with qualifying children, conditional on the preparation method they used in 2010. Table 6 shows the results, where each of the three panels is separated by the claimant's initial preparation method.[13] Looking at the top panel for users of a credentialed preparer in 2010, we find that, when they claimed the credit again in subsequent years, they gradually shifted to using assistance from an unregulated preparer or, to a lesser extent, self-prepared their returns. Of the taxpayers who used a credentialed preparer in 2010 and who also claimed the credit in 2018, less than one half—about 45 percent—used a credentialed preparer in 2018 whereas 31 percent used an unregulated preparer and 23 percent did not use a preparer at all.

[13] In this table, the decline in sample size over time reflects the number of taxpayers in the 2010 panel who no longer claimed the EITC with qualifying children in the following years.

TABLE 6. Shares of Return Preparation Methods in 2011–2018 Among EITC Claimants with Qualifying Children in 2010

Tax Year	Credentialed Preparer (1)	Unregulated Preparer (2)	VITA/TCE (3)	Self-Preparation (4)	Sample Size
Used a Credentialed Paid Preparer in 2010					
2010	1				63,298
2011	0.752	0.183	0.004	0.062	47,400
2012	0.669	0.222	0.007	0.103	41,058
2013	0.614	0.247	0.008	0.130	36,640
2014	0.570	0.262	0.010	0.157	32,703
2015	0.536	0.273	0.010	0.181	29,172
2016	0.507	0.285	0.011	0.198	25,899
2017	0.475	0.301	0.010	0.213	23,105
2018	0.449	0.308	0.010	0.232	20,515
Used an Unregulated Paid Preparer in 2010					
2010		1			211,029
2011	0.060	0.838	0.006	0.096	164,239
2012	0.071	0.766	0.009	0.153	144,610
2013	0.077	0.719	0.012	0.193	131,320
2014	0.081	0.679	0.013	0.227	119,274
2015	0.079	0.649	0.013	0.260	109,319
2016	0.077	0.634	0.013	0.275	99,338
2017	0.077	0.621	0.012	0.290	90,657
2018	0.075	0.607	0.012	0.307	82,730
Used No Paid Preparer in 2010					
2010			0.054	0.946	146,453
2011	0.024	0.103	0.043	0.830	108,282
2012	0.032	0.131	0.041	0.796	94,761
2013	0.038	0.147	0.039	0.776	85,196
2014	0.043	0.160	0.039	0.758	76,337
2015	0.045	0.173	0.034	0.748	68,624
2016	0.048	0.191	0.033	0.728	61,514
2017	0.049	0.204	0.029	0.718	55,553
2018	0.049	0.214	0.027	0.710	49,921

NOTES: The sample includes a panel of 2 percent of EITC claimants with qualifying children in 2010.
N=420,780 in 2010.

Of users of an unregulated preparer in 2010, a higher share continued to use this type of paid preparer for their subsequent claims—about 60 percent used an unregulated preparer in 2018 when they claimed the credit. Taxpayers who switched claim preparation methods predominately began to self-prepare, with 30 percent of those who claimed the credit in 2018 not using a preparer.[14] A small but persistent portion of the initial users of

[14] The extent to which these returns were prepared by a "ghost preparer," i.e., a preparer who was paid for the return preparation service but refused to sign on the taxpayer's return, is unknown.

an unregulated preparer switched to using a credentialed preparer or the VITA or TCE service for their subsequent EITC claims. Finally, those who did not use a paid preparer in 2010 were the most likely of the three types of taxpayers shown in this table to keep using the same preparation method for their subsequent EITC claims. Over 70 percent of those taxpayers who claimed the EITC in 2018 still self-prepared their returns. If they switched, they were disproportionately more likely to use an unregulated preparer than to use a credentialed preparer.

Overall, Table 6 offers a few explanations for the increasing use of the self-preparation method among EITC claimants through the lens of transition analysis. First, the continuation rate among those who do not use a paid preparer is high; that is, once taxpayers self-prepare their returns, they are not likely to seek help from a paid preparer in the subsequent years. Second, a nontrivial share of taxpayers who previously used an unregulated preparer transitioned into the self-preparation method for their subsequent EITC claims, and a smaller portion transitioned into the self-preparation method from prior use of a credentialed preparer.

Next, we turn to the question of what causes a taxpayer to seek professional help for preparing her EITC claim in a given year. We hypothesize that the decision is affected by the taxpayer's prior return preparation method, her familiarity with the credit, and individual and neighborhood characteristics. Table 7 shows the regression results for the panel of taxpayers, who filed an EITC claim with qualifying children in 2010, during the period between 2010 and 2018 when they filed a tax return. We run the regressions on a few specifications, column (1) for linear probability model over the entire time period, column (2) for a linear probability model with individual fixed effects, columns (3) only for years in which an EITC claim was made, and columns (4) and (5) for selected later years only.

TABLE 7. Use of a Paid Preparer

Item	Full 2010 Panel		Only EITC Claims	Full 2010 Panel	
	OLS 2010–2018 (1)	FE LPM 2010–2018 (2)	FE LPM 2010–2018 (3)	FE LPM 2013–2018 (4)	FE LPM 2015–2018 (5)
Preparer Use, Prior Year	0.664***	0.231***	0.199***	0.139***	-0.030***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Log # of Prior Preparer Uses	0.103***	0.345***	0.398***	0.516***	0.860***
	(0.000)	(0.002)	(0.003)	(0.003)	(0.006)
EIC Claim, Prior Year	0.021***	0.022***	0.022***	0.020***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log # of Prior EITC Claims	-0.049***	-0.140***	-0.248***	-0.138***	-0.145***
	(0.000)	(0.002)	(0.004)	(0.003)	(0.006)
Log of AGI	-0.002***	0.010***	0.011***	0.013***	0.013***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Children	0.012***	0.019***	0.009***	0.016***	0.016***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Log of Avg AGI, ZIP	0.001	-0.004**	-0.003*	-0.008***	-0.010***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Log of Taxpayers, ZIP	0.006***	0.004***	0.005***	0.004***	0.002**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Paid Preparer Rate, ZIP	0.088***	0.212***	0.244***	0.145***	0.157***
	(0.010)	(0.023)	(0.028)	(0.031)	(0.043)
E-file Rate, ZIP	-0.155***	0.017	0.075***	-0.001	0.035
	(0.010)	(0.023)	(0.028)	(0.032)	(0.044)
VITA/TCE Rate, ZIP	-0.026**	-0.193***	-0.208***	-0.222***	-0.259***
	(0.012)	(0.027)	(0.033)	(0.034)	(0.046)
2012 Dummy	-0.008***	-0.037***	-0.031***		
	(0.001)	(0.001)	(0.001)		
2013 Dummy	-0.010***	-0.066***	-0.052***	0.164***	
	(0.001)	0.001)	(0.001)	(0.002)	
2014 Dummy	-0.012***	-0.090***	-0.070***	0.124***	
	(0.001)	(0.001)	(0.002)	(0.001)	
2015 Dummy	-0.016***	-0.115***	-0.089***	0.084***	0.152***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
2016 Dummy	-0.011***	-0.129***	-0.095***	0.058***	0.100***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
2017 Dummy	-0.013***	-0.145***	-0.105***	0.030***	0.050***
	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)
2018 Dummy	-0.017***	-0.165***	-0.117***		
	(0.001)	(0.002)	(0.003)		
N	2,364,313	2,364,313	1,586,320	1,701,694	1,100,778

* p<0.10, ** p<0.05, *** p<0.01.

NOTES: The sample includes a panel of 2 percent of EITC claimants with qualifying children in 2010. Regressions also include a constant term. Standard errors are clustered on Taxpayer Identification Number (TIN) .

The variables of interest are the four coefficients for prior experiences: having a recent EITC claim, having recently used a paid preparer, (log) total prior EITC claims since 2000, and (log) total prior returns done by a paid preparer since 2000. In column (1), both of the paid preparer-related coefficients are positive, with the one for recent paid preparer use being particularly large. This result suggests repeated uses of paid preparers, likely driven by the group using unregulated preparers given the pattern shown in Table 6. In contrast, while the variable for a recent EITC claim is positive, the coefficient for total past EITC claims is negative. It is possible that the need to seek assistance from a paid preparer is lower once taxpayers accumulate experience with making EITC claims.

When the regression controls for underlying unobservable characteristics, column (2) shows a much smaller magnitude for the coefficient of recent paid preparer and a somewhat larger magnitude for the coefficient of the total prior claims using a paid preparer. This indicates that selection into paid preparer use is correlated with individual preferences or other long-run characteristics. Given our interest in the effects of the taxpayer's past experiences in using a preparer and claiming the EITC, we consider fixed effect model a preferred specification for isolating the effects of these variables. In addition, other variables that are positively associated with the decision to use a paid preparer include taxpayer income, number of children, and the share of taxpayers in the ZIP code using a paid preparer.

Column (3) of Table 7 shows the results of restricting our fixed-effect linear probability model to those returns making an EITC claim for qualifying children. Comparing the results to those in column (2), we see that the differences are relatively minor. The aforementioned effects concerning the decision to use a paid preparer are generally applicable to EITC claims. Another issue of interest is the possible change in behavior across different points in time due to differences in experiences and/or other exogenous factors. Column (3) combines all observations from 2010 through 2018, but it is possible that the coefficients that would be estimated for behavior in 2011 are quite different from the ones that would be estimated for 2018. For example, the choice of preparation methods may have become more dynamic in recent years or the marginal value of using a paid preparer may have changed after repeated uses of a paid preparer by the taxpayer.

Columns (4) and (5) show the regression results of limiting our regression to years after 2012 and 2014, respectively. We can see different effects across different points in time over the panel, based on the smaller, or even a negative, coefficient of the preparer use in the previous year. On the other hand, the coefficient on total prior preparer experience is much larger than that in column (3). It is possible that this larger effect comes from long-term users of paid preparers, as taxpayers included in columns (4) and (5) have a longer look-back period than do those in column (3) for counting prior preparer uses. In contrast, the effect of either recent or total past EITC claims does not appear to change over time for any of these specifications.

c. Impacts of Prior Claims and Preparer Use on EITC Errors—Panel Data Analysis

Better compliance (as measured by our two indicators) may be more likely for those who have claimed the EITC in the past, as taxpayers become more familiar with the complex rules through repetition. In Table 3, we find that new claimants are more likely to break a DDb rule or have a high DIF score. In this section, we further explore the relationship between past EITC claims and current compliance using the longitudinal nature of the 2-percent EITC panel.

The 2-percent EITC panel tracks taxpayers using their tax identification number, regardless of claiming status. Therefore, we can observe years, prior to the current one, in which a particular taxpayer claimed or did not claim the EITC. To construct a measure of past claiming for each EITC claim, we look back 6 years and count the number of years in which the EITC was claimed.[15] In addition, we differentiate between past claims that broke a DDb rule or scored a high DIF score. We use indicators to represent both counts of past EITC claiming in a regression where the dependent variables are our measures of noncompliance, with and without controls.[16]

[15] The 2-percent sample contains Tax Years 2010 through 2018, and we use returns from Tax Years 2004 through 2018 to construct a 6-year look-back period for each taxpayer. To ensure a consistent 6-year look-back window for all ages, we limit the analysis to taxpayers who are 24 or older.

[16] Taxpayer controls include indicators for 5-year age bins, filing status, number of dependents under 24, $10,000 AGI bins, tax years, and the 3-digit ZIP code. We do not use taxpayer fixed effects as controls in this part of the analysis because the covariates of interest have a long look-back period, and therefore may not be time-varying over the period of interest.

TABLE 8. Past EITC Claiming Experience and Current Claim Noncompliance

Item	Violating Any DDb Rule (1)		DIF Score in Top Decile (2)		Violating Any DDb Rule (3)		DIF Score in Top Decile (4)	
Number of Past EITC Claims Without a DDb or DIF10 Flag								
0, omitted								
1	-0.044	(0.001)	-0.041	(0.001)	-0.060	(0.001)	-0.026	(0.000)
2	-0.068	(0.001)	-0.062	(0.001)	-0.095	(0.001)	-0.034	(0.001)
3	-0.092	(0.001)	-0.078	(0.001)	-0.114	(0.001)	-0.038	(0.001)
4+	-0.125	(0.001)	-0.092	(0.001)	-0.095	(0.001)	-0.044	(0.001)
Number of Past EITC Claims With a DDb or DIF10 Flag								
1	0.190	(0.001)	0.073	(0.001)	0.115	(0.001)	0.079	(0.001)
2	0.294	(0.001)	0.082	(0.001)	0.189	(0.001)	0.111	(0.001)
3	0.338	(0.003)	0.100	(0.002)	0.226	(0.002)	0.145	(0.002)
4+	0.259	(0.005)	0.222	(0.005)	0.189	(0.003)	0.272	(0.004)
Controls	No		No		Yes		Yes	

NOTES: The sample includes a panel of 2 percent of taxpayers 24 years or older claiming the EITC with qualifying children. Past EITC claims are measured from 2004, and the regression sample limited to 2010-2018. Controls include indicators for 5-year age bins, filing status, number of dependents, AGI bin, tax years, and 3-digit ZIP code. Standard errors are clustered on TIN.

Table 8 contains the results. We find that having past EITC claims without a violation of DDb rules or a DIF score in the top decile is negatively associated with a violation of a DDb rule or a high DIF score in the current year. Furthermore, this negative association is generally increasing in absolute value with the number of past EITC claims, consistent with either learning from experience or persistence in behavior by compliant individuals. Controlling for taxpayer characteristics and ZIP code fixed effects mitigates this effect. The effects of past claims in columns (3) and (4) range from -6 to -11 percentage points where the DDb flag is the outcome and -3 to -4 percentage points where DIF score is the outcome. In contrast, having past claims that had possible compliance issues is associated with a higher likelihood in the current year of violating a DDb rule or scoring a high DIF score. Interestingly, this association is increasing in the number of past claims that contained a possible compliance issue. These effects are large: having three or more past claims with a compliance issue increases the probability of current noncompliance by more than 20 percentage points.

Taken together, these associations are consistent with past experience exerting some influence over claiming and noncompliance in the current period. These results may indicate that taxpayers are gaining knowledge about how to claim the EITC without compliance issues if they have previously done so successfully. However, the opposite may also be true; the more taxpayers claim the EITC with a violation of a DDb rule or high DIF score, the more likely they will do so in the future. This may be due to claim errors not being corrected in some way (through a correspondence audit, for example), or because taxpayers are getting better at working around the compliance rules.

Next, we explore how past preparer use, in addition to EITC claiming, might be associated with our measures of current noncompliance. To differentiate the previous results by preparer type, we observe how past EITC claiming interacts with preparer use and type. Because the preparer-type data began in 2010, we restrict our analysis to claims made from 2015 to 2018 with a look-back period beginning in 2010. In particular, we construct a set of indicator variables equal to 1 if at any time during the look-back period a taxpayer claimed the EITC without violating a DDb rule or scoring a high DIF score and self-prepared, used a credentialed preparer or used an unregulated preparer. We do the same for those with a past violation of a DDb rule or high DIF score. We add these indicators to the previous regression.

Table 9 presents the results. Regardless of preparer type, the association between past EITC claims without a possible compliance issue and current possible noncompliance is small and estimates are nearly 0 when we add taxpayer controls. For those with past EITC claims with possible compliance issues, we find different results based on the noncompliance measure we use. Focusing on the results with controls (columns (3) and (4)), those who self-prepared or used an unregulated preparer in past EITC claims with compliance issues were more likely to trip a DDb rule in the current year. However, the association is negative when DIF score is used to construct the outcome. In column (3), we can see that those who used a credentialed preparer in past EITC claims with compliance issues were less likely to violate a DDb rule in their current claims. Overall, the coefficients are much smaller in Table 9 relative to what we found in Table 8, suggesting that the preparation method used in a prior claim may be less important than the overall past experience of claiming the EITC.

TABLE 9. Past Preparer Use and Current Claim Noncompliance

Item	Violating Any DDb Rule (1)		DIF Score in Top Decile (2)		Violating Any DDb Rule (3)		DIF Score in Top Decile (4)	
Indicator of Past EITC Claims Without a DDb or DIF10 Flag if That Return…								
was self-prepared	0.010	(0.001)	-0.014	(0.001)	0.000	(0.001)	0.000	(0.001)
had a credentialed preparer	-0.009	(0.001)	0.007	(0.001)	0.002	(0.001)	0.003	(0.001)
had an unregulated preparer	0.029	(0.001)	-0.002	(0.001)	0.000	(0.001)	-0.001	(0.001)
Indicator of Past EITC Claims With a DDb or DIF10 Flag if That Return…								
was self-prepared	0.075	(0.004)	-0.044	(0.003)	0.029	(0.003)	-0.033	(0.003)
had a credentialed preparer	-0.038	(0.005)	0.007	(0.003)	-0.020	(0.003)	0.001	(0.003)
had an unregulated preparer	0.076	(0.004)	-0.054	(0.003)	0.031	(0.003)	-0.043	(0.003)
Controls		No		No		Yes		Yes

NOTES: The sample includes a panel of 2 percent of taxpayers 24 years or older claiming the EITC with qualifying children. Past EITC claims are measured from 2010, and the regression sample limited to 2015-2018. Controls include indicators for 5-year age bins, filing status, number of dependents, AGI bin, tax years, and 3-digit ZIP code. Standard errors are clustered on TIN.

VI. Conclusions

In this paper, we analyze EITC noncompliance through the lens of claim experience and uses of preparation methods. We find that taxpayers' prior experience matters. At the first glance, it appears that those without prior claims are more likely to make EITC errors than are those with prior claim experiences. When taking a closer look, we find that the effect of prior experience depends on whether the claim made in the past contained possible compliance issues. Specifically, having prior compliance issues increases the probability of noncompliance in the current year whereas having prior claims without compliance issues reduces the probability of current noncompliance. Moreover, return preparation methods or types of preparers are positively or negatively associated with EITC noncompliance, depending on the types of errors and types of preparers, but the effect of prior preparation methods is generally small compared to the effect of the taxpayer's claim experience.

These findings help explain why EITC error rates have been relatively stable over time. Each year, nearly a quarter of all claimants with children did not claim the credit in the previous year and thus may need to acquire or update their knowledge about the credit. In addition, of those who make repeated claims, past claim compliance is predictive of current claim compliance, making substantial changes in the error rate associated with this portion of the claiming population to be unlikely. Also, while we find persistence in return preparation behavior in taxpayers' decisions to use a paid preparer, there have been significant changes in return preparation methods among EITC claimants. These changes either have offsetting effects on EITC noncompliance (i.e., shifting from using of a credentialed preparer to using an unregulated preparer or self-preparing) or make little difference (i.e., shifting from using an unregulated preparer to self-preparing) due to different types of errors associated with different preparation methods.

These findings highlight several areas for future work which would help in understanding and assisting in tax compliance. First, uncredentialed tax preparers are an important component of the tax preparation economy, but one about which we have little information. The ability to further classify tax preparers in terms of skills or professional experience would allow a better understanding of their relationship with noncompliance. Second, there is a strong relationship between past and current EITC compliance. This suggests that identifying errors early in a taxpayer's history of EITC claims may aid in either providing taxpayer education or anticipating future noncompliance. Finally, although this paper explores the dynamics in how taxpayers switch between preparation types, it does not explain why they choose to do so. Understanding the large shifts we see away from regulated preparers and towards unregulated preparers and software-aided self-preparation may help in providing outreach to taxpayers.

References

Ackerman, Deena, Janet Holtzblatt, and Karen Masken (2009). "The Pattern of EITC Claims Over Time: A Panel Data Analysis." *IRS Research Bulletin* 197–226. Washington, DC: Internal Revenue Service.

Beer, Sebastian, Matthias Kasper, Erich Kirchler, and Brian Erard (2019). *Do Audits Deter or Provoke Future Tax Noncompliance? Evidence on Self-employed Taxpayers.* IMF Working Paper #19/223. Washington, DC: International Monetary Fund.

Book, Leslie (2007). "Study of the Role of Preparers in Relation to Taxpayer Compliance with Internal Revenue Laws." *National Taxpayer Advocate 2007 Annual Report to Congress* 2: 44–74. Washington, DC: Internal Revenue Service.

Chetty, Raj, John N. Friedman, Peter Ganong, Kara E. Leibel, Alan H. Plumley, and Emmanuel Saez (2012). *Taxpayer Response to the EITC: Evidence from IRS National Research Program.* Retrieved October 9, 2018, from http://www.rajchetty.com/chettyfiles/eitc_nrp_tabs.pdf.

DeBacker, Jason, Bradley Heim, Anh Tran, and Alex Yuskavage (2018a). "Once Bitten, Twice Shy? The Impact of IRS Audits on Filer Behavior." *Journal of Law and Economics* 61(1): 1–35.

DeBacker, Jason, Bradley Heim, Anh Tran, and Alex Yuskavage (2018b). "The Effects of IRS Audits on EITC Participants." *National Tax Journal* 71(3): 451–484.

DeBacker, Jason, Bradley Heim, Anh Tran, and Alex Yuskavage (2021). *Seeking Professional Help: How Paid Preparers Decrease Tax Compliance.* Working paper.

Dowd, Tim, and John B. Horowitz (2011). "Income Mobility and the Earned Income Tax Credit: Short-Term Safety Net or Long-Term Income Support." *Public Finance Review* 39(5): 619–652.

Erard, Brian, 1993. "Taxation with representation: An analysis of the role of tax practitioners in tax compliance." *Journal of Public Economics* 52 (2): 163–197.

Gemmell, Norman, and Marisa Ratto (2012). "Behavioral Responses to Taxpayer Audits: Evidence from Random Taxpayer Inquiries." *National Tax Journal* 65(1): 33–57.

Goldin, Jacob, Tatiana Homonoff, Rizwan Javaid, and Brenda Schafer (2021). *Tax Filing and Take-Up: Experimental Evidence on Tax Preparation Outreach and EITC Participation.* NBER Working Paper #28398. Cambridge, MA: National Bureau of Economic Research.

Government Accountability Office (GAO) (2006). *Paid Tax Return Preparers: In a Limited Study, Chain Preparers Made Serious Errors.* GAO-06–563T. Washington, DC.

Government Accountability Office (GAO) (2016). *Refundable Tax Credits: Comprehensive Compliance Strategy and Expanded Use of Data Could Strengthen IRS's Efforts to Address Noncompliance.* GAO-16–475. Washington, DC.

Guyton, John, Pat Langetieg, Day Manoli, Mark Payne, Brenda Schafer, and Michael Sebastiani (2017). "Reminders and Recidivism: Using Administrative Data to Characterize Nonfilers and Conduct EITC Outreach." *American Economic Review* 107(5): 471–475.

Guyton, John, Kara Leibel, Dayanand S. Manoli, Ankur Patel, Mark Payne, and Brenda Schafer (2018). *The Effects of EITC Correspondence Audits on Low-Income Earners.* NBER Working Paper #24465. National Bureau of Economic Research, Cambridge, MA.

Hotz, V. Joseph, and John Karl Scholz (2008). "Can Administrative Data on Child Support Be Used to Improve the EITC? Evidence from Wisconsin." *National Tax Journal* 61(2): 189–204.

Internal Revenue Service (IRS) (2014). *Taxpayer Compliance for the Earned Income Tax Credit Claimed on 2006–2008 Returns.* Internal Revenue Service Report 5162 (8–2014). Washington, DC: Internal Revenue Service.

Internal Revenue Service (IRS), various years. *Internal Revenue Service Data Book.* Washington, DC: Internal Revenue Service.

Klepper, Steven, Mark Mazur, and Daniel Nagin (1991). "Expert Intermediaries and Legal Compliance: The Case of Tax Preparers." *Journal of Law and Economics* 34(1): 205–229.

Klepper, Steven, and Daniel Nagin (1989). "The Role of Tax Preparers in Tax Compliance." *Policy Sciences* 22(2): 167–194.

Langetieg, Patrick, Mark Payne, and Melissa Vigil (2013). "Return Preparer Industry Analysis." *2013 IRS Research Bulletin, Tax Administration at the Centennial: An IRS-TPC Research Conference*, edited by Alan Plumley, 17–44. Washington, DC: Internal Revenue Service. Retrieved October 9, 2018, from https://www.irs.gov/pub/irs-soi/13rescon.pdf.

Leibel, Kara, Emily Y. Lin, and Janet McCubbin (2020). *Social Welfare Considerations of EITC Qualifying Child Noncompliance.* Office of Tax Analysis Working Paper #120. U.S. Department of the Treasury.

Lin, Emily Y. (2020). "Recent Changes in the Paid Return Preparer Industry and EITC Compliance." *IRS Research Bulletin, Internal Revenue Service Publication 1500 (Rev. 6-2020), pp. 107–127.* Internal Revenue Service Publication 1500 (Rev. 6-2020). Washington, DC: Internal Revenue Service.

Maag, Elaine, Michael Pergamit, Devlin Hanson, Caroline Ratcliffe, Sara Edelstein, and Sarah Minton (2015). *Using Supplemental Nutrition Assistance Program Data in Earned Income Tax Credit Administration: A Case Study of Florida SNAP Data Linked to IRS Tax Return Data.* Urban Institute Research Report. Washington, DC: Urban Institute.

Mortenson, Jacob A., and Andrew Whitten (2020). "Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule." *American Economic Journal: Economic Policy* 12(3): 402–432.

Pergamit, Michael R., Elaine Maag, Devlin Hanson, Caroline Ratcliffe, Sara Edelstein, and Sarah Minton, (2015). *Pilot Project to Assess Validation of EITC Eligibility with State Data.* Urban Institute Research Report. Washington, DC: Urban Institute.

Saez, Emmanuel (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2(3): 180–212.

Treasury Inspector General for Tax Administration (2008). *Most Tax Returns Prepared by a Limited Sample of Unenrolled Preparers Contained Significant Errors.* 2008-40–171. Washington, DC: Treasury Inspector General for Tax Administration.

U.S. Department of the Treasury, various years. *Agency Financial Report: Fiscal Year.* Washington, DC: U.S. Department of the Treasury.

Appendix

TABLE A1. EITC Error Rate and Amount by Year

Fiscal Year	EITC Improper Payments		
	Rate (%)	Amount ($B)	Data Year
2003	25 - 30	9.5 - 11.5	1999
2004	27 - 32	8 - 10	1999
2005	23 - 28	9.6 - 11.4	2001
2006	23 - 28	9.8 - 11.6	2001
2007	23 - 28	10.4 - 12.3	2001
2008	23 - 28	11.1 - 13.1	2001
2009	23 - 28	11.2 - 13.3	2001
2010	23.9 - 28.7	15.3 - 18.4	2006
2011	23.5	15.2	2007
2012	22.7	12.6	2008
2013	24.0	14.5	2009
2014	27.0	17.7	2010
2015	23.8	14.8	2011
2016	24.0	16.5	2012
2017	23.87	16.2	2013
2018	25.06	18.4	2014
2019	25.26	17.3	2015
2020	23.53	16.0	2016

SOURCE: Department of the Treasury Annual Financial Report.
NOTE: The lower bounds of the ranges for improper payment rates before 2011 were consistent in estimation methodology with the point estimates in later years.

2

∇

Impacts of Variations in Process

Brown ◆ Kenchington ◆ White

Reinhart ◆ Dacal ◆ Kaylor ◆ Wooten ◆ Curtiss

Dennis ◆ Ferris ◆ Gagin ◆ Ozer-Gurbuz ◆ Shiller ◆ Williams

Local Sales Tax Administration and the Real Economy

Jennifer L. Brown, David G. Kenchington, and Roger M. White (Arizona State University)[1]

Introduction

The effect of tax administration on the behavior of businesses is an important research topic in accounting, but the focus has primarily been on federal income taxes (El Ghoul, Guedhami, and Pittman (2011); Hoopes, Mescall, and Pittman (2012); De Simone, Sansing, and Seidman (2013); Bozanic, Hoopes, Thornock, and Williams (2017); Ayers, Seidman, and Towery (2019); Gallemore and Jacob (2020)). In contrast, we examine how the administration of state and local sales taxes affect businesses. Specifically, we examine whether local government administration of sales taxes (as opposed to state government administration) affects business decision making in the real economy. A long literature documents how sales tax rates affect the real economy. Higher sales taxes reduce retail employment (Thompson and Rohlin (2012); Wong (1996)), consumer spending (Baker, Johnson, and Kueng (2020)), firms' capital investment (Hageman, Bobek, and Luna (2015); Jacob, Michaely, and Müller (2019)), commodity flows among geographically close states (Fox, Luna, and Schaur (2014)), and local home prices (Shon and Chung (2018)). This literature focuses almost solely on the level of sales taxes, with little attention paid to how sales taxes are administered. We explore this latter issue and, controlling for sales tax levels, examine whether sales tax administration policy affects the real economy.

In most states, the state department of revenue administers local sales taxes, typically in exchange for a small percentage of the revenue collected. Some states, however, either do not offer this service to local governments or provide local governments a choice between state administration or local administration of local sales taxes. For example, Georgia and Alabama, like most states, impose state-level sales taxes that are administered by the state department of revenue. Also like most states, Georgia and Alabama permit municipalities and counties to impose their own local-level sales tax. Important differences emerge in the sales tax administration policies of Georgia and Alabama, however.

In Georgia, the state department of revenue administers local sales taxes, so a hypothetical business operating in 15 of the 162 different local sales tax jurisdictions in Georgia need only file a single sales tax return (Georgia Department of Revenue (2020)). The state revenue department then remits local sales taxes to the appropriate local government. By comparison, Alabama allows local governments to administer their own local sales taxes (e.g., the 3 percent city-level sales tax rate in Tuscaloosa, Alabama, is paid to and administered by the Tuscaloosa municipal government). Accordingly, a business that operates in 15 of the 327 different local sales tax jurisdictions in Alabama may need to file 16 different sales tax returns (one for the state and one for each local sales tax jurisdiction) (Institute for Professionals in Taxation (2016)).

Beyond the increase in returns that have to be filed, businesses may face other costs when local governments administer sales taxes. For example, information regarding which local jurisdictions self-administer their taxes, the tax rates, or even whom to contact for such information is frequently not available from a central source. Further, businesses are required to register, often for a fee, with each taxing jurisdiction. The registration may include details about where the taxpayer conducts operations within the jurisdiction and information about the officers of the business—all of which must be kept current. In addition, there may be differences between the state sales tax base and the bases for locally administered taxes (as well as differences in tax bases among local governments within a state), which practitioners suggest may be the most difficult compliance

[1] For helpful feedback that much improved this paper, we thank Dane Christensen, Josh Ogle, Jennifer Stratton, Maximilian Todtenhaupt, Ryan Wilson, conference participants at the IRS/TPC Joint Research Conference on Tax Administration and 11th EIASM Conference on Current Research in Taxation, and workshop participants at the University of Oregon. We are also grateful to Mason Snow for very capable research assistance.

issue presented by locally administered taxes (Institute for Professionals in Taxation (2016)). Also, the administrative and procedural requirements of locally administered taxes often vary from those for the state and across local jurisdictions within a given state. Such requirements include tax registration, return filing dates, protest periods, penalty provisions, and exemption certificate requirements—each of which complicates the task of tax compliance. Finally, local administration also means that any enforcement or collection procedures, such as audits, notices, and appeals, are commonly dealt with on a jurisdiction-by-jurisdiction basis.

Clearly, decentralized local administration imposes higher compliance costs on businesses (i.e., taxpayers responsible for paying sales taxes). For example, in a recent report KPMG examined four states with locally administered sales taxes (Alabama, Arizona, Colorado, and Louisiana) and estimated that businesses in these states spent $190 million per annum on *incremental* sales tax compliance costs due to decentralized local sales tax administration (Institute for Professionals in Taxation (2016)). Incremental labor cost comprises nearly two-thirds of the $190 million (e.g., additional hours spent complying with locally administered sales taxes), with the remaining one-third coming from non-labor activities such as the cost of software acquisition, contract software development and maintenance, etc.

To date, the only research related to these costs and any potential accompanying market frictions consists of discussion pieces describing different policies of tax administration (e.g., local vs. centralized) (Due and Mikesell (1983); Mikesell (2007)), a KPMG whitepaper commissioned by an industry group highlighting the onerous compliance costs of locally administered taxes (Institute for Professionals in Taxation (2016)), and state rankings of tax complexity that also stress the compliance hassle of locally administered taxes (Afonso (2019)). Despite 25 states allowing or requiring local governments to administer their own consumption taxes, no existing study provides an empirical test of whether decentralized tax administration imposes incremental compliance costs that are meaningful enough to have measurable effects on the real economy.

We provide such an analysis and examine whether the incremental compliance costs of local tax administration affect the real economy. We test this question using Florida's Tourist Development Tax, which is a sales tax on short-term property rentals (e.g., hotels rooms, private homes), levied by individual Florida counties. The state imposes and administers a 6 percent tourist sales tax on rentals, and individual counties can also impose similar taxes via the Tourist Development Tax (hereafter, "tourism tax" or "hotel tax"). Counties have the option of having the state administer this county-level tourism tax (along with the state tax), or the county can administer the county tourism tax locally.

Counties in Florida that choose to locally administer the tourism tax likely do so for a variety of reasons (Florida Tourist Development Tax Association (2001a and 2001b)). For instance, counties receive revenues from the tax more quickly when it is locally administered. Further, local administration allows counties, rather than the state, to capture the jobs related to the collection, enforcement, and auditing of the tax. In addition, local governments may decide to administer the tax because it allows them more control of enforcement. For example, they can choose whom to audit. Relatedly, some taxes are locally administered because of the belief that state-level revenue officers do not closely monitor for evasion by privately owned rental properties, where tourist taxes apply for stays of less than six months (e.g., beach houses or condos rented through magazines, websites, Airbnb, VRBO, etc.) (Swoboda (2017); Moffa (2019); Callihan (2019)).

Over the course of our 30-year sample period, we identify 33 counties in Florida that switch from state to local administration of local tourism taxes, and a few counties switching *back* to state administration after years of county-level tourism tax administration (usually after determining that the cost of administering the local tourism tax does not offset the incremental tax revenue). We use these switches to examine the real economic effects of sales tax administration on the hotel industry in Florida. Important for identification purposes, this county-level choice to administer the county-level tourism tax locally or delegate the administration to the state is not endogenously determined by hotel characteristics (such as a lack of compliance with the tourism tax by hotels).[2] Rather, as mentioned, these policy decisions stem mostly from a desire to increase the number of employees at the county government level, speed up receipt of the tourism tax, increase control

[2] We are unaware of any evidence that a lack of compliance with the tourism tax by hotels influences the county-level switching decision. It would be difficult for hotels to avoid the tourism tax because they are hard to conceal, given that hotels are typically registered with and regulated by various government bodies that oversee issues such as food and fire safety, building codes, zoning, property taxes, corporate income taxes, etc.

over the audit process, and curtail tourism tax evasion by operators of private rental properties. Consequently, any negative impact of local tourism tax administration on the hotel industry can be attributed to a (plausibly) exogenous, unintended consequence of locally administered taxes increasing tax compliance costs.[3]

We test for such unintended consequences by predicting county-year hotel industry size (number of hotel employees, total hotel payroll, number of hotel establishments) as a function of local tourism tax administration policy, as well as a host of county-year level proxies for the health of other industries (to capture local economic conditions) (e.g., Wong (1996)). If the incremental compliance costs of locally administered taxes hamper the real economy, then we would expect to see local tax administration negatively predict hotel industry size in our models. Such a relation would suggest that the hotel industry is smaller than otherwise expected in county-years in which local tourism taxes are locally administered, relative to county-years in which local tourism taxes are administered by the state in conjunction with the state tourism tax (in which the local portion of the tax is later remitted to the county). This pattern in the data would emerge if higher tourism tax compliance costs lead existing hotels to cut costs or raise prices, both of which could result in fewer employees and lower payroll. Likewise, investors could be less likely to build or continue operating hotels in affected counties due to these higher compliance costs.

We find evidence of such a relation in our data. Controlling for local tourism tax rates, we find that local administration of local tourism taxes negatively predicts the county-year total hotel payroll and number of hotel employees. Given that our models include county and year fixed effects, county-year measures of non-hotel industry development (to proxy for general county-year economic conditions), and controls for adjacent counties' hotel industry characteristics, most alternative explanations can be safely ruled out. Instead, our results indicate that the compliance costs of locally administered sales taxes depress total wages, employment, and investment in affected industries.

Beyond our direct contribution to the overall discussion of the real economy costs of sales tax decentralization, we also add to literature in accounting examining the effect of tax administration on the behavior of businesses. The primary focus of this literature has been on the administration of federal income taxes. For example, when enforcement by the IRS is higher, businesses undertake less aggressive income tax strategies (Hoopes, Mescall, and Pittman (2012)). Further, IRS monitoring seems to reduce expropriation of corporate resources by insiders, as evidenced by a lower cost of debt and equity financing when a firm's probability of a face-to-face IRS audit is higher (El Ghoul, Guedhami, and Pittman (2011); Guedhami and Pittman (2008)). Varying levels of tax administrator enforcement also affect firms' public disclosure behavior (Bozanic, Hoopes, Thornock, and Williams (2017)) and financial reporting quality (Hanlon, Hoopes, and Shroff (2014)). We add to this literature by examining how businesses respond to tax administration choices about the sales tax. Another line of research in accounting focuses on how tax administration affects revenue collection (Kubick, Lockhart, Mills, and Robinson (2017); Nessa, Schwab, Stomberg, and Towery (2020); Tang, Mo, and Chan (2017)). We further this research by examining how tax administration affects the real economy.

Finally, our results also speak to the broader literature on tax complexity (of which compliance costs are a significant component). Most literature on tax complexity focuses on how taxpayers are less likely to comply with complex taxes, due to either mistakes or resentment (see Carnes and Cuccia (1996); Cox and Eger (2006); Cuccia and Carnes (2001)). However, other work suggests that complex tax systems create deadweight compliance costs that slow economic development (Laffer *et al.* (2011)). This conjecture is intuitive, but testing it is difficult, as it requires a setting in which tax complexity can be measured distinctly *and* isolated from the effects of the tax rate. We are unaware of any other paper using observational data that has been able to identify and test such a case. Thus, we view our study as contributing to this intuitive but hard-to-examine hypothesis in public economics regarding the optimal design of tax systems. We note that laboratory experiments (Boylan and Frischmann (2006)) and survey research (Matarirano, Chiloane-Tsoka, and Makina (2019)) do provide evidence supporting this conjecture, that high costs of complying with taxes can distort real economic activity. However, the nuances of econometrically identifying and testing this prediction in observational settings have until now precluded corroborative findings in real-world data.

[3] Conversely, any positive impact could be attributed to local hotel tax administration increasing costs for hotels' competitors (private vacation rentals) and subsequently improving hotels' relative competitiveness in lodging markets.

In the following section, we provide background and institutional details. We then describe our data and predictions, present results, and briefly conclude.

Setting and Background

In the late 1970s, when tourism in Florida began to quickly grow (Harrington, Chi, and Gray (2017)), the Florida legislature passed the "Local Option Tourist Development Act," which allowed counties to impose local taxes on transient rentals such as hotel stays and vacation rentals, provided that the proceeds went to boost Florida tourism (advertise local tourist destinations, build and maintain public convention centers, fund tourist information centers, etc.) (Wenner (2020)). Since that time, legislation has expanded the permissible use of these tax proceeds to include funding beach maintenance, public safety, docks, boat ramps, and erosion control.

As originally stipulated by the legislature, local option tourist development taxes in Florida were administered and collected by the state-level Florida Department of Revenue. This was largely a matter of convenience, as the Florida Department of Revenue was *already* collecting and administering a state-level tourist tax on hotel stays and other transient rentals. Given that this bureaucratic framework was already in place, counties introducing local-level tourist taxes were able to simply graft the collection and administration of these taxes onto existing state-level processes. For taxpayers, mostly businesses such as hotels and vacation property management companies, complying with these new local tourist taxes was straightforward and involved the same forms, deadlines, and filing procedures as the existing state-level tourist taxes. After the Florida Department of Revenue collected tourist taxes via this process, county-level taxes were remitted to the individual counties.

In the late 1980s, the legislature amended the "Local Option Tourist Development Act" to allow counties to self-administer the county-level tourist development tax. For counties choosing to take advantage of this option, state-level tourist taxes would continue to be collected and administered by the state, but the county-level portion of the tax would be collected and administered by county officials using whatever forms, deadlines, procedures, and filing criteria the county selected (Florida Tourist Development Tax Association (2001a and 2001b)).

For county-level officials, taking over "home rule" of the county-level tourist tax was and is appealing for a number of reasons. First, under state-level administration, counties typically receive their county-level taxes after a two-month delay (processing time, state-level audits, bank transfer time, etc.). Local administration eliminates this waiting period and makes funds available to local officials more quickly. Second, under county-level administration, counties are permitted to keep 3 percent of county-level tourist taxes to offset administrative costs. Accordingly, county-level administration permits county officials an opportunity to increase their local workforce and scope of control. Third, state-level administration removes enforcement power from the hands of county-level officials, and state-level enforcement efforts focus primarily on ensuring compliance by large taxpayers (i.e., hotels, campgrounds, and other operating businesses with an existing regulatory profile). Smaller taxpayers, most notably individuals renting lone vacation homes for short-term stays, frequently fly under the radar and can evade state-administered tourist taxes. County officials are both more informed about these smaller operators and have stronger incentives to enforce tax compliance. Accordingly, county-level administration can lead to higher county-level tourist tax receipts. Overall, 43 of Florida's 67 counties have at some point elected to self-administer county-level tourist taxes, and more closely enforcing the tax on vacation rental homes, such as those rented through Airbnb, is frequently given as a primary motivation for this decision (Moffa (2019); Swoboda (2017)).[4]

While county officials may see benefits to this decision to self-administer the local (county-level) tourist tax, the change does burden local taxpayers with increased compliance costs. For example, consider a small hotel chain that operates in both Miami-Dade and Broward counties, neighboring Atlantic coast beach destinations. In 1985, both counties had county-level tourist taxes in place, but since the state Department of

[4] In addition to benefits, local governments may incur various costs when electing to self-administer county-level tourist taxes. For instance, data processing systems may need to be updated to ensure the privacy of taxpayer information. Local governments also need to decide who will be responsible for the collection, enforcement, and audit of the tax. Along with this, decisions need to be made about what will be reported and how public and confidential requests for information will be handled.

Revenue (DOR) handled administration, the example hotel chain needed only to file one monthly tourist tax form and remit a single payment for tourist taxes, after which the state DOR would remit respective taxes to officials in Broward and Miami-Dade counties. By 1995, however, both Miami-Dade and Broward had taken over county-level hotel sales tax administration, so the example hotel chain would, every month, need to file and remit tourist taxes to the state DOR, Broward County, and Miami-Dade County. In essence, the decision of Miami-Dade and Broward to self-administer county-level tourist taxes *tripled* some aspects of the compliance process for the example hotel chain. Since tourist taxes are paid and remitted monthly, instead of 12 tax payments and reporting processes (as in 1985 under state-level administration), the 1995 tourist tax compliance process for the example hotel chain involves 36 different payment and filing processes (every month, a filing and payment must be made for the state DOR, Broward County, and Miami-Dade County).

This added complexity in the tax compliance process could impose real costs for hotel firms, including dedicating more of managers' time to the compliance process as well as paying higher fees to the law and accounting firms that must prepare and review a higher volume of tax filings. In addition, local governments' tourist tax audit procedures vary, which requires incremental time and expense for affected taxpayers (relative to state-level audits with consistent procedures and expectations). Our focus in this paper is on whether these higher compliance costs induced by local tax administration have an appreciable effect on the real economy. If these higher compliance costs meaningfully cut into profit margins or prospective profit margins, then existing hotel firms could respond to increased compliance cost pressure by trying to cut other costs or by raising prices, both of which could result in fewer employees and lower payroll. Likewise, investors could be less likely to build and/or operate hotels in affected counties.

Data and Empirics

We test these predictions using data from the Quarterly Census of Employment and Wages series (QCEW), published by the U.S. Bureau of Labor Statistics (BLS). The QCEW data is collected via quarterly surveys of U.S. businesses, and the series includes counts of establishments (places of business), total wages, and employees (full-time equivalent). Approximately 95 percent of all jobs are captured by the QCEW, and importantly for our purposes, the BLS reports QCEW data by year, county, and industry. We define the county-year number of hotel establishments, hotel employees, and total hotel industry wages as the QCEW count of establishments, employees, and total wages for hotels (NAICS Subsector Code = 721) by county-year. As the county-year QCEW data coverage begins in 1990, we begin our sample with that year (and end in 2019, for a 30-year sample).

Our primary tests examine whether hotel industry size is predicted by the move from state administration to local administration of county-level tourist taxes. If this move increases tourist tax compliance costs to the point of affecting the real economy, then we would expect to see fewer employees, lower payrolls, and fewer hotels *after* counties make the move to self-administering county-level tourist taxes.

To identify this effect in our regression models, we use as our treatment variable an indicator labeled *Locally Administered Local Hotel Tax (Dummy)*. This indicator loads with a value of one in county-years in which the local hotel tax is locally administered. To isolate this treatment effect from any systematic differences between county-years with and without a local hotel tax in place, we include another indicator variable titled *Any Local Hotel Tax (Dummy)*. We also control for the local hotel (tourist) tax rate, so that our models capture the *incremental* effect of higher tax compliance costs and are not polluted by tax rate effects (Baker *et al.* (2020); Hageman *et al.* (2015); Jacob *et al.* (2019); Thompson and Rohlin (2012); Wong (1996)).

Our models take the following form, where subscript *c* indexes county and subscript *t* indexes year:

$$\textit{Hotel Industry Size}_{c,t} = \alpha + \beta_1 \text{ x } \textit{Locally Administered Local Hotel Tax (Dummy)}_{c,t} + \beta_2 \text{ x } \textit{Any Local Hotel Tax (Dummy)}_{c,t} + \beta_3 \text{ x } \textit{Local Hotel Tax Rate (\%)}_{c,t} + \Sigma\ \textit{Controls}_{c,t}$$

A significant, negative coefficient on β_1 would suggest, after adjusting for the level of the county-year hotel tax rate (with the main effect β_3) and any differences between county-years with versus without local hotel taxes (captured by the main effect β_2 on the *Any Local Hotel Tax* indicator variable), that a shift to local sales tax administration corresponds to slower growth in affected industries (i.e., lower payroll, fewer employees, or fewer establishments).

To better isolate a causal effect, we include county and year fixed effects, county-year measures of non-hotel industry development (to proxy for general county-year economic conditions), *and* hotel industry size measures from adjacent (border) counties. Accordingly, our β_1 treatment effect is identified by time series variation in the local administration of tourist taxes *within* each county. In our sample of Florida's 67 counties from 1990 to 2019, we observe 33 counties moving from statewide administration of tourist taxes to local administration of tourist taxes. We also observe four counties switching *back* from local administration to state administration of tourist taxes, after realizing that the incremental revenue from stronger tourist tax enforcement for (mostly) non-commercial vacation rentals does not offset the costs of collecting and administering the tax locally.

Beyond county and year fixed effects, to control for county-year economic conditions we include measures of non-hotel industry size from the QCEW. A shift to local administration of hotel taxes could be partially motivated by broad county-year economic booms or recessions, and to avoid correlated omitted variable problems of this nature we control for the size of two large industries with a presence in most Florida counties, finance (NAICS sector code = 52) and manufacturing (NAICS sector codes = 31, 32, 33). We examine three different measures of hotel industry size in our models (total wages, number of employees, and number of establishments), and we use the matching metric for these control industries in our models (e.g., we control for total county-level manufacturing wages in models estimating the total hotel wages in the county).

To control for regional hotel industry trends, we also control for the size of the hotel industry in adjacent counties. We identify border counties using the NBER County Adjacency file.[5] Combining this county adjacency data with the QCEW industry size data, we construct a panel of hotel industry size measures for each focal county's border counties for each year in our sample. Including these controls in our regressions allows us to rule out correlated omitted variable issues that could motivate a move to local administration of hotel taxes *and* affect the regional hotel industry (the opening of a major tourist attraction, a hurricane hitting a certain area of the Florida coast, etc.). For each model specification we include the relevant border-county control (e.g., in models estimating hotel industry employment in a county-year, we include the total combined hotel industry employment for border counties in the county-year).

Finally, we also include county-level sales tax and population figures as control variables, to adjust for county-level growth or local government revenue demands not captured by other controls. We draw population figures from the U.S. Census Bureau, and we obtain county-year local sales tax rates from the Florida Department of Revenue.[6] Finally, we cluster standard errors by county to adjust for potentially correlated error terms between same-county observations from different years (Cameron and Miller (2015)).

Results

Univariate Statistics

Before estimating regressions, we first present in the Appendix excerpts from a Florida Department of Revenue publication titled "History of Local Sales Tax and Current Rates." This publication reports the dates that county-level hotel taxes (officially titled as "Local Option Tourist Development Tax") are implemented and raised (there are no decreases in our sample period), and, most importantly for our purposes, the date when a county took over local administration of the county-level hotel tax.[7]

In Exhibit A of the Appendix, we excerpt the record for Hamilton County. Hamilton County introduced the county-level hotel tax in 1996 at a 2 percent rate and increased the rate to 3 percent in 2002. Hamilton County is in the minority of Florida counties in that the state department of revenue has administered the local

[5] https://www.nber.org/research/data/county-adjacency.

[6] https://floridarevenue.com/Forms_library/2019/dr15DSS_%20r11_19.pdf.

[7] https://floridarevenue.com/taxes/Documents/flHistorySalesTaxRates.pdf.

hotel tax on behalf of Hamilton County since its inception. As a result, hotels in Hamilton County need only file a single monthly sales tax return (with the state). The state then remits the Hamilton County tax proceeds to the county government.

Exhibit B of the Appendix reports the record for Marion County. Marion County introduced a 2 percent county-level tourist tax in 2005 and increased the tax to 4 percent in late 2015. In 2008, however, Marion County took over administering this county-level tourist tax. After this policy shift, Marion County hotels need to file two sales tax returns monthly, one with the state (for the state-level tourist tax) and one with Marion County officials (to pay the county-level tourist tax). In addition, following the switch to local administration, hotels in Marion County could be subject to enforcement and collection procedures by both the state and local taxing authorities.

Exhibit C of the Appendix reports a similar Table for Highlands County, where a county-level hotel tax was introduced in 2003 and increased in 2018. Local administration of this county-level hotel tax began in 2014 but ended in 2018, when Highlands County determined that the costs of local administration outweighed any incremental revenue from local administration. As mentioned previously, we observe only four such changes *back* from local administration to state administration of county-level hotel taxes. Accordingly, in our following discussion we phrase the interpretation of treatment effects in terms of moving from state administration to local administration (which is about ten time more common).

We use these Florida Department of Revenue records to build our treatment indicator, *Locally Administered Local Hotel Tax (Dummy)*. In Table 1, we condense these county-level records to report the initial date a county-level tourist tax was enacted in each county and the year (if applicable) that county-level administration began.

Next, we report the summary statistics for our independent variable of interest, our dependent variables, and our control variables. Many of these important measures come from the QCEW. While the QCEW data provides granular, county-industry-year-level data, for privacy and competitive purposes this data is limited in small counties. In county-years in which only one hotel operates, for example, county-year-level data for the hotel industry could disadvantage the single firm for which data is reported. Therefore, for some industry-years in smaller counties, the QCEW does not report data. Our original sample size is 2,010 county-years (30 years x 67 counties), but this QCEW data restriction leaves us with a usable sample of 1,556 county-years. Of these, 1,392 (89 percent) county-years have a county-level hotel tax in place, and for 1,003 (72 percent) of those the local hotel tax is locally administered. We report this sample breakdown in Table 2.

Table 3 reports summary statistics for the 1,556 county-years for which we have usable data. The median county-year involves a 3 percent local hotel tax, a population of 159,000 residents, and 837 hotel employees working in 37 hotels with a total payroll of about $13.6 million. These variables are skewed and have long right tails, so we use the natural log of these figures as dependent variables in our models (to ensure normally distributed error terms).[8]

[8] That is, the models we report in Tables 4–6 use as dependent variables the natural log of one plus our hotel industry size proxies. However, in untabulated results we confirm all of these findings using the raw, untransformed data. Our results persist at consistent or stronger levels when using this untransformed data.

TABLE 1. County-level Timeline of Local Hotel Sales Taxes and Beginning of Local Administration

County	County-level Tourist Tax Enacted	County-level Administration of Tourist Tax Begins	County	County-level Tourist Tax Enacted	County-level Administration of Tourist Tax Begins
Alachua	1987	2001	Lee	1982	1988
Baker	2000	2000	Leon	1988	1994
Bay	1986	1994	Levy	2003	n/a
Bradford	1990	n/a	Liberty	n/a	n/a
Brevard	1986	1992	Madison	1999	n/a
Broward	1980	1994	Manatee	1981	1989
Calhoun	n/a	n/a	Marion	2005	2008
Charlotte	1984	1990	Martin	2002	2002
Citrus	1986	1991	Miami-Dade	1978	1988
Clay	1989	1989	Monroe	1981	1991
Collier	1990	1993	Nassau	1989	1989
Columbia	1984	n/a	Okaloosa	1989	1992
DeSoto	2011	n/a	Okeechobee	1993	n/a
Dixie	2011	n/a	Orange	1978	1992
Duval	1979	1990	Osceola	1977	1992
Escambia	1980	1989	Palm Beach	1982	1993
Flagler	1986	2018	Pasco	1991	2019
Franklin	2005	n/a	Pinellas	1978	1990
Gadsden	2003	n/a	Polk	1986	1994
Gilchrist	2007	n/a	Putnam	1993	1999
Glades	2009	n/a	St. Johns	1986	1988
Gulf	1999	2001	St. Lucie	1984	1991
Hamilton	1996	n/a	Santa Rosa	1992	1994
Hardee	2017	n/a	Sarasota	1988	1992
Hendry	2003	n/a	Seminole	1989	1993
Hernando	1993	1993	Sumter	2005	n/a
Highlands	2003	2014	Suwannee	1991	2001
Hillsborough	1978	1992	Taylor	1998	2006
Holmes	2005	n/a	Union	n/a	n/a
Indian River	1987	2000	Volusia	1978	1990
Jackson	1999	n/a	Wakulla	1995	1996
Jefferson	2007	n/a	Walton	1986	1991
Lafayette	1991	n/a	Washington	2001	n/a
Lake	1984	1998			

TABLE 2. Sample Selection Steps

Sample Set	n
Total County-years in Sample Period (30 years x 67 counties)	2,010
County-years with Usable Data in BLS QCEW	1,556
Sample County-years with Local Hotel Tax	1,392
Sample County-years with Locally Administered Local Hotel Tax	1,003

TABLE 3. Summary Statistics

Variable	Mean	Std. Dev.	Min	25th %	50th %	75th %	Max
Locally administered local hotel tax (dummy)	0.63	0.51	0.00	0.00	1.00	1.00	2.50
Any local hotel tax (dummy)	0.80	0.40	0.00	1.00	1.00	1.00	1.00
Local hotel tax rate (%)	3.06	1.52	0.00	2.00	3.00	4.00	6.00
Population (thousands)	325.46	450.67	7.06	56.93	158.94	357.25	2,812.13
No. of hotel employees	2,970.59	6,710.96	0.00	161.00	837.00	2,435.00	57,933.00
No. of adjacent county hotel employees	15,784.77	18,750.21	68.00	2,462.00	7,237.00	19,201.00	76,792.00
No. of manufacturing employees	7,426.27	11,848.58	0.00	852.00	2,279.00	7,790.00	82,431.00
No. of finance employees	8,759.08	15,910.31	0.00	564.00	2,371.00	6,848.00	76,339.00
No. of hotel establishments	76.31	94.81	0.00	15.00	37.00	96.00	592.00
No. of adjacent county hotel establishments	386.78	311.18	10.00	133.00	282.00	585.00	1,428.00
No. of manufacturing establishments	325.31	540.00	0.00	38.00	119.00	395.00	3,637.00
No. of finance establishments	1,021.93	1,686.01	0.00	96.00	404.00	965.00	10,986.00
Total hotel payroll ($K)	68,627.99	179,174.11	0.00	2,399.45	13,623.29	50,437.72	1,884,999.10
Adjacent county total hotel payroll ($K)	364,323.90	499,709.39	713.99	42,560.13	135,791.96	464,370.68	2,425,220.68
Total manufacturing payroll ($K)	311,936.03	497,945.20	0.00	26,402.47	83,307.35	293,273.44	2,594,915.13
Total finance payroll ($K)	446,617.38	961,853.21	0.00	15,511.05	81,403.66	269,634.44	7,108,561.46

Number of observations for all variables: 1,556.

Modeling Hotel Industry Size

We first estimate regressions of hotel industry size by county-year in Table 4, where we use *Hotel Payroll* (at the county-year level) as our dependent variable. Column 1 reports a baseline specification omitting our treatment effect indicator (*Locally Administered Local Hotel Tax (Dummy)*) and the indicator for county-years with any state-administered or locally administered local hotel tax in effect (*Any Local Hotel Tax (Dummy)*), but controlling for the county-level tourist tax rate, the county-level general sales tax rate, population, adjacent county hotel payroll, same county manufacturing industry payroll, and same county finance industry payroll (banks, insurance agencies, etc.). The only statistically significant predictor in this model is the hotel payroll of adjacent counties, where the coefficient suggests that a 1 percent increase in the combined hotel payrolls in adjacent counties predicts to a 0.5 percent increase in focal county hotel payroll. The local sales tax rate and local hotel sales tax rate load with negative coefficients, but not at statistically significant levels. These controls, and the county and year fixed effects, predict almost all of the variation in county-year hotel payroll (R^2 = 94 percent).

TABLE 4. Models Estimating *Total Hotel Payroll*

$$Ln(Hotel\ payroll_{c,t}) = \alpha + \beta_1 \times Locally\ administered\ local\ hotel\ tax_{c,t} + \beta_2 \times Any\ local\ hotel\ tax_{c,t} + \beta_3 \times Local\ hotel\ tax\ rate_{c,t} + \Sigma\ Controls_{c,t}$$

	1	2
Locally administered local hotel tax (dummy)$_{c,t}$		-0.1401* (-1.81)
Any local hotel tax (dummy)$_{c,t}$		0.2673 (1.24)
Local hotel tax rate (%)$_{c,t}$	-0.0249 (-0.64)	-0.0546 (-1.25)
Local sales tax rate (%)$_{c,t}$	-0.1142 (-1.23)	-0.1099 (-1.26)
Ln(Population$_{c,t}$)	0.1977 (0.79)	0.1644 (0.65)
Ln(Adjacent county hotel payroll$_{c,t}$)	0.4969** (2.17)	0.5452** (2.37)
Ln(Manufacturing payroll$_{c,t}$)	0.1105 0.90)	0.0878 (0.66)
Ln(Finance payroll$_{c,t}$)	-0.1052 (-0.57)	-0.0996 (-0.56)
County and Year fixed effects	Yes	Yes
Observations	1,556	1,556
R^2	94.13%	94.18%

OLS models predicting *Total Hotel Payroll* per county-year (c, t) for Florida's 67 counties over the 1990-2019 window (for county-years with valid QCEW data). Fixed effects for county and year are included. Standard errors are clustered at the county level.
* and ** correspond to two-tailed statistical significance at the p < 0.10 and p < 0.05 levels, respectively.

Column 2 reports a fully specified model that includes our treatment effect indicator, Locally Administered Local Hotel Tax (Dummy), as well as Any Local Hotel Tax (Dummy) (to capture differences between county-years with and without a local hotel tax that are unrelated to local administration).9 Our treatment effect identifies whether the local administration of local hotel taxes predicts a smaller hotel industry incremental to the level of local hotel taxes (already captured in the main effect of Local Hotel Tax Rate (%)). Given that we estimate this model using a within-county framework, the -0.14 coefficient (t-statistic = -1.81) on Locally Administered Local Hotel Tax (Dummy) suggests that hotel payrolls are lower by about 14 percent after a county shifts the administration of local hotel taxes from the state level to the local level.

We see few systematic decreases in the hotel payroll in Florida counties over the course of our sample period (1990–2019), as over these years tourism in Florida grows significantly (Harrington *et al.* 2017). Rather, our models suggest that counties that shift from state-administered to locally administered hotel sales taxes see slower growth and development of the hotel industry, at least as measured by industry payroll. Note that since our models control for hotel sales tax rates, we are not concerned about tax *levels* influencing our result. Rather, our finding indicates that more complicated sales tax administration policies can boost sales tax compliance costs to the point that it measurably discourages the growth and development of affected industries.

[9] We include this measure in all of our models for completeness, but we note that its omission does not affect the magnitude or statistical significance of our treatment effects.

In Table 5, we examine whether the local administration of hotel taxes also negatively corresponds to the size of the hotel industry workforce (as opposed to payroll). Again, we estimate two iterative regressions, with a baseline specification in Column 1 and Column 2 adding *Any Local Hotel Tax (*Dummy) and our treatment effect, Locally Administered Local Hotel Tax (Dummy).

As with Table 4, the fixed effects models we estimate explain almost all the variation in the count of hotel industry employees by county-year, as R2 measures exceed 96 percent in both regressions. Also similar to Table 4, our treatment effect Locally Administered Local Hotel Tax (Dummy) loads with a negative and statistically significant coefficient (-0.11) in Column 2. This negative, significant coefficient on our treatment effect suggests that a move from state-administered to locally administered county-level hotel taxes predicts an 11 percent drop in the aggregate number of workers in local hotels.

This indicates that the increased compliance costs of locally administered hotel sales taxes contribute measurably to the hotel industry employing fewer workers than expected in a given county after the county moves from state-level administration to county-level administration of county-level hotel sales taxes.

We further test this result in Table 6, where we model the total number of hotel establishments for each county-year. The number of establishments is somewhat more static than employment and payroll figures, so we are less confident in a result in this setting. The Column 2 treatment effect, *Locally Administered Local Hotel Tax (Dummy)*, is negative but not statistically significant (t-statistic = -0.43).

TABLE 5. Models Estimating *the Number Hotel Employees*

$$\text{Ln}(1 + \#\ \text{Hotel employees}_{c,t}) = \alpha + \beta_1 \text{ x Locally administered local hotel tax}_{c,t} + \beta_2 \text{ x Any local hotel tax}_{c,t} + \beta_3 \text{ x Local hotel tax rate}_{c,t} + \Sigma\ \text{Controls}_{c,t}$$

	1	2
Locally administered local hotel tax (dummy)$_{c,t}$		-0.1088*
		(-1.76)
Any local hotel tax (dummy)$_{c,t}$		0.186
		(1.51)
Local hotel tax rate (%)$_{c,t}$	-0.0387	-0.0575*
	(-1.27)	(-1.79)
Local sales tax rate (%)$_{c,t}$	-0.1005	-0.0977
	(-1.29)	(-1.27)
Ln(Population$_{c,t}$)	0.2845	0.2499
	(1.66)	(1.43)
Ln(1 + # Adjacent county hotel employees$_{c,t}$)	0.3479	0.3857*
	(1.58)	(1.80)
Ln(1 + # Manufacturing employees$_{c,t}$)	-0.0284	-0.0616
	(-0.25)	(-0.50)
Ln(1 + # Finance employees$_{c,t}$)	-0.1823	-0.1373
	(-1.27)	(-1.01)
County and Year fixed effects	Yes	Yes
Observations	1,556	1,556
R_2	96.19%	96.23%

OLS models predicting # *Number Hotel Employees* per county-year for Florida's 67 counties over the 1990–2019 window (for county-years with valid QCEW data). Fixed effects for county and year are included. Standard errors are clustered at the county level.

* corresponds to two-tailed statistical significance at the p < 0.10 level.

TABLE 6. Models Estimating # *Number Hotel Establishments*

$Ln(1 + \# \text{ Number hotel establishments}_{c,t}) = \alpha + \beta_1 \times \text{Locally administered local hotel tax}_{c,t} + \beta_2 \times \text{Any local hotel tax}_{c,t} + \beta_3 \times \text{Local hotel tax rate}_{c,t} + \Sigma \text{ Controls}_{c,t}$		
	1	2
Locally administered local hotel tax (dummy)$_{c,t}$		-0.0168 (-0.43)
Any local hotel tax (dummy)$_{c,t}$		0.0522 (0.60)
Local hotel tax rate (%)$_{c,t}$	-0.0116 (-0.66)	-0.0185 (-0.99)
Local sales tax rate (%)$_{c,t}$	0.052 (1.32)	0.0538 (1.39)
Ln(Population$_{c,t}$)	0.1759 (0.83)	0.168 (0.79)
Ln(1 + # Adjacent county hotel establishments$_{c,t}$)	0.5515*** (2.89)	0.5579*** (2.80)
Ln(1 + # Manufacturing establishments$_{c,t}$)	0.035 (0.34)	0.0335 (0.33)
Ln(1 + # Finance establishments$_{c,t}$)	0.0365 (0.22)	0.0393 (0.23)
County and Year fixed effects	Yes	Yes
Observations	1,556	1,556
R^2	97.36%	97.37%

OLS models predicting # Number Hotel Establishments per county-year for Florida's 67 counties over the 1990–2019 window (for county-years with valid QCEW data). Fixed effects for county and year are included. Standard errors are clustered at the county level.
*** corresponds to two-tailed statistical significance at the p < 0.01 level.

Discussion and Conclusion

Long literatures in economics, public policy, and accounting document the myriad effects that sales tax rates have on the economy. In this paper, we move beyond the first-order effect of sales tax *rates* and explore whether sales tax *administration* policy can likewise affect the real economy. Other research has investigated this issue using experimental markets (Boylan and Frischmann (2006)) and surveys (Matarirano *et al.* (2019)), but we are the first to do so using real-world observational data.

Our interest in this question is driven by commentary from the tax industry suggesting that compliance costs are noticeably higher in settings in which municipalities or counties, instead of states, administer sales taxes (Institute for Professionals in Taxation (2016); Moffa (2019)). We test this hypothesis in the Florida hotel industry, in which county-level hotel sales taxes have largely shifted, one county at a time, from being administered by the state to being administered by individual counties.

Counties often prefer local administration as it allows for a larger local government workforce because local-level administration means county governments can keep a portion of tax proceeds internally to administer the tax, as opposed to paying the state government for the service (St. Johns County Tourist Development Council (2019)). Also, county-level administration puts enforcement under the purview of local officials, who tend to be better informed than state-level officials about smaller short-term rental operations such as vacation homes or beach condos (property that, unlike hotels, lack a large regulatory profile and can sometimes avoid state-administered sales taxes) (Florida Tourist Development Tax Association (2001a and 2001b)).

We examine whether counties switching to this type of "home rule" administration of local hotel sales taxes impose incremental compliance costs that meaningfully affect local employment and investment decisions. To do so, we control for the overall level of local hotel sales taxes, and then use fixed effects regressions to examine whether the move from state-level administration to county-level administration of local hotel sales taxes affects the size of the hotel industry within-county.

Controlling for tax rates, we find that shifting from state to local administration of county-level hotel sales taxes has a negative effect on the size of the hotel industry as measured by countywide number of hotel employees and countywide hotel industry payroll. Importantly, our OLS models use both county and year fixed effects. The inclusion of county-level fixed effects translates to our analysis being within-county, and our identification subsequently stems from counties that move from state-level administration of local hotel sales taxes to county-level administration (as opposed to exploiting differences between counties).

The main remaining threat to causality and our interpretation of results in this econometric setup comes from some alternative channel through which local administration of county-level hotel taxes could negatively affect hotels besides increased compliance costs. One such potential alternative channel could be stronger enforcement. This is one reason counties switch to local administration, but commentary from the profession suggests that this stronger enforcement is targeted at smalltime vacation rental operations that are sometimes able to avoid notice from tax officials due to their small size and minimal regulatory footprint (e.g., Florida Tourist Development Tax Association (2001a and 2001b); London (2018); Swoboda (2017)). Hotels, by comparison, are unable to fly under the radar and avoid sales taxes even under more relaxed enforcement by state-level officials. This inability of hotels to avoid sales taxes stems from their sizable regulatory profile. Unlike vacation rentals, hotels are paying commercial property taxes, have special purpose zoning, and are subject to a variety of state and local government inspections (health, fire, food service, etc.).

Given this reasoning and associated industry commentary, we have no reason to believe that hotels face any stricter enforcement of local hotel sales taxes under county-level administration. If anything, hotels should see some operational benefit if county-level local hotel sales tax administration translates into stricter tax compliance from vacation rentals that compete with hotels. With these institutional details in mind, and with our experimental design exploiting dozens of staggered within-county policy changes, we endorse a causal interpretation of our treatment effects.

Our result has clear implications for those who set sales tax policy. Controlling for the level of sales taxes, adding complexity such as local administration to the compliance process can actively depress economic development and employment in affected industries. These consequences should be weighed against the benefits of local tax administration, such as higher government employment and potentially stronger enforcement.

This issue is particularly pertinent to the hotel industry, as 19 states allow local administration of hotel taxes. Arizona, Colorado, Idaho, and Louisiana also require local administration of *general* local-level sales taxes, even though the state department of revenue in each case is also administering a parallel state-level sales tax. KPMG estimates that in Alabama alone, businesses collectively spend an *extra* $60 million per year in compliance costs necessitated by the local administration of sales taxes.

This may seem to be a modest figure on a statewide basis, but this sales tax compliance cost burden falls disproportionately on young, small businesses (Evans *et al.* (2014)). These are exactly the firms that drive an outsized share of employment growth (Neumark, Wall, and Zhang (2011)), which likely contributes to our findings regarding hotel workforce and payrolls being negatively affected by local tax administration. This existing research, in conjunction with our findings, offers a clear recommendation for policymakers looking to boost economic development: simplify the sales tax compliance process. The accompanying compliance cost savings, at least in our setting, map to more jobs and more payroll.

References

Afonso, W. B. (2019). "The Barriers Created by Complexity: A State-by-State Analysis of Local Sales Tax Laws in Light of the Wayfair Ruling." *National Tax Journal* 72(4): 777–799.

Ayers, B., J. Seidman, and E. Towery (2019). "Tax reporting behavior under audit certainty." *Contemporary Accounting Research* 36(1): 326–358.

Baker, S. R., S. Johnson, and L. Kueng (2020). "Shopping for Lower Sales Tax Rates." *American Economic Journal: Macroeconomics.*

Boylan, S. J., and P. J. Frischmann (2006). "Experimental Evidence on the Role of Tax Complexity in Investment Decisions." *Journal of the American Taxation Association* 28(2): 69–88.

Bozanic, Z, J. Hoopes, J. Thornock, and B. Williams (2017). "IRS Attention." *Journal of Accounting Research* 55(1): 79–114.

Callihan, R. (2019). "Airbnb doesn't have to pay tourism tax in Florida, judge says. That's the host's job." *Bradenton Herald*, October 29.

Cameron, A. C., and D. L. Miller (2015). "A Practitioner's Guide to Cluster-Robust Inference." J*ournal of Human Resources* 50(2): 317–372.

Carnes, G. A., and A. D. Cuccia (1996). "An analysis of the effect of tax complexity and its perceived justification on equity judgments." *The Journal of the American Taxation Association* 18(2): 40–56.

Cox, S. P., and R. J. Eger (2006). "Procedural Complexity of Tax Administration: The Road Fund Case." *Journal of Public Budgeting, Accounting & Financial Management* 18(3): 259–283.

Cuccia, A. D., and G. A. Carnes (2001). "A closer look at the relation between tax complexity and tax equity perceptions." *Journal of Economic Psychology* 22(2): 113–140.

De Simone, L., R.C. Sansing, and J. K. Seidman (2013). "When are enhanced relationship tax compliance programs mutually beneficial?" *The Accounting Review* 88(6): 1971–1991.

Due, J., and J. Mikesell (1983). *Sales Taxation: State and Local Structure and Administration*. Baltimore: The Johns Hopkins University Press.

El Ghoul, S., O. Guedhami, and J. Pittman (2011). "The Role of IRS Monitoring in Equity Pricing in Public Firms." *Contemporary Accounting Research* 28(2): 643–674.

Evans, C., A. Hansford, J. Hasseldine, P. Lignier, S. Smulders, and F. Vaillancourt (2014). "Small Business and Tax Compliance Costs: A Cross-Country Study of Managerial Benefits and Tax Concessions." *eJournal of Tax Research* 12(2): 453–482.

Florida Tourist Development Tax Association (2001a). *Guide for Self-Administering Counties*. Orlando, FL: Florida Tourist Development Tax Association.

———. (2001b). *Self-Administration: Will It Benefit Your County*? Orlando, FL: Florida Tourist Development Tax Association.

Fox, W., L. Luna, and G. Schaur (2014). "Destination taxation and evasion: Evidence from U.S. inter-state commodity flows." *Journal of Accounting and Economics* 57(1): 43–57.

Gallemore, J., and M. Jacob (2020). "Corporate Tax Enforcement Externalities and the Banking Sector." *Journal of Accounting Research* 58(5): 1117–1159.

Georgia Department of Revenue (2020). *Georgia Sales and Use Tax Rate Chart*. Atlanta, GA.

Guedhami, O., and J. Pittman (2008). "The Importance of IRS Monitoring to Debt Pricing in Private Firms." *Journal of Financial Economics* 90(1): 38–58.

Hageman, A., D. Bobek, and L. Luna (2015). "The Influence of State Sales and Use Taxes on Manufacturers' Capital Expenditures and Employment." *Public Finance Review* 43(4): 458–484.

Hanlon, M., J. Hoopes, and N. Shroff (2014). “The Effect of Tax Authority Monitoring and Enforcement on Financial Reporting Quality.” *Journal of the American Taxation Association* 36(2): 137–170.

Harrington, J., H. Chi, and L. P. Gray (2017). Florida Tourism. In *Florida’s Climate: Changes, Variations, & Impacts*, edited by J. W. Jones, E. P. Chassignet, J. Obeysekera, and V. Misra, 297–309. Gainesville, FL: Florida Climate Institute.

Hoopes, J. L., D. Mescall, and J. A. Pittman (2012). “Do IRS audits deter corporate tax avoidance?” *The Accounting Review* 87(5): 1603–1639.

Institute for Professionals in Taxation (2016). *Locally Administered Sales and Use Taxes*. Sponsored Research. Atlanta, GA.

Jacob, M., R. Michaely, and M. A. Müller (2019). “Consumption Taxes and Corporate Investment.” *The Review of Financial Studies* 32(8): 3144–3182.

Kubick, T. R., G. B. Lockhart, L. F. Mills, and J. R. Robinson (2017). “IRS and corporate taxpayer effects of geographic proximity.” *Journal of Accounting and Economics* 63(2/3): 428–453.

Laffer, A. B., W. H. Winegarden, and J. Childs (2011). *The economic burden caused by tax code complexity*. Austin, TX: The Laffer Center for Supply-Side Economics.

London, A. (2018). “Airbnb officials say Flagler about to lose bed tax revenue.” *Daytona Beach News-Journal,* July 1.

Matarirano, O., G. E. Chiloane-Tsoka, and D. Makina (2019). “Tax compliance costs and small business performance: evidence from the South African construction industry: research.” *South African Journal of Business Management* 50(1): 1–9.

Mikesell, J. L. (2007). “Developing Options for the Administration of Local Taxes: An International Review.” *Public Budgeting & Finance* 27(1): 41–68.

Moffa, J. (2019). *Pasco County Tourist Development Tax Reported to County*. Fort Lauderdale, FL: Moffa, Sutton, & Donnini, P.A.

Nessa, M., C. M. Schwab, B. Stomberg, and E. M. Towery (2020). “How do IRS Resources Affect the Corporate Audit Process?” *The Accounting Review* 95(2):311–338.

Neumark, D., B. Wall, and J. Zhang (2011). “Do Small Businesses Create More Jobs? New Evidence for the United States from the National Establishment Time Series.” *The Review of Economics and Statistics* 93(1): 16–29.

Shon, J., and I. H. Chung (2018). “Unintended Consequences of Local Sales Tax: Capitalization of Sales Taxes into Housing Prices.” *Public Performance & Management Review* 41(1): 47–68.

St. Johns County Tourist Development Council (2019). Frequently Asked Questions. St. Augustine, FL: St. Johns County Government.

Swoboda, L. (2017). “County may collect bed tax in-house.” *The Apalachicola Times*, November 29.

Tang, T., Mo, P. L. L., and Chan, K. H. (2017). “Tax collector or tax avoider? An investigation of intergovernmental agency conflicts.” *The Accounting Review* 92(2): 247–270.

Thompson, J., and S. Rohlin (2012). “The Effect of Sales Taxes on Employment: New Evidence from Cross-Border Panel Data Analysis.” *National Tax Journal* 65(4): 1023–1041.

Wenner, K. (2020). *Diverting Tourist Development Tax Revenue. A Florida TaxWatch Briefing*. Tallahassee, FL: Florida TaxWatch.

Wong, J. D. (1996). “The Impact of Local Option Sales Taxes on Retail Sales, Employment, Payrolls, and Establishments: The Case for Kansas.” *The Review of Regional Studies* 26(2): 166–176.

Appendix

Exhibit A. Hamilton County

34 Hamilton	10.0%	Living and Sleeping Accommodations
	7.0%	All Other Taxable Transactions
		Local Option Tourist Development
	2%	November 1, 1996 – December 31, 2001
	3%	January 1, 2002 –
		Local Government Infrastructure Surtax
	1%	July 1, 1990 – June 30, 2005
		Small County Surtax
	1%	July 1, 2005 – December 31, 2019
		Expiration date extended to December 31, 2029

Exhibit B. Marion County

52 Marion	11.0%	Living and Sleeping Accommodations
	7.0%	All Other Taxable Transactions
		Local Option Tourist Development
	2%	January 1, 2005 – October 31, 2015
	4%	November 1, 2015 –
		Local Administration of Tourist Development Tax – Began April 1, 2008
		Local Government Infrastructure Surtax
	1%	January 1, 2003 – December 31, 2004
	1%	January 1, 2017 - December 31, 2020
		School Capital Outlay Surtax
	0.5%	January 1, 2005 – December 31, 2009

Exhibit C. Highlands County

38 Highlands	11.5%	Living and Sleeping Accommodations
	7.5%	All Other Taxable Transactions
		Local Option Tourist Development
	2%	January 1, 2003 – July 31, 2018
	4%	August 1, 2018 –
		Local Administration of TDT Began January 1, 2014 – Ended March 31, 2018
		Local Government Infrastructure Surtax
	1%	November 1, 1989 – October 31, 2004
		Expiration date extensions to: December 31, 2018; and December 31, 2033
		School Capital Outlay Surtax
	0.5%	January 1, 2017 - December 31, 2036

Using Discrete Event Simulations To Understand the Impact of Changes to IRS Processes

Deandra Reinhart, Rafael Dacal, Patrick Kaylor, and Ariel S. Wooten (IRS Small Business/Self-Employed Research) and Jonathan Curtiss (MITRE Corporation)

Introduction

Background

The automated underreporter system (AUR) is a Small Business/Self-Employed (SB/SE) Division program that "systemically identifies [potential cases] through [the] computer matching of tax returns with corresponding Information Returns Master File (IRMF) payer information documents" (IRM 4.19.3.1.1). Those cases that are systematically identified are further reviewed by a tax examiner to evaluate underreporting and/or over-deduction discrepancies.

Prior to using a simulation model to test an organizational process change, SB/SE may disrupt the existing system to conduct a pilot (test) to assess the impact of the change on the process's outcomes and measures. These pilots may result in "down time" as employees are trained and gain efficiency on the proposed process changes. SB/SE Research recognized that a discrete event simulation (DES) could evaluate the impact of changes relating to resource allocation, employee scheduling, and capacity planning on process performance with minimal, if any, disruptions to existing operations.

In short, the DES allowed SB/SE Research to predict the impact of potential resource allocation or process changes on a particular SB/SE operation. Additionally, this project worked as a proof of concept to determine whether DES could be expanded to other SB/SE work units, helping to conduct experiments without disturbing their operations.

Project Objective

The AUR DES objectives are:

1. To assess the efficacy of the DES, and
2. To determine how changes in AUR resources or processes might impact AUR operations under different experimental scenarios.

Market Segment and Sample Frame

The internal IRS market segment for this project included all AUR clerical employees and tax examiners. The external IRS market segment included all taxpayers impacted by the AUR process. This included taxpayers in recipient of AUR Notices (i.e., CP2000,[1] Notice CP2501,[2] CP3219A,[3] and Reconsideration). Because of the complexity of the AUR program, the sampling frame is limited to AUR Philadelphia Campus employees and taxpayers in the Philadelphia Campus AUR inventory for Tax Years 2016 through 2019.

---

[1] CP 2000—*Request for Verification of Unreported Income, Payments, and/or Credits.*

[2] CP 2501—*Initial Contact to Resolve Discrepancy Between Income, Credits, and/or Deductions Claimed on Return & Those Reported by Payer.*

[3] CP3219A—*Statutory Notice of Deficiency.*

Limitation

In the development of this proof of concept, SB/SE decided to focus on the Philadelphia campus. Philadelphia was selected after case complexity, related cost, time availability, and the type(s) of inventory worked were evaluated across all AUR sites. Due to these specifications and local constraints, this exact model cannot be applied to other campus locations. The Philadelphia DES model, however, is currently being modified to accommodate differences across campuses. After these adjustments are completed, we will repeat the verification, calibration, and validation process to ensure that the updated model retains reliability and fidelity.

Method

Simulation models are commonly used across industries.[4] Yet, what is a simulation? A simple example, with which most people are familiar, comes from the news: the weather forecast is derived from simulations of the weather system. The AUR simulation attempts to imitate a dynamic system, defined as a system that changes over time, rather than simulating something that does not change: a static system.[5] Since we are simulating an operational system within the IRS, we used the model to predict AUR's performance under different sets of inputs, for example comparing the process changes after the building closed, due to COVID-19, to normal business conditions. To this end, we conducted a computer-based simulation: "An imitation (on a computer) of a system as it progresses over time."[6] The goals of the AUR simulation were twofold. Our first goal was to better understand AUR. Second, and more importantly, we wanted to use the simulation to identify the impact of changes within the system on overall process output and measures. This leads us to Robinson's definition of simulation:

> *Experimentation with a simplified imitation (on a computer) of an operations system as it progresses through time, for the purpose of better understanding and/or improving that system.*[7]

There are four types of simulations: DES, Monte Carlo simulations, systems dynamics, and agent-based simulation. SB/SE Research, with the help of the MITRE Corporation, selected a DES because it is a good option for use with queueing systems. The system is modeled as entities; in AUR's instance, this is a taxpayer going from one timing delay to another. These timing delays are termed activities, and this model is appropriate because the AUR process can be considered as a list of chronologically sequenced events through which a taxpayer must proceed to rectify potential underreporting of income and payment of additional taxes due.

A DES is also useful for estimating the impact of rare events such as shutdowns or major disasters, such as weather events, on operations. The IRS has experienced several weather events and federal government shutdowns in recent years. For example, there were three government-wide shutdowns over the last 10 years lasting between three and 35 days. In addition, severe weather events (such as hurricanes Maria[8] and Harvey[9]) resulted in extended U.S. Government office closures lasting from one week to three months. The COVID-19 pandemic closure was administratively treated like the 2018–2019 government shutdown. These government shutdowns or office closures can be examined with a DES model similar to how DES is used in industry for factory shutdowns to update an assembly line.

Another inherent benefit of a DES model is its ability to simulate process adjustments and shifting resources on the basis of stakeholder suggestions. This flexibility allows the IRS to proactively predict potential system outcomes resulting from changing agency policies and procedures. Finally, a DES can assist IRS operations in predicting consequences to policy and procedure changes without impacting tax administration or the quality of service provided to taxpayers.

---

4 The following section is adapted from Robinson (2019).

5 Systems can best be thought of as a group of components brought together for a larger purpose.

6 Robinson (2019).

7 *Ibid.*, p. 5.

8 "GSA makes progress in recovery from Hurricane Maria | GSA" (https://www.gsa.gov/about-us/regions/welcome-to-the-northeast-caribbean-region-2/region-2-newsroom/feature-stories/gsa-makes-progress-in-recovery-from-hurricane-maria).

9 "Feds in Harvey's Path Are Sleeping in Their Offices to Keep Agencies Operational" (https://www.govexec.com/pay-benefits/2017/08/feds-harveys-path-are-sleeping-their-offices-keep-agencies-operational/140685/).

The following subsections describe how the model was developed and the steps taken to evaluate the overall model.

Business Process Model Development

Developing the Business Process Model (BPM)

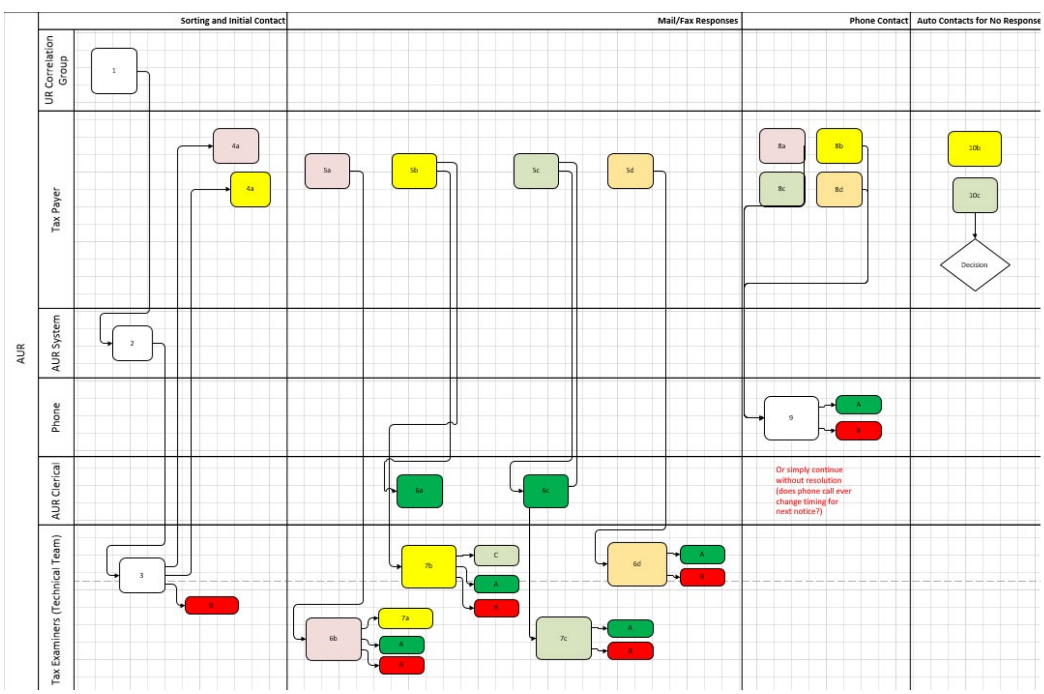
The first step in completing a successful DES is creating a BPM of an operation that is being modeled. According to Banks *et al.* (2010), a common approach to DES simulations is the process-interaction approach. While using this approach,

> *The analyst defines the simulation model in terms of entities or objects and their life cycle as they flow through the system, demanding resources and queueing to wait for resources. More precisely, a process is the life cycle of one entity.*[10]

In our model, the entity is an individual taxpayer account, but we may also use the terms "case" or "module." To develop an accurate BPM, we accessed information describing how cases/modules move between processes. This information was retrieved initially from the Internal Revenue Manual (IRM) and refined by subject matter experts (SMEs) familiar with the operation.

Once a process flow was developed, we decomposed it into lower levels of detail to support different phases of the simulation. Figure 1 shows the "top level" view of the AUR process flow. The initial business process flow contained the most detailed information and specified all activities involved in the process that was modeled. After the final version of the "top level" BPM was developed, SB/SE Research validated it with SMEs familiar with the activities being modeled. SB/SE Research used existing BPMs, flowcharts, IRMs, and onsite visits (when possible) to produce the final product.

FIGURE 1. Business Process Model – Process Flow



The type of information within the model shows how cases/modules advance from activity to activity (i.e., from point A to point B, then from B to C), the paths cases/modules follow through the process, along with

[10] Banks *et al.* (2010).

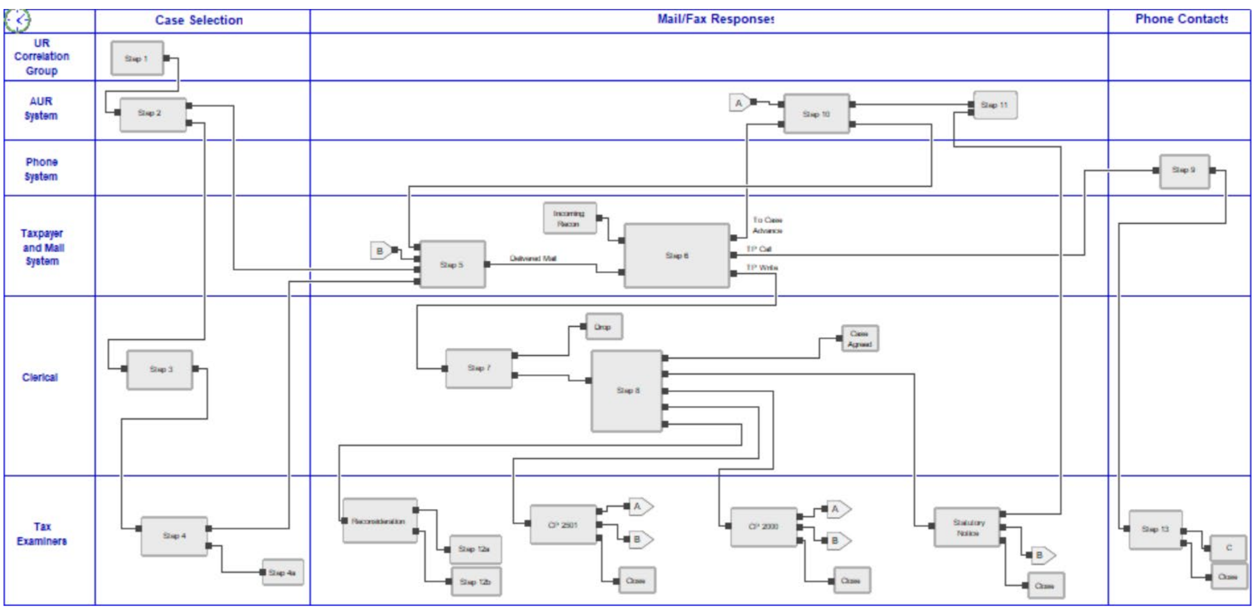
the frequency of taking each path. When developing the model, SB/SE Research considered and documented all the potential paths by which a case may travel through the process.

SB/SE Research drafted the business process model using commercial off-the-shelf (COTS) software, in this case Microsoft Visio. In addition, SB/SE Research documented the process flows following business process modeling and notation (BPMN) standards. BPMN provided a standard format for documenting how the case cycled through the process, what occurred at each activity, and how cases/modules advanced or failed to advance from activity to activity. At a minimum, SB/SE Research strived to clearly describe what happened at each activity, who performed the activity, and how the cases/modules were routed so that this information could be used later as inputs in the simulation software. After the business process model (BPM) flow was constructed, SB/SE Research evaluated the configuration of all inputs, path decisions, timing, resources, and outputs.[11]

Developing the Executable BPM

Once the BPM was completed, SB/SE Research moved the model into the simulation software to complete the DES. Figure 2 shows the structure of the executable business process model developed for the AUR simulation. Note how the activities are contained in "swim lanes," horizontal rows indicating ownership of the activities. This model was constructed using the ExtendSim software program.

FIGURE 2. Business Process Model—Executable Model



Fidelity

Simulation models seek to duplicate real-world sequences of events to predict potential outcomes. The Fidelity Definition and Metrics Implementation Study Group[12] defined fidelity as:

> *The degree to which a model or simulation reproduces the state and behavior of a real world object or the perception of a real world object, feature, condition, or chosen standard in a measurable or perceivable manner; a measure of the realism of a model or simulation; faithfulness.*[13]

[11] The methods by which this evaluation is completed are detailed later in the paper.

[12] The Fidelity Definition and Metrics Implementation Study Group was created by Simulation Interoperability Standards Organization (SISO) on the basis of the interest of the simulation research community to standardize simulation definitions and fidelity. The SISO site is https://www.sisostds.org. See also Schricker *et al.* (2001).

[13] SIW Fidelity Report (99S-SIW-167) (n.d.). Retrieved from https://vva.msco.mil/Ref_Docs/Ref_Docs/99s-siw-167.pdf.

In other words, fidelity in the context of this project is how well the BPM depicts the inner working (event, timelines, sequence, production actuals, etc.) of AUR operations.

Verification, Validation, and Calibration

According to Banks *et al.* (2010),[14] model developers need to work closely with end-users throughout model development so that the model satisfactorily represents reality, and the model results can be used later in the decision-making process. Sargent (2011) reminds us, however, that developers and end-users need to determine at what point a model is sufficiently accurate to accomplish its intended purposes. Once this accuracy is agreed upon, the simulation model needs to be judged by that metric, and validation and verification are methods that we use to evaluate whether the standard has been met.[15] Through this process, developers need to calibrate their model as well, as an ongoing check of the input parameters developers have included.

The Department of Defense (DoD) has formally defined both verification and validation. Verification is defined as "the process of determining that a model or simulation implementation and its associated data accurately represent the developer's conceptual description and specifications."[16] The verification process attempts to ensure that the model is constructed correctly, and that the data accurately represent the conceptual model. This can be addressed through the following procedures:

- Debug the computer program; begin with a simple model and add complexity as needed.
- Ensure group involvement among developers to check one another's work.
- Conduct hand calculation checks.
- Conduct simple model runs, where the assumptions or conditions are known to be true, with added complexity.

Validation is defined as "the process of determining the degree to which a model or simulation and its associated data are an accurate representation of the real world from the perspective of the intended users [sic] of the model."[17] The model should represent a system's behavior such that end-users can make informed decisions from its results, while maintaining that outcomes cannot be predicted with certainty.

Sargent (2011) argues that developers must go through two activities while validating a model. The first is conceptual model validation. During this phase, the modeler and end-users agree that the model's theories and assumptions are correct, and that the model is a reasonable representation of the process under investigation. The second activity is operational validation, whereby the model's accuracy is judged to be sufficient. These are iterative processes and must be repeated during model development. Sargent (2011), Strickland (2012), and Banks *et al.* (2010) agree upon the following methods necessary to validate a model:

- Ensure that the model has high face validity, and is approved by end-users or other experts, or is consistent with theory relating to a specific process.
- Test the model's assumptions empirically through goodness of fit tests and sensitivity analyses.
- Test the model's output with real world data outputs.

While validating a model, we should think in terms of how we may need to calibrate it. Calibration is an iterative process whereby modelers compare the model to the system being simulated, but also adjust the model parameters when they are incorrect (Banks *et al.* (2010)). This process is done repeatedly: an initial model is developed, which is compared to the actual system, and then a first revision of the model is completed. This process is not complete until an agreed-upon level of model accuracy is achieved. Banks *et al.*

[14] Banks *et al.*, ibid.

[15] This decision on the accuracy level should be made early in model development.

[16] Department of Defense Instruction 5000.61. (December 9, 2009) Incorporating Change 1, October 15, 2018 (whs.mil).

[17] *Ibid.*

(2010) summarize a three-step approach to operational validation and calibration. First, work with end-users to build a model with high face validity. Next, continuously validate the model's assumptions. Finally, compare the model's input and output values with data collected from the real-world processes being modeled.

AUR Model Inputs

For the AUR model, inputs were inserted into the executable model in phases. The phases were dependent on where in the AUR model the information input was necessary to continue to the next step in the AUR process. A simple example is the following: mail volume data were used before taxpayer response rates because of the timing of these events, i.e., mailing the notice precedes the taxpayer response to the notice. The input variables were obtained from SMEs, published IRMs, or AUR reports. Examples of the AUR simulation inputs include the following:

- AUR inventory metrics, including:
  - Taxpayer response rate(s)
  - Volumes of mailed notices (CP2000, CP2501, Statute Notice, and Reconsideration)
  - Inventory age
- Resource pools (a group of resources available to work specific parts of a process). Resource pools identified the number and type of resources available for each task. For AUR, we listed resource availability by week.
- Time measures, such as:
  - Shifts that determined when resources were available during the week. They showed the times of day resources were working and the days of the week.
  - Time it took for resources to complete an activity. It can be the amount of time a resource takes to perform some activity or an analysis of how many transactions were performed over time. These types of inputs determined how long, on average, a single resource took to perform each identified activity.
  - Expected time for the IRS to compile and mail a notice
  - Expected time for the taxpayer to respond
  - Time intervals not associated with any resource
- External factors, such as USPS mailing time or macroeconomic variables if applicable

Output Logs

The outputs logs are the executable model's estimates acquired at the beginning and the end of identified activities. Examples of these estimates are volume of notices processed or mailed, time to complete an activity such as batching, and time to complete a screening. The output logs were defined similarly[18] to the inputs and include the following:

- AUR inventory metrics, including:
  - Taxpayer response rate
  - Mailing notice volumes (CP2000, CP2501, Statute Notice, and Reconsideration)
  - Inventory age
- Expected time for activities
- Time to complete activities
- Related call times

[18] SB/SE Research used "similarly" here because not all input and output variables are exactly defined. For example, a SME or an AUR document may provide an input variable based on a population parameter when actual production data are not available. Therefore, the input is a scalar, but the output is a vector in the model. For those variables for which this definition does not apply, the inputs and outputs define the same way.

The output log data were used to evaluate the model to ensure that the previously agreed-upon accuracy level was met. The output logs were also used to calculate changes in AUR inventory metrics (e.g., compare actuals to estimated mailing volumes).

Findings

As a result of this project, SB/SE Research found that DES models can provide insights to identify and remove bottlenecks, forecast the impact of changes to work volume and staffing, and ensure the optimum use of resources to meet work demands, without disrupting existing work processes.

SB/SE Research conducted two experiments to validate the AUR DES model:

1. Post-COVID restart of campus operations, and
2. Impact of a notice redesign.

These experiments evaluated the impact on tax examiners' workload and overall output without causing operational disruption. The post-COVID restart experiment allowed AUR leadership to assess bottlenecks in the process created by campus closures due to COVID. In this experiment, SB/SE Research forecast the backlog of correspondence and calculated the time required by the AUR staff to clear and process the backlog, given different restart assumptions. SB/SE Research also calculated the potential impact that the campus reopening could have on taxpayers' responses and downstream impact on tax examiners' case workload (including correspondence volume, case age, and screening).

In the notice redesign experiment, the experiment assumed that the redesign of the CP2501 and CP2000 would increase taxpayers' response rates. SB/SE Research evaluated the impact of a 10-, 20-, and 30-percent increase in taxpayer response to IRS notices. This assumption was integrated into the DES model by increasing the volume of taxpayer responses and increasing case resolution rates. The results of this experiment complemented other IRS studies, such as the redesign of AUR notices issued to taxpayers. The results from this research demonstrated DES' effectiveness at illustrating workflows, visualizing taxpayers' interactions, and assessing IRS caseload completion to improve taxpayer compliance and operational efficiency.

General Application

In this section, we provide potential examples of how a DES model could be applied. It is also important to indicate that the development of a DES model is expensive and time-intensive. Therefore, this type of effort should be applied only in areas where it can be most used.

Process changes

The DES method can be used to simulate changes to processes. The AUR simulation, for example, could estimate how changes to the automated routing calling system or new batching process may impact the overall process flow, and therefore predict future changes to downstream work. The DES can also simulate different campus staff work-from-home policies to help leadership determine the best policy based on the simulations' results. In addition, this method could be used to simulate changes to taxpayer response rate as a result of notice redesign (which assumes an improvement in taxpayer response rate). This type of experiment could assist the operation with identifying hiring needs and the impact on inventory and life cycle.

Resource allocation

DES models can simulate changes to resources allocation to estimate how such changes may impact processing time, time to completion, and inventory. For example, AUR may run multiple DES experiments to estimate where new hires might best be placed for optimal processing time.

Conclusion

In conclusion, the development of the AUR Philadelphia simulation model and subsequent experiments validated that the use of DES can be used to assess changes to complex processes while minimizing disruptions to existing systems. Furthermore, the AUR DES provides SB/SE leadership with data-driven analysis of process and resource changes.

For the post-COVID-19 campus restart, the results provided AUR leadership with the ability to identify potential risks and develop mitigation strategies to minimize inventory issues. As dynamics within the IRS quickly changed between March 2020 and August 2020 due to COVID-19, the DES model was shown to be a useful tool. Making modifications within the DES model allowed leadership to assess the impact of changes within this environment to minimize service disruption to taxpayers. For the notice redesign experiment, the information provided by the model helped the IRS understand how changing the notices to increase taxpayer responsiveness may affect both the volume of correspondence and the downstream effects. In summary, the development of this model and the accompanying experiments showed that this method is a useful tool for analysis throughout the tax administration community.

References

Banks, J., *et al*. (2010). *Discrete-Event System Simulation,* 5th Ed. New York: Prentice Hall.

Robinson, Stewart (2019). *Simulation: The Practice of Model Development and Use.* London: Red Globe Press.

Sargent, R. (2011). *Verification and Validation of Simulation Models.* 2011 Winter Simulation Conference. https://www.informs-sim.org/wsc11papers/016.pdf.

Schricker, B., *et al*. (2001). "Considerations for selective-fidelity simulation," *Proceedings SPIE 4367, Enabling Technology for Simulation Science V* (19 September 2001); https://doi.org/10.1117/12.440055.

Strickland, J. (2012). *Discrete Event Simulation Using ExtendSim 8.* Colorado Springs: Simulation Educators.

Effects of Forms 1040-X and 1120-X Post-filing Adjustments on SOI Estimates[1]

Derrick Dennis, Jennifer Ferris, Chloe Gagin, Tuba Ozer-Gurbuz, Julia Shiller, and Christopher Williams (IRS, Research, Applied Analytics, and Statistics, Statistics of Income Division)

Introduction

Historically, Statistics of Income (SOI) research has been based on data collected from initial tax return filings.[2] Although initial tax return filings are often final, taxpayers may elect to amend their income tax returns.[3] Common reasons for amending an initial return include tax over- or underassessments, or the taxpayer filing an amendment after identifying errors on the original return.[4]

Taxpayers have an option to amend their initial return after the tax filing period has ended. Under Internal Revenue Code (IRC) § 6511, taxpayers must file a claim for a credit or refund within 3 years from the time the initial tax return was filed or 2 years from the time the tax was paid, whichever is later. In some instances, extensions may be granted for filing an amended return.

Post-filing changes can influence net revenue, which consequently impacts fiscal research and policy. However, to publish statistics in a timely manner, SOI does not wait for the processing of amended return data. Thus, amendments are excluded from the SOI survey population sampling frame and are not reflected in SOI estimates. Additionally, incorporating these changes into SOI statistics would require substantial SOI resources in terms of time and labor, which may not be justifiable depending on the magnitude of the impact.

Nevertheless, it may be valuable to IRS stakeholders to understand the difference between currently published statistics and taxpayers' finalized tax statistics. This study aims to contribute to a better understanding of the difference by measuring the impact of including amended returns to a subset of SOI products.

Information in this article is presented as follows. First, the scope of the project is defined and previously published data are summarized. Next, new results are presented for amendments to both individual and corporate income tax returns. Conclusions, as well as a discussion on the limitations to this study and suggestions for future work, follow. Additional information is listed in the appendix.

Scope of Project

This study aims to estimate the effect of including amended tax returns in SOI products. It focuses on individual and corporate returns—a U.S. Individual Income Tax Return (Form 1040) or a U.S. Corporation Income Tax Return (Form 1120)—filed in a selected tax year that were amended using either an Amended U.S. Individual Income Tax Return (Form 1040-X) or an Amended U.S. Corporation Income Tax Return (Form 1120-X).[5]

The base year for this study is SOI Year 2013. This means that this study measures changes made in subsequent years to returns that were filed for SOI Year 2013. SOI Year 2013 was chosen because it is a relatively recent year with 3 years of post-filing data. There are a number of exemptions to the general rule of 3 years, so this study identified all amendments to SOI Year 2013 returns at the time the data were pulled.

---

[1] This article was also published in the IRS *SOI Bulletin* Volume 40, Number 3. https://www.irs.gov/pub/irs-prior/p1136--2021.pdf

[2] For some studies, SOI replaces initially filed tax returns with superseding returns that are filed within the same SOI study year.

[3] Taxpayers can add, delete, update, or change their income, exemptions, deductions, credits, filing status, etc., reported on their original tax returns, including claiming tax credits and deductions that were not previously claimed.

[4] In some cases, post-filing adjustments may start out as an overassessment and turn into an underassessment (or vice versa).

[5] Although there are other ways to amend Forms 1040 and 1120, this study is limited to Forms 1040-X and 1120-X for amendment estimates.

This study uses IRS Master File-level data as well as the SOI samples to capture changes to individual and corporate returns. Tax adjustments were identified in the Master File data and linked to tax returns in the individual and corporate SOI samples. The matched data were weighted using the respective sample weights from the individual and corporate studies to represent adjustments in the overall population. Finally, the weighted sample estimates were compared to total changes in the population data to verify the appropriateness of SOI sample weights for this study.

Existing Estimates and Published Data

The estimates reported in this section are examples of published data that include information on amendments.

The IRS annually reports data about Form 1040-X and 1120-X filings in Calendar Year Return Projections for the United States and IRS Campuses (Publication 6186).[6] This publication includes the number of returns that are processed during the given calendar year as well as prior years, and projections of returns to be filed for future years.[7] As reported in the 2014 release of Publication 6186, the IRS processed 145,851,813 individual income tax returns and 6,685,423 corporate income tax returns in Calendar Year 2013. For this same calendar year, the IRS processed 3,951,923 amended individual income tax returns (Forms 1040-X) and 7,000 amended corporate income tax returns (Forms 1120-X).

The data published in Publication 6186 provide estimates for the number of amended returns processed (based on Forms 1040-X and 1120-X) in a given year. However, the amended returns reported on in this publication could be associated with initial filings from any prior tax year. Currently, no published data reflect the number of amended returns associated with a fixed initial filing tax year. Existing studies do not follow the returns from a fixed initial filing tax year to capture the related amended returns. Therefore, published data cannot shed light on how these changes would affect SOI's tax year estimates.

Methodology

Datasets containing original population data for individuals and corporations were selected from Master Files and other IRS databases. These datasets contain multiple years of transactions regarding filed income tax returns, including data on subsequent adjustments to the initial filings. Master File data on Forms 1040-X and 1120-X include net total tax change amounts.

Population data were matched to the SOI Form 1040 and Form 1120 sample data. The data for individual income tax returns were matched using the Taxpayer Identification Number (TIN). The data for corporate income tax returns were matched using a corporation's Employer Identification Number (EIN) and accounting period.[8] It is necessary to use accounting period in addition to EIN since some corporations could have multiple original returns in a file due to short tax year corporate returns[9] or other reasons. The SOI sample data for this research were used to produce the 2013 Individual Income Tax Returns Complete Report (Publication 1304) and the 2013 Corporation Income Tax Returns Complete Report (Publication 16), otherwise known as the Complete Reports for individual and corporate returns.

The individual sample selection occurred during Calendar Year 2014. This study included all filed Forms 1040-X that corresponded to a return selected for the 2013 individual sample. The corporate sample selection differs from the individual sample selection. The corporate sample selection occurred during the period July 2013 to June 2015, which allowed for the capture of returns filed in different accounting periods, as well as those that had changes within the limitation period, including filing extensions. By limiting the accounting period in

[6] Publication 6186 reports the number of tax returns filed for the United States and IRS processing campuses by major return categories for each calendar year. Data on amended returns can be found in Publication 6186, Table 2, "Total Number of Returns Filed by Type for the United States."

[7] The calendar year runs from January 1 to December 31.

[8] Corporations can choose to adopt a calendar year, fiscal year, or 52-53-week year. If a corporation adopts the calendar year, it should maintain its books and records and report income and expenses from January 1 through December 31 of each year. A fiscal year is 12 consecutive months ending on the last day of any month except December. This study considers 52-53-weeks year filers as fiscal year filers.

[9] A short tax year is a tax year of less than 12 months. A short period tax return may be required when a corporation is not in existence for an entire tax year, or it decides to change an accounting period. Tax on a short period tax return is figured differently for each situation.

accordance with the SOI sample, this study sought to include filed Forms 1120-X that corresponded to returns selected for the 2013 SOI corporate sample.[10]

Results

The following estimates are based on Master File data that were not edited or reviewed for quality. Where results are presented in this section, Forms 1040-X and 1120-X are referred to as amended individual and corporate returns, and Forms 1040 and 1120 are referred to as individual tax returns and corporate tax returns. Similarly, Form 1040 and Form 1120 filing populations are referred to as individual and corporate filing populations, respectively.

Individual Income Tax Returns

For Tax Year 2013, taxpayers filed 142.9 million individual income tax returns, of which 3.2 million (2.3 percent) were amended at least once after the initial return was filed (Table 1). Of the amended returns, tax increases totaled $2.5 billion and tax decreases totaled $4.6 billion, resulting in a net decrease of $2.1 billion.

Incorporating these adjustments from amended individual income tax returns into SOI statistics decreased the estimated total amount of reported tax liability from Form 1040 filers by 0.16 percent for Tax Year 2013.

In the analysis by adjusted gross income (AGI), taxpayers were separated into two groups, married filing jointly and filing nonjointly (i.e., single, married filing separately, head of household, and qualifying widow(er)).[11]

Population demographics

With respect to filing status, single, married filing separately, and widowed taxpayers represented roughly the same proportion of amended individual tax returns (Forms 1040-X) as they did individual tax returns (Forms 1040) (Figure A). Married filing jointly taxpayers filed 44.8 percent of all amended individual returns (Forms 1040-X) for Tax Year 2013, despite comprising 53.4 percent of the population of individual tax returns (Forms 1040) for the year. Conversely, head of household filers represented 16.6 percent of amended individual returns but only comprised 10.8 percent of individual tax returns.

Taxpayers with certain filing statuses are more likely to amend their initial returns (Figure B). Specifically, 4.6 percent of returns filed by qualifying widow(er)s were amended, compared to 2.9 percent of returns filed by those married filing separately and 2.5 percent filed by a head of household, versus only 1.7 percent of returns filed by single filers and 1.4 percent of returns filed by those married filing jointly.

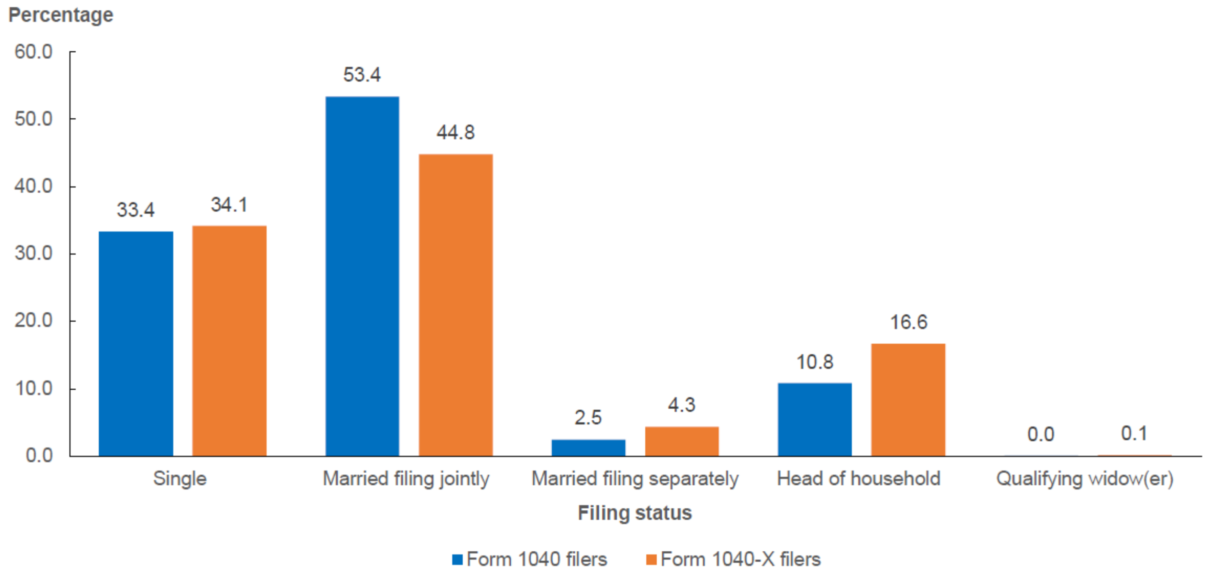
The age distribution among individual taxpayers who filed amended returns is roughly the same as among those filing initial tax returns (Figure C). Taxpayers between ages 30 and 70 represented a marginally higher proportion of amended returns than individual tax returns. Individuals in the under 30 and over 70 age categories accounted for a marginally lower proportion of amended returns than individual tax returns. The largest difference was in the 40-under-50 age group; these taxpayers filed 19.7 percent of the amended returns and 18.1 percent of the individual tax returns for Tax Year 2013. There were no age groups whose returns were markedly more likely to be amended (Figure D).

Nonjoint returns with an AGI between $1 and $25,000 made up a smaller proportion of amended returns than they did individual tax returns, accounting for 48.5 percent of amended returns and 52.9 percent of individual tax returns for the year (Figure E). Individual filers with an AGI less than $1 filed amended returns in nearly the same proportion as individual tax returns, 1.4 percent and 1.5 percent, respectively. Nonjoint returns with an AGI in the $100,000-under-$200,000 category were more likely to be amended (2.8 percent of filers)

[10] This study was unable to categorically exclude superseding corporate returns.

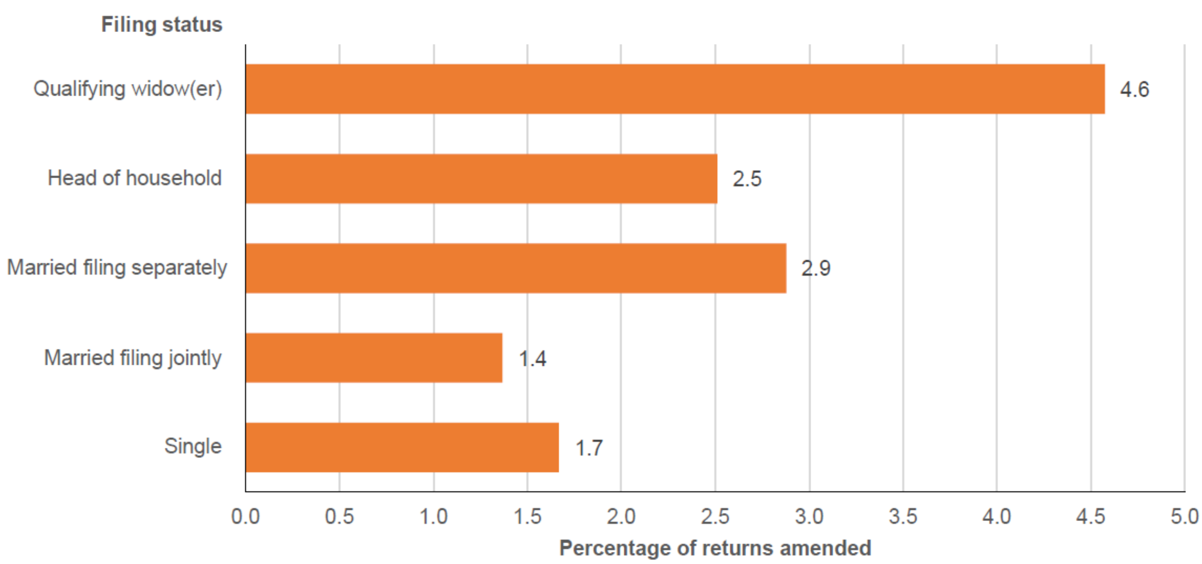
[11] The data were divided into married filing jointly and otherwise to avoid miscounting or erroneous interpretations. A married filing jointly couple with $60,000 in AGI cannot be counted as two taxpayers who each make $60,000, nor as one taxpayer who makes $60,000, nor as two taxpayers who make $30,000 each.

Figure A
Proportion of Form 1040 and Form 1040-X Returns Filed by Filing Status, Tax Year 2013



SOURCE: IRS, Statistics of Income, Weighted Sample, February 2021.

Figure B
Form 1040-X Filing Rates by Filing Status, Tax Year 2013



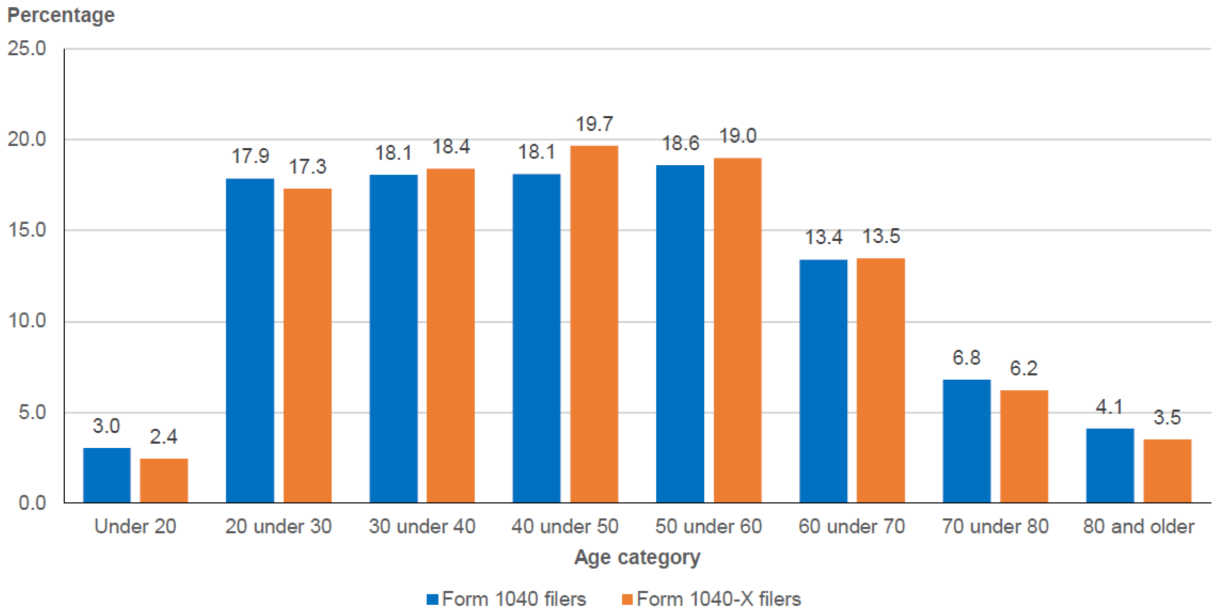
SOURCE: IRS, Statistics of Income, Weighted Sample, February 2021.

than returns of other AGI groups (Figure F). Filers with an AGI between $1 and $25,000 were the least likely to amend their returns (1.7 percent of filers).

Married filing jointly taxpayers with an AGI of less than $1 through the under $100,000 AGI category class of AGI filed a proportionally lower volume of amended returns than individual tax returns (Figure G). A significantly larger proportion of couples with an AGI of over $200,000 filed amended returns, comprising 14.0 percent of all amended returns and 8.9 percent of individual tax returns. These same taxpayers were more likely to file an amended return (around 2.2 percent of returns in this AGI group) than other AGI groups (Figure H).

Figure C

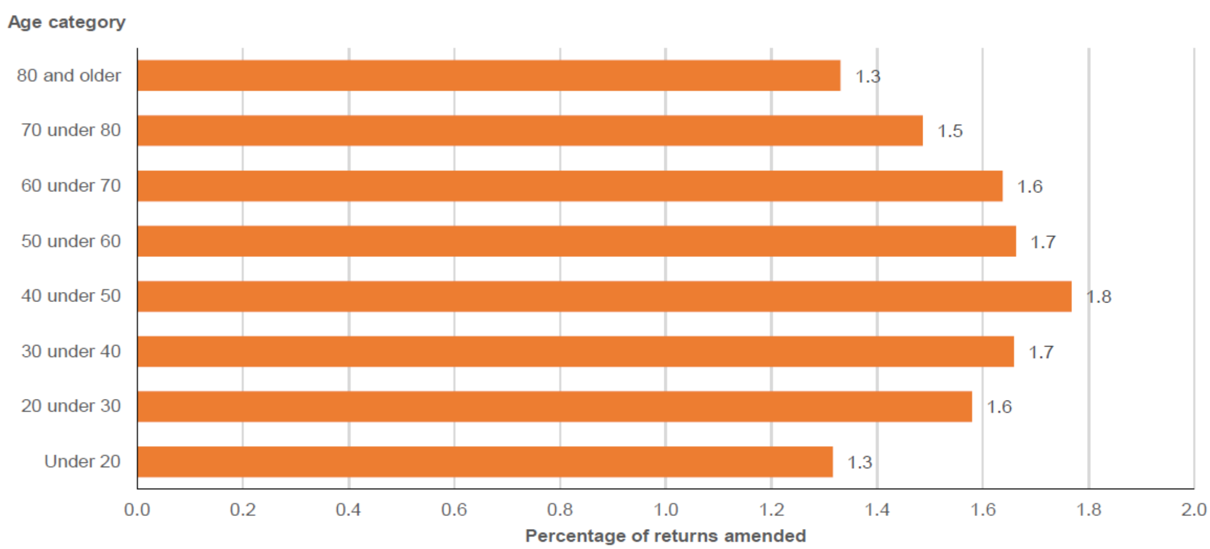
Proportion of Form 1040 Filers and Form 1040-X Filers by Age of Taxpayer, Tax Year 2013



SOURCE: IRS, Statistics of Income, Weighted Sample, February 2021.

Figure D

Form 1040-X Filing Rates by Age of Taxpayer, Tax Year 2013

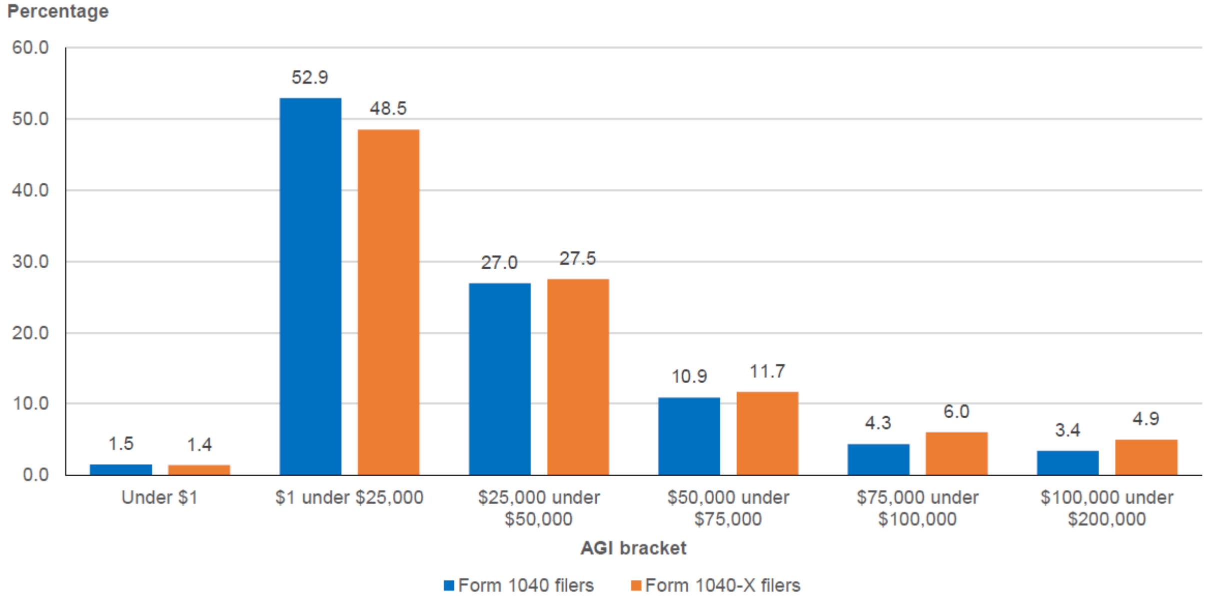


NOTE: Percentage labels were rounded and therefore may not align exactly with the vertical percentage lines.
SOURCE: IRS, Statistics of Income, Weighted Sample, February 2021.

Corporate Income Tax Returns

Of the 1.6 million corporate tax returns included in the 2013 SOI study year, 8,200 (0.52 percent) were amended after the initial return was filed (Table 2). These filers reported a net tax overpayment of $87.8 million. Tax increases for 1,300 amended corporate tax returns totaled $30.4 million while tax decreases for 1,800 returns totaled $118.2 million. There were 5,100 amended corporate tax returns with no tax change. Incorporating adjustments from amended returns into SOI statistics would decrease the estimated total amount of taxes paid by these filers by $87.8 million (0.03 percent) for Tax Year 2013.

Figure E
Proportion of Form 1040 and Form 1040-X Filers, Nonjoint Returns, by Selected Adjusted Gross Income Bracket, Tax Year 2013



NOTES: This figure excludes married filing jointly filers. This figure only includes filers with a marital status of single, head of household, widow(er), and married filing separately. Taxpayers whose AGI exceeds $200,000 were excluded from this analysis due to too few returns in that category.
SOURCE: IRS, Statistics of Income, Weighted Sample, February 2021.

Figure F
Form 1040-X Filing Rates, Nonjoint Returns, by Selected Adjusted Gross Income Bracket, Tax Year 2013

NOTES: This figure excludes married filing jointly filers. This figure only includes filers with a marital status of single, head of household, widow(er), and married filing separately. Percentage labels were rounded. Taxpayers whose AGI exceeds $200,000 were excluded from this analysis due to too few returns in that category.
SOURCE: IRS, Statistics of Income, Weighted Sample, February 2021.

Classification

The majority of amended corporate income tax returns were electronically filed. The IRS requires that if the original Form 1120 is e-filed, then the Form 1120-X must also be e-filed. Approximately 70 percent of Forms 1120 were e-filed compared to 74 percent of Forms 1120-X.

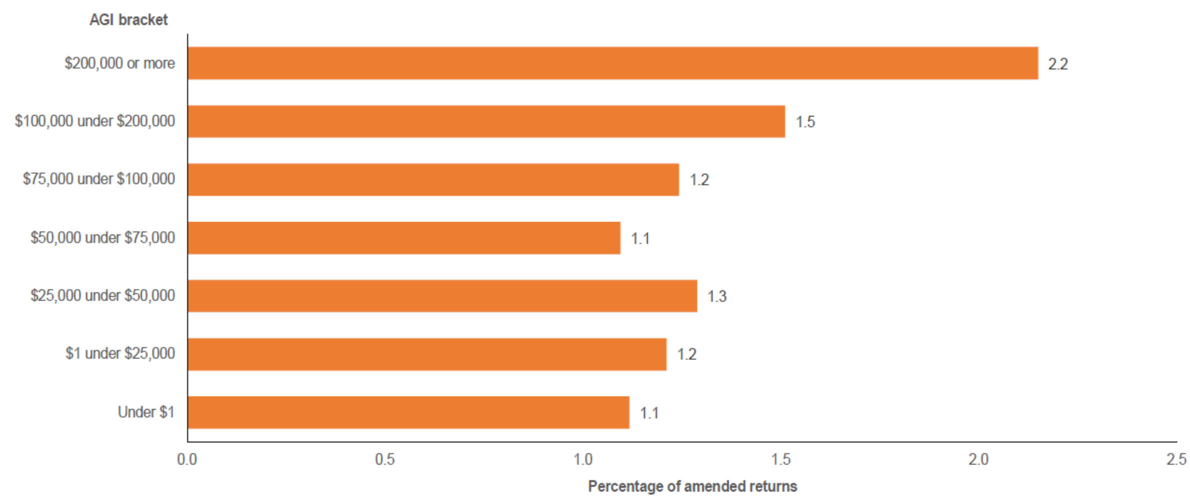
Figure G

Proportion of All Married Filing Jointly Form 1040 Filers and Form 1040-X Filers by Adjusted Gross Income Bracket, Tax Year 2013

SOURCE: IRS, Statistics of Income Division, Weighted Sample, February 2021.

Figure H

Form 1040-X Filing Rates, Married Filing Jointly Filers, by Selected Adjusted Gross Income Bracket, Tax Year 2013

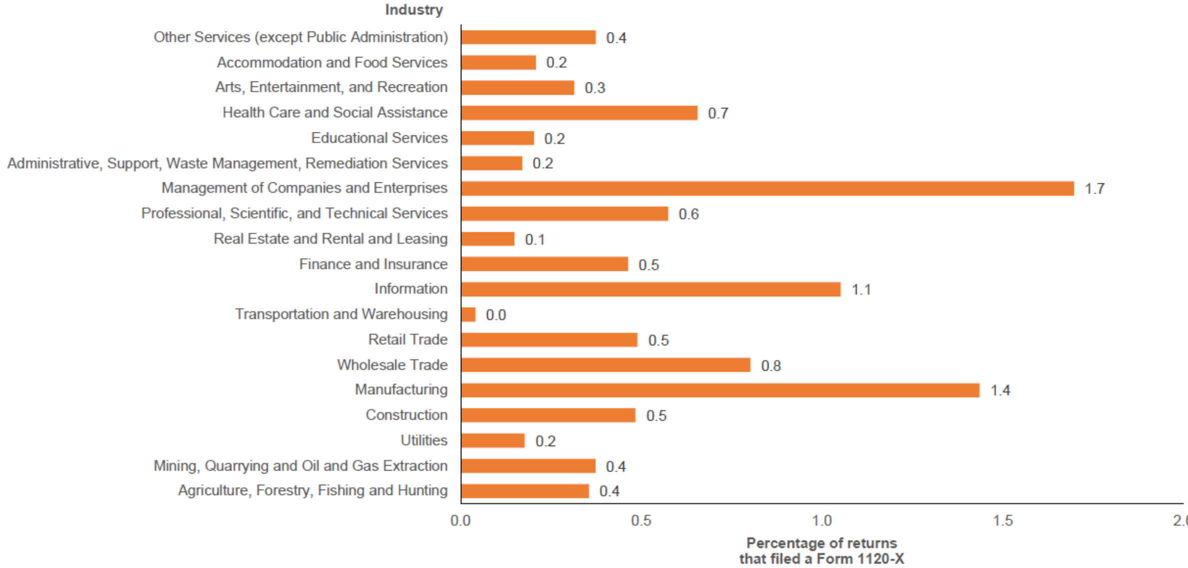


NOTE: Percentage labels were rounded.
SOURCE: IRS, Statistics of Income Division, Weighted Sample, February 2021.

Based upon SOI-assigned industry codes, the industrial sectors that had the highest rates of amended corporate tax returns were Management of Companies and Enterprises (1.7 percent), Manufacturing (1.4 percent), and Information (1.1 percent) (Figure I). The sectors with the lowest amendment rates were Transportation and Warehousing (0.04 percent), Real Estate and Rental and Leasing (0.1 percent), and Utilities (0.18 percent).

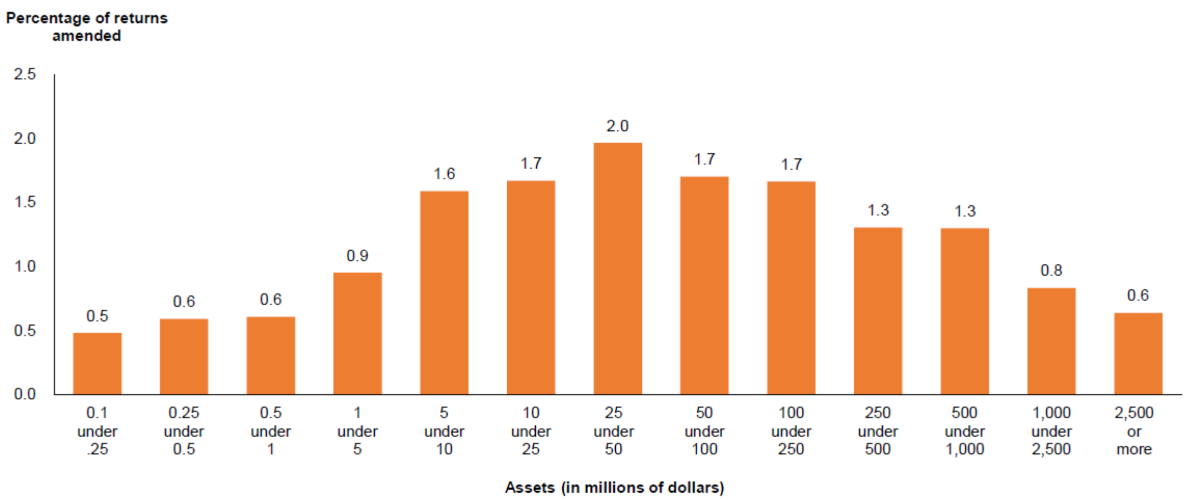
Corporations with assets between $100,000 and $250,000 had the lowest amendment rate (0.5 percent, Figure J). The amendment rates rose as assets increased up to the peak—corporations with reported assets between $25 and $50 million—where the amendment rate was approximately 2 percent, and then declined from there.

Figure I
Form 1120-X Filing Rates by Industry, SOI Year 2013



NOTE: Percentage labels were rounded.
SOURCE: IRS, Statistics of Income Division, Weighted Sample, February 2021.

Figure J
Form 1120-X Filing Rates by Size of Total Assets, SOI Year 2013



NOTE; Percentage labels were rounded.
SOURCE: IRS, Statistics of Income Division, Weighted Sample, February 2021.

Conclusion/Summary

In conclusion, this study sought to measure the impact that amendments would have on SOI tax return statistics. Based on the results, there is an important observation about SOI estimates that can be made for individual and corporate income returns. In this preliminary study looking at 2013, some 2.3 percent of individual income tax returns (Forms 1040) and 0.52 percent of corporate income tax returns (Forms 1120) were amended (using Forms 1040-X and 1120-X). These amendments decreased the estimated total tax from individual filers by only $2.1 billion (0.16 percent) and decreased the estimated total tax from corporate filers even less, by $87.8 million (0.03 percent).

Thus, the study's review of individual and corporate tax returns from 2013 suggests that while SOI tax return statistics would be affected by post-filing adjustments, current SOI statistics do not significantly change.

TABLE 1. Effects of Post-filing Adjustments on Forms 1040, SOI Year 2013
[Money amounts are in millions of dollars]

Item	Amendments [1]
	Recommended tax change
Increases to tax	$2,500
Decreases to tax	$4,600
Net impact on SOI estimates [2]	$-2,100
Percent effect on SOI estimates	-0.16%

[1] Amendments refers to post-filing adjustments induced by a Form 1040-X.
[2] This row is calculated from approximately 3.2 million amended returns.
SOURCE; IRS, Statistics of Income Division, Weighted Sample, February 2021.

TABLE 2. Effects of Post-filing Adjustments on Forms 1120, SOI Year 2013
[Money amounts are in thousands of dollars]

Item	Amendments [1]	
	Number of returns	Recommended tax change
Increases to tax	1,251	$30,397
Decreases to tax	1,778	$118,200
Net impact on SOI estimates [2]	3,029	$-87,803
Percent effect on SOI estimates	0.16%	-0.03%

[1] Study includes returns initially filed with Form 1120.
[2] Returns with no change to tax are not included.
SOURCE; IRS, Statistics of Income Division, Weighted Sample, February 2021.

Discussion

Study challenges and limitations

There are a number of legal requirements, data quality issues, and other study limitations that may affect the quality of the results. This section discusses some of these limitations, but is not exhaustive, and focuses on the adjustments to corporate income tax returns.

Legislative and regulatory changes in years after 2013 could affect adjustments to 2013 returns if later changes in tax liability are carried back to 2013. In such cases, a corporation may request a Tentative Allowance of a carryback of a credit or loss within the statutory period of limitation. Typically, corporations use Corporation Application for Tentative Refund (Form 1139) for this purpose.[12] There are some cases in which the taxpayer prefers to file Form 1120-X over Form 1139. Although some Forms 1120-X might include both a carryback adjustment and a general tax adjustment claim, only a general tax adjustment is within the scope of this study. Thus, the carryback adjustments are not included in the study calculations.

At this time, some data posted to the IRS Master File do not clearly distinguish recommended changes from paid amounts. In the tables used as part of this study, the posted tax change is a recommendation, not a paid amount. In some postings, the information posted is part of an installment payment (and does not reflect

[12] Corporations should file Form 1120-X or another amended return instead of Form 1139 when carrying back to years that have a section 965(a) inclusion ("965 year"); a prior year minimum tax credit released due to a net operating loss (NOL), or net capital loss carryback; a prior year foreign tax credit released due to an NOL or net capital loss carryback; a prior year general business credit released because of the release of the foreign tax credit; or an NOL for a year for which a corporation has a 965(a) inclusion.

the full amount). In processing, there may be tax adjustments which are posted in multiple installments (i.e., a partial tax adjustment is posted and later the remainder of the tax adjustment is posted).

Large changes after filing could result in a sample that is no longer weighted correctly to reflect the total population. Any large post-filing season changes could affect the value of total assets or other amounts used in the weighting scheme. Therefore, it is important to identify alterations to the size of total assets, income, or gross receipts that could trigger a subsequent assignment of the return to a higher or lower sample class. For example, a reclassification to a higher sample class may result in the movement of returns from noncertainty strata to certainty strata, which initiates the process of recalculating the weights.

For this study, weights were not recalculated after amendments.[13] Instead, the weighted sample data were compared to the true population. This verified that the weights were appropriate for the study and that the strata were not significantly changed. Future studies should verify weights and consider recalculating them if the differences are significant. One impediment to recalculating weights is that there may be errors in the Master File data. SOI study weights are based on SOI data, which go through editing and quality review, but data posted to the IRS Master File are as reported by the taxpayer.

The Master File data used in this study were aggregated net total tax change amounts. This study does not look at individual record-level and/or item-level tax differences between the matched sample and amended return record. Making such a comparison might produce a different adjustment estimate because of the changes that are made as part of the SOI editing process to the original return. Additional resources would be required to edit returns and follow the SOI study processes. In the meantime, this study can be used as a way to estimate and then to determine if record-level or item-level detailed research is necessary for a given year.

There are additional ways to amend Form 1120, but they are outside the scope of this study. The corporate part of this study only collected data from Form 1120-X to estimate amendments to Form 1120. Future studies may consider additional ways of filing an amendment.

[13] All corporate dollar values and counts presented in this paper are population estimates that have been weighted in accordance with the SOI 2013 corporate sample weighting scheme based in part on size of total assets.

Appendix

Sampling details

All dollar values and counts presented in the Results section of this paper for individual and corporate returns are population estimates that have been weighted in accordance with the sample weighting scheme for the 2013 Complete Reports for individual and corporate returns.

Individual sample

Individual income tax statistics for this article are based on a sample of individual income tax returns (Forms 1040, 1040A, and 1040EZ, including electronically filed returns) filed for Tax Year 2013. SOI stratified the returns in the sample based on the (1) larger of positive income or negative income (absolute value); (2) size of business and farm receipts; (3) presence or absence of specific forms or schedules; and (4) usefulness of returns for tax policy modeling purposes. SOI then selected returns at rates ranging from 0.10 percent to 100 percent.

Corporate sample

Corporate income tax statistics for this article are based on a sample of corporate income tax returns (Forms 1120) filed during SOI Year 2013. SOI stratified the returns in the sample based on (1) the size of the total assets, and (2) the size of the "proceeds," which is the measure of income for this form. SOI then selected returns for Form 1120 at rates ranging from 0.40 percent to 100 percent.[14]

Definitions

Amended return—A return that was amended using Form 1040-X or 1120-X.

Form 1040-X—A form filed to change a previously filed Form 1040 or Form 1040-X. It must be filed on paper within 3 years of the original return's filing date or within 2 years of the tax being paid, whichever is later. Special exemptions for the statute of limitations exist for net operating loss, loss carrybacks, foreign tax credits, and bad debt.

Form 1120-X—A form filed to change a previously filed corporate return. It can be either e-filed or paper filed within 3 years of the original return's filing date or within 2 years of the tax being paid, whichever is later. In the event of bad debt or a worthless security, the statute of limitations is extended to 6 years.

Publication 16—Corporation Income Tax Returns Complete Report—The Corporation Complete Report publication contains complete corporate income tax data. The statistics are based on a sample of corporate income tax returns selected before examination. The report contains data on the number of returns, total assets, total receipts, and net income (less deficit), income subject to tax, total income tax before credits, and total income tax after credits for active corporations.

Publication 1304—Individual Income Tax Returns Complete Report—The Individual Complete Report publication contains complete individual income tax data. The statistics are based on a sample of individual income tax returns, selected before examination, which represents the population of Forms 1040, 1040-A, and 1040-EZ, including electronic returns filed for a given tax year. The report contains data on sources of income, adjusted gross income, exemptions, deductions, taxable income, income tax, modified taxable income, tax credits, self-employment tax, and tax payments.

Publication 6186—Calendar Year Return Projections for the United States and IRS Campuses—This publication provides the total number of returns filed historically and future projections of returns to be filed, ordered by form and type of return.

[14] See 2013 Corporation Income Tax Returns Complete Report, Section 3: Description of the Sample and Limitations of the Data. "Corporation Returns: Number Filed, Number in Sample, and Sampling Rates, by Selection Class," https://www.irs.gov/pub/irs-soi/13coccr.pdf.

Acknowledgments

The authors would like to thank David Paris, Melanie Patrick, Mike Strudler, Laura Rasmussen, Jeff Matsuo, Kathleen A. MacDonald, Bobby Hodges, Kevin Pierce, Victoria L. Bryant, Heather Parisi, and many others for their support and insightful feedback.

This article was written by Derrick Dennis, formerly an operations research analyst with the Servicewide Support Section, under the direction of Timothy Castle, Chief; Jennifer Ferris, formerly an economist with the Individual and Tax Exempt Special Studies Section, under the direction of Michael Weber, Acting Chief; Chloe Gagin, an economist with the International Returns Analysis Section, under the direction of Chris Carson, former Chief (now retired); Tuba Ozer-Gurbuz, an economist with the Partnership & Special Projects Section, under the direction of Joe Koshansky, Chief; Julia Shiller, an economist with the Corporation Returns Analysis Section, under the direction of Martha Harris, Chief; and Christopher Williams, an economist with the Individual Returns Processing and Data Perfection Section, under the direction of Scott Hollenbeck, Acting Chief. This article was written under the direction of Barry Johnson, Statistics of Income Division Director and Acting IRS Chief Research and Analytics Officer, RAAS.

3

∇

Developments in Technology and Analytics

Hertz ♦ Langetieg ♦ Payne ♦ Plumley ♦ Jones

Millard ♦ Whitfield ♦ Smolenski ♦ Stavrianos ♦ Szczerbinski

Yismaw ♦ Johns ♦ Hussain ♦ Creem ♦ Risler

Estimating the Extent of Individual Income Tax Nonfiling[1]

Thomas Hertz, Pat Langetieg, Mark Payne, and Alan Plumley (IRS, Research, Applied Analytics, and Statistics), and Marggie R. Jones (U.S. Census Bureau)

1. Introduction

The Internal Revenue Service (IRS) has been estimating the extent of individual income tax nonfiling for many years—both the Voluntary Filing Rate (VFR, the percentage of required returns filed on time) and the nonfiling tax gap (the tax not paid on time by those who do not file on time or at all). To estimate the extent to which an individual contributes to the individual income tax nonfiling tax gap, we need accurate information about his income and family situation. Family information, such as marital status, the age of the taxpayer(s), and the number and ages of dependents, influence key determinants of tax liability such as the standard deduction, the applicable marginal tax rate schedule, and common tax deductions and credits. The IRS receives income information (and reports on certain deductible amounts) from third parties. The IRS also receives the date of birth associated with each Social Security Number (SSN) from the Social Security Administration. However, unless someone files a tax return, the IRS does not have information about a person's marital status or number of dependents for any given tax year.

One way to obtain information about nonfilers is to conduct a comprehensive enforcement program focused on this group (Erard and Ho (2001)), but such an approach has its own estimating disadvantages. As an alternative, we had previously gained some demographic information on nonfilers by incorporating Annual Social and Economic Supplement of the Current Population Survey (CPS-ASEC) data into our estimates (Erard *et al.* (2012)), taking advantage of the limited tax administrative data for the entire population of individual tax records that the IRS routinely provides to Census under Internal Revenue Code section 6103(j).[2] The Census survey data could be linked to the IRS administrative data because both sets of records were assigned valid unique identifiers, known as Protected Identification Keys (PIKs) (see Wagner and Layne (2012)). Aside from the demographic information, a major advantage of this matched "(j)" data was the ability to identify which CPS-ASEC respondents were linked to a filed tax return and which were not. However, as summarized in Table 1, this approach also has drawbacks—the primary one being that the income self-reported on Census surveys is often much less than what is reported to the IRS by third parties for those persons. To gain a more complete picture of the income of those who did not file, we imputed additional income to the Census records (Erard *et al.* (2014)).

The IRS took this "Census Method" approach when estimating the individual income tax nonfiling gap for Tax Years 2008–2010 (Langetieg *et al.* (2016)). However, we also produced corresponding estimates using an "Administrative Data Method," which did not make use of linked data, and then averaged the two sets of estimates due to the uncertainties associated with each method (see Table 1). That Administrative Data Method generally followed the approach of Lawrence *et al.* (2011), but applied it to population data instead of to a sample. However, the Administrative Data Method also has its limitations. First, not all income earned by taxpayers is reported on third-party information documents. And in some cases, such as capital gains and nonemployee compensation, it is difficult to translate the amount reported on third-party documents into earnings net of expenses and costs that would be reported on a tax return. Second, as mentioned above, with this method one does not have information about the types of tax units (marital status and number of dependents) to which each potential nonfiler belongs.

---

1 Most of the estimates included in this paper are derived from data protected by Title 13 and/or Title 26 of the U.S. Code. The numbers have been approved for public release by the Internal Revenue Service and the U.S. Census Bureau (approval numbers CBDRB-FY2021-CES005-003, CBDRB-FY2021-CES005-017, and CBDRB-FY2021-CES005-020). All statements in this paper are the opinions of the authors and do not necessarily represent the position of either the Census Bureau or the IRS.

2 Joint research between the IRS and U.S. Census Bureau on nonfilers is performed for statistical purposes only and is not used for enforcement.

To partially address the first limitation, we imputed net self-employment income as in the Census Method. However, it is still unknown to what extent models based on patterns observed among filers are applicable to nonfilers.

For the second challenge, potential nonfilers in the Administrative Method are combined into tax units and assigned children through an algorithmic approach that targets aggregate counts by filing status, number of dependents, and age groups resulting from the subtraction of filers from CPS-ASEC population profiles. In the end, one does not know precisely what impact this fairly unguided process of assembling tax units is having on the nonfiling tax gap or Voluntary Filing Rate estimates.

Finally, when the IRS subsequently estimated the individual income tax nonfiling tax gap for Tax Years 2011–2013, it abandoned the Census Method because even with the income imputations, the income reported on the CPS-ASEC by nonfilers still fell short of the income reported to the IRS by third parties (Internal Revenue Service (2019)).

The approach described in this paper attempts to get the best of both worlds: relying on detailed micro information on income from tax administrative data that are linked at the person level with detailed demographic data from the Census. This became possible through a special short-term IRS research project created under the authority of Internal Revenue Code section 6103(n). The records are linked using the same anonymous identification system used for the far more limited data, but the linked records now benefit from the much more detailed income information from IRS records. So, one major contribution of this paper is that—for the first time—we have been able to link detailed demographic information for a sample of individuals from Census survey records to detailed income information for the same individuals from tax records (including filed Form 1040s as well as information documents such as Form W-2 and Form 1099 for nonfilers) to generate stronger estimates of the individual income tax nonfiling tax gap.[3]

We still need to impute net self-employment income in this method because it is not fully documented by third parties. Until now, however, these imputations were based on formulas estimated from amounts reported on filed tax returns. This paper uses data from the IRS National Research Program (NRP)—a stratified random sample of returns that were selected for a full audit. Instead of using the amount *reported* by taxpayers, the imputations are now based on the values of net self-employment income as *corrected* by the auditor. While the corrected self-employment amounts are closer to the "true" earnings than what is reported on returns, significant amounts of this income remain undetected by the auditors. But we still are assuming that the self-employment income of timely filers can be used as a basis for imputing such income to similarly situated nonfilers, as well.

Because not all Census survey records could be assigned a PIK with high confidence, the usable linked sample is smaller than the full sample of people who appear in both the survey and the administrative data. Moreover, a portion of the Census survey data are fully imputed records, which we do not consider valid for our estimation purposes, further reducing the usable sample. We observe that the survey records that lacked valid PIKs were disproportionately located at the lower end of the income distribution (see Table 2), which suggests that nonfilers will be under-represented in the usable data. In the (j) data environment, the potential nonfiler population could not be identified from the IRS administrative data provided, so the weights were designed to make the portion of the CPS-ASEC with PIKs look like the full CPS-ASEC population. In contrast, the second major contribution of this paper is to apply an improved re-weighting process made possible by the (n) data, so that the linked sample of potential nonfilers is representative of the full population of potential nonfilers in the administrative data. We show that this re-weighting method is more successful in making the matched potential nonfiler sample look like the full population of potential nonfilers in the administrative data than the method used when only the (j) data were available.

After applying these innovations, we use a tax calculator to identify which potential nonfiler tax units were required to file an income tax return, and to estimate their tax liability and unpaid balance due (i.e., their contribution to the nonfiling tax gap) at the payment due date. We conclude by incorporating similar tabulations of late filers from IRS administrative records to derive a comprehensive updated estimate of the Tax Year 2010 individual income tax nonfiler tax gap.

[3] Linking the Census data to the tax data allows us to more accurately assign individuals to tax units (especially married couples), assign the appropriate tax filing status (e.g., Married filing jointly, Head of Household, or Single), and to identify dependents within the tax unit, all of which affect tax liability, and therefore our tax gap estimates. This is the major reason for linking these data sources for nonfiler analysis.

The remainder of this paper is organized as follows: section 2 describes the (n) data; section 3 describes our new method for re-weighting the linked records and how we used and supplemented the (n) data to generate counts and income estimates for all potential nonfilers in the administrative tax data; section 4 describes how we estimated a tax filing obligation among the potential nonfilers, as well as their contribution to the nonfiling tax gap; section 5 examines the distributional make-up of the nonfiling tax gap; and section 6 offers some concluding observations and priorities for future research.

A Word on Nomenclature

We use the term nonfiler to include anyone who was required to file an income tax return for the tax year in question (Tax Year 2010 for this paper) but did not do so by the filing deadline (as extended, if applicable). There are two types of nonfilers: late filers (who filed after their deadline, and before the end of a specified calendar year—2011 for this paper), and not-filers (who did not file at all by December 31, 2011). We prefer to set the date to distinguish between late filers and not-filers as 4 years after the year in which the return was originally due (because it's helpful to base our estimates on actual tax returns), but because we had only 1 year of (n) data available at Census for this paper, we were constrained to set the deadline for considering late-filed returns as the end of 2011. In reality, a significant share of the late filers file their return by the end of that first year (unless they do so in response to some enforcement action), so our tabulations of late filers for this paper still make a significant contribution to the total nonfiling tax gap.

TABLE 1. Pros and Cons of Tax Administrative Data and Census Survey Data for Estimating the Extent of Nonfiling

Data Sources	Advantages	Disadvantages	How To Mitigate Disadvantages	Remaining Weaknesses
Tax administrative data	• Micro data on all income reported by third parties • Population data	Generally lacks demographic information for nonfilers Some income is not reported or deficiently reported on third-party documents	Impute filing status and dependents to nonfilers using aggregate Census profiles; Impute net self-employment income	Demographic imputations very approximate. Not all self-employment income is detected in NRP† audits. Nonfilers may not look like filers.
Census CPS-ASEC	• Complete demographic information	Significantly understates some income types	Impute key income amounts to Census records on the basis of patterns in the tax administrative data	Does not account for overstated amounts in Census records
		No indication if taxpayer filed a tax return	Use CPS-ASEC only to estimate all required returns; subtract filers using IRS data	Mixes aggregates from two incompatible definitions
CPS-ASEC linked to 6103(j) tax data*	• Can identify nonfilers • Have some tax-related income values	Not all Census records can be assigned PIKs, so linked sample not representative of all nonfilers	Re-weight the PIKed subset to represent the full CPS-ASEC population	The full CPS-ASEC population doesn't represent well the population of nonfilers in the administrative data
		Very limited information on income amounts from (j) data, so re-weighting is very constrained	Use wages from W-2s and use 1099s to impute pension, social security, unemployment compensation, and self-employment income to Census records on the basis of patterns in the tax administrative data	For the self-employment imputation, the characteristics of filers may not necessarily represent the income of nonfilers well
CPS-ASEC linked to 6103(n) tax data**	• Complete tax-related income values • Can estimate new weights to reflect income distribution of nonfilers in the administrative data	Self-employment income is generally not reported by third parties	Impute this income on the basis of patterns among filers as corrected under audit and limited third-party self-employment income reporting	Imputations based on the characteristics of filers may not necessarily represent the income of nonfilers well.
		Sometimes one spouse can be linked and the other one not	Use the Census income of the nonlinked spouse	May ignore other problems with the Census income of the nonlinked spouse
		Income reported to Census does not always agree with income in the tax data	Except for nonlinked records, use only the income from the tax administrative data (pending further study)	May ignore some income correctly reported to Census but not to the IRS

* Limited micro information for the entire population of individual tax records that the IRS routinely provides to Census under Internal Revenue Code section 6103(j). The Census records can be anonymously linked to these tax records using a unique identification number applied to both sets of data by the Census Bureau.

** Greatly expanded income and other information for the entire population of individual tax records provided to Census for a short-term IRS research project under Internal Revenue Code section 6103(n).

† The IRS National Research Program (NRP) conducts audits on stratified random samples of individual tax returns, documenting all adjustments made by the auditors.

2. The 6103(n) Data

Section 6103 of the Internal Revenue Code protects the confidentiality of tax records. Sub-section (a) describes the General Rule: "Returns and return information shall be confidential, and except as authorized by this title [no one] shall disclose any return or return information obtained by him in any manner in connection with his service as such an officer or an employee or otherwise or under the provisions of this section."

Sub-section (j) provides for the statistical use of federal tax records by the Department of Commerce:

> Upon request in writing by the Secretary of Commerce, the Secretary [of the Treasury] shall furnish—
>
> (A) such returns, or return information reflected thereon, to officers and employees of the Bureau of the Census, and
>
> (B) such return information reflected on returns of corporations to officers and employees of the Bureau of Economic Analysis, as the Secretary may prescribe by regulation for the purpose of, but only to the extent necessary in, the structuring of censuses and national economic accounts and conducting related statistical activities authorized by law.

The IRS regularly provides limited data to the Census Bureau under this sub-section per detailed regulations.

Sub-section (n) further authorizes the release (e.g., to the Census Bureau) of protected tax information for the purposes of tax administration specifically:

> Pursuant to regulations prescribed by the Secretary [of the Treasury], returns and return information may be disclosed to any person, including any person described in section 7513(a), to the extent necessary in connection with the processing, storage, transmission, and reproduction of such returns and return information, the programming, maintenance, repair, testing, and procurement of equipment, and the providing of other services, for purposes of tax administration.

Because estimating the extent and drivers of income tax nonfiling furthers tax administration, and because the data available under 6103(j) are inadequate for that purpose, the IRS entered into an agreement with the Census Bureau to transmit to Census a much more complete set of individual income tax records for a special short-term tax administration research project. This paper is one of the first studies based on these expanded (n) data.

Whereas the (j) data contained several indicators of the presence of certain types of income and very few income amounts, the (n) data include the amounts reported on income tax returns as well as the amounts reported on third-party information returns for virtually every type of income reported on such returns to the IRS. It is these latter amounts that allow us to estimate the filing obligations, tax liabilities, and balance due (or refund) for each person not represented on a timely filed tax return. Moreover, this level of income detail also allowed us to re-weight the linked records not connected with a filed return so that they represent not the original CPS-ASEC population, but rather the population of potential nonfilers among the tax records. The (n) data used for this paper pertained only to IRS Processing Year 2011 and Tax Year 2010.

Imputing Self-employment Income

Given that only a small portion of self-employment income is reported on Form 1099-MISC (miscellaneous),[4] we use regression models to impute this income to not-filers on the basis of the net self-employment income amounts that should have been reported on filed returns as corrected by NRP audits. Our imputations were restricted to the corrected amounts of net sole proprietorship income reported on Schedule Cs. We estimated the likelihood that a not-filer has self-employment earnings falling into one of the following three categories: (a) negative net self-employment earnings; (b) net self-employment earnings between $1 and $433; and (c)

[4] Payment card and third-party transaction reporting on Form 1099-K was not in place for Tax Year 2010.

net self-employment earnings in excess of $433 (since taxpayers with more than $433 in net self-employment earnings are required to file a tax return and pay self-employment tax).[5]

The econometric framework involved three separate models. The first was a probit specification for the likelihood that an individual has nonzero self-employment earnings:

$$SE^* = \gamma' x + \mu \tag{1}$$

where SE^* is a latent variable describing the propensity for net self-employment earnings to be present, x is a vector of explanatory variables, and γ is a vector of coefficients to be estimated. The explanatory variables include five age categories, region, indicators for the presence of key income types as represented on third-party information documents (wages, interest, dividends, taxable state and local tax refunds, nonemployee compensation, capital gains, pensions, Schedule E, Social Security, unemployment compensation, and other income), and the log of each of these income amounts. The error term μ is assumed to follow the standard normal distribution. Estimation of this model permits us to develop a prediction equation for the unconditional likelihood that an individual has net income from self-employment. Each individual was assigned a random number from a uniform distribution, and if the value of this number was below the predicted probability then the person was determined to have some net self-employment income.

Our second model was an ordered probit specification for the dollar amount category that net self-employment earnings fall into when they are present (negative, $1 to $433, or over $433):

$$I^*_{SE} = \delta' x + v \tag{2}$$

where I^*_{SE} is a latent variable for the propensity for net self-employment earnings to fall into one of these categories, x is the same set of explanatory variables used in the probit model, δ is a coefficient vector to be estimated, and v is a standard normal random disturbance. The model also includes a limit parameter l to be estimated.[6] The indicator I_{SE} for the net self-employment earnings category is assigned as follows:

$$I_{SE} = \begin{cases} 1 & net \;\; earnings \; < \; \$0 \\ 2 & \$0 \; < \; net \;\; earnings \; \leq \; \$433 \\ 3 & net \;\; earnings \; > \; \$433. \end{cases} \tag{3}$$

Our third model is a regression specification for the magnitude of net self-employment earnings when they exceed $433. Our specification is:

$$\ln(SE) = \beta' x + \varepsilon, \tag{4}$$

where ln(SE) represents the natural log of net self-employment earnings, x is the same set of explanatory variables used in the preceding models, β' is a vector of coefficients to be estimated, and ε is assumed to be a normal random error term with mean zero and standard deviation σ. Under this specification, the distribution of self-employment earnings is assumed to be log normal. Each individual is assigned a random number from a normal distribution, which is multiplied by the root mean squared error and added to the predicted log amount. Furthermore, we have imposed the constraint that the imputed self-employment income amount cannot exceed the amount corresponding to the 99.99th percentile of self-employment income on filed returns.

Re-weighting the Linked Records To Match the Full Administrative Population

For the re-weighting scheme centered on the administrative (n) data, our target population is the set of all individuals for whom the IRS received information documents[7] indicating economic activity for Tax Year 2010, but who do not appear on a timely filed Form 1040 by the end of 2011, as either a primary filer or a spouse on a married-filing-jointly return. These are potential nonfilers; they are "potential" because they might not have a filing requirement (or a tax liability) due to having income that is less than the relevant filing thresholds.

[5] Schedule SE indicates that if net profit times 92.35 percent is less than $400, no self-employment tax is due.

[6] This parameter serves as a threshold for separating the various levels of the response variable.

[7] The information returns considered were Forms W-2, W-2G, 1099-R, 1099-SSA, 1099-INT, 1099-DIV, 1099-MISC, 1099-LTC, 1099-PATR, 1099-Q, 1099-G, 1099-C, 1099-OID, 1099-S, 1041-K1, 1120S-K1, 1065-K1, 5498, 5498-SA, 1098, 1098-E, and 1098-T.

After limiting to records with valid Taxpayer Identification Numbers (TINs, essentially, valid Social Security numbers) and a PIK, the target population is about 54.8 million. This is necessary because without the assurance that the TINs are valid, and belong to a unique individual, it is not possible to ascertain whether the information document in question can be linked to an individual in an unambiguous way. This does, however, imply that some people who are working under invalid Social Security numbers, including some foreign-born workers not lawfully present in the United States, will be dropped from our count of nonfilers. This imparts a downward (conservative) bias to our estimates.

We next require the TIN to be in the master list of individuals provided to the IRS by the Social Security Administration, and the payee on the form to have a birthdate after 1900 (allowing them to be up to 110 years old in 2010) and no data of death prior to 2010. These restrictions further reduce the potential nonfiler population to about 50.7 million.[8]

To see why it is necessary to re-weight the linked records using the income information available in the (n) data, consider the information tabulated in Table 2. Column A reports the population counts by personal income decile in the 2011 March CPS-ASEC (representing Tax Year 2010), for people of all ages using the standard CPS-ASEC sample weights. In column B, we see that, because the bottom 30 percent of the population has no positive income, the bottom three deciles collapse into one, and the fourth decile is slightly underpopulated so as to ensure that 40 percent of the population is indeed in the bottom four deciles. Thereafter, each decile contains 10 percent of the population, as expected.

In column C, we eliminate the CPS-ASEC records with income amounts that are fully imputed, leaving us with 87.6 percent of the sample. This share varies by decile between 87.2 percent at the top to 88.4 percent at the bottom. The percentage of these that have valid PIKs (column D), however, is higher at the top of the distribution (94.8 percent) than at the bottom (88.1 percent). As a result of these two factors combined, the share of records that are usable is somewhat higher at the top (82.6 percent) than at the bottom (77.9 percent). This is visible in column E, as well as in column F, which reports the decile distribution of the usable data, which is very slightly top heavy.

Next, we link the CPS-ASEC and the tax data. Column G shows that, as expected, relatively few individuals (27.5 percent) in the bottom 30 percent of the personal income distribution as measured by the CPS-ASEC are found in the IRS data, whether on a tax return or on an information document. Overall, 75.9 percent of the usable cases are found in the IRS data. Note that we do not expect this number to be 100 percent; it should reflect the share of the full population that is economically active, or otherwise appears, as a filer or spouse, on a tax return.

In columns H through J, we use the new set of weights, which correct for the share of the CPS-ASEC data that is fully imputed or lacks a PIK. These weights rebalance the usable sample to have the same shares as the full CPS-ASEC in all the survey's sample strata, and they adjust the total population to match that of the full CPS-ASEC. We now see that the CPS-ASEC represents approximately 231 million people on third-party information documents, of whom 189 million are on a Form 1040 and about 42 million are not. This indicates that even after re-weighting the data to preserve the CPS-ASEC stratification proportions, potential nonfilers are under-represented in the usable CPS-ASEC dataset and remain well short of our target of 50.7 million.

The final column presents results that are re-weighted to represent our target population size, taking full advantage of the detailed income data available in the (n) data. The weights were constructed using a probit model run on the full administrative population of potential nonfilers, with an outcome variable equal to 1 for records that were linked with a respondent in the CPS-ASEC, and 0 otherwise. The predictors in the model

[8] This could potentially be overcome through an alternative to the re-weighting methodology we developed for this paper. The alternative approach would create cells in both the IRS and CPS-ASEC datasets that are defined by the control variables used in our econometric method. The weight for any given cell would then simply be the ratio of the number of individuals in that cell in the IRS data to the number in the CPS-ASEC. This wouldn't require the IRS records to have a valid PIK.

include age, gender, and the presence and decile locations (with respect to the full potential nonfiler population in the IRS administrative data) for each of the following types of income: wages, interest, dividends, capital gains, the imputed amount of net self-employment income, farm, unemployment compensation, Social Security, pension, rents and royalties, tax refunds, and other. This was found to be the preferred specification for replicating the aggregate counts and amounts of income in the full population of potential nonfilers as found in the administrative data. Finally, each record in the linked data is assigned a weight equal to the inverse of the predicted probability of being linked. The overall average of that weight would be the total number of potential returns in the (n) data divided by the number of linked records from the CPS-ASEC data, but the weight for any given linked record would be somewhat higher or lower than that average based on the values of the predictor variables for that record.

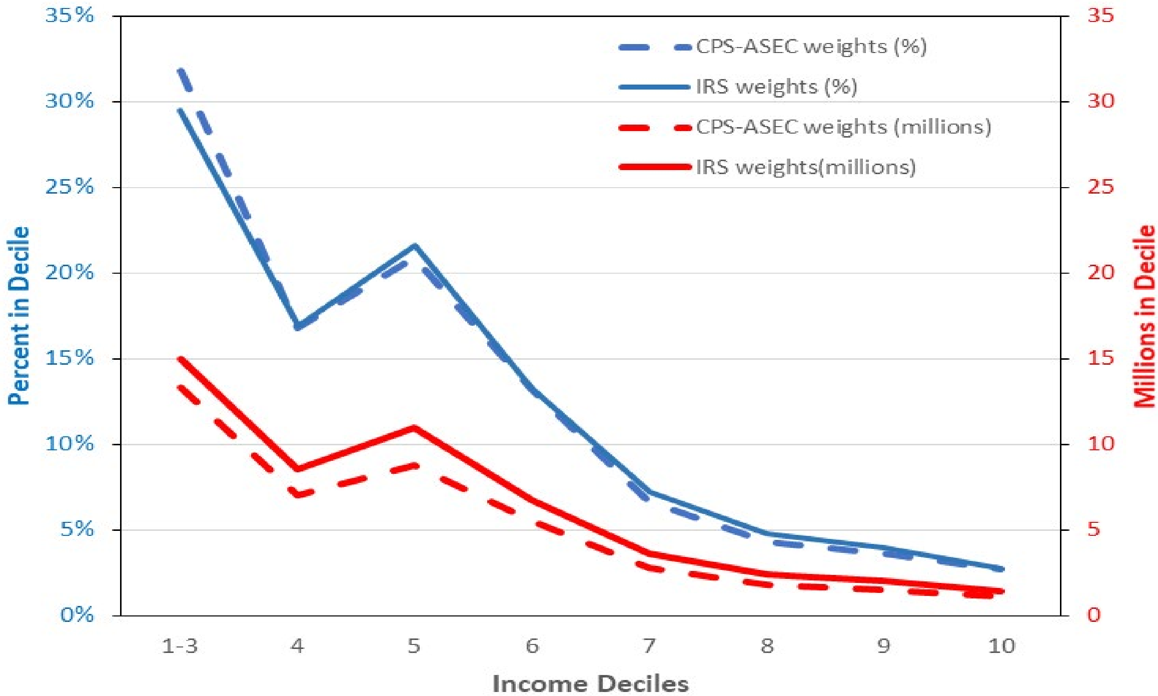
TABLE 2. The Extent to Which CPS-ASEC Records Linked to IRS Administrative Data Reflect the Distribution of the Population Across Income Deciles Using Alternative Weights, Tax Year 2010

Decile*	Full CPS-ASEC Population (Millions)	Full CPS-ASEC Decile Share	Percent Not Fully Imputed	Percent PIKed	Percent Usable (C x D)	Usable CPS-ASEC Decile Shares	Percent of Records Found in IRS Data	Shares in Linked Data			
								Adjusted CPS-ASEC weights			IRS re-weights
								All	On a 1040	Not on a 1040	Not on a 1040
	A	B	C	D	E	F	G	H	I	J	K
1-3	95.0	31.0%	88.4%	88.1%	77.9%	30.2%	27.5%	11.1%	6.5%	31.8%	29.5%
4	27.4	9.0%	86.4%	91.4%	78.9%	8.8%	89.4%	10.5%	9.1%	16.8%	16.9%
5	30.6	10.0%	87.6%	91.7%	80.3%	10.0%	92.2%	12.2%	10.3%	20.9%	21.6%
6	30.6	10.0%	87.4%	91.0%	79.5%	9.9%	97.7%	12.8%	12.7%	13.2%	13.3%
7	30.6	10.0%	87.3%	92.3%	80.5%	10.1%	98.7%	13.1%	14.6%	6.7%	7.2%
8	30.6	10.0%	87.3%	93.7%	81.8%	10.2%	99.5%	13.4%	15.4%	4.3%	4.8%
9	30.6	10.0%	87.5%	94.0%	82.2%	10.3%	99.7%	13.4%	15.6%	3.6%	4.0%
10	30.6	10.0%	87.2%	94.8%	82.6%	10.3%	99.7%	13.4%	15.8%	2.7%	2.8%
All	306.1	100.0%	87.6%	91.3%	80.0%	100.0%	75.9%	100.0%	100.0%	100.0%	100.0%
Weighted Population Counts (Millions)								230.9	188.9	42.0	50.8

* Deciles of Total CPS-ASEC Income. Census Bureau Disclosure Review Board release authorization number CBDRB-FY2021-CES005-020.

Column K of Table 2 shows that these weights bring us very close to the target population size of 50.7 million (by construction), and that they alter the income distribution of the linked dataset somewhat. In particular, the share of individuals in the bottom three deciles (who have negative or zero income) falls by about 2 percentage points. Still, the similarity of the shares in columns J and K implies that re-weighting the data to match the IRS administrative income data (along many dimensions) does not have a dramatic effect on the income distribution (as measured by CPS-ASEC). But given that the distribution across deciles is similar yet the projected population total in column K is significantly larger than in column J, the most important consequence of re-weighting is to make up for the fact that the potential nonfilers linked to CPS-ASEC respondents appear to be systematically under-represented within each income decile of the usable linked data. That is, the new weights help to correct the overall reported income in the CPS-ASEC (not just at the low end) (see Figure 1).

FIGURE 1. Distribution of Potential Nonfilers Across Deciles Using Alternate Weights, Tax Year 2010



SOURCE: Analysis of CPS-ASEC records linked to IRS administrative data.

3. Potential Nonfilers

Having assigned a new weight to each record in the linked sample, we can easily tabulate a variety of income-related statistics for the population of individuals who are potential nonfilers. In Table 3A, we show the counts of potential nonfilers with different types of income and the sum of the amounts of income for each income type using different weights and different sources for income information. Table 3B shows the percentage of potential nonfilers with the different types of income and the mean amounts following the same column ordering as in Table 3A.

Column A shows the counts and amounts for the full population of potential nonfilers using income found on third-party documents, except that net self-employment income is imputed on the basis of formulas developed using amounts on filed returns as corrected by National Research Program audits.[9] In terms of income, this is the administrative data baseline that we would like to replicate using the linked sample (which is necessary in order to obtain greater micro-level accuracy in the creation of tax units and in the assignment of dependents).

Columns B through E re-weight the observations using the inverse probability weights (adjusted CPS-ASEC weights) developed using a probit model incorporating CPS-ASEC strata variables. Column B uses only the income amounts reported by CPS-ASEC survey respondents. Column C starts with these responses but adds imputations for Social Security, pension, and unemployment compensation using formulas estimated on the basis of amounts on Form 1099s, and for net self-employment income using formulas estimated on the basis of amounts reported on filed Schedule Cs.[10] Since this was available in the (j) data, this column also uses

[9] Remember that this population excludes those for whom a valid PIK could not be assigned to the taxpayer identified on the third-party documents.

[10] This assumes that the CPS-ASEC income amounts can be used in lieu of the amounts reported on third-party information returns, which are not generally present in the (j) data. Also, this imputation was an older one, based on amounts reported on filed returns, not amounts as corrected by NRP audits; however, this doesn't change the basic conclusions that can be drawn from the table.

Form W-2 wage income instead of wages reported by CPS-ASEC respondents.[11] Column D shows the counts and amounts based only on the detailed income reported on third-party documents, which is included in the (n) data. Column E is the same as column D, except that self-employment income uses the imputed amount[12] instead of just the amount reported on Form 1099-MISC.

Columns F through H use the weights described in the previous section, which are directly aimed at re-weighting to the potential nonfiler population identified through third-party documents in the (n) data. Column F reports the counts and amounts of income using the same combination of CPS-ASEC reported numbers, imputations, and Form W-2 wages as found in column C. The numbers in column G are based on third-party documents exclusively, while those in column H include the imputed amounts for net self-employment income.

As shown in the previous section, the new weights based on the (n) data are better at matching the counts of the full potential nonfiler population found in the IRS administrative data. While the adjusted CPS-ASEC weights result in a population of 41.5 million, the IRS weights lead to a population of 50.8 million, which (by design) closely matches the full population in column A. The counts of the total numbers of individuals with each income type are also closer to the population counts based on IRS administrative data when using these weights. For instance, while there are 11.8 million persons with Form W-2 wages when the adjusted CPS-ASEC weights are used, these counts grow to 15.8 million when using the IRS weights, closely mirroring the corresponding count in column A. The incidence of interest, Social Security, pension, unemployment compensation, and self-employment income is also brought up to that of the full population when the IRS weights are used (see Figure 2).

Most importantly, the total income of the potential nonfilers is considerably closer to that of the IRS potential nonfiler population (column A) when the new weights are used (Figure 3). Column H, which most closely replicates the income concepts used in the population column (column A) and uses the new IRS weights, matches the population the best overall with $656.2 billion compared to $645.1 billion. This income source and weighting combination also results in similar amounts for each type of income. Only for capital gains is the percentage difference substantial, but this type of income is fairly uncommon among potential nonfilers, and is poorly accounted for anyway in third-party reporting, particularly in 2010. Considering the IRS administrative data with the self-employment imputation, the new weights add about $132 billion to the total income of potential nonfilers compared with the adjusted CPS-ASEC weights (columns E and H).

The larger amount of total income is obtained using the IRS weights, in part because of the larger number of potential nonfilers with various forms of income. The estimated number of potential nonfilers with significant income types increased by between 20 percent and 45 percent when the IRS weights were used, translating to an increased number of potential nonfilers for a given income type ranging from very few (e.g., farm income) to almost 4 million (wages). The average amounts increased significantly for wages and unemployment compensation (10 percent and 14 percent).

[11] Note that this approach differs from our previous method for estimating the nonfiler portion of the tax gap; that earlier method used the larger of the amount of wages reported on Forms W-2 compared with the amount of wages reported to Census by the respondent, which probably included some nonwage income reported as wage income.

[12] This imputation was based on NRP adjustments.

TABLE 3A. Estimates of the Number and Income of Potential Nonfilers (Person Level) Using Alternative Weights and Income Sources, Tax Year 2010

Data source:	Administrative population	CPS-ASEC linked to IRS administrative data for potential nonfilers						
Weights:	None needed	Adjusted CPS-ASEC weights				IRS re-weights		
Income source:	Administrative (n) data with SE imputation	Raw CPS-ASEC	CPS-ASEC, with imputations using (j) data	Raw administrative (n) data	Administrative (n) data with SE imputation	CPS-ASEC, with imputations using (j) data	Raw administrative (n) data	Administrative (n) data with SE imputation
Type of income	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Aggregate Counts (Millions)								
Total population	50.70	41.53	41.57	41.53	41.53	50.81	50.76	50.76
Wages	15.67	9.63	11.80	11.79	11.79	15.86	15.84	15.84
Interest	11.92	6.50	6.51	10.54	10.54	8.06	11.91	11.91
Dividends	5.60	1.42	1.42	5.18	5.18	1.71	5.59	5.59
Capital gains	0.58	N/A	N/A	0.57	0.57	N/A	0.58	0.58
Pensions	6.96	2.05	6.20	5.65	5.65	7.79	7.11	7.11
Social Security	21.93	13.48	14.26	18.60	18.60	17.52	22.11	22.11
Self-employment	4.85	1.20	4.71	1.86	3.71	6.09	2.56	5.01
Positive	4.27	1.13	3.56	1.86	3.25	4.61	2.56	4.43
Negative	0.58	0.07	1.15	N/A	0.46	1.48	N/A	0.58
Unemployment compensation	2.78	1.34	2.29	2.06	2.06	3.01	2.90	2.90
Schedule E	0.75	0.58	0.58	0.66	0.66	0.83	0.79	0.79
Farm	0.06	0.13	0.13	0.08	0.08	0.18	0.11	0.11
Alimony	N/A	0.04	0.04	N/A	N/A	0.06	N/A	N/A
Other income	1.82	N/A	N/A	1.44	1.44	N/A	1.83	1.83
Aggregate Amounts ($ Billions)								
Total income	$645.1	$488.2	$504.7	$488.1	$524.6	$649.1	$616.9	$656.2
Wages	$247.0	$250.5	$184.0	$183.5	$183.5	$246.7	$245.9	$245.9
Interest	$3.7	$7.1	$7.1	$3.2	$3.2	$8.9	$4.0	$4.0
Dividends	$3.0	$4.1	$4.0	$2.4	$2.4	$4.7	$3.1	$3.1
Capital gains	$0.7	N/A	N/A	$0.3	$0.3	N/A	$1.0	$1.0
Pensions	$54.7	$20.5	$47.6	$46.0	$46.0	$63.4	$58.4	$58.4
Social Security	$226.5	$158.1	$183.5	$198.7	$198.7	$220.2	$229.7	$229.7
Self-employment	$75.5	$31.6	$55.1	$27.2	$63.7	$72.9	$38.6	$77.9
Positive	$78.4	$32.0	$65.4	$27.2	$65.2	$87.0	$38.6	$80.5
Negative	-$2.9	-$0.4	-$10.3	N/A	-$1.6	-$14.0	N/A	-$2.6
Unemployment compensation	$23.4	$11.3	$18.4	$17.4	$17.4	$24.2	$24.2	$24.2
Schedule E	$10.0	$3.9	$3.9	$9.1	$9.1	$6.0	$11.4	$11.4
Farm	$0.6	$0.8	$0.8	$0.3	$0.3	$1.5	$0.6	$0.6
Alimony	N/A	$0.3	$0.3	N/A	N/A	$0.6	N/A	N/A
Other income	$20.7	N/A	N/A	$16.0	$16.0	N/A	$22.0	$22.0

N/A—Not Applicable.
Census Bureau Disclosure Review Board release authorization numbers CBDRB-FY2021-CES005-003 and CBDRB-FY2021-CES005-020.

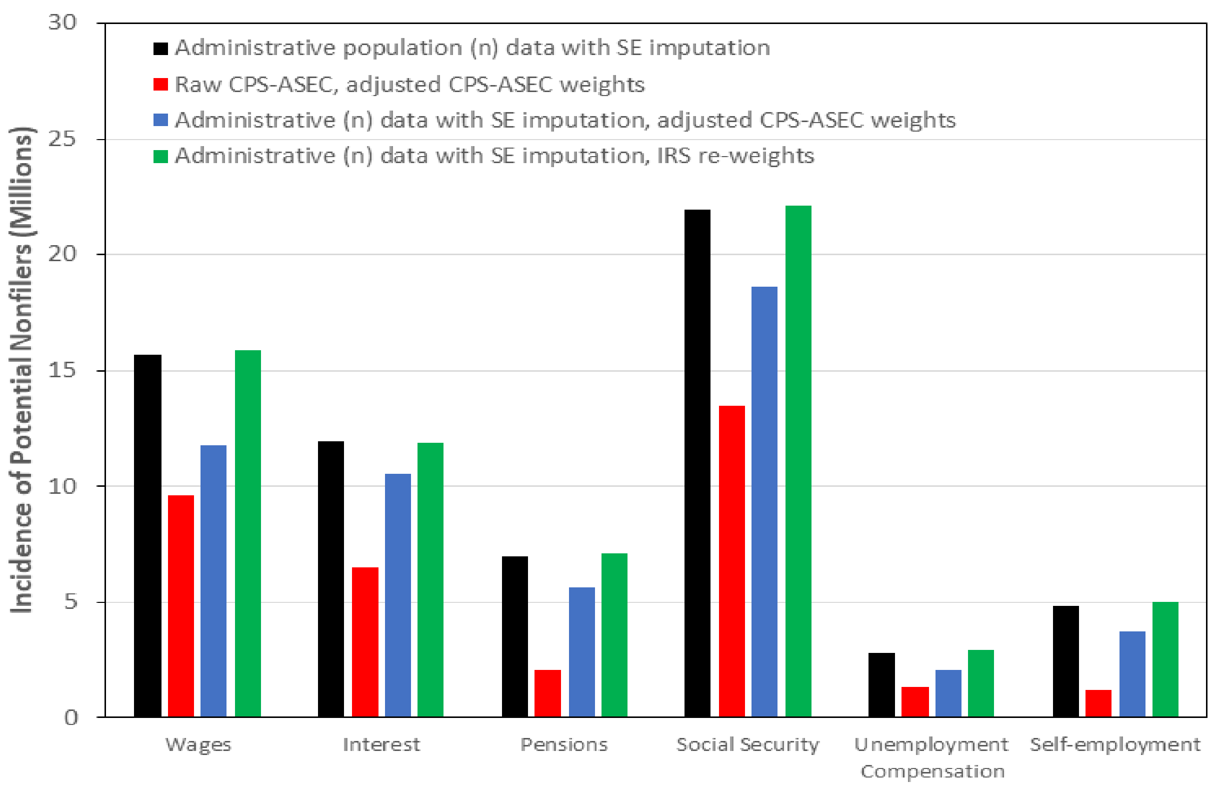
TABLE 3B. Estimates of the Incidence and Average Amount of Income of Potential Nonfilers (Person Level) Using Alternative Weights and Income Sources, Tax Year 2010

Data source:	Administrative population	CPS-ASEC linked to IRS administrative data for potential nonfilers						
Weights:	None needed	Adjusted CPS-ASEC weights				IRS re-weights		
Income source:	Administrative (n) data with SE imputation	Raw CPS-ASEC	CPS-ASEC, with imputations using (j) data	Raw administrative (n) data	Administrative (n) data with SE imputation	CPS-ASEC, with imputations using (j) data	Raw administrative (n) data	Administrative (n) data with SE imputation
Type of income	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Incidence								
Wages	30.9%	23.9%	28.4%	28.4%	28.4%	31.2%	31.2%	31.2%
Interest	23.5%	15.9%	15.7%	25.4%	25.4%	15.9%	23.5%	23.5%
Dividends	11.0%	3.5%	3.4%	12.5%	12.5%	3.4%	11.0%	11.0%
Capital gains	1.1%	N/A	N/A	1.4%	1.4%	N/A	1.1%	1.1%
Pensions	13.7%	5.2%	14.9%	13.6%	13.6%	15.3%	14.0%	14.0%
Social Security	43.3%	34.1%	34.3%	44.8%	44.8%	34.5%	43.6%	43.6%
Self-employment	9.6%	2.8%	11.3%	4.5%	8.9%	12.0%	5.0%	9.9%
Positive	8.7%	2.7%	8.6%	4.5%	7.8%	9.1%	5.0%	8.7%
Negative	1.2%	0.2%	2.8%	N/A	1.1%	2.9%	N/A	1.2%
Unemployment compensation	5.5%	3.4%	5.5%	5.0%	5.0%	5.9%	5.7%	5.7%
Schedule E	1.5%	1.4%	1.4%	1.6%	1.6%	1.6%	1.6%	1.6%
Farm	0.1%	0.3%	0.3%	0.2%	0.2%	0.4%	0.2%	0.2%
Alimony	N/A	0.1%	0.1%	N/A	N/A	0.1%	N/A	N/A
Other income	3.6%	N/A	N/A	3.5%	3.5%	N/A	3.6%	3.6%
Average Amounts ($)								
Wages	$4,871	$6,302	$4,426	$4,418	$4,418	$4,855	$4,845	$4,845
Interest	$73	$175	$171	$78	$78	$175	$80	$80
Dividends	$60	$101	$97	$58	$58	$92	$60	$60
Capital gains	$13	N/A	N/A	$7	$7	N/A	$19	$19
Pensions	$1,080	$522	$1,145	$1,108	$1,108	$1,248	$1,150	$1,150
Social Security	$4,468	$3,997	$4,413	$4,785	$4,785	$4,334	$4,525	$4,525
Self-employment	$1,489	$737	$1,326	$655	$1,533	$1,435	$761	$1,535
Positive	$1,547	$746	$1,573	$655	$1,571	$1,712	N/A	$1,585
Negative	-$58	-$9	-$247	N/A	-$38	-$276	N/A	-$51
Unemployment compensation	$462	$283	$443	$420	$420	$477	$478	$478
Schedule E	$197	$99	$94	$220	$220	$117	$225	$225
Farm	$5	$21	$20	$7	$7	$30	$11	$11
Alimony	N/A	$8	$8	N/A	N/A	$12	N/A	N/A
Other income	$409	N/A	N/A	$386	$386	N/A	$433	$433

N/A—Not Applicable.
Census Bureau Disclosure Review Board release authorization numbers CBDRB-FY2021-CES005-003 and CBDRB-FY2021-CES005-020.

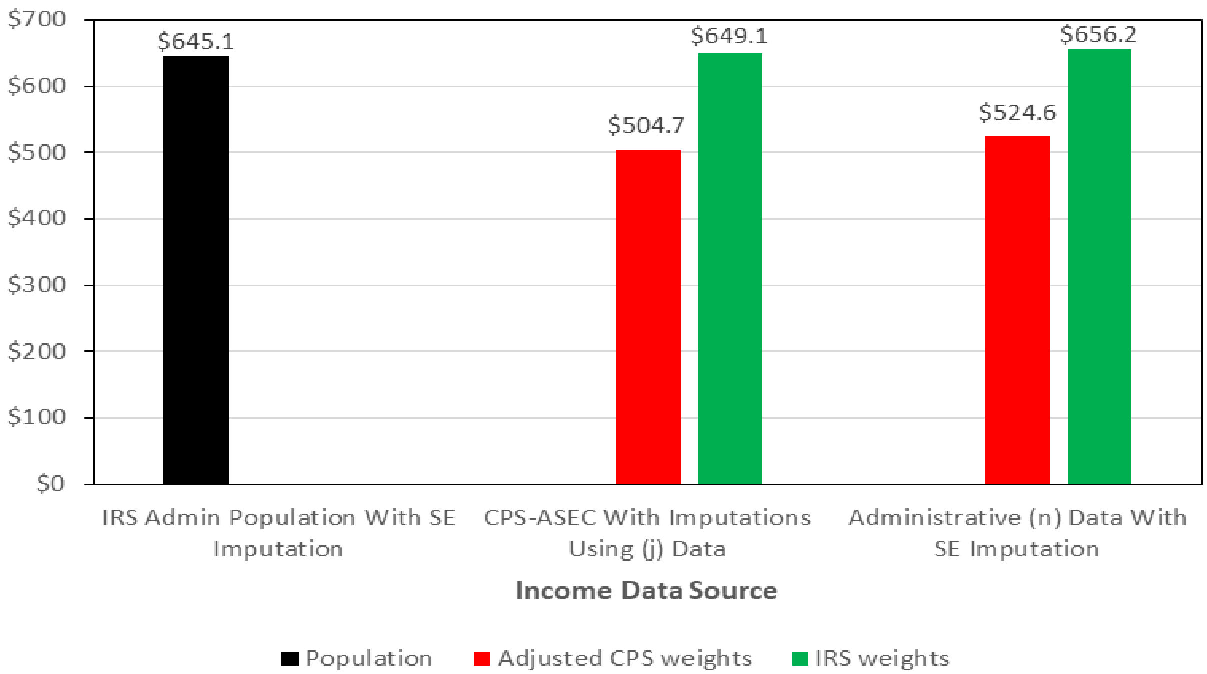
The imputations add significant amounts of pension, Social Security, unemployment compensation, and self-employment income to that reported by CPS-ASEC respondents. But a larger amount of wages is reported on the Census survey than is found on Forms W-2 for this population (possibly due to survey respondents who characterize nonwage income as wages or include unreported tips). The greater amount of wages in the survey responses is derived from far fewer individuals, so the average is much greater ($6,302 compared with $4,426). As a consequence, the total income for columns using CPS-ASEC survey responses, whether the raw reported income amounts (column B) or as a starting point to which imputed amounts are added (column C), ends up similar because the lower amount of wages on the Forms W-2 of potential nonfilers offsets the increased amounts for the income types that are imputed (Figure 4). The range of total income for all of the income definitions using the adjusted CPS-ASEC weights goes from a low of $488.2 billion for the raw CPS-ASEC to $524.6 billion for the administrative (n) data with the self-employment imputation—an increase of 6.2 percent. By contrast, switching between the two re-weighting methods results in a 25.1-percent increase in total income when considering the administrative (n) data with the self-employment imputation (columns E and H) and a 28.6-percent increase when considering the CPS-ASEC with imputations using (j) data (columns C and F in Table 3A).

FIGURE 2. Estimated Number of Potential Nonfilers Having a Given Income Type with Four Combinations of Data, Imputations, and Weights, Tax Year 2010



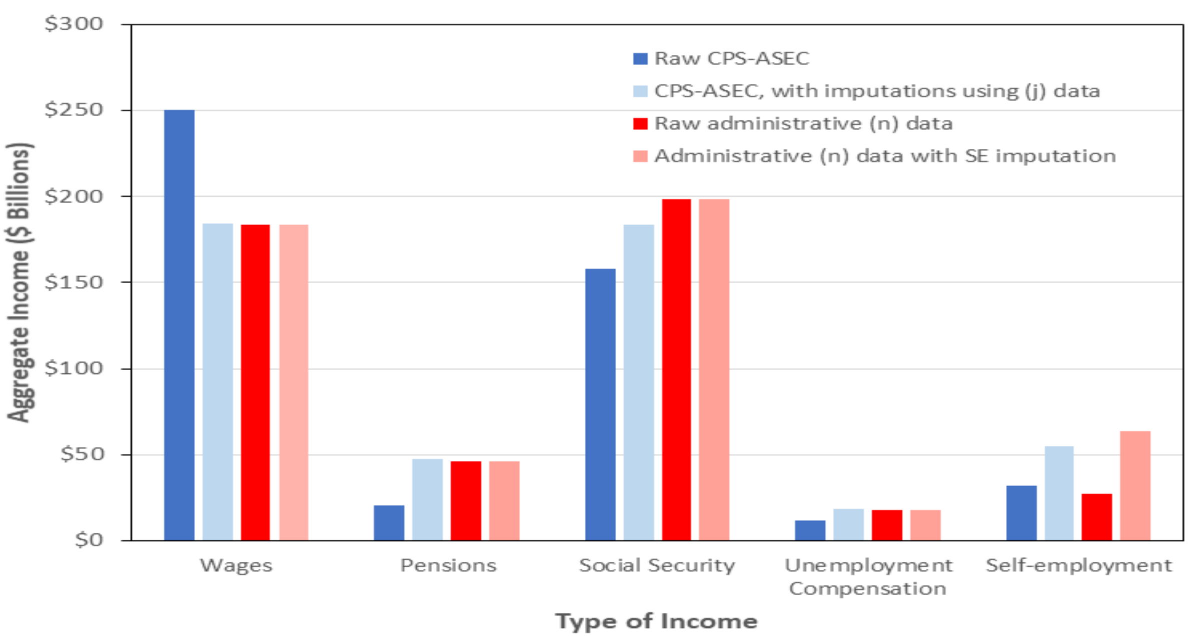
SOURCE: Amounts tabulated in Table 3A.

FIGURE 3. Total Income of Potential Nonfilers Using Different Income Sources and Weights, Tax Year 2010



SOURCE: Amounts tabulated in Table 3A.

FIGURE 4. Estimated Total Income of Potential Nonfilers Using the Adjusted CPS-ASEC Weights With and Without Imputations, Comparing the (j) Data with the (n) Data for Selected Income Types, Tax Year 2010



SOURCE: Amounts tabulated in Table 3A.

Focusing again on Table 3B (column H), we can see that about 31 percent of potential nonfilers have wage income, while about 44 percent have Social Security income and 10 percent have self-employment income. Even though fewer potential nonfilers have wage income, the typical amounts are larger than for Social Security income so that overall wages and Social Security have fairly similar average amounts ($4,845 and $4,525, respectively). These two income sources are far and away the largest sources of income among potential nonfilers. Pensions and self-employment income come next, but at a much lower level of importance, with the average amounts around $1,150 and $1,535. Nonetheless, self-employment income plays a disproportionate role in determining requirement to file and tax liability for low-income earners since any amount above $433 is subject to self-employment tax.

4. Contributions to the Nonfiling Tax Gap

Having identified the potential nonfilers in the linked data, and having re-weighted those records to represent the full population of potential nonfilers, we were able to adapt the standard tax model we had developed for use with the (j) data to assign the potential nonfilers into tax units.[13] But instead of using just the CPS-ASEC records matched to the IRS third-party documents, as explained later, in some cases we also use CPS-ASEC records with PIKs that are not found on a filed return. To build the tax units, we (1) combined the records of spouses;[14] (2) assigned them the Married Filing Jointly filing status; and (3) assigned all others to either Single or Head of Household filing status, depending on their demographics in the CPS-ASEC. Using family demographics this way at the micro level is a key benefit of linking the tax administrative data to the CPS-ASEC survey records.

After creating tax units in this way, we were then able to identify which of them appear to have had a filing requirement. There were two key filing thresholds for Tax Year 2010. First, tax units with more than $433 of net self-employment income are required to file a return to report that income; even if they do not have an income tax liability, they may have a self-employment tax liability. This is one reason we impute self-employment income to the potential nonfilers.[15]

The more common filing threshold applies to everyone. In Tax Year 2010, people were required to file a tax return if their gross income exceeded the sum of their standard deduction, any additional standard deduction on account of being 65 or older or blind, and the value of the personal exemption(s) of the primary taxpayer (and spouse, if any).[16] For Tax Year 2010, the standard deductions were: $5,700 for Single filing status; $11,400 for Married Filing Jointly; and $8,400 for Head of Household. Each age and/or blind additional deduction was $1,400 for Single and Head of Household, and $1,100 for Married. Finally, the value of each personal exemption was $3,650. The corresponding gross income filing thresholds (for those under 65 and not blind) were $9,350 for Single, $18,700 for Married Filing Jointly, and $12,050 for Head of Household. Again, someone's gross income could give them a filing requirement even if they do not have any income tax liability, but the offsetting costs, expenses, deductions, credits, etc., that would reduce their tax liability to zero would need to be reported on a filed tax return.

We also assigned a number of children to these tax units using the information in the CPS-ASEC. These would reduce income tax liability through dependent exemptions, credits, etc. As in our previous estimates derived from the (j) data linked to the CPS-ASEC,[17] the tax model computes self-employment tax,[18] the adjustment

[13] Note that this first set of estimates relates only to individuals who were not represented on a filed tax return by the end of calendar 2011. Some of these may have filed after that date and some of those may have actually filed "on time" due to special extensions to the original due date of their return (e.g., due to a disaster declaration or to being deployed in a combat zone). Our published tax gap estimates (e.g., Langetieg *et al.* (2016) and Internal Revenue Service (2019)) and the estimates provided in Tables 5A and 5B account for these late filers.

[14] In the case in which one spouse could be linked to the (n) data and the other could not, for the purposes of this paper we used the income reported in the CPS-ASEC as that spouse's income. This may partially overcome the fact that not all information documents in the (n) data had been assigned a reliable PIK (see footnote 4). However, the IRS re-weights do not account for that missing spouse among the potential nonfilers.

[15] Note that we impute net self-employment earnings. Moreover, for the tax unit analyses that do not use the detailed income from the (n) data, we impute self-employment income to and include some individuals who cannot be linked to any third-party information document. This is another reason (see footnote 10) the IRS re-weights are not strictly consistent with the records and income included in our tax unit analyses.

[16] This threshold does not take into account any dependents; those need to be claimed on a tax return.

[17] See Langetieg *et al.* (2016, pp. 4-5).

[18] For each of these imputations, a two-step model is applied. In the case of deductions, the first model estimates the likelihood that the taxpayer itemizes deduction rather than taking the standard deduction. For adjustments and credits, the first model estimates the likelihood that the taxpayer has a positive, nonzero amount. In all three cases, the second model estimates the log amount for each of the aggregate line items.

for one-half of the self-employment tax, exemptions, the standard and itemized deductions, taxable income, tentative tax, nonrefundable credits, refundable credits, and tax balance due after credits. Included in these computations are imputations for deductions, nonrefundable credits other than the Child Tax Credit, and adjustments other than the adjustment for one-half of self-employment tax. Subtracting the amount of tax withheld (per Forms W-2 and other information documents when using the (n) data and imputed withholding amounts otherwise), we arrive at an estimate of the net balance due or refund. We aggregate the positive balance due amounts and other quantities on the mock tax return to the population of nonfilers using either the adjusted CPS-ASEC weights or the new IRS re-weights.

Table 4A presents the estimated number of tax units associated with each filing status and the counts of tax units having the given income types. Table 4B presents the corresponding dollar amounts for the income and tax categories. However, there are several differences between the person-level estimates in Table 3A and the tax unit estimates in Tables 4A and 4B—beyond the facts that the tax unit estimates combine spouses into tax units and all tax units are restricted to those that have a filing obligation. Unlike in Table 3A, many of the columns in Tables 4A and 4B also use income reported on the CPS-ASEC when no information returns in the (n) data are matched to the individual (see footnotes 8 and 9). The tax units in Tables 4A and 4B potentially include everyone in the CPS-ASEC with a valid PIK who was not matched to a Form 1040 and associated with a tax return that was estimated to be required. On the other hand, the person-level estimates in Table 3A include only those for whom the CPS-ASEC record matches to at least one third-party information document in the (n) data. For columns F and H, which use the IRS weights, CPS-ASEC income is included only for spouses who are not linked to any IRS third-party data. For columns C and E, which use the adjusted CPS-ASEC weights, CPS-ASEC income is included for all individuals with a PIK who are not linked to any IRS third-party data, whether they are single or married. In each case, the columns using the raw administrative (n) data (columns D and G) do not use any income from CPS-ASEC. The column labeled "Raw CPS-ASEC" (column B) uses only CPS-ASEC reported income for all records with PIKs.

Note from column H that the 50.8 million potential nonfilers from Table 3A reduces to 10.3 million nonfiler tax units. This is partly because around 3 million spouses were combined into one tax unit per couple, but mostly because we estimate that about 37 million potential nonfilers did not have a filing obligation. Notice also that only around 29 percent of nonfiler tax units were married (as opposed to about 37 percent of filers[19]). The 10.3 million tax units in column H (based on the (n) data weighted by our new re-weights) is greater than the 8.9 million estimated from the population in column A. Recall that the method using the population data in column A imputes filing status (either married-joint or single) to potential nonfiler tax units based on aggregate Census demographic breakdowns, whereas filing status in column H is based on specific demographic characteristics at the micro level and includes Heads of Household. Because the imputation method for column A "marries" people to others who also appear on an information document, it ignores the fact that many are actually married to spouses who do not appear on an information document. Thus, it understates the number of single tax units. Moreover, our method for imputing self-employment income is more likely to assign such income to both spouses if they both appear on an information document than if one of them does not, so more married tax units are likely to be above the self-employment income filing threshold under this method. We therefore expect that the estimates in column H are more accurate than those in column A because the requirement to file depends so heavily on filing status.

The total income of nonfilers is estimated in column H of Table 4B to be $426.9 billion, which is about 65 percent of the corresponding $656.2 billion shown in Table 3A for all potential nonfilers. The total tax liability of the nonfilers (after nonrefundable credits) is estimated to be $56.9 billion, $23.3 billion of which was covered by withholding and refundable credits, leaving an aggregate balance due (contribution to the tax gap) of $33.7 billion. This compares very closely with column A, which shows the results that come from applying the IRS Administrative Method. In that approach, total income is $424.5 billion, total tax liability is $56.6 billion, total payments is $24.8 billion, and the aggregate balance due is $31.7 billion[20] (see Figure 5). This gives us some confidence that the balance due estimates based exclusively on IRS administrative data are not dramatically different from those using the matched data, despite the different methods used for composing tax units and the differences in the share of married households.

[19] See IRS, *2010 Estimated Data Line Counts Individual Income Tax Returns,* https://www.irs.gov/pub/irs-soi/10inlinecount.pdf.

[20] Note that this estimate takes estimated payments into account, while column H does not.

TABLE 4A. Estimates of the Number and Income Types of Nonfilers (Tax Unit Level) Using Alternative Weights and Income Sources, Tax Year 2010

Data source:	Administrative population	CPS-ASEC linked to IRS administrative data for potential nonfilers						
Weights:	None needed	Adjusted CPS-ASEC weights				IRS re-weights		
Income source:	Administrative (n) data with SE imputation	Raw CPS-ASEC	CPS-ASEC, with imputations using (j) data	Raw administrative (n) data	Administrative (n) data with SE imputation*	CPS-ASEC, with imputations using (j) data	Raw administrative (n) data	Administrative (n) data with SE imputation*
Item	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Number of Tax Units (Millions)								
All	9.30	7.93	10.30	6.10	10.88	10.48	8.36	10.27
Single	5.41	5.08	7.15	3.23	7.53	6.36	3.92	6.17
Married-Joint	3.90	1.98	2.16	1.48	2.29	2.98	2.38	2.94
Head of Household	N/A	0.87	0.99	1.40	1.05	1.14	2.07	1.17
Wages	6.94	5.85	4.72	4.09	5.60	6.50	5.78	6.39
Interest	2.94	2.19	2.36	1.60	2.21	2.90	2.07	2.56
Dividends	1.39	0.69	0.74	0.80	1.04	0.93	1.03	1.25
Capital gains	0.24	N/A	N/A	0.18	0.20	N/A	0.22	0.25
Pensions	2.22	0.79	1.65	1.33	1.61	2.33	1.90	2.17
Taxable Social Security	1.42	0.58	0.84	0.55	0.77	1.09	0.70	0.90
Positive self-employment	4.30	1.45	5.61	1.61	5.22	4.60	2.20	4.13
Unemployment compensation	1.97	0.91	1.65	1.17	1.62	1.88	1.64	1.82
Schedule E	0.52	0.38	0.38	0.31	0.35	0.56	0.42	0.46
Farm	0.05	0.11	0.10	0.05	0.07	0.14	0.08	0.08
Alimony	N/A	0.03	0.03	N/A	0.03	0.04	N/A	0.03
Other income	1.22	N/A	N/A	0.69	0.76	N/A	0.91	1.02

N/A—Not Appicable.

* This column always uses amounts from the CPS-ASEC for alimony and uses them for other income types for persons who do not have a PIK but are married to an individual who does have a PIK.

Census Bureau Disclosure Review Board release authorization numbers CBDRB-FY2021-CES005-003 and CBDRB-FY2021-CES005-020.

TABLE 4B. Estimates of the Income and Tax Characteristics of Nonfilers (Tax Unit Level) Using Alternative Weights and Income Sources, Tax Year 2010

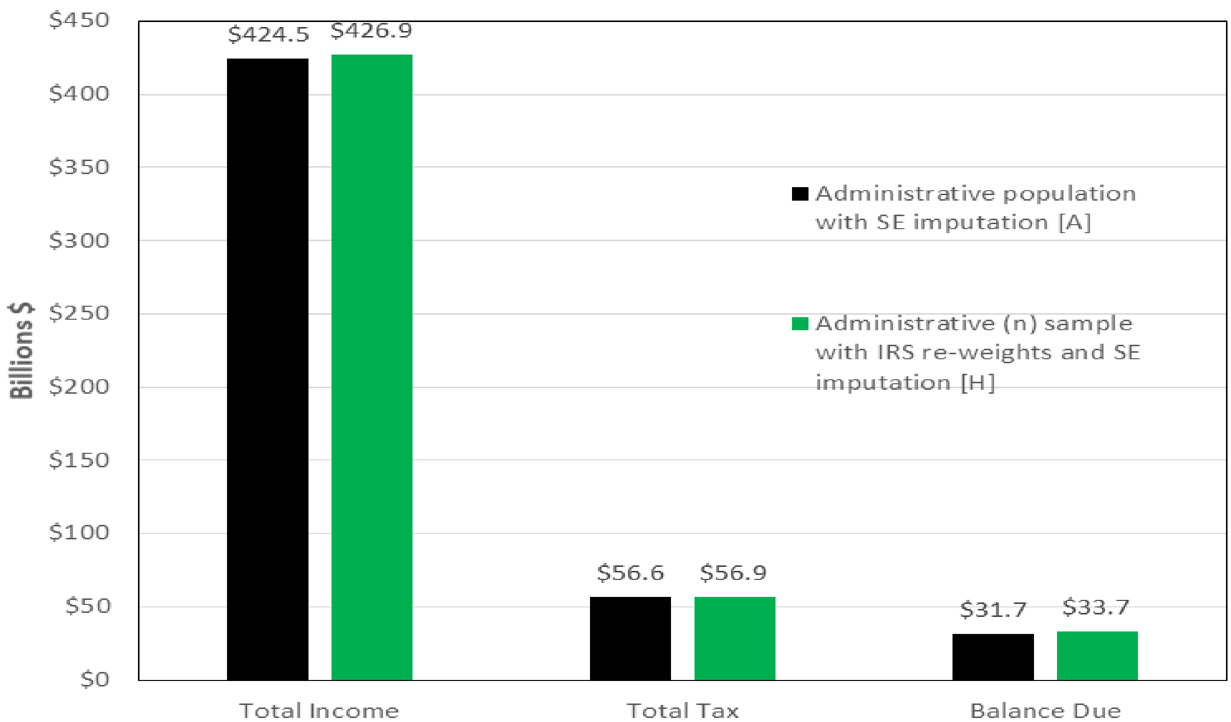
Data source:	Administrative population	CPS-ASEC linked to IRS administrative data for potential nonfilers						
Weights:	None needed	Adjusted CPS-ASEC weights				IRS re-weights		
Income source:	Administrative (n) data with SE imputation	Raw CPS-ASEC	CPS-ASEC, with imputations using (j) data	Raw administrative (n) data	Administrative (n) data with SE imputation*	CPS-ASEC, with imputations using (j) data	Raw administrative (n) data	Administrative (n) data with SE imputation*
Item	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Aggregate Amounts ($ Billions)								
Wages	$235.6	$257.7	$166.2	$154.8	$194.2	$229.4	$216.6	$230.9
Interest	$2.4	$5.6	$5.7	$1.1	$1.6	$7.1	$1.6	$1.9
Dividends	$3.3	$3.7	$3.7	$1.5	$1.7	$4.3	$1.9	$2.2
Capital gains	$2.1	N/A	N/A	$0.3	$0.3	N/A	$0.9	$0.9
Pensions	$33.5	$17.9	$31.3	$28.4	$32.5	$41.3	$38.6	$40.7
Taxable Social Security	$11.1	$4.8	$7.6	$5.1	$6.8	$10.1	$6.4	$8.1
Positive self-employment	$80.5	$40.0	$81.4	$24.1	$95.2	$77.8	$35.0	$89.4
Unemployment compensation	$19.5	$9.2	$17.1	$12.4	$16.2	$18.8	$17.3	$18.6
Schedule E	$15.3	$3.7	$3.7	$8.1	$8.5	$5.6	$13.3	$12.0
Farm	$0.4	$0.9	$0.9	$0.3	$0.3	$1.6	$0.6	$0.6
Alimony	N/A	$0.3	$0.3	N/A	$0.3	$0.5	N/A	$0.5
Other income	$18.2	N/A	N/A	$13.2	$13.4	N/A	$18.6	$18.9
Total income	$424.5	$343.6	$337.6	$249.4	$378.4	$415.5	$350.4	$426.9
Adjusted gross income	$412.7	$336.5	$325.6	$245.3	$366.5	$403.0	$344.6	$415.3
Deductions	$98.3	$74.8	$78.9	$55.7	$85.2	$94.4	$79.5	$93.5
Exemptions	$50.9	$42.1	$36.9	$33.0	$42.5	$49.5	$50.0	$50.9
Taxable income	$263.6	$219.6	$209.7	$156.7	$238.8	$259.1	$215.0	$271.0
Tentative tax	$46.6	$38.9	$36.8	$27.2	$42.7	$45.5	$36.7	$47.9
Nonrefundable credits	$1.1	$2.0	$1.7	$1.4	$1.9	$2.6	$2.5	$2.7
Income tax	$45.5	$36.9	$35.1	$25.8	$40.8	$42.9	$34.2	$45.3
Self-employment tax	$11.1	$5.2	$13.4	$3.4	$13.2	$12.4	$5.0	$11.7
Total tax	$56.6	$42.0	$48.5	$29.2	$54.0	$55.3	$39.2	$56.9
Withholding	$24.1	N/A	$12.1	$15.2	$18.0	$16.2	$20.1	$22.0
Refundable credits	$0.7	$0.6	$1.0	$0.5	$1.2	$1.0	$0.8	$1.3
Balance due	$31.7	$41.5	$35.4	$13.5	$34.8	$38.1	$18.4	$33.7

N/A—Not Appplicable.

* This column always uses amounts from the CPS-ASEC for alimony and uses them for other income types for persons who do not have a PIK but are married to an individual who does have a PIK.

Census Bureau Disclosure Review Board release authorization numbers CBDRB-FY2021-CES005-003and CBDRB-FY2021-CES005-020.

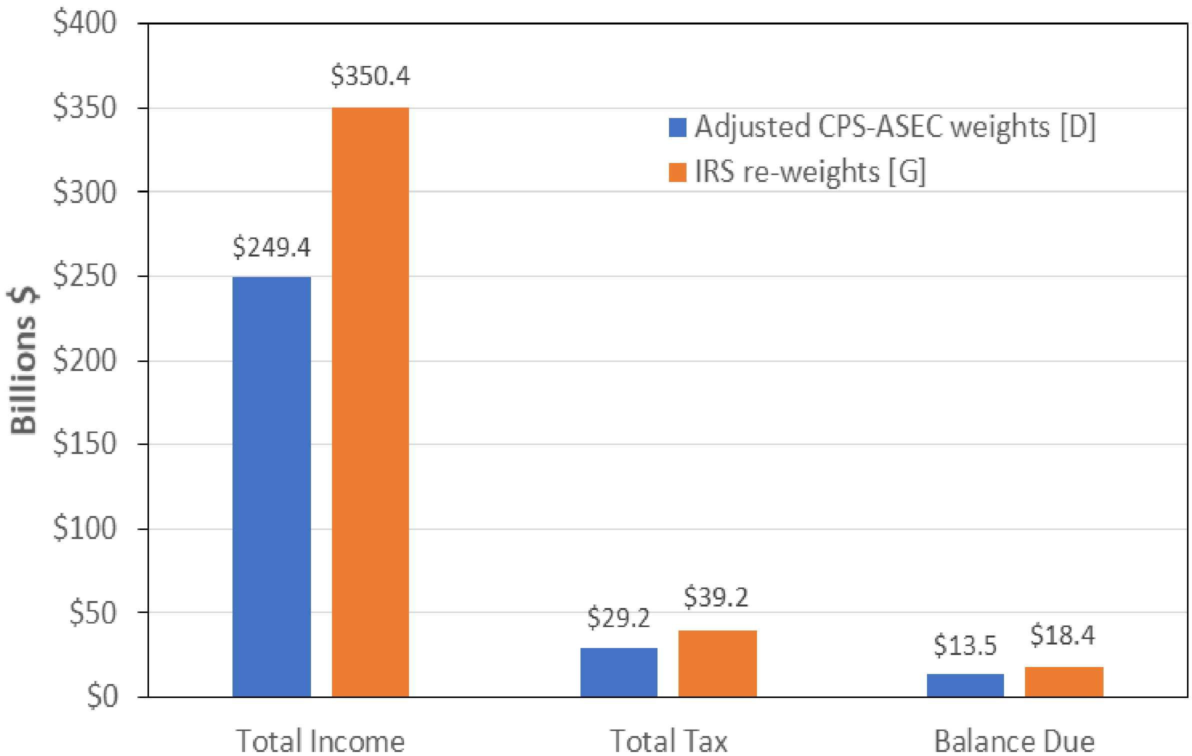
FIGURE 5. Selected Aggregate Income and Tax Values Among Nonfilers Estimated from the CPS-ASEC Records Linked to the (n) Data Weighted by IRS Re-weights, Compared with Corresponding Estimates from Population Data, Tax Year 2010



SOURCE: Amounts tabulated in Table 4B.

A comparison of columns D and G, which are both based on the raw administrative (n) data (thus not including the self-employment income imputation or any CPS-ASEC income), is informative in comparing the impact of using the new IRS weights as opposed to the adjusted CPS-ASEC weights. We can see in Table 4A that the number of tax units increases from 6.1 million to 8.4 million and in Table 4B that income is just over $100 billion more when using the IRS weights. This results in about $10 billion more in total tax and a balance due that is $4.9 billion larger (see Figure 6). By design, and as shown also in Table 3A, the IRS weights are better at re-weighting the matched potential nonfilers to the full population of potential nonfilers (identified in the (n) data) than the CPS-ASEC weights.

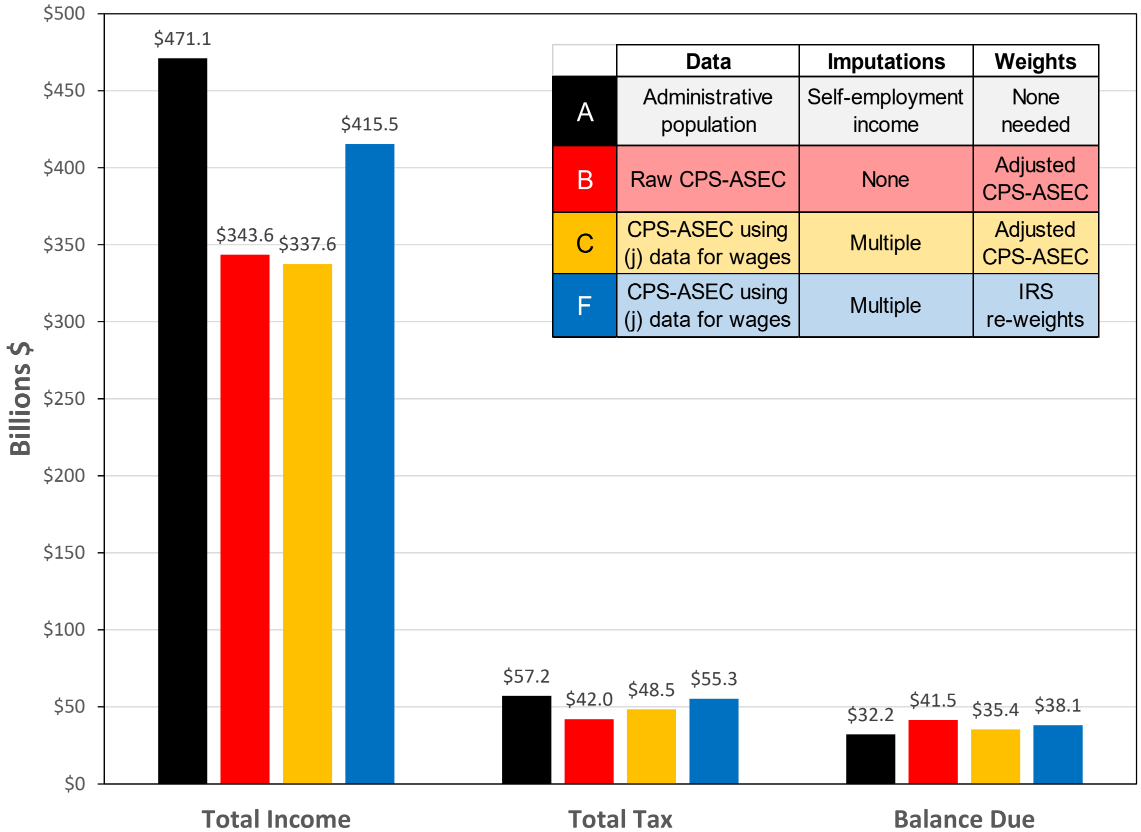
FIGURE 6. Selected Aggregate Income and Tax Values Among Nonfilers Estimated from the Raw (n) Data Using Two Weighting Methods, Tax Year 2010



SOURCE: Amounts tabulated in Table 4B.

While column H apparently does a good job of replicating the mock tax return estimates (income and tax) derived from population data at the IRS (column A), it is not making much use of the CPS-ASEC income information, which might be able to complement the incomplete view of income provided by third-party documents. Columns B, C, and F make more explicit use of CPS-ASEC income data, supplemented by imputations for several income items. In the case of columns B and C, all records not linked to a filed return with a PIK are included in the estimate while in column F only those records that are linked to a spouse who matches to the (n) data are included. But what we have not yet done is to consider both sources of income with respect to a single record and choose how much (if any) of the CPS-ASEC income to use according to reasonable selection rules or by imputing an amount on the basis of both data sources.

FIGURE 7. Selected Aggregate Income and Tax Values Among Nonfilers Estimated with Four Combinations of Data, Imputations, and Weights, Tax Year 2010

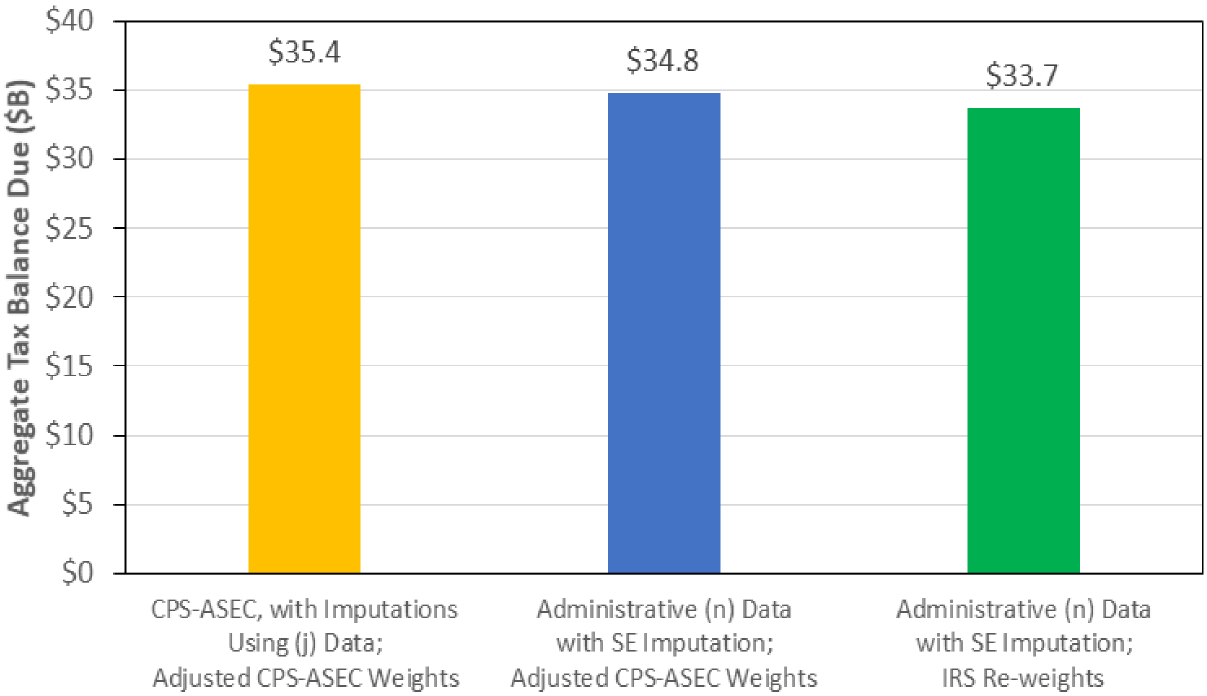


SOURCE: Amounts tabulated in Table 4B.

Table 4B and Figure 7 show that the columns for which the CPS-ASEC income information is used—either in its raw form or with imputations based in the (j) data environment—show larger balance due amounts than those based on the IRS administrative (n) data. In the case of the raw CPS-ASEC estimates (column B), the balance due amount of $41.5 billion arises largely from the fact that nothing has been done to impute withholding so none of the estimated tax liability is offset. As we saw in the discussion of person-level income (Table 3A), the larger amount of wages reported by respondents than appears on Forms W-2 offsets the deficient amounts of income from self-employment, Social Security, pensions, and unemployment compensation found in the raw CPS-ASEC data. Therefore, when these income items are imputed in column C, the use of the smaller wage amount from Forms W-2 in lieu of the CPS-ASEC wage reports results in almost a wash in terms of total income ($337.6 billion vs. $343.6 billion). But again, the comparatively larger balance due amount in column C ($35.4 billion) is not because of unusually high income, but because the amount imputed for withholding appears to understate the amount that shows up on Forms W-2 and other third-party documents (as is evident from looking at the columns that rely mainly or exclusively on administrative (n) data). For our previous tax gap estimates, we addressed this problem by applying an adjustment that accounted for known aggregate withholding. Column F also shows a comparatively large balance due ($38.1 billion). In this case, the relatively large amount of total income ($415.5 billion) arises from the use of IRS weights (which properly re-weight the wage and other income to the full potential nonfiler population identified in the (n) data) and the inclusion of CPS-ASEC raw and imputed income for spouses who are not linked to the (n) data. The larger balance due amount compared with the administrative (n) data column using the same weights (column H) and the IRS administrative population (column A) appears to arise largely because of smaller amounts of withholding.

As Figure 8 illustrates, a comparison of columns C and E suggests that using the (n) data instead of the less comprehensive (j) data results in a $0.6 billion drop in the estimated nonfiler balance due when using the adjusted CPS-ASEC weights, and a comparison of columns E and H suggests that the estimated balance due is further reduced by about $1.1 billion when the IRS weights are used on the (n) data, resulting in an overall reduction of $1.7 billion compared with our previous methodology.[21] As the foregoing discussion explained, the "reduction" due to using the (n) data for income (column E vs. column C) is a consequence of accounting for too little tax withholding when using the (j) data—something we corrected for in the final tax gap estimates with an aggregate adjustment.

FIGURE 8. The Estimated Aggregate Tax Balance Due Among Nonfilers Based on Alternative Income Sources and Weighting Methods, Tax Year 2010



SOURCE: Amounts tabulated in Table 4B.

There are also an estimated 3.7 million late filers (those who filed their 2010 tax return after the deadline, but during 2011) with a requirement to file whose balance due needs to be factored into the nonfiling tax gap (Table 5A). These taxpayers are estimated to have a total income of $223.4 billion compared to the $426.9 billion total income of the 10.3 million not-filers (those who did not file before the end of 2011).[22] By adding the balance due of late filers ($5.1 billion, column I) to that of not-filers, we find that the estimated balance due for all nonfilers is $38.8 billion (or about $36.8 billion if the estimated tax payments of not-filers were taken into account). The comparable estimate including late filers for the IRS administrative population method (column A) would also be $36.8 billion when estimated payments are factored in.

Despite the similarity of the estimates, the composition of the tax units is quite different between the two methods. For example, there are a lot more married required filers in the full population method (which uses

[21] Note that a net balance due contributes to the nonfiling gap, while an expected refund does not reduce the tax gap. However, it is possible that some nonfilers who appear to be due a refund may actually have taxable income not reported on third-party information documents that we have not accounted for (e.g., beyond what we imputed using our self-employment income imputation). Therefore, our aggregate balance due estimate is likely to be a lower bound on the true figure.

[22] The method used to estimate the portion of the nonfiling tax gap attributed to late filers is detailed in Langetieg *et al.* (2016, pp. 11–15).

aggregate Census counts) compared to the matched sample (which uses CPS-ASEC demographic information at the micro level to form tax units, presumably making this the superior method). As noted earlier, the aggregate proportion of married tax units used in the Administrative Method doesn't predict well the number of required tax units by filing status.

TABLE 5A. Nonfiler Estimates Including Late Filers, Tax Year 2010

Scope:	Not-filers	Late Filers	All Nonfilers
Data source:	Matched data (Census)	Administrative population (IRS CDW)	
Weights:	New IRS weights	None needed	
Income source:	Administrative (n) data with NRP SE imputation*	Administrative (n) data	Total
Item	**H**	**I**	**J**
Number of Tax Units (Millions)			
All	10.27	3.73	14.00
Single	6.17	2.35	8.52
Married-Joint	2.94	0.89	3.83
Head of Household	1.17	0.49	1.66
Wages	6.39	2.95	9.34
Interest	2.56	1.25	3.81
Dividends	1.25	0.53	1.78
Capital gains	0.25	0.38	0.63
Pensions	2.17	0.65	2.82
Taxable Social Security	0.90	0.43	1.33
Positive self-employment	4.13	0.85	4.98
Unemployment compensation	1.84	0.41	2.25
Schedule E	0.46	0.46	0.92
Farm	0.08	0.03	0.11
Alimony	0.03	0.01	0.04
Other income	1.02	0.30	1.32

Census Bureau Disclosure Review Board release authorization number CBDRB-FY2021-CES005-020.

* This column always uses amounts from the CPS-ASEC for alimony and uses them for other income types for persons who do not have a PIK but are married to an individual who does have a PIK.

TABLE 5B. Estimates of the Income and Tax Characteristics of Nonfilers (Tax Unit Level) Using the Revised Self-Employment Imputation, Tax Year 2010

Scope:	Not-filers	Late Filers	All Nonfilers
Data source:	Matched data (Census)	Administrative population (IRS CDW)	
Weights:	New IRS weights	None needed	
Income source:	Administrative data with NRP SE imputation*	Administrative (n) data	Total
Item	**H**	**I**	**J**
Aggregate Amounts ($ Billions)			
Wages	$230.9	$137.7	$368.6
Interest	$1.9	$10.1	$12.0
Dividends	$2.2	$5.7	$7.9
Capital gains	$0.9	$16.6	$17.5
Pensions	$40.7	$13.6	$54.3
Taxable Social Security	$8.1	$4.0	$12.1
Positive self-employment	$89.4	$13.0	$102.4
Unemployment compensation	$18.6	$3.8	$22.4
Schedule E	$12.0	$13.1	$25.1
Farm	$0.6	-$0.4	$0.2
Alimony	$0.5	$0.2	$0.7
Other income	$18.9	-$1.7	$17.2
Total income	$426.9	$223.4	$650.3
Adjusted gross income	$415.3	$218.3	$633.6
Deductions	$93.5	$48.7	$142.2
Exemptions	$50.9	$21.0	$71.9
Taxable income	$271.0	$148.6	$419.6
Tentative tax	$47.9	$30.9	$78.8
Nonrefundable credits	$2.7	$1.7	$4.4
Income tax	$45.3	$29.1	$74.4
Self-employment tax	$11.7	$1.9	$13.6
Total tax	$56.9	$31.2	$88.1
Withholding	$22.0	$22.7	$44.7
Refundable credits	$1.3	$3.4	$4.7
Balance due†	$33.7	$5.1	$38.8

* This column always uses amounts from the CPS-ASEC for alimony and uses them for other income types for persons who do not have a PIK but are married to an individual who does have a PIK.

† Before applying estimated payments

Census Bureau Disclosure Review Board release authorization number CBDRB-FY2021-CES005-020.

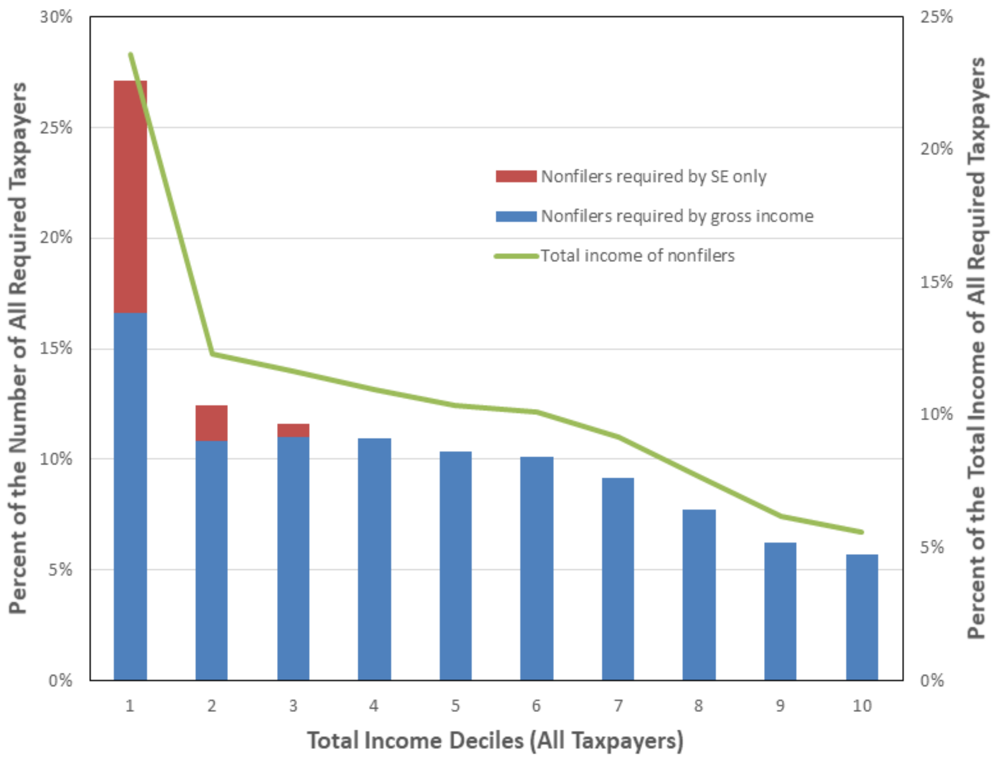
5. Distribution of the Nonfiler Tax Gap

In this section we examine how the counts, tax, and balance due of nonfilers are distributed across income, age, and other characteristics. These estimates are based on the IRS Administrative Method and describe the full nonfiler population; that is, including late filers.

Figure 9 shows that nonfilers make up a larger percentage of all required taxpayers in the lower deciles of total income and particularly in the first decile, where about 27 percent of required taxpayers fail to file a tax return on time. Similarly, about 27 percent of all nonfilers are found in the first income decile, while slightly more than 5 percent are found in the highest decile. A large share of nonfilers in the first decile of total income is estimated to be required to file because they earn enough net self-employment income that they need to pay self-employment tax (>$433), even though their gross income by itself would not require them to file a tax return. While the rate of nonfiling falls by about 7 percent between the 2nd and 10th decile of total income, it remains about 6 percent even among those in the top income decile.

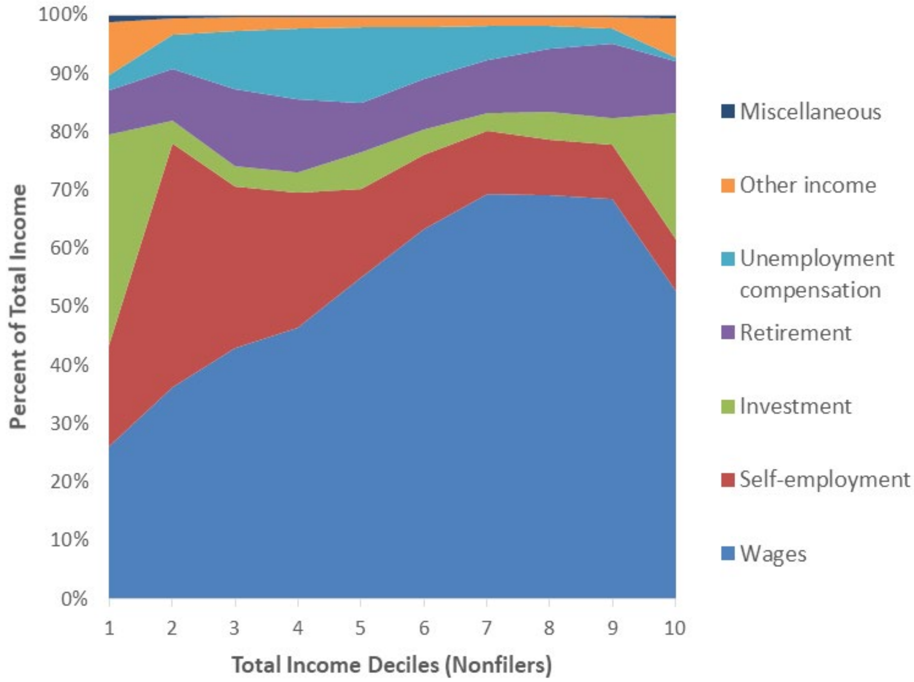
Given the larger rate of nonfiling in the lower income deciles, it is not surprising that the total income of nonfilers makes up a larger share of the total income of all taxpayers in that decile as well. The total income of nonfilers is about 28 percent of the total income of all taxpayers in the first decile, but only about 7 percent in the 10th decile.

FIGURE 9. Nonfilers as a Percent of the Number of All Required Taxpayers and Total Income of Nonfilers as a Percent of the Total Income of All Required Taxpayers by Total Income Deciles Computed Using All Required Taxpayers, Tax Year 2010



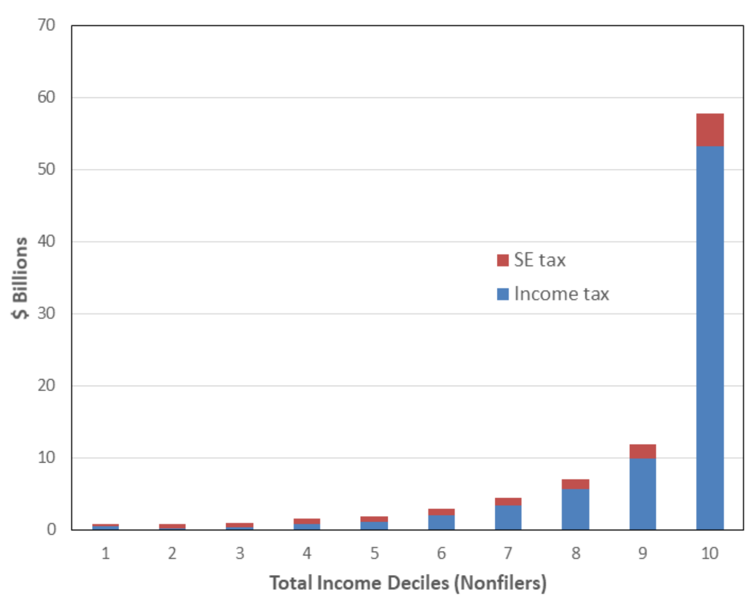
Wages and self-employment income make up the largest portion of the total income of required nonfilers (Figure 10). Of these two, self-employment income is especially important for lower-income nonfilers while the share of wage income becomes much larger at higher incomes. Through much of the distribution, retirement income (including pensions, IRA distributions, and Social Security) and unemployment compensation are the next most prominent sources of income. Unemployment compensation reaches about 13 percent of total income in the 5th decile. At very low and high incomes, the share of investment income (interest, dividends, capital gains and rents and royalties) is fairly significant (about 36 percent in the 1st decile and 22 percent in the 10th decile), but its share is relatively small (between 3 and 6 percent) in the rest of the income range.

FIGURE 10. Income Type Share of Total Income by Total Income Decile of Nonfilers, Tax Year 2010



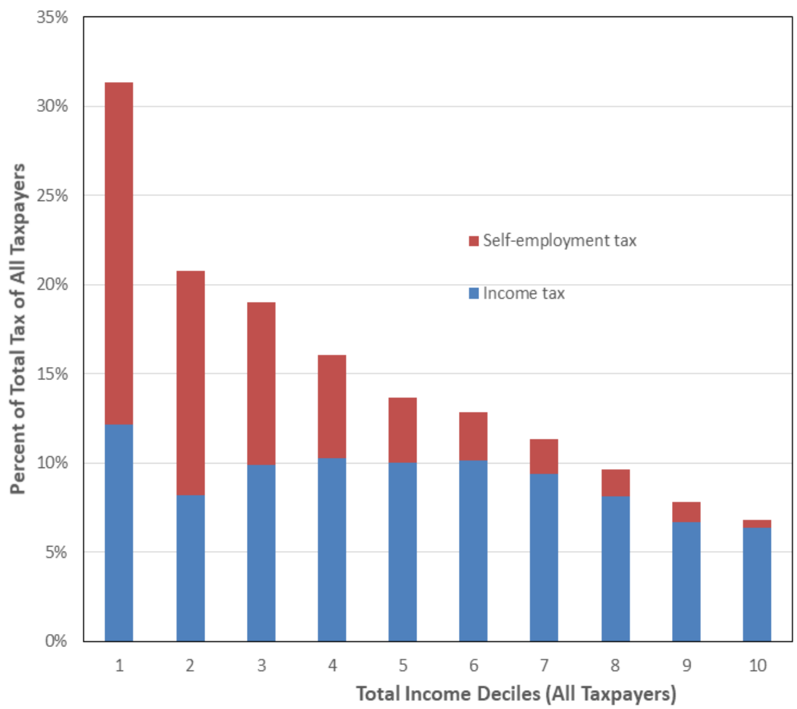
From Figure 11, one can see that a large share of total tax liability of nonfilers is attributed to those in the upper income deciles. This is especially clear for those in the top decile, whose tax liability is about $59 billion and about 64 percent of the total tax attributable to nonfilers. A significant proportion of the tax estimated to be owed by nonfilers in the lower income deciles is self-employment tax, but income tax makes up an increasing share as income increases. Much of the self-employment tax owed by low income nonfilers is derived from the self-employment income that is imputed to these taxpayers.

FIGURE 11. Income Tax and Self-Employment Tax by Total Income Deciles of Nonfilers, Tax Year 2010



From Figure 12 we can see that the total tax liability of nonfilers is around 31 percent of the tax of all required taxpayers in the 1st decile, but only around 7 percent of this amount in the highest income decile. A significant share of the tax liability of nonfilers in the first decile is attributable to those who are estimated to be required to file on the basis of having net self-employment income greater than $433 rather than having gross income above the threshold amounts for their filing status.

FIGURE 12. Income Tax and Self-Employment Tax of Nonfilers as a Percent of Total Tax of All Taxpayers by Total Income Deciles, Tax Year 2010



About 60 percent of the nonfiler tax gap can be attributed to those in the top income decile (Figure 13). Late filers account for a noticeable share (about 16.5 percent) of the total estimated balance due from nonfilers in this decile and close to 10 percent of the balance due of all nonfilers. But the balance due of nonfilers in the lower income deciles make up a larger percentage of the total tax of all required taxpayers than in the upper income deciles (Figure 14). While in the first three deciles of total income the balance due of nonfilers is more than 12 percent of the total tax of all required taxpayers, in the highest income decile it is less than 3 percent. Similar to the case for total tax (Figure 12), a significant proportion of the balance due of those in the first three deciles of total income is attributable to nonfilers who are estimated to be required just on the basis having net self-employment income above the low $433 threshold, rather than having sufficient gross income to meet that higher threshold.

FIGURE 13. Share of Nonfiler Tax Gap by Total Income Deciles of Nonfilers, Tax Year 2010

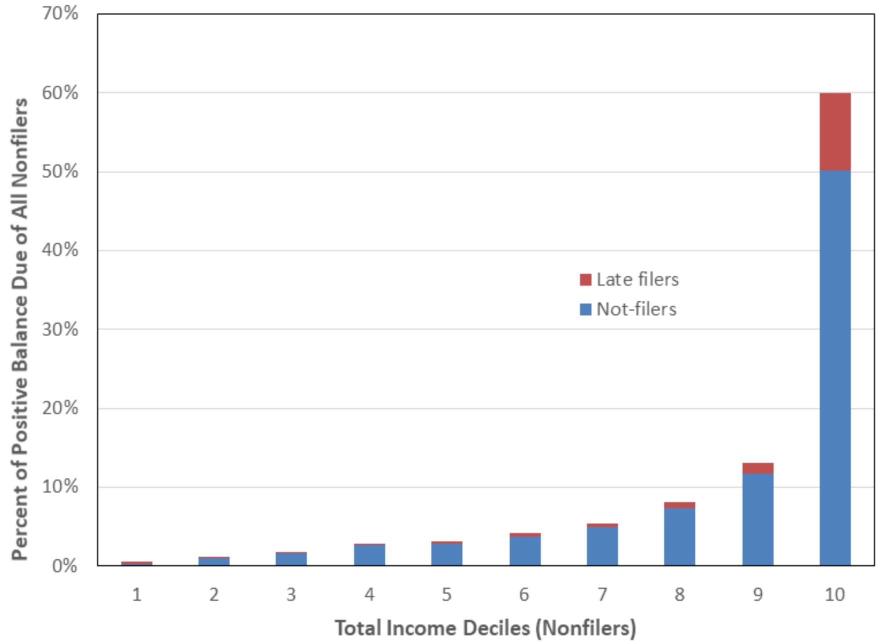


FIGURE 14. Balance Due of Nonfilers as a Percentage of the Total Tax of All Required Taxpayers by Total Income Deciles, Tax Year 2010

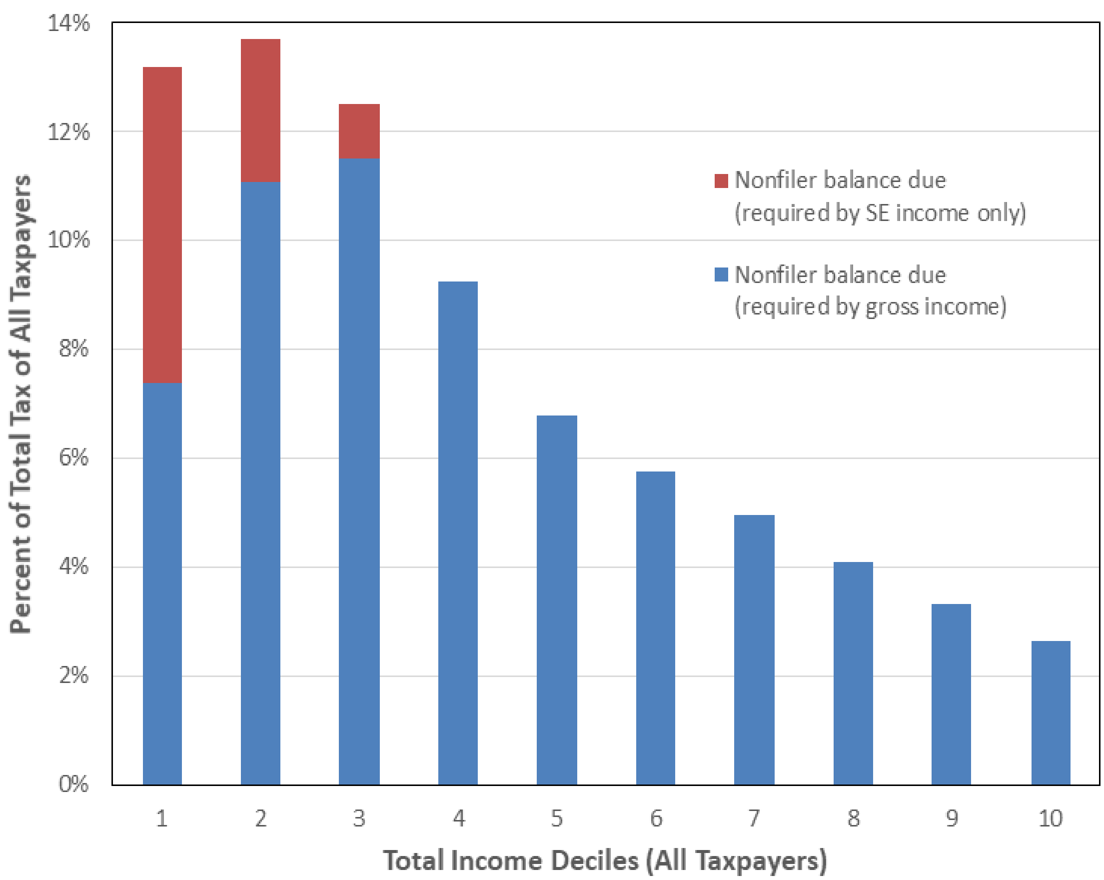


Figure 15 shows that the VFR climbs from around 70 percent for 18-year-olds to 91 percent for 25-year-olds. After that, it dips slightly down to around 89 percent for those in their middle ages (38 to 54) and then climbs back to around 91 percent until taxpayers reach their mid-70s. Beyond that point, the VFR begins to fall again. Given that the rate of nonfiling is higher for younger taxpayers, the balance due as a share of the total tax of all required taxpayers is also significantly larger for this group (Figure 16). This proportion falls precipitously between the ages of 18 and 22 from around 41 percent to 8 percent, and then continues dropping more gradually throughout the age distribution until it is barely above 2 percent for taxpayers in their 60s.

FIGURE 15. Voluntary Filing Rate and Numbers of Not-filers and Late Filers by Age, Tax Year 2010

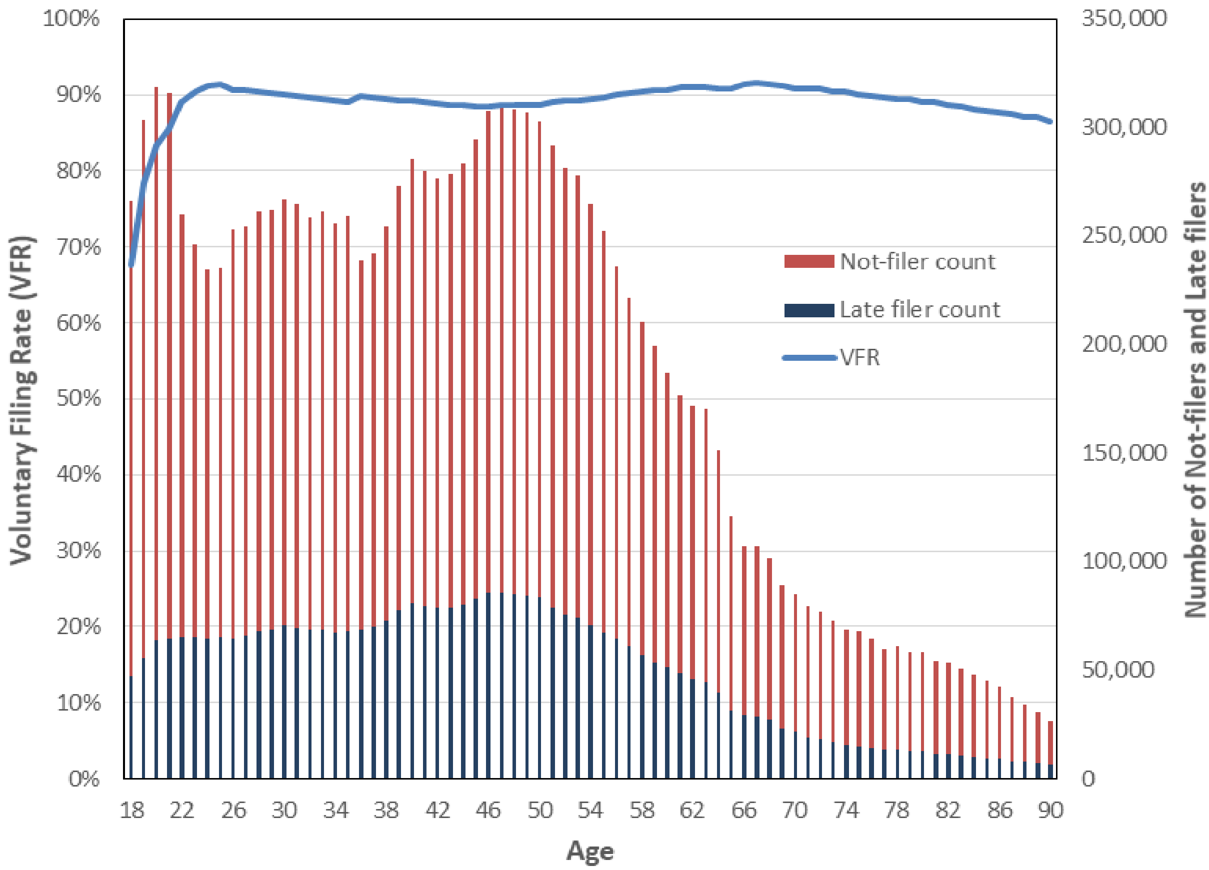
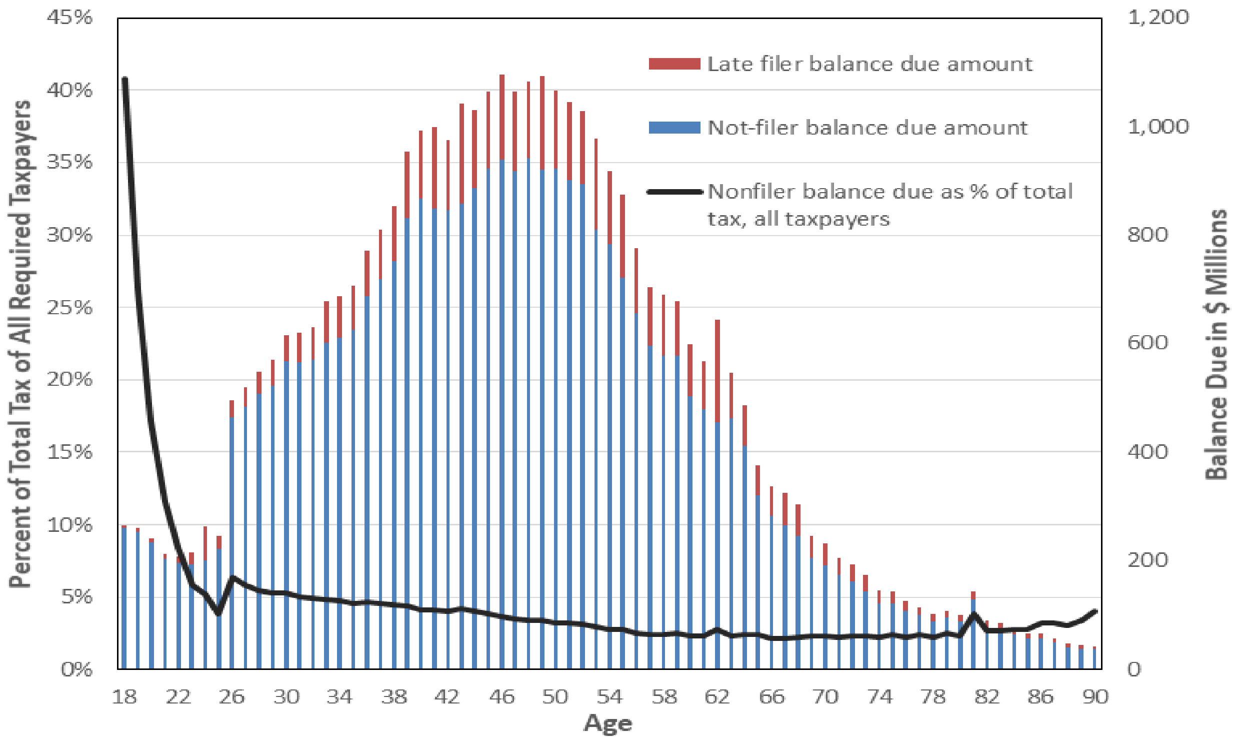
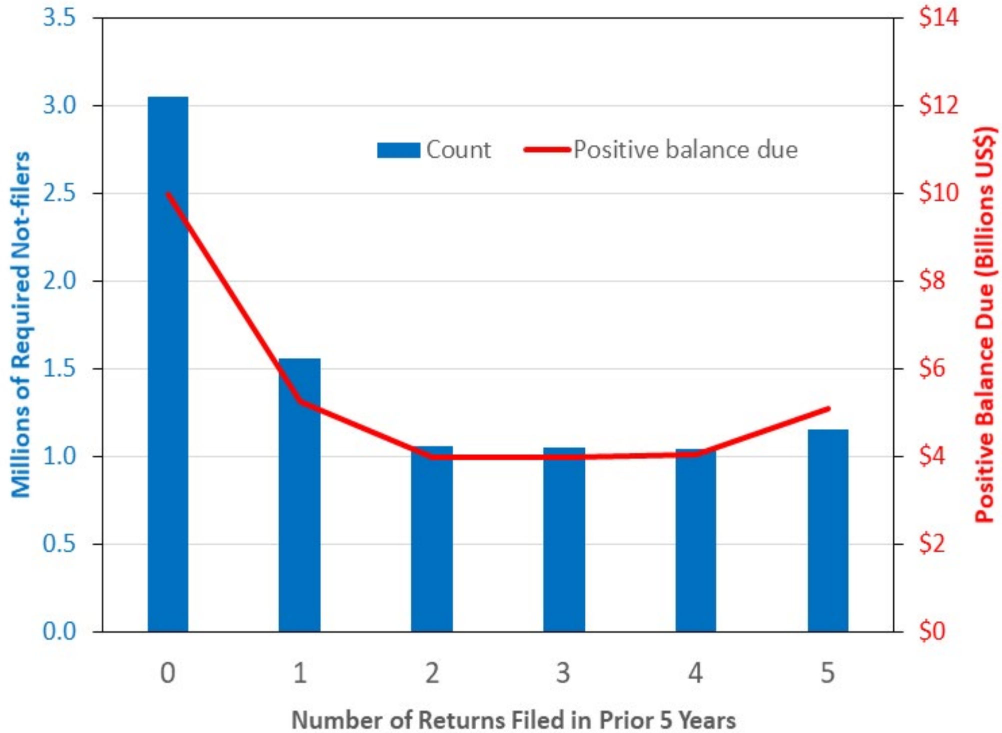


FIGURE 16. Balance Due of Nonfilers as a Percent of Total Tax of All Taxpayers by Age, Tax Year 2010



A significant share (34 percent) of not-filers did not file a return in the first processing year for any of the previous 5 tax years (Figure 17). Fewer than half filed more than one return in the previous 5 years. Similarly, about 47 percent of the total balance due of not-filers is attributed to those who filed in just one or none of these years. Some of these taxpayers were likely not required to file in some of the previous tax years. But in general, this suggests that nonfiling is a recurrent pattern for many taxpayers.

FIGURE 17. Counts and Positive Balance Due of Required Not-filers by How Many of the Previous 5 Tax Years They Filed a Tax Return Within One Processing Year, Tax Year 2010



6. Concluding Remarks

This analysis suggests that, at least for Tax Year 2010, the new method for estimating the extent of nonfiling using the (n) data yields a very similar—and likely better—estimate compared with the IRS Administrative Method, which is based on IRS population data. Given the previously mentioned factors that could be driving the estimates apart in different directions (e.g., the slightly different starting populations, the use of some CPS-ASEC data in the new method, and the more accurate formation of tax units when using the (n) data), it is not possible yet to be certain which method is more accurate. It seems reasonable to expect that greater micro-level accuracy in assigning individuals to tax units provides some benefits over the algorithmic approach. In this context, we would like to test whether the factors we have suggested that may be contributing to the larger number of married filers in the algorithmic approach to creating tax units fully explain the difference. It will be valuable to have additional years of (n) data available at Census in order to assess how the estimates using the two methods trend over time. But there are clear advantages to undertaking these estimates using the (n) data for other tax administration research purposes as well, as discussed below.

Both estimates are lower than the estimates derived from the prior method whereby we used CPS-ASEC data with imputations drawing on the (j) data and applied the adjusted CPS-ASEC weights. Using those weights, and despite imputations for various kinds of income, we find that total income is significantly lower in the aggregate than what we find for the IRS administrative population or using the re-weighted (n) data. But it turns out that the estimated balance due amount is larger under that prior method because the imputed aggregate withholding amount does not offset as much tax in those methods.

In future work, we would like to explore an alternative re-weighting approach that incorporates those with PIKs who are not linked to tax returns or to any third-party information documents. In the current method using the IRS re-weights, this group is not considered as part of the population of potential nonfilers that the weights target. Only those who are spouses of those in the matched nonfiler sample are included in the estimates and the weight used is that of the matched record. But it would be valuable to explore a re-weighting scheme that would properly re-weight the matched records and these nonfiling CPS-ASEC records to the full potential nonfiler population defined by those on third-party documents and also those with income not reported on third-party documents. Because of the likelihood of inaccurate or incomplete matching going in both directions, there is a risk of double counting as well as undercounting and it is unclear how far to go in either direction.

We would also like to explore to what extent income reported on CPS-ASEC might be used in conjunction with the (n) data for estimating the income for each of the matched records.

Moreover, it would be instructive to replicate the current method by linking data from the American Community Survey (ACS) to the (n) data to see how the estimates compare. So far, we have focused on CPS-ASEC because it has more detailed income questions that are focused on the prior calendar year, which aligns well with the tax year for most individual filers. In addition, the fact that it is administered in March means that respondents are more likely to have in mind (or in hand) their income tax return and other tax documents when answering the income-related questions. The ACS, in contrast, has more limited income questions and is implemented using a rolling sample with questions focused on the previous 12 months rather than a particular calendar year. But despite these shortcomings, the ACS has a few advantages—particularly if we don't need to rely on the surveys for income information. First, it is much larger, with about 3 million households per year compared with the CPS-ASEC's 100,000. Second, since it is mandatory (like the Decennial Census), it has a better response rate at the unit and item level. Third, the sampling frame is updated more frequently and includes individuals in institutional and noninstitutional group quarters. Thus, it is worth exploring how the measures would compare using the ACS, and potentially taking advantage of the larger survey to study nonfiling behavior among smaller segments of the population.

Erard (2020) and Erard *et al.* (2020) analyze the drivers of nonfiling using a calibrated qualitative choice estimation methodology, drawing on 13 years of cross-sectional samples from IRS tax return data and the CPS-ASEC. Using this estimation methodology permitted an examination of drivers of trends as well as of individual differences in filing behavior. But a limitation of this methodology is that one must restrict the set of explanatory variables to those that are comparably measured in both data sources. When multiple years of (n) data become available at Census, we would like to apply a more conventional probit or logistic regression methodology, which will permit the inclusion of many more variables that likely influence filing behavior and a more robust and complete examination of the drivers of nonfiling. Our aim will be to provide IRS decision-makers with information that will help them to address nonfiling through improved services and outreach efforts to potential nonfilers, as well as to conceive of more efficient enforcement efforts where those might be needed.

References

Erard, Brian (2020), "Modeling Qualitative Outcomes by Supplementing Participant Data with General Population Data: A New and More Versatile Approach." MPRA Paper 99887, University Library of Munich, Germany. Revised April 26, 2020.

Erard, Brian, and Chin-Chin Ho (2001), "Searching for Ghosts: Who Are the Nonfilers and How Much Tax Do They Owe?" *Journal of Public Economics.* Vol. 81 pp. 25–50.

Erard, Brian, Pat Langetieg, Mark Payne, and Alan Plumley (2014), "Missing Returns vs. Missing Income: Estimating the Extent of Individual Income Tax Filing Noncompliance from IRS and Census Data," *Proceedings of the National Tax Association Annual Conference,* November 2014.

Erard, Brian, Pat Langetieg, Mark Payne, and Alan Plumley (2020), "Flying under the Radar: Ghosts and the Income Tax," *CESifo Economic Studies,* January 13, 2020. https://doi.org/10.1093/cesifo/ifz021.

Erard, Brian, Mark Payne, and Alan Plumley (2012), "Advances in Nonfiling Measures," *2012 IRS Research Bulletin,* Internal Revenue Service, Publication 1500, Washington, DC, pp. 79–89. Available at https://www.irs.gov/pub/irs-soi/12resconadvnonfile.pdf.

Hertz, Tom, Pat Langetieg, Mark Payne, and Alan Plumley (2020), "Estimating the Extent of Individual Income Tax Nonfiling," *Proceedings of the National Tax Association Annual Conference,* November 2020.

Internal Revenue Service (2019), *Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011–2013*, Publication 1415 (Rev. 9-2019). Washington, DC, September 2019. Available at https://www.irs.gov/pub/irs-pdf/p1415.pdf.

Langetieg, Pat, Mark Payne, and Alan Plumley (2016), "The Individual Income Tax and Self-Employment Tax Nonfiling Gaps for Tax Years 2008–2010," *2016 IRS Research Bulletin*, Internal Revenue Service, Publication 1500, Washington, DC, pp. 38–59. Available at https://www.irs.gov/pub/irs-soi/16resconpayne.pdf.

Langetieg, Pat, Mark Payne, and Alan Plumley (2017), "Counting Elusive Nonfilers Using IRS Rather Than Census Data," *2017 IRS Research Bulletin,* Internal Revenue Service, Publication 1500, Washington, DC, pp. 197–222. Available at https://www.irs.gov/pub/irs-soi/17resconpayne.pdf.

Lawrence, Josh, Michael Udell, and Tiffany Young (2011), "The Federal Tax Position of Persons Not Filing Returns for Tax Year 2005," *2011 IRS Research Bulletin,* Internal Revenue Service, Publication 1500 (Rev. 4-2012), pp. 143–155. Available at https://www.irs.gov/pub/irs-soi/11rescontaxposition.pdf.

Wagner, D., and Layne, M. (2012), "Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' Record Linkage Software," Washington, DC: Center for Administrative Records Research and Applications Internal Document, U.S. Census Bureau.

Using Uplift Modeling To Improve ACS Case Selection and Compliance Outcomes

Jan Millard (IRS Research, Applied Analytics, and Statistics) and Travis Whitfield, Sarah Smolenski, Michael Stavrianos, and Lauren Szczerbinski (ASR Analytics, LLC)

Introduction

This paper introduces the use of uplift modeling to improve case selection processes, building on the improvements in compliance outcomes realized by incorporating behavioral insights (BI)[1] in prior notice redesign efforts. With a growing desire to be data-driven and optimize IRS resources, the Automated Collection System (ACS) wanted to test the use of predictive modeling to score the impact of a proposed treatment for each potential case so decision makers can determine how to allocate treatment resources.

In this paper, the authors present findings from an uplift modeling test using data generated by the 2019 LT11 Notice Redesign Pilot. Uplift models for dollars recovered and incoming telephone calls were used to predict the potential treatment effects for taxpayers in ACS if they were selected to receive an LT11 notice. These uplift predictions will be used to implement a case selection methodology based on the resource constrained optimization of uplift in dollars collected subject to a cost constraint primarily driven by uplift in incoming telephone calls.

Building upon the lessons learned from prior pilots, in this iteration predictive models were built to specifically identify taxpayers who are expected to receive a refund offset in the near future, and ensure their cases are <u>not</u> selected for receipt of a notice. A key goal of this effort was to predict latent, underlying features of taxpayers which are likely to have a large impact on their behavior, primarily whether they will receive a refund offset.

Background

In 2016, as part of the IRS-wide Future State Initiative, the ACS Optimization team began a notice redesign effort focused on collection notices issued through ACS. The notice redesign initiative comprises three major ACS priorities:

1. **Improve Taxpayer Experience** by directing taxpayers towards self-service tools such as Online Payment Agreement (OPA) and Voice Balance Due (VBD).
2. **Increase Taxpayer Compliance** by designing notices encouraging taxpayers to pay their balances due through full payments, partial payments, or installment agreements.
3. **Reduce IRS Costs** by decreasing the volume of telephone calls and correspondence the IRS receives.

The first notice redesigned as part of this effort was the LT16, through the 2017 LT16 Notice Redesign Pilot. The LT16, an ACS notice, warns a taxpayer of a potential downstream levy if their delinquent debt is not resolved. During this pilot, six different redesigned notices were tested against the existing LT16, which served as a control notice, in a randomized controlled trial (RCT). Various compliance metrics—such as full payment and installment agreement (IA) rates, dollars collected, and self-service rates—were compared across the different notices to determine which notice created the most lift in compliance relative to the control notice. Results from the initial LT16 redesign demonstrated that redesigned notices can increase compliance rates and reduce costs (e.g., an increase of up to 11 percent in overall compliance rate relative to the control notice, and a

[1] See Appendix A for a list of acronyms.

decrease of 7.5 percent to 19.5 percent in average cost per case resolution). Crucially, this pilot did not include a non-treatment group, and thus it was not possible to develop uplift models based on this pilot alone.

The 2019 LT16 Model Test Pilot followed the initial LT16 Notice Redesign Pilot study. This pilot tested a preliminary case selection process using non-uplift predictive modeling and ran a second RCT to verify the results of the initial LT16 Notice Redesign Pilot and to provide a non-treatment group for further uplift modeling and analysis. The lessons learned from this pilot heavily influenced the uplift modeling analysis described in this paper. Those lessons learned and the results of the 2019 LT16 Model pilot are reviewed in more detail below.

The 2019 LT11 Notice Redesign Pilot followed and was similar to the original 2017 LT16 Notice Redesign Pilot. The LT11 is another ACS notice which informs taxpayers of their Collection Due Process rights and that a levy will be issued if their delinquent debt is not resolved. A taxpayer will only receive an LT11 if they have identified assets which can be levied, and thus LT11 eligible taxpayers correspond to a population with different characteristics than taxpayers who are eligible to receive an LT16. This pilot consisted of two separate RCTs, one which compared four different redesigned variants against a control notice for individual master file (IMF) taxpayers, and one which compared a single redesigned variant against the control notice for business master file (BMF) taxpayers. Each RCT also included a non-treatment group. The non-treatment group for individual taxpayers was used to develop the uplift models described in this paper.

Uplift Modeling

Because taxpayer interventions, such as issuing a notice or initiating an enforcement action, are costly to implement and not always effective, it is desirable to restrict their application to cases in which they are likely to have a significant beneficial impact. Uplift modeling involves the application of supervised learning techniques to estimate the incremental impact (or "uplift") of a specified treatment on individual recipients.

Taxpayers can generally be thought of as falling into one of the following four categories based on how they would respond to a treatment: *Sure Things*, *Do Not Disturbs*, *Persuadables*, and *Lost Causes*. These four categories correspond to how a given taxpayer would respond to receiving a treatment, or not receiving a treatment, and are described in Figure 1.

FIGURE 1. Uplift Category Matrix

Taxpayer Behavior If Sent a Notice

Taxpayer Behavior If Not Sent Notice	Payment	No Payment
Payment	**Sure Thing** Taxpayers who "self-cure" (i.e., full pay regardless of whether they receive a notice). Includes refund offsets.	**Do Not Disturb** Taxpayers for whom a notice has a negative impact on the likelihood to full pay. Also represents taxpayers who only call upon receipt of a notice.
No Payment	**Persuadable** Taxpayers who full pay only if they receive a notice.	**Lost Cause** Taxpayers who will not full pay regardless of notice receipt. Includes taxpayers who are not financially capable or unlikely to successfully receive the notice.

Uplift modeling intends to help identify which taxpayers fall into which category. For instance, if a given taxpayer has a predicted uplift close to zero, this indicates they are expected to fall into the *Sure Thing* or *Lost*

Cause category and should therefore not be sent a notice. A taxpayer with a high predicted uplift has a tendency to fall into the *Persuadable* category, which indicates they would have the best response to receiving a notice (i.e., they are the most likely to pay if and only if they receive a notice). A taxpayer with a negative predicted uplift is potentially in the *Do Not Disturb* category, which means that sending a notice will reduce their probability of paying. We expect that the *Do Not Disturb* category represents a small segment of the population, as it seems unlikely a taxpayer who would otherwise pay is persuaded not to make a payment due to the receipt of a notice. On the other hand, there could be a large segment of taxpayers who are only prone to call the IRS if they receive a notice. If there are not enough resources to handle the expected telephone calls, these taxpayers would be avoided.

Uplift Modeling Implementation

Uplift modeling is used here to develop algorithms that predict uplift with respect to business objectives, in this case increasing revenue and decreasing incoming telephone calls. The case selection methodology will prioritize cases based on their predicted uplift in compliance per hour (or dollar) of notice-induced telephone utilization.

An overview is provided below of options for estimating the uplift associated with the issuance of an LT11 or LT16 notice, and the methodology employed to use the model predictions to prioritize cases for issuing notices.

Predicting Uplift

The uplift associated with a specified intervention or treatment (such as issuing a collection notice to a taxpayer) is defined as the incremental change in the likelihood an individual will undertake a specified action (such as making a payment of delinquent tax) in response to the treatment. Let T represent a binary indicator of whether a taxpayer has received the treatment and allow Y to represent a binary indicator of whether the taxpayer has undertaken the specified action. The uplift associated with this treatment may then be expressed as:

$$P(Y=1 \mid T=1)-P(Y=1|T=0),$$

namely, the difference between the probability a taxpayer will take an action after receiving the treatment and the probability the individual will take the action if there is no treatment.

Estimating uplift is challenging, in part, because an individual cannot both receive and not receive a given treatment. Consequently, one is unable to directly observe the counterfactual outcome if the individual's treatment assignment had been different. The best one can do is compare the outcomes of individuals receiving one treatment to the outcomes of observationally similar individuals who received a different treatment (or no treatment at all). Ideally, the basis of such estimates is data from an RCT as discussed in this paper. This approach ensures the underlying characteristics (X) of the individuals in each group will be comparable.

There are a variety of alternative ways to develop uplift estimates. An overview of three popular approaches is provided in Figure 2: the two-model approach, the single-model approach, and the class-transformation method.

FIGURE 2. Uplift Modeling Approaches

Approach	Description
Two-Model	Under the two-model approach to estimating uplift, one develops separate models to predict the likelihood of compliance for a taxpayer with a given set of characteristics (e.g., age, income, location, etc.) under the scenarios in which the taxpayer does and does not receive a notice. The difference between these two estimated probabilities then serves as the measure of uplift for that taxpayer. Using the results from these two models, one can predict, for any member of the eligible population, the probability of compliance under the scenarios where the taxpayer does and does not receive the notice. The difference between these two predicted probabilities serves as a measure of uplift for that case. Taxpayers in the eligible population for which this measure is high can then be selected to receive a notice. The two-model approach has several appealing features. First, it requires no specialized data preprocessing, so it is relatively straightforward to implement. Second, it can be implemented with most popular supervised machine learning algorithms (e.g., logistic regression, support vector machines, decision trees, etc.), thereby allowing for a robust variety of tuning options. Third, it is the basis for some more elaborate approaches and therefore serves as a useful baseline for performance comparisons. At the same time, there are two potential drawbacks to the two-model approach. First, the reliance on two separate models to estimate uplift may be associated with some loss of precision. Second, the uplift measure is likely to be biased unless care is taken to ensure the taxpayers in the control and treatment groups are comparable in all relevant respects (such as through randomized assignment).
Single-Model	Instead of fitting separate classification models to samples of treated and untreated cases (taxpayers who did and did not receive notices) to predict the outcome under the two alternative treatment scenarios, the single-model approach applies a single classification model to the combined sample of treated and untreated cases. Under this approach, the set of characteristics applied under the two-sample approach is expanded to include a binary indicator (T) of treatment status, where T=1 for cases from the treatment group (i.e., taxpayers who received notices) and T=0 for cases from the control group (i.e., taxpayers who did not). As with the two-model approach, the results of estimation can be used to predict, for any eligible case in the overall population, the probability of compliance under the alternative scenarios of receiving and not receiving a notice. The difference between these two predicted values then serves as a measure of uplift to prioritize cases for selection. There are two main drawbacks to this approach. First, it is less flexible in addressing different forms of heterogeneity in treatment effects. Second, it is more prone to overfitting than the two-model approach. However, the single-model approach does have some advantages. It tends to perform better than the two-model approach when the treatment has no uplift within a large share of the eligible population—for instance, if a significant proportion of taxpayers ignore notices. Furthermore, while it can be implemented using standard supervised learning models like the two-model approach, only one model needs to be tuned instead of two.

FIGURE 2. Uplift Modeling Approaches—Continued

Approach	Description
Class-Transformation	The third approach is the class-transformation method. Like the single-model method, the class-transformation method relies on applying a single classification estimator to the combined sample of treated and untreated cases. However, class assignment under this approach is based on a newly defined target variable (Z) whose value depends on both the outcome of interest (Y) and the treatment status (T) associated with an observation: $$Z=T^*Y+(1-T)^*(1-Y)$$ There are only two possible ways for Z to take the value 1 in the combined sample: either the observation is from the treatment group and has an outcome of Y=1 (a taxpayer received a notice and complied) or the observation is from the control group and has an outcome of Y=0 (the taxpayer did not receive a notice and did not comply). Together, these two cases represent the necessary conditions for uplift to occur. Intuitively, then, it makes sense to prioritize those cases for treatment for which the indicator Z has a high likelihood of taking the value 1; conversely, it would be a poor use of resources to prioritize cases where the taxpayer is unlikely to comply after receiving a notice (Y=0|T=1) or where a taxpayer is still likely to comply without receiving a notice (Y=1|T=0))—both of which result in Z taking the value zero. This approach assumes an equal number of cases are randomly assigned to the treatment and control groups, defining uplift as: $$uplift=2^*P(Z=1 \mid X)-1$$ A standard machine learning classification model can be employed for estimation. Using this model, one estimates the probability that Z is equal to 1 in the combined data sample, conditional on a set of features (X), and then uses the estimation results to predict uplift for the cases within the population eligible for treatment (i.e., cases eligible to receive notices). The strengths of this approach potentially outweigh its weaknesses. Like the other approaches, the class-transformation method is straightforward to apply using standard classification learning methods, and it has been shown to outperform the two-model method in certain cases. While the standard class-transformation model does require a balanced treatment design involving random assignment of an equal number of cases to treatment and control, extensions to the methodology can be applied in cases involving unbalanced designs.*

* In the case where datasets are imbalanced, modifications to this approach can be applied. For further discussion, see Gutierrez and Gerardy (2016).

Estimating Uplift for Multiple Treatments

To this point, the discussion of uplift estimation methods has focused on applications involving a single treatment. However, results from past notice redesign pilots demonstrate that different groups of taxpayers may respond better to different versions of a notice. With multiple treatments, not only would the most promising taxpayers be selected to receive a notice, but the most effective notice version for each of those taxpayers would also be selected. This approach requires a generalization of the uplift estimation methods reviewed in the preceding section, known as multi-treatment uplift modeling. Under this generalization, the training sample includes samples from each of the treatment groups as well as the non-treatment group. A model is then trained to predict the uplift associated with each of the treatments. The prediction algorithm is applied to members of the eligible population, and the treatment associated with the highest predicted uplift is then identified for each population member.

Evaluating an Uplift Model

Because a single taxpayer cannot both receive a notice and not receive a notice, an uplift model must be evaluated in aggregate to determine its performance. We cannot know for certain whether any individual taxpayer was motivated to make a payment due to receiving their notice or whether they were going to make a payment anyway; therefore, we cannot know if our uplift model predicted that individual's uplift correctly. Instead, we look at individuals from the RCT and compare the aggregated results for taxpayers from the treatment and non-treatment groups which our uplift models scored similarly.

In uplift modeling literature, there are several different methods for evaluating uplift model performance, but they generally follow the same approach:

1. The uplift model is trained on a training set (using a two-model, single-model, class-transformation, or other uplift modeling approach).
2. The holdout test set (which the models were not trained on) is split into treatment and non-treatment groups.
3. The uplift model is used to predict the uplift scores for members of both groups in the holdout test set.
4. Both groups are sorted by predicted uplift, from highest to lowest.
5. The normalized, cumulative uplift is calculated at each coverage rate, through the following three steps:
   a. Calculate the cumulative target value at each unique value of uplift (for both the treatment and non-treatment groups).
   b. Normalize the cumulative target value by multiplying it with the corresponding proportion of individuals with the same treatment status up to that coverage rate.
   c. Calculate the normalized, cumulative target value for the treatment group minus the same value for the non-treatment group.
6. This provides an uplift curve which can be plotted and analyzed similarly to a receiver operating characteristic (ROC) curve.
7. The area under the uplift curve can be normalized by dividing it by the area under the "perfect" model curve.

This evaluation approach is notionally shown in Table 1 and Figure 3.

TABLE 1. Uplift Evaluation Example

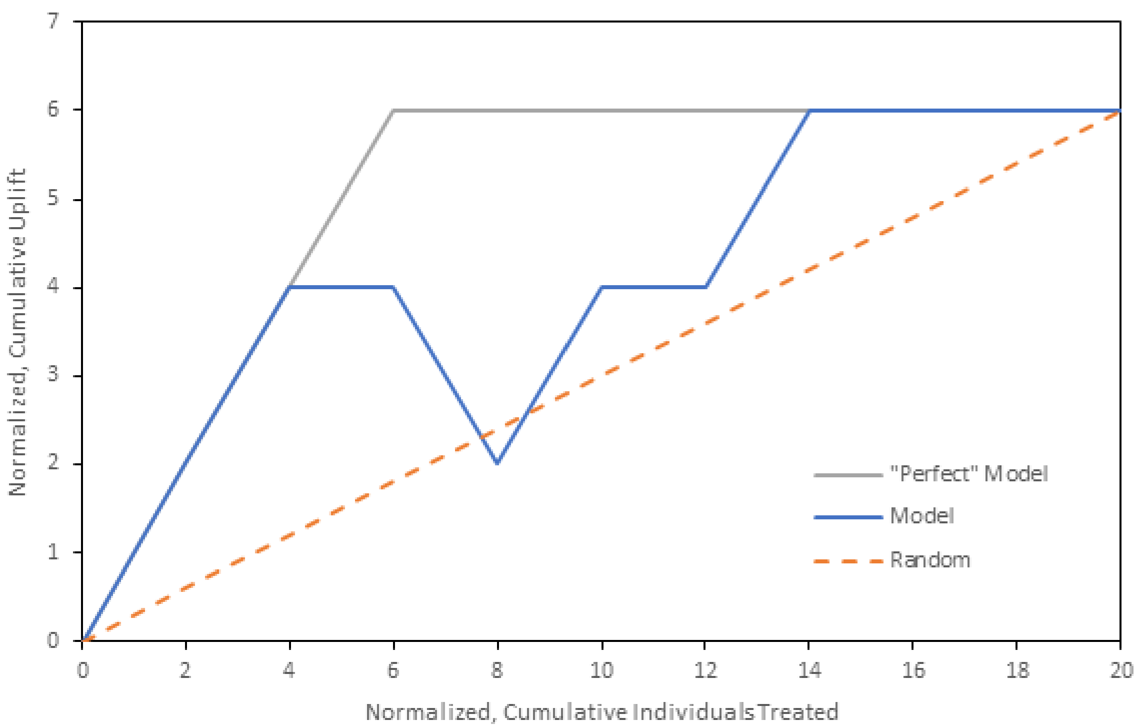
Treatment Group				Non-treatment Group				Normalized, Cumulative Uplift
Predicted Uplift	Target Value	Cumulative Target Value	Normalized, Cumulative Target Value	Predicted Uplift	Target Value	Cumulative Target Value	Normalized, Cumulative Target Value	
0.75	1	1	2	0.75	0	0	0	2
0.65	1	2	4	0.65	0	0	0	4
0.5	1	3	6	0.5	1	1	2	4
0.45	0	3	6	0.45	1	2	4	2
0.4	1	4	8	0.4	0	2	4	4
0.3	0	4	8	0.3	0	2	4	4
0.1	1	5	10	0.1	0	2	4	6
0	0	5	10	0	0	2	4	6
-0.15	0	5	10	-0.15	0	2	4	6
-0.2	0	5	10	-0.2	0	2	4	6

As described in the steps above, Table 1 shows both the treatment and non-treatment groups sorted by predicted uplift, from highest to lowest. For simplicity, the treatment group and the non-treatment group have the same set of predicted uplift scores; however, this methodology generalizes to groups of different sizes and with different distributions of predicted uplift.

The target value column indicates the specific value occurring for the individual in that row, and the cumulative target value is simply the cumulative sum of the target value column. The normalized, cumulative target value column is the normalization of the cumulative target value column—which in this example is equal to 2 times the cumulative target value, because the predicted uplift values and group sizes are the same. In practice, the coefficient used to normalize each record will vary depending on the distribution of predicted uplift values and group sizes between the two groups.

Finally, the normalized, cumulative uplift column is calculated as the difference between the treatment group and non-treatment group normalized, cumulative target variable. This column is plotted against the normalized, cumulative number of treated individuals, as shown in Figure 3.

FIGURE 3. Example Uplift Chart



In Figure 3, it is clear the "Perfect" model selection outperforms the Model selection, which in turn performs better than the Random selection. However, in comparing different models to one another or evaluating the improvement in performance of a single model while tuning hyperparameters, it is helpful to have a single numeric value instead of a graphical representation of the performance. In this case, the area under the uplift curve is used, typically normalized by dividing it by the area under the "Perfect" uplift curve, which provides a score similar to an ROC score. Unlike an ROC score, however, the normalized uplift Area Under the Curve (AUC) score can be greater than 1. This can occur due to individuals in the *Do Not Disturb* category, who are more likely to have a positive outcome if they are not treated. In our calculations of uplift AUC, we do not include any notion of *Do Not Disturb* individuals in the "Perfect" model, so our models can potentially have a score greater than 1. A score greater than 1 does not indicate a truly optimal model, as the optimal model score could be significantly higher than 1. It is also possible to normalize by a "Perfect" model which does include the notion of *Do Not Disturb* individuals (meaning that each individual in the non-treatment group who has a positive outcome is assumed to be a *Do Not Disturb* individual), which would ensure that the uplift AUC is always less than or equal to 1.

Prioritizing Cases Using Uplift

For each eligible taxpayer, the uplift model will estimate the predicted impact of receiving a notice on two outcomes—payment compliance and inbound telephone call resource utilization. To understand how cases will be selected under this approach, consider selection under the metric of the predicted increase in total dollars collected. Let $\hat{P}_i$ represent the predicted uplift in payments by taxpayer *i* when the taxpayer receives the notice, and let $\hat{C}_i$ represent the predicted uplift in inbound telephone call resources expended on taxpayer *i* as a result of the notice. The case selection objective is to find the subset of eligible taxpayers for whom the predicted uplift in overall payments is maximized subject to a specified threshold for notice-related inbound telephone call resources. This subset is identified as the solution to the following optimization problem:

$$\max_{a_1,a_2,\ldots,a_n} \sum_{i=1}^{n} a_i\hat{P}_i \text{ subject to } \sum_{i=1}^{n} a_i\hat{C}_i = \bar{C},$$

where $\bar{C}$ represents the total dollar threshold on notice-related inbound telephone call resource utilization, and are binary indicators for each of the n eligible taxpayers of whether they will be selected to receive a notice. The solution to this problem is represented by the following conditions:

$$a_j = \begin{cases} 1 & \text{if } \dfrac{\hat{P}_j}{\hat{C}_j} \geq \lambda \\ 0 & \text{otherwise} \end{cases} \quad j = 1, \ldots, n.$$

$$\sum_{i=1}^{n} a_j\hat{C}_j = \bar{C}.$$

The parameter λ represents the threshold for the marginal uplift in payment compliance per dollar (or hour) of notice-induced inbound telephone utilization that an eligible taxpayer must satisfy in order to be issued a notice under this selection scheme. The value of this parameter is implicitly determined by the equality constraint in the second equation. Intuitively, these conditions imply that one should rank cases in descending order of the predicted uplift in payments per dollar (or hour) of inbound telephone call utilization. One should then select cases in which to send notices in rank order, up to the point where the cumulative predicted uplift in inbound telephone call utilization for the selected cases is no longer below the specified threshold.

This method for prioritizing cases explicitly accounts for the predicted uplift in inbound telephone call resource costs associated with issuing a notice. Existing notice redesign studies consider such costs in an indirect way, through a subjective preference for notice variants and case selection approaches associated with a relatively low incidence of inbound telephone calls. In contrast, the methodology described above directly ranks cases in terms of their predicted uplift in payment compliance per dollar (or hour) of notice-related inbound telephone call utilization, thereby maximizing the expected uplift in payment compliance achievable with available resources. Moreover, since cases are ranked under this selection scheme according to their marginal payoff per dollar (or hour) of additional telephone resources required, it is straightforward to make real-time adjustments to the number of notices issued in response to changes in inbound telephone call resources.

Table 2 illustrates how cases are selected using the predicted uplift in dollars collected and telephone costs. Here, a notional cost constraint of $24 is imposed, with the goal of optimizing the total dollars collected up to that cost constraint. By calculating the ratio of predicted uplift in dollars collected to predicted uplift in call cost, and sorting by that ratio, an optimal case selection approach is achieved.

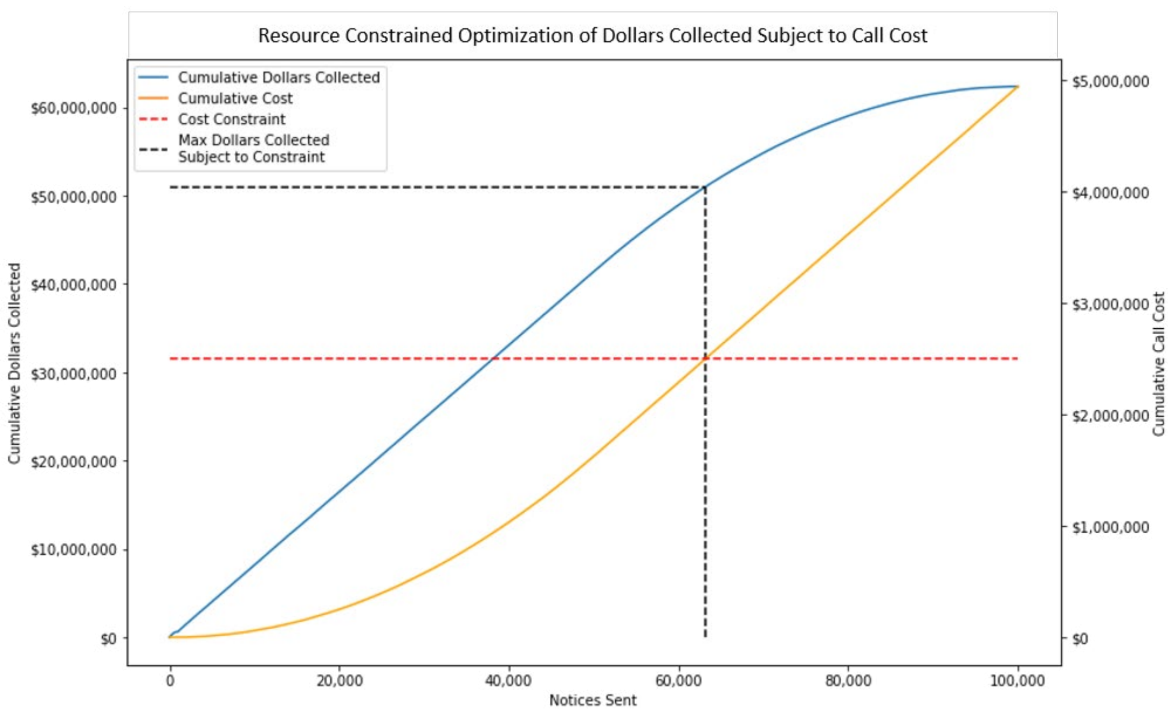
TABLE 2. Case Prioritization Example

Predicted Uplift in Dollars Collected	Predicted Uplift in Call Cost	Ratio
$50	$1	50
$10	$1	10
$15	$2	7.5
$100	$20	5
$20	$5	4
$15	$5	3
$1	$1	1
$1	$10	0.1

Figure 4 shows how this approach would work in practice. Here, a notional additional telephone call cost constraint (above the baseline call volume which would occur if no LT11 notices were sent) is imposed at $2.5 million, and the goal is to maximize the total additional dollars collected until that cost constraint is met. Cases are presorted by the ratio of predicted uplift in dollars collected to predicted uplift in call cost, so the optimal approach is achieved by simply determining how many of the available cases to treat. This is done by finding where the cumulative cost curve (orange line) intersects with the cost constraint (red dashed line). Then, the total number of notices to send can be calculated (black dashed line), as well as the predicted total dollars collected by finding where the cumulative dollars collected curve (blue line) intersects with the notices sent line (black dashed line).

In this notional example, given a cost constraint of $2.5 million, our case selection methodology would recommend issuing 63,000 notices to taxpayers, and those notices would result in $51 million in predicted collections.

FIGURE 4. Resource Constrained Optimization Example



2019 LT16 Model Test Pilot Review

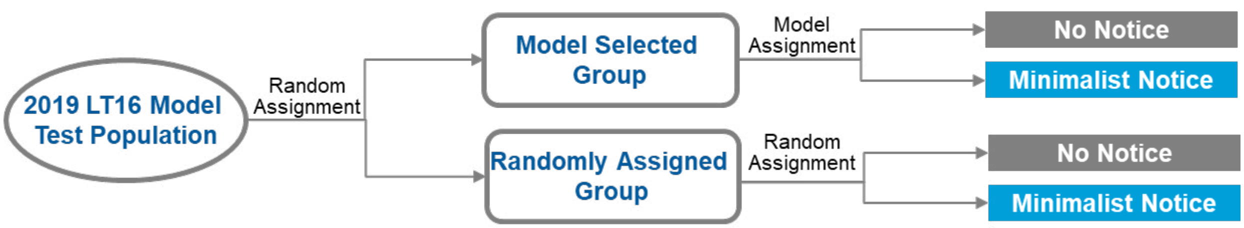
The LT16 Model Test Pilot provided the first look at using a model-based case selection approach to choose which taxpayers would be sent an LT16. Lessons learned from this pilot provided the starting point for further research into using predictive models to select taxpayers for treatment.

Pilot Overview

The LT16 Model Test Pilot developed a non-uplift modeling approach to attempt to improve upon the legacy ACS case selection methodology. A machine learning model was trained on data from the 2017 LT16 Notice Redesign Pilot, which only included taxpayers who had received an LT16. The lack of a non-treatment group precluded the ability to train an uplift model; thus, the model could only predict whether a given taxpayer was likely to full pay conditional on receipt of an LT16 notice (i.e., the model could not predict whether a certain taxpayer would full pay without receiving a notice at all). The case selection methodology used the outputs of this model to prioritize taxpayers for treatment by sorting them according to their predicted probability to full pay, from highest to lowest.

The 2019 LT16 Model Test Pilot was then run to evaluate the performance of this new, model-based case selection approach. As shown in Figure 5, taxpayers were first randomly assigned to either the model selected group or the randomly assigned group. Taxpayers in the model selected group were evaluated using the machine learning model to predict their probability of making a full payment, sorted from highest to lowest, and the top 50 percent of those taxpayers were selected to receive the Minimalist notice. In the randomly assigned group, 50 percent of taxpayers were randomly selected to receive the Minimalist notice.

FIGURE 5. 2019 LT16 Pilot Study Groups



Each week during the test, a new list of LT16-eligible taxpayers was generated to ensure that taxpayers involved in the experiment were LT16-eligible when they entered the test (notice eligibility and module[2] balance within ACS can change week to week). To balance baseline characteristics, taxpayers were randomly assigned with equal probability to either the model selected group or the randomly assigned group using a random number generator. Each week, the similarity and consistency of the two groups was verified across several characteristics (entity balance, number of open modules, entity age, age of debt, etc.) to ensure that any differences in outcomes could be attributed to the different case selection strategies.

LT16 Pilot Results

The LT16 Model Test Pilot was effective at predicting which taxpayers were likely to full pay within three months of receiving an LT16 notice. As seen in Table 3, in the model selected group, taxpayers who were selected to receive the Minimalist notice had an overall full pay rate of 31.74 percent, compared to 23.82 percent among the randomly assigned group, a difference of 7.92 percent. However, the overall full pay rate of the no notice recipients in the model selected group was only 12.74 percent, compared to the 21.15 percent among no notice recipients in the randomly assigned group, a difference of -8.41 percent. When comparing averages for

[2] A tax module refers to a single year's tax return as it moves through the IRS systems. A taxpayer in ACS may have one or several tax modules which have a balance due remaining, corresponding to one or several years' tax returns.

the model selected group overall (22.24 percent) and the randomly assigned group overall (22.49 percent), the model selected group performed slightly worse (-0.24 percent).

TABLE 3. 2019 LT16 Pilot Full Pay Outcome Comparison

Study Group	Overall Full Pay Rate	Full Refund Offset Full Pay Rate	Partial Refund Offset Full Pay Rate	Non Refund Offset Full Pay Rate	Other Full Pay Rate
Model Selected Group					
Minimalist	31.74%	18.19%	7.17%	5.36%	1.95%
No Notice	12.74%	4.43%	2.27%	4.80%	2.58%
Approx. Average	22.24%	11.31%	4.72%	5.08%	2.27%
Randomly Assigned Group					
Minimalist	23.82%	11.53%	4.75%	6.45%	2.19%
No Notice	21.15%	11.45%	4.65%	4.11%	2.03%
Approx. Average	22.49%	11.49%	4.70%	5.28%	2.11%
Study Group Comparison (model selected average minus randomly assigned average)					
Difference	**-0.24%**	**-0.18%**	**0.02%**	**-0.20%**	**0.16%**

As shown in Table 4, refund offsets were responsible for a large portion of full payments among both the model selected group and the randomly assigned group. The model selected group minimalist recipients had a far higher rate of full pays due to refund offsets (18.19 percent) than the no notice recipients in that same group (4.43 percent), or either the Minimalist (11.53 percent) or no notice recipients (11.45 percent) in the randomly assigned group. The difference in refund offset full pay rates between those selected by the model to receive a Minimalist notice and those selected to receive no notice was 13.76 percent, which indicates the model was selecting taxpayers who were far more likely to receive a refund offset than the baseline. While this finding did not support the core goal of the pilot (i.e., identifying taxpayers who would be positively impacted by a LT16 treatment), it demonstrated the IRS's ability to predict refund offsets.

TABLE 4. 2019 LT16 Pilot Refund Offset Comparisons

Study Group	Full Refund Offset Full Pay Rate	Partial Refund Offset Full Pay Rate
Model Selected Group		
Minimalist	18.19%	7.17%
No Notice	4.43%	2.27%
Difference (Minimalist—No Notice)	13.76%	4.90%
Randomly Assigned Group		
Minimalist	11.53%	4.75%
No Notice	11.45%	4.65%
Difference (Minimalist—No Notice)	0.08%	0.10%
Study Group Comparison (model selected difference minus randomly assigned difference)		
Overall Difference	13.68%	4.80%

This result pointed to the logical next step of conducting the uplift modeling effort described in this paper. If taxpayers who are going to have a refund offset applied to their account can be accurately predicted, the IRS can focus its limited notice resources on other taxpayers who are more likely to be influenced by receipt of the notice.

Enhanced Modeling Approach

Taking the lessons learned from the prior LT16 Model Test Pilot, an enhanced uplift modeling approach was developed for the current study of case selection for LT11 notice receipt. This approach involved three main components: 1) a refund offset model to predict whether a taxpayer is likely to have a refund offset applied to their account (and thus should not be selected for a notice); 2) an uplift model to predict the uplift in dollars collected upon receipt of a notice; and 3) an uplift model to predict the uplift in incoming telephone calls upon receipt of a notice, which is paired with a cost constraint to identify how many notices can be sent in a given period of time. These three components were developed and experimentally tested on historical data, as described in the following sections.

Refund Offset Modeling Effort

Data Used

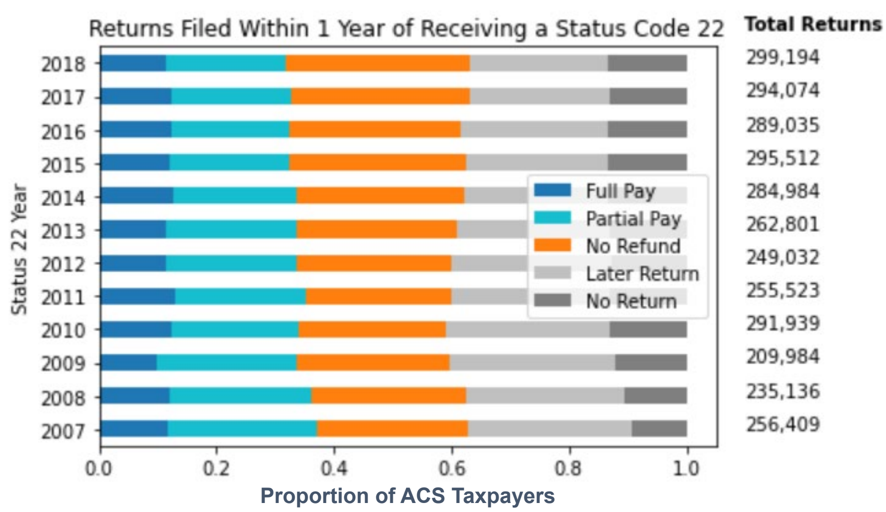
We examined historical data for ACS taxpayers from 2007 to 2018. Tax years from 2019 onward were not included in this general analysis, as our goal was to evaluate trends and build a model to predict refund offsets during normal tax years, rather than the anomalous ones generated by the COVID-19 pandemic.

To this end, we collected historical data on individual taxpayers who entered ACS between 2007 and 2018. We analyzed their behavior during the 365-day period after one of their modules entered ACS (received a status code (ST) 22), which indicates that a taxpayer has entered ACS.

Exploratory Analysis

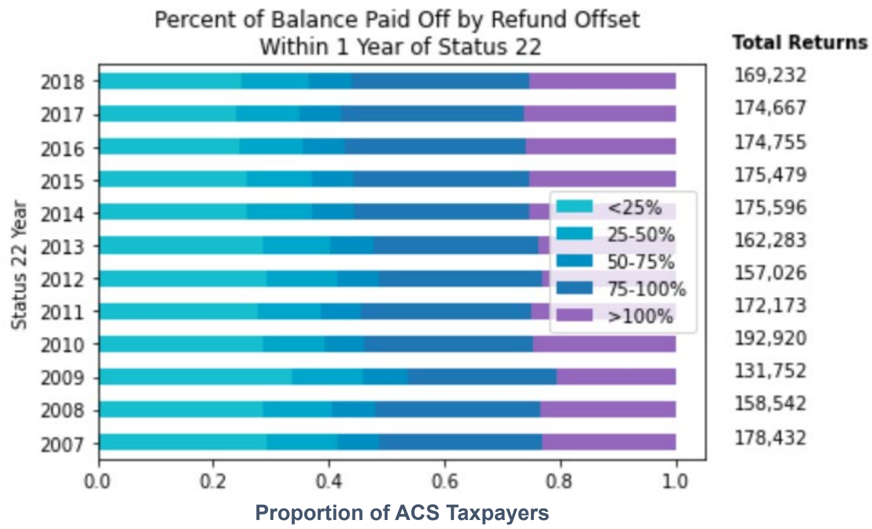
Initial exploration of the data focused on determining the proportion of ACS taxpayers that received any tax refund in the 365-day period after receiving an ST 22 on any module (Figure 6). Though roughly 75-80 percent of the general population received a tax refund during this time period,[3] our analysis indicates that less than 40 percent of ACS taxpayers did so over the same period (see Figure 6). Furthermore, of those who received a refund, approximately 10 percent received a refund large enough to pay off their full balance within a given year. The distribution of refund amounts as a percent of the taxpayers' account balance for taxpayers who received a refund is shown in Figure 7. Of the remaining ~60 percent of taxpayers, roughly 25-30 percent filed a return with a balance due, and the others filed no returns within one year of having a module enter ACS.

FIGURE 6. Returns Filed Within 1 Year



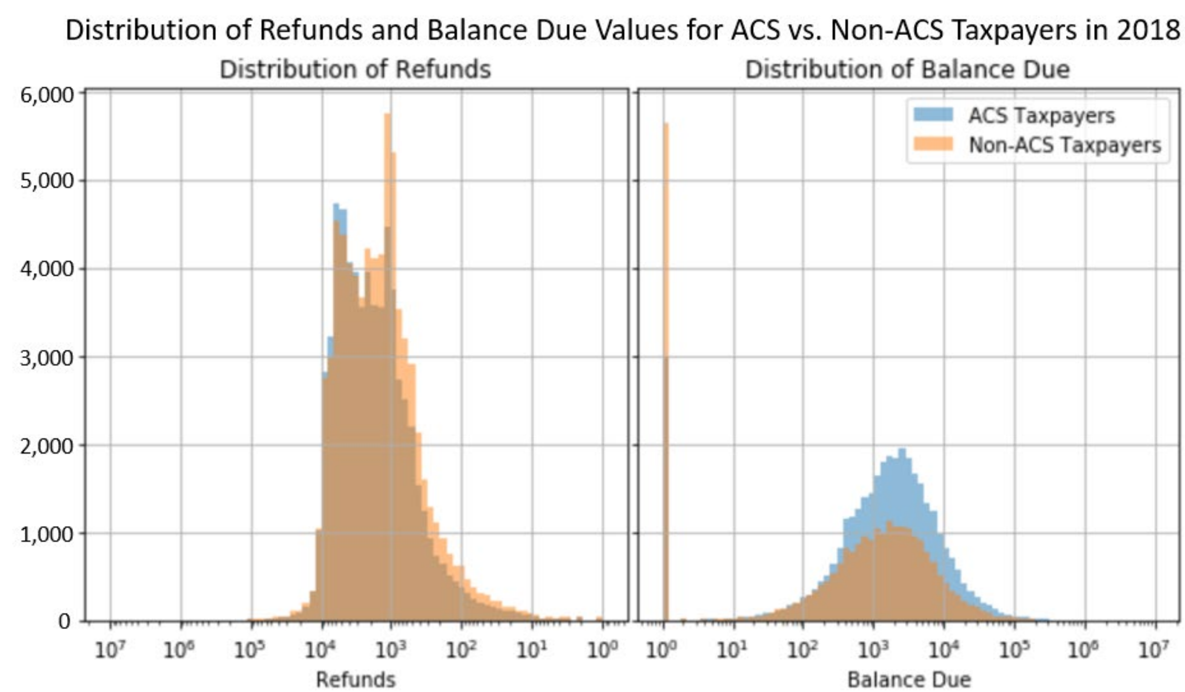
[3] Detailed analysis of the general population is not shown. General population represents all individual taxpayers with a filing requirement from 2007 to 2018, where a consistent proportion of taxpayers typically file returns with a tax refund each year.

FIGURE 7. Percent of Balance Paid Off by Refund Offset



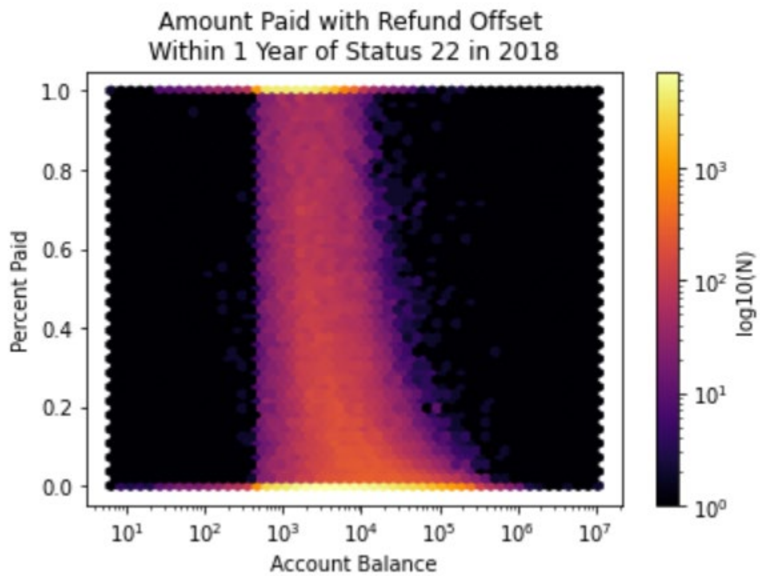
Further analysis comparing the distribution of returns filed between the general population and ACS taxpayers indicates that returns filed by ACS taxpayers are more likely to have a balance due, and the amount they owe is likely to be larger than among the general population. Though the distribution of refunds between the ACS population and the general population is similar, ACS taxpayers are less likely to have a zero-balance return than non-ACS taxpayers. These distributions were analyzed for the same time period as shown in the above graphic with consistent results. (For simplicity, only one year is shown in Figure 8.)

FIGURE 8. Distribution of Refunds and Balance Due Values



Since the account balance of taxpayers in ACS varies greatly, we also examined the relationship between account balance and percent of account balance potentially paid off by a tax refund within a 365-day period after entering ACS. As shown in Figure 9, taxpayers with a smaller account balance are more likely to reduce or eliminate that balance through a tax refund within one year. However, most taxpayers will either have a refund large enough to pay off 100 percent of their account balance or have no refund at all, leaving relatively few taxpayers in between.

FIGURE 9. Refund Offset Amount vs. Account Balance



Our analysis indicates that roughly 10 percent of ACS taxpayers can be excluded from LT11 notice selection because they are expected to receive a tax refund large enough to pay off their account balance. This number is significant enough to make a difference in the ultimate Uplift model, provided the refund offset prediction model can accurately distinguish between taxpayers who will and will not receive a large refund within 365 days.

Modeling Techniques

Models were trained on historic tax data for taxpayers in ACS. Since production models will ultimately be used to predict future behavior based on the previous behavior of other taxpayers, models were trained on one year of historic data and tested on the following year. Taxpayers selected for this modeling effort had a module receiving an ST 22 within the past year (e.g., models trained on 2015 data were trained on a sample of modules having an ST 22 in 2015, and then tested on a sample of modules having an ST 22 in 2016). Later, models were tested further in the future to evaluate model robustness.

Our exploratory analysis suggests that taxpayers tend to file one type of return (with either a balance due or a refund) year after year, only rarely changing behavior. Thus, a baseline model was developed to predict whether a taxpayer would file a refund return within the next year based on their most recent return. This simple Logistic Regression model assigned a probability based on the magnitude of a taxpayer's most recent balance due or refund. It achieved an ROC AUC of 0.665.

Next, we developed a similar model to predict which taxpayers would receive a refund large enough to pay off their full account balance within 365 days of entering ACS. Much like the baseline model, this model included only a single feature—the balance due or refund amount from the taxpayer's most recent return. The Logistic Regression model was able to predict whether a taxpayer would receive a refund large enough to pay off their account balance with an ROC AUC of 0.737; thus, a high level of predictive power comes from the most recent year's tax return.

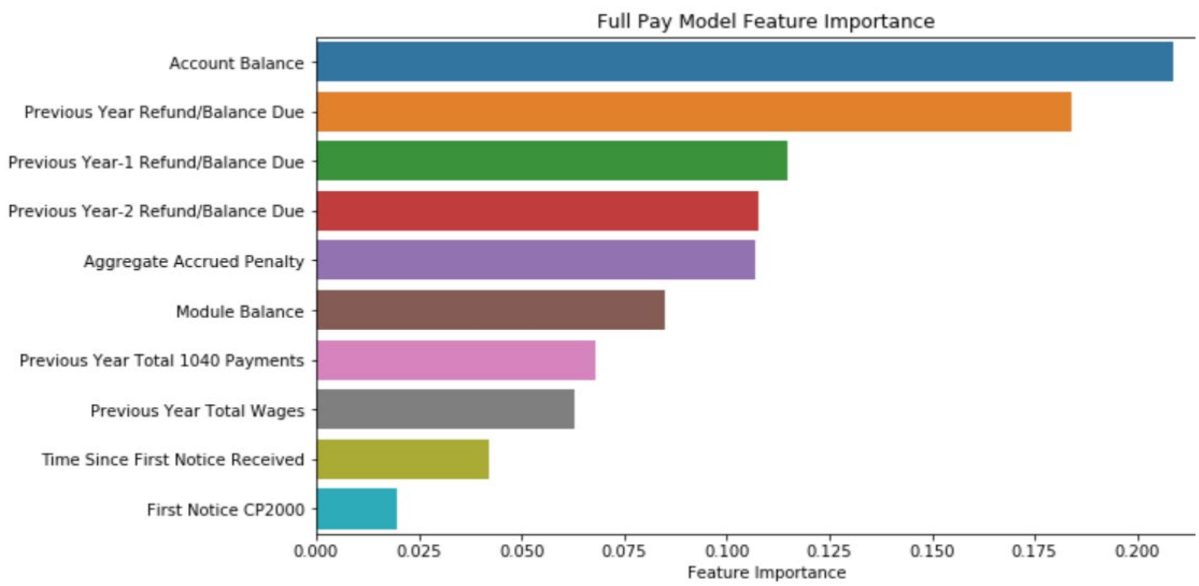
To develop a production model, two model types (Logistic Regression and Random Forest Classifier), several feature scalers, and a variety of additional features were tested. During this process, additional historic data were added to the feature set, including up to ten years of a taxpayers' previous returns, their account information (such as account balance and accrued interest), and which notice they first received. As in the uplift modeling, categorical features were One Hot Encoded. This broad feature set was reduced using Recursive Feature Elimination (RFE) and further reduced by removing features with a high Variance Inflation Factor (VIF) to eliminate multicollinearity in the dataset.

Within this reduced feature set, missing values were imputed with the constant value of zero and hyperparameters were optimized on ROC AUC score with cross-validation within the training year.

Modeling Results

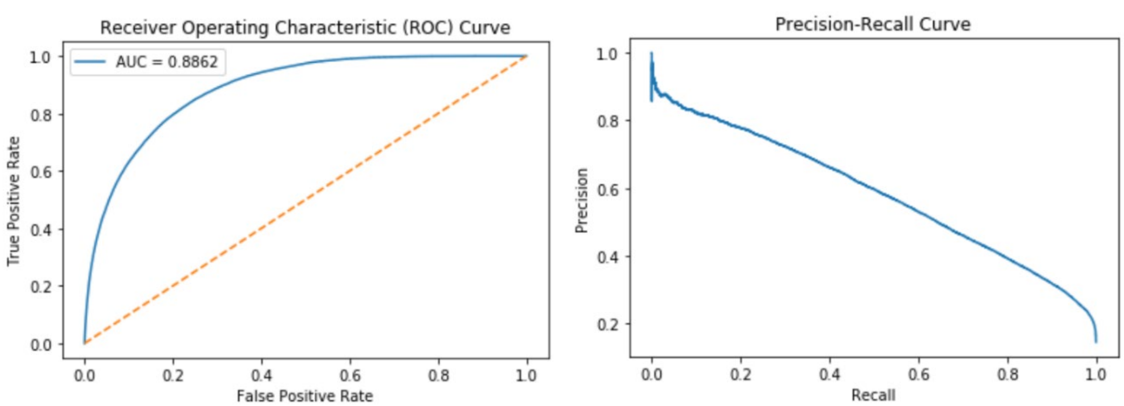
The final set of ten features and their respective importance in the Random Forest Classifier is shown in Figure 10.

FIGURE 10. Full Pay Model Feature Importance



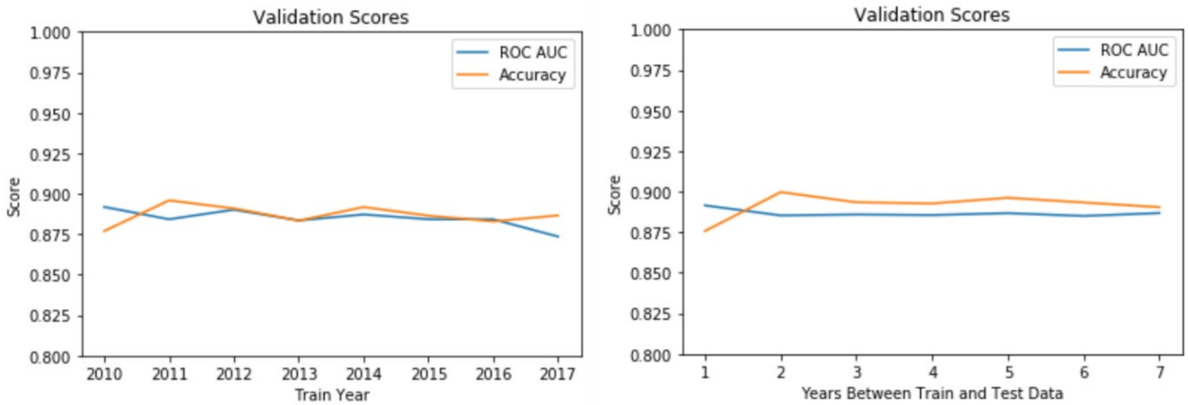
With a training year of 2015 and a testing year of 2016, this model achieved an ROC AUC score of 0.8861 and an accuracy score of 0.8863 (see Figure 11).

FIGURE 11. ROC and Precision-Recall Curves



As shown in Figure 12, the ROC AUC scores remained between 0.875 and 0.900 in subsequent models, regardless of which year was used for training data, when predicting one year in the future. Furthermore, a model trained on 2010 data remained relevant when used to predict results up to seven years in the future, demonstrating thatthese models are robust and not sensitive to the year training data are selected from, barring any significant economic events.

FIGURE 12. Year-Over-Year Validation Scores



LT11 Uplift Modeling Effort

Following on the notice redesign and case selection efforts, a new uplift modeling effort was undertaken to develop an optimized case selection methodology for taxpayers in ACS. Data from the LT11 Notice Redesign Pilot were used to train uplift models for both total dollars collected and incoming telephone calls. These models will eventually be used in a case selection pilot of taxpayers eligible to receive an LT11 to experimentally determine the effectiveness of this uplift modeling approach.

Data Used

While LT11 notices are issued to both individual and business taxpayers, for this iteration of the uplift modeling effort only individual taxpayers from the LT11 Notice Redesign Pilot were considered. In order to maximize dollars collected up to a certain cost constraint, the total dollars collected within 365 days and the probability of calling the IRS within 365 days are the target variables used in the two uplift models. Table 5 shows the distribution of modules and taxpayers between the different test groups along with the target variables of interest for those groups.

TABLE 5. LT11 Pilot Characteristics and Outcomes

Notice Version*	Modules	Taxpayers	Average Dollars Collected in 365 Days	Proportion Who Call Within 365 Days
LT11A	16,880	14,488	$1,947	42.8%
LT11C	17,018	14,586	$1,959	36.3%
LT11D	16,969	14,535	$1,867	36.1%
LT11E	16,955	14,587	$1,903	35.4%
LT11F	17,086	14,538	$1,954	36.3%
No LT11	292,896	252,795	$1,611	19.5%

* See Appendix C for descriptions of LT11 notice prototypes.

The non-treatment population was identical to the population which received one of the LT11 variants, so it is possible to evaluate in the aggregate the incremental impact of receiving LT11s on taxpayers in comparison to not receiving a notice. In order to simplify the modeling and eventual pilot effort, the LT11 variant data were condensed into two groups: 1) LT11 redesign recipients (LT11C, LT11D, LT11E, and LT11F); and 2) taxpayers who did not receive an LT11. This led to the two groups of modules and taxpayers (see Table 6), which were used to train the uplift models.

TABLE 6. LT11 Pilot Condensed Dataset

Notice Version	Modules	Taxpayers	Average Dollars Collected in 365 Days	Proportion Who Call Within 365 Days
LT11	68,028	58,246	$1,921	36.0%
No LT11	292,896	252,795	$1,611	19.5%

Exploratory Analysis

Prior to modeling, exploratory analysis was performed to investigate available features for potential links to the target variables and to aid in determining the predictive power of the dataset. During the exploration phase, we examined several outcomes of receiving a notice, including 1) how many times a taxpayer calls the IRS; 2) how much money the taxpayer pays; 3) whether the taxpayer resolves the module and moves to ST 12; 4) whether the taxpayer enters into an IA; and 5) whether a taxpayer pays at least 95 percent of the balance due. Outcomes were examined within different time frames from 35 days to 365 days.

After 35 days, LT11 recipients paid (whether voluntarily or involuntarily) almost 250 percent more than non-LT11 recipients. This decreases to 83 percent more at 180 days and only 18 percent more at 365 days. Similarly, LT11 recipients made nearly 400 percent more telephone calls to the IRS within 35 days after receiving a notice than non-LT11 recipients. This decreases to 240 percent more at 180 days and 120 percent more at 365 days.

In light of the results (i.e., total dollars received), much of the additional exploration in preparation for the modeling effort was directed toward refund offsets, as described in the Refund Offset Modeling Effort Exploratory Analysis section above.

Modeling Techniques

Following the exploratory analysis effort, initial modeling attempts focused on the two-model uplift modeling approach, though alternative approaches may be explored in the future. A complete feature set was built from the data explored in the phase described above, containing only historical data for the LT11 pilot modules from the period before the notice was sent. Target variables were generated from the relevant outcomes after a 365-day period. Though the length of this period makes it more difficult to discern if outcomes such as voluntary payments or phone calls are a direct result of the notice, it ensures thatany refund offset applied to the taxpayer's account will be included as an outcome. As our models were built to predict the potential of a refund offset being applied, accounting for this is essential to measuring model success.

The broad feature set containing historical 1040 data, historical notice data, and account information (such as account balance and penalties accrued) was split into three groups for training, validation, and testing of models. Categorical features, such as initial notice received, were One Hot Encoded to generate binary features for model training. The training set was used to train two separate Random Forest models with cross validation. One model was trained only on taxpayers who received a notice during the LT11 Notice Redesign Pilot, while the other was trained on taxpayers who received no notice.

The two-model uplift approach did not allow for direct optimization of uplift itself during model training, and instead several model metrics were assessed for their effectiveness in generating the best uplift on the validation set. Once the metric corresponding to the highest validation uplift was determined, hyperparameters for the two Random Forest models were separately optimized for that metric.

The optimum number of features for the models was determined through Recursive Feature Elimination, which trained a secondary Extra Trees model and generated a model score based on the number of features included. The precise number of features was found to be variable between different outcomes, as was the specific features with the highest predictive power. However, consistently important features included current account balance, previous years' tax refund or balance due amounts, and the receipt of a CP2000 notice.[4]

Once important features were selected and model hyperparameters were tuned with a cross validated Randomized Search to optimize uplift on the validation set, the final models were then tested against the hold-out test set to determine how well the model would generalize to previously unseen data.

Modeling Results

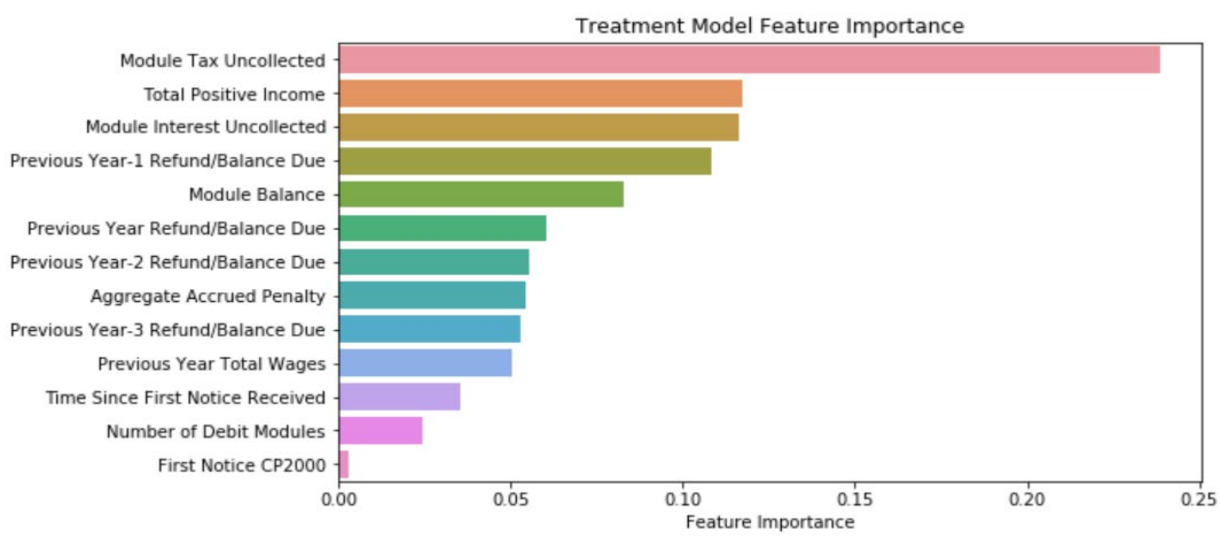
LT11 Total Dollars

To predict the total dollars a taxpayer will pay within the 365-day period following receipt of an LT11 notice, two Random Forest Regression models were trained using the approach outlined above. Additionally, data preprocessing and several target transformation methods were tested and assessed for their ability to produce a higher uplift.

As total dollars collected can vary greatly and can be difficult to predict accurately, several approaches to decrease the range of the target variable were tested, including predicting total dollars as a percent of account balance, log-transforming the target variable, winsorizing the target variable, predicting a binary target for any dollars collected, and dropping outliers within the top 1 percent of all account balances. Among these approaches, winsorizing the target variable produced the greatest boost to uplift between the models.

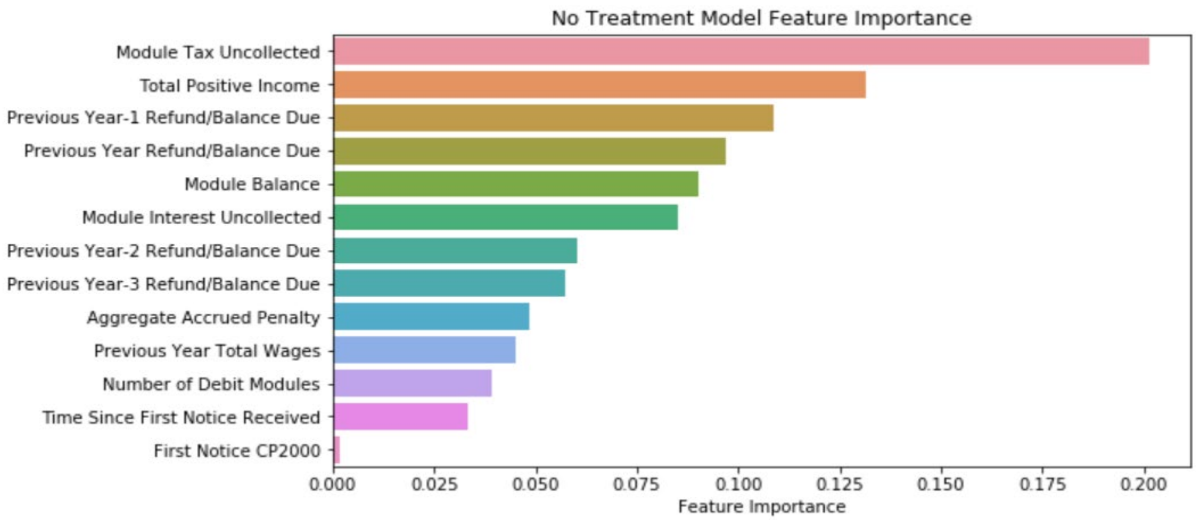
The highest performing uplift model consists of two Random Forest Regression models trained on a winsorized total dollars target variable. Each model utilized the thirteen features determined to have the most predictive power, shown in Figure 13 and Figure 14.

FIGURE 13. Treatment Model Feature Importance



[4] CP2000 notices are issued to taxpayers who underreported income on their prior year Form 1040 filing, which caused an increase or decrease to their balance.

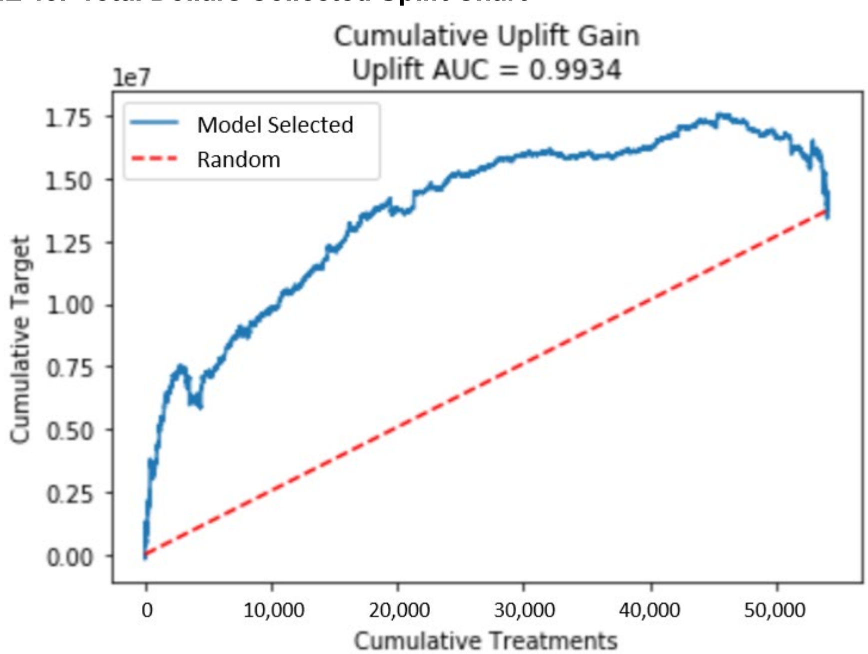
FIGURE 14. No-Treatment Model Feature Importance



The model trained on the treatment group (i.e., the subset of taxpayers who received an LT11 notice) resulted in an average mean squared error of -3.95E6 on the cross-validation test set, while the non-treatment model (trained on the taxpayers who received no notice) had an average mean squared error of -3.37E6. Though these large mean squared error values indicate a large discrepancy (on the order of 1 million) in the average squared difference between the predicted and actual values for both models, the mean squared error is not the desired optimization metric. Individually, these metrics are not as significant as the final uplift. When combined into a single uplift model and tested on the validation set, these models produced an uplift AUC of 1.03.

The models as trained above were then used to make predictions on the holdout test set, previously unseen by the models. The uplift AUC on the holdout set was calculated to be .9934 as shown in Figure 15.

FIGURE 15. Total Dollars Collected Uplift Chart



The small decrease in uplift AUC between the validation set and the holdout test set indicates that this model should generalize well to previously unseen data without a significant drop in performance. The large gap between the randomly selected group (red dashed line) and the model selected group (blue solid line) in Figure 15 above indicates that using this pair of models to determine which taxpayers to send an LT11 notice will have a significant effect on the total dollars collected. For example, if 30,000 LT11 notices were sent, selecting the taxpayers with this uplift model would result in 8 million more dollars being collected than if the notices were sent randomly.

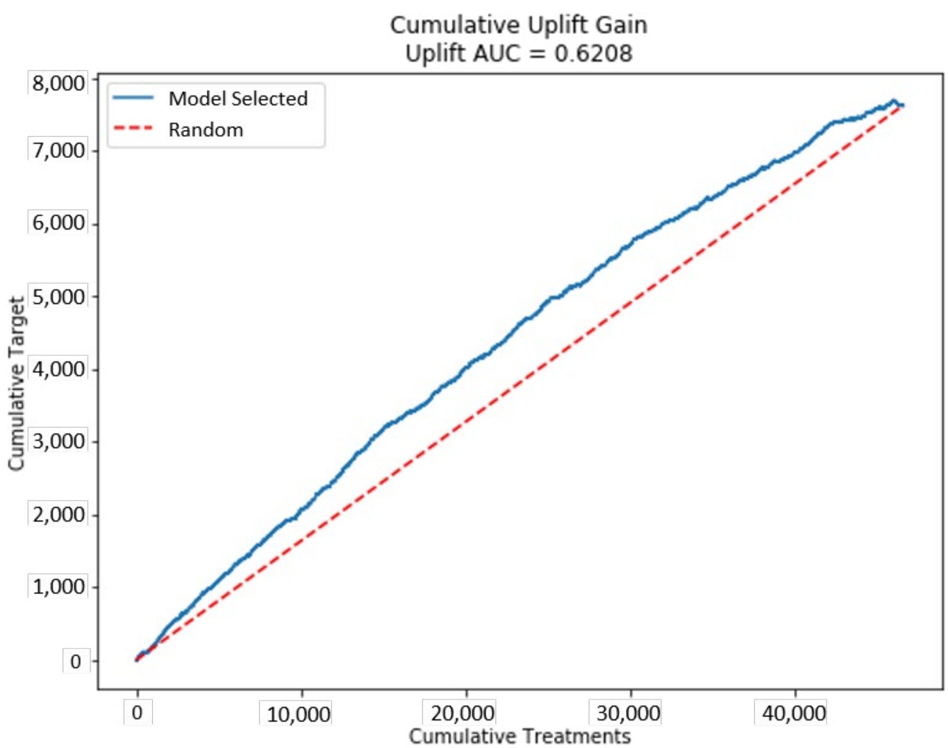
LT11 Telephone Calls

To predict the probability of calling the IRS within the 365-day period following receipt of an LT11 notice, two Random Forest classification models were trained using the uplift modeling approach described above.

The model trained on the treatment group (i.e., the subset of taxpayers who received a LT11 notice) resulted in an ROC score of 0.630 on the cross-validation test set, while the non-treatment model (trained on the taxpayers who received no notice) had an ROC score of 0.650. When combined into a single uplift model and tested on the validation set, these models produced an uplift AUC of 0.6315.

The models trained, as described above, were then used to make predictions on the holdout test set, previously unseen by the models. The uplift AUC on this holdout set was calculated to be .6208.

FIGURE 16. Incoming Telephone Calls Uplift Chart



As shown in Figure 16, the uplift AUC of .6208 and corresponding gap between the randomly selected group (red dashed line) and the model selected group (blue solid line) indicates that there is some predictive power in the model, and it can be used to identify taxpayers more likely to call the IRS after receipt of an LT11. While comparing uplift scores or curves from different models usually requires a lot of caveats, the predictive power of the telephone call uplift models is less than that of the total dollars collected models.

Discussion

Based on the model results presented above, uplift modeling is a promising approach to optimizing the ACS case selection processes. This is demonstrated in an uplift curve for total dollars collected by the LT11, which shows that sending notices to approximately one third of the eligible taxpayers can result in nearly the same amount of dollars collected as sending the notice to all eligible taxpayers. While the uplift models for incoming telephone calls showed a smaller potential benefit, they too indicate that using uplift modeling would enable more productive allocation of the IRS's limited resources.

While uplift modeling in these experiments was only possible with access to data from a prior RCT that included a non-treatment group, other methods could be used to achieve some of the improvement shown here without a dedicated RCT. The promising results from the refund offset modeling show it is possible to identify taxpayers who are likely to self-cure, even if an RCT with a non-treatment group is not available. There are also a variety of causal inference methodologies which can be employed using data not derived from a dedicated RCT, provided there exists both a treatment and non-treatment group. For example, these methodologies could be employed on historical ACS data, where there is a substantial portion of taxpayers who are eligible to receive an LT11 or LT16 but do not receive one due to resource constraints. Because the non-treatment groups in these populations are not necessarily drawn from the same underlying distribution as are the treatment groups, care must be taken to account for differences in the characteristics of those populations, but it is possible to draw some conclusions from this type of data and potentially improve on current methodologies using uplift modeling.

Future Case Selection Pilot Test

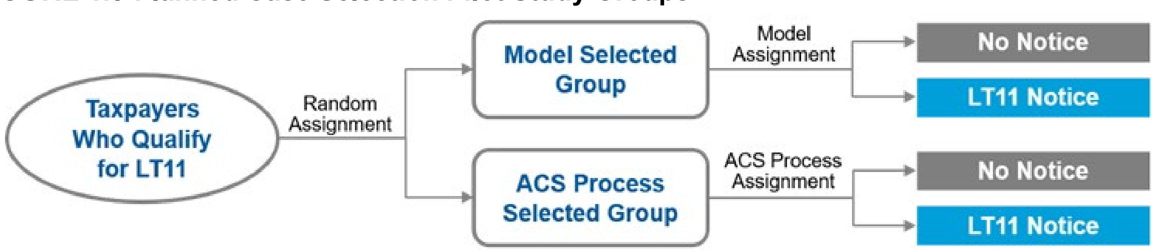
The uplift modeling results discussed here indicate there is a considerable amount of predictive power available to predict the uplift in total dollars collected for potential LT11 recipients. A case selection methodology which makes use of these uplift models will likely outperform traditional case selection methods. To test this claim, the IRS plans to conduct a pilot test to compare ACS's current approach to case selection against one based on the uplift models presented above.

Pilot Test Plan

An RCT will be employed to test whether a new case selection method based on uplift modeling outperforms the existing methodology. To facilitate this test, the pilot will include one experimental arm for the existing approach, and one arm for the new case selection methodology.

To assess the relative performance of the various arms of the study, we will randomly allocate taxpayers eligible to receive an LT11 to each arm. In this way, the two case selection methods will be applied to a comparable group of individuals as shown in Figure 17. Each arm will use its associated case selection method to rank taxpayers and select which within their study arm will receive a notice.

FIGURE 17. Planned Case Selection Pilot Study Groups



Case Selection Evaluation Approach

The two case selection approaches will be evaluated by comparing the total dollars collected by each approach, up to a specified cost constraint. In order to compare the dollars collected at the same associated cost for the two arms, the specified cost constraint should be less than or equal to the total cost associated with the arm with the lowest cost.

To account for differences in costs between the two case selection approaches, only a portion of the recipients in one or both groups will be used in the evaluation calculations. This will be determined in the following manner:

1. Cases are ranked in the order in which they would be selected by their respective case selection methodology.
2. The total dollars collected, and the total cost associated with the next highest ranked case (in which the taxpayer received a notice) and the next lowest ranked case (in which the taxpayer did not receive a notice) are iteratively added to the total for each methodology.
3. This is done until the total cost associated for that methodology has reached the imposed cost constraint.

Table 7 shows a notional pilot example, comparing the case selection methodologies of an uplift model selected group and an ACS process selected group. Here, the notional cost constraint is $20, and the point at which each group reaches that cost constraint is shown at the red dashed line. The total dollars collected, and total cost correspond to the amount added by both the next highest ranked case (in which the taxpayer is issued a notice) and the next lowest ranked case (in which the taxpayer is not issued a notice). For the uplift model selected group, the cumulative dollars collected is $60 at the point that reaches the cost constraint which corresponds to five notices issued. For the ACS process selected group, the cumulative dollars collected is $40 at the point that reaches the same cost constraint, which corresponds to only four notices issued as the cost associated with each notice is higher among the ACS process selected group in this example.

Thus, the relative uplift of the proposed case selection methodology in this scenario would be (60 – 40)/40 = 50 percent.

TABLE 7. Proposed Case Selection Evaluation Example

	Notices Issued	Total Dollars Collected	Total Cost	Cumulative Cost	Cumulative Dollars Collected	Relative Uplift in Total Dollars
Model Selected Group	1	20	5	5	20	50%
	2	15	5	10	35	
	3	10	3	13	45	
	4	5	2	15	50	
	5	10	5	20	60	
	6	5	3	23	65	

	Notices Issued	Total Dollars Collected	Total Cost	Cumulative Cost	Cumulative Dollars Collected	Relative Uplift in Total Dollars
ACS Process Selected Group	1	10	10	10	10	-
	2	10	3	13	20	
	3	5	4	17	25	
	4	15	3	20	40	
	5	5	5	25	45	
	6	2	3	28	47	

The two case selection methodologies will be evaluated against one another using this approach to determine whether the results seen in uplift modeling efforts to date can generalize to real world applicability.

References

Gutierrez, P., and J.-Y. Gerardy (2016), "Causal Inference and Uplift Modeling: A Review of the Literature," *JMLR: Workshop and Conference Proceedings* 67.

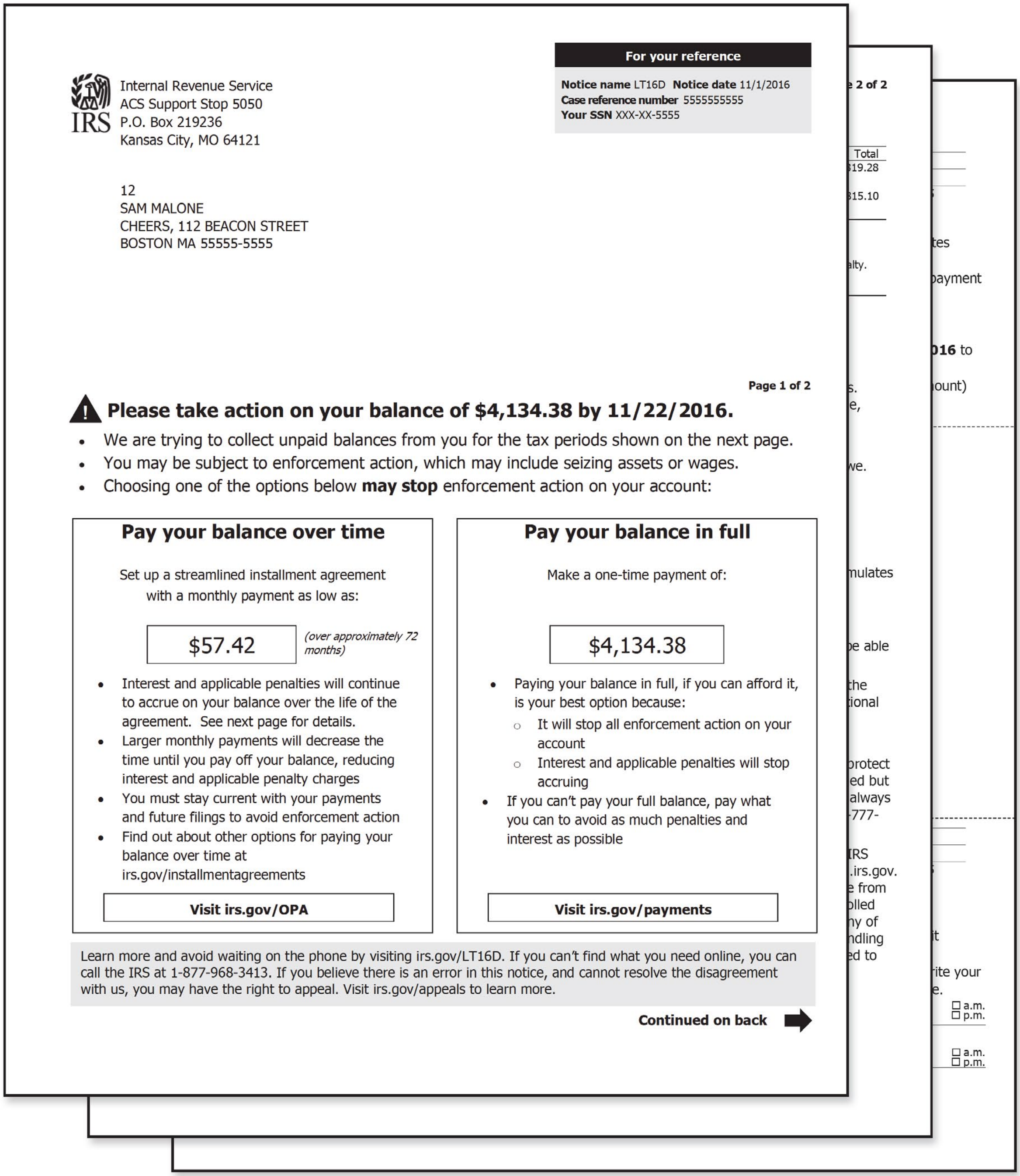
Herlache, A., J. Millard, A. Miller, and M. Theel (2018), "Using Behavioral Insights in Notice Design To Improve Taxpayer Responses and Achieve Compliance Outcomes," *2018 IRS Research Bulletin* (Publication 1500).

Appendix A. Acronyms

Acronym	Definition
ACS	Automated Collection System
AUC	Area Under the Curve
BI	Behavioral Insights
BMF	Business Master File
IA	Installment Agreement
IMF	Individual Master File
IRS	Internal Revenue Service
OPA	Online Payment Agreement
RCT	Randomized Controlled Trial
RFE	Recursive Feature Elimination
ROC	Receiver Operating Characteristic
TIN	Taxpayer Identification Number
VBD	Voice Balance Due
VIF	Variance Inflation Factor

Appendix B. LT16 Notice Prototypes

LT16 Installment Agreement (IA) Notice



For your reference

Notice name LT16D **Notice date** 11/1/2016
Case reference number 5555555555
Your SSN XXX-XX-5555

Internal Revenue Service
ACS Support Stop 5050
P.O. Box 219236
Kansas City, MO 64121

12
SAM MALONE
CHEERS, 112 BEACON STREET
BOSTON MA 55555-5555

Page 1 of 2

Please take action on your balance of $4,134.38 by 11/22/2016.

- We are trying to collect unpaid balances from you for the tax periods shown on the next page.
- You may be subject to enforcement action, which may include seizing assets or wages.
- Choosing one of the options below **may stop** enforcement action on your account:

Pay your balance over time

Set up a streamlined installment agreement with a monthly payment as low as:

$57.42 *(over approximately 72 months)*

- Interest and applicable penalties will continue to accrue on your balance over the life of the agreement. See next page for details.
- Larger monthly payments will decrease the time until you pay off your balance, reducing interest and applicable penalty charges
- You must stay current with your payments and future filings to avoid enforcement action
- Find out about other options for paying your balance over time at irs.gov/installmentagreements

Visit irs.gov/OPA

Pay your balance in full

Make a one-time payment of:

$4,134.38

- Paying your balance in full, if you can afford it, is your best option because:
  - It will stop all enforcement action on your account
  - Interest and applicable penalties will stop accruing
- If you can't pay your full balance, pay what you can to avoid as much penalties and interest as possible

Visit irs.gov/payments

Learn more and avoid waiting on the phone by visiting irs.gov/LT16D. If you can't find what you need online, you can call the IRS at 1-877-968-3413. If you believe there is an error in this notice, and cannot resolve the disagreement with us, you may have the right to appeal. Visit irs.gov/appeals to learn more.

Continued on back

LT16 Minimalist Notice

IRS
Internal Revenue Service
ACS Support Stop 5050
P.O. Box 219236
Kansas City, MO 64121

For your reference

Notice name LT16A **Notice date** 11/1/2016
Case reference number 5555555555
Your SSN XXX-XX-5555

12
SAM MALONE
CHEERS, 112 BEACON STREET
BOSTON MA 55555-5555

Page 1 of 2

Please take action on your balance of $4,134.38 by 11/22/2016.

We are trying to collect unpaid balances from you for the tax periods shown on the next page.

Your account may be subject to enforcement action, which may include seizing assets or wages.

Following the instructions under the "What you need to do" section **may stop** enforcement action.

Please visit irs.gov/LT16A for more information.

What you need to do right now

- **Pay as much of your balance as you can now**: Visit irs.gov/payments to pay online or mail in a check or money order with the payment stub below.
- **If you can't pay in full right now:** You can pay your remaining balance over time if you are current with your filing obligations. Visit irs.gov/OPA to learn more.
- **If you are currently facing financial hardship:** See the next page to learn about options available to you.

Learn more and avoid waiting on the phone by visiting irs.gov/LT16A. If you can't find what you need online, you can call the IRS at 1-877-968-3413. If you believe there is an error in this notice, and cannot resolve the disagreement with us, you may have the right to appeal. Visit irs.gov/appeals to learn more.

Continued on back

IRS

Payment

SAM MALONE
CHEERS, 112 BEACON STREET
BOSTON MA 55555-5555

Notice	LT16A
Notice date	11/1/2016
Your SSN	XXX-XX-5555

- Make your check or money order payable to the United States Treasury.
- Write your Social Security number (XXX-XX-5555) on your payment and any correspondence.

Select your payment amount:

- ☐ Pay the full amount due of **$4,134.38 by 11/22/2016** to avoid future interest and applicable penalties
- ☐ Make a partial payment of $____________ (enter amount)

Internal Revenue Service
ACS Support Stop 5050
P.O. Box 219236
Kansas City, MO 64121

Appendix C. LT11 Notice Prototypes

Notice Version	Description
Control (LT11A)	The Control notice is the standard notice which was then distributed to LT11 recipients. The Control notice is used as a benchmark to measure the impact of each alternative notice across various performance metrics (e.g., dollars collected, full pay rates).
Baseline for individual taxpayers (LT11F)	The Baseline notice incorporates 11 main design changes relative to the Control notice. The changes reflect best practices from prior notice redesign studies and leverage a wide range of BI principles, including: • **Salience**: Drew attention to important information by using bold text (e.g., amount due immediately) • **Reduced cognitive load**: Provided clear step-by-step online payment information • **Simplification**: Replaced dense legal jargon with easy-to-understand clear language • **Urgency of messaging**: Used urgent words such as “immediately”
Deliberate Choice (LT11C)	Using the Baseline notice as a starting point, this notice adds targeted language to characterize failure to comply with the LT11 as a deliberate choice and eliminate omission as an excuse for noncompliance: “If you do not make your payment now, we will consider your noncompliance an active choice and you could face a levy.”
Social Norm (LT11D)	Social norm messages aim to increase the “moral benefit” of compliance by emphasizing the norm among American taxpayers. Using the Baseline notice as a starting point, this notice adds targeted language to encourage the taxpayer to align with social norms: “Each year, more than 8 out of 10 taxpayers pay their taxes. You are part of a minority that has not yet fulfilled that duty.”
Personalization (LT11E)	Personalization can increase the perception that the IRS may take a specific compliance action. Using the Baseline notice as a starting point, this notice adds targeted language to personalize the notice to the taxpayer’s situation: “We identified the following source for the levy: [*major levy source*].”
Baseline for business taxpayers (LT11B)	The LT11B (baseline notice for business taxpayers) is nearly identical to the LT11F (baseline notice sent to individual taxpayers). Both notices contain a wide range of design changes. The only substantive difference between the two notices is that the LT11B provides the following business-specific guidance in the “What to Do Immediately” portion of page 1: • In the “Gather this Information” portion of page 1, LT11B recipients are instructed to gather a slightly different set of information from that for LT11F recipients (e.g., business taxpayers are not asked to gather information on filing status or address for outstanding tax years). • In the “Pay directly online” portion of page 1, LT11B recipients are instructed to make payments via EFTPS rather than Direct Pay.

Enhancing Return Risk Assessment for Examination: Recent DIF (Discriminant Function) Model Updates

Getaneh Yismaw, Drew Johns, Taukir Hussain, Jonathan Creem, and Mary-Helen Risler (IRS, Research, Applied Analytics and Statistics)[1]

1. Introduction

The mission of the Internal Revenue Service (IRS) is to:

> *Provide America's taxpayers top quality service by helping them understand and meet their tax responsibilities and enforce the law with integrity and fairness to all.*

To fulfill its mission, the IRS works to promote voluntary compliance by providing services and education to help taxpayers who want to comply with the tax law while also strategically enforcing the law to ensure that taxpayers who are unable or unwilling to comply ultimately pay their tax obligations.

The IRS is responsible for administering and enforcing the law for all taxpayers. To that end, the IRS employs a balanced approach to compliance enforcement. Part of that approach is ensuring that the IRS has an enforcement presence across all types of tax returns. This promotes fairness and horizontal equity by ensuring that all noncompliant returns have some probability of being audited.[2] Fairness and vertical equity are promoted by ensuring that tax returns that are most likely to have significant underreported tax are audited at a relatively higher rate.[3]

Another part of the balancing is employing a variety of compliance strategies and tools commensurate with the nature of the noncompliance. Key compliance strategies employed by the IRS to promote reporting compliance include:

1. math error corrections during return processing,
2. matching return information to third-party information reporting documents such as wages on the Form W-2 or interest income on a Form 1099-INT, and
3. examinations.

This paper focuses on recent updates to DIF (Discriminant Function) models that the IRS uses to meet its goals with respect to balancing *examination* coverage while also pursing the most egregious noncompliance.

1.1 Background on Three Key IRS Enforcement Strategies

Math error corrections involve computerized checks for mathematical and clerical errors during return processing of originally filed tax returns. This generally involves checks for arithmetic mistakes and errors in reading tax and Earned Income Tax Credit (EITC) tables, but also includes checks for valid taxpayer identification numbers and certain other inconsistencies outlined in the statute. If there are errors on the return, the taxpayer is issued a math error notice indicating the nature of the error and any adjustments to taxes or credits. Since

---

[1] The authors are analysts and their manager in the Knowledge Development & Application Division, Compliance Modeling Lab. The views expressed in this paper do not necessarily represent the views of the Department of the Treasury or the Internal Revenue Service.

[2] Horizontal equity in this context is the notion that similarly situated taxpayers should pay the same amount of tax.

[3] "Vertical equity" has two applicable meanings in this context. While consistent with the progressivity of the income tax, where marginal tax rates generally increase with income, it is also consistent with the notion that more egregious misreporting (in absolute terms) should be audited at a relatively higher rate.

this is done during processing of the return, it is particularly useful for math errors related to overclaimed refundable credit amounts. Since math errors are identified during the processing of the return, the amount overclaimed is not actually paid. In this sense, math error corrections reduce the need for the IRS to conduct an examination to correct simple mathematical mistakes and other types of straightforward "clerical" errors.

Third-party information reporting provides a service to many taxpayers by providing them with the information they need to complete a large portion of their tax return. And since that information also is provided to the IRS, it provides an incentive for taxpayers to accurately report that income. The incentive to accurately report income is only credible, though, if the IRS incorporates that third-party information into its compliance strategy. The IRS compliance strategy with respect to third-party information reporting makes extensive use of document matching through the Automated Underreporting Program (AUR). AUR notices are not examinations. However, they could lead to an examination, depending on the nature of the response from the taxpayer and the egregiousness of the omission.

For many compliance issues, the IRS does not have third-party information reporting or may only have limited or incomplete third-party information reporting. In these instances, the IRS must initiate an examination to ascertain the accuracy of the return and ensure that the taxpayer pays the correct amount of tax. The IRS essentially has three methods of examining tax returns that are primarily dependent on the nature and severity of the noncompliance:

1. campus correspondence examinations conducted by a tax examiner,
2. office examinations conducted by a tax compliance officer (TCO), and
3. field examinations conducted by a revenue agent (RA).

Campus examinations focus on a very limited set of issues that can be resolved by the taxpayer sending documentation to the IRS. These types of examinations typically require only a small amount of time from both the taxpayer and IRS campus staff; therefore, they represent a relatively high volume of IRS examinations. For more complicated returns and issues, such as returns with relatively small amounts of sole proprietor income, the examination may be conducted by a TCO in an IRS office at a scheduled appointment. The most complicated examinations, including examinations of business returns and individual high income returns with complex issues, are conducted by a revenue agent who examines the taxpayer's books and records and can visit the taxpayer's place of business and meet with the taxpayer outside of an IRS office setting. These field examinations take the longest to complete, partially due to their complexity and taxpayer due process rights. Therefore, it takes many revenue agents to maintain a significant compliance presence across all types of business returns and complex individual returns.

The IRS has many strategies and methods for selecting returns for an examination. The IRS may have information from a whistle blower or suspect that a taxpayer is engaged in an abusive transaction or scheme. Some examinations are initiated through referrals from state agencies or to validate a large taxpayer claim. Ultimately, most returns are selected based on some indication of compliance risk.[4]

In the absence of complete and accurate third-party reporting, the IRS generally does not know what the truth is unless it conducts an examination. In order to make the best use of the information that the IRS does have, the IRS often uses modeling to predict compliance risk and prioritize returns. Modeling could involve expert rules and filters, but often it involves sophisticated machine-learning techniques. There are many models currently deployed across the IRS using a variety of techniques. Many of those models are tuned to a specific compliance problem—for example, identifying returns with a particular compliance issue.[5] Other models are intended to assess the overall compliance risk of the whole tax return.

In particular, the IRS has historically used DIF as one method to assess compliance risk for individual income tax, small corporation income tax, S corporation, and partnership returns. DIF helps the IRS fulfill

[4] This includes returns selected as part of the National Research Program (NRP). NRP uses statistical selection methods that ensure every return covered by the program has some positive probability of being selected. Selection probabilities depend on return characteristics that are correlated with noncompliance and on IRS goals for improving compliance risk models.

[5] Models, rules, and filters are used in correspondence examinations to identify when EITC eligibility rules are likely broken, for example.

its mission by ranking returns based on the relative compliance risk of similar tax returns. To ensure fairness through balanced coverage across types of tax returns, the IRS examination function uses activity codes to group tax returns. DIF models are then developed specific to each activity code to ensure that the IRS can select the most noncompliant returns of a given return type when meeting its goals for balancing coverage.

Although DIF models (also informally called formulas) are available for individual income tax returns, small Form 1120 corporation income tax returns, Form 1120-S corporation returns, and Form 1065 partnership returns, this paper focuses exclusively on model updates made to the individual income tax return DIF models. This paper summarizes recent updates and improvements to the DIF models that use data collected by the National Research Program (NRP) in model development.[6]

2. DIF Background and Overview

Beginning in 1963, automation in tax return processing allowed the IRS for the first time to start using a set of automated expert rules (criteria) that were developed from the expertise of experienced examiners as well as historical data to rank returns for potential audit selection.[7] This shift was a huge improvement from an audit selection process that was based on manual return review and classification by experienced examiners (called classifiers). Although there were established general guidelines for the earlier manual review process, these were challenging to apply with consistency and uniformity.

Meanwhile, in a quest for a more effective and objective return selection method, the IRS began a research project in the mid-1960s to explore statistical methods for selection of returns for examination. This research project led to the use of linear discriminant analysis (LDA). After evaluation of this new approach showed superior results over the automated expert rules method, a modified linear discriminant function algorithm, from then on referred to as DIF, was first implemented for return scoring and selection in 1969. DIF is a supervised machine-learning technique[8] that predicts the likelihood of a significant tax change at the tax return level.

The effectiveness of DIF has been widely recognized by internal agency research, reviews of actual field examination results, and independent technical reviews.[9] Non-IRS independent technical evaluations have also shown its methodological soundness and effectiveness for noncompliance risk scoring and ranking. This pioneering method and scoring model have been refined and improved on an ongoing basis since the initial development and deployment.

DIF models are developed using tax return information and examination results from statistically representative samples of tax returns. These data currently are obtained from the National Research Program (NRP) and previously were from NRP's predecessor, the Taxpayer Compliance Measurement Program (TCMP). With appropriate weighting, estimates from the NRP risk-based stratified samples of returns are generalizable to the population of returns from which the samples are drawn. The representativeness of the NRP data ensures that DIF models based on those data produce return-level risk scores that are objective and representative of the population of all tax returns. This means that the predictions produced by the training data reflect the expected predictions when the DIF models are applied to the population.

DIF models are implemented as part of tax return processing, where returns are assigned a score that then becomes part of administrative data systems and can be used in downstream systems and applications for examination selection.

---

[6] Returns are selected for examination by NRP using statistical methods that allow the IRS to use the examination results to estimate noncompliance for the population and to effectively build risk-based statistical/machine-learning models to predict and/or rank compliance risk.

[7] See Fox (1986).

[8] Supervised machine learning techniques use data that include known values of the outcome variable to train a predictive model in order to predict future outcomes. These techniques are different from unsupervised techniques whereby the outcome of the variable of interest for the training data is unknown. In supervised machine learning, it is possible to assess the quality of predictions and inferences before models are deployed.

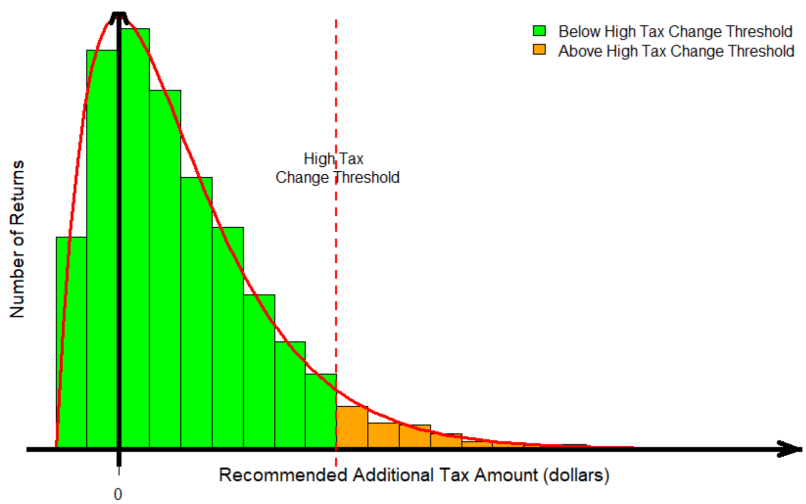
[9] "The higher the DIF score on a return, the more likely it is that an audit of that return will result in tax change. We believe that this analysis provides one of the more convincing arguments for DIF effectiveness." See GAO (1976), page 34.

3. DIF Methodology

DIF is a form of a multivariate statistical technique that is designed to classify observations into two or more groups based on a characteristic of interest. DIF is used to develop predictive models to separate returns that have a high likelihood of a high tax change (at or above a given threshold level) from those that do not.

The distribution of tax misreporting is extremely right skewed. Analysis and estimation using statistically representative data have shown that many returns have little or no underreported tax while a small number of returns have large amounts of underreported tax. The level and extent of skewness and amount of tax misreported, in general, vary by the complexity of the return and the level of income. For this reason, tax returns are stratified into mutually exclusive groups of classes called activity codes.[10] A return is assigned to an activity code during return processing. In general, the distribution of misreported tax looks like the stylized distribution in Figure 1 below.

FIGURE 1. Stylized Distribution of Tax Misreporting



Under an objective to select returns with the largest tax underreporting, the goal is to select returns from the extreme (high tax change) tail of the misreporting distribution (the orange region in Figure 1). Based on extensive analysis of NRP data, a "high tax change threshold" level is set for each activity code as part of pre-processing of NRP data during the model development process. These threshold levels may vary across activity codes.

The basic approach is to identify those predictor variables[11] that, in combination, do the best job of predicting returns with the largest amount of underreported tax.

3.1 Difference Between DIF and Standard Linear Discriminant Analysis

The model parameters are estimated, just as in any other LDA[12] technique, by inverting the pooled covariance matrix for the two groups, returns that are above and those that are below the high tax change threshold. This solution produces the estimated coefficients of predictor variables.

While the technique of solving for the model parameters of DIF is almost identical to that of LDA, the former is not the same as the latter. DIF is a modified application of LDA in which the DIF algorithm transforms

[10] The IRS stratifies returns into mutually exclusive groups of classes known by activity codes. Returns within an activity code are as homogenous as possible while operationally feasible to manage and administer. Individual income tax returns are classified by 12 activity codes. Complexity, nature, and amount of underreporting vary by activity code. See the Appendix for the definitions of the activity codes.

[11] Predictor variables, also referred to as "features" in machine-learning lexicon, in DIF model development are return line items and combinations of line items.

[12] See Hastie *et al.* (2009), pp. 106-116, for detailed discussion of linear discriminant analysis (LDA) classification technique.

predictor variables into a set of discrete likelihood ratios. This step is called binning or intervalization. Predictor variables are preprocessed into wide ordinal intervals.

For each predictor variable, a likelihood ratio—the likelihood that returns in that interval are above the high tax change threshold relative to the likelihood that returns in that interval are below the high tax change threshold—is computed for each interval. For a given variable and interval, a likelihood ratio equal to one is considered to be neutral—that is, the expected tax misreporting for returns whose value of the predictor variable is in that interval is the same as the expected tax misreporting for the overall population. A likelihood ratio less than one implies a lower proportion of returns with misreporting than in the overall population, and a likelihood ratio greater than one implies a higher proportion of returns with misreporting than in the overall population.

During DIF model development, likelihood ratios replace all potential predictor variables that are extracted or calculated from return line items. Returns whose value for a predictor variable fall in the same interval share the same likelihood ratio for that variable.

This binning (intervalization) and transformation of inputs into likelihood ratios has been very effective. It is well suited for data that are as complex as tax return data[13] and uniformly transforms potential predictor variables into discrete likelihood ratios. It is well suited to handle return line items that have large numbers of zero entries (such as certain tax credits claimed on only a few returns). Above all, extreme outlier predictor variables are "naturally" handled by the intervalization process because the bins on either end of skewed variables can be open-ended. Only the existence of the outlier, not its extreme magnitude, influences the model. This treatment of outliers contributes to the robustness of the models by ensuring that a small number of extreme outliers are unlikely to overly influence the predictions.

We find this transformation technique effective at reducing noisy and uninformative input variables. Those variables with a likelihood ratio close to "one" across their intervals will be eliminated during preprocessing for their inability to explain the variability in underreporting. We find that reducing such uninformative potential predictor variables reduces the possibility of overfitting[14] the model. We consider likelihood transformation to be a key factor that contributes to the robustness[15] of DIF performance and its durability over time.

3.2 Interpretation and Distribution of DIF Scores

As the last step in finalizing each DIF model, a model is refined so that the output scores are scaled (normalized) to fall within a specified range. Returns assigned a high DIF score by the models have a higher likelihood of significant *overall* tax change. However, the relative difference in the magnitude of two DIF scores is not necessarily reflective of the relative difference in their compliance risk. Case selection for examination is designed to be "top-down" from high to low ranking by DIF score.

Separate DIF models are developed for each activity code. The distribution of DIF scores is not necessarily the same across activity codes. DIF scores cannot be compared across activity codes or between different DIF model versions (different tax filing years) within an activity code. It is the rank of a return's DIF score within an activity code that matters for risk ranking and application.

3.3 Model Testing (Evaluation) and Selection Criteria

There are three evaluation criteria used to compare candidate models during model development as well as to assess and monitor the effectiveness of implemented scoring models. These are the following:

1. *Average Tax Change:* the overall average examination's recommended additional tax; recommended additional tax will sometimes be referred to simply as the "tax change" for brevity in this paper.

---

[13] From a statistical modeling perspective, tax return data are rather complex with a mixture of continuous, binary, categorical, and ordinal variables.

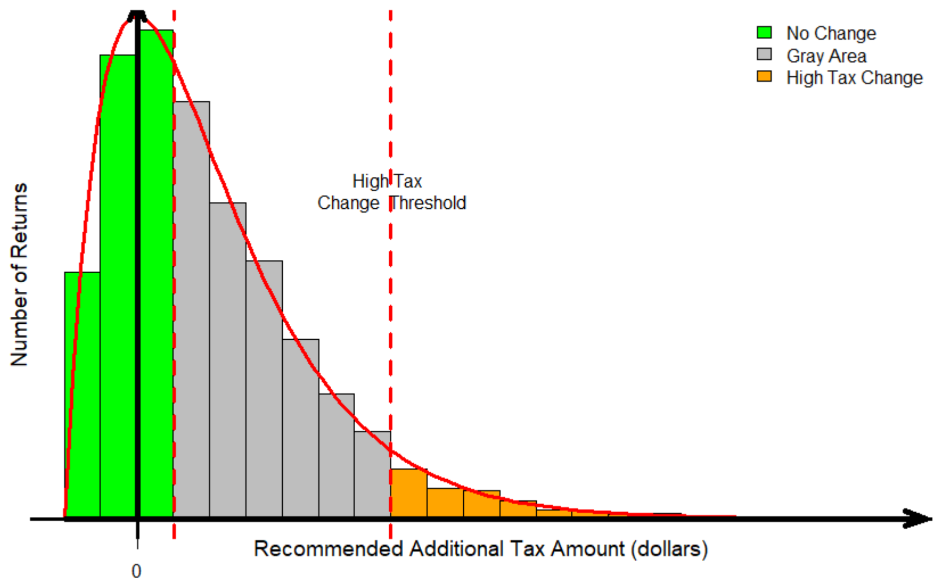
[14] Overfitting refers to the failure of a model to develop accurate prediction for observations that are not part of the training data set.

[15] Robustness in predictive modeling is the ability of the model to work well even when the underlying conditions of the data, with which the model was developed, change. For example: robustness relative to economic or tax law changes.

2. *No-Change Rate*: the share of returns expected to have a recommended additional tax less than a "no-change threshold," which is a specified positive tax change level used across all activity codes.
3. High *Tax Change (or "Hit") Rate*: the share of returns expected to result in a recommended additional tax above a "high tax change" threshold, which may vary across activity codes.

Figure 2 illustrates these concepts.

FIGURE 2. DIF Model Evaluation Thresholds[16]



These three criteria are computed at various coverage levels (cumulatively), say at 1%, 2%, …, 20%. As predicted examination coverage increases (and results from lower DIF score returns are included), the overall measures decline as the additional returns with lower productivity are included in the calculation of the evaluation measures.

In practice, even though examination coverage varies by activity code, the IRS only audits a small percentage of returns. Accordingly, DIF models are evaluated and selected using an expected examination coverage level at the tails of the misreporting distribution. For this reason, DIF models are, by design, effective at the upper tail of the examination coverage levels in ranking returns by their likelihood of high tax change potential.

Once a potential new DIF model is selected from among all competitor models,[17] it is then compared against the existing DIF model that is in production.[18] If the performance of the new model is significantly better than the model that is in production, a decision will be made to implement the new model. This starts a process in which staff from the IRS Information Technology (IT) organization reprogram the relevant return processing computer code that classifies a return into the correct activity code and assigns a DIF score as part of return processing.

Figure 3 presents an example of a DIF model evaluation. In this example, the evaluation graphs show that the competitor model significantly outperforms the existing model across all three model evaluation criteria. However, there are circumstances in which a model shows significant improvement in one criterion but not in

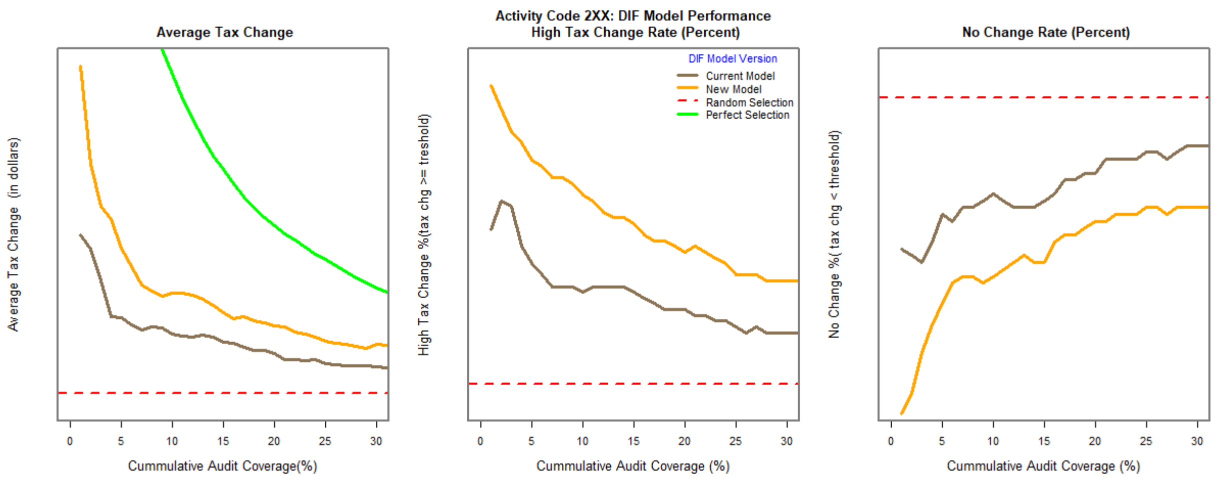
[16] The gray area is the area between "no-change" and "high tax change" areas where returns have tax change higher than at the "no-change" threshold but lower than at the "high tax change" threshold.

[17] DIF model developers usually develop multiple models and select a final candidate model among competitors based on the three-model selection criteria, thus using a champion/challenger strategy.

[18] "In production" in this context describes the DIF model that is currently being used to score returns during return processing.

another. For example, a model may be very good at picking outlier cases and hence exhibit better average tax change numbers, but this may come at the expense of increased no-change rates. These tradeoffs between the no-change rate and the average tax change often require additional model redevelopment and further testing and evaluations before final model selection. If, based on a qualitative collective assessment and careful consideration of potential tradeoffs between lower/higher expected no-change rate versus lower/higher expected average tax change, it is determined that the new model is not materially better than the existing one, implementation will be deferred and the existing model will remain in production.

FIGURE 3. Example of DIF Model Evaluation: Comparison of Existing DIF Model and Competitor Model Using a Holdout Data Set



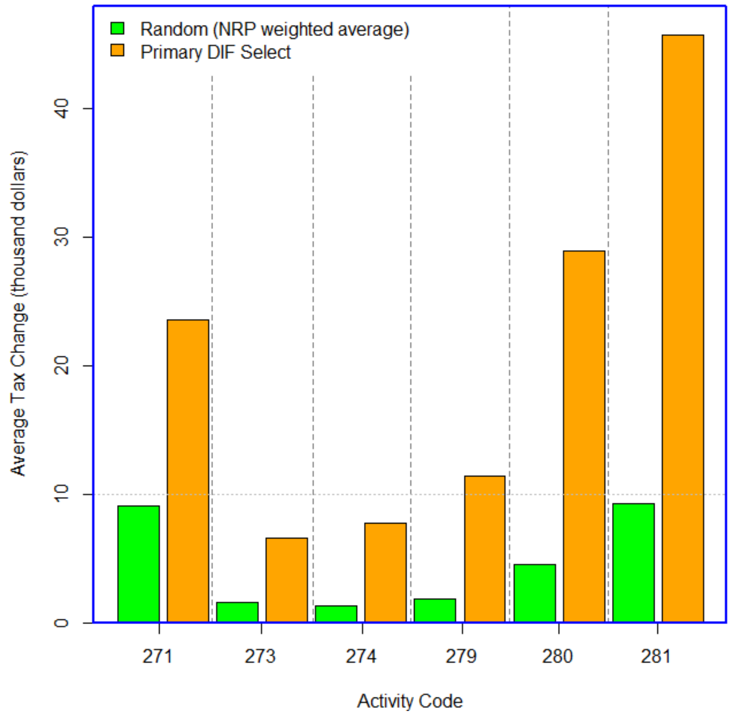
For effective and optimal DIF score utilization, DIF model developers work with the relevant examination subject-matter experts, review DIF score rankings on an annual basis, and monitor and analyze model effectiveness using audit outcomes of returns selected for examination by DIF.

3.4 Effectiveness

One way to observe whether DIF is effective at ranking the risk of a high tax change is to compare it, on relevant measures, with the outcome of selecting returns using simple random sampling from the population.[19] To demonstrate this, additional average recommended tax amounts are compared using NRP data for Tax Year 2015 (Tax Year 2015 returns are filed predominantly in Calendar Year 2016) and data from examinations of primary DIF selected returns filed in 2016 that were audited and closed. Since NRP samples are statistically representative of the filing population, the weighted average additional recommended tax amounts are statistically unbiased population-level estimates of additional tax amounts. As Figure 4 shows, examination results from DIF selected returns have much higher average additional recommended tax amounts than do the estimates of additional recommended tax for the population of the respective activity code. In some activity codes, this difference is many times higher. Since return complexity varies by activity code, the relative effectiveness of DIF also varies by activity code.

[19] Comparing against simple random sampling gives an indication of how well a DIF model selects a group of returns that are more noncompliant than the population average.

FIGURE 4. Average Additional Recommended Tax Amounts: Primary DIF Selected Returns Compared with Population Average (Random Selection) for Selected Activity Codes



4. DIF Utilization

DIF for individual income tax returns is used mainly by the Small Business/Self-Employed (SBSE) division examination offices where high DIF scored returns are selected, classified, and assigned to an RA for a field examination or to a TCO for an office examination.[20] Over the past decade, the IRS budget has seen reductions[21] and, as a result, the number of examinations has fallen steadily. Similarly, the number of DIF selected returns has declined substantially over the years. For examinations that were closed as of approximately April 2021, Table 1 shows the number of returns selected using DIF plus DIF-related[22] and non-DIF methods for examinations conducted by RAs and TCOs by the year in which the return was filed.

[20] The focus is on the use of DIF scores to select returns for field examinations because DIF scores are generally not used to select returns for correspondence examinations.

[21] "The IRS's appropriations have fallen by 20 percent in inflation-adjusted dollars since 2010, resulting in the elimination of 22 percent of its staff. The amount of funding and staff allocated to enforcement activities has declined by about 30 percent since 2010. Since 2010, the IRS has done less to enforce tax laws. Between 2010 and 2018, the share of individual income tax returns it examined fell by 46 percent, and the share of corporate income tax returns it examined fell by 37 percent. The disruptions stemming from the 2020 coronavirus pandemic will further reduce the ability of the IRS to enforce tax laws." CBO (2020).

[22] Primary DIF cases are returns that were selected specifically because those returns had a high DIF score. Sometimes an examination of a return selected for examination on the basis of its DIF score may lead to the selection of other returns for examination. Common examples include prior or subsequent year returns for the same taxpayer or related passthrough businesses in which the taxpayer is a shareholder or partner. These examinations are referred to as DIF-*related* examinations.

TABLE 1. Number of Closed Examinations by Revenue Agents and Tax Compliance Officers by Return Filing Year

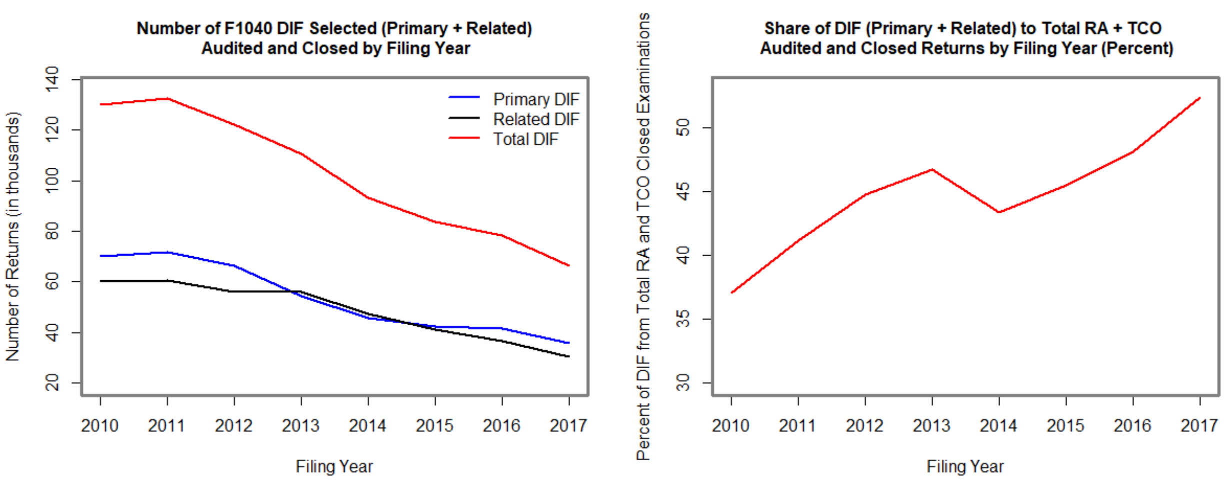
Filing Year	Total Audits	Correspondence Audits	Field + Office Audits		
			Non-DIF	Primary DIF + DIF-Related	Total
2010	1,540,504	1,188,828	220,695	129,970	350,665
2011	1,460,707	1,138,156	189,543	132,596	322,139
2012	1,311,556	1,038,371	150,686	122,254	272,940
2013	1,221,850	985,030	126,006	110,557	236,563
2014	1,051,727	836,617	121,416	93,161	214,577
2015	918,544	734,192	100,338	83,746	184,084
2016	897,541	734,846	84,432	78,211	162,643
2017	777,586	651,241	60,084	66,159	126,243

SOURCE: Audit Information Management System (AIMS). Cases closed through March 2021.

As shown in Figure 5, compared to the overall decline in total RA and TCO examined returns, the proportion of DIF selected returns to that of total RA and TCO returns increased from a little over a third for Filing Year 2010 to about half of the reduced RA and TCO total for 2016.

Overall, DIF remains one of the main sources of return selection for RA and TCO audits. For Filing Years 2010–2017, about one-third to one-half of returns audited by RAs and TCOs were selected directly or indirectly using DIF (primary and related pickups).

FIGURE 5. Trends in Individual Income Tax DIF Utilization for Returns Audited by Revenue Agents and Tax Compliance Officers: Number and Share of Closed Examinations for Filing Years 2010–2017



SOURCE: Audit Information Management System (AIMS). Closed examinations as of March 2021.

5. Continuous Model Improvement

All individual income tax DIF models used to be updated at the same time due to the nature of TCMP data collection and availability of new data. TCMP studies of individual income tax returns were conducted regularly every 3 years from 1961 to 1988. TCMP was discontinued for 13 years and NRP was launched for Tax Year

2001 with a stratified sample of about 41,000 individual income tax returns. Completed data were made available in 2005. DIF models were first updated under the prior individual income tax return activity codes and implemented in Filing Year 2006 as an interim measure while the categorization of returns was being revised on the basis of data analysis, eventually leading to 12 new activity codes. New DIF scoring models were then developed for all 12 activity codes and made operational for Filing Year 2007.

Beginning with Tax Year 2006, NRP restarted its individual income tax studies with a redesigned sample. NRP redesigned its stratified sample design from one large periodic sample to a smaller annual sample design.[23] The annual sample was one-third of the sample size needed to achieve the statistical precision previously achieved through a larger, periodic sample. The new design required combining multiple NRP study years and waiting until sufficient data had been accumulated before improved models could be obtained.

Since it takes time to complete the audit process and collect and process detailed NRP data, the first DIF model development took place in 2011–2012 for implementation in Filing Year 2013 and used data from completed NRP sample returns for Tax Years 2006–2008.

Unlike with the previous DIF model development cycles under TCMP and the NRP Tax Year 2001 data, there were not enough data to develop improved models for all activity codes and update them all at once. We developed a data-driven approach based on the strength of the existing DIF model and the availability of new data to identify which activity codes to prioritize for model development based on the potential and need for improvement.

In the first development cycle using the newly available NRP data, we identified and developed models for five of the 12 activity codes. Of these, we implemented new models for two activity codes. The following year, with an additional year of sample data for Tax Year 2009, we developed competitor models for five activity codes and implemented four for Filing Year 2014.

By Filing Year 2019, after four rounds of model development and implementation, all 12 individual income tax activity codes had been updated with at least one new scoring model. Of these, five activity codes have had models updated twice.

Additional new models will be implemented for filing season 2022. NRP data for Tax Years 2009–2015 were used to develop competitor models for nine activity codes. Six were chosen for implementation. Three of these models are for the newly created activity codes 282–284 which were created by splitting activity code 281.

Table 2 summarizes the development and implementation activity related to model implementation for Filing Years 2013–2022.

TABLE 2. DIF Model Development and Implementation, Selected Filing Years 2013–2022

Implementation Filing Year	NRP Data Used	Number of Activity Codes	
		Developed	Implemented
2013	TY2006–2008	5	2
2014	TY2006–2009	5	4
2016	TY2006–2010	7	6
2019	TY2006–2013	6	5
2022*	TY2009–2015	9	6

* For Filing Year 2022, the IRS is increasing the number of individual income tax activity codes from 12 to 14. Activity code 281 will be divided into three new activity codes (each with its own new DIF scoring models). The new activity codes are as follows:

282: Total positive income ≥ $1,000,000 and < $5,000,000
283: Total positive income ≥ $5,000,000 and < $10,000,000
284: Total positive income ≥ $10,000,000.

[23] Beginning with Tax Year 2006, the individual income tax NRP sample size needed to achieve the same level of precision as that of the larger annual periodic sample size was divided equally over 3 years. Each annual NRP sample is independent and statistically representative of the filing population from which it is selected.

With the annual NRP data strategy, as more data are accumulated over time, we are able to monitor model effectiveness, identify activity codes for potential for model improvement, and adopt a more flexible regime for continuous model development and updating.

6. Model Improvements and Examination Results

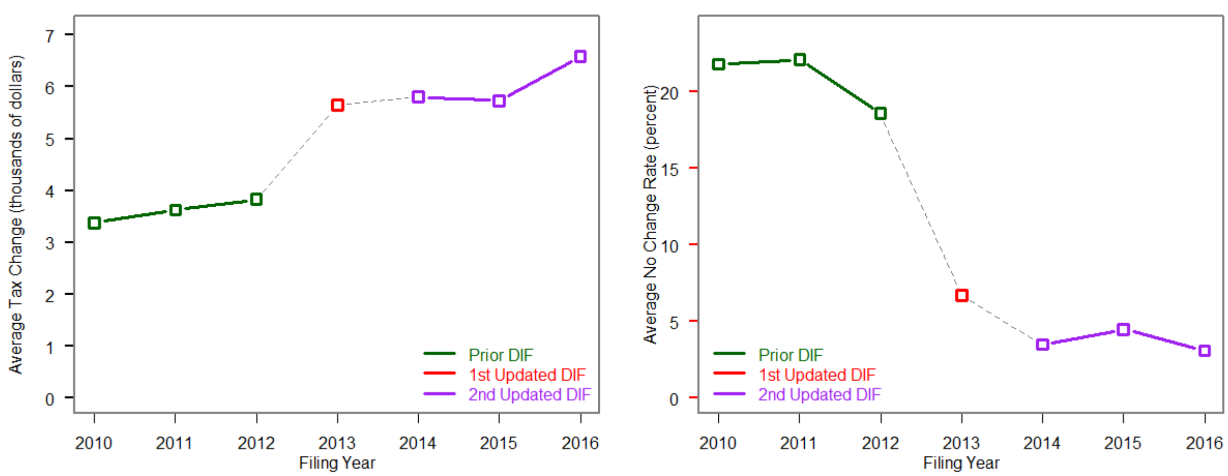
Due to the nature of the examination process and the duration of examinations, it takes multiple years for the outcome of most audits selected under the same new scoring model to become available for analysis. In general, returns with the largest proposed examination adjustments take longer to complete, and therefore there is a considerable lag before the effect of new models is fully known.

This section shows trends in examination results from the first three rounds of DIF updates (for Filing Years 2013, 2014, and 2016). To do this, we analyze the results of closed examinations for Filing Years 2010-2016. This gives us 3 to 4 years before a model was updated and another 3 or 4 years after model updates to compare trends.

The following five graphs, Figures 6-10 illustrate the results of primary DIF select returns[24] from Filing Years 2010–2016, comparing trends in average additional recommended tax amounts and no-change[25] rates before and after DIF model updates for each of the five activity codes whose DIF scoring models were updated twice.

Figure 6 shows primary DIF performance trends for activity code 273. The updated DIF scoring models for activity code 273 were implemented in Filing Year 2013 and again in Filing Year 2014. As can be seen in the graph, the average recommended additional tax per return is substantially higher than what the trendline before model updates would expect. More notable, the no-change rate is much lower—it was about 20% before the model updates and is about 5% after the model updates. This is a substantial reduction in unnecessary contacts with compliant taxpayers—it is a reduction from about one in five to fewer than one in 20 after new models were implemented.

FIGURE 6. Average Tax Change and No-Change Rates for Activity Code 273* Primary DIF Select Examined Returns by Filing Year



SOURCE: Audit Information Management System (AIMS). Examinations are closed as of March 2021.

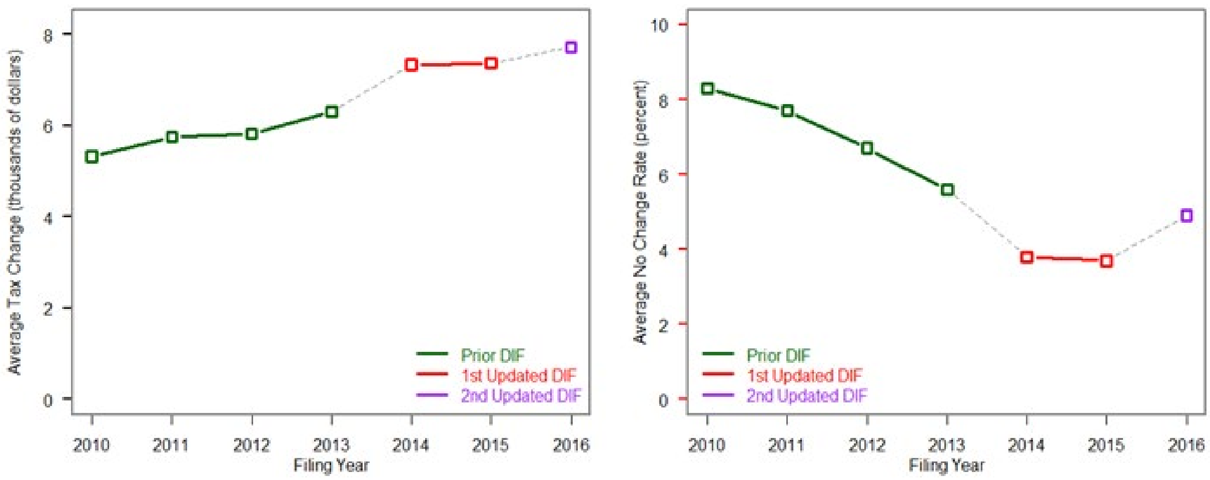
*Activity code 273 includes non-business returns with total positive income under $200,000; with no schedule C/F; and with Schedule E or Form 2106 present.

[24] The primary objective of DIF is to score returns on the basis of the likelihood of a specific return with a high tax change if audited. Therefore, the evaluation focuses on returns that were selected specifically because those returns had a high DIF score (primary DIF select returns). Once a primary DIF select return has been selected for an examination, examiners are responsible for ensuring that all other required returns have been filed and if any of the returns controlled by the taxpayer should be selected for a DIF-related examination. These DIF-related returns are not included in this analysis.

[25] The operational definition of a no-change return is one that is closed with the Examination Disposal code of "2" (audit closed with no adjustment).

Figure 7 shows primary DIF performance trends for activity code 274. The updated DIF models for activity code 274 were implemented into production in Filing Year 2014 and again in Filing Year 2016. The prior DIF scoring model for activity code 274 returns was one of the strongest with operational no-change rates in the single digits. The updated DIF model improved the strong trendlines both in additional recommended taxes and no-change rates.

FIGURE 7. Average Tax Change and No-Change Rates for Activity Code 274* Primary DIF Select Examined Returns by Filing Year

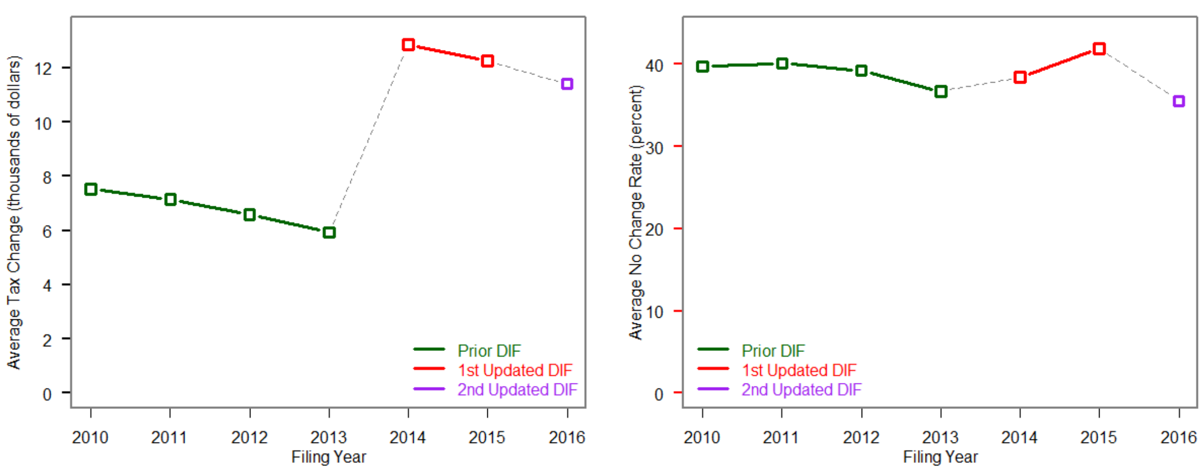


SOURCE: Audit Information Management System (AIMS). Examinations are closed as of March 2021.

*Activity code 274 includes returns with total positive income < $200,000; with Schedule C; and with total gross receipts < $25,000.

Figure 8 shows DIF model performance trends for activity code 279. Updated DIF models for activity code 279 were implemented in Filing Year 2014 and again in Filing Year 2016. Average additional recommended taxes sharply increased by more than 80 percent from the previous 3-year trendline the year the first model update was implemented. While the no-change rate has seen modest uptick for Filing Years 2014–2015, it has been reversed with the second model update for Filing Year 2016.

FIGURE 8. Average Tax Change and No-Change Rates for Activity Code 279* Primary DIF Select Examined Returns by Filing Year

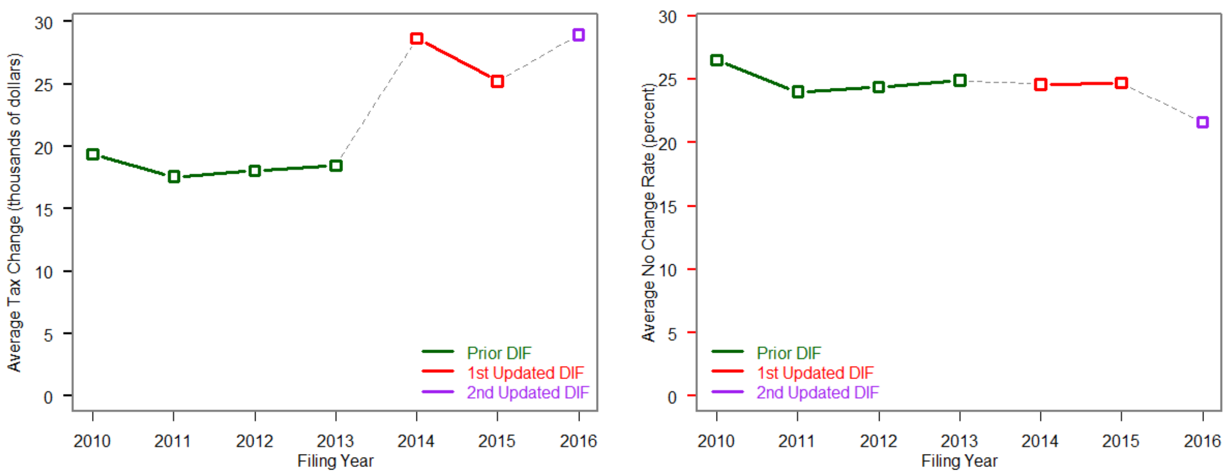


SOURCE: Audit Information Management System (AIMS). Examinations are closed as of March 2021.

*Activity code 279 includes non-business returns with total positive income at or above $200,000 but less than $1 million; with no Schedule C/F

Figure 9 shows primary DIF performance trends for activity code 280. Updated DIF scoring models for activity code 280 were implemented in Filing Year 2014 and again in Filing Year 2016. For returns filed the year the new model was implemented, average additional recommended taxes per return were more than 50 percent higher than expected on the basis of the previous 3-year trendline. Average additional recommended tax amounts increased from about $18,000 to about $29,000.

FIGURE 9. Average Tax Change and No-Change Rates for Activity Code 280* Primary DIF Select Examined Returns by Filing Year

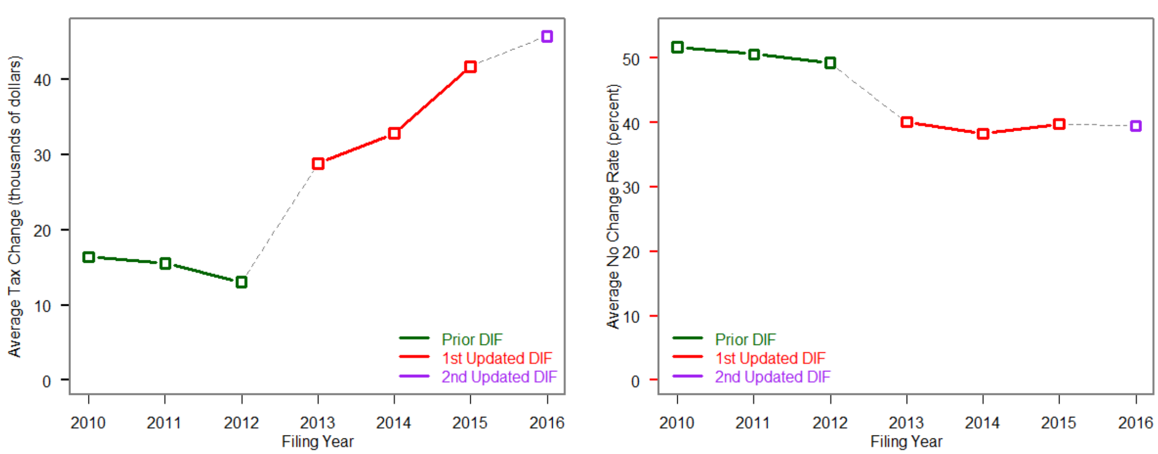


SOURCE: Audit Information Management System (AIMS). Examinations are closed as of March 2021.

*Activity code 280 includes business returns with total positive income at or above $200,000 and below $1 million; and with Schedule C/F.

Figure 10 shows primary DIF performance trends for activity code 281. Updated DIF scoring models for activity code 281 were implemented in Filing Year 2013 and again in Filing Year 2016. Average additional recommended taxes per return doubled from the previous 3-year trendline from about $15,000 to about $29,000 the year the updated model was implemented. In the 4 years since the updated model was implemented, the average additional recommended tax amount more than tripled from $13,000 in 2012 to over $45,000 per case for returns filed in 2016. Meanwhile the average no-change rate went down from about 50% with the retired model to about 40%, a reduction of 20% from the prior baseline trend.

FIGURE 10. Average Tax Change and No-Change Rates for Activity Code 281* Primary DIF Select Examined Returns by Filing Year



SOURCE: Audit Information Management System (AIMS). Examinations are closed as of March 2021.

*Activity code 281 includes returns with total positive income at or above $1 million.

7. Exploring Enhancements to Return-Level Risk Modeling

Two general lines of research for enhancing return-level risk scoring are being conducted in parallel with the ongoing development of improved DIF models. These are exploring alternative analytical techniques and expanding the set of potential predictor variables. This section discusses work related to the former. Advanced data analytics to inform decision-making and improve operational outcomes is an IRS strategic goal. Data and analytics are front and center to drive decisions based on data insights and analytics.[26] In recent years, the IRS's data and analytics capabilities have greatly expanded, making possible more in-house research into other advanced analytical techniques, among them various supervised and unsupervised machine-learning techniques, active learning, and text analytics.

Earlier research[27] into alternative analytical techniques[28] generally was undertaken by external parties under contract. The earliest of this research included development of competitor models to DIF using a discriminant analysis technique along the lines of the DIF approach. Other work involved the development of various two-stage models. The first stage model involved the prediction of some variation of "no-change" vs. "change" using techniques such as logistic or probit regression or some type of classification technique. The second stage involved the prediction of tax change using some form of regression technique. The first-stage probabilities of "tax change" are then multiplied with the predicted tax change to rank returns by *expected* tax change. One contract explored the development of early versions of neural networks and another explored Classification and Regression Trees (CART).

These efforts provided useful insights into other modeling techniques. When the performance of the models developed through these efforts was compared to DIF, in general DIF had better performance. Some two-stage models had slightly better expected tax changes but at the expense of unacceptably high predicted no-change rates. Some DIF-type modified linear discriminant models developed by an outside contractor performed closer to DIF in a few activity codes but came short in many others.

The most recent externally contracted work was completed in the fall of 2020. For that project, the contractor team extensively explored a comprehensive suite of machine-learning algorithms—including various parametric regression models, tree-based models, and other classification algorithms—for two activity codes. The contractor team developed a variety of models[29]—along with a variety of pipeline[30] steps—and identified their top models for each of the three evaluation criteria that IRS DIF model developers use in development and evaluation and the top models based on a contractor-specified criterion.

IRS staff then selected the six top performing models for each activity code and compared the performance of those models to the performance of the current DIF model in production. The comparison was done on an out-of-sample basis, i.e., the test data was independent and not used to develop either the current DIF scoring models or the contractor's top performing models. The comparison showed that no single analytical technique is consistently superior to another. While the contractor's top-performing models, in general, have slightly lower expected no-change rates than do DIF models, the current DIF scoring models in general have higher average recommended additional tax amounts.

Our DIF team also has been exploring the development and performance of models using analytical techniques other than DIF. The earliest of this work explored the development of random forest (RF) models.

---

[26] See the IRS strategic plan for Fiscal Years 2018–2022 in "Advance Data Access, Usability and Analytics to Inform Decision-Making and Improve Operational Outcomes," Advanced Data & Analytics, https://www.irs.gov/about-irs/strategic-goals/advance-data-analytics.

[27] The IRS also pursued major research projects to augment DIF scores by developing issue (line item)-level risk scoring/ranking models. For example, in 1984 the IRS conducted "Issue Scoring of Returns (ISCOR)"—a major project to identify risk score and rank line items (issues) that have high DIF scores. In 2006 and 2007, the IRS pursued the issue-based research "Issue Based Models (IBM)" project that resulted in development of 11 issue-based scoring models that were piloted. Research on issue-based modeling is an ongoing project. The application of early versions of Neural Networks and Classification and Regression Tree (CART) was also pursued in the 1990s (see Asner (1993)).

[28] All the alternative analytical techniques explored are exclusively supervised predictive algorithms—as it is possible to assess the accuracy and reliability of the models in question by using test data with known audit outcomes, such as NRP data.

[29] The types of models developed include the following: parametric regression (ordinary least squares, Ridge, Lasso, and elastic net), tree-based methods (decision tree, random forest, extra tree, gradient boosting, and adaptive boosting), and Support Vector Machine (K-nearest neighbors, logistic regression, and ensemble methods).

[30] Some of the preprocessing pipeline steps used include feature creation, feature selection, feature scaling/normalization, and feature and target transformations.

The most recent work has been done as part of redesigning the sample selection approach for the individual income tax NRP. The general approach for that is to move away from a design-based sample. The NRP redesign hypothesizes that better integration of risk modeling with the exploration necessary for learning will allow the IRS to more efficiently collect the data that it needs to effectively estimate compliance risk. The foundation of the approach is the development of a risk model for use in the sample stratification. The analytical techniques explored in the process of developing a risk model for the NRP redesign pilot include multiclass multistage random forest (RF) classifier and regression models, ensemble modeling, model-based stratification, and active learning concepts.

The alternative techniques explored for risk modeling as part of the NRP redesign, on balance, provide comparable performance in ranking returns at the top of the compliance distribution as DIF. Given the early stages of the NRP redesign, the results are promising, particularly for some activity codes. However, as prior research has shown, DIF is very effective and robust. Also consistent with prior research results, there are often tradeoffs in lowering the no-change rate at the cost of missing some cases with high unreported tax.

8. Future Work

Future work will include continuing to update the DIF models as new data become available and continuing to explore alternative techniques and expanding the universe of potential predictor variables. Results from the research to date into alternative techniques suggest that no one analytical technique yields models that are vastly superior to all others on all relevant criteria. Additional work will continue to refine models developed using an assortment of analytical techniques. The work will explore ways in which models developed using different analytical techniques might complement one another as part of a process for return selection that could improve the overall strategy for examination return selection.

References

Asner, Lance S. (1993), "Neural Networks and Discriminant Function: Alternative Techniques of Selecting Tax Returns for Audit,. *IRS Research Bulletin*, pp. 118–127.

Congressional Budget Office (CBO) (2020). "Trends in the Internal Revenue Service's Funding and Enforcement," Publication 56467, July 2020. https://www.cbo.gov/publication/56467.

Fox, Norman H. (1986). "Examination: Allocating Resources," Federal-State Research Conference, November 19–21 .

Government Accountability Office (GAO, formerly Comptroller General of the United States) (1976). "How the IRS Selects Individual Income Tax Returns for Audit," *Report to the Joint Committee on Taxation*, GGD-76-55, November 5.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*, Second Edition, Springer-Verlag.

Hiniker, John P. (1987). "The Selection of Returns for Audit by the IRS," *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 294–299.

Appendix

Form 1040 Activity Code Definitions (From 2007 to 2021)*

Activity Code	Definition
270	EIC > 0; TGR < $25,000
271	EIC > 0; TGR >= $25,000
272	TPI < $200,000 and No Schedule C, E, F, or 2106
273	TPI < $200,000; No Schedule C/F; Schedule E or 2106 Present
274	Schedule C; TGR < $25,000; TPI < $200,000
275	Schedule C; $25,000 <= TGR < $100,000; TPI < $200,000
276	Schedule C; $100,000 <= TGR < $200,000; TPI < $200,000
277	Schedule C; $200,000 <= TGR; TPI < $200,000
278	Schedule F Not Classified Elsewhere; TPI < $200,000
279	No Schedule C/F; $200,000 =< TPI < $1,000,000
280	Schedule C/F; $200,000 <= TPI < $1,000,000
281	TPI >= $1,000,000

TPI = Total Positive Income
TGR = Total Gross Receipts
EIC = Earned Income Tax Credit
* In Filing Year 2022, the IRS is in the process of changing the number of F1040 Activity Codes from 12 to 14. This will be done by splitting activity code 281 into three classes (each with its own new DIF scoring model):

282: TPI > $999,999 and < $ 5,000,000
283: TPI > $ 4,999,9999 and < $10,000,000
284: TPI > $9,999,999

4

▽

Enhancing Taxpayer Customer Experience

Herlache ◆ Turner ◆ Hall ◆ Bell

Bonhomme ◆ Holland ◆ Hu

Sun ◆ Needham

Increasing Take-up of the American Opportunity Tax Credit[1]

Anne Herlache,[2] Mary Clair Turner,[3] Crystal Hall,[3,4] and Elizabeth Bell[3,5]

Introduction

To promote college access and affordability, education tax credits help with the cost of college by reducing the amount of tax owed. Despite the American Opportunity Tax Credit (AOTC) being the most generous Federal higher education tax credit, national estimates find that many eligible college students or their families do not claim it (Guyton *et al.* (2017a)). AOTC has a maximum credit of $2,500 per eligible student and is partially refundable; thus, eligible students who do not claim the credit are missing out on a potentially large tax benefit, including a refund of up to $1,000 if the credit reduces their tax owed to $0 (Internal Revenue Service (2021)). The Internal Revenue Service (IRS) FY 2018-2022 Strategic Plan includes an objective to empower and enable all taxpayers; increasing take-up of AOTC among those eligible would support this goal by helping taxpayers receive the tax credits to which they are entitled (Internal Revenue Service (2018)).

Letter and postcard campaigns that aim to increase awareness and provide information have shown promising results at scale across a number of domains. These communications interventions have increased take-up of the Earned Income Tax Credit (EITC) (Bhargava and Manoli (2015); Guyton *et al.* (2017b)), take-up of nutrition benefits through the Supplemental Nutrition Assistance Program (SNAP) (Finkelstein and Notowidigdo (2019)), increased enrollment at highly selective colleges among high-achieving, low-income students (Hoxby and Turner (2013)), and increased enrollment in health insurance plans (Goldin *et al.* (2021)), to name a few examples.

Many of these information campaigns draw on what we know from the social and behavioral sciences about how people make decisions and process information to help the reader more easily act on the information provided and follow through on a desired behavior. These "behaviorally informed" communications implement insights such as simplifying information, providing clear actions steps, or cuing relevant social norms. Beyond communications, researchers and policymakers have used this approach to redesign other aspects of program implementation (e.g., changing the default option on a form, requiring a user to make an active choice, or changing when information is requested). On average, behaviorally informed program changes, which include information-based communications, are a relatively low-cost strategy to increase take-up of benefits and tend to have small but meaningful effects (DellaVigna and Linos (2022)).

Yet, despite this success at scale and across domains, these communications campaigns are not uniformly effective. Other studies call into question whether information alone is sufficient to move the needle on priority outcomes. For example, a handful of recent studies that tested multiple variations of behaviorally informed communications—with sample sizes ranging from approximately 800,000 to over one million individuals—rule out even small effects of these interventions on take-up of EITC (Linos *et al.* (2020)), submissions of the Free Application for Federal Student Aid (FAFSA) (Bird *et al.* (2021)), and changes to college enrollment patterns (Gurantz *et al.* (2021)). Instead, other strategies that have successfully increased college enrollment and take-up of aid, such as coupling information with personal assistance (Bettinger *et al.* (2012); Carrell and Sacerdote (2017)) or reducing uncertainty with guarantees for aid (Dynarski *et al.* (2021)), may be necessary to increase take-up of aid or change college enrollment decisions.

1 The opinions expressed in this paper are those of the authors alone and do not necessarily represent the views of the Internal Revenue Service, the U.S. Treasury Department, or the U.S. General Services Administration.

2 IRS Research, Applied Analytics, and Statistics

3 U.S. General Services Administration, Office of Evaluation Sciences

4 University of Washington

5 Florida State University

The effectiveness of behaviorally informed communication and other program changes often depends on the characteristics of the sample of interest, the actions the outreach is trying to encourage, and the behaviorally informed strategy that is applied, among other dimensions of sample, context, and intervention (DellaVigna and Linos (2022); Guyton *et al.* (2017a)). The primary aim of this study is to build evidence on the effectiveness of behaviorally informed tax credit communications on promoting take-up of AOTC among current and recently enrolled college students. A secondary aim is to explore the extent to which the effectiveness of the intervention depends on the sender of the communications (i.e., messenger effects).

To reach these aims, the Office of Evaluation Sciences (OES) at the U.S. General Services Administration; the Research, Applied Analytics, and Statistics (RAAS) division of the IRS; and the University of Illinois at Chicago collaborated to design, implement, and evaluate the effectiveness of a multi-modal, coordinated communications intervention aimed at increasing AOTC take-up. To increase the likelihood that students would engage with the materials, the content of the messages incorporated feedback from student focus groups and insights from the behavioral sciences (e.g., timely reminders, action steps, loss aversion, and easy-to-use links to free tax software). Moreover, to increase the likelihood that eligible taxpayers open and read the materials, the communications bundle included multiple modes of communications (emails and a letter) and included multiple target audiences (student and primary tax filer). For students assigned to the IRS-letter group, the communications bundle also included multiple senders of the communications (University Bursar and IRS). In this paper, we report preliminary results of the effects of the communications bundle on AOTC take-up and the effects on AOTC take-up for the University- and IRS-letter groups separately.

Even though this study was implemented during the start of the COVID-19 pandemic (during the 2019-2020 academic year), the experimental design allows for credible causal estimates of the effects of the intervention. Absent the intervention, the randomized design ensures that in expectation the effects of the pandemic on student outcomes should be similar across assignment groups. Still, the external validity of the results and the applicability of the results to other contexts will need more careful consideration. Future research will explore these concerns and examine the effects on additional outcomes, including whether the communications prompted filing among those who otherwise would not file, either because they fall below the income threshold or for other reasons, and potential downstream effects of the intervention on college outcomes.

Related Research

Earlier work on higher education tax credits finds that many eligible students and families with college-age children do not claim AOTC (Guyton *et al.* (2017a)). Among eligible students, take-up of AOTC is lowest among independent students and highest among recent high school students, with dependent students falling in between (Guyton *et al.* (2017a)). When considering the broader college-age population, high-income households are more likely to take up the credit—in part because their children are more likely to enroll in college and spend more on eligible college expenses, but likely also because they have more resources with which to navigate the complexity of the tax system and file a return (Bulman and Hoxby (2015)). As with take-up of tax benefits and other forms of aid (described above), researchers have explored the extent to which behaviorally informed communications are an effective way to increase take-up of higher education tax benefits and promote college enrollment. Our study builds most closely upon two studies that use outreach to increase awareness about higher education tax benefits and measure the effects of these communications efforts on take-up of benefits (in one case) and college enrollment (in both cases). We focus the remainder of this review on those studies and how our study complements this earlier work.

Working with ApplyTexas, a widely used portal for applying to colleges and universities in Texas, Bergman *et al.* (2019) provide evidence that lack of awareness of higher education tax benefits is not the primary barrier limiting tax benefits for higher education from improving college outcomes. In the field experiment, ApplyTexas sent email and letter outreach about higher education tax benefits and FAFSA to a random sample of over one million students, who were split into three student segments: high school students who applied to college, students who had enrolled in college the prior year, and students who applied to a public college or university in Texas but did not enroll. To support their claim that the intervention reduced barriers to awareness, the authors document engagement with the email communications: forty-three percent of the treatment sample opened at least one email and of those, one-third opened an email multiple times. While the authors cannot measure take-up of higher education tax benefits directly, their large sample size allows them to rule out even small effects of the communication campaign on college enrollment outcomes for each of the student segments.

Like Bergman *et al.* (2019), Guyton *et al.* (2017a) study the effects of outreach about higher education tax benefits and split their sample into three segments: high school seniors, college students claimed as dependents for tax purposes, and college students who filed their taxes independently. Also similar to the ApplyTexas study, the outreach included information about three tax credits (AOTC, the Lifetime Learning Credit, and EITC). However, in this field experiment, the IRS implemented the outreach. The IRS mailed letters to those in the treatment group twice (once in February 2015 and again in March 2015) and sent no communications to those in the control group. Using IRS tax data to measure outcomes, the authors find that the letters have small effects on take-up of these tax benefits. The effects on AOTC take-up ranged from just under 1 percentage point to 2 percentage points, with the largest effects observed among college students who filed their taxes independently. While these effects are similar in size to the average effects of behaviorally informed program changes implemented by government research units (DellaVinga and Linos (2022)), the letters impacted filing behaviors—nearly a year after the mailings—when families completed their 2015 tax returns. Still, consistent with Bergman *et al.* (2019), the authors find no effects of the letters on college enrollment among any of the student segments.

Our study builds on this work in a couple of ways. First, we examine the effects of outreach among a sample drawn from the University partner, which allows us to examine the effects of outreach among current or recently enrolled college students, rather than among those who appear eligible for higher education tax benefits based on their age or prior interest in applying to college. Moreover, we are able to measure the effects of outreach on AOTC take-up directly, as in Guyton *et al.* (2017a), but during the same tax year as the intervention was implemented. Second, working with both the IRS and a University partner allows us to examine the effects of outreach from both messengers and explore possible messenger effects of this outreach. Whereas the University Bursar sent all email components of the interventions, we randomly varied whether the IRS or the University Bursar sent the letter component. Still, like Bergman *et al.* (2019), we are able to address potential coordination problems between students and primary tax filers by including outreach to both students and adults registered as authorized payers on their university accounts. By implementing a multi-modal, multi-sender communication study, we add to a knowledge base on outreach for take-up of tax benefits that primarily reflects single sender, and often single mode, communications. The results of this study leave both the IRS and colleges and universities better informed on ways to help students or their families take advantage of tax benefits to which they are entitled.

Data and Method

We conducted a randomized field trial during the 2019–2020 academic year, centered on filing tax year (TY) 2019 returns, during the calendar year 2020 filing season. Our analysis uses two administrative data sources: University records and IRS tax data. Using administrative data from the University, we determined the sample and generated treatment assignment groups within random assignment blocks. Later, we matched these data to IRS tax data to measure outcomes.

Sample. The sample includes undergraduate students who are U.S. citizens and enrolled at the University at least half-time for one or more academic terms in 2019 (Spring 2019, Summer 2019, and/or Fall 2019; N=19,071). To limit the sample to students likely eligible for AOTC, we excluded students with 120 or more transfer credits, students only enrolled in graduate programs or non-degree programs, students enrolled less than half-time each academic period, and students missing key information (e.g., Social Security Number).

Experimental Design. To improve precision, we randomized among students who shared characteristics on three dimensions: (a) likelihood of eligibility for AOTC, (b) likelihood of take-up of AOTC, and (c) likelihood of engagement with the communications materials. Since at the time of random assignment we only had access to data from the University, we constructed exact blocks based on students' year in school (including whether they transferred to the University or earned a bachelor's degree in 2019); their enrollment status at the University in Fall 2019; and whether they had registered an authorized payer to receive University emails.[6]

Random assignment occurred within these exact blocks with treatment probability consistent across blocks. Within each block, we randomly assigned 50 percent of the sample to a business-as-usual group and 50 percent of the sample to a communications-bundle group (N=9,530 and 9,541, respectively). Among students randomly assigned to the communications-bundle group, we further randomized the sender of the letter to be either the IRS or the University (25 percent of the overall sample each). We describe the sample and treatments in *Table 1* below.

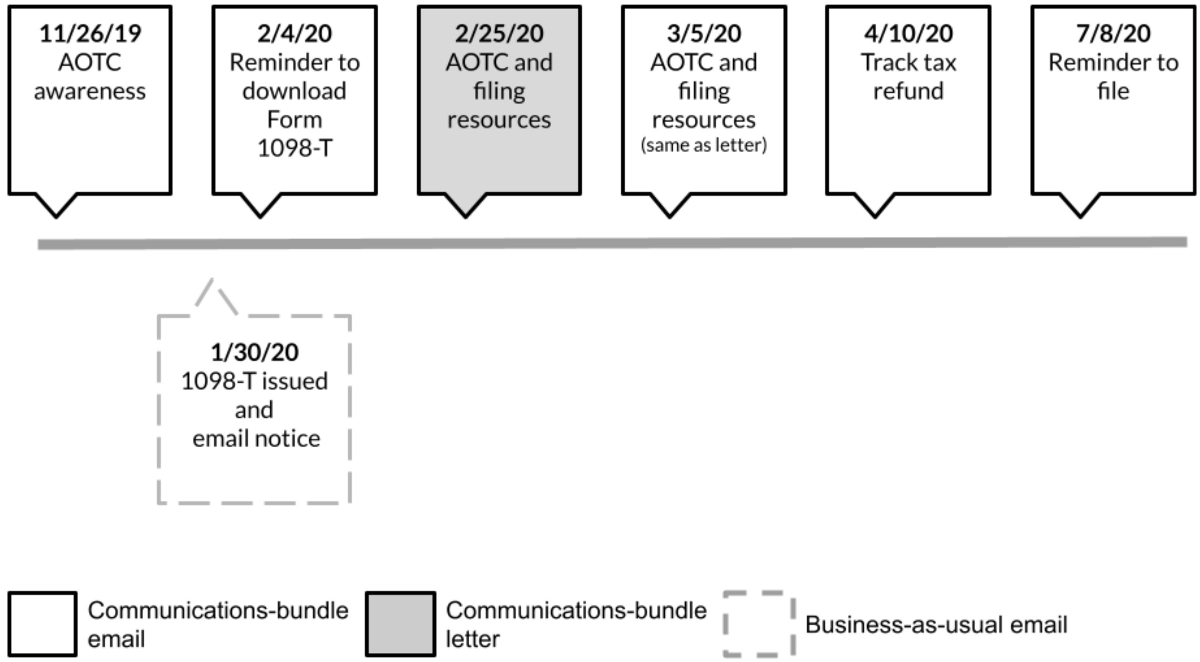
TABLE 1. Sample and planned treatments

Treatment Assignment	Description	Planned Sample Size
Business as usual	The University sends an email to students and (if applicable) authorized payers that the Form 1098-T is ready to download.	9,530
Communications bundle with IRS letter	*The business-as-usual communication:* The University sends an email to students and (if applicable) authorized payers that the Form 1098-T is ready to download. *The email bundle:* The University sends five AOTC emails to students and (if applicable) authorized payers in November 2019-July 2020. The IRS letter: IRS sends one letter about claiming AOTC to students' permanent address (or current address if they have no permanent address listed).	4,765
Communications bundle with University letter	*The business-as-usual communication:* The University sends an email to students and (if applicable) authorized payers that the Form 1098-T is ready to download. *The email bundle:* The University sends five AOTC emails to students and (if applicable) authorized payers in November 2019-July 2020. *The University letter:* The University sends one letter about claiming AOTC to students' permanent address (or current address if they have no permanent address listed).	4,776
Total communications-bundle sample		9,541
Total sample		19,071

Communications. The communications used in this study were designed to address specific behavioral barriers. The original study plan called for four emails to be sent over the course of the 2019–2020 academic school year. However, we adjusted the content and timeline for the emails to account for changes to the academic calendar and filing season due to the COVID-19 pandemic. Namely, we added a fifth email to the communications bundle that was sent in July 2020 that reminded students of the new filing deadline. In Figure 1, we summarize these elements.

[6] Because at the time of analysis we also had access to IRS tax records, we include additional covariates that predict likelihood of AOTC take-up and eligibility in our regression models, explained in further detail in the next section.

FIGURE 1. Communications bundle summary



Analysis

In the primary analysis, we compare AOTC take-up on TY 2019 returns between students in the business-as-usual and communications-bundle groups. In the exploratory analysis, we examine potential messenger effects by comparing AOTC take-up between students in the communications bundle group with an IRS letter and students assigned to the communications bundle group with a university letter.

> **Research Question #1.** *Does sending a coordinated communications bundle about AOTC to enrolled or recently enrolled college students increase take-up of AOTC?*

To answer this research question, we compare the likelihood that a student claims AOTC on their 2019 tax return among students sent the communications bundle compared to students sent only business-as-usual communications. We tested this hypothesis using the following OLS regression:

$$(1)\ Y_{ib} = \beta_0 + \beta_1(T_{ib}) + \pi Z'_{ib} + \alpha_b + \varepsilon_{ib}$$

where *i* indexes students in blocks *b,* and

Y_{ib} is the outcome of interest;

T_{ib} is an indicator for random assignment to a communications-bundle group;

Z'_{ib} is a vector of student-level characteristics;

α_b are block fixed effects; and

ε_{ib} is an error term.

The coefficient β_1 is the ITT estimate of the average treatment effect for being sent the communications bundle (including either letter) compared with not being sent the communications bundle. The null hypothesis is that $\beta_1 = 0$.

Blocking variables (α_b) capture:

- Whether a student was registered at the University in Fall 2019;
- Whether a student had a registered authorized payer; and

- A student's year in school, which also accounts for status as a first-time or transfer student.

The vector of additional covariates (Z'_{ib}) to improve precision includes:

- Dichotomous indicators for history of prior take-up of AOTC (i.e., first-time eligibility for AOTC based on 1098-T enrollment history, previously eligible and never claimed AOTC, claimed AOTC one to three times, and claimed AOTC four or more times);[7]
- Dichotomous indicators for tax filing behavior in TY 2018 (i.e., having filed taxes as an independent tax filer, having been claimed as a dependent, and not having filed taxes);
- Dichotomous indicators of income in TY 2018 (i.e., missing adjusted gross income (AGI) reported AGI less than $60,000; reported AGI between $60,000–$120,000; and reported AGI greater than $120,000); and
- A dichotomous indicator for employment at the University during TY 2019.[8]

Research Question #2. *Are there messenger effects on AOTC take-up of being sent a coordinated communications bundle with a letter from the IRS compared to being sent a coordinated communications bundle with a letter from the University?*

We test this hypothesis using a less restricted version of the OLS regression in equation (1), in which:

$$(2)\ Y_{ib} = \beta_0 + \beta_1(IRS_{ib}) + \beta_2(University_{ib}) + \pi Z'_{ib} + \alpha_b + \varepsilon_{ib}$$

where *i* indexes students in blocks *b*, and

Y_{ib}	is the outcome of interest;
IRS_{ib}	is an indicator for random assignment to the communications-bundle group with an IRS letter;
$University_{ib}$	is an indicator for random assignment to the communications-bundle group with a University letter;
Z'_{ib}	is a vector of student-level characteristics;
α_b	are block fixed effects for student enrollment and financial characteristics; and
ε_{ib}	is an error term.

The quantity $\beta_2 - \beta_1$ is an estimate of the difference in the average treatment effect for those sent a communications bundle with a university letter compared with those sent a communications bundle with an IRS letter. The null hypothesis is that $\beta_2 - \beta_1 = 0$. Again, our primary outcome is a dichotomous measure for claiming AOTC in TY 2019. The covariates (Z'_{ib}) and block fixed effects (α_b) are the same as in equation (1).

Preliminary Findings

In our preliminary analysis, we find a statistically significant effect of the communications bundle on AOTC take-up. AOTC take-up was 46.9 percent in the business-as-usual group compared to 48.5 percent in the communications-bundle group (p = 0.02; 95% CI [0.2 - 2.8]). Turning to messenger effects, AOTC take-up was 48.6 percent in the IRS-letter group and 48.3 percent in the University-letter group. When compared to the business-as-usual group, the effect is statistically significant for the IRS-letter group (p=0.04; 95% CI[0.1 - 3.3]) and marginally statistically significant for the University-letter group (p=0.09; 95% CI[-0.2 - 3.0]). However, there is no statistically discernible difference in effects on AOTC take-up between the IRS- and University-letter groups (p=0.74).

[7] The limit for claiming AOTC is four times; the last category reflects people who are no longer eligible from having reached (or attempted to surpass) this limit.

[8] Note that the categorical covariates described are mutually exclusive, and one group will be dropped from the regression due to collinearity. All are included in the analysis plan for completeness.

Conclusions and Future Research

In this pilot, we find that a multi-modal coordinated communications intervention can increase take-up of AOTC among current or recently enrolled college students. Moreover, the magnitude of these effects—1.5 percentage points—is similar to the average treatment effect of other behaviorally informed program changes implemented and evaluated at scale by government research units (DellaVigna and Linos (2022)). While our experimental design allows for credible causal estimates of the effects of the intervention, the implications of the results should be considered in light of the fact that the study coincided with the start of the COVID-19 pandemic. For instance, the events surrounding the pandemic could have shaped students' responses to the communications bundle and financial stability, including access to other sources of financial aid, which could have influenced the effects of the intervention on college outcomes that we plan to explore in future work.

The disruptions to the academic calendar and other pandemic-induced disruptions could have increased or decreased engagement with the intervention. On one hand, the cognitive load induced by the pandemic could have limited attention to tax-related communications and the ability of students to act on this information. On the other hand, the timely reminders included in the communications bundle could have been all the more effective as students struggled to juggle additional demands brought on by the pandemic. Perhaps most notably, as with most universities, the students in our sample switched to virtual learning in March 2020, and as a result many students likely moved back in with their families. Because of their new living arrangement, they may have been more likely to discuss tax filing with their families and coordinate their returns with the primary tax filer. Because these disruptions may have increased or decreased engagement with the communications, it is unclear whether a similar intervention would be more or less effective under more typical circumstances.

Two policy responses to the pandemic may also influence the external validity of our findings. First, the IRS extended the filing deadline for the 2019 tax year to July 15, 2020. Because the intervention reminded students of this later filing deadline, the intervention may have induced some students who otherwise would not have filed or claimed AOTC to file their taxes over the summer and receive their refund shortly thereafter. Second, Congress passed the Coronavirus Aid, Relief, and Economic Security (CARES) Act on March 27, 2020, with implementation beginning shortly after passage. Many students in the sample were likely eligible for aid through the CARES Act through Economic Impact Payments (EIP).[9] While students in our sample who were claimed as dependents for tax filing purposes were too old to qualify for the first and second EIP payments, students who filed their taxes as independent tax filers were eligible for up to $1,200 from the first EIP, distributed in April 2020, and $600 from the second EIP, distributed in December 2020. The intervention also may have increased take-up of these additional benefits by encouraging early filing and including steps for claiming AOTC as an independent tax filer. We plan to investigate the timing of filing and receipt of EIP in future analysis. The timing of aid receipt and amount of aid received could also influence college enrollment decisions, which we also plan to explore in future work.[10]

Exploring filing, preparation method, and the filing status of the student (dependent or independent) will also help situate the impacts of the coordinated communications bundle in the broader tax landscape. Similarly, the inclusion of recent graduates in the study sample provides an opportunity for additional insight and is, to our knowledge, a novel design component. Studying the effects of the coordinated communications bundle on taxpayers who had recently graduated, but were still potentially eligible to claim AOTC, can aid in understanding a subgroup that is often overlooked in studies of people entitled to higher education tax benefits.

Further exploring potential messenger effects could help inform future studies. For instance, colleges and universities would likely find it beneficial to understand the extent to which they could achieve similar results by implementing these communication efforts themselves or whether the IRS letter is necessary to prompt

---

[9] In addition, many students were likely eligible for emergency aid directly from the University via Higher Education Emergency Relief Funds (HEERF), another component of the CARES Act.

[10] We plan to measure progress toward degree completion at the intensive margin, measured by cumulative academic credits earned at the University between Spring 2020 and Fall 2020, and at the extensive margin, measured by degree completion or enrollment at the University as of Fall 2020. Our full analysis plan can be found here: https://oes.gsa.gov/projects/aotc/.

credit take-up. Even with the external validity concerns discussed above, follow-up analyses could also investigate the level of engagement with the various emails, as shown through open rates and click-through on embedded links. Assessing the level of engagement could help refine guidance for future communications and inform follow-on studies.

The IRS aims to help taxpayers understand their rights and responsibilities through proactive education and outreach (Internal Revenue Service (2018)). The results of this study suggest that partnering with academic institutions may be an effective way to deliver important messaging to a relevant population. While this paper covers only preliminary results, this study offers several opportunities to add to the broader literature. Future exploration of additional research questions can help inform the role of communications bundles in the larger tax and college landscape.

References

Bergman, Peter, Jeffrey T. Denning, and Dayanand Manoli (2019). "Is information enough? The effect of information about education tax benefits on student outcomes." *Journal of Policy Analysis and Management* 38, no. 3 (2019): 706–731.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu (2012). "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment." *The Quarterly Journal of Economics* 127, no. 3 (2012): 1205–1242.

Bhargava, Saurabh, and Dayanand Manoli (2015). "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment." *American Economic Review* 105, no. 11 (2015): 3489–3529.

Bird, Kelli A., Benjamin L. Castleman, Jeffrey T. Denning, Joshua Goodman, Cait Lamberton, and Kelly Ochs Rosinger (2021). "Nudging at scale: Experimental evidence from FAFSA completion campaigns." *Journal of Economic Behavior & Organization* 183 (2021): 105–128.

Bulman, George B., and Caroline M. Hoxby (2015). "The returns to the federal tax credits for higher education." *Tax Policy and Economy* 29, no. 1 (2015): 13–88.

Carrell, Scott, and Bruce Sacerdote (2017). "Why do college-going interventions work?" *American Economic Journal: Applied Economics* 9, no. 3 (2017): 124–151.

DellaVigna, Stefano, and Elizabeth Linos (2022). "RCTs to scale: Comprehensive evidence from two nudge units." *Econometrica* 90, no. 1 (2022): 81–116.

Dynarski, Susan, C. J. Libassi, Katherine Michelmore, and Stephanie Owen (2021). "Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students." *American Economic Review* 111, no. 6 (2021): 1721–1756.

Finkelstein, Amy, and Matthew J. Notowidigdo (2019). "Take-up and targeting: Experimental evidence from SNAP." *The Quarterly Journal of Economics* 134, no. 3 (2019): 1505–1556.

Goldin, Jacob, Ithai Z. Lurie, and Janet McCubbin (2021). "Health insurance and mortality: Experimental evidence from taxpayer outreach." *The Quarterly Journal of Economics* 136, no. 1 (2021): 1–49.

Gurantz, Oded, Jessica Howell, Michael Hurwitz, Cassandra Larson, Matea Pender, and Brooke White (2021). "A national-level informational experiment to promote enrollment in selective colleges." *Journal of Policy Analysis and Management* 40, no. 2 (2021): 453–479.

Guyton, John, Dayanand Manoli, Brenda Schafer, Michael Sebastiani, and Nick Turner (2017a). "Credits for College." *Proceedings of Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association* 110, pp. 1–20. National Tax Association, 2017.

Guyton, John, Pat Langetieg, Dayanand Manoli, Mark Payne, Brenda Schafer, and Michael Sebastiani (2017b). "Reminders and recidivism: Using administrative data to characterize nonfilers and conduct EITC outreach." *American Economic Review* 107, no. 5 (2017): 471–475.

Hoxby, Caroline, and Sarah Turner (2013). "Expanding college opportunities for high-achieving, low-income students." *Stanford Institute for Economic Policy Research Discussion Paper* 12 (2013): 014.

Internal Revenue Service (2018). *IRS Strategic Plan FY2018-2022*, Internal Revenue Service, https://www.irs.gov/about-irs/irs-strategic-plan.

Internal Revenue Service (2021). "American Opportunity Tax Credit," Internal Revenue Service, https://www.irs.gov/credits-deductions/individuals/aotc.

Linos, Elizabeth, Allen Prohofsky, Aparna Ramesh, Jesse Rothstein, and Matt Unrath (2020). *Can Nudges Increase Take-up of the EITC? Evidence from Multiple Field Experiments*. Working Paper No. w28086. National Bureau of Economic Research.

The Customer Experience for Small Business and Self-Employed Taxpayers

Kahoa Bonhomme, Kathleen Holland, and Janice Hu (IRS, Small Business/Self-Employed Division)

Introduction

In response to Executive Order 12862 and the Restructuring and Reform Act of 1998, the IRS has been measuring customer satisfaction after critical interactions and journeys with taxpayers. Customer satisfaction is the result of a series of customer experiences. This research is part of the effort to move from a customer satisfaction framework to a more inclusive customer experience framework. The objective of this paper is to provide an overview and glimpse into the small business and self-employed (SB/SE) taxpayer experience.

OMB Circular A-11, Section 280, defines Federal Government customer experience as a combination of factors that result from touchpoints between an individual, business, or organization and the Federal Government over the duration of an interaction and relationship.[1] To better understand this experience for SB/SE taxpayers, we reviewed taxpayer profiles, surveys, journey maps, call analysis, and focus groups. Our findings include the preferences, expectations, and opinions of SB/SE taxpayers, as well as the high points and challenges these taxpayers face. Results from this research provide the information necessary to make improvements to services and to increase taxpayer satisfaction and compliance.

We begin this paper with a review of the general SB/SE taxpayer population, followed by a closer look at five subsets of the population. Next, we discuss achieving a better understanding of the future customer experience through the integration of multiple data and research sources. We end this paper with conclusions and ideas for future research.

The SB/SE Taxpayer

There are approximately 57 million small business/self-employed taxpayers, making up roughly one-third of the overall taxpayer base (which includes non-business individuals, large corporations, exempt organizations, etc.).

Taxpayer Profile

The following bullet items characterize the SB/SE taxpayer population:[2]

- Individuals filing Form 1040 (U.S. Individual Income Tax Return), Schedules C, E, F, or Form 2106 (Employee Business Expenses)
    - ~ 41 million self-employed persons
    - ~ 7 million filers of employment, excise, estate and gift returns
    - ~ 9 million small businesses with assets of less than $10 million
- All other businesses with assets under $10 million

The primary source of material for the remainder of this section is the 2020 Customer Experience, Expectations, and Needs (CEEN) survey.[3] Details about the CEEN survey are provided in Box 1.

---

1 OMB Circular No. A-11, Section 280, Office of Management and Budget, accessed April 6, 2021, www.whitehouse.gov/wp-content/uploads/2018/06/s280.pdf.

2 "Small Business/Self-Employed Division At-a-Glance," Internal Revenue Service/IRS.gov, last modified April 8, 2021, https://www.irs.gov/about-irs/small-business-self-employed-division-at-a-glance.

3 ICF (February 2021).

Box 1: Customer Experience, Expectations, and Needs Survey

Background

In 2020, the inaugural Customer Experience, Expectations, and Needs survey was administered, with the purpose of capturing the needs, expectations, and experiences of SB/SE taxpayers and their third-party preparers. Outside of the CEEN survey, SB/SE surveys have been administered to taxpayers who had interactions with the IRS and were focused on obtaining satisfaction ratings and feedback. The CEEN survey goes beyond these existing surveys by including SB/SE taxpayers regardless of whether they contacted the IRS. In doing so, the CEEN survey includes a large segment of the population that traditional SB/SE customer satisfaction surveys miss.

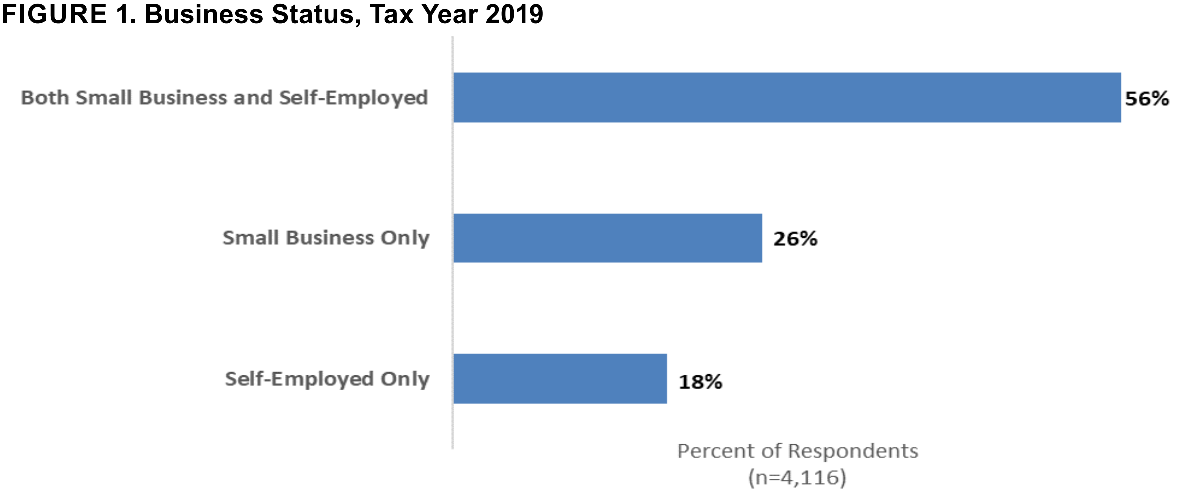
Survey Details

The CEEN survey was administered from August 20 to November 9, 2020, and used NORC's AmeriSpeak probability panel[4] and Lucid's non-probability online panel for sampling and data collection. There were 4,116 small business and/or self-employed respondents, all of whom filed a 2019 Federal tax return in 2020. Respondents took the survey online (4,088 respondents) or over the telephone (28 respondents). Calibration weighting was employed to reduce potential bias and yield more accurate SB/SE population estimates. AmeriSpeak Calibration combines probability-based AmeriSpeak and non-probability online samples using calibrating statistical weights derived from AmeriSpeak. Since not all sampled panel members respond to the survey screening questions, an adjustment is needed to account for survey screener non-response. This weight for eligible screener respondents is used to derive the benchmark for the adult independent contractor or small business owner taxpayer totals. ICF uses Race, Age, Education, Gender, Census Division, Internet Access, Urbanization, and screener question for weighting. ICF also worked with the IRS to identify additional weighting criteria from IRS specific data sources (e.g., type of tax return filed, schedules used) or internal administrative records.

Business Characteristics

Figure 1 shows the breakdown of small business and/or self-employed status among respondents to the CEEN survey.

FIGURE 1. Business Status, Tax Year 2019

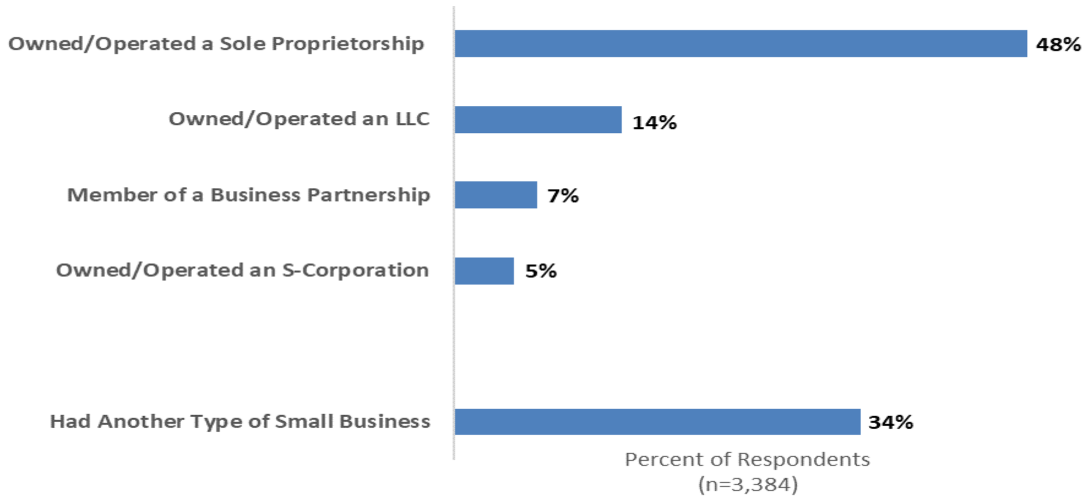


SOURCE: ICF (February 2021).

[4] The probability panel was constructed by drawing a representative sample to cover over 97 percent of U.S. households; the AmeriSpeak panel also includes non-internet households.

Figure 2 shows the types of businesses owned by small business respondents.

FIGURE 2. Small Business Type, Tax Year 2019



SOURCE: ICF (February 2021).

Of those who owned a small business, almost half (48 percent) owned or operated a sole proprietorship. Of the 34 percent that had another type of small business, several of the CEEN survey respondents indicated that they were performing gig work (e.g., tutoring or driving for Uber).

Table 1 lists the industries in which SB/SE respondents worked.

TABLE 1. Small Business Industry, Tax Year 2019

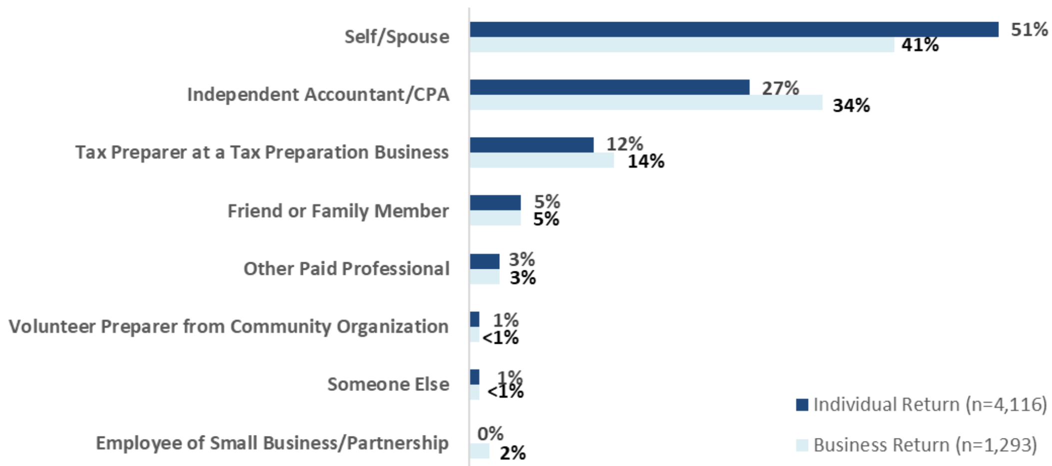
Industry	Percent (n=4,044)
Professional, scientific, technical and educational services	12%
Arts, entertainment, and recreation	10%
Retail or wholesale trade	10%
Construction	8%
Real estate and rental and leasing	8%
Health care and social assistance	6%
Accommodation and food services	4%
Transportation and warehousing	4%
Agriculture, forestry, fishing, and hunting	3%
Finance and insurance	3%
Manufacturing	3%
Other	28%

SOURCE: ICF (February 2021).

The professional, scientific, and educational services industry had the highest percentage of respondents (12 percent), followed by arts, entertainment, and recreation (10 percent) and retail or wholesale trade (10 percent). From the 28 percent who selected the “other” industry option, write-in responses included sales, writing, and technology-related work (web hosting, web design, and information technology (IT)).

Regarding preparation of their Federal income tax returns, the largest group of respondents either self-prepared or had a spouse prepare the return, as shown in Figure 3. This figure also shows that compared to individual returns (51 percent), separate business returns were less likely to be self/spouse-prepared (41 percent).

FIGURE 3. Preparer of Federal Income Tax Returns: Individual and Business,[5] Tax Year 2019



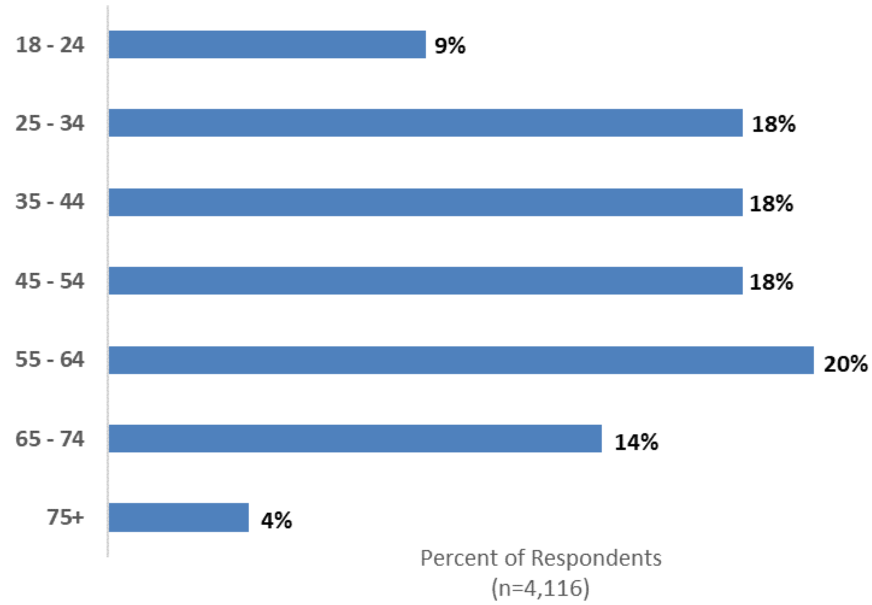
SOURCE: ICF (February 2021).

Demographics

Knowing the demographics of the SB/SE population increases our understanding of the customer experience and may lead to improved services and outreach efforts. Figures 4, 5, and 6 provide respondents’ age, education, and income, respectively.

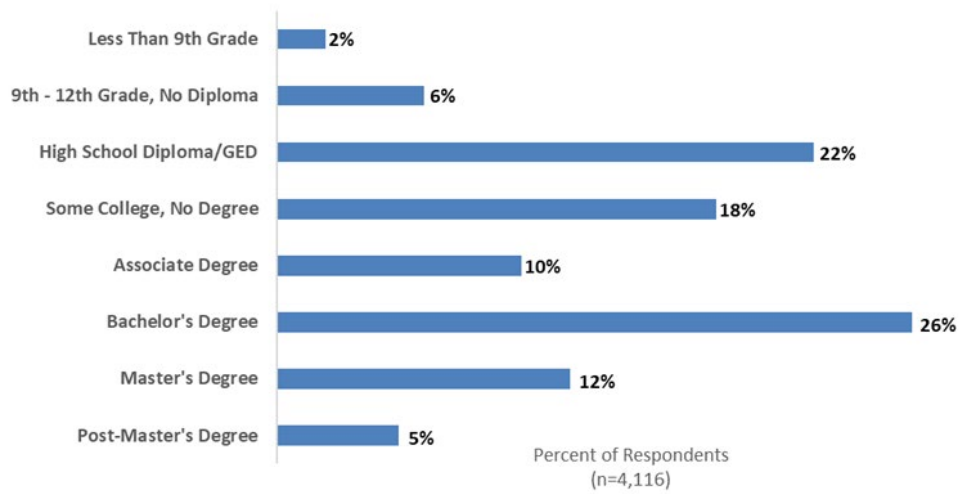
[5] Among CEEN respondents, 39 percent filed a business income tax return that was separate from their personal income tax return. The lower bars on Figure 3 show results for these respondents.

FIGURE 4. Age of Respondents, Tax Year 2019



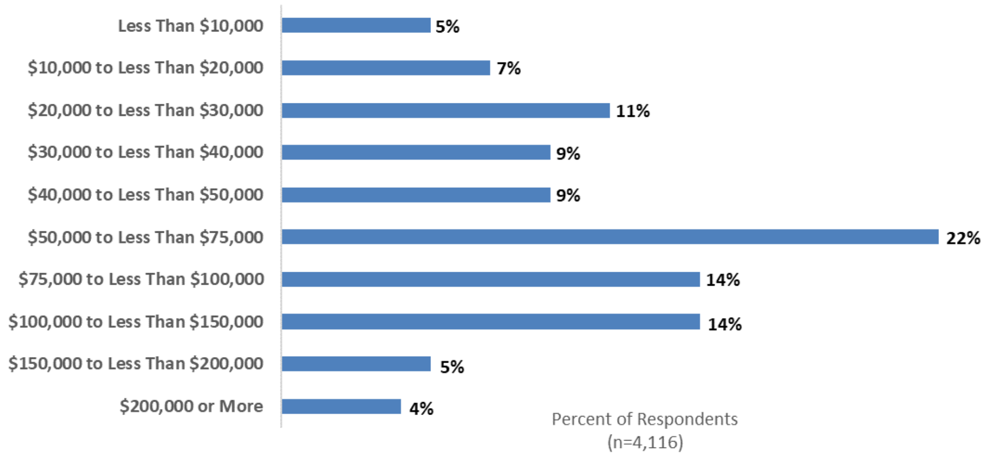
SOURCE: ICF (February 2021).

FIGURE 5. Education of Respondents, Tax Year 2019



SOURCE: ICF (February 2021).

FIGURE 6. Household Income of SB/SE Respondents, Tax Year 2019



SOURCE: ICF (February 2021).

Over half of the CEEN respondents (56 percent) were 45 years or older, and 45 percent were between 18 and 44 years old. As to highest level of education obtained, 43 percent of the respondents had a bachelor's degree or higher; an additional 28 percent had at least some post-high school education. Finally, 41 percent of the respondents had a household income of less than $50,000, 36 percent had income between $50,000 and $100,000, and 23 percent had income above $100,000.

Taxpayer Experience

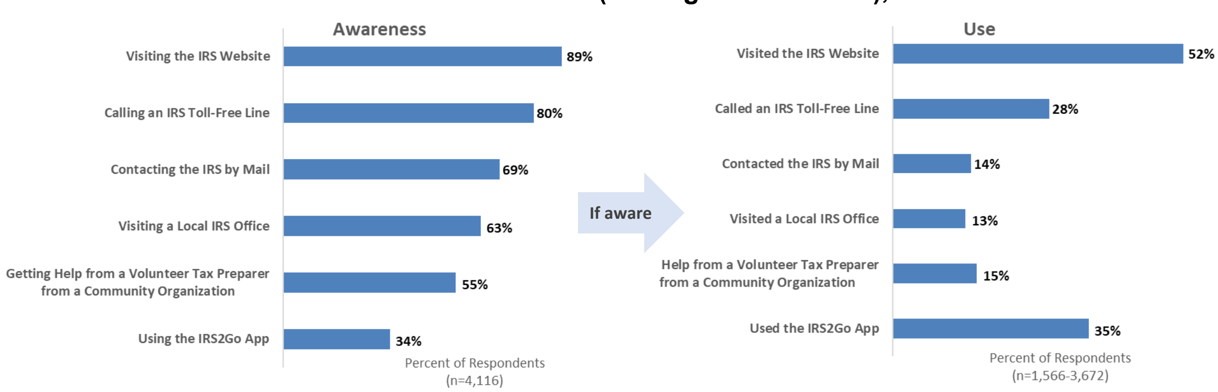
In addition to the business and demographic characteristics described above, the 2020 CEEN survey also captured multiple aspects of SB/SE taxpayers' pre-filing, filing, and post-filing experiences.

Communication with the IRS

For many taxpayers, communication is a major component of their IRS experience. The CEEN survey captured respondents' awareness of, use of, and satisfaction with IRS communication channels and tools.

Figure 7 shows the percentage of SB/SE respondents that were aware of and used various IRS channels. Figure 8 displays the satisfaction ratings for each channel.

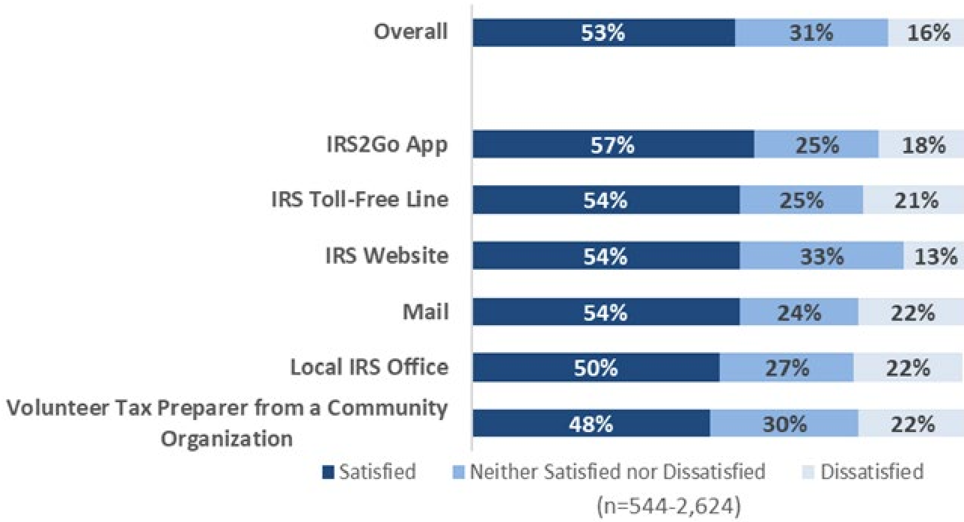
FIGURE 7. IRS Channels: Awareness and Use (Among Those Aware), Tax Year 2019



SOURCE: ICF (February 2021).

Overall, awareness of IRS channels is high. More than half of the respondents are aware of the various ways to contact the IRS, except for IRS2Go (34 percent awareness). The IRS website had the highest use among those aware of the site (52 percent use). In thinking about the gap between awareness and use, it is important to remember that not all taxpayers need to contact the IRS beyond filing their return.

FIGURE 8. IRS Channels: Satisfaction,[6] Tax Year 2019



SOURCE: ICF (February 2021).

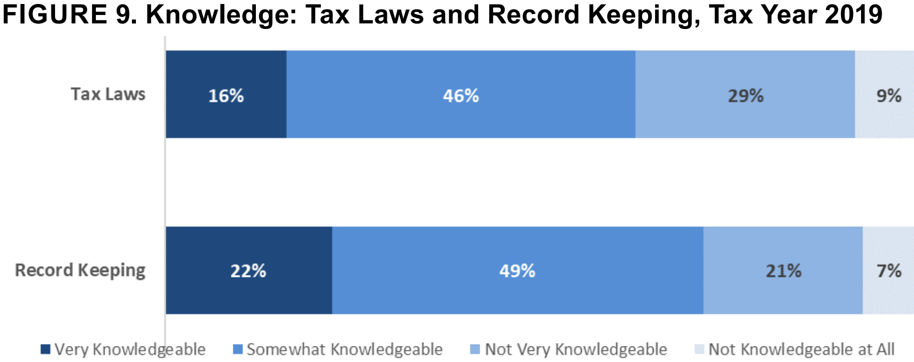
Among those who contacted the IRS in the last 12 months, overall satisfaction with IRS contact was 53 percent. Respondents who used the IRS2Go application (app) were most satisfied (57 percent). Respondents who got help from a volunteer preparer from a community organization were least satisfied (48 percent); however, there were no big differences in satisfaction among the modes of contact.

Tax Issues: Knowledge, Preferences, and Beliefs

An essential element of creating a successful experience is a solid understanding of the taxpayer viewpoint. The CEEN survey captured SB/SE taxpayers' knowledge about tax issues, preferences for getting general tax information, and beliefs about tax-related issues.

Figure 9 shows the self-reported level of respondents' knowledge about tax laws and record keeping for their self-employment or small business situation.

FIGURE 9. Knowledge: Tax Laws and Record Keeping, Tax Year 2019



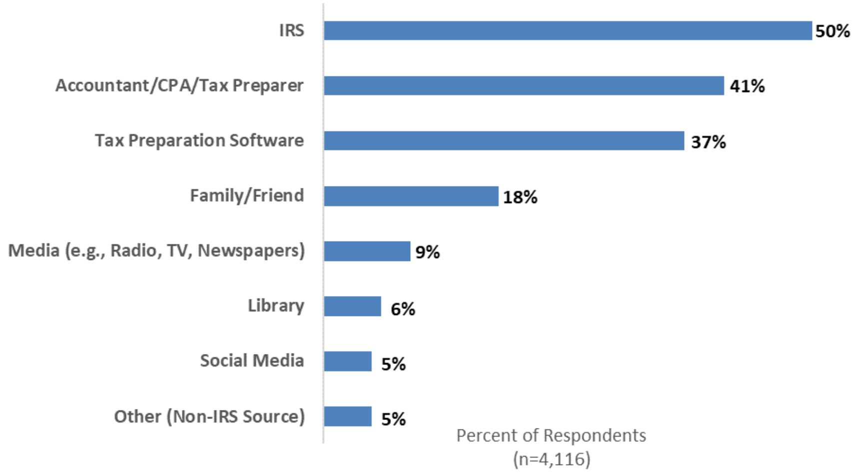
SOURCE: ICF (February 2021).

[6] Among CEEN respondents, 59 percent contacted the IRS in the last 12 months. Figure 8 shows results for these respondents.

Sixty-two percent of respondents felt either very or somewhat knowledgeable about the tax laws for their self-employment or small business situation, while 38 percent felt that they were not knowledgeable. As to record keeping, 71 percent felt either very or somewhat knowledgeable, while 28 percent felt that they were not knowledgeable.

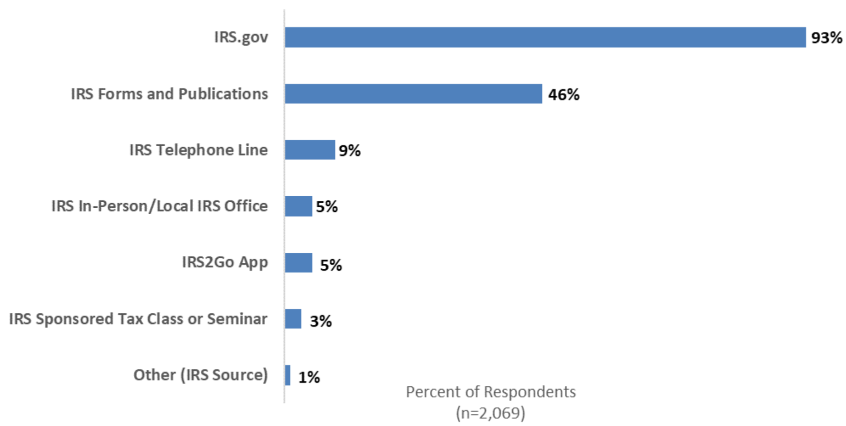
Another set of questions on the CEEN survey related to getting general tax information. Figures 10 and 11 show the current sources of general tax information, from all sources (Figure 10) and from only IRS sources (Figure 11).

FIGURE 10. Sources of General Tax Information, Tax Year 2019



SOURCE: ICF (February 2021).

FIGURE 11. IRS Sources of General Tax Information (Among Those Who Use IRS Sources), Tax Year 2019

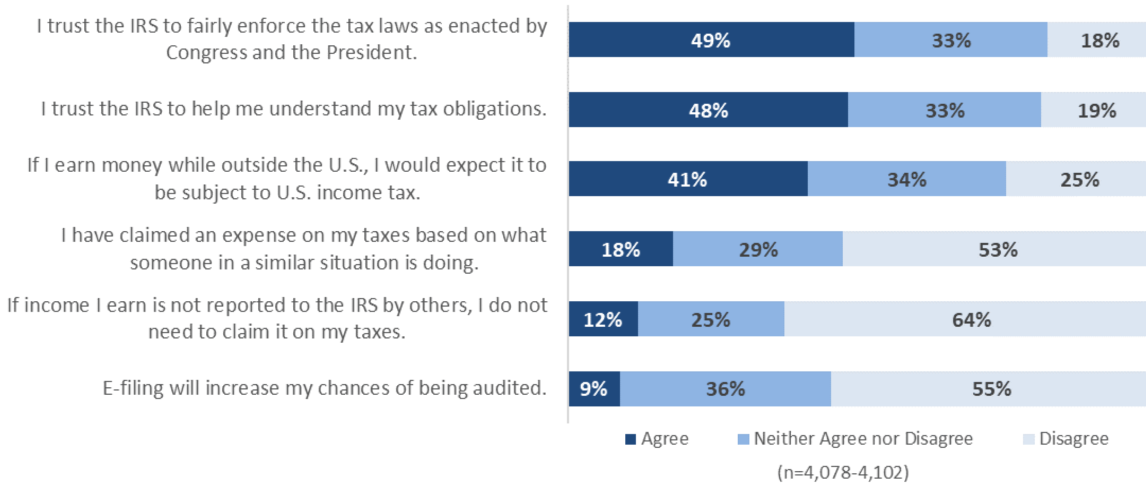


SOURCE: ICF (February 2021).

Half of all respondents got their general tax information from IRS sources. For those respondents, the top IRS sources of information were IRS.gov (93 percent) and IRS forms and publications (46 percent). Among non-IRS sources of information, the most common sources used were an accountant/CPA/tax preparer (41 percent) and tax preparation software (37 percent). When asked for their preferred sources of information, the majority of respondents (86 percent) stated that they prefer to get this information from the same sources they currently use (data not shown on graph).

To understand taxpayer attitudes and beliefs on specific tax-related matters, the CEEN survey asked respondents to assess a series of tax-related statements. Figure 12 summarizes the responses.

FIGURE 12. Taxpayer Attitudes and Beliefs: Agreement with Statements



SOURCE: ICF (February 2021).

Nearly half of all respondents (49 percent) agreed that they trust the IRS to enforce the tax laws enacted by Congress and the President, and a similar 48 percent agreed that they trust the IRS to help them understand their tax obligations. Forty-one percent of respondents agreed that money earned overseas is subject to U.S. income tax, and more than half of the respondents (55 percent) disagreed that e-filing increases the chances of being audited.

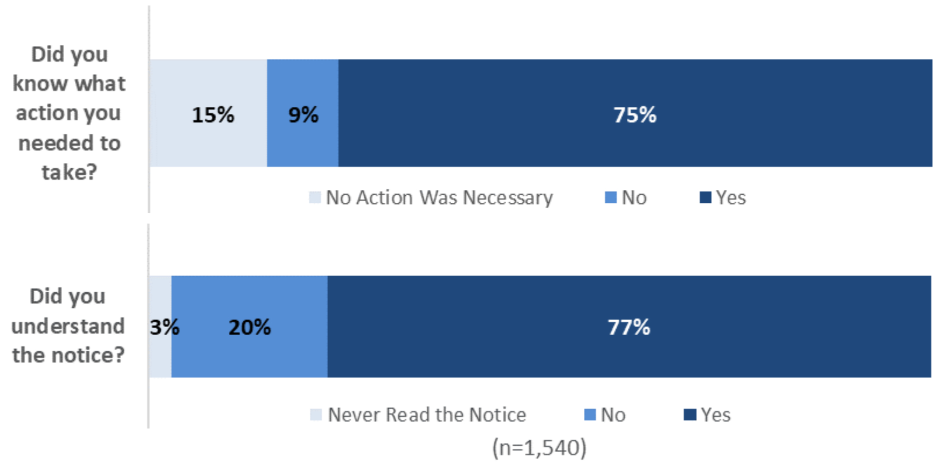
Notices

Satisfaction during critical touchpoints often leads to a successful customer experience; when individual touchpoints work well, the overall experience is more likely to be a positive one. IRS notices, which can serve as a proxy for compliance, are important touchpoints in the taxpayer experience. IRS notices are linked to specific issues on a taxpayer's federal tax return or account, such as changes to the account, requests for more information, or requests for payment. Reviewing taxpayer understanding of both the notice and the next steps and addressing the common reasons for the notice may improve the experience and prevent future non-compliance.

Nearly one-third (31 percent) of CEEN respondents had ever received a notice. Fourteen percent received a notice within the past 12 months, another 14 percent received a notice 1–2 years previously, and 23 percent received a notice 3 or more years before.

For respondents who had ever received a notice, two follow-up questions were asked: "Did you understand why you received the notice?" and "Did you know what action you needed to take?" Figure 13 shows the responses to these two questions.

FIGURE 13. IRS Notices: Understanding the Notice and Knowing Next Steps, Tax Year 2019

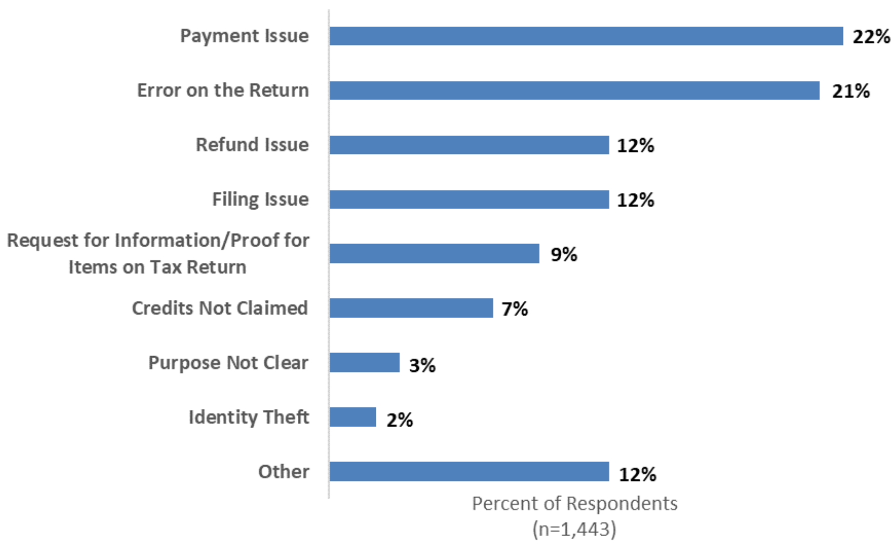


SOURCE: ICF (February 2021).

Three percent of notice recipients never read the notice. The majority of respondents (77 percent) who had received an IRS notice understood why they received their most recent notice. Among respondents who read the notice, 75 percent knew what action was needed.

Among the reasons for receiving the notice, the most common were payment issues (22 percent) and errors on the return (21 percent). Figure 14 provides these and the remaining reasons that respondents received their most reent notice.

FIGURE 14. IRS Notices: Reasons, Tax Year 2019



SOURCE: ICF (February 2021).

SB/SE Taxpayer Subgroups

In addition to studying the SB/SE population as a whole, we identified five important subpopulations of interest based on the distribution of responses and subpopulations that trend upward year over year.[7] Studying these subpopulations, whose knowledge, needs, and preferences often differ from those of the general SB/SE

[7] These five SB/SE subpopulations are not mutually exclusive.

population, allows us to uncover important details unique to each group and enables us to identify improvement opportunities that are tailored to each subgroup. This section covers five SB/SE subgroups: balance due taxpayers, employment tax filers, gig workers, Spanish-preferred taxpayers, and new SB/SE taxpayers.

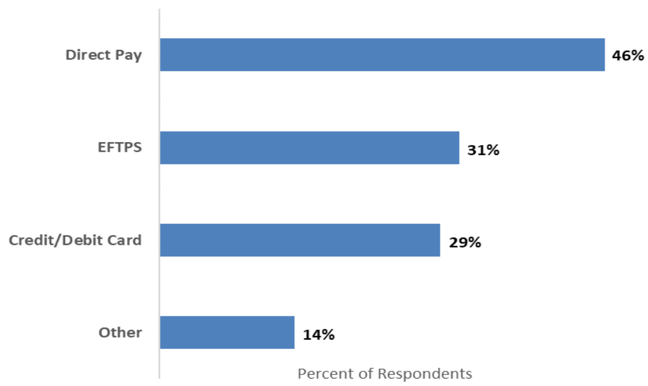
Balance Due Taxpayers

One subgroup of the SB/SE population is made up of taxpayers with a balance due that was not paid at the time of filing.

Business Characteristics

In Tax Year 2019, balance due taxpayers made up 17 percent of SB/SE respondents to the CEEN survey; of these taxpayers, 78 percent made payments toward their balance. Figure 15 shows the method used to make payments toward the balance due.

FIGURE 15. Balance Due Taxpayers: Method of Payment, Tax Year 2019



SOOURCE: ICF (February 2021).

Among those who made payments, the most common method was Direct Pay (46 percent).

Fifty-nine percent of CEEN respondents with a balance due had a payment plan. Among those with a payment plan, most (79 percent) indicated that it was either very easy or somewhat easy to track the balance of that payment plan.

The CEEN survey balance due respondents were asked about barriers to making timely payments to the IRS. Many respondents indicated that they did not face any barriers to making timely payments to the IRS. Of those who mentioned barriers, the most common were financial issues/lack of money to pay and hardships (such as Covid-19 impact or a death in the family).

Satisfaction

Balance due taxpayers' level of satisfaction with IRS interactions can be found in the annual SB/SE Customer Satisfaction (CSAT) survey reports. Additional information about the CSAT surveys is provided in Box 2.

Box 2: SB/SE Customer Satisfaction Surveys

Designed to measure external customer satisfaction, the SB/SE Customer Satisfaction surveys are administered to customers who have had recent interactions with the IRS. Surveys are conducted by mail or by telephone. The results from these surveys provide information on taxpayers' satisfaction with their IRS experience and include verbatim comments about the interaction. Taxpayer demographic information is also captured on the surveys.

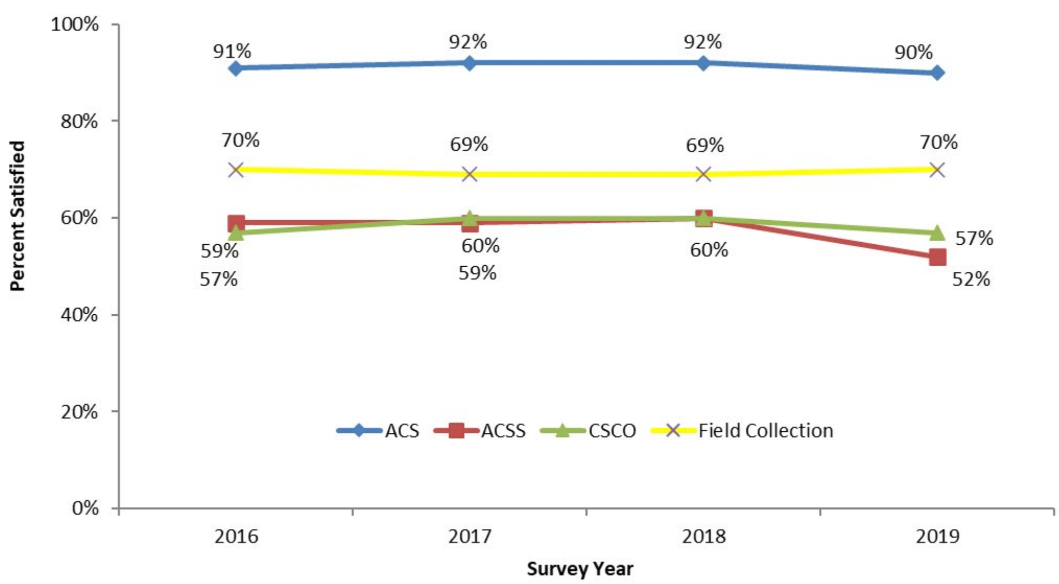
Each CSAT survey is tailored to the specific IRS function with which the taxpayer interacted. Balance due taxpayers might interact with any of the following four functions:

- Automated Collection System (ACS)
- Automated Collection System Support (ACSS)
- Compliance Services Collection Operations (CSCO)
- Field Collection

A random selection of taxpayers from each function is invited take the survey after the interaction. Data are weighted on the basis of distribution of cases among IRS sites/campuses.

Each CSAT survey captured the overall satisfaction of its respondents. Figure 16 shows the overall satisfaction of balance due taxpayers, broken up by function, over the past four survey years (SY16–SY19).[8]

FIGURE 16. Balance Due Taxpayers: Overall Satisfaction, Survey Years 2016–2019 (Percent Satisfied)



SOURCE: SB/SE Customer Satisfaction Survey Data (ACS, ACSS, CSCO, Field Collection), 2016–2020.

Overall satisfaction among balance due taxpayers varies by IRS function but is stable across years for each function. Satisfaction was highest for ACS taxpayers in each survey year from 2016 to 2019. For the most recent survey year (SY2019), overall satisfaction was highest for ACS taxpayers (90 percent), followed by taxpayers interacting with Field Collection (70 percent), CSCO (57 percent), and ACSS (52 percent).

[8] The survey year begins in April and ends in March of the following year.

In addition to overall satisfaction, the CSAT surveys captured information about specific aspects of the taxpayer experience. One area of high satisfaction among all four groups of balance due respondents was the treatment by the IRS representative in survey year 2019.

> For ACS, 96 percent of all respondents were satisfied with multiple aspects of the representative, including professionalism, courtesy, and willingness to help.
>
> For ACSS, CSCO, and Field Collection respondents, satisfaction with the representative ranged from 66 percent to 77 percent, which was higher than satisfaction percentages for other aspects of the interaction.

The following aspects had the highest levels of dissatisfaction in survey year 2019 (dissatisfaction percentages in parentheses):

- Clarity and tone of the notice received (ACS: 18 percent, CSCO: 57 percent)
- Time spent throughout the case, including
    - Length of time to complete the collection process (ACSS: 32 percent)
    - Length of time it took to get through to an employee (CSCO: 54 percent)
    - Length of time spent on the issue (Field Collection: 28 percent)

The CSAT survey results point to one of the challenges faced in integrating multiple data sources: large differences among a single group of taxpayers. The difference in satisfaction levels between ACS and the other three balance due groups could likely be the result of the surveys' timing. "Transaction" surveys measure the customer perspective after a single, stand-alone interaction, while "journey" surveys measure the customer perspective after completion of a multistage process. The ACS transaction survey is taken at the point of service, administered on the telephone immediately after the interaction. In contrast, the ACSS, CSCO, and Field Collection journey surveys are sent by mail after the case has closed; often, an extended period of time has passed between a series of IRS interactions and the receipt of the survey. Surveys taken directly after the service have traditionally had higher satisfaction levels than surveys taken later.

Preferences

Insights about the preferences of SB/SE balance due taxpayers come from two SB/SE Research projects: *IRS Balance Due Taxpayer Experience: Current State Analysis* (July 2020) and *IRS Balance Due Taxpayer: Taxpayer Vocabulary* (August 2020). Details about these projects are provided in Box 3.

Box 3: SB/SE Research Balance Due Taxpayer Projects

IRS Balance Due Taxpayer Experience: Current State Analysis

The objectives of the Current State Analysis project were to understand the current state of the balance due taxpayer experience and to explore best customer experience practices of both government and non-government entities. To achieve the objectives, SB/SE Research reviewed a wide range of literature, internal and external, pertaining to the balance due experience.

IRS Balance Due Taxpayer: Taxpayer Vocabulary

The objectives of the Taxpayer Vocabulary project were to summarize important characteristics of balance due taxpayer telephone calls and to identify improvement opportunities for the taxpayer telephone experience. To achieve the objectives, SB/SE Research listened to and analyzed 323 telephone calls made to the IRS by balance due taxpayers.

IRS Balance Due Taxpayer Experience: Current State Analysis

Our research uncovered a strong balance due taxpayer preference: completing tasks online, as opposed to over the telephone or in person. The bullet items below summarize the responses to several questions about IRS service channels asked on the 2020 Taxpayer Experience Survey (TES).[9]

- For finding information about making payments:
  - IRS.gov was used by 48 percent of TES respondents.
  - The IRS telephone line was used by 15 percent.
  - A local IRS office was used by 2 percent.
- For finding information about an IRS notice:
  - IRS.gov was used by 52 percent of TES respondents.
  - The IRS telephone line was used by 32 percent.
  - A local IRS office was used by 5 percent.

IRS Balance Due Taxpayer: Taxpayer Vocabulary

Taxpayers who run into difficulties with online tools often call the IRS telephone line for assistance; representatives are often unable to help. An analysis of taxpayer telephone calls, uncovered the following caller actions:

> Callers requested the telephone representative's help with diagnosing the online issue and/or completing their online tasks. Telephone representatives were unable to assist with online issues.
>
> Callers requested a telephone number for assistance with online tasks. Telephone representatives were unable to provide a number for online help.

These findings point to a balance due taxpayer need: support for online issues. Specifically, these taxpayers need a telephone number to call for online issues and/or a telephone representative trained to help with online issues.

Employment Tax Filers

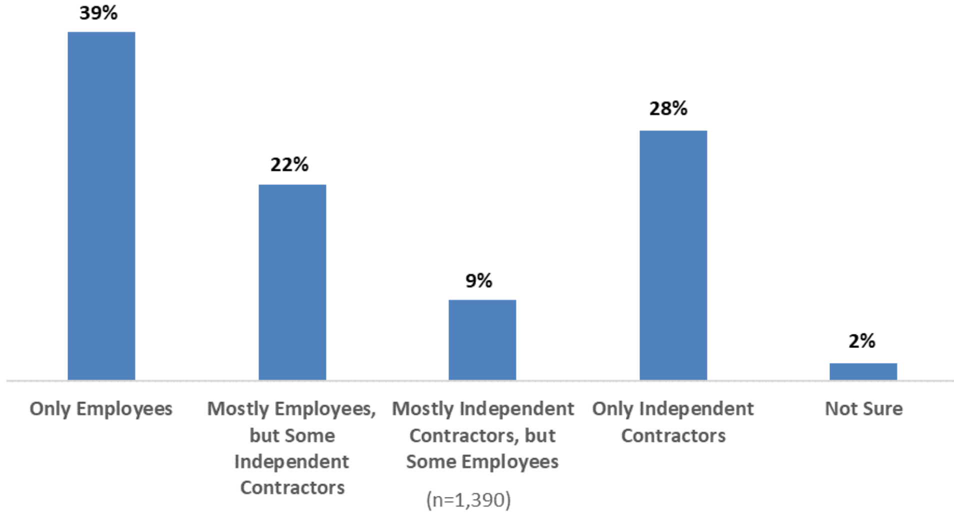
A second subgroup of the SB/SE population is made up of taxpayers who file employment tax returns.

Business Characteristics

Among CEEN respondents, 25 percent paid employees or independent contractors. A majority of these respondents (62 percent) did not use a payroll company. Among the 25 percent who paid workers (employees or independent contractors), the largest group (39 percent) paid only employees. Figure 17 provides this information, as well as the percentages for all other types of paid workers.

[9] These TES questions were asked of Form 1040 filers.

FIGURE 17. Employment Tax: Types of Workers Paid, Tax Year 2019



SOURCE: ICF (February 2021).

Respondents who paid workers were asked about several key employment tax forms. Table 2 shows information about these respondents' awareness and filing of Forms 940, 941, and 944.

TABLE 2. Employment Tax: Awareness and Filing of Forms, Tax Year 2019

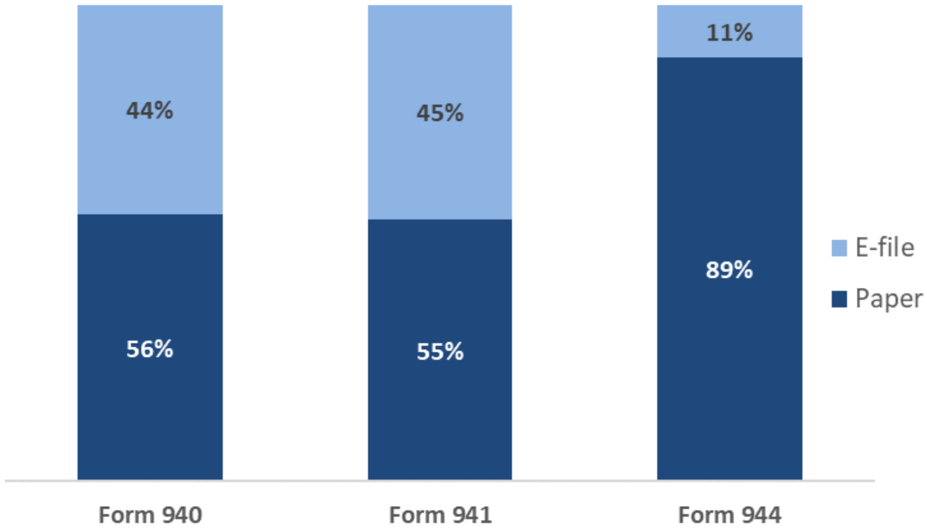
Employment Tax Form	Aware of Form	Filed Form *(among those aware)*
Form 940	53%	63%
Form 941	53%	64%
Form 944	55%	62%
	(n=1,391)	(n=880-924)

SOURCE: ICF (February 2021).

A little more than half of the respondents (53 percent to 55 percent) were aware of employment tax forms 940, 941, and 944. Of those aware of these forms, 62 percent to 64 percent filed them in 2019.

Although employment tax returns represent the largest overall volume of business returns, the Form 94X series continues to exhibit the lowest e-file rates among business returns. Figure 18 provides operational data on Calendar Year 2018 filing methods for Forms 940, 941, and 944.

FIGURE 18. Employment Tax: Filing Method, Calendar Year 2018



SOURCE: RAAS Publication 6149 2019 Final: Calendar Year Return Projections by State 2019–2026.

Slightly less than half of Form 940 (44 percent) and Form 941 (45 percent) filers e-filed their returns in 2018. The e-file rate for Form 944 was even lower, at 11 percent. These percentages all fall short of the IRS e-filing goal of 80 percent for all forms.

A question in the 2020 CEEN survey was directed to respondents who filed their employment tax returns by paper in the mail and asked for the reason behind this choice. Many respondents said that filing by paper was easier, that it was just their preference, or that their accountant made the choice. Other responses are listed below.

> *"Hard copy is easier to keep for records."*
>
> *"I tried to sign up for electronic filing but kept getting errors/having problems, so I gave up."*
>
> *"In this pandemic year, I was not sure about online [filing]. So I chose to file by paper."*
>
> *"I had to because of past identity theft."*
>
> *"I feel more secure that way."*

Additionally, results from an employment tax e-file survey[10] show that 36 percent of paper filers and 25 percent of e-filers dislike using a non-IRS website to e-file.

In addition to these personal reasons, there are outside factors that may impact the filing method of employment tax returns. A recent SB/SE Research report highlighted several possible external e-file factors. Information about this report is provided in Box 4.

Box 4: Employment Tax Environmental Scan

The objective of the Employment Tax Environmental Scan was to better understand external factors that impact the current low e-filing rate of employment tax returns (Form 94X series). To achieve the objective, the team scanned external literature, used internal and external data sources, and reviewed prior F94X research studies. Discussions with subject-matter experts also informed this research. External e-file factors covered in the report include payroll software, state-level influence, paid preparers, Covid-19, emerging markets, and technology drivers.

[10] Employment Tax E-File Survey for Tax Practitioners (Form 14710, OMB Number 1545-1432), November 2015.

Table 3 summarizes the key findings from a selection of the external factors explored in the Environmental Scan report.

TABLE 3. Employment Tax: External Factors Affecting E-file Behavior

External Factor	Finding
Payroll Software	With no direct e-file option through the IRS, taxpayers who want to e-file must engage with a third-party modernized e-file (MeF) provider to input and submit their forms. Navigating the many options to find an appropriate provider presents a challenge for taxpayers.
	Most software packages have a base cost between $159 and $1,695, with additional charges per e-filed form of $1.60-$22 per form. These costs may be prohibitive for employment tax filers.
Paid Preparers	Returns prepared by a paid preparer have a lower average e-file rate than that of the overall population. Large preparers are more likely to e-file returns (26 percent in TY2018) than are small preparers (11 percent in TY2018).[11]
	For paid preparers, the additional cost of e-filing is a barrier to e-filing.
	Preparers and taxpayers perceive limited incentives to e-file the prepared returns. In addition, they do not find their current paper filing process prohibitively expensive.

SOURCE: Employment Tax Environmental Scan, April 2021.

Taxpayer Journey

In 2020, SB/SE Research completed an Employment Tax Journey Map project which details the journeys of multiple employment tax personas. More information about this project is provided in Box 5.

Box 5: SB/SE Employment Tax Journey Maps

The SB/SE Employment Tax Journey Map project outlines the customer experience from the taxpayer's perspective. The objective of this research was to understand the employment tax return filing environment and what drives taxpayers' choice of filing method. Journey maps were created for five personas, including a new employer, an experienced employer, an employer who uses a payroll company, an employer who hires both employees and independent contractors, and an unbanked employer. Topics covered in the maps include on-boarding new employees, withholding, depositing, reporting, paying employment taxes, and filing Forms W-2 with the Social Security Administration.

One of the first steps of the employment tax filer's journey is the search for information. Many employers rely on online searches, which presents challenges due to both the numerous sources of information online and the potential for unreliable material. It is likely that before taxpayers see the IRS website, they will see links for ads; examples of ads include those used by paid preparers to solicit business and by financial institutions to offer loans. Highlighting this point, 88 percent of CEEN survey respondents indicated that they used a resource outside of the IRS to determine their estimated employment tax payment requirements. The comments below, offered by 2020 CEEN focus group participants, demonstrate some of the difficulties taxpayers face as they search for information.

> *"When I did a Google search on how to start your own business, they had all these different websites of people with blogs that were writing about what you should do, go to the IRS website, go to the state website, go to the city website to make sure there are no violations with the address where you're running your business from."*

> *"You have to search for yourself, there's nobody you can really trust, there's no reputable tax source."*

[11] Preparers were segmented on the basis of number of F94X prepared in TY2018. Preparers with more than ten F94X returns were classified as "large," and preparers with ten or fewer F94X returns were classified as "small."

Fundamental next steps in the taxpayer journey are the preparation and filing of tax returns. Several questions in the CEEN survey were asked of respondents who filed employment tax forms. The sentences below summarize the responses.

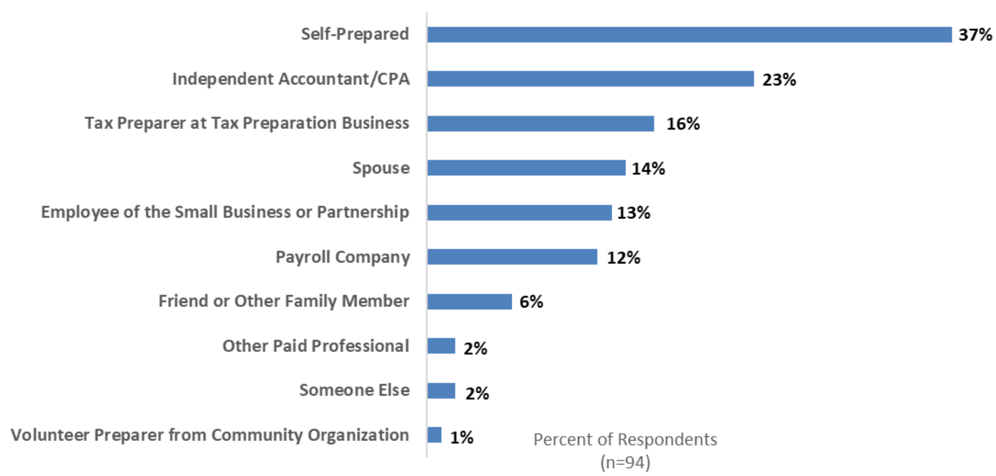
The majority of respondents (78 percent) felt that it was either very easy or somewhat easy to determine which employment (payroll) tax forms to file.

Most respondents (84 percent) felt that it was either very easy or somewhat easy to prepare their Federal employment (payroll) tax forms.

The majority of CEEN respondents (85 percent) used the same preparer for their Federal employment tax and income tax returns.

Figure 19 provides employment tax preparer information for CEEN respondents who used a different preparer for their employment tax and income tax returns.

FIGURE 19. Employment Tax: Preparer of Employment Tax Forms, Tax Year 2019



SOURCE: ICF (February 2021).

Of the respondents who used different employment tax and income tax preparers, 37 percent self-prepared their employment tax forms. An additional 23 percent used an independent accountant/CPA.

Another key step in the employment tax filer's journey is making estimated tax payments. Several questions in the CEEN survey were asked of taxpayers who paid quarterly taxes (n=1,298). The sentences below summarize the responses.

Almost half of all respondents (45 percent) said that they knew they needed to make quarterly tax payments because their tax preparer told them so.

The most common methods for paying quarterly estimated taxes were Direct Pay (44 percent of respondents) and Electronic Federal Tax Payment System (EFTPS) (27 percent of respondents).

Journey mapping has shown that employers can find it difficult to make quarterly payments on time. This pain point is felt more acutely by taxpayers who rely on third-party providers, such as the 38 percent of CEEN respondents who use a payroll company to prepare Federal employment tax forms. If a third-party provider fails to make payments, the taxpayer is likely to be liable for all taxes, penalties, and interest due. Often, the taxpayer is initially unaware of the failure, and is thus unable to take immediate action.

Finally, an important touchpoint in the taxpayer journey is contact with the IRS. Contacting the IRS by telephone generates a pain point that is common to employment tax filers: long wait times. This pain point extends beyond employment tax filers. Telephone data[12] and satisfaction survey data[13] from multiple types of taxpayers contain many examples of taxpayer dissatisfaction with the excessive wait times on IRS telephone calls. When asked about this issue, a 2020 CEEN focus group participant had the following input:

> *"I tried to reach out, but they say if you call, you're going to be on hold forever so there could be other ways to communicate with the IRS when we have an issue or a question."*

A 2020 CEEN survey respondent offered suggestions for this pain point.

> *"Text chat, online chat, and easier ways to contact and communicate with the IRS. Waiting on hold for 3 hours makes it difficult to be able to hold that long or have time to wait."*

Gig Workers

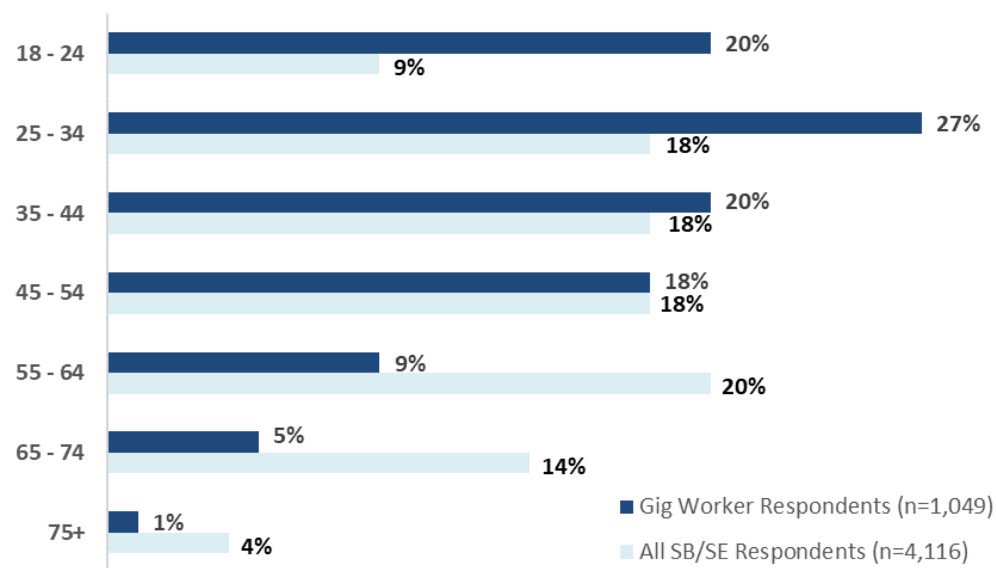
A third subgroup of the SB/SE population is made up of taxpayers who work in the gig economy.[14]

In Tax Year 2019, 15 percent of CEEN survey respondents reported having income from gig work.[15] The number of gig workers is likely to grow, evidenced by the addition of 6 million gig workers from 2010 to 2020.[16]

Demographics

Figures 20, 21, and 22 provide CEEN survey gig worker respondents' age, education, and income, respectively. The corresponding demographic information for all SB/SE CEEN respondents is also included in each figure.

FIGURE 20. Gig Workers: Age of Respondents, Tax Year 2019



SOURCE: ICF (February 2021).

[12] SOURCE: "IRS Balance Due Taxpayer: Taxpayer Vocabulary" report.

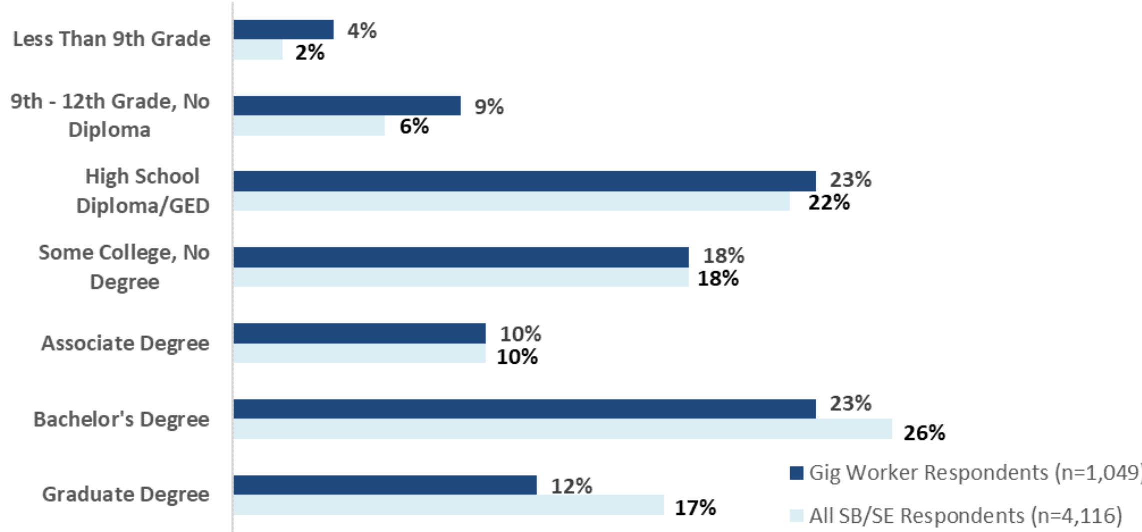
[13] SOURCE: CSAT survey data.

[14] There is no standard definition of a gig worker. Findings about gig workers from this report should not be compared to those from surveys with differing definitions of gig workers.

[15] The CEEN survey question used to determine whether a respondent was a gig worker was "Did you have any income from 'gig' or 'sharing economy' work in 2019, such as rideshare driver, pet sitting, rental host (such as Airbnb), or other on-demand work, services, or goods?"

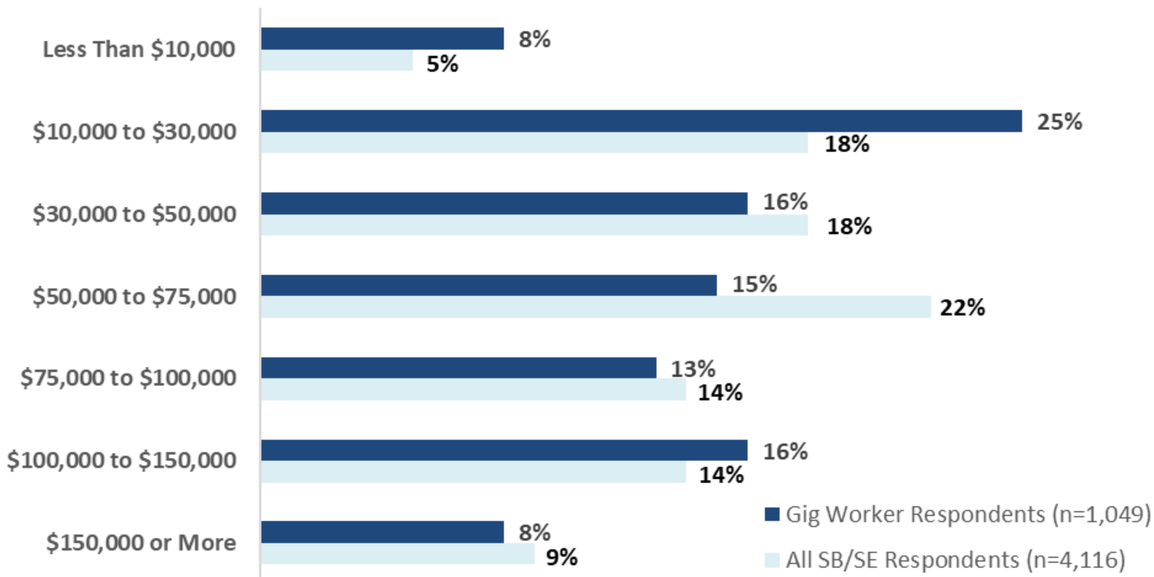
[16] ADP Research Institute (2020).

FIGURE 21. Gig Workers: Education of Respondents, Tax Year 2019



SOURCE: ICF (February 2021).

FIGURE 22. Gig Workers: Household Income of Respondents, Tax Year 2019



SOURCE: ICF (February 2021).

Almost half of the gig worker respondents (47 percent) were under 35 years old, compared to 27 percent of all SB/SE respondents. For the highest level of education obtained, the biggest difference between gig workers and all SB/SE respondents was the percentage with a graduate degree: 12 percent of gig workers vs. 17 percent of all respondents. Finally, 33 percent of gig workers had a household income of less than $30,000, compared to 23 percent of all respondents.

Tax Issues: Knowledge, Challenges, and Needs

Working in the gig economy can take many forms, including working in multiple industries, for multiple payers simultaneously, and in addition to a full-time job or school.[17] The wide range of scenarios and multiple sources of information may limit gig workers' understanding of their tax requirements and could hinder their ability to comply with their tax obligations.

One possible repercussion of this uncertainty is an increased likelihood of having a balance due. Among CEEN respondents, gig workers are more likely than non-gig workers to have a balance due. In Tax Year 2019, 34 percent of gig workers had an unpaid balance due to the IRS that was not paid at the time of filing, compared to 14 percent of non-gig workers.

To determine the level of knowledge among gig workers, the CEEN survey contained questions about specific tax issues. Asked of respondents with gig income,[18] these questions centered around the ease or difficulty of (1) figuring out how to report the gig income on their taxes and (2) figuring out what expenses from the gig work could be claimed on their taxes. Of respondents with gig income, the majority felt that it was either very easy or somewhat easy to make these determinations: 79 percent for income reporting and 74 percent for claiming expenses.

To gain a deeper understanding of gig workers' unique challenges and needs, SB/SE Research recently held focus groups with gig workers. More information about these focus groups is provided in Box 6.

Box 6: Gig Economy Worker Focus Groups

In February 2021, SB/SE Research conducted focus groups with gig economy workers. A total of 56 gig workers from across the country participated in one of six focus group sessions. The purpose of the focus groups was to better understand the experiences of gig workers, including knowledge about, challenges with, and needs for filing tax returns for gig income. Information gained from this study will help the IRS target communication and outreach efforts to increase tax awareness, knowledge, and compliance among workers in the gig economy.

The major findings from the six gig worker focus groups are listed in the sentences below.

- Gig workers received little to no tax-related assistance from their app-based company.
- Instead of using IRS resources, gig workers turned to Google, social media, tax professionals, and friends or family for answers to their questions.
- Participants wanted to know what they can claim as deductions when filing taxes, what they should be tracking throughout the year, and how best to track that information.
- Knowledge of the option to file quarterly taxes was not widespread.
- Participants found Publication 5369 to be a good overview of filing taxes as a gig worker but suggested improvements with specific details.

The gig worker participants brought up five of their biggest challenges, listed below.

- Finding and validating information for all allowable expenses
- Identifying what is considered taxable income
- Failing to receive tax documents, tax-related education, and filing information from many of the platform companies
- Understanding basic tax language in IRS resources, such as "employment tax"
- Learning to document detailed and accurate books and records

[17] IRS, TM1-0347 (April 2021).

[18] Examples of gig work roles provided in the CEEN survey included rideshare driver, pet sitter, and rental host (such as Airbnb).

Focus group participants stated several important needs, listed below.

- More details and hyperlinked resources in Publication 5369
- A step-by-step webinar or video to train gig workers on filing their taxes
- More guidance on what qualifies as acceptable write-offs/deductions, including a list of common deductions
- Advice on record keeping
- Specific information about filing taxes for different types of gig work, for workers holding multiple gigs, and for workers doing gig work in addition to a main job

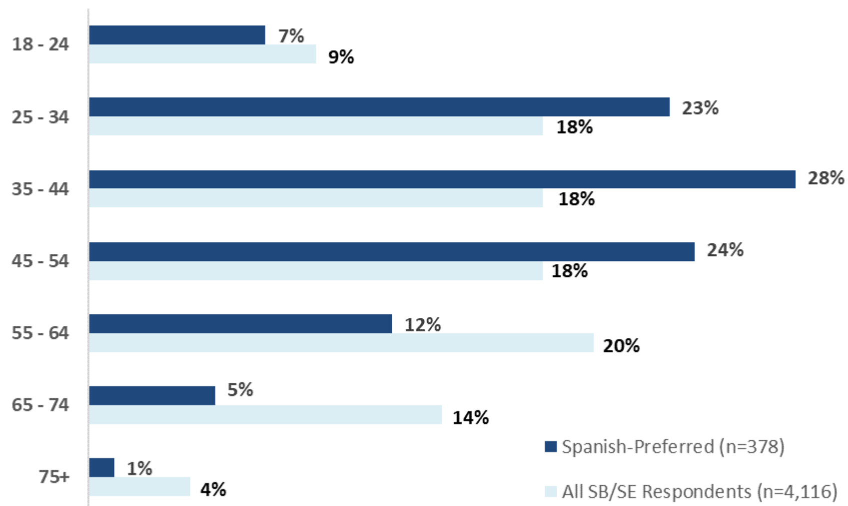
Spanish-Preferred Taxpayers

A fourth subgroup of the SB/SE population is made up of Spanish-speaking taxpayers. For the 2020 CEEN survey, 378 respondents chose to take the CEEN survey in Spanish. Results for this Spanish-preferred group[19] are presented in this section of the report.

Demographics

Figures 23, 24, and 25 provide CEEN Spanish-preferred respondents' age, education, and income, respectively. The corresponding demographic information for all SB/SE CEEN respondents is also included in each figure.

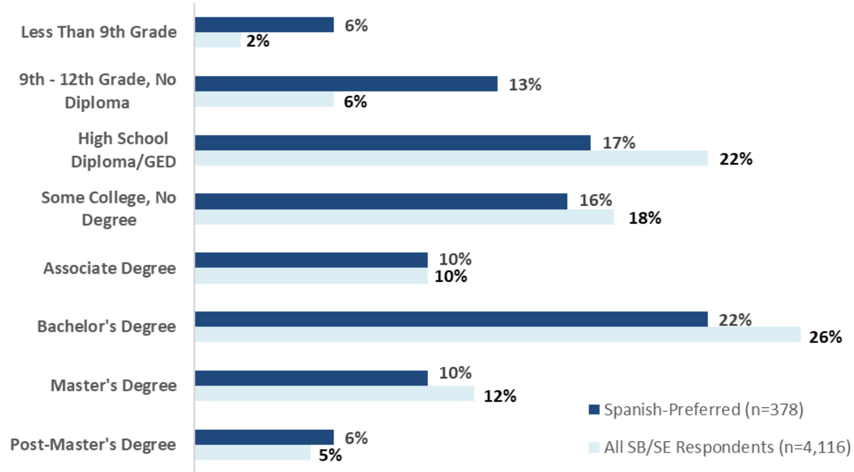
FIGURE 23. Spanish-Preferred: Age of Respondents, Tax Year 2019



SOURCE: ICF (February 2021).

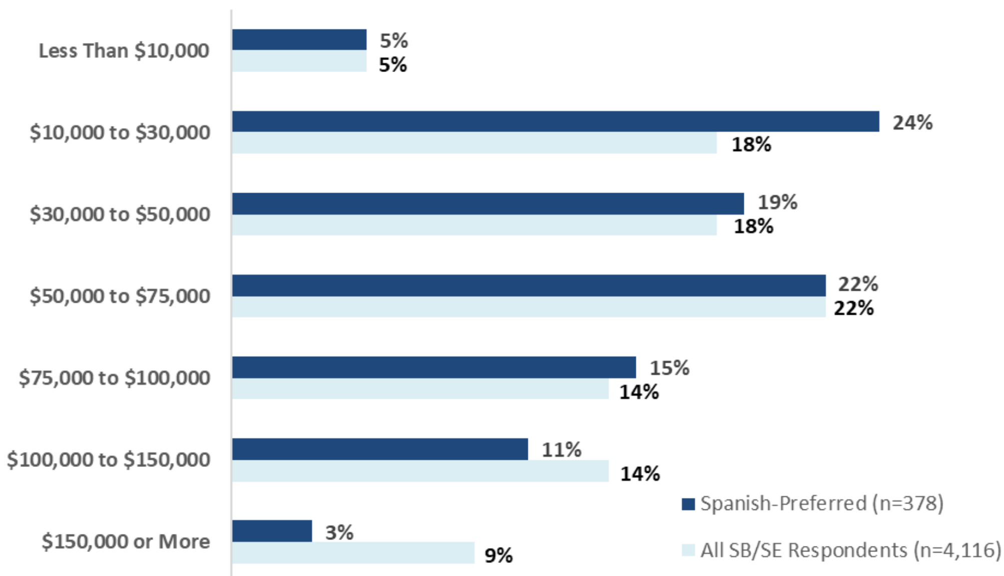
[19] Note: Spanish-preferred is not equivalent to Spanish Limited English Proficiency (LEP).

FIGURE 24. Spanish-Preferred: Education of Respondents, Tax Year 2019



SOURCE: ICF (February 2021).

FIGURE 25. Spanish-Preferred: Household Income of Respondents, Tax Year 2019



SOURCE: ICF (February 2021).

There was a larger percentage of Spanish-preferred respondents in the 25-54 age range (75 percent) than among all SB/SE respondents (54 percent). Nineteen percent of Spanish-preferred respondents had less than a high school education, compared to 8 percent of all respondents. Finally, 48 percent of Spanish-preferred respondents had a household income of less than $50,000, compared to 41 percent of all respondents.

Communication with the IRS

Awareness of most IRS channels was similar between Spanish-preferred and all CEEN respondents. The channels with the highest awareness among both groups were the IRS website (84 percent for Spanish-preferred and 89 percent for all respondents) and the toll-free IRS telephone line (76 percent for Spanish-preferred and 80 percent for all respondents).

Use of each IRS channel differed between Spanish-preferred and all respondents, with the exception of the IRS website (51 percent for Spanish-preferred and 52 percent for all respondents). The biggest difference was between the percentages using the toll-free IRS phone line. Among Spanish-preferred respondents who were aware of this channel, 47 percent used it. Among all respondents, 28 percent of those aware of this channel used it.

With regard to contacting the IRS, 67 percent of Spanish-preferred respondents contacted the IRS in the last 12 months, compared to all respondents (59 percent). Among the respondents who contacted the IRS in the last 12 months, a larger percentage of Spanish-preferred (62 percent) were satisfied with contacting the IRS, compared to all respondents (53 percent).

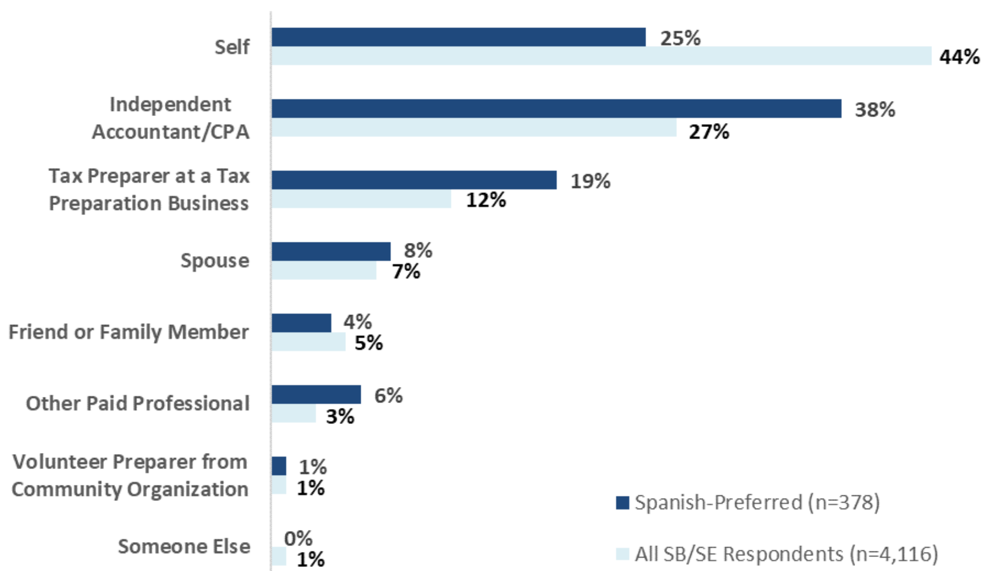
Several questions on the CEEN survey asked for respondents' assessments of communication-related statements. The sentences below summarize the responses.

- A larger percentage of Spanish-preferred respondents (66 percent) want the IRS to conduct outreach about tax filing requirements, compared to all SB/SE respondents (28 percent).
- A higher percentage of Spanish-preferred respondents (33 percent) agreed that contacting the IRS will increase their chances of being audited, compared to all SB/SE respondents (15 percent).
- A higher percentage of Spanish-preferred respondents (26 percent) agreed that e-filing will increase their chances of being audited, compared to all SB/SE respondents (9 percent).

Tax Preparation

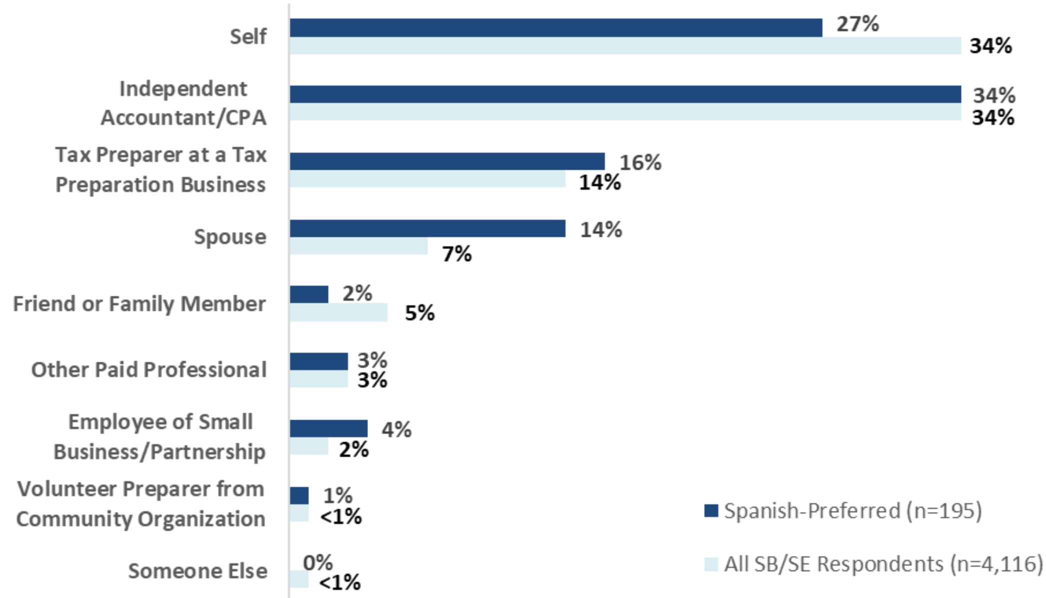
Spanish-preferred taxpayers are more likely than the general SB/SE taxpayer to use a paid preparer. This information is shown in Figures 26 (individual returns) and 27 (business returns), which also show the remaining preparer options.

FIGURE 26. Spanish-Preferred: Preparer of Federal Individual Income Tax Return, Tax Year 2019



SOURCE: ICF (February 2021).

FIGURE 27. Spanish-Preferred: Preparer of Federal Business Income Tax Return,[20] Tax Year 2019



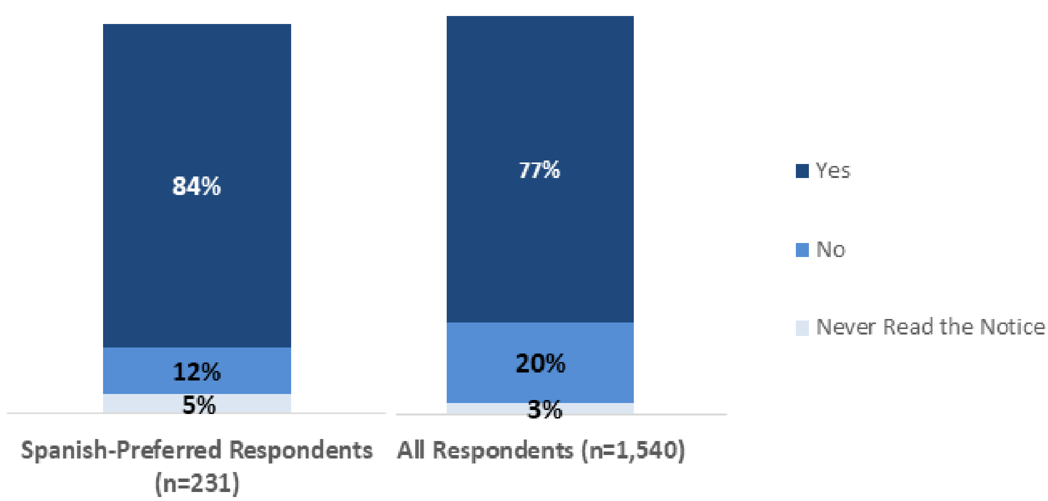
SOURCE: ICF (February 2021).

Notices

Nearly one-fourth (24 percent) of Spanish-preferred CEEN respondents received a notice within the past 12 months, compared to 14 percent of all respondents. When asked if they had ever received a notice (regardless of timeframe), 43 percent of Spanish-preferred respondents stated that they had, compared to 31 percent of all respondents.

For respondents who had ever received a notice, two follow-up questions were asked: "Did you understand why you received the notice?" and "Did you know what action you needed to take?" Figures 28 (understanding) and 29 (action) show the responses to these two questions for Spanish-preferred and all respondents.

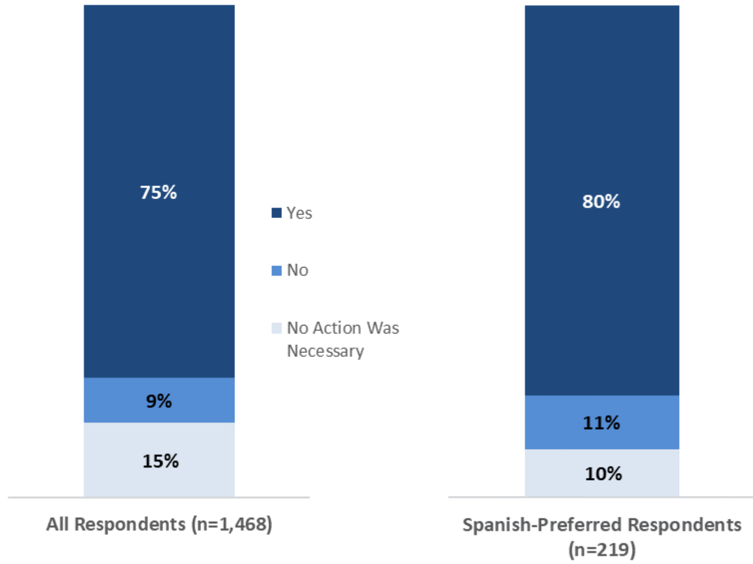
FIGURE 28. Spanish-Preferred: Understanding the IRS Notice, Tax Year 2019



SOURCE: ICF (February 2021).

[20] Among CEEN Spanish-preferred respondents, 73 percent filed a business income tax return that was separate from their personal income tax return. Figure 27 shows results for these respondents.

FIGURE 29. Spanish-Preferred: Knowing Next Steps After Notice, Tax Year 2019



SOURCE: ICF (February 2021).

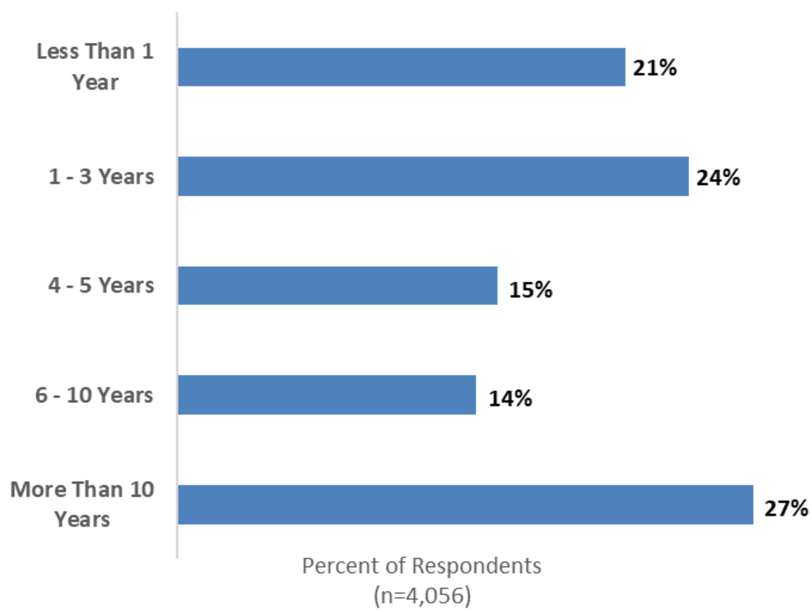
The majority of respondents who had received an IRS notice in both groups understood why they received their most recent notice and what action to take.

New SB/SE Taxpayers

The final SB/SE subgroup covered in this report is made up of taxpayers who have been a small business owner/partner or self-employed for less than one year.

Among all CEEN respondents, 21 percent had been in business for less than one year. This information is shown in Figure 30, which provides the breakdown of CEEN respondents by years in business.

FIGURE 30. Years in Business, Tax Year 2019



SOURCE: ICF (February 2021).

Almost half of all respondents (45 percent) have been an owner/partner in a small business or self-employed for 3 years or less. The percentage jumps to 60 percent when including respondents who have been in business or self-employed for 5 years or less.

The U.S. Small Business Administration provides the following statistics on the survival rate of new small businesses in the U.S.:[21]

- From 2008 to 2018, an average of 78.7 percent of new establishments survived one year.
- About half of all establishments survive five years or longer.
- About one-third of all establishments survive at least 10 years.

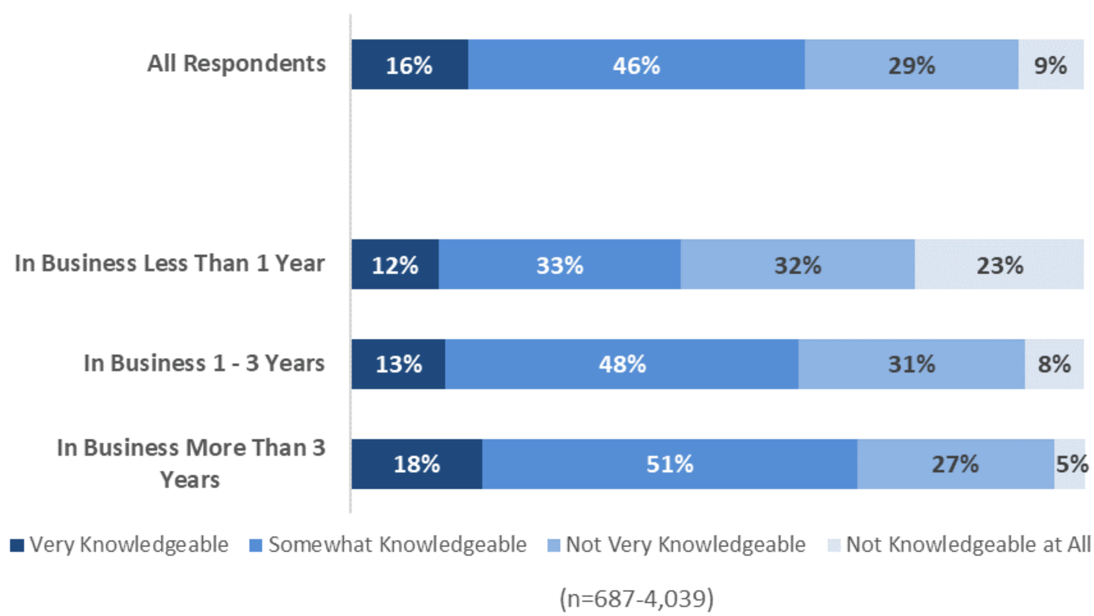
Lack of financial self-efficacy and/or knowledge may factor into why many new small businesses do not survive past the first several years. The following sections provide information on the knowledge, confidence, and pain points of new SB/SE taxpayers.

Taxpayer Journey

For many new small businesses, the taxpayer journey begins with a search for tax information. The Employment Tax Journey Map shows that a new employer who is seeking tax information is likely to encounter multiple perspectives; information from these sources is likely to have varying degrees of accuracy. Sources may include preparers, financial institutions (regarding guidance, loans, etc.), friends, family members, and other business owners. The search for information is often more challenging for new taxpayers than for more experienced ones.

Figure 31 provides the CEEN respondents' self-reported knowledge about the tax laws regarding their self-employment and/or small business situation. In addition to responses from new SB/SE taxpayers, responses for those who have been in business longer and responses for the overall SB/SE respondent population are provided.

FIGURE 31. New SB/SE: Knowledge of Tax Laws, Tax Year 2019



SOURCE: ICF (February 2021).

Self-reported knowledge of tax laws is lower for new SB/SE taxpayers. For those who have been in business for less than 1 year, 55 percent were either not very knowledgeable or not knowledgeable at all, compared to 39 percent of those in business for 1 to 3 years and 32 percent of those in business for more than 3 years.

[21] U.S. Small Business Administration Office of Advocacy: Frequently Asked Questions about Small Business, September 2019.

Another step in the taxpayer journey is making employment decisions. For new employers, trying to determine whether their paid workers are employees or independent contractors is often a difficult one. Sixteen percent of CEEN respondents found it either somewhat difficult or very difficult to make this determination. Respondents who felt it was difficult offered suggestions on how the IRS could make it easier. Examples are listed below.

> *"Have a simple criteria [sic] for employee/contract definitions."*

> *"Improve the language on [IRS.gov] to make it easier to understand and search for questions I have."*

> *"Put more scenarios on the website."*

Less than half of all CEEN respondents who paid workers (47 percent) were aware of Form SS-8. Among those aware of Form SS-8,[22] less than half (43 percent) had ever used it. Respondents who had used this form had specific suggestions for improvement.

> *"Make it easier to understand and have more room to write numbers."*

> *"The form is quite complicated and hard to understand without getting help from a tax professional."*

The next critical step in the new taxpayer journey is meeting tax obligations. An important component of this step is understanding tax requirements and deposit schedules. Sixteen percent of CEEN respondents felt that it was either somewhat difficult or very difficult to determine whether they had to pay estimated quarterly taxes. It may become very easy for new businesses to inadvertently leave their balance due unpaid at the time of filing. Several CEEN focus group participants had feedback on this issue.

> *"I'm sometimes left behind in the dark for a little bit longer than it needs to be."*

> *"It takes a lot of time away from trying to run a business when dealing with the overly complicated requirements and forms."*

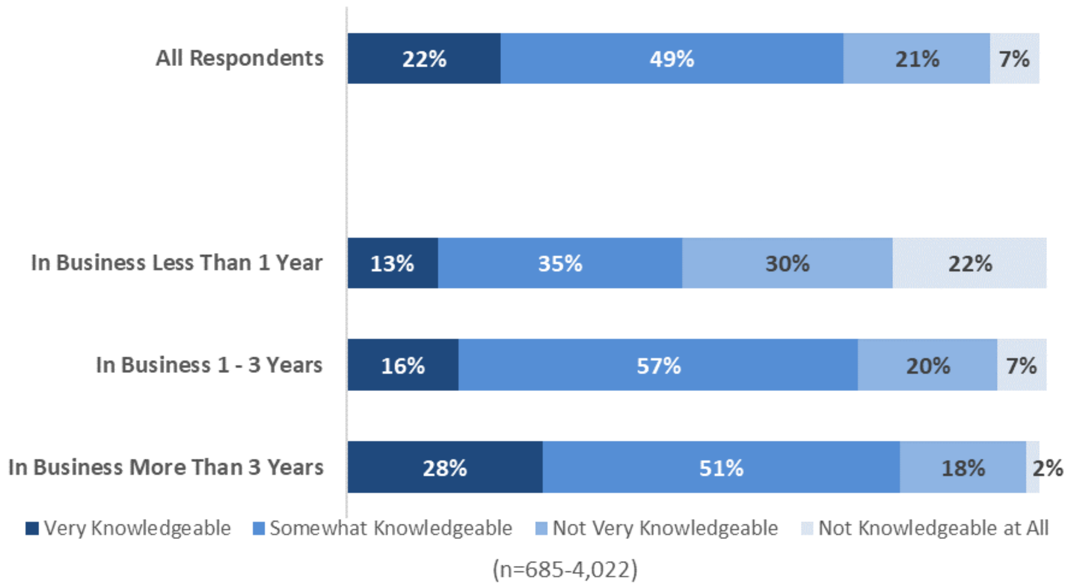
Record keeping is another major component of meeting tax obligations. CEEN respondents were asked several questions about record keeping for their self-employment or small business situation, the first of which was about the ease of determining what records to keep for their taxes. The sentences below summarize the responses to this question.

- Among respondents in business for **less than 1 year**, 54 percent felt it was either very easy or somewhat easy to make the determination.
- Among respondents in business for **1 to 3 years**, 73 percent felt it was either very easy or somewhat easy to make the determination.
- Among respondents in business for **more than 3 years**, 78 percent felt it was either very easy or somewhat easy to make the determination.
- Among **all respondents**, 72 percent felt it was either very easy or somewhat easy to make the determination.

Respondents were also asked about their knowledge and confidence in record keeping. Figure 32 shows the responses to the question "How knowledgeable are you about the record keeping requirements for your self-employment and/or small business taxes?"

[22] IRS Form SS-8 requests that the IRS make the determination about whether a worker is an employee or an independent contractor.

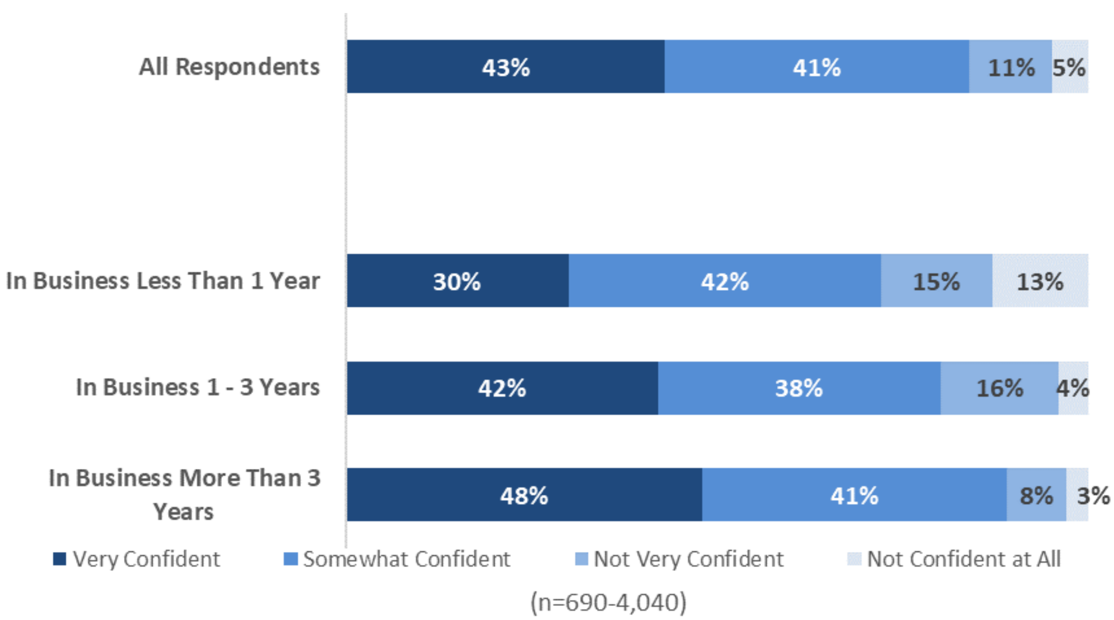
FIGURE 32. New SB/SE: Knowledge of Record Keeping, Tax Year 2019



SOURCE: ICF (February 2021).

Figure 33 shows the responses to the question "How confident are you that you could provide all necessary records, receipts, and documents if you were applying for a loan or responding to an audit?"

FIGURE 33. New SB/SE: Confidence in Record Keeping, Tax Year 2019



SOURCE: ICF (February 2021).

Among all respondents, the majority were knowledgeable about record keeping requirements (71 percent) and confident that they would provide all necessary documentation if needed (84 percent). Knowledge and confidence in record keeping was lower among those with new businesses. Of those in business for less than 1 year, 48 percent were knowledgeable and 73 percent were confident, compared to 79 percent knowledgeable and 89 percent confident among those in business for more than 3 years.

Among the last steps of the taxpayer journey are the preparation and filing of tax forms. The following sentences summarize CEEN respondents' views on these steps:

- Twenty-three percent felt that it was either somewhat difficult or very difficult to **determine which employment tax form to file**.
- Sixteen percent felt that it was either somewhat difficult or very difficult to **prepare their employment tax returns**.
- Sixteen percent felt that it was either somewhat difficult or very difficult to **determine whether they needed to pay quarterly taxes**.

Several CEEN focus group participants voiced their thoughts on tax preparation and filing.

> *"I wish I knew how important it was to keep receipts, keep your paperwork, and keep everything filed."*

> *"Most people aren't purposely making mistakes -most of these people don't have an accounting background and are hard-working people. They are not making mistakes on purpose in most cases. There should be more leniency with forgiveness and education."*

New business taxpayers must also consider the turnaround time for accessing e-pay and e-file systems for the first time. Form 94X journey maps cite the processing time for EFTPS PIN applications (7 business days) and E-file Online Signature PIN applications (45 days) as a pain point for new business owners attempting to use electronic services to file and pay timely.

Research Integration

The SB/SE population contains a wide range of taxpayer types with a diverse set of needs, preferences, and perspectives. Capturing the multitude of scenarios, decisions, actions, and emotions attached to a taxpayer experience illustrates some of the challenges involved in obtaining a thorough understanding of the future SB/SE experience.

Journey mapping is a useful tool for meeting these challenges. Because no single data source or research product can cover the multi-faceted and evolving taxpayer experience, integrating results from different pieces of research is key to understanding the SB/SE experience from the taxpayer's perspective. Journey maps are the compilation of multiple, often disparate, sources of information which portray a customer's key interactions, as well as associated emotions, pain points, and decisions. Journey maps can provide a big-picture overview of the experience and pinpoint specific aspects of the journey. With a storytelling element added to data points, journey maps create a holistic and visual view of the customer experience.

Benefits of Research Integration

One of the most important benefits of combining multiple research studies is gaining a richer understanding of the complete taxpayer experience, including deeper insights that are not possible from using only one study. A related benefit is the ability to overcome the limitations of one study with the strengths of another. In this report, the CEEN survey provided quantitative results which were broad and robust. The taxpayer telephone calls and focus group discussions provided qualitative data, which included rich and detailed information about taxpayer preferences and pain points. Each of these separate pieces enabled us to come to a solid overall understanding of the experience.

Journey map insights can easily be leveraged for user stories, requirement descriptions, design objectives, or measurement opportunities. This helps to communicate and distribute the knowledge gained from departments, teams, and stakeholders. In addition to prioritization, the output of a journey map can serve as a backbone for strategic recommendations and initiatives.

Another benefit of integrating multiple studies is the discovery of emerging themes. In compiling information on the SB/SE taxpayer, we identified a subpopulation that has previously gone unexplored: new small businesses. Through review of multiple studies, we uncovered insights about this group, including many stark differences between this group and more experienced businesses.

Challenges of Research Integration

The biggest challenge we faced from combining different sources of data was discovering large differences among similar taxpayers. As seen in the Balance Due Taxpayers section of this report, characteristics varied widely within this single SB/SE subgroup, creating difficulties in interpreting and generalizing results about this group. In general, variation in results may stem from the differences among studies in any of the following: data type, respondent type, timeframe covered, or as in our case with balance due, survey mode.

Another challenge that comes with research integration is the variation in definitions among research products. One survey may define a group or term differently from another survey (e.g., the definition of a gig worker can vary among surveys), or there may be differences between the IRS and taxpayer definitions of an issue. Combining different pieces of research without standard and clear definitions may make it difficult to draw meaningful conclusions about the group of interest.

Conclusions and Future Research

The research findings in this report build a comprehensive depiction of the SB/SE taxpayer experience from a customer point of view. The insights gained contribute to a clearer understanding of this population and lay the groundwork for improvements in the overall experience for SB/SE taxpayers.

An important contribution of this research is the identification and exploration of several subgroups within the SB/SE population. Each of the five SB/SE groups we studied has its own characteristics, needs, preferences, and expectations. Knowing these group-specific attributes allows us to address each subpopulation with focused improvement strategies. Such targeted efforts can have a much more substantial impact on satisfaction with the taxpayer experience than inexact, general ones.

Understanding the SB/SE experience is an ongoing effort. Several areas of future research are listed below.

The Taxpayer First Act (TFA)

- The TFA was passed by Congress in 2019 with an overarching objective of improving IRS operations. A major component of the TFA is the requirement that the IRS develop a comprehensive customer service strategy. Understanding the impact of the TFA on the taxpayer experience has important consequences for creation of future initiatives.
- Together, the TFA and OMB Circular No. A-11[23] move us toward a holistic view of the taxpayer experience. For future research, this shift in perspective requires continued integration of research projects and thoughtful development of research programs.

Additional SB/SE subgroups

In this paper, we covered several important SB/SE subpopulations. Identifying and learning about additional groups within SB/SE can inform and enhance targeted outreach and other improvement efforts.

New SB/SE taxpayers

Our review of new SB/SE taxpayers uncovered characteristics and journeys that are markedly different from those of other SB/SE groups. Due to the nature of this paper, we were unable to explore this group in detail. Future research into this group should include journey mapping, which will allow a comprehensive, in-depth look at these taxpayers.

Differences between small business and self-employed taxpayers

SB/SE taxpayers include two broad types that are not mutually exclusive: small businesses and self-employed taxpayers. One limitation in this paper is the lack of distinction between these two key groups, whose differences may be substantial. Future research should address potential disparities in the customer experience by assessing the needs of taxpayers who interact with both Individual Master File (IMF) and Business Master File (BMF) systems and services.

[23] OMB Circular No. A-11 Section 280 is centered around managing the customer experience and improving service delivery. Specific elements include establishing a customer experience-mindful culture across Federal Government services and improving customer satisfaction with and trust in the Federal Government.

References

Cantor, Brian, Leong, N., & Copcutt, S., 2018 CCW Market Study: Performance & Metrics, *Customer Contact Week Digital*, March 2018.

ICF International Inc., *Customer Experience, Expectations, and Needs Survey: 2020 National Report*, February 2021.

ICF International Inc., *Taxpayer Experience Survey (TES): 2020 National Report*, March 2021.

IRS SB/SE Research, Team 1, Gig Economy Worker Focus Groups, Project ID: TM1-0347, April 2021.

IRS SB/SE Research, Team 4, 2020 SB/SE 94X Employment Tax Journey Map, Project ID: TM4-0536, December 2020.

IRS SB/SE Research, Team 4, Employment Tax: Environmental Scan, Project ID: TM4-0521, April 2021.

IRS SB/SE Research, Team 4, IRS Balance Due Taxpayer Experience: Current State Analysis, Project ID: TM4-0525, July 2020.

IRS SB/SE Research, Team 4, IRS Balance Due Taxpayer Experience: Taxpayer Vocabulary, Project ID: TM4-0542, August 2020.

IRS SB/SE Research, Team 4, Journey Maps—What Are They?, SB/SE Research Showcase, February 2020.

IRS SB/SE Research, Team 4, SB/SE ACS IVR Customer Satisfaction Report: Survey Year 2019, Project ID: TM4-0555, September 2020.

IRS SB/SE Research, Team 4, SB/SE ACSS Mail Customer Satisfaction Report: Survey Year 2019, Project ID: TM4-0559, October 2020.

IRS SB/SE Research, Team 4, SB/SE CSCO Mail Customer Satisfaction Report: Survey Year 2019, Project ID: TM4-0561, November 2020.

IRS SB/SE Research, Team 4, SB/SE Employment Tax Mail Customer Satisfaction Report: Survey Year 2019, Project ID: TM4-0568, February 2021.

IRS SB/SE Research, Team 4, SB/SE Field Collection Mail Customer Satisfaction Report: Survey Year 2019, Project ID: TM4-0566, January 2021.

Starcher, Jeremy, Callback—The Science Behind the Art of Customer Relationships, Virtual Hold Technology White Paper, 2018.

Yildirmaz, Ahu, Goldar, M., & Klein, S., Illuminating the Shadow Workforce: Insights Into the Gig Workforce in Businesses, ADP Research Institute, February 2020.

Are Annual Federal Employment Tax Returns Effective? Using Tax Administrative Data To Examine Filing, Reporting, and Payment Compliance Associated with Forms 943 and 944

Yan Sun and Stephanie D. Needham (IRS, Research, Applied Analytics, and Statistics)[1]

1. Introduction

Federal law requires employers to withhold federal income tax, Social Security and Medicare taxes from the employees' pay, and to pay and report this liability to the Internal Revenue Service (IRS) by filing the Form 941 (Employer's Quarterly Federal Tax Return).

To ease the burden of filing quarterly employment tax returns, the IRS provides certain groups an alternative filing option, i.e., annual instead of quarterly. Specifically, employers who employ agricultural workers are permitted to file the annual Form 943 (Employer's Annual Federal Tax Return for Agricultural Employees); and small employers with an annual tax liability less than $1,000 are permitted to file the annual Form 944 (Employer's Annual Federal Tax Return).

The goal of the annual Federal Employment Tax returns is to reduce taxpayer burden for those with low employment taxes, and to improve tax compliance. However, more and more issues regarding Form 944 returns emerge each year, including incorrect forms filed, untimely filing, and repeated improper tax filing activities from the taxpayers, etc.[2] IRS subject matter experts (SMEs) have noted that despite numerous attempts to rectify internal processing issues, the problems have persisted.[3] In addition, many tax experts at IRS have started to express concern about the annual employment tax forms and agree that the burden associated with the implementation and processing of the annual filing forms has outweighed its benefits.

However, there has been limited research conducted on the annual 943 and 944 employment tax forms. In 2007, the Treasury Inspector General for Tax Administration (TIGTA) published a report (TIGTA, 2007) and found that fewer Form 944 returns than anticipated were filed. The most recent study available is the internal report (IRS, 2013), released in 2013 by Small Business/Self-Employed (SB/SE), which suggested to increase the threshold for annual tax liability of Form 944.

In alignment with the IRS mission to provide top quality service to taxpayers, and in response to the problems identified by IRS business operating divisions (BODs), the IRS established the Innovation Lab 1.0 in 2019,

[1] The paper was prepared as an extension of the IRS Innovation Lab 1.0 project: "Assessment of the Effectiveness of Forms 943 and 944". The Executive Champions for the project were Imelda Deniz-Vazquez (Program Manager, Large Business & International, IRS) and Holly Donnelly (Director, Strategy Business Solutions, RAAS, IRS). The authors acknowlege significant contributions from the Innovation Lab Project Team subject matter experts (SMEs) Barbara Sandstrom, Mel Hadley, Terry Manzi, Lindsay Schrock, and Tiffany Zerangue. The authors are grateful to Cindy Ferwerda, Amy Salmon, Crystal Stinson, Joseph Tiberio, Ann Willis, and Barbara Wulf, for their support and valuable feedback on the paper. Special thanks to Rania Abumeri, Christine Glass, Christine Oehlert, Alan Plumley, and Anne Parker, for their valuable help and advice. The authors also would like to acknowledge the comments from IRS SB/SE Specialty Exam Policy Program, Manager and Analysts, Scott Irick (Director, Examination, SB/SE, IRS), and Maha Williams (Deputy Director, Examination, SB/SE, IRS). The authors would like to thank the participants in the 11th Annual Internal Revenue Service (IRS)-Tax Policy Center (TPC) Joint Research Conference on Tax Administration in 2021, the discussant Dr. Jacob Goldin, and the anonymous reviewers for helpful comments.

[2] This observation is based on authors' personal conversation with Mel Hadley, Field Collection Territory Manager with SB/SE. It has been confirmed by Specialty Exam Policy Team in Employment Tax Policy at SB/SE of IRS.

[3] Sandstrom (2020) provides a detailed explanation on the complicated implementation of Form 944 compliance program.

with a focus on Employment Tax. One of the projects selected for this pilot effort, was the Assessment of the Effectiveness of Forms 943 and 944.

As part of this broad project, this paper analyzes the tax return data of annual employment tax forms 943 and 944. We explore the data both historically and cross-sectionally, to identify the factors causing the operational ineffectiveness. The "effectiveness" in the context of this project, mainly focuses on two spheres: first, the ability to identify and address noncompliant taxpayers; and second, the ability to administer the programs without increasing the operational cost (TIGTA, 2008). We systematically examine the tax filing, tax liability, operational costs, and tax compliance for annual employment tax returns. We further develop an empirical model to estimate the compliance risk and enforcement in the annual filing.

We aim to provide data-supported evidence for potential regulatory change. We do this in three steps. First, the paper adopts a comparative analysis approach to evaluate the effectiveness of annual employment tax filing of Forms 943 and 944. Second, the study provides sufficient statistics to characterize the tax filing and compliance in depth and in breadth and updates the general understanding of annual employment tax returns. It fills the knowledge gap in examining tax reporting and payment behavior by utilizing the administrative data, while no survey data exists. Third, we develop an empirical model and provide empirical evidence to quantify the compliance risk.

The rest of the paper is organized as follows. Section 2 describes the background of the annual employers' federal tax returns and a brief literature review. The filing condition and tax liability of Forms 943 and 944 as well as tax compliance are examined in Sections 3, 4, and 5. Section 6 develops an empirical model to analyze the payment compliance, and Section 7 summarizes the findings.

2. Background and Literature Review

2.1 Background

Employers use Form 941[4] to report wages and tips paid, federal income tax withheld, both the employer's and the employee's share of Social Security and Medicare taxes (and additional Medicare tax withheld from employees, if any), adjustments, and qualified small business payroll tax credits.[5] After an employer files the first Form 941, the employer must file the return for each quarter, even if there are no taxes to report, unless the employer files a final return, or is an employer of seasonal, household, or agricultural employees. Under the withholding system in the U.S., employers must withhold a certain amount of federal income tax, Social Security tax, and Medicare tax. Form 941 is filed every 3 months and is generally due on the last day of the following months: April, July, October, and January of the following tax year.[6]

The Form 943 was implemented in 1956, and it can only be filed by employers that pay wages to one or more farmworkers and if those wages are subject to Social Security, Medicare, and federal income tax withholdings.[7] Hired farmworkers include field crop workers, nursery workers, livestock workers, graders and sorters, agricultural inspectors, etc. According to the Economic Research Services in the U.S. Department of Agriculture, the majority of farmworkers are wage and salary employees, who are hired directly by farmers, or agricultural service companies. Form 943 is filed annually, with a due date of January 31 of the following calendar year. For example, for Tax Year 2017, the Form 943 should be filed by January 31, 2018 (footnote 5 applies as well).

Form 944 was published by the IRS 50 years later, in 2006, and was designed for the smallest employers, (those whose annual tax liability for Social Security, Medicare, and withheld federal income taxes is $1,000 or less). Unlike the Form 941 that is filed quarterly, the Form 944 is filed once a year. If an employer is required to

[4] The snapshots of Forms 941, 943, and 944 in Tax Year 2017 are shown in Appendix I.

[5] Based on Specialty Exam Policy team, 1956 is the oldest version that the Form 941 repository has on file.

[6] If any due date for filing falls on a Saturday, Sunday, or legal holiday, you may file your return on the next business day.

[7] All cash wages that an employer pays to an employee during the year for farm work are subject to Social Security and Medicare taxes and federal income tax withholding if either of the two tests are met: the $150 test or the $2,500 test. The $150 test applies if an employer pays an employee cash wages of $150 or more; and the $2,500 test applies if the total cash and noncash wages an employer pay to all farmworkers is $2,500 or more.

file Form 941 but estimates their upcoming annual taxes to be $1,000 or less, then the employer may request to file Form 944 instead of Form 941 by either calling IRS or sending a written request to the IRS. The employer must receive written notice from IRS to file Form 944 before the employer is permitted to file this form. In other words, if the employer does not receive this notice, the employer must continue to file Form 941 for that calendar year.[8] The employers who are eligible to file Form 944 are required to file and pay employment taxes on or before January 31 following the end of the tax year. For example, the first Forms 944 were due January 31, 2007. Form 944 At A Glance is listed in Appendix II, briefly describing the implementation of Form 944 in a historical order.

2.2 Literature Review

We only found two formal studies on the annual employment tax returns, and both were focused solely on Form 944. The first study was conducted by the Treasury Inspector General for Tax Administration (TIGTA) one year after Form 944 was enacted. This report reviewed the development and implementation of the employer's Annual Federal Tax Return (Form 944 Program). The 2007 study showed that the IRS had estimated that approximately 1 million taxpayers would be eligible to file Form 944. However, only about 326,000 employers had filed the Form 944 by June 2007 (TIGTA, 2007).

The second study, conducted by the IRS SB/SE, was conducted in 2013, and examined Forms 941 and 944 returns that were filed from 2008 to 2011. This report suggested that the tax liability threshold of Form 944 should be increased from $1,000 to $2,500 or more, and that a semiannual filing option should be permitted for a selected group of Form 941 filers.

3. The Filing Population

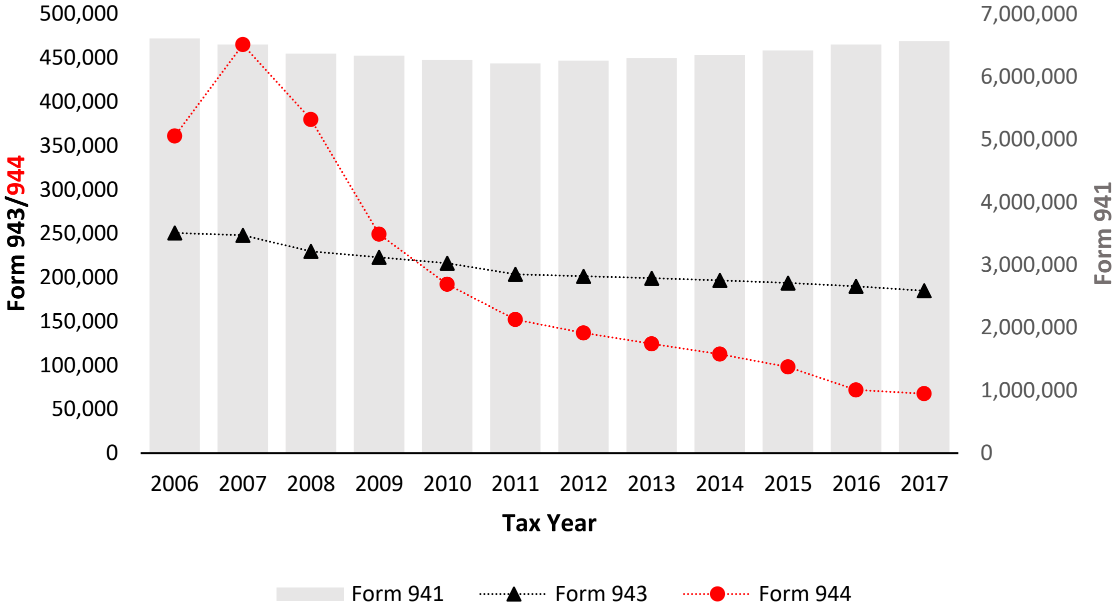
The filing volumes as well as their industry sectors are analyzed to provide an understanding of the filing population of the employment tax returns. The quarterly filing of Form 941 is also included in the analysis to complete the full employment tax population.

3.1 Filing Volumes

In order to understand the filing behavior of the different forms, we present a historical trend for the filing volume of Forms 941, 943, and 944, from 2006, when Form 944 was introduced, up to 2017. It is striking that we observe two distinct patterns: a rapid declining trend, represented by Form 944; and relatively stable trends for Form 943 and Form 941 (Figure 1).

[8] This is also true if the employers are Form 944 filers and their liability is now more than $1,000. They still have to continue to file Form 944. Also, the IRS can only change the filing requirement during the first quarter of the year. Based on the Specialty Exam Policy, this is one of the biggest unpostable issues the team faces: the taxpayers are trying to self-correct but they can't.

FIGURE 1. The Total Number of Employers Filing Employment Tax Forms 941, 943, or 944, Tax Years 2006–2017



The number of employers who have filed the annual Form 944 has been declining consistently over the past 10 years. When Form 944 was first implemented in 2006, approximately 360,000 employers filed this newly introduced annual form (represented by the red dot in Figure 1) for that tax year.[9] In the following tax year (2007), the number of annual Form 944 filings jumped up almost thirty percent, to around 465,000.

However, since Tax Year 2008, the volume of Form 944 filings has declined rapidly. In Tax Year 2017, only 67,000 employers filed the employment tax return using Form 944; approximately one-sixth of the initial volume of Form 944's filed in Tax Year 2006. The average annual percent change rate for Form 944 filing volumes from its inception date to 2017 is negative 13 percent, verifying the observation of a continuously falling filing volume.

Compared with Form 944 filing, the trend for the annual filing of Form 943 (represented by the black dots in Figure 1) is much flatter, though with a slightly falling trend. The average annual percent change for the filing population of Form 943 is negative 3 percent over the period from 2006 to 2017.

Two different trends of Forms 943 and 944 are observed in Figure 1, though both are annual filing forms. In recent years, the divergence has grown wider. When Form 944 was implemented in 2006, the number of employers who filed Form 944 was greater than the number of the employers who filed the annual Form 943. This status is demonstrated by the left segment in Figure 1, where Form 944 filing volumes (represented by the red dots) is above Form 943 filing volumes (represented by the black dots), before the two lines cross. However, since Tax Year 2010, the position of the two lines has reversed, and the number of employers who filed Form 943 has exceeded the number of employers who filed Form 944. This is shown by the second half of the segment in the right side (where the black line is above the red line after the intersection). Compared to the rapid diminishing of Form 944 filing population, the volume of employers who filed Form 943 has been relatively stable.

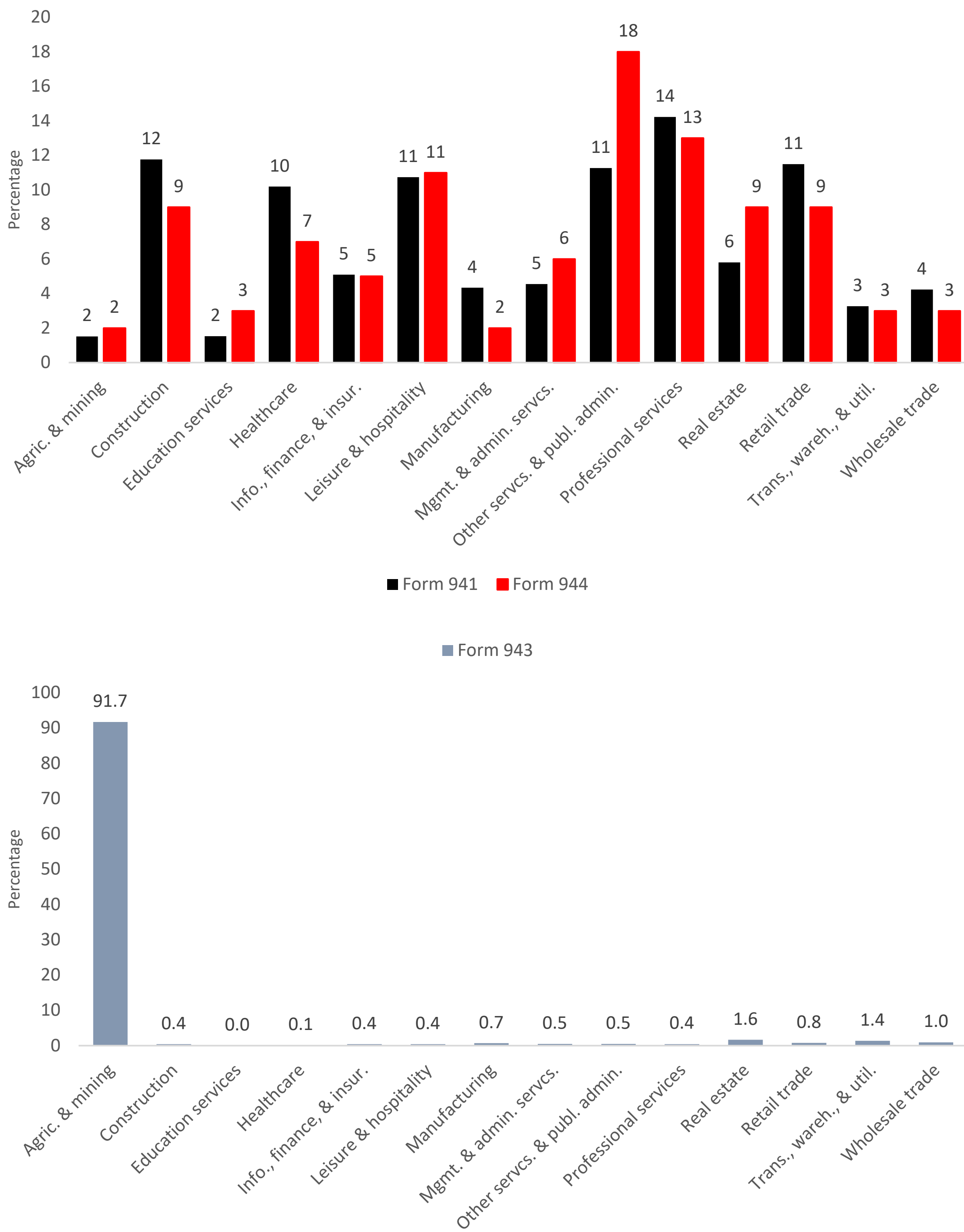
The number of employers who filed the quarterly employment Form 941 from 2006 to 2017 has remained at a relatively stable level of 6.6 million (represented by a gray bar in the background of Figure 1), though with a slight dip in the middle (around 2011). This observation implies that the majority of employers filed the quarterly employment tax form in the last decade, even though there were other options available (e.g., Form 944).

[9] The data were accessed in January of 2021.

3.2 *Filing Industries*

In this section, we examine the industry sector of the businesses to obtain a better understanding of the population who file Forms 941, 943, and 944. Figure 2 presents the distribution of filers by industry sectors, for Forms 941, 943, and 944 filers. The industries of Form 943 filers are illustrated separately since this form has been designed for agricultural employers. Fourteen industrial sectors are evaluated for the filing of each form.

FIGURE 2. The Industrial Sectors of Form 941, 943, and 944 Filers, Tax Year 2017



Forms 941 and 944 filing share a similar industry distribution. First, they are distributed among all the 14 sectors. The proportion of each industry varies, from 2 percent to 18 percent. Second, though the volume of employers filing Form 944 is much smaller than the volume of employers filing Form 941, we find they have a comparable industrial allocation: approximately 75 percent of Form 941 and 76 percent of Form 944 filers, respectively, are composed of the following seven industries: Construction, Healthcare, Leisure and hospitality, Professional services, Real estate, Retail trade, and Other services. We do not find any specific industry dominating Form 941 or Form 944.

On the contrary, the Form 943 has a completely different industry distribution, as one would expect. Agriculture is a prevailing sector among Form 943 filers, accounting for almost 92 percent of the total Form 943 filing population. The remaining small portion of the Form 943 employers are distributed throughout five sectors: Manufacturing, Real estate, Retail trade, Wholesale trade, and Transportation and utilities, accounting for 6 percent of the total. Each share is less than 2 percent. This distribution is consistent with the original implementation goal of the Form 943, to serve the employers who hire agricultural workers.

In terms of industry sectors, Form 943 filers have a completely different structure from Form 944 filers, even though both of them are annual filing forms. Conversely, despite the different filing frequencies (annual versus quarterly), Forms 941 and 944 filers share a similar filing population portfolio. While the Form 944 eligibility is limited by the amount of tax liability, Form 943 qualification is restricted by the industry sector. Among the employers who file Form 941 or Form 944, we do not find that any particular industry sector prefers one form over another.

4. Tax Liability

The tax liabilities[10] of different employment tax forms are examined in depth in this section. We divide the filers of each tax form into three tax liability groups based on the employer's tax liability. We then focus on the annual employment tax Forms 943 and 944 to investigate the characteristics of the filers and their tax filing patterns.

4.1 Tax Liability Groups

To better understand the filing pattern at a more granular level, we stratify the filers by their tax liabilities. Table 1 presents the classification of the employers by tax form they filed and by tax liability, in Tax Year 2017. The following two thresholds are selected for examination: $1,000 and $2,500. The $1,000 threshold is chosen since $1,000 is Form 944 tax eligibility limit, while the $2,500 is the federal tax deposit threshold[11] for Forms 941 and 944.

The filers of the three forms are composed of employers of all groups: small employers (annual tax liability of $1,000 or less), medium-sized employers (annual tax liability between $1,000 and $2,500), and large employers (annual tax liability over $2,500). Table 1 presents the breakdown of the employment tax filers by size and liability group.

In Tax Year 2017, a total of 6.8 million employers filed either the quarterly filing Form 941 or the annual Forms 943 or 944. Among them, Table 1 (row 1) shows that the majority (around 6.6 million, accounting for 96 percent) of employers filed Form 941. Conversely, only 4 percent of employers filed an annual employment tax return; 3 percent (around 185,000) filed Form 943; and 1 percent (around 67,000) filed Form 944.

[10] Tax liability in this study is defined as total annual tax settlement amount.

[11] If the liability for the annual taxes is $2,500 or more, the employers are generally required to deposit the taxes instead of paying them with the Form 944 filing.

TABLE 1. Employment Tax Filing In Tax Year 2017, by Tax Form and by Tax Liability

Annual tax liability	Form 941		Form 943		Form 944	
	No. of employers	Share	No. of employers	Share	No. of employers	Share
All	6,562,139	100%	184,737	100%	66,897	100%
<=$1,000	605,407	9%	39,651	21%	43,786	65%
$1,001 - $2,500	406,493	6%	21,878	12%	6,513	10%
>$2,500	5,550,239	85%	123,208	67%	16,598	25%

Form 944

Though Form 944 was originally designed for the smallest employers, it is observed that some large-size employers filed their tax return by using this form (Table 1, column Form 944). In Tax Year 2017, some 16,598 employers with tax liability over $2,500, accounting for 25 percent of the filing population of Form 944, filed Form 944.

Conversely, it is observed that the majority of small employers filed their employment taxes quarterly instead of annually, by filing Form 941 (Table 1, column Form 941). In Tax Year 2017, some 605,407 employers whose annual tax liability was $1,000 or less filed Form 941. This is more than 10 times the number of employers with an annual tax liability less than $1,000 who filed the annual Form 944.

Form 943

A significant portion of Form 943 filers (around 67 percent) had an employment tax liability over $2,500 in Tax Year 2017 (Table 1, column Form 943). Approximately 21 percent of the Form 943 filers are small-size farmers with annual tax liability less than $1,000.[12]

The division and examination of tax liability groups of the three forms reveals that 1) Form 941 is the most commonly filed employment tax form by employers of all sizes; and 2) the majority of small businesses that are eligible to file the annual Form 944, are electing to file quarterly Form 941, instead of the annual Form 944. This highlights the observation that the implementation goal of the annual Form 944 to reduce small business taxpayer burden is not being met.

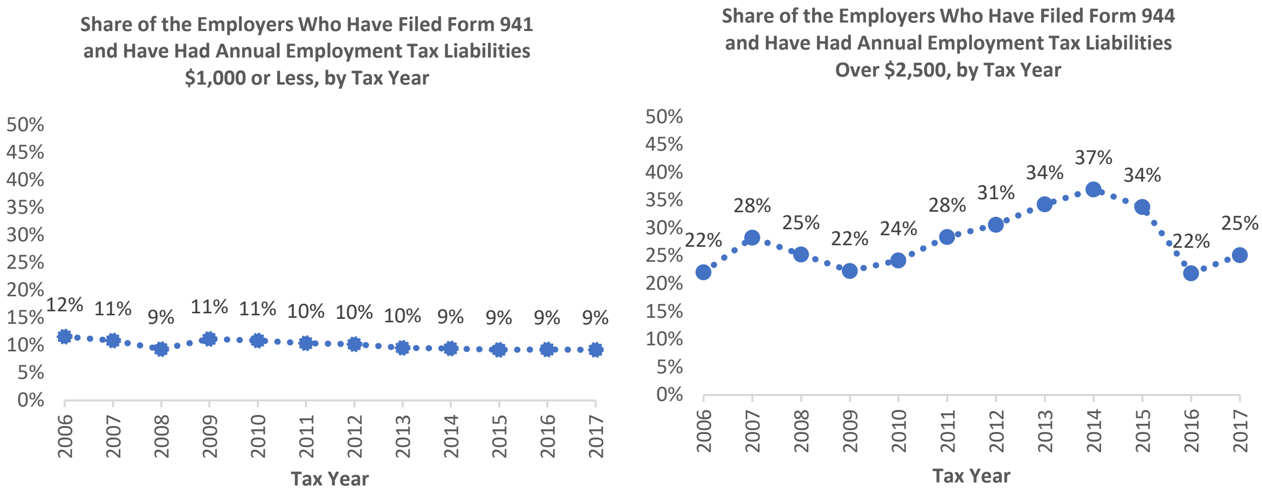
4.2 Tax Filing

As stated above, the original intention of Form 944 was to reduce the burden of small employers, by enabling them to file once a year instead of four times a year. However, we have observed that a consistent portion of 10 percent of Form 941 filers, who were small employers with an employment tax liability of $1,000 or less, opted to file Form 941. Figure 3 (left panel) shows that, from 2006 to 2017, approximately 600,000 employers on average, who filed Form 941, had an employment tax liability of $1,000 or less.

Although the tax burden for some small businesses was not reduced, because approximately 600,000 small employers opted not to file the annual Form 944, some large employers with an annual employment tax liability over $2,500 have filed the annual Form 944. During the period from Tax Years 2006 to 2017, on average, 28 percent of employers who have filed Form 944 have had an annual employment tax liability over $2,500. The historical rate has fluctuated slightly and has ranged between 22 percent to 37 percent. The filing of Form 944 by larger businesses may be due to the business' failure to follow IRS rules and IRS internal processing, or the fast expansion of the business.

[12] For the detailed requirements for Form 943, please see footnote 7.

FIGURE 3. Employment Tax Filing Pattern for Forms 941 and 944, Tax Years 2006–2017



The tax filing pattern demonstrated in Figure 3 has been established over the last decade. Specifically, the majority of small employers (with employment tax liability under $1,000), continued to file Form 941, even though the annual Form 944 was an option. And a nontrivial number of large businesses continued to file Form 944, even though they had an annual tax liability over $2,500. We also present a more robust analysis in Appendix IV and calculate the Form 944 take-up rate. The observation from the take-up rate is consistent with the findings in Figure 3: participation in Form 944 program (e.g., filing Form 944) has been low over the last decade.

4.3 Tax Liability of Form 944

In order to detect the source of incorrect filing identified earlier, in this section we focus on two types of taxes: federal income tax withheld, and total Social Security tax and Medicare tax. Figure 4 presents the distribution of the Federal income tax withheld (left panel) and total Social Security tax and Medicare tax (right panel) by quantiles, for Form 944 filing in Tax Year 2017.

The federal income tax withheld in Form 944 is highly dispersed. A significant portion (around 60 percent) of Form 944 filers had zero federal income tax withheld in Tax Year 2017. However, the 99th percentile of the federal income tax withheld surges to over $50,000, making the latter part of the distribution curve very steep.

In addition, the total Social Security tax and Medicare tax displays a similar skewed distribution: 1) a wide gap in the taxes is observed among the filers; 2) half of filers are paying very little Social Security tax and Medicare tax, and a few (around 5 percent) of the employers had large Social Security and Medicare tax liabilities (over $10,000). In Tax Year 2017, around 20 percent of Form 944 employers had zero Social Security and Medicare tax, but the 99th percentile for the total Social Security and Medicare tax is over $60,000.

Figure 4 displays the polarization of the tax liabilities among Form 944 filers, for both the federal income tax withheld and the total Social Security and Medicare tax, though the distribution of the two types of taxes mirror each other. The heterogeneity of the tax liabilities of the Form 944 filing population contradicts the original goal of Form 944, to be filed by the smallest employers.

FIGURE 4. The Distribution of the Federal Income Tax Withheld and Social Security Tax and Medicare Tax in Form 944, by Percentile, Tax Year 2017

While Form 944 does not ask for the number of employees employed, as Form 943 does, we use total wages and compensation as a proxy for the employer size. We then examine the taxes by the ranking of this size proxy. Figure 5 presents the average amount of tax liability for Form 944 in Tax Year 2017, by deciles of the total wages and compensation.

Due to the large variation in tax amounts, we split Figure 5 into two parts: Panel A and Panel B, with different scales on the vertical axis. Panel A displays the tax for the bottom six deciles, and Panel B shows the top four deciles. The two panels combined provide us with a complete picture of the distribution of the tax among the filers. It is observed: 1) as the employer hires more and more workers, the total tax liability goes up; 2) approximately 60 to 70 percent of Form 944 employers had an annual tax liability $1,000 or less, which is consistent with the statistics in Table 1; 3) the average tax amount for the top decile of Form 944 filers is an outlier, and is much larger than the summation of the average tax amount of the bottom 9 deciles (Panel B), about 10 times the average tax amount of the filers in the 9th decile.

By segmenting the data, Figure 5 shows that the top decile behaves very differently from the rest of Form 944 filers, with a disproportionately high tax liability, deviating from the main body of Form 944 filers. Thus Form 944 filing population could be sorted into three groups: the bottom 6 deciles (with tax liability $1,000 or less), the 7th to 9th deciles (with tax liability between $1,000 and $10,000), and the top decile with a spike in the tax liability.

FIGURE 5. Average Tax Liability in Form 944, by Deciles of Wages and Compensation, Tax Year 2017

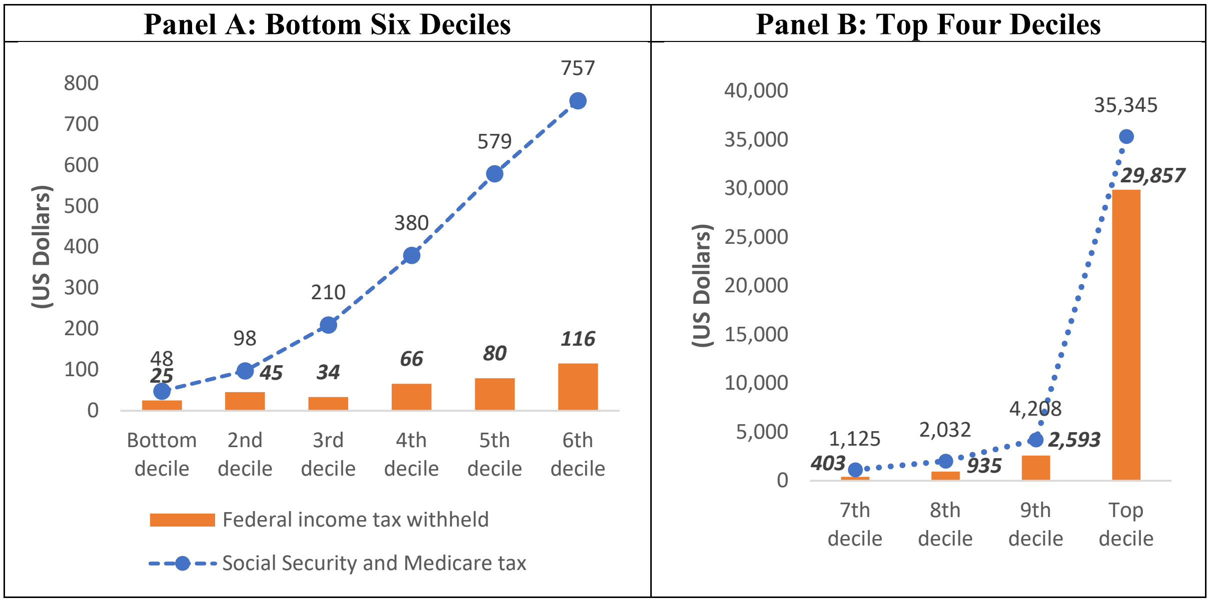
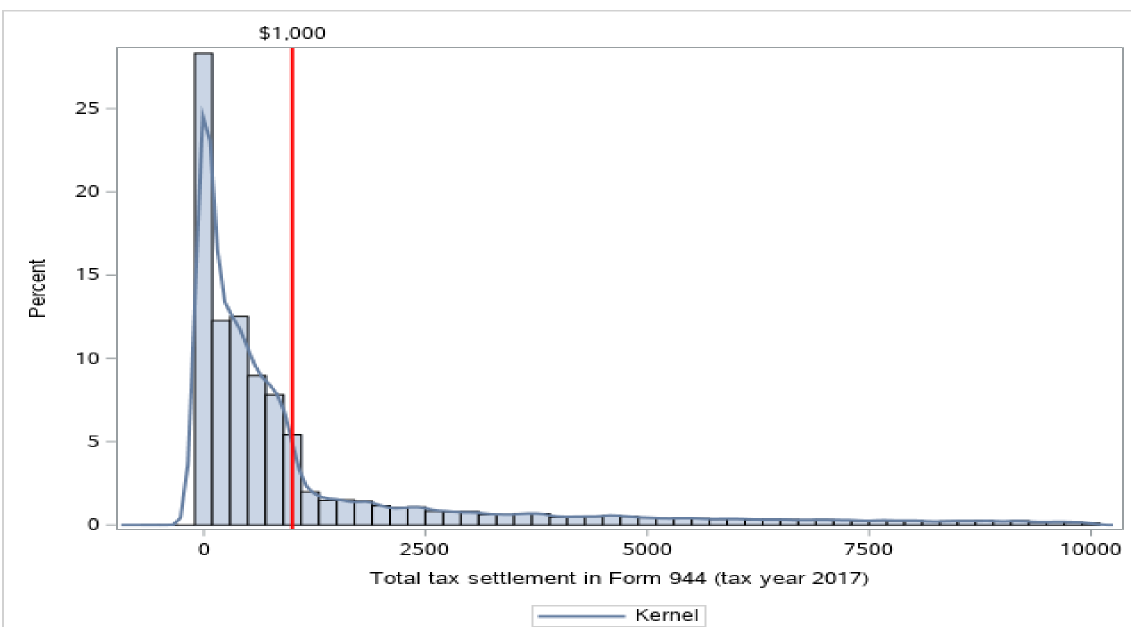


Figure 6 shows a truncated distribution of tax liabilities for Form 944 filers in Tax Year 2017, with a tax liability of $10,000 or less. The maximum tax liability is 10 times Form 944 eligibility threshold (of $1,000), which is exhibited by a red line in Figure 6. The truncated distribution covers about 88 percent of Form 944 filers, implying that approximately 12 percent of Form 944 filers have an annual tax liability over $10,000.

FIGURE 6. Truncated (≤ $10,000) Distribution of the Total Tax Settlement of Form 944 in Tax Year 2017



The graph also shows that while there is a significant percentage (around 65 percent) of Form 944 filers positioned before the $1,000 line, there is also a nontrivial percentage of large employers filing Form 944, as indicated by the long thin tail of the distribution. We also note that the distribution is positively skewed. This reflects the fact that the large employers lying in the tail of the distribution have driven the average Form 944 annual tax liability up.

Form 944 was designed for the smallest employers, with a tax liability threshold set at $1,000 in 2006. This threshold has remained at that level and has not been changed since then. However, the average wage level has risen over the last 10 years. Based on the statistics from the Federal Reserve,[13] the average earnings per hour for the private sector[14] has gone up from $20 in 2006, up to $26 in 2017, a 30 percent increase. Thus, we consider the $1,000 limit to no longer be appropriate.

4.4 Tax Liability of Form 943

In this section, we examine the tax liability of Form 943, to detect if any issues arose in another annual tax filing population. Besides the farming sector, Form 943 does not have any other restriction on the tax returns. Figure 7 presents the volume of Form 943 filers by their tax liability[15] category (Panel A), as well as total tax liability of Form 943 by tax year (Panel B).

The number of large employers filing Form 943 has been stable in recent years, while the number of small employers filing Form 943 has diminished. Figure 7 Panel A displays a deviating trend: while the number of small employers filing Form 943 with tax liability $1,000 or less (blue line) has been declining, the volume of Form 943 filers with an annual tax liability over $2,500 (grey line) has remained the dominant employers since Tax Year 2006. The total count of the small employers filing Form 943 has dropped from about 90,000 in Tax Year 2006, to only around 40,000 in Tax Year 2017, down 50 percent. On the contrary, the volume of Form 943 filers with an annual tax liability over $2,500 has remained at around 120,000 for more than 10 years, with little fluctuation.

Combining Figure 7 (Panel A) with Figure 1, we find that the slight decrease of Form 943 filing volumes has mainly been driven by the persistent shrinking of the small employers filing Form 943. In Tax Year 2017, 70 percent of Form 943 employers that filed Form 943 had an annual tax liability over $2,500, and only 20 percent of these filers had an annual tax liability of $1,000 or less.

The total Form 943 tax liability has grown, as shown in the Panel B of Figure 7, from approximately $5 billion in Tax Year 2006, to almost $8 billion in Tax Year 2017. This observation is consistent with the increasing trend in the wage and salary employment in agriculture since 2010[16] and gradually rising wages for hired farm workers over the past 30 years.[17]

The steady growth of Form 943 total tax liability demonstrated in Figure 7 does not contradict the observation of a slightly decreasing number of Form 943 filings in Figure 1. The declining volume of Form 943 filers, with a small amount of tax liability, has insignificant impact on the total Form 943 tax liability, which is mainly impacted by the Form 943 filers with tax liability over $2,500. The decomposition of Form 943 total tax liabilities is listed in the Appendix Table A1, which shows that 99 percent of Form 943 tax liability came from the employers with tax liability over $2,500 over the last decade. In contrast, the total tax liability for Form 944 filing is only $0.6 billion in Tax Year 2017, accounting for less than 10 percent of the total tax liability of Form 943 filing (around $7.8 billion). In terms of tax revenue generation, Form 943 returns are more impactful than Forms 944.

---

[13] Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org.

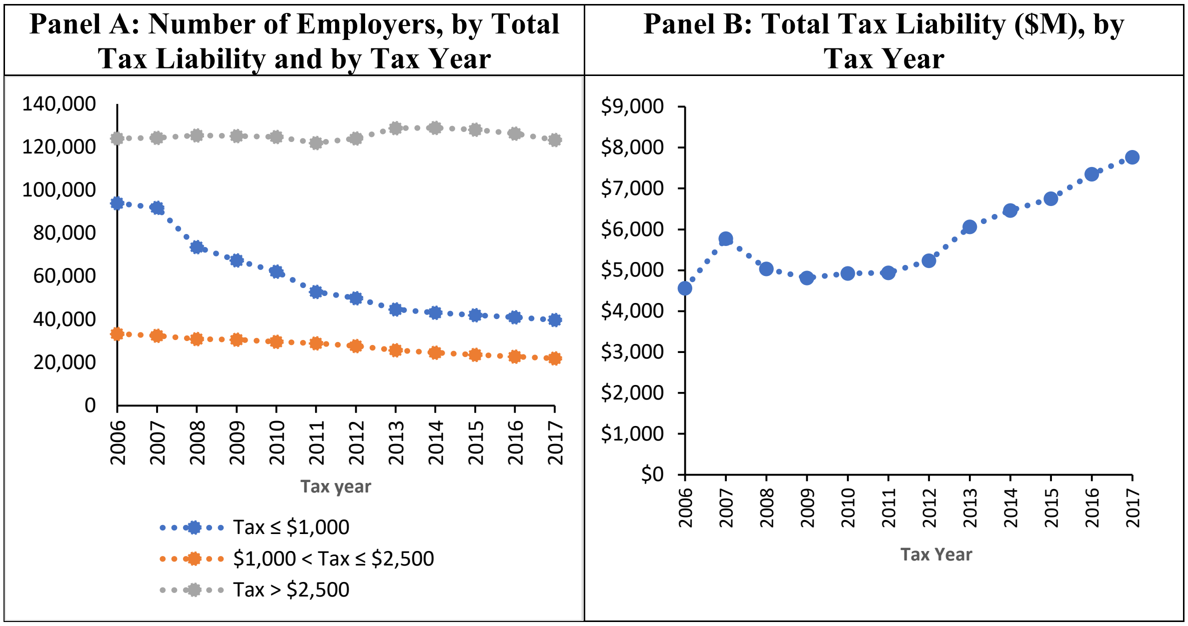
[14] Average hourly earnings of all employees, total private, dollars per hour, monthly, seasonally adjusted.

[15] The total tax liability is defined as total tax settlement in Form 943, and it is measured in nominal dollars.

[16] Based on U.S. Department of Agriculture, the U.S. wage and salary employment in agriculture has been stabilized since 2000 and had increased from 1.07 million in 2010 to 1.18 million in 2019.

[17] The wages for hired farmworkers have risen, both in real terms and in relation to wages for the average nonsupervisory worker in a nonfarm occupation (Economic Research Service, https://www.ers.usda.gov). The wages for nonsupervisory crop and livestock workers rose at an average annual rate of 1.1 percent per year between 1990 and 2019. The real farm wages grew 2.8 percent per year in the past 5 years due to the fact that workers were harder than usual to find.

FIGURE 7. Number of Form 943 Employers and Total Tax Liabilities, Tax Years 2006–2017



While large employers filing Form 944 seems to be one of the major causes of the problems for Form 944, the large employers filing Form 943 seem to be the reason for the stability of Form 943, especially in terms of tax revenue generation and maintaining the filing volume.

5. Tax Compliance

Tax compliance is examined in this section. We utilize the administrative data to study employers tax filing behavior and evaluate whether filers of Forms 943 and 944 comply with the tax rules of IRS, such as tax reporting, filing a return, and paying the tax due in a timely manner. A failure to follow IRS regulations generates significant negative consequences for taxpayers, as well as processing costs for the IRS. One issue associated with Form 944 filing is that the employers file the tax return without strictly following the IRS filing requirements. This may cause the tax return to be "unposted" in IRS systems. This "unpostable" tax return must be corrected manually by a tax agent at a later stage, requiring extra IRS resources. Thus, we use "corrected unpostable"[18] as one measure to assess IRS internal processing costs, generated by filers' noncompliant tax behavior. Though the measure is not in a monetary term, it does provide us the extra operational cost resulting from noncompliance quantitatively.

5.1 Tax Reporting

When a business owner establishes an entity (other than a sole proprietorship) or hires employees, it must obtain an Employer Identification Number (EIN) from IRS. Employers use EINs to file employment tax returns and IRS uses this number to track their payroll tax remittance.

When a business applies for an EIN (SS-4 Form, Application for Employer Identification Number), the application information entered (e.g., name, address, type of entity, principal business activity, etc.) is disclosed to IRS. In addition, if a business expects their employment tax liability to be $1,000 or less in a full calendar year, the business must check Form 944 option box on the SS-4 Form.

[18] When the filing requirement of a specific tax form is not met, the tax return will not be posted in the IRS system. However, a tax examiner may re-examine the case and manually post the tax return. Under such circumstance, this tax return is "corrected unpostable". This type of "unpostable return that was corrected" is examined in this study. Those that are never corrected or assessed are not discussed in this analysis.

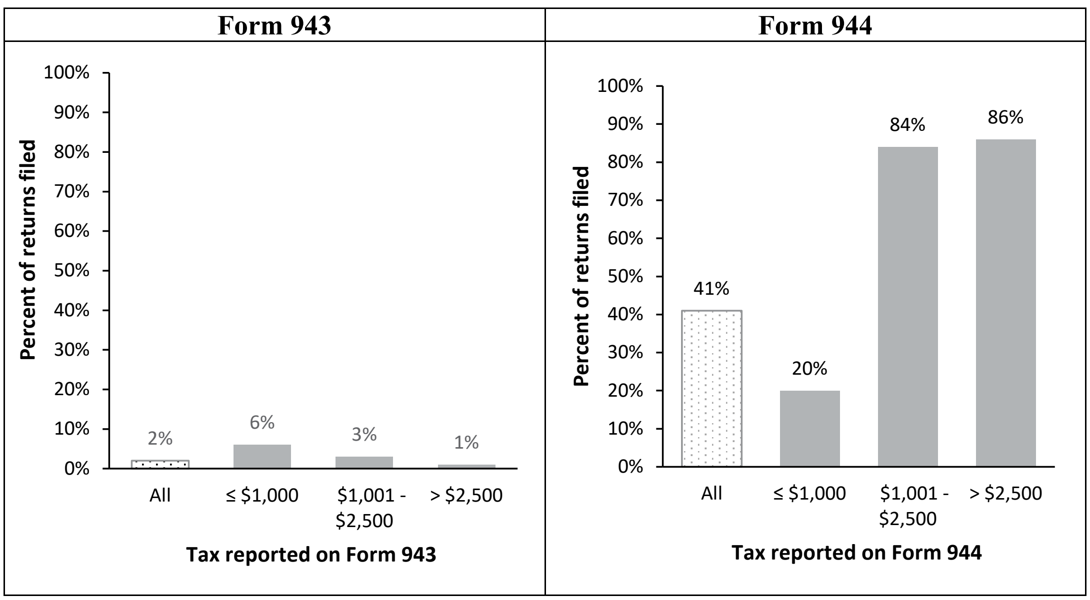
Thus, the "filing requirement" (used in IRS systems to track if taxpayers file the correct tax form) for the business is set, based on their SS-4 Form submitted. If later a business wants to change the tax form used to file the employment tax, the business must submit a change request to IRS.

For Form 943, if an employer has paid wages to one or more farmworkers and the wages are subject to Social Security and Medicare taxes or federal income tax withholding,[19] the "filing requirement" is set as "Form 943."

In general, once the IRS has notified an employer of its filing requirement, the employer must file the identified form even if there are no taxes to report, or the tax liability has changed. For example, if an employer's annual tax liability drops below $1,000, and it now would like to file Form 944 instead of Form 941, the employer must contact IRS to request to file the annual form.[20] However, until the business receives written notification from the IRS that the filing requirement has been changed, the employer must continue to file the quarterly Form 941.

The annual Form 943 filers are more likely to follow the IRS tax filing requirement than Form 944 filers. Figure 8 (left panel, first column) shows that in Tax Year 2017, only around 2 percent of the Form 943 employers filed Form 943 but had no Form 943 filing requirement. Conversely, 41 percent of Form 944 employers filed the Form 944 but had no Form 944 filing requirement (Figure 8, right panel, first column). The large difference in incorrect filing requirements highlights the high degree of confusion or noncompliance among Form 944 filers as compared to Form 943 filers.

FIGURE 8. Filing Forms 943 and 944 Without a Filing Requirement, by Tax Liability, Tax Year 2017



Form 943 filers with annual tax liability over $2,500 have the lowest rate of filing the incorrect form, at around 1 percent. This rate is lower than the rate of Form 943 filers with tax liability $1,000 or lower, at 6 percent, or the rate of Form 943 filers with tax liability between $1,000 and $2,500, at 3 percent. Accounting for two-thirds of the volume of Form 943 filers, the majority of Form 943 filers with tax liabilities over $2,500 file the right tax form. This significantly lowers the average rate of incorrect filing for Form 943 to only 2 percent.

In contrast, Form 944 employers with annual tax liability over $2,500, have the highest rate of filing the incorrect tax form. In Tax Year 2017, approximately 86 percent of the Form 944 filers with annual tax liability

[19] Please refer to footnote 6 on the explanation of $150 test or the $1,500 test.

[20] The request can only be submitted in the first quarter of the year.

over $2,500 filed the Form 944 but had no Form 944 filing requirement. Among Form 944 filers with annual tax liability of $1,000 or less, the rate of filing with incorrect form is 20 percent, much less than the next tax category, the filers with tax liability between $1,000 and $2,500, at 84 percent. This tax reporting noncompliance of the overall Form 944 filers is mainly driven by those whose tax liability exceeds $1,000.

We expect the tax reporting behavior of Forms 943 and 944 would be similar, conditional on the fact that both filers file their employment tax return annually (in comparison with Form 941 filers who file quarterly). However, the high contrasting difference in the filing requirement accuracy between Form 944 filers and Form 943 filers demonstrates that Form 944 filers behave very differently from Form 943 filers, in terms of filing behavior. Among Form 944 filers, Form 944 filers with tax liability over $2,500 are the least likely to have filed with the correct tax form.

In Tax Year 2017, it is also observed that approximately 34 percent of the employers who filed Form 944, had a Form 941 filing requirement.[21] This highlights the confusion/noncompliance among the Form 944 employers. Employers only choose one employment tax form to file: Form 941 or Form 944, but not both. This confusion/noncompliance does not appear to be the case for Form 943 employers, who follow IRS requirements to file Form 941 for regular employees, and Form 943 for the agricultural employees.[22]

The analysis on reporting and filing requirement reveals there is confusion or noncompliance in selecting employment tax forms among Form 944 filers. Reporting tax without following IRS requirements creates a significant cost for both the taxpayers and the IRS. For taxpayers, filing the incorrect form complicates and delays the tax return; for the IRS, filing the incorrect tax form generates extra work. While the tax return without the correct filing requirement will be rejected by the Integrated Data Retrieval System (IDRS), the human effort to try to manually revise the incorrectly filed tax return drains the already limited IRS resources.

5.2 Timely Filing

Timely filing of tax returns is an essential component of tax compliance. Filing delinquency not only generates interest and penalties, it also produces more work and cost for the tax administration side. In this section we will examine whether the Forms 941, 943, and 944 are filed timely.

Form 941 is required to be filed every quarter, even if the employer has no taxes to report (unless the employer is a seasonal employer). The due dates of the four quarters are April 30, July 31, October 31, and January 31 of the following year. Both of the annual filing of Forms 943 and 944 are required to be filed by January 31 following the close of the tax year.

Employers that file employment tax returns late will incur penalties and interest charges, which can significantly impact those employers with large tax liabilities. However, large businesses can be impacted even more by filing delinquencies, due to the time lag in processing the annual tax returns (annual tax delinquencies are not identified until the following year's return is processed), thus allowing a full year's penalties to accrue.

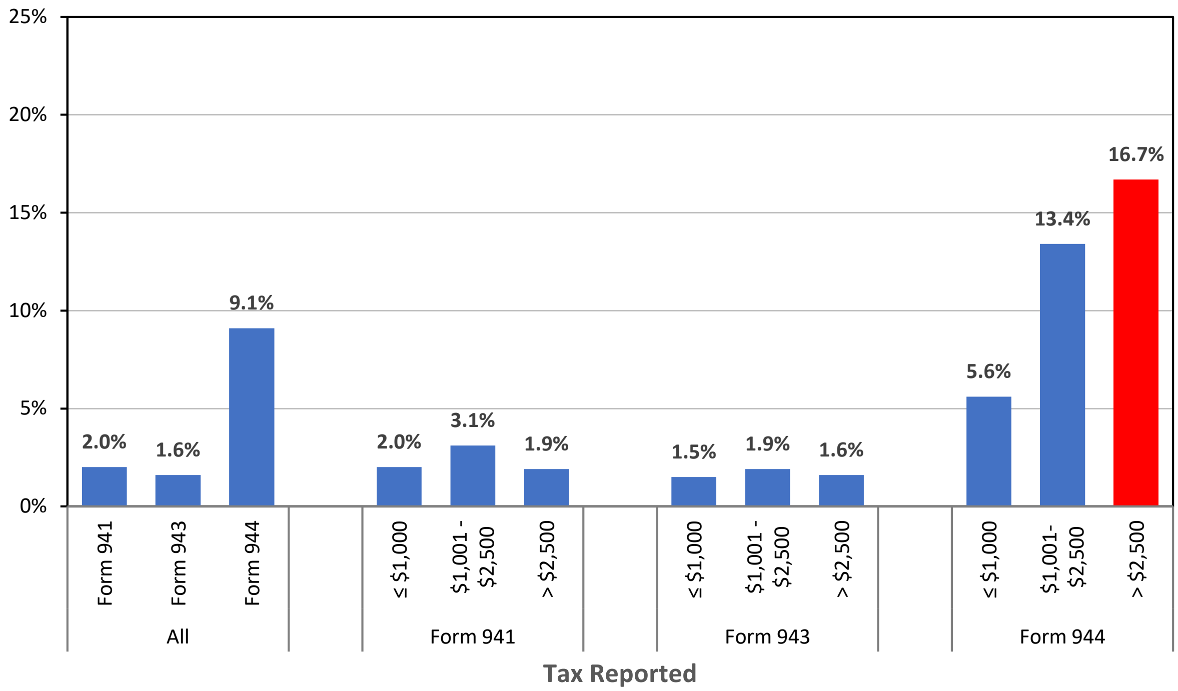
Figure 9 demonstrates the share of employers whose employment tax return was processed late,[23] by tax forms and by tax liability. For Tax Year 2017, the due date for both Forms 943 and 944 filing is January 31, 2018. In this study, we use the share of employers whose tax return was posted in 2019, which is one year after the tax due date, as a proxy for the filing delinquency.

[21] The statistics are not shown in the figure.

[22] We find in Tax Year 2017 that approximately 3 percent of the employers who filed the Form 943 also filed Form 941.

[23] The "late tax return" in this study is defined as the cycle posted date one year after the tax due date.

FIGURE 9. Share of Employers Who File Form 941/943/944 One Year After the Due Date, Tax Year 2017



We notice that the filing delinquency rate is the highest for Form 944, among the three employment tax forms. In Tax Year 2017, the filing delinquency rate for Forms 941 and 943 are similar, around 2 percent, much smaller than the rate for Form 944, around 9 percent. While both Forms 943 and 944 are annual filing forms, Form 944 filers have much higher probability to file late than Form 943 filers. It implies that the annual filing itself is not the barrier to filing a tax return timely.

Form 944 filers with higher tax liability are less compliant in terms of timely filing. In Figure 9 we further refine the analysis by breaking down the taxpayers by their tax liability. We find that there is a small variation in terms of filing delinquency among the filers with different tax liabilities, for both Forms 941 and 943. Specifically, the filing delinquency ranges from 2 to 3 percent for Form 941 and Form 943 filers, across all tax liabilities.

However, we observe that Form 944 employers with higher tax liabilities are more likely to file the tax late, than the Form 944 filers with lower tax liability. Approximately 6 percent of Form 944 filers with annual tax liability equal $1,000 or less filed over a year after the due date, but the rate for Form 944 filers with annual tax liability over $2,500 is as high as 17 percent. Further, comparing the filers within the same tax liability group, but with different forms, the filing delinquency for Form 944 is the highest.

Annual Form 944 filers with large tax liabilities will suffer higher penalties if they file their taxes late, versus the quarterly Form 941 filer, since the filing delinquency would be identified earlier, and therefore less interest and penalties would be accumulated.

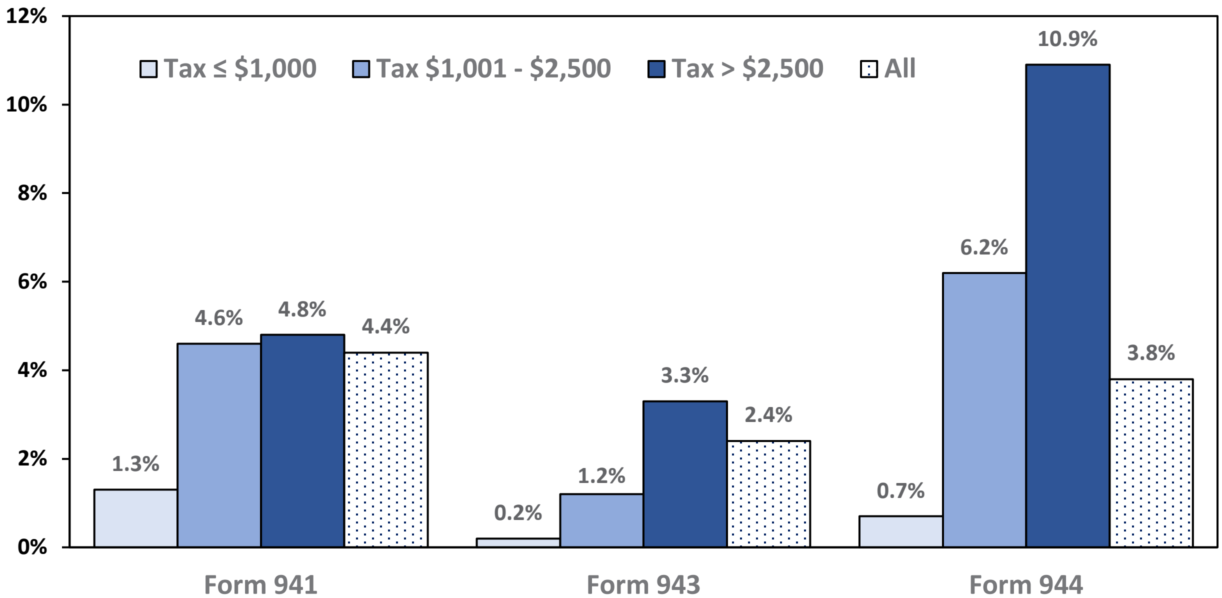
5.3 Tax Due Payment

When employers fail to pay their employment taxes, this produces a balance due on their tax accounts, which generates penalties, interest, and also jeopardizes their employees' future Social Security and Medicare benefits. Figure 10 presents the tax payment delinquency for Forms 941/943/944 filings in Tax Year 2017.[24]

[24] The CDW data are used to analyzing on-time payment compliance.

Form 943 filers have the lowest payment delinquency rate. Figure 10 shows that in Tax Year 2017, on average approximately 2 percent of the Form 943 filers had a balance due in their account. During the same period, the payment delinquency rate for the Forms 944 and 941 filers was double, at around 4 percent.

FIGURE 10. Payment Delinquency Rate, by Tax Forms and by Tax Liability, Tax Year 2017



We further disaggregate the employers into three subgroups for each tax form, based on their tax liability: $1,000 or less, between $1,001 and $2,500, and over $2,500. We find that the tax liability correlates with the payment delinquency likelihood: on average, employers with higher tax liabilities are more likely to have a balance due.

Filers with higher tax liabilities have higher tax payment delinquency rates, for each of the three employment tax forms. Taking Form 944 as an example, in Tax Year 2017, around 1 percent of Form 944 filers with tax liability of $1,000 or less had not paid their tax liability; the rate went up to 6 percent when the tax liability went up, for the subgroup with tax liability between $1,001 and $2,500; the rate continues to rise up to nearly 11 percent, for the subgroup with tax liability over $2,500. Same pattern is detected for both Forms 941 and 943 filing as well, with smaller magnitude.

Form 944 filers with tax liability over $2,500 had the highest rate of payment delinquency, among the nine tax subgroups shown in Figure 10. While Form 943 employers and Form 941 employers with annual tax liability over $2,500 had a payment delinquency rate of 3 and 5 percent, respectively, the payment delinquency rate for Form 944 with annual tax liability over $2,500 is much higher, reaching a two-digit number at nearly 11 percent.

Large employers that file the annual Form 944 late will incur higher penalties and interest on collectable accounts than large quarterly Form 941 filers and small annual Form 944 filers. One suggestion to help the large employers that file Form 944 become more compliant, and thus reduce their tax burden, is to ensure these large businesses file the correct form in a timely manner.

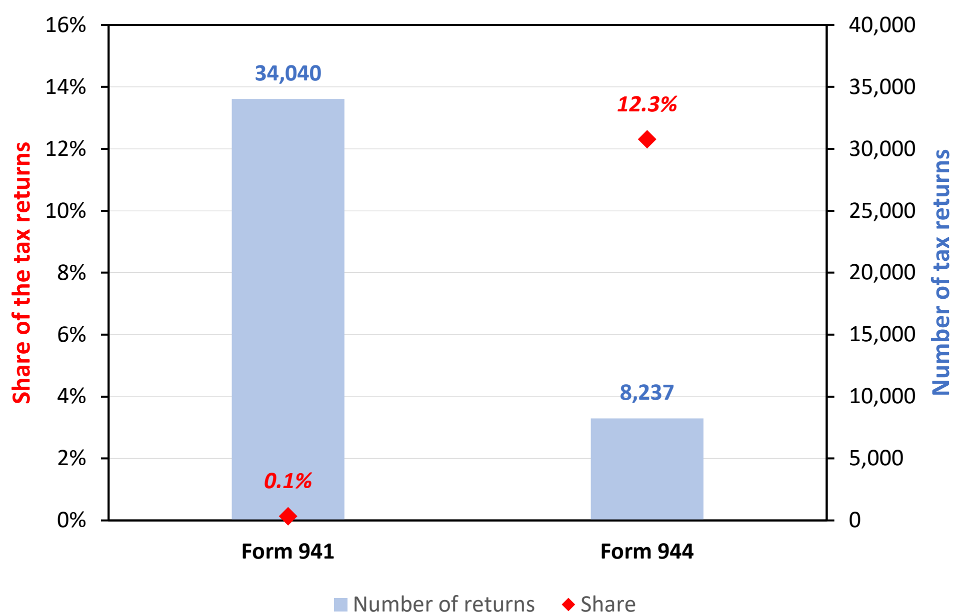
5.4 Consequence of Tax Reporting Noncompliance

When a taxpayer files the tax return correctly, the tax returns will be processed and posted in the IRS system. But, when there is an error in reporting (e.g., the filing requirement is not followed), the tax return is automatically rejected. However, such errors could be adjusted if a tax examiner corrects and hence the tax return is posted in the IRS system.[25] Thus, we use the record of the "unpostable correction"[26] captured in the Compliance Data Warehouse (CDW), as an indicator of taxpayer reporting noncompliance and burden to tax administration.

Figure 11 presents the unpostable returns,[27] corrected, posted, and captured by CDW, for Forms 941 and 944.[28] The left axis displays the share of tax returns being corrected (out of total tax returns filed in that tax year), and right axis shows the number of returns. We see that in Tax Year 2017, there were 34,000 unpostable corrected tax returns for Form 941, accounting for approximately 0.1 percent (out of 24 million Form 941 returns). Though there were only around 8,000 unpostable corrected tax returns for Form 944 in Tax Year 2017, the rate of unpostable corrected returns for Form 944 is 12 percent, due to the small number of Form 944 filings (around 67,000).

The high rate of the unpostable corrected for Form 944 not only generates a high internal cost and burden for IRS to process Form 944, but it also produces burden for taxpayers as well. Based on Sandstrom (2020), the IRS often will contact the taxpayers to resolve the issue and request that they file the correct return, resulting in time cost as well as monetary cost for taxpayers. In addition, the returns that are not corrected are often considered noncompliant by IRS nonfiler programs. This causes a burden for both taxpayers and the tax administration: for taxpayers, they filed employment tax returns, but end in a noncompliant program; for tax administration, it is the loss of uncollectable tax revenue.

FIGURE 11. Unpostable Corrected Tax Returns for Forms 941 and 944, Tax Year 2017



[25] For detail, please see Sandstrom (2020).

[26] The tax returns that were rejected and have not gotten corrected are beyond the scope of this study. Based on Sandstrom (2020), CDW does not pull in all the TRDBV data, and so information on the volume of rejected returns that could not be corrected is not available in CDW.

[27] The CDW data are used to estimate the CDW captured unpostable correction.

[28] There are no data regarding the unpostable corrections in CDW for Form 943 in Tax Year 2017.

6. Empirical Analysis of Tax Payment Noncompliance

6.1 The Model

To achieve a better insight of the tax noncompliance of the annual filing, a limited-dependent variable technique, logistic regression is used to relate the characteristics of annual filers to the probability of their tax payment delinquent behavior. We apply the model to the annual filing employers who filed Form 943 or Form 944 in Tax Year 2017.

$$ln\frac{P_i^f}{1-P_i^f} = \beta_0^f + \sum_{j=1}^{k}\beta_j^f x_{ij}^f \tag{1}$$

where f = *Form 943, Form 944*

P_i^f denotes the probability of tax noncompliance of employer i for Form f. In this model, the tax payment delinquency rate is used as an indicator for tax noncompliance.

Taking the antilog of Equation (1) on both sides, we derive an equation to predict the probability of tax noncompliance as follows:

$$P_i^f\left(Y_i^f = 1 \middle| X_i^f = x_i^f\right) = \frac{e^{\beta_0^f + \sum_{j=1}^{k}\beta_j^f x_{ij}^f}}{1 + e^{\beta_0^f + \sum_{j=1}^{k}\beta_j^f x_{ij}^f}} \tag{2}$$

The outcome Y_i^f we observed is a dichotomous variable. It takes a value of 1 if the employer i who filed Form f had a payment balance due, and a value of zero if not.

The independent variables included in the models are the tax liability category, size of the employers, industry sectors of the employers, and geographical location. Since Form 944 does not have information of the size of employers, we use the total wages and compensations of employees as a proxy for employer size. A logarithmic value is used. There are three tax liability groups: tax liability of $1,000 or less, between $1,001 and $2,500; and over $2,500.

6.2 The Result

Table 2 lists the coefficient estimates, their standard errors, the Wald Chi-Square statistic, the associated p-value, and odds ratio, for Forms 943 and 944 filing, respectively. The Wald Statistics testing the hypothesis that each coefficient equals zero in the logistic regression and total number of observations are also reported.

TABLE 2. Logistic Regression of Probability of Tax Payment Delinquency

Explanatory variable	Parameter estimate	Standard error	Wald Chi-square	Pr>Chi-square	Odds ratio
Panel A: Form 943					
Intercept	-7.8071	0.4356	321.19***	0.0001	
Log of total wages and compensations	0.1875	0.0132	200.93***	0.0001	1.21
Dummy, Tax liability ≤ $1,000 (base)					
Dummy, Tax liability between $1,001 and $2,500	0.6867	0.1525	20.29***	0.0001	**1.99**
Dummy, Tax liability above $2,500	1.2580	0.1398	80.97***	0.0001	**3.51**
Industry sectors fixed effects			Yes		
Geographical location			Yes		
Observations			184,737		
Chi-Square Statistics of overall significance of the regression = 2305.35 (p-value=0.0001)					
Panel B: Form 944					
Intercept	-5.6035	0.3760	222.06***	0.0001	
Log of total wages	0.1961	0.0207	89.63***	0.0001	1.22
Dummy, Tax liability ≤ $1,000 (base)					
Dummy, Tax liability between $1,001 and $2,500	1.6441	0.0916	322.19***	0.0001	**5.17**
Dummy, Tax liability above $2,500	1.9289	0.0943	418.72***	0.0001	**6.88**
Industry sectors fixed effects			Yes		
Geographical location			Yes		
Observations			66,897		
Chi-Square Statistics of overall significance of the regression = 2199.24 (p-value=0.0001)					

NOTE: *** Statistical significance at p-value = 0.001

In Panel A, the Chi-square statistic for the overall significance of the regression tests the hypothesis that all variable coefficients are simultaneously zero. The likelihood ratio Chi-square of 2,305 with a p-value of 0.0001 indicates that the model is statistically significant. The coefficients for employer size and tax liability groups are statistically significant.

In Panel A for Form 943 filing, the size of employer is positively associated with the probability of the tax delinquent behavior. The employers with larger size and more employees are more likely to have tax payment delinquency, if compared to the small employers with other factors unchanged. The estimated coefficient of 0.1875 specifies for each unit increase of the logarithm value of total wages, the log odds of tax payment delinquency increases (versus no tax delinquency) by 0.1875 unit.

The coefficients for the two dummy variables for tax liability groups are positive, indicating the large tax liability exhibits a positive contribution to filer's probability of tax balance due. For example, the estimated coefficient of 1.2580 for the dummy variable with tax liability over $2,500 implies that with tax liability over $2,500, versus tax liability of $1,000 or less, the log odds of tax payment noncompliance are increased by 1.2580.

The Logistic regression in Panel B for Form 944 filing displays a similar pattern: the filers with more employees and with more tax liability are more likely to have tax payment delinquent behavior.

The odds ratios for the Logistic regressions are also reported. Form 943 filers with tax liability between $1,000 and $2,500 has an odds ratio of 1.99, which means for Form 943 filing, the predicted odds of the tax payment delinquency rate for this group is 1.99 times the odds for the filers with tax liability of $1,000 or under. The Form 943 filers with tax liability above $2,500 have an odds ratio of 3.51, indicating for employers with tax liability over $2,500, the odds of being tax delinquent are 3.51 times larger than the odds for filers with tax liability $1,000 or under in Form 943 filing. In other words, the tax liability increases the odds of tax payment delinquency for Form 943 filing.

The odds ratio in the Logistic regression for Form 944 filing could be interpreted in the same pattern. The predicted odds of the tax delinquency rate for Form 944 filers with tax liability between $1,001 and $2,500 is 5.17 times the odds for the filers with tax liability $1,000 or less. And the odds ratio for Form 944 filers with tax liability over $2,500 is even higher, at the value of 6.88.

6.3 Model Interpretation

From the model, we have obtained two important insights. First, it shows that employers with bigger employees' size and larger tax liability tend to be more delinquent in tax payment, in both Forms 943 and 944 filing. This is consistent with our previous observations in the previous section: as tax liability moves from low tax group to high tax group, the tax payment delinquency rate rises, for both Forms 943 and 944.

Secondly, the employers filing Form 944 have higher odds ratio for tax payment delinquency than the employers filing Form 943. For the employers with tax liability over $2,500, the odds ratio is 3.51 in the regression for Form 943 filing, and it approximately doubles (6.88) in the regression for Form 944 filing, for the same tax group. This implies, in Form 944 filing, employers are more likely to have tax payments due than the employers in Form 943 filing, when they are in the same tax group.

Form 944 was introduced to serve small businesses with tax liability of $1,000 or under. However, we find that a significant share of filers with tax liability above this eligibility threshold filed this form. Our empirical model shows that these employers are more likely to have tax noncompliant behavior. Form 943 filers do not have this eligibility limitation for tax liability, though they do have requirements for farming employees. Though the larger size employers filing Form 943 tend to have more tax noncompliance, the tendency is much lower than the filing in Form 944.

The implementation and processing of Form 944, whose eligibility is based on the employer's annual tax liability, is more complex than Form 943, whose eligibility is based on the industry sector. Form 944 eligibility is based on annual tax liability, and the tax liability is sensitive to a lot of factors, including expansion or shrinkage of the production due to the environment change, covariate shocks, idiosyncratic shocks, or other risk factors. However, the setting of the filing requirement of the tax administration is less responsive to these factors. It sometimes requires multiple cycles to complete the change of the filing requirement of the taxpayer's requested change. Whenever these two conditions do not match, it produces a lot of difficulties and burden for both taxpayers and the tax administration, such as taxpayers filing of incorrect forms, or repeated self-correction, and unpostables on the side of tax administration, etc.

7. Summary

Urged by the tax experts and championed by the IRS Innovation Lab, we use the administrative tax return data to assess the effectiveness of two annual employment tax forms: Forms 943 and 944. We examined the filing population, tax liability, and tax compliance for the two annual employment reporting forms, together with the quarterly filing Form 941.

Different trends, for the filing populations of the annual forms, are observed: while Form 944 filing is declining rapidly, Form 943 filing has been stable. The goal of introducing Form 944 was to reduce the burden of small businesses, however this tax reporting form has not been widely adopted since its inception. In Tax Year

2017, there were around 600,000 employers with annual employment tax liability $1,000 or below, that chose to file Form 941, instead of Form 944. Conversely, the total Form 943 filings has remained relatively stable: driven mainly by the consistent filing of large employers with farm workers. Furthermore, the total Form 943 tax liability has gradually grown over the last decade, due to wage increases of farm workers. In Tax Year 2017, while the total tax liability of Form 943 reached $7.8 billion, the total tax liability of Form 944 was only $0.6 billion.

Though the eligibility to file Form 944 is an annual employment tax liability of $1,000 or lower, we find that the top 10 percent of Form 944 filers have tax liabilities over this threshold. In Tax Year 2017, approximately 12 percent of Form 944 filers had an annual tax liability over $10,000. Over the last decade, a significant portion of Form 944 filers (on average 28 percent) had an annual employment tax liability over $2,500.

Form 944 filers are observed to have higher incidence of tax noncompliance, in reporting, timely filing, and tax due payments, than Form 943 filers. Form 944 filers, especially the Form 944 filers with annual employment tax liability over $2,500, have a higher rate of filing the tax form without the filing requirement, than Form 944 filers with lower tax liability. In Tax Year 2017, while 2 percent of Form 943 employers filed the Form 943, without having a Form 943 filing requirement, approximately 41 percent of Form 944 employers filed Form 944 without a Form 944 filing requirement. Among Form 943 filers with tax liability over $2,500, only 1 percent of them filed the incorrect tax form. In contrast, among Form 944 filers with tax liability over $2,500, around 86 pwecent of them filed the incorrect tax form.

Form 944 filers are more likely to file their employment tax return late. The contrast is more drastic among Form 943/944 filers with tax liability over $2,500: While less than 2 percent of Form 943 filers with tax liability over $2,500 filed the tax return late. Around 17 percent of Form 944 filers with tax liability over $2,500 filed the tax return late.

Form 944 filers tend to have a higher rate of payment delinquency than Form 943 filers. The tax payment delinquency rate is around 3 percent for Form 943 filers with tax liability over $2,500, and the rate tripled for Form 944 filers with tax liability over $2,500.

Our research shows the annual filers of Form 944 have more noncompliance in terms of tax reporting, timely filing, and tax payment in employment tax filing, in comparison with Form 943 filers. The implementation complexity of Form 944 requires more IRS resources and is less cost friendly.

The format of annual filing *per se* is not the main cause of the ineffectiveness. However, the combination of the annual filing setting and other factors together impact the effectiveness in implementation and processing of the annual forms. The other factors include rules and regulations regarding the eligibility threshold (e.g., eligibility based on employees' wages, or tax liability, or industry sector), taxpayers' correct understanding of the tax rules, the communication between the taxpayers and the IRS, the flexibility and responsiveness of the IRS when the underlying conditions of the tax eligibility have changed during the year, and the constraints of the funding resources of the tax administration, etc. The combination of these factors generates the complexity in processing, administration, and implementation of Form 944.

Finally, for some small employers with limited resources, it may be difficult for them to have a full understanding of the market, to have precision estimates of their profits and tax liabilities, to make decisions quickly, and to have a up-to-date knowledge of tax rules and regulations. Given this, the IRS may consider implementing education strategies to support small businesses with the tax literacy necessary for successful implementation and continuation of programs such as Form 944.

References

Internal Revenue Service (2013). Mimeo. "Burden Reduction for Certain Filers of Forms 944 and 941." SB/SE Finance, Research and Strategy. DEN0223. Internal Revenue Service.

Sandstrom, Barbara (2020). Mimeo. SB/SE. Internal Revenue Service.

Treasury Inspector General for Tax Administration (TIGTA) (2007). "The Employer's Annual Federal Tax Return (Form 944) Program Was Effectively Planned and Implemented, but Fewer Returns Than Anticipated Were Filed." Reference Number: 2007-30-182. Treasury Inspector General for Tax Administration.

Treasury Inspector General for Tax Administration (TIGTA) (2008). *Semiannual Report to Congress*. https://www.treasury.gov/tigta/semiannual/semiannual_mar2008.htm.

Appendix I. Forms 941, 943, and 944

Form **941 for 2017: Employer's QUARTERLY Federal Tax Return** 950117
(Rev. January 2017) Department of the Treasury — Internal Revenue Service OMB No. 1545-0029

Employer identification number (EIN) ☐☐ – ☐☐☐☐☐☐☐

Name *(not your trade name)*

Trade name *(if any)*

Address
Number Street Suite or room number
City State ZIP code
Foreign country name Foreign province/county Foreign postal code

Report for this Quarter of 2017
(Check one.)

☐ **1:** January, February, March
☐ **2:** April, May, June
☐ **3:** July, August, September
☐ **4:** October, November, December

Instructions and prior year forms are available at *www.irs.gov/form941*.

Read the separate instructions before you complete Form 941. Type or print within the boxes.

Part 1: Answer these questions for this quarter.

1	Number of employees who received wages, tips, or other compensation for the pay period including: *Mar. 12* (Quarter 1), *June 12* (Quarter 2), *Sept. 12* (Quarter 3), or *Dec. 12* (Quarter 4)			1	
2	Wages, tips, and other compensation			2	.
3	Federal income tax withheld from wages, tips, and other compensation			3	.
4	If no wages, tips, and other compensation are subject to social security or Medicare tax				☐ Check and go to line 6.
		Column 1		Column 2	
5a	Taxable social security wages	.	× 0.124 =	.	
5b	Taxable social security tips	.	× 0.124 =	.	
5c	Taxable Medicare wages & tips	.	× 0.029 =	.	
5d	Taxable wages & tips subject to Additional Medicare Tax withholding	.	× 0.009 =	.	
5e	Add Column 2 from lines 5a, 5b, 5c, and 5d			5e	.

Form **943**

Department of the Treasury
Internal Revenue Service

Employer's Annual Federal Tax Return for Agricultural Employees

► Go to *www.irs.gov/Form943* for instructions and the latest information.

OMB No. 1545-0035

2017

Type or Print

Name (as distinguished from trade name) | Employer identification number (EIN)

Trade name, if any

Address (number and street)

City or town, state or province, country, and ZIP or foreign postal code

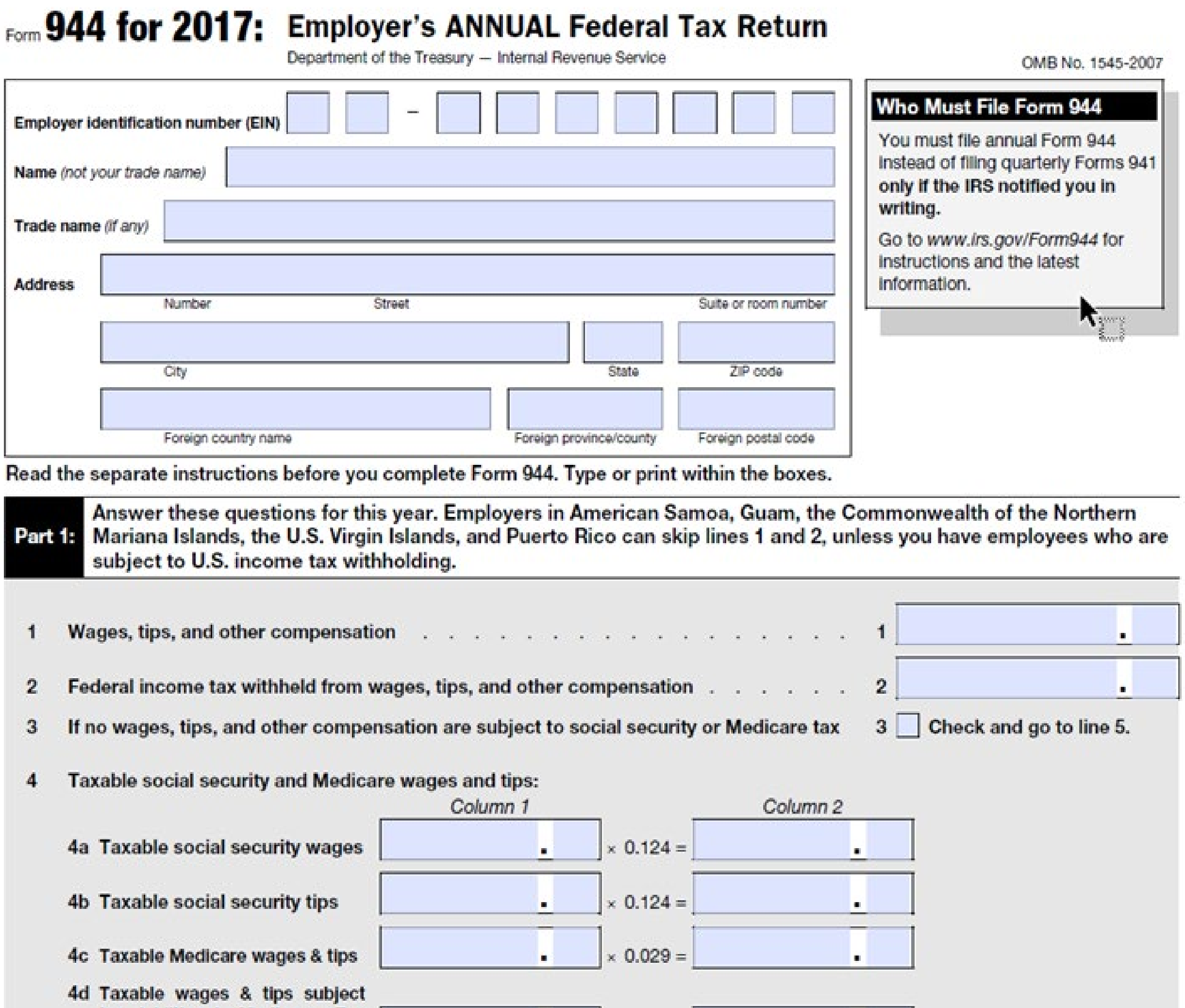
If you don't have to file returns in the future, check here . . . . ► ☐

If address is different from prior return, check here. ► ☐

1	Number of agricultural employees employed in the pay period that includes March 12, 2017 ►		1
2	Total wages subject to social security tax	2	
3	Social security tax (multiply line 2 by 12.4% (0.124))		3
4	Total wages subject to Medicare tax	4	
5	Medicare tax (multiply line 4 by 2.9% (0.029))		5
6	Total wages subject to Additional Medicare Tax withholding	6	
7	Additional Medicare Tax withholding (multiply line 6 by 0.9% (0.009))		7
8	Federal income tax withheld		8
9	Total taxes before adjustments. Add lines 3, 5, 7, and 8		9
10	Current year's adjustments		10
11	Total taxes after adjustments (line 9 as adjusted by line 10)		11
12	Qualified small business payroll tax credit for increasing research activities. Attach Form 8974		12
13	Total taxes after adjustments and credits. Subtract line 12 from line 11		13
14	Total deposits for 2017, including overpayment applied from a prior year and Form 943-X		14
15	**Balance due.** If line 13 is more than line 14, enter the difference and see the instructions ►		15
16	**Overpayment.** If line 14 is more than line 13, enter the difference ► $		Check one: ☐ Apply to next return. ☐ Send a refund.

- **All filers:** If line 13 is less than $2,500, **don't** complete line 17 or Form 943-A.
- **Semiweekly schedule depositors:** Complete Form 943-A and check here ► ☐ • **Monthly schedule depositors:** Complete line 17 and check here ► ☐

17	**Monthly Summary of Federal Tax Liability.** (**Don't** complete if you were a semiweekly schedule depositor.)		
	Tax liability for month	**Tax liability for month**	**Tax liability for month**



Form **944 for 2017: Employer's ANNUAL Federal Tax Return**

Department of the Treasury — Internal Revenue Service

OMB No. 1545-2007

Employer identification number (EIN) [ ][ ] – [ ][ ][ ][ ][ ][ ][ ]

Name *(not your trade name)*

Trade name *(if any)*

Address

Number | Street | Suite or room number

City | State | ZIP code

Foreign country name | Foreign province/county | Foreign postal code

Who Must File Form 944

You must file annual Form 944 instead of filing quarterly Forms 941 **only if the IRS notified you in writing.**

Go to *www.irs.gov/Form944* for instructions and the latest information.

Read the separate instructions before you complete Form 944. Type or print within the boxes.

Part 1: **Answer these questions for this year. Employers in American Samoa, Guam, the Commonwealth of the Northern Mariana Islands, the U.S. Virgin Islands, and Puerto Rico can skip lines 1 and 2, unless you have employees who are subject to U.S. income tax withholding.**

1 Wages, tips, and other compensation . . . . 1 [.]

2 Federal income tax withheld from wages, tips, and other compensation . . . . . . 2 [.]

3 If no wages, tips, and other compensation are subject to social security or Medicare tax 3 [ ] **Check and go to line 5.**

4 Taxable social security and Medicare wages and tips:

	Column 1		*Column 2*
4a Taxable social security wages	.	× 0.124 =	.
4b Taxable social security tips	.	× 0.124 =	.
4c Taxable Medicare wages & tips	.	× 0.029 =	.
4d Taxable wages & tips subject			

Appendix II. Form 944 At A Glance[29]

Form 944 program was implemented in Tax Year 2006 to reduce small business burden by allowing certain taxpayers to file an annual employment return instead of quarterly returns. Initially beginning with Tax Year 2006, the IRS determined who should file Form 944. In 2010, that changed. The taxpayer could opt out of filing Form 944 by requesting to do so.

Tax Years 2006–2009

Form 944 filing requirement criteria for Tax Years 2006 to 2009 was as follows:

- Had an annual liability for Social Security, Medicare and withheld federal income tax of $1,000 or less, and/or
- They filed Form 944 for the prior year and reported $1,000 or less in total tax liability.

Tax Year 2010 and After

For tax years beginning January 1, 2010, Rev. Proc. 2009-51 expanded an 'opt-out' option by allowing employers to opt out of Form 944 program for any reason when their filing requirement was for Form 944. Previously, the exceptions for Form 944 were narrowly defined and consisted mainly of Schedule H and Form 943 filers. Taxpayers could now opt out of filing Form 944 by calling the Service by April 1 of the current year, or by sending a written request postmarked by March 15 of the current year and obtaining approval to file Forms 941. The request must be timely, but there is no other criteria required to approve the request.

Currently, if a taxpayer has a filing requirement for Form 944:

- The Service set it when Form 944 was originally implemented, and they have not opted out by notifying the Service, and/or the taxpayer's previously filed Form 944 did not exceed the dollar threshold;
- The taxpayer tax was below the threshold and requested that their filing requirement be for Form 944; or
- As a first-time filer, the taxpayer set their filing requirement for Form 944 when submitting Form SS-4.

[29] Appendix II is drafted based on Sandstrom (2020).

Appendix III. Table

TABLE A1. Form 943 Tax Liability, by Tax Category and by Tax Year

Tax Year	Tax ≤ $1,000	$1,000 < Tax ≤ $2,500	Tax > $2,500	Total
2006	24,312,360	57,028,660	4,476,422,606	4,557,763,626
2007	23,183,382	55,544,654	5,697,717,550	5,776,445,586
2008	21,600,352	53,040,386	4,960,647,911	5,035,288,649
2009	20,859,242	52,779,151	4,738,449,079	4,812,087,473
2010	19,838,548	51,253,257	4,855,218,282	4,926,310,088
2011	18,733,453	49,730,691	4,867,844,770	4,936,308,913
2012	17,815,834	47,660,682	5,167,503,097	5,232,979,613
2013	16,615,344	44,659,796	6,001,947,812	6,063,222,952
2014	15,783,762	42,891,240	6,406,175,414	6,464,850,417
2015	15,293,027	41,253,006	6,690,764,198	6,747,310,231
2016	14,863,968	39,833,536	7,296,074,537	7,350,772,042
2017	14,323,281	38,696,577	7,712,783,809	7,765,803,667

Appendix IV. Take-up Rate for Form 944

Take-up rate is the percentage of persons who are eligible for some benefit or compensations and who take advantage of it. In this study, take-up rate is defined as the number of employers who are eligible for Form 944 and who also file this form as a percentage of the estimated number of employers who were eligible for Form 944. That is:

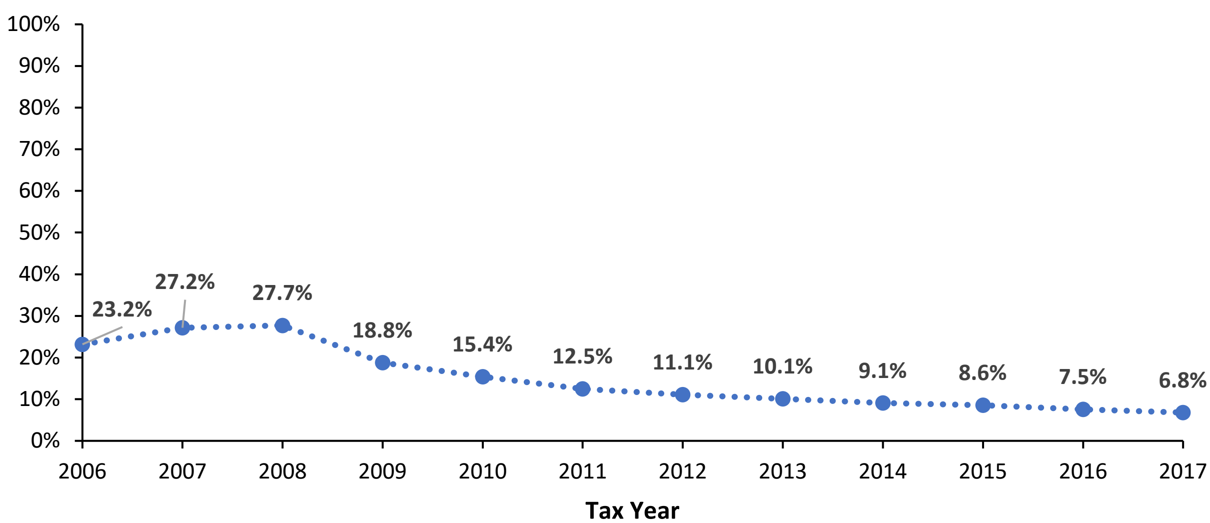
$$\textbf{\textit{Form 944 take-up rate}} = \frac{\textit{Number of employers who filed Form } 944\ (\textit{and eligible})}{\textit{Number of employers eligible for Form } 944} \qquad (3)$$

When we add Form 944 eligibility threshold into the formula, the above equation becomes:

$$\textbf{\textit{Form 944 take-up rate}} = \frac{\textit{Number of employers who filed Form } 944\ (\textit{with annual tax liability } \leq \$1{,}000)}{\textit{Number of employers filed Form } 941 \textit{ or Form } 944\ (\textit{with annual tax liability } \leq \$1{,}000)} \qquad (4)$$

The Form 944 take-up rate is then computed based on Equation (4). The result is presented in Figure A1. The take-up rate was at a moderate level when it was first implemented, however, it has been declining shortly after. While the average take-up rate was around moderate 26 percent in the first three consecutive years since Form 944 was introduced, it quickly dropped approximately 10 percent points in the following year, from 27.7 percent in Tax Year 2008, to only 18.8 percent in Tax Year 2009. The take-up rate has continued to plummet since then. In Tax Year 2017, Form 944 take-up rate was only 6.8 percent. The low take-up rate implies that most employers (e.g., 93 percent in Tax Year 2017) have not participated Form 944 program, even though they are eligible.

FIGURE A1. Form 944 Take-up Rate, Tax Years 2006–2017



5

∇

Appendix

Conference Program

11th Annual IRS-TPC Joint Research Conference on Tax Administration
Virtual Conference
June 24, 2021

Program

9:00–9:40 Opening

Eric Toder (Co-Director, Urban-Brookings Tax Policy Center) and

Barry Johnson (Acting IRS Chief Research and Analytics Officer)

Charles Rettig (Commissioner of Internal Revenue)

9:40-11:10 Improving Individual Taxpayer Compliance

Moderator: *Robert McClelland (Urban-Brookings Tax Policy Center)*

- Audit Contagion? Investigating the General Indirect Effect of Audits Through Tax Preparer Networks
  Ellen Badgley, Kyle Furlong, Lucia Lykke, Abby Ng, and Leigh Nicholl (MITRE); Alan Plumley (IRS, RAAS)
- Do Collateral Sanctions Work? Evidence from the IRS' Passport Certification and Revocation Process
  Paul R. Organ and Joel Slemrod (Univ. of Michigan); Alex Ruda and Alex Turk (IRS, RAAS)
- EITC Noncompliance: Examining the Roles of the Dynamics of EITC Claims and Paid Preparer Use
  Emily Y. Lin, Ankur Patel, and Alexander Yuskavage (Office of Tax Analysis, U.S. Department of the Treasury)

Discussant: *Tatiana Homonoff (New York University)*

11:10–12:20 Impacts of Variations in Process

Moderator: *Mary-Helen Risler (IRS, RAAS)*

- Sales Tax Administration and the Real Economy
  Roger M. White (Arizona State University)
- Using Discrete Event Simulations to Understand the Impact of Changes to IRS Processes
  Deandra Reinhart, Rafael Dacal, and Ariel Wooten (IRS, SB/SE); Jonathan Curtiss (MITRE)
- Effects of Post-filing Adjustments on Statistics of Income (SOI) Estimates
  Derrick Dennis, Jennifer Ferris, Chloe Gagin, Tuba Ozer-Gurbuz, Julia Shiller, and Christopher Williams (IRS, SOI)

Discussant: *Jennifer Stratton (U.S. Government Accountability Office)*

12:20–12:30 Break

12:30–1:15 Keynote Speaker

Mark Mazur (Acting Assistant Secretary for Tax Policy, US Department of the Treasury)

1:15–2:45 Developments in Technology and Analytics

Moderator: *Terry Ashley (IRS, Taxpayer Advocate Service)*

- New Approaches to Estimating the Extent of Nonfiling
  Tom Hertz, Pat Langetieg, Mark Payne, and Alan Plumley (IRS, RAAS); Maggie R. Jones (U.S. Census Bureau)
- Using Uplift Modeling to Improve ACS Case Selection and Compliance Outcomes
  Jan Millard (IRS, RAAS); Michael Stavrianos, Sarah Smolenski, Lauren Szczerbinski, and Travis Whitfield (ASR Analytics)
- Recent IRS Discriminant Function (DIF) Model Improvements
  Getaneh Yismaw, Taukir Hussain, Drew Johns, Jon Creem, and Mary-Helen Risler (IRS, RAAS)

Discussant: *Brian Erard (Brian Erard & Associates)*

2:45–4:15 Enhancing Taxpayer Customer Experience

Moderator: *Christine Oehlert (IRS, RAAS)*

- Increasing Take-up of the American Opportunity Tax Credit
  Crystal Hall (University of Washington), Anne Herlache (IRS, RAAS), and Mary Clair Turner (U.S. General Services Administration)
- Customer Experience for Small Business and Self-Employed Taxpayers
  Kahoa Bonhomme, Kathleen Holland, and Janice Hu (IRS, SB/SE)
- Are Annual Federal Employment Tax Returns Effective? An Economic Analysis of Filing, Reporting, and Payment Compliance Associated with Forms 943 and 944
  Yan Sun and Stephanie Needham (IRS, RAAS)

Discussant: *Jacob Goldin (Stanford Law School)*

4:15–4:20 Wrap-up

Eric Toder (Co-Director, Urban-Brookings Tax Policy Center)