Individual Nonfilers and IRS-Generated Tax Assessments: Revenue and Compliance Impacts of IRS Substitute Assessments When Taxpayers Don't File

Saurabh Datta, Stacy Orlett, and Alex Turk, Small Business/Self-Employed Division, Internal Revenue Service[1]

Background and Introduction

The U.S. income tax system relies on taxpayers voluntarily filing tax returns when required, and reporting and paying their tax liabilities. Each year, a fraction of taxpayers fail to file required returns. After the filing season, the IRS identifies potential nonfilers and attempts to secure returns via a series of notices and other contacts. When the taxpayers fail to respond by filing a return, the IRS can file a "substitute for return" that creates a tax assessment based on prior-year information and information obtained from third parties. Many of the substitute assessments are made via the IRS's Automated Substitute for Return (ASFR) process.[2]

The number of delinquent returns worked in the ASFR process varies from year to year. In recent years, the ASFR program has experienced a noticeable decline in resources and a corresponding decline in the number of delinquent returns worked by the ASFR process. This is partly attributable to a general decline in IRS budgets and partly due to the reallocation of nonfiler resources to other areas as IRS responsibilities expand. One criticism of the ASFR process is that in some cases the assessments are overstatements of the taxpayer's true liability. Most of the deductions and exemptions a taxpayer may be entitled to claim can be obtained only if the taxpayer files the return. Thus, the ASFR assessments may overstate the true amount of unpaid tax. However, not making the ASFR assessments leads, in many cases, to an understatement of the unpaid tax that is owed to the U.S. Government. Another criticism is that many of the ASFR assessments can be difficult to collect.

To make the best use of the resources available to the IRS, it is critical that the IRS and policy makers understand the impacts of the ASFR program on collecting delinquent taxes and fostering future filing and payment compliance. To explore these impacts, we develop models of the potential collection of ASFR assessments and then predict the impact of the ASFR program on subsequent filing compliance. We can then use these models to estimate, in terms of dollars collected and the numbers of delinquent returns, the opportunity costs of reductions to the number of cases worked in the ASFR program.

In the next section, we broadly highlight some of the recent research papers in the area of taxpayer compliance and behavior, which will help us in understanding the noncompliance issues and then formulate the economic and empirical model from ASFR's perspective.

Literature Review

The literature on taxpayer compliance is varied and has witnessed growth in recent years due to research in academics, public policy, and in legal enforcement.

---

[1] The views and opinions presented in this paper reflect those of the authors. They do not necessarily reflect the views or the official position of the Internal Revenue Service.

[2] ASFR is an automated process that generates notices and an automated "Substitute for Return" tax assessment if taxpayers do not resolve their delinquent returns. However, labor resources are needed to work with the taxpayer when the taxpayer responds to one of the letters or the default assessment.

Many economic models have been built to understand the interaction between taxpayers and tax authorities. The models developed in the literature are based on a principal-agent framework with highly simplified assumptions that fail to understand the complex relationship between the two parties. Understanding a taxpayer's behavior, psychology, moral, and social influences are critical elements in studying taxpayer compliance. Underestimating these factors has resulted in greatly overestimating noncompliance (Andreoni, *et al.* (1998)). Taxpayers' behavior may often appear to be "unethical," selfish, or irrational. Different taxpayers may behave differently under distinct circumstances, and some taxpayers may behave differently inter-temporally. However, taxpayers are not always driven by "unethical" traits, but are constrained by "bounded rationality" and underestimate the consequences of noncompliance (Alm and Torgler (2011)).

In recent years, attempts have been made to develop game-theory-based models intended to be more realistic, taking into account the repeated interactions between taxpayers and the tax authority. Considering the channels of interaction between taxpayers and the tax authority through notices and telephonic conversations may provide a more realistic model formulation and precise estimation (Andreoni, *et al.* (1998); Hashimzade, *et al.* (2013)). These models have enabled researchers to estimate both compliance and enforcement aspects simultaneously and address the problem of endogeneity in enforcement activities. However, on the empirical side, these models have broadly ignored some key aspects such as 100 percent document matching with third-party data (Plumley (1996); Kleven, *et al.* (2011)). The noncompliance rate and underreporting are high in cases of self-reported income, but in cases of income reported by third parties the tax evasion rate is very low. Therefore, supplementing tax administration data with household-level surveys and other governmental sources help in the detection of underreporting and improving voluntary compliance. Since data are collected for purposes other than for tax administration, the households and businesses are perhaps more likely to report their correct income and income sources. The magnitude of underreporting for self-employed businesses was estimated to be nearly 25 percent using U.S. data compared with tax administration data (Hurst, *et al.* (2014)). Moreover, most of the models assume audit rates to be fixed rather than endogenous. Audits as an endogenous tool of enforcement has resulted in greater compliance among self-reported taxpayers in a household study based in Denmark (Kleven, *et al.* (2011)).

The prime motivation of our research paper comes from Erard and Ho (2001), who specifically look at the issue of nonfilers with less restrictive assumptions. The authors incorporate nonfiling strategies adopted by nonfilers in a standard neoclassical theoretical model. The theoretical model accounts for sequential steps involved in a taxpayer's decision-making process that makes him decide whether to be compliant or noncompliant looking at the expected payoff from each decision point. Based on the theoretical model, the paper estimates a simultaneous equation model where simultaneity exists between the probability that a taxpayer will file a return and the likelihood of the taxpayer being located. The paper uses a 25 percent random sample of the IRS TCMP Phase III Survey, which has 54,000 individual income tax returns for Tax Year 1988. This research identifies key behavioral, demographic, and financial factors that influence a taxpayer's decision to file a tax return. A comparison between filers and nonfilers suggests that nonfilers have relatively fewer offsets and itemized deductions compared to the former. Unlike the income sources of the filers, a nonfiler's income mostly comprises business income and capital gain receipts. Moreover, the taxpayers who are close to the filing threshold seem to be deterrent to filing as the burden appears to outweigh the benefits.

Deterrence in noncompliance accrues when a nonfiler is treated directly by the tax authority for noncompliance in the form of audit or imposition of penalties and interests in addition to the unpaid tax liabilities. This is also known as the "direct" effect. The direct effect of enforcement has positive effect on voluntary compliance in the subsequent years for the treated nonfiler. Additionally, a change in compliance behavior may be triggered if the general population becomes aware of a change in enforcement level by the tax authority. This may result in deterrence in noncompliance. This effect is termed as "indirect" or "induced" effect in the existing literature (Plumley, 1996). Plumley (1996) has shown that the deterrence effect of audits on taxpayers is 11 times larger than the audits by themselves in his study involving data over a 10-year period. However, the estimation of "indirect" effect has been mainly confined to the tax audit literature. This concept has important economic and policy significance in the realm of filing and payment compliance and the ASFR program.

The aim of this research paper is not only to study the direct impact of ASFR on revenue collection and subsequent voluntary compliance, but also to estimate the indirect effect of this treatment on the general

nonfiler population. In doing so, the paper improves and extends the analysis of previous research in this area and contributes to the literature of taxpayer compliance by outlining a more realistic theoretical model considering imperfect information in a principal-agent framework. The empirical estimation from this research paper uses comprehensive taxpayer data, which are a substantial improvement over the predecessors who have been largely deficient in their analysis due to limited data availability.

Theoretical Model

We develop a theoretical model following tax compliance models proposed by Allingham and Sandmo (1972), and Andreoni, Erard, Feinstein (1998) using an expected utility maximizing framework within a principal-agent model setting—the principal being the tax authority and the agent being the taxpayer.

We assume that a representative (partly noncompliant) taxpayer who doesn't file an income tax return (nonfiler), but who has total income Y (some of which is unknown to IRS), has an amount θW of tax withheld from his "visible" income W (known to IRS). In the second step, only a certain portion of his income is reported on information returns (Y_r) while he suppresses the residual component of his income Y_s. However, the IRS, through its direct and indirect intervention, comes to know a proportion α of suppressed income Y_s. In the next step, if the nonfiler's reported income satisfies the criteria of the ASFR program, the case may be assigned to the ASFR treatment stream, and the IRS may identify another portion of the undisclosed income. Based on this logical structure, we propose to formulate a nonfiler's optimization problem and identify the instruments available to the IRS to promote taxpayer compliance.

Mathematically, we can formulate this problem as follows:

Let a nonfiler's total income equal:

$$Y = W + Y_s + Y_r$$

Where:

W = Income known to the IRS (on which tax is withheld);

Y_s = Income suppressed by the nonfiler; and

Y_r = Income reported for the nonfiler on information returns.

Assuming that α (0<α<1) proportion of suppressed income Y_s is known to the IRS through its direct and indirect interventions,[3] one can decompose his suppressed income level as $Y_s = \alpha Y_s + (1-\alpha)Y_s$, which means

$$Y = W + \alpha Y_s + (1-\alpha)Y_s + Y_r \text{ where } \alpha = \alpha(IRS) \text{ with } \alpha' > 0$$

In other words, $Y_r = Y - W - \alpha Y_s - (1-\alpha)Y_s$

$Max\ [Y_r] = (Y - W)$ if $Y_s = 0$. This means the nonfiler's income is completely transparent to the IRS.

$Min\ [Y_r] = 0 \Rightarrow Y_s = (Y - W)$ This is a nonfiler who is suppressing all non-withheld income. Our presumption is that even then, the IRS can approximate the nonfiler's true income and attempt to extract the rightful taxes due through successive efforts.

Now, if p is the probability that a given nonfiler who has suppressed a part of his income is assigned to the ASFR treatment stream, then one can reasonably assume that:

p = Probability that the given nonfiler is selected to be treated in ASFR, where $p = p(ASFR)$ with $p' > 0$.

Due to ASFR intervention, the IRS may be able to get more information about the suppressed portion of the taxpayer's income. Let the proportion of suppressed income from Y_s as identified by ASFR be β, where:

[3] Direct and indirect interventions include obtaining information about wages, income, interest, dividend, pension and social security incomes, etc. from direct and third-party reported sources.

$\beta = \beta(ASFR)$ with $\beta'>0$ and $0<\beta<1$ This means at this stage that revelation of additional income through ASFR efforts is $\beta(1\text{-}\alpha)Y_s$

Based on ASFR's intervention, the implied tax liability of a taxpayer would now be:

$$\text{Implied Tax Liability} = \theta(W+Y_r) + (\theta+\gamma)\alpha Y_s + T_1 + T_2 + (\rho+\theta)\beta(1-\alpha)Y_s$$

Here θ is the proportional tax rate, γ is the proportional penalty rate at the time of IRS assessment, ρ is the proportional penalty rate imposed by ASFR on the revealed portion of the suppressed income, T_1 and T_2, are flat-rate penalties assessed by the IRS in the beginning while matching direct and third-party reported documents and ASFR, respectively. For simplicity, assume this is a one-period static model, where everything is taking place within the same period, so there is no interest charge imposed on the taxpayer.

$\theta(W+Y_r)$ is the potential tax liability in the first stage on the withheld and revealed portion of the taxable income as reported for the nonfiler.

$(\theta+\gamma)\alpha Y_s + T_1$ is the tax and penalty charged by the IRS on the identified part of the suppressed portion of the income in the second stage (beyond the withholding stage).

$(\rho+\theta)\beta(1\text{-}\alpha)Y_s + T_2$ is the tax and penalty charged by the ASFR treatment stream on the further identified part of the suppressed income in the third stage.

Based on all this available information, a nonfiler's optimization problem can be stated as:

$$\begin{aligned}\underset{w.r.t.\ Y_s}{Max}\ EU(C) = {} & p(ASFR)\bullet U[Y-\{\theta(W+Y_r)+(\theta+\gamma)\alpha Y_s + T_1 + T_2 + (\rho+\theta)\beta(1-\alpha)Y_s\}] + \\ & (1\text{-}\ p(ASFR))\bullet U[Y-\{\theta(W+Y_r)+(\theta+\gamma)\alpha Y_s + T_1\}] \\ = {} & p(ASFR)\bullet U[Y-T-\theta\{Y-Y_s(1-\alpha-\beta(1-\alpha))\}-Y_s\{\gamma\alpha+\rho\beta(1-\alpha)\}] + \\ & (1\text{-}\ p(ASFR))\bullet U[Y-T_1-\theta\{Y-(1-\alpha)Y_s\}-\gamma\alpha Y_s]\end{aligned}$$

$$\text{Let } A = Y-T-\theta\{Y-Y_s(1-\alpha-\beta(1-\alpha))\}-Y_s\{\gamma\alpha+\rho\beta(1-\alpha)\}$$
$$B = Y-T_1-\theta\{Y-(1-\alpha)Y_s\}-\gamma\alpha Y_s$$

We assume $T = T_1 + T_2$. An underlying assumption is that the tax liability of the nonfiler is greater than their withheld amount. That is,

$$\text{E (Tax liability with penalties and interest)} \geq \theta W$$

Also, for simplicity, assume that a nonfiler has a Constant Relative Risk Aversion (CRRA) utility function (Ljungqvist and Sargent, 2000):

$$U(\text{C}) = \frac{C^{1-\mu}}{1-\mu}$$

where $0<\mu<1$, and μ signifies risk aversion, with $U'(C) > 0$ and $U''(C) < 0$.

Based on the choice of utility function, the first order conditions can be derived as:

First Order Condition w.r.t. Y_s:

$$p(ASFR)\bullet A^{-\mu}[\{\theta(1-\alpha-\beta(1-\alpha)\}-\{\gamma\alpha+\rho\beta(1-\alpha)\}]+(1\text{-}\ p(ASFR))\bullet B^{-\mu}[\theta(1-\alpha)-\gamma\alpha]=0$$

The first order condition needs to be solved for Y_s given the values of ρ, θ, α, β, p, T_1, T_2, and Y.

Y_s will be a nonlinear equation expressed in terms of the exogenous parameters and the instruments under IRS-ASFR's control, which can be solved using numerical simulations.

$$\text{Let } C = \{\theta(1-\alpha-\beta(1-\alpha)\}-\{\gamma\alpha+\rho\beta(1-\alpha)\}$$
$$D = \theta(1-\alpha)-\gamma\alpha$$

Second Order Condition w.r.t. Y_s can be written as:

$$p(ASFR) \bullet \mu A^{-\mu-1} \bullet C^2 - (1 - p(ASFR)) \bullet \mu B^{-\mu-1} \bullet D^2 < 0$$
$$\Rightarrow -[p(ASFR) \bullet \mu A^{-\mu-1} \bullet C^2 + (1 - p(ASFR)) \bullet \mu B^{-\mu-1} \bullet D^2] < 0$$

The above expression is true since all the elements within the parentheses are positive quantities by construction.

In this model, there are two special cases, which deserve attention:

1. A solution value of $Y_s = 0$ suggests the IRS is in a position to set up parameter values that produce 100-percent compliance.
2. $Y_s = (Y\text{-}W)$ suggests the IRS has no income information beyond W; that is $Y_r = 0$. This is a case of a nonfiler. The IRS has to work hard through several successive steps to obtain an estimate of Y and impose tax on the nonfiler.

The objective of the IRS and ASFR may be viewed as:

Min $Y_s = Y_s\ (\rho, \theta, \alpha, \beta, p, T_1, T_2)$ subject to the resource constraints and based on available information on Y. A combination of the parameters $\rho, \theta, \alpha, \beta, p, T_1, T_2$ and a realistic estimate of Y will minimize the value of Y_s.

Interestingly, this model captures the natural conflicts of interest—while a taxpayer may be interested in maximizing his utility through choice of suitable value of Y_s, the IRS would like to see a minimum, even a zero value of Y_s, and thus would like to choose an appropriate combination of values of the parameters under its control. The present model formulation does not explicitly cover the process through which the IRS chooses the right combination of values of these parameters (i.e., how the IRS chooses Y_s as a function of these parameters from the taxpayer's reaction function as derived above). The final values of Y_s and the underlying parameters, which may be interpreted as Nash equilibrium, will invariably be the result of a convergent sequence of interactions (provided it exists) between the two sides over time (once we introduce explicit and discrete time lags based on observed behavior). However, the IRS never operates with unlimited budget to choose any combination of values of the underlying parameters (including a predicted value of Y). So, more often than not, the model formulation from the IRS side must explicitly incorporate how over time the IRS manages to ease its budget constraint, besides undertaking certain necessary reforms to improve effectiveness of its systems and process.

However, an achieved Nash equilibrium is still not a Pareto-optimal move. In order to achieve Pareto optimality or Nash bargaining equilibrium, the IRS and the taxpayer must be engaged in a 'Coasian' negotiation process (Coase, 1960; Milgrom and Roberts, 1992).

The above stated theoretical model suggests the purpose of the ASFR program is to minimize Y_s and thereby maximize tax revenue (or Dollars Collected) and then promote voluntary subsequent compliance in the successive years. The factors that minimize Y_s help in identifying the taxpayer's characteristics that maximize the expected revenue for the IRS based on this principal-agent framework.

Now, suppose ASFR treatment is applied to n_1 taxpayers out of n, where n were eligible for ASFR treatment.[4] In the first stage of the estimation process, we estimate the probability of a taxpayer's case being selected for ASFR treatment (p). The probability can be estimated based on observable taxpayer characteristics. In the second stage, given the probability that ASFR has worked the case, we estimate the revenue collection over a period of the next 3 years.[5] We assume that the dollars collected over the next 3 years depends on ASFR treatment, the probability of being worked by the ASFR treatment stream, and other observable taxpayer characteristics beyond those captured in the probability measure. In other words, there are certain factors that affect

[4] Since ASFR doesn't have unlimited resources, it selects to work certain cases based on the observable characteristics of the taxpayer.

[5] Three years is selected based on the maximum time an ASFR case generally takes time to resolve under normal circumstances.

revenue collection that do not affect the probability of being selected in ASFR, which are suitably introduced in this model specification.

The variable signifying ASFR treatment captures the direct treatment effect of ASFR on revenue collection, whereas the probability measure estimates the indirect effect of ASFR. The indirect effect captures the additional productivity of the case due to the reason that it has been worked by ASFR. This measure provides an empirical evidence of ASFR's additional effectiveness in collecting revenue by employing its collection instruments beyond its direct effects.

Furthermore, we estimate the future voluntary compliance of taxpayers who have been through the ASFR treatment process earlier. We specifically estimate the voluntary compliance of these taxpayers 2, 3, and 4 tax years later.[6] The factors affecting future voluntary compliance depend on a taxpayer's past observable characteristics, previous treatment by ASFR, probability that the case was assigned earlier to ASFR, and whether the taxpayer self-corrected before the next tax return was due. We argue that the taxpayer who self-corrected before the next tax return was due demonstrated greater willingness to be compliant and therefore needs to be controlled for suitably in the empirical model. Analogous to the earlier model, the indicator that ASFR worked this taxpayer's case previously and the probability of being worked by ASFR earlier measures the direct and indirect effects of ASFR on taxpayer's future voluntary compliance respectively.

In the following sections, we explain in detail the delinquent process, explain the data sources and variables selected, and then estimate the empirical models.

Summary of the IRS Return Delinquency Process

The IRS individual Return Delinquency process is illustrated in Figure 1. The process begins by identifying individual taxpayers who may be required to file but have not filed a tax return (e.g., Form 1040) by the Return Due Date.[7] A primary method the IRS uses to identify these taxpayers is the individual Return Delinquency case creation process. The case creation process is critical for identifying income and other information for these taxpayers, which is reported by third parties to the IRS on several types of information returns (e.g., Form W-2, Form 1099-R, etc.).[8] Using this information alone will not identify all nonfilers, but it will identify those who have some sort of income that is reported to the IRS. We can refer to these as the *known* nonfilers. An example of an *unknown* nonfiler would be an individual who has only cash income that is not reported on information returns and therefore is not identified during the case creation process.

After some additional compliance checks, a portion of these known nonfilers will be identified as required to file and will go into the Return Delinquency notice process. The nonfilers entering the notice process will receive up to two notices requesting them to file their tax return. During the notice process, a taxpayer has up to 14 weeks to respond. If a taxpayer does not respond to these notices, the case may then proceed to Taxpayer Delinquent Investigation (TDI) status.

The types of treatment a nonfiler in a TDI status receives varies based on case characteristics. A TDI may be treated by various functions including the Automated Collection System or Call Site (ACS), a Revenue Officer from a Field Collection office (FC), the ASFR program, and/or others. The ASFR program is an important program to the IRS for enforcing filing compliance by determining and assessing a tax liability when the taxpayer has not come forward with a return.[9] Some TDI cases go directly to the ASFR inventory after the notice process and some of the ASFR inventory is transferred to the ASFR function from other functions (e.g., ACS or FC) after unsuccessful attempts to secure or otherwise resolve the delinquent return, as long as the case meets specific ASFR eligibility criteria.[10]

[6] The very next tax year may be too short of a period of time to measure voluntary compliance. Therefore we selected 2 tax years later. Also, the results for 3 tax years later are consistent with the 2 years later results.

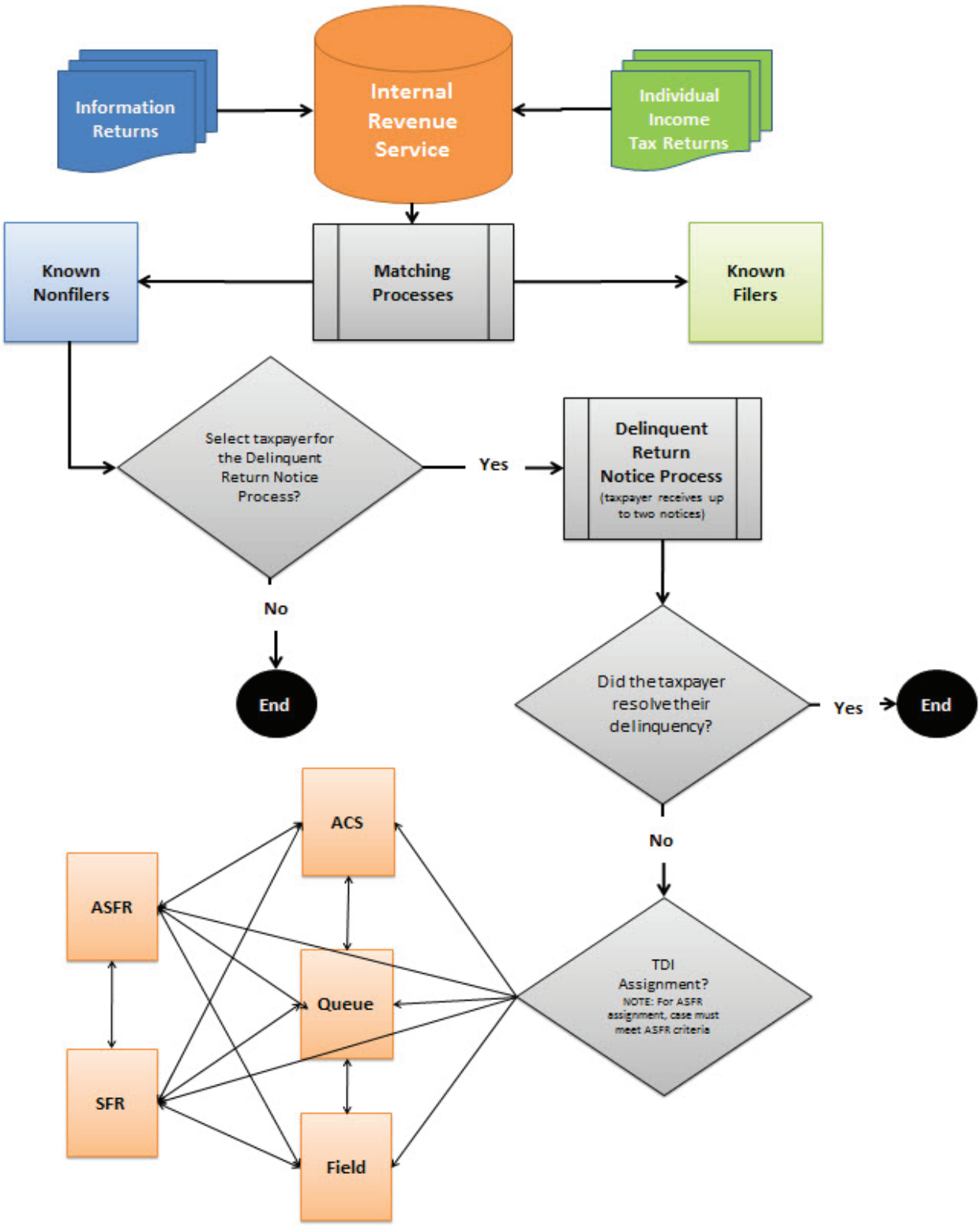
[7] Internal Revenue Service. Internal Revenue Manual 5.19.2.1 (01-16-2015) "What is the IMF Return Delinquency Program?"

[8] Internal Revenue Service. Internal Revenue Manual 5.19.2.4.1 (01-16-2015) "IRP Income."

[9] Taxpayers must file a return if they wish to demonstrate that their tax liability is different than the ASFR assessment. Payments of assessed amounts can be made by the taxpayer, offset from refunds claimed by the taxpayer for other tax years, or generated from liens on the taxpayer's assets or levies on the taxpayer's income sources.

[10] Internal Revenue Service. Internal Revenue Manual 5.18.1.3.1 (12-09-2014) "ASFR Criteria."

FIGURE 1. IRS Return Delinquency Process



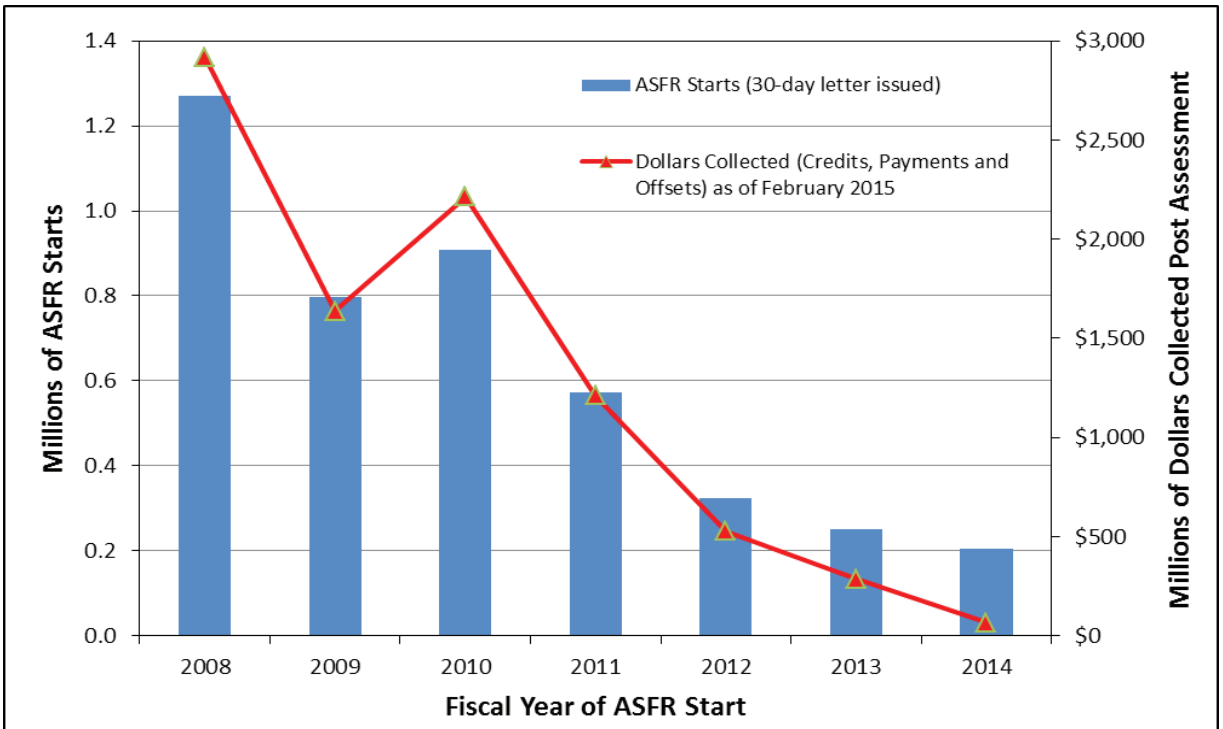
Source: Internal Revenue Service, Internal Revenue Manual 5.19.2.1

Cases received by the ASFR program are prioritized by Refund Hold, Tax Year, and Net Tax Due.[11] For purposes of our research, we are excluding Refund Hold cases worked by ASFR to focus on the discretionary TDI cases worked and available to be worked. Refund Hold cases must be worked within a designated amount of time and differ from non-Refund Hold because the service is holding a refund for that taxpayer from another tax return. The taxpayer will be notified that the Service is holding the refund, and that he must resolve all of his delinquent returns within the last 5 years prior to the current year before the IRS releases the refund. By holding the refund, the taxpayer, arguably, has different motivations to file the delinquent tax returns compared to the non-Refund Hold cases.

The ASFR process begins with various compliance checks. If the taxpayer passes these checks, a 30-day letter (Letter 2566) will be systemically sent to the taxpayer giving him 30 days to respond. If there is not a sufficient response from the taxpayer to the 30-day letter, then ASFR generates a Statutory Notice of Deficiency (also known as the 90-day letter) sent by certified mail. If there is still not a sufficient response from the taxpayer after the 90-day letter is issued, then the ASFR process will systemically request a default assessment based on the proposed tax assessment.

Figure 2 provides the number of cases started by ASFR, where a 30-day letter was issued, over the last 7 fiscal years. Historically, ASFR has had more cases available to initiate than available resources to work them. As ASFR resources have declined over the years, fewer cases have been started. In FY 2008, over 1.2 million cases were started by ASFR compared to only approximately 200,000 in FY 2014. The line on the figure represents the dollars collected to date for respective ASFR starts in that fiscal year. Since the cases started in FY 2008 have had more time to collect compared to cases worked in FY 2014, the decline in dollars collected to date is due to both the decline in the number of cases and the shorter time (so far) to collect. On cases ASFR started in FY 2008, the IRS has collected over $2.5 billion to date. The decline in cases worked also means fewer returns secured and lower dollar amounts collected on these delinquent returns.

FIGURE 2. ASFR Starts (30-day letters issued) During Fiscal Years 2008-2014



Source: IRS, Compliance Data Warehouse Individual Master File Transaction History, as of February 2015 (cycle 201508)

[11] Internal Revenue Service. Internal Revenue Manual 5.18.1.3.2 (06-20-2012) "ASFR Prioritization."

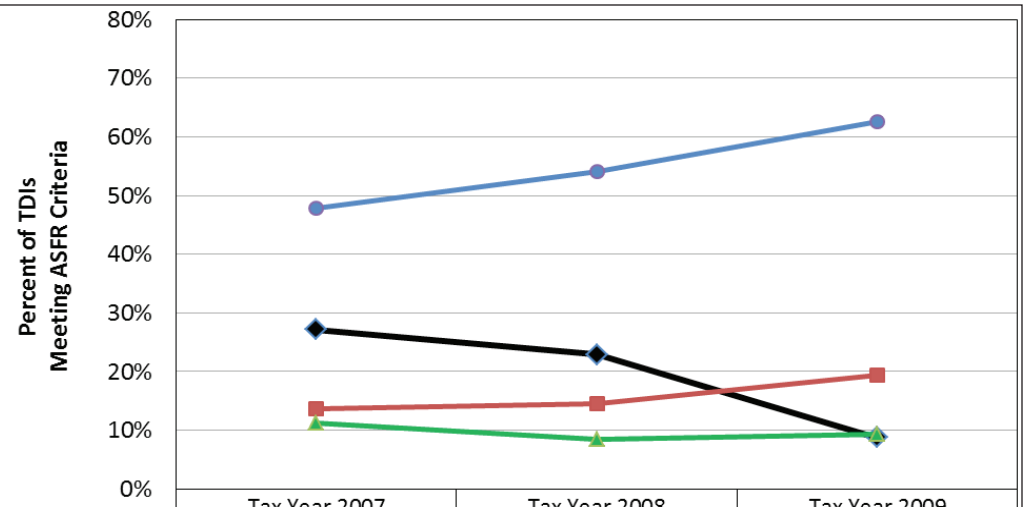
Research Design

Available ASFR Inventory

To develop our models, we identified a set of cases that would be available for ASFR to work. To begin, we identified individual delinquent tax returns for Tax Years 2007, 2008, and 2009 that eventually became TDIs.[12] Using these tax years allows us to capture various changes from year to year, and enough time to evaluate the compliance behavior from TDI status in terms of dollars collected and subsequent filing compliance. Next, we identified the TDIs that met the ASFR criteria.

Figure 3 provides the number of TDIs for the 3 years that met ASFR criteria and the types of treatments received, if any. Of the cases that met ASFR criteria, the percent of inventory being treated by ASFR declined from TY 2007 to TY 2009. As ASFR treats fewer cases, the percent of cases not treated or assigned to ACS and/or Field increased.

FIGURE 3. Treatments and Assignments of Taxpayer Delinquent Investigations Meeting ASFR Criteria, Tax Years 2007-2009



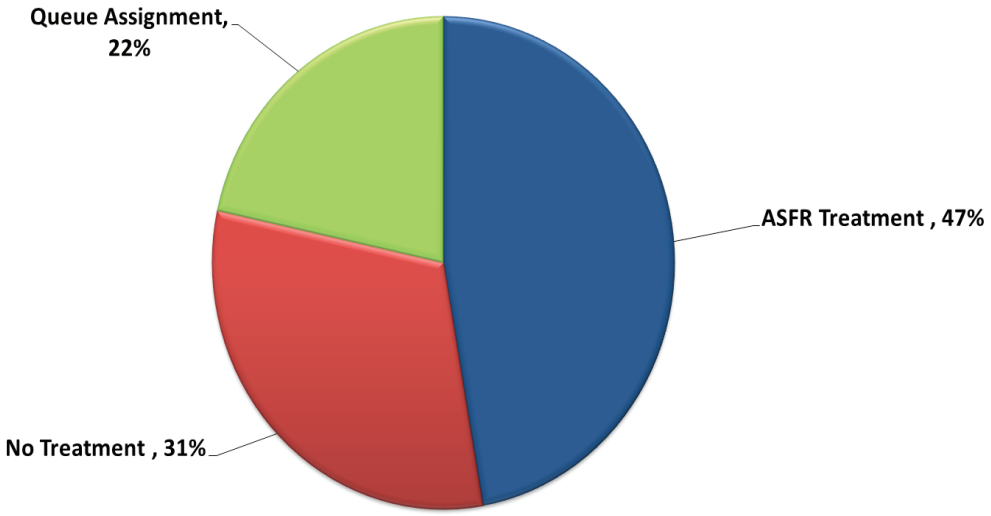
	Tax Year 2007	Tax Year 2008	Tax Year 2009
ASFR Treatment	27%	23%	9%
No Treatment	14%	15%	19%
Queue Assignment	11%	9%	9%
ACS and/or Field Assignment	48%	54%	63%

For this paper, we analyzed the TDIs that met ASFR criteria and were available for ASFR to work. TDIs available for ASFR to work were identified as those not assigned to ACS or Field functions. This left us with a set of TDIs available for ASFR that were either treated by ASFR, given no treatment, or assigned to the Queue.[13] Figure 4 provides the percentage of the available TDIs that fall into each of these three categories for Tax Years 2007–2009. TDIs treated by ASFR accounted for 47 percent of the TDIs available for ASFR over these 3 tax years.

[12] ASFR inventory includes different types of cases that are classified into distinct priority categories based on their observable characteristics. Each year a stratified sample representing each priority category is selected to be worked from this inventory. Furthermore, the ASFR resources have varied widely across years and thus there has been a large variation in the types of cases selected in the ASFR program each year. We use this randomized representation of ASFR inventory to identify ASFR treatment effects.

[13] While these TDI Cases are not assigned randomly, the only factors that influence the likelihood of a case being worked in the ASFR process is a narrow set of case characteristics and available resources. Available resources have varied widely across the years included in this study and are unrelated to taxpayer behavior.

FIGURE 4. Taxpayer Delinquent Investigations Meeting ASFR Criteria and Available for ASFR To Work, Tax Years 2007–2009



Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

We used several IRS databases to identify the TDIs available for ASFR to work.[14] All results in this paper are based on a 10 percent random sample.[15]

Dependent Variables

This research focuses on compliance behavior in terms of both payment and filing compliance.

1. **Dollars Collected.** We modeled the dollars collected related to the delinquent return. We aggregated, to the case level, the dollars collected over the 3 years following TDI status. This provided a consistent time frame for everyone in our research design.

2. **Subsequent Filing Compliance.** We define subsequent voluntary filing compliance as a dichotomous outcome. We assigned a "1" if the taxpayer voluntarily filed a subsequent return and a "0" if not. A voluntarily filed return is defined as one filed by the return's due date (e.g., April 15), or by the requested extension date, without any subsequent delinquent return treatments. To be consistent across all cases in our study, we noted whether the taxpayer voluntarily filed the tax return associated with tax periods 2, 3, or 4 tax years following the delinquent return tax period. For example, for a taxpayer with a TY 2007 delinquent return in our study, we identified if they voluntarily filed a TY 2009, TY 2010, or TY 2011 returns.[16]

For dollars collected, 28 percent of the TDIs treated by ASFR made a payment within 3 years of becoming a TDI. On average, $1,454 dollars were collected per case treated by ASFR. Cases without treatment or assigned to the Queue had a lower percentage of taxpayers with a payment and on average fewer dollars collected compared to cases treated by ASFR. The table below provides a summary of the dollars collected on the TDIs available for ASFR to work for Tax Years 2007–2009.

[14] Data was gathered from Individual Case Creation Nonfiler Identification Process, Individual Master File Status and Transaction History, and Individual Return Transaction File databases stored on the IRS Compliance Data Warehouse.

[15] The ASFR starts in our modeling population that had been previously assigned to other treatment streams, namely Automated Collection System (ACS) and/or Collection Field Function (CFf) accounted for 26 percent. These are indirect assignments to ASFR. This characteristic is suitably controlled in the empirical model.

[16] We did not control for taxpayers not having a filing requirement in the subsequent voluntary filing compliance models. Taxpayers may have had economic circumstances removing their filing requirement.

TABLE 1. Dollars Collected Within Three Years from TDI Status on TDIs Available for ASFR (Tax Years 2007–2009)

Type of Treatment	Percent with a Payment	Average Dollars Collected (all cases)	Cases with a Payment		
			25th Percentile Dollars Collected	50th Percentile Dollars Collected	75th Percentile Dollars Collected
ASFR Treatment	28%	$1,454	$805	$2,147	$4,914
No Treatment	19%	$804	$723	$1,975	$4,150
Queue Assignment	6%	$384	$491	$1,557	$4,996

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

Note that cases were not randomly assigned to these three treatments, so the differences cannot be attributed solely to the treatments.

For subsequent compliance, we chose to look at the second to the fourth years following the delinquent return. This time frame was chosen based on the amount of time it takes the IRS to identify a delinquent return and go through the delinquent return process before becoming a TDI. Take, for example, a TY 2007 return with a return due date of April 15, 2008, for which the taxpayer filed for an extension to October 15, 2008. During this time, the IRS received information returns from third parties providing income and other information about taxpayers. Once any granted extensions have passed, the IRS begins the case creation process of identifying delinquent returns that appear required to be filed based on income reported on those information returns. From this process, the IRS can identify a set of potential nonfilers and select cases to put into the delinquent return notice process. Recall, this notice process can take approximately 14 weeks if the taxpayer does not respond. After that, a portion of these cases will then become TDIs. For a delinquent TY 2007 return, in general, the notice process was ongoing at the time the TY 2008 return was due. Therefore, we structured our study to look at the next TY return due, which was TY 2009, to allow time for the delinquent return process to begin. In addition, we wanted to look at the taxpayer's filing compliance for the subsequent 2 years: TY 2010 and TY 2011.

When analyzing compliance, a taxpayer's past compliance behavior can be a good predictor of his future compliance behavior. Therefore, we looked at how many of the taxpayers in our study had later filed their delinquent return prior to the due date of subsequent tax returns. Of the available TDIs not treated by ASFR, a portion eventually filed their return without any treatment. The table below provides the percentage of cases that filed the delinquent return prior to the second, third or fourth subsequent tax years. For example, of the cases with no treatment, at least 20 percent had filed their delinquent return prior to the due date of the third and fourth subsequent tax years.

TABLE 2. Return Filed After TDI Status on TDIs Available for ASFR Not Treated (Tax Years 2007–2009)

Type of Treatments for TDI Before Due Date of Subsequent Return	Percent of Cases That Filed Return on TDI Before Due Date of Subsequent Return		
	Two Tax Years After TDI	Three Tax Years After TDI	Four Tax Years After TDI
No Treatment	10%	20%	23%
Queue Assignment	4%	7%	8%

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

Note that cases were not randomly assigned to these three treatments, so the differences cannot be attributed solely to the treatments.

Next, we identified subsequent voluntary filing compliance using the same groups identified in Table 2. Table 3 provides the percentage within each group that voluntarily filed a subsequent return. We found that taxpayers who eventually filed their delinquent return for the TDI before the due date of the subsequent return

had a higher rate of voluntary subsequent filing compliance compared to taxpayers treated by ASFR or taxpayers not filing their delinquent return. Of the cases ASFR treated, 39 percent voluntarily filed a subsequent tax return 4 years after the tax year of the TDI.

TABLE 3. Subsequent Voluntary Filing Compliance on TDIs Available for ASFR (Tax Years 2007–2009)

Type of Treatments for TDI Before Due Date of Subsequent Return	Filed Return of the TDI Before Due Date of Subsequent Return	Percent of Cases That Voluntarily Filed a Subsequent Return		
		Two Tax Years After TDI	Three Tax Years After TDI	Four Tax Years After TDI
ASFR Treatment	No	33%	38%	39%
No Treatment	No	37%	42%	42%
	Yes	71%	69%	61%
Queue Assignment	No	22%	25%	26%
	Yes	65%	63%	56%

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

Note that cases were not randomly assigned to these three treatments, so the differences cannot be attributed solely to the treatments.

Independent Variables

Independent variables for our models included a dummy and other controls for ASFR treatment, and a variety of explanatory variables about the taxpayer and delinquent return gathered at the time the delinquent return case was created. To create a nonfiler case, the IRS identifies income and payments from various types of information returns, plus information reported and compliance characteristics on prior tax returns, and then computes a potential balance due.[17]

For modeling both dollars collected and subsequent compliance, we used the dummy for ASFR treatment to capture the average direct effect. To identify indirect effects, we created an instrumental variable to control for the probability the case is worked by ASFR. The instrumental variable was developed using a probit regression with the dependent variable being ASFR treatment. We used tax year as the instrument.

We also included in all the models a variable to control for the number of cycles from TDI status to ASFR treatment.

Empirical Model

We estimate three sets of regressions to assess the impact of ASFR on dollars collected and subsequent filing compliance. First, we estimate the probability of ASFR working a case from available ASFR inventory. Second, we estimate the dollars collected within the next 3 years following TDI status using a linear specification and a Tobit specification.

We then estimate a third set of regressions to assess the impact of ASFR on voluntary subsequent filing compliance 2, 3, and 4 years after the delinquent tax year of the TDI.

Model: Probability of ASFR Selection

Since the cases were not assigned to the groups randomly, the first regression provides an estimate of the probability that a case is selected for ASFR treatment using case characteristics and proxies for ASFR resources/level of treatment. The specification of the regression is as follows:

[17] The potential tax liability is based on the assumption that the taxpayer is single with no itemized deductions, adjustments, or credits.

$$\textit{Model 1: } P(ASFR=1)_i = \Phi(X_{im}\,\beta_m)$$

In Model 1, the probability of ASFR working a case *i* is estimated as a function of a set of *m* exogenous variables, including a taxpayer's observable characteristics obtained from the case creation process, represented as X_{im}. The observable characteristics include various income and income sources of the taxpayer, past tax compliance behavior, and the nature of the taxpayer (Federal employee, small business, etc.), etc. In addition, there are dummies for the tax year of the TDI to serve as proxies for ASFR resources/level of treatment. Recall, this paper looks over multiple tax years, and during the time these tax years were available to treat the resources and levels of ASFR treatment changed as shown in Figure 2. As mentioned in the literature review, prior research has explored changes in compliance behavior as a result of changes in enforcement levels by the tax authority (Plumley (1996); Bloomquist (2004)). By computing the probability of ASFR working a case, regardless of treatment, we are capturing the indirect effect of changes in ASFR's level of enforcement and how that impacts a taxpayer's subsequent compliance behavior perhaps based on the taxpayer's perception of potential ASFR treatment.[18] For example, as a result of the probability or likelihood of ASFR treating a taxpayer who hasn't filed their return declines does the taxpayer's future compliance behavior also decline?

The β coefficients estimate the impact of changes in X on the probability. We obtain these predicted probabilities using a probit regression.

Model: Dollars Collected

We estimate two specifications for the dollars collected. The first, a linear model of dollars collected, is:

$$\textit{Model 2A: } Y_i = \beta_1 ASFR_i + \beta_2 P(ASFR)_i + X_{ik}\beta_k + e_i$$

Y_i indicates dollars collected within 3 years of TDI status for case *i*, *ASFR* is an indicator for ASFR treatment, and *P(ASFR)* is the predicted probability from *Model 1*. X_{ik} is a vector of additional *k* observable taxpayer characteristics over and above the characteristics captured by X_{im}. The model is estimated using Ordinary Least Squares (OLS) regression. We tested for heteroscedasticity using White's Heteroscedasticity test (White, 1980) and used Heteroscedasticity Consistent Covariance Matrix to correct the standard errors. Since this is an OLS equation and the dependent variable is dollars collected, β_1 and β_2 capture the marginal direct and indirect effects of ASFR treatment on dollars collected from delinquent returns in the available inventory.[19]

As shown in Table 1, there are a number of cases in our sample for which the IRS has not received any payments from the delinquent taxpayer once they are in TDI status. This results in left censoring of the payment variable at zero dollars collected. In such a situation, the OLS estimates are inconsistent, the slope is biased upward, and the intercept is biased downward. A Tobit estimate using maximum likelihood estimation is consistent (Amemiya (1973)). Hence, we estimate another variant of Model 2A below using a Tobit regression:

$$\textit{Model 2B: } Y_i = \beta_1 ASFR_i + \beta_2 P(ASFR)_i + X_{ik}\beta_k + e_i$$

Where Y_i is a latent variable: $Y_i^*=0$ if $Y_i \leq 0$ and $Y_i^*=Y_i$ if $Y_i>0$. We use the same exogenous variables as specified in Model 2A. The parameters β_i reflect the marginal impacts of each variable on the latent variable, Y_i^*. The marginal impact on dollars collected is given by:

$$\frac{\partial(Y_i^*)}{\partial x_i} = \beta_i \Phi\left(\frac{(\beta_1 ASFR_1 + \beta_2 P(ASFR)_i + X_{ik}\beta_k)}{\sigma_U}\right)$$

Where x_i is a specific element of the set $[ASFR_i, P(ASFR)_i, X_{ik}]$, $\Phi()$ is the normal distribution function and σ_U is the scale parameter.

[18] Our model assumes taxpayers do not update their expectation of the probability they are selected as new tax years come due, especially when the taxpayer has not yet been selected.

[19] We have controlled for the characteristics used for case selection. We assume there are no other idiosyncratic factors associated with the taxpayer that influence the likelihood of selection. Thus, the random component in the selection process is not related to the error in Model 2 or Model 3. If the random component in the selection process and the errors were in fact correlated with taxpayer behavior, we would need to use the probability of treatment as an instrumental variable for the direct treatment effect.

Model: Subsequent Filing Compliance

The third set of regression models estimate the impact of both direct and indirect effects of ASFR on subsequent voluntary filing compliance for ASFR inventory in subsequent years. We model subsequent voluntary filing compliance for tax period $t+j$ as:

$$\textit{Model 3: } P(File_{t+j})_i = F(\alpha_1 ASFR_i + \alpha_2 P(ASFR)_i + X_{ij}\alpha_j).$$

The variable $File_{t+j}$ represents whether the taxpayer timely filed their $t+j$ tax return, X_{ij} represents case characteristics at time that return is due, and $F()$ is a logistic probability distribution function.[20] Model 3 provides estimates of ASFR direct and indirect effects on subsequent compliance j tax years after a delinquent return. In our case, j =2, 3, and 4.[21] We estimate separate models for each j. This model is estimated using a logistic regression. Since the parameter estimates obtained from a logistic regression are obtained in the form of log-odds ratio, we compute marginal effects of ASFR treatment and the probability of ASFR working a delinquent case on subsequent compliance. These marginal effects are important in testing our hypothesis that ASFR treatment has both direct and indirect effects on subsequent voluntary filing compliance.

Results

In this section we report our regression results of ASFR treatment on dollars collected and on subsequent voluntary filing compliance. Our sample consists of cases that satisfy the requirements of ASFR treatment and were available for ASFR to treat. However, due to resource constraints and prioritization within ASFR, only a portion of those identified cases are worked in ASFR.[22] In the first step, we estimate the probability of ASFR working a case from its available inventory. The probability is estimated based on available observable characteristics of the taxpayer. The results are reported in the Appendix.[23]

Table 4 reports the regression results for the dollars collected model for specifications 2A and 2B. The dependent variable for these regressions is defined as the net dollars and offsets[24] collected on a module for the 3 years starting when the delinquent return enters TDI status. The direct impact of ASFR treatment on dollars collected is captured by the indicator that the module has been worked by ASFR. The coefficient from the OLS[25] model and the marginal effect from Tobit models suggest that a case treated by ASFR, on average, yields $672 and $1,640 more, respectively, than a module that is not treated by ASFR, while keeping other factors fixed.[26] Recall the coefficient on probability of ASFR working a module captures the indirect effects of ASFR. In this case, the coefficient suggests that treating an additional case in ASFR provides an indirect increase in dollars collected for the OLS and Tobit model of $194 and $1,187 respectively,[27] all else equal. These two coefficients provide an estimate of both direct and indirect effects of ASFR working a module based on the representative sample we have used to estimate our models. For dollars collected, the direct effect of ASFR is larger than the indirect effects. Also the Tobit estimates are much larger than the OLS estimates. These marginal effects will be explored later via a simulation of working additional cases for Tax Year 2009.

---

[20] There is potential for future related research in subsequent voluntary filing compliance to control for taxpayers having a filing requirement based on the information reported for the taxpayer. Our definition of subsequent voluntary filing compliance does not account for when a taxpayer is not required to file subsequent tax years.

[21] We do not select $j=1$ due to the lag in the ASFR treatment and the impact of ASFR treatment on subsequent compliance.

[22] While there is some prioritization within ASFR, the only factors that influence the likelihood of a case being worked in the ASFR process is a narrow set of case characteristics and available resources. Available resources have varied widely across the years included in this study and are unrelated to taxpayer behavior.

[23] We consider Fiscal Year dummies instead of Tax Year dummies as instruments in this regression, but the results are insensitive to this choice.

[24] The net dollars and offsets includes any applied payments and offsets minus any reversed payments or offsets; such as misapplied payments, bad checks, etc.

[25] OLS estimates are marginal effects.

[26] We provide both OLS and Tobit model results to show the effect of when you do or do not control for the censoring.

[27] The indirect effect for a specific case is $\beta_2 * 1/N$. Since the indirect effect impacts all cases, the aggregate indirect effect of an additional case is β_2.

TABLE 4. Models 2A & 2B: Expected Dollars Collected Three Years from TDI Assignment

Dependent Variable: Dollars Collected Three Years from TDI Assignment Explanatory Variables	Model 2A: OLS Estimates	Model 2B: Tobit	
		Estimates	Marginal Effect
Indicator of ASFR Treatment (ASFR)	$672.44 (14.60)***	$11,385.00 (182.15)***	$1,639.59
Predicted Probability of ASFR Working a Case (P(ASFR))	$193.92 (37.32)***	$8,241.28 (340.38)***	$1,186.86
Number of Cycles to ASFR Treatment (30-day letter issued)	-$6.55 (0.24)***	-$103.76 (3.26)***	-$14.94

Source: Internal Revenue Service Individual Master File Status and Transaction History, and Case Creation Nonfiler Identification Process. Data extracted February 2015.
Notes: Not all explanatory variables shown. See Appendix.

N = 277,314

*p<0.1; **p<0.05; ***p<0.01; Standard errors reported in parentheses; the standard errors for the OLS model are Heteroscedasticity Consistent Standard Errors. Marginal Effects are calculated at the sample means.

The estimated parameter for the number of cycles to the 30-day letter (the start of the ASFR process) is negative and significant in both models. One interpretation of this may be that the sooner the ASFR process is started after TDI status, the greater the amount of unpaid tax collected within the 3-year time frame. Thus, starting the ASFR process early will result in more dollars collected. While it seems intuitive that this would be the case, it may not be the appropriate conclusion with the model as specified. Due to data limitations, we cannot track all ASFR cases for a consistent amount of time between the ASFR case closing and any unpaid tax being assessed. Thus, cases that are started later have a shorter window of time for us to observe payments following ASFR treatment. The "cycle-to-30-day letter" measure is controlling for the varied amount of time following treatment to observe payments on each case worked by ASFR. The other control variables in the regression are significant. They are reported in the Appendix.

The individual t-statistics of each parameter estimate of these regressions are significant at a 5 percent or lower level of significance in a two-tailed t-test. The overall F statistics of the OLS and Log-likelihood ratios for the Tobit models are found to be significant. For the OLS specification, we perform the White test of heteroscedasticity on the residuals obtained from this regression. We reject the null hypothesis of homoscedasticity and correct the standard errors of the estimates using heteroscedasticity consistent standard errors (White, 1980). This correction results in consistent and efficient parameter estimates. We report the heteroscedasticity consistent standard errors in our results.

The next aspect of this paper is to look at subsequent compliance 2, 3, and 4 tax years after the ASFR treatment. The main objective is to estimate the impact of the ASFR treatment, both directly and indirectly, on subsequent voluntary filing compliance. The estimates obtained from the three logistic regressions are expressed in terms of log-odds ratios, which are not very intuitive for interpretation purposes. Therefore, we estimate marginal effects at the sample means for interpretational convenience. The results are reported in Table 5.

TABLE 5. Model 3: Voluntarily Filing a Return Two, Three, and Four Years from TDI Assignment

Dependent Variable: Taxpayer Voluntarily Filed a Tax Return *'j'* Tax Years Later; *j*=2,3 and 4	Two Tax Years After		Three Tax Years After		Four Tax Years After	
Explanatory Variables	Coeffi-cients	Marginal Effects	Coeffi-cients	Marginal Effects	Coeffi-cients	Marginal Effects
Indicator of ASFR Treatment	0.42 (0.02)***	0.09	0.25 (0.01)***	0.06	0.18 (0.01)***	0.04
Predicted Probability of ASFR Working a Case	0.51 (0.02)***	0.11	0.89 (0.03)***	0.21	1.14 (0.03)***	0.27
Number of Cycles to ASFR Treat-ment (30-day letter issued)	-0.01 (0.001)***	-0.002	-0.004 (0.0003)***	-0.001	-0.003 (0.0002)***	-0.0006
Self-correction: Taxpayer Filed Return on TDI Prior to Due Date of Tax Return *j*	1.55 (0.02)***	0.35	1.32 (0.02)***	0.31	1.01 (0.02)***	0.24

Source: Internal Revenue Service Individual Master File Status and Transaction History, and Case Creation Nonfiler Identification Process. Data extracted February 2015.

Notes: Not all explanatory variables shown. See Appendix.

*p<0.1; **p<0.05; ***p<0.01; Marginal Effects are calculated at the sample means.

Based on the estimates reported in Table 5, the direct impact of the ASFR treatment on voluntarily filing a return 2, 3, and 4 tax years after ASFR treatment is positive. The likelihood of filing increases by 9, 6, and 4 percentage points, respectively, keeping other variables fixed. This direct effect appears to be decreasing over time, which seems reasonable. The indirect effects are also positive but larger than the direct effects: 11 percent, 21 percent, and 27 percent, respectively for this period. This is consistent with the hypothesis that the indirect effect is stronger than the direct effect and has lasting effect over the subsequent years. The marginal impact of "cycle to 30 days" declines over time.

The indicator for self-correction is defined as the taxpayer who voluntarily filed his tax return after ASFR treatment but before the next tax return was due. It is a measure that captures the "willingness" of the taxpayer to be compliant post-ASFR treatment. We find a positive and significant impact on future voluntary filing compliance. The marginal impact is found to be strong and ranges from 24 percent to 35 percent during the period of 2 to 4 years. The impact of other control variables on subsequent voluntary filing compliance is reported in the Appendix.

Simulation

It seems intuitive that working more ASFR cases will result in increased tax revenue and more voluntarily filed returns. Both results come from the direct impact of IRS enforcing filing requirements, and indirect effects of the level of enforcement. We use our estimated models of payment and subsequent filing to simulate the impacts of working more of the available nonfiler cases in ASFR.

Using the Tax Year 2009 cases in our study, we estimate the increase in dollars collected and the increase in returns subsequently filed voluntarily in the simulated counterfactual scenario of ASFR working an additional 100,000 cases from the available inventory. To create our counterfactual data (i.e., ASFR working 100,000 more cases), we replicate the Tax Year 2009 data and randomly designate 100,000 of the unworked cases as being treated by ASFR.[28] We assume that the initial ASFR letter was issued immediately; thus the number of cycles to 30-day letter would remain as zero. In order to include the indirect effects, P(ASFR) in the subsequent filing model is increased to reflect the increase in the proportion of available inventory worked. Working 100,000

[28] We selected a random sample of unworked cases to be conservative in our estimates compared to using the ASFR prioritization criteria to select the next 100,000 unworked cases.

more cases for Tax Year 2009 would increase the proportion of available inventory worked by 0.13. Thus, 0.13 is added to the P(ASFR) for each of the cases in the counterfactual scenario, but it is constrained to be no more than one.[29]

Let

$E(P_{ai})$ be the predicted dollars collected for taxpayer *i* based on the actual data;

$E(P_{ci})$ be the predicted dollars collected for taxpayer *i* based on the counterfactual data;

$P(R_{ai})$ be the predicted probability of filing a return for 2011 based on the actual data; and

$P(R_{ci})$ be the predicted probability of filing a return for 2011 based on the counterfactual data.

For the OLS model of payments the expected payments are calculated as

$$E(P) = X\hat{\beta}.$$

For the Tobit model of payments the expected payments are calculated as

$$E(P) = (X\hat{\beta} + \sigma\lambda)\Phi\left(\frac{X\beta}{\sigma}\right).$$

Then, the estimated increase in payments would be

$$\text{Increase in payments} = \sum_{\forall i}(E(P_{ci}) - E(P_{ai})).$$

The increase in returns subsequently filed would be

$$\text{Increase in returns} = \sum_{\forall i}(P(R_{ci}) - P(R_{ai})).$$

The results of the simulated increase in enforcement are reported in Table 6. These estimates are to some degree conservative because we do not estimate with our analysis the indirect effects on the cases that were never in a post notice delinquent return status. It is reasonable to expect that increased enforcement might increase the cases responding to delinquent return notices or may choose to file their return timely. That impact is beyond the scope of this paper. Also, we are assuming that the additional cases are being selected randomly. To a degree, the IRS can prioritize the "best" case to select; therefore, the increases in dollars and/or returns would be somewhat larger.

TABLE 6. Simulated Total Impact of Working 100,000 More ASFR Cases for Tax Year 2009

Model	Total Increase	Increase Per ASFR Case Started
Increase in Payments (Linear Model)	$118,077,994	$1,181
Increase in Payments (Tobit Model)	$326,192,842	$3,262
Increase in Voluntarily Filed Returns in 2011	19,469	0.19
Increase in Voluntarily Filed Returns in 2012	24,563	0.25
Increase in Voluntarily Filed Returns in 2013	29,166	0.29

Based on data reported in internal CFO cost accounting, the cost per case closed varied between $53 and $80 per case for FY 2009 to FY 2013.[30] Using the $80 as an estimated cost per case, revenue collected relative to the cost would be just under 15:1 using the linear model estimate and just over 40:1 using the Tobit model

[29] No case had a probability high enough to invoke the constraint.

[30] Since ASFR is an automated system, the cost per working an ASFR case remains fairly stable for direct ASFR assignments.

estimate. In addition, every $110 spent on the ASFR program results in an additional voluntarily filed return ($80/(.19+.25+.29)), based on our model estimates.

Conclusions and Direction for Further Research

In this paper, we develop a model of taxpayer payments of unpaid taxes associated with delinquent returns and the subsequent decision to file future tax returns timely. We focus specifically on the impact of IRS enforcement via the ASFR process. Our model provides estimates of both the direct effects of the IRS enforcement and the indirect effects of that enforcement.

Our estimates suggest significant direct and indirect impacts of enforcing filing compliance via the ASFR process, for both payment and subsequent filing compliance. The indirect effects are somewhat smaller than the direct effects for payment of taxes on delinquent returns. However, the indirect effects on subsequent filing compliance are large relative to the direct effects. In addition, the relative magnitude of the estimates is similar to effects reported in other studies of an audit's effect on reporting compliance. While we do not examine data around cost of an ASFR case, cost estimates from other studies compared to our model estimates suggest that the return on investment is relatively high.

Based on our estimates, it is clear that, given the downward trend in the number of ASFR cases worked, there has been and will be significant declines in enforcement revenue from the ASFR program and decreases in the number of returns voluntarily filed. Clearly, the IRS is devoting fewer resources to the ASFR program. Some of this decline may be the result of shifting resources to other programs that may be more productive or important, and thus would actually result in overall improvement in accomplishing the goals of tax administration. However, it is likely that much of the decline is the result of the IRS response to decreasing real budgets and increasing responsibilities. It may well be that the decreases in the ASFR starts were part of an optimal strategy to absorb the budget shocks. Our estimates suggest the ASFR resource reductions result in both significant decreases in enforcement revenue and reduced voluntary filing compliance.

This paper is a first attempt to estimate the impacts of nonfiling enforcement on the gross and net tax gap. With more years and a broader set of data, this study could be enhanced in many ways. We make some fairly restrictive assumptions about how taxpayers form their expectations of the "likelihood" that IRS will start an ASFR case. Our model assumes taxpayers do not update their expectation of the probability they are selected as new tax years come due, especially when the taxpayer has not yet been selected. Our research could be extended to consider all the nonfiler treatment streams and impact on all taxpayers, including those who have always filed timely or at least have always resolved in the notice process. This would, however, dramatically increase scope and complexity of the analysis.

References

Allingham, M. G., and Sandmo, A. (1972). "Income tax evasion: A theoretical analysis." *Journal of Public Economics,* 1:3–4, 323–38.

Alm, J., and Torgler, B. (2011). "Do ethics matter? Tax compliance and morality." *Journal of Business Ethics,*101(4), 635–651.

Amemiya, T. (1973). "Regression analysis when the dependent variable is truncated normal." *Econometrica: Journal of the Econometric Society,* 997–1016.

Andreoni, J., Erard, B., and Feinstein, J. (1998). "Tax compliance." *Journal of Economic Literature,* 818–860.

Coase, R. H. (1960). "The problem of social cost." *Journal of Law & Economics,* 3, 1.

Erard, B., and Ho, C.(2001). "Searching for ghosts: Who are the nonfilers and how much tax do they owe?"*Journal of Public Economics,* 81(1), 25–50.

Greene, W. H. (2008). *Econometric Analysis.* Granite Hill Publishers.

Hashimzade, N., Myles, G. D., and Tran-Nam, B. (2013). "Applications of behavioural economics to tax evasion." *Journal of Economic Surveys,* 27(5), 941–977.

Hurst, E., Li, G., and Pugsley, B. (2014). "Are household surveys like tax forms? Evidence from income underreporting of the self-employed." *Review of Economics and Statistics,* 96(1), 19–33.

Kleven, H. J., Knudsen, M. B., Kreiner, C.T., Pedersen, S., and Saez, E. (2011). "Unwilling or unable to cheat? Evidence from a randomized tax audit experiment in Denmark." *Econometrica,* 79(3), 651–692.

Lederman, L., and Sichelman, T. M. (2013). "Enforcement as substance in tax compliance." *Washington and Lee Law Review,* 70, 1679–1749.

Ljungqvist, L., and Sargent, T. J. (2000). *Recursive Macroeconomic Theory.* Granite Hill Publishers.

Milgrom, P., and Roberts, J. (1992). *Economics, Organization and Management.* Engelwood Cliffs, NJ: Prentice Hall.

IRS Office of the Chief Financial Officer, Financial Management, Office of Cost Accounting Cost-Based Performance Measures Automated Substitute for Return (ASFR) FY2009—FY2013, Unpublished internal CFO document, 2014.

Plumley, A. H. (1996). *The determinants of individual income tax compliance: Estimating the impacts of tax policy, enforcement, and IRS responsiveness.* Internal Revenue Service. Publication 1916 (Rev. 11–96).

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: Journal of the Econometric Society,* 817–838.

Appendix

Complete Model Results

TABLE A1. Probability of ASFR Working a Case, Probit Model

Explanatory Variables*	Parameter Estimate	Standard Error	Wald Statistic	P-value
Intercept	-0.955	0.024	1546.944	<.0001
Log of wages reported on Form W-2	0.006	0.001	55.396	<.0001
Log of Non-Employment Compensation reported on Form 1099-MISC	0.020	0.001	451.147	<.0001
Log of the balance due calculated for a potential Substitute for Return (SFR) return.**	0.051	0.003	336.260	<.0001
Indicator = 1 if the delinquent return is for Tax Year 2007	0.796	0.007	14731.298	<.0001
Indicator = 1 if the delinquent return is for Tax Year 2008	0.750	0.006	13439.036	<.0001

n=277,314

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

*Some explanatory variables have been suppressed.

** To SFR Potential Tax Assessment (IRPSFRTX), add the sum of Advance Earned Income Credit (AEIC), computed tax on premature distributions, and Self-Employment (SE) tax, less the sum of current year credit balance, excess FICA, and withholding. It is set to zero if the result is a negative.

TABLE A2. Predicted Dollars Collected, OLS Regression

Explanatory Variables	Parameter Estimate	Std. Error	t Value	P-value	Heteroscedasticity Consistent Statistics			Tolerance	Variance Inflation
					Std. Error	t Value	Pr > |t|		
Intercept	-8759.950	156.695	-55.900	<.0001	534.600	-16.39	<.0001		0
Indicator = 1 if ASFR treated the case within 3 years of TDI Assignment	841.834	51.601	16.310	<.0001	48.776	17.26	<.0001	0.39823	2.51114
Probability of ASFR working the case	331.310	94.200	3.520	0.0004	120.122	2.76	0.0058	0.7405	1.35044
For ASFR treated cases, this is the number of weeks from TDI Assignment to ASFR 30-day letter; else = 0	-8.589	0.903	-9.520	<.0001	0.875	-9.82	<.0001	0.43236	2.31287
Indicator = 1 if the taxpayer filed for an extension along with a payment.	1792.805	102.999	17.410	<.0001	302.685	5.92	<.0001	0.96361	1.03776
Indicator = 1 if Federal employee	293.454	70.985	4.130	<.0001	50.888	5.77	<.0001	0.89953	1.11169
Indicator = 1 if the taxpayer had either interest reported on Form 1099-INT or dividends reported on Form 1099-DIV	525.954	41.646	12.630	<.0001	29.672	17.73	<.0001	0.66814	1.49669
Indicator = 1 if the taxpayer had pensions or annuity benefits reported on Form 1099-R	-226.482	40.306	-5.620	<.0001	40.841	-5.55	<.0001	0.7708	1.29736
Indicator = 1 if the taxpayer had unemployment benefits reported on Form 1099-G	-173.513	71.438	-2.430	0.0151	25.913	-6.7	<.0001	0.94645	1.05658
Indicator = 1 if the taxpayer had any withholding reported on information returns received by the IRS.	330.025	55.980	5.900	<.0001	74.708	4.42	<.0001	0.38716	2.58291
Indicator = 1 if the taxpayer reported a Married Filing Joint Filing Status on their prior-year return.	250.959	44.370	5.660	<.0001	62.702	4	<.0001	0.86498	1.1561
Indicator = 1 if the taxpayer had a filing requirement code of not required to file therefore indicating no tax form package be mailed to the taxpayer.	-276.580	43.919	-6.300	<.0001	33.924	-8.15	<.0001	0.83508	1.19749
Log of the balance due calculated for a potential Substitute for Return (SFR) return.*	962.347	18.423	52.240	<.0001	62.981	15.28	<.0001	0.76381	1.30923
Log of wages reported on Form W-2	21.747	5.115	4.250	<.0001	7.347	2.96	0.0031	0.39261	2.54706
Indicator = 1 if the taxpayer had more than two Forms W-2	-536.156	67.825	-7.910	<.0001	54.341	-9.87	<.0001	0.88362	1.13171
Log of the number of information returns received for the taxpayer.	450.948	26.162	17.240	<.0001	35.342	12.76	<.0001	0.54001	1.85181
Log of the total positive income reported on the prior-year return filed.	30.122	3.617	8.330	<.0001	5.444	5.53	<.0001	0.76085	1.31431

n=277,314

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

* To SFR Potential Tax Assessment (IRPSFRTX), add the sum of Advance Earned Income Credit (AEIC), computed tax on premature distributions, and Self-Employment (SE) tax, less the sum of current year credit balance, excess FICA, and withholding. It is set to zero if the result is a negative.

TABLE A3. Predicted Dollars Collected, Tobit Regression Censored at Zero

Explanatory Variable	Parameter Estimate	Standard Error	t Value	P-value
Intercept		556.250	-97.500	<.0001
Indicator = 1 if ASFR treated the case within 3 years of TDI Assignment	11385.000	182.148	62.500	<.0001
Probability of ASFR working the case	8241.278	340.377	24.210	<.0001
For ASFR treated cases, this is the number of weeks from TDI Assignment to ASFR 30-day letter; else = 0	-103.764	3.258	-31.850	<.0001
Indicator = 1 if the taxpayer filed for an extension along with a payment.	5883.410	296.835	19.820	<.0001
Indicator = 1 if Federal employee	1968.786	221.887	8.870	<.0001
Indicator = 1 if the taxpayer had either interest reported on Form 1099-INT or dividends reported on Form 1099-DIV	2935.717	142.671	20.580	<.0001
Indicator = 1 if the taxpayer had pensions or annuity benefits reported on Form 1099-R	1590.049	137.725	11.550	<.0001
Indicator = 1 if the taxpayer had unemployment benefits reported on Form 1099-G	-3073.823	269.743	-11.400	<.0001
Indicator = 1 if the taxpayer had any withholding reported on information returns received by the IRS.	1998.573	195.552	10.220	<.0001
Indicator = 1 if the taxpayer reported a Married Filing Joint Filing Status on their prior-year return.	1365.964	144.111	9.480	<.0001
Indicator = 1 if the taxpayer had a filing requirement code of not required to file therefore indicating no tax form package be mailed to the taxpayer.	-3657.524	175.781	-20.810	<.0001
Log of the balance due calculated for a potential Substitute for Return (SFR) return.*	1902.135	62.939	30.220	<.0001
Log of wages reported on Form W-2	143.076	17.401	8.220	<.0001
Indicator = 1 if the taxpayer had more than two Forms W-2	-4130.596	253.805	-16.270	<.0001
Log of the number of information returns received for the taxpayer.	2306.890	91.902	25.100	<.0001
Log of the total positive income reported on the prior-year return filed.	669.091	12.431	53.830	<.0001

n=277,314

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

* To SFR Potential Tax Assessment (IRPSFRTX), add the sum of Advance Earned Income Credit (AEIC), computed tax on premature distributions, and Self-Employment (SE) tax, less the sum of current year credit balance, excess FICA, and withholding. It is set to zero if the result is a negative.

TABLE A4. Probability of Voluntarily Filing a Subsequent Tax Return Two Tax Years Later, Logistic Regression

Explanatory Variables	Parameter Estimate	Standard Error	Wald Statistic	P-value
Intercept	-1.167	0.043	750.824	<.0001
Probability of ASFR working the case	0.512	0.025	429.598	<.0001
Indicator = 1 if ASFR treated the case prior to the due date of the tax return 2 years later.	0.420	0.019	484.754	<.0001
For ASFR treated cases, this is the number of weeks from TDI Assignment to ASFR 30-day letter; else = 0	-0.009	0.001	230.103	<.0001
Indicator = 1 for non-ASFR treated TDIs if the taxpayer filed their return before the due date of the return due 2 tax years later.	1.554	0.019	6978.168	<.0001
Indicator = 1 if the case is classified as SB/SE	0.144	0.009	256.123	<.0001
Indicator = 1 if the taxpayer filed for an extension on the unfiled return.	0.138	0.011	155.763	<.0001
Indicator = 1 if the taxpayer had interest reported on Form 1099-INT	0.118	0.009	157.036	<.0001
Indicator = 1 if the taxpayer had pensions or annuity benefits reported on Form 1099-R	0.062	0.010	40.752	<.0001
Indicator = 1 if the taxpayer had unemployment benefits reported on Form 1099-G	-0.113	0.018	38.399	<.0001
Indicator = 1 if the taxpayer had any withholding reported on information returns received by the IRS.	0.106	0.014	57.165	<.0001
Indicator = 1 if the taxpayer reported a Married Filing Joint Filing Status on their prior-year return.	0.058	0.011	27.618	<.0001
Indicator = 1 if the taxpayer had wages reported on a Form W-2	0.380	0.014	777.412	<.0001
Indicator = 1 if the taxpayer had a delinquent prior-year tax module that was closed and received a TC59x.	0.860	0.016	3081.011	<.0001
Indicator = 1 if the taxpayer had a zero-dollar tax assessment on their prior-year tax module indicating either no tax due or a Substitute for Return (SFR).	-0.209	0.014	233.139	<.0001
Log of the balance due calculated for a potential SFR return.*	-0.026	0.005	31.367	<.0001
Number of years from the delinquent return and the taxpayer's filed prior-year tax module.	-0.034	0.001	867.929	<.0001
Log of the total positive income reported on the prior-year return filed.	0.027	0.001	583.954	<.0001

n=277,314

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

* To calculate SFR Potential Tax Assessment (IRPSFRTX), add the sum of Advance Earned Income Credit (AEIC), computed tax on premature distributions, and Self-Employment (SE) tax, less the sum of current year credit balance, excess FICA, and withholding. It is set to zero if the result is a negative.

TABLE A5. Probability of Voluntarily Filing a Subsequent Tax Return Three Tax Years Later, Logistic Regression

Explanatory Variables	Parameter Estimate	Standard Error	Wald Statistic	P-value
Intercept	-1.592	0.042	1466.150	<.0001
Probability of ASFR working the case	0.893	0.025	1269.457	<.0001
Indicator = 1 if ASFR treated the case prior to the due date of the tax return 3 years later.	0.251	0.014	308.130	<.0001
For ASFR treated cases, this is the number of weeks from TDI Assignment to ASFR 30-day letter; else = 0	1.323	0.016	6574.608	<.0001
Indicator = 1 for non-ASFR treated TDIs if the taxpayer filed their return before the due date of the return due 3 tax years later.	-0.004	0.000	230.569	<.0001
Indicator = 1 if the case is classified as SB/SE	0.180	0.009	418.407	<.0001
Indicator = 1 if the taxpayer filed for an extension on the unfiled return.	0.186	0.011	297.219	<.0001
Indicator = 1 if the taxpayer had interest reported on Form 1099-INT	0.083	0.009	80.643	<.0001
Indicator = 1 if the taxpayer had pensions or annuity benefits reported on Form 1099-R	0.024	0.010	6.411	0.0113
Indicator = 1 if the taxpayer had unemployment benefits reported on Form 1099-G	-0.140	0.018	62.174	<.0001
Indicator = 1 if the taxpayer had any withholding reported on information returns received by the IRS.	0.115	0.014	70.395	<.0001
Indicator = 1 if the taxpayer reported a Married Filing Joint Filing Status on their prior-year return.	0.069	0.011	40.971	<.0001
Indicator = 1 if the taxpayer had wages reported on a Form W-2	0.465	0.013	1221.670	<.0001
Indicator = 1 if the taxpayer had a delinquent prior-year tax module that was closed and received a TC59x.	0.813	0.015	2816.524	<.0001
Indicator = 1 if the taxpayer had a zero-dollar tax assessment on their prior-year tax module indicating either no tax due or a Substitute for Return (SFR).	-0.212	0.014	246.543	<.0001
Log of the balance due calculated for a potential SFR return.*	0.017	0.004	13.659	0.0002
Number of years from the delinquent return and the taxpayer's filed prior-year tax module.	-0.027	0.001	602.749	<.0001
Log of the total positive income reported on the prior-year return filed.	0.025	0.001	526.521	<.0001

n=277,314

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

* To calculate SFR Potential Tax Assessment (IRPSFRTX), add the sum of Advance Earned Income Credit (AEIC), computed tax on premature distributions, and Self-Employment (SE) tax, less the sum of current year credit balance, excess FICA, and withholding. It is set to zero if the result is a negative.

TABLE A6. Probability of Voluntarily Filing a Subsequent Tax Return Four Tax Years Later, Logistic Regression

Explanatory Variables	Parameter Estimate	Standard Error	Wald Statistic	P-value
Intercept	-1.167	0.043	750.824	<.0001
Probability of ASFR working the case	0.512	0.025	429.598	<.0001
Indicator = 1 if ASFR treated the case prior to the due date of the tax return 4 years later.	0.420	0.019	484.754	<.0001
For ASFR treated cases, this is the number of weeks from TDI Assignment to ASFR 30-day letter; else = 0	-0.009	0.001	230.103	<.0001
Indicator = 1 for non-ASFR treated TDIs if the taxpayer filed their return before the due date of the return due 4 tax years later.	1.554	0.019	6978.168	<.0001
Indicator = 1 if the case is classified as SB/SE	0.144	0.009	256.123	<.0001
Indicator = 1 if the taxpayer filed for an extension on the unfiled return.	0.138	0.011	155.763	<.0001
Indicator = 1 if the taxpayer had interest reported on Form 1099-INT	0.118	0.009	157.036	<.0001
Indicator = 1 if the taxpayer had unemployment benefits reported on Form 1099-G	-0.113	0.018	38.399	<.0001
Indicator = 1 if the taxpayer had any withholding reported on information returns received by the IRS.	0.106	0.014	57.165	<.0001
Indicator = 1 if the taxpayer reported a Married Filing Joint Filing Status on their prior-year return.	0.058	0.011	27.618	<.0001
Indicator = 1 if the taxpayer had wages reported on a Form W-2	0.380	0.014	777.412	<.0001
Indicator = 1 if the taxpayer had a delinquent prior-year tax module that was closed and received a TC59x.	0.860	0.016	3081.011	<.0001
Indicator = 1 if the taxpayer had a zero-dollar tax assessment on their prior-year tax module indicating either no tax due or a Substitute for Return (SFR).	-0.209	0.014	233.139	<.0001
Log of the balance due calculated for a potential SFR return.*	-0.026	0.005	31.367	<.0001
Number of years from the delinquent return and the taxpayer's filed prior-year tax module.	-0.034	0.001	867.929	<.0001
Log of the total positive income reported on the prior-year return filed.	0.027	0.001	583.954	<.0001

n=277,314

Source: IRS, Compliance Data Warehouse Individual Master File Status and Transaction History, and Individual Case Creation as of February 2015 (cycle 201508)

* To calculate SFR Potential Tax Assessment (IRPSFRTX), add the sum of Advance Earned Income Credit (AEIC), computed tax on premature distributions, and Self-Employment (SE) tax, less the sum of current year credit balance, excess FICA, and withholding. It is set to zero if the result is a negative.