Estimating the Specific Indirect Effect for Multiple Types of Correspondence Audit

Ben Howard, Lucia Lykke, and Leigh Nicholl (The MITRE Corporation),
and Alan Plumley (IRS, RAAS)

Introduction

Tax enforcement actions have a direct revenue effect: the tax collected from (or refunded to) the contacted taxpayer pertaining to the year that was the subject of the contact. These enforcement actions undoubtedly also have an indirect effect on revenues: a change in the current or future behavior of taxpayers who either have experienced an enforcement contact themselves (the "specific" indirect effect) or have some knowledge or perception about others' tax enforcement experience (the "general" indirect effect). This study seeks to estimate and compare the magnitudes of the specific effects on taxpayers following one of three types of correspondence audit, using longitudinal taxpayer data obtained by the United States Internal Revenue Service (IRS) through operational audits conducted on tax returns filed within the Tax Year 2006 through 2012 period. We study the effects of correspondence audits that examine three different types of taxpayers: those who were audited based on business expenses, itemized deductions, and self-employment tax, respectively. In each case, we compare the subsequent-year reporting on total tax and audit-specific line items between the audited group and a not-audited taxpayer "control" group who were otherwise eligible for the audit according to all operational eligibility criteria. In this way, we advance prior literature by estimating the specific indirect effect for three categories of audit, examining taxpayers who are designated for audit eligibility by operational criteria that vary by the audit category.

Comparing the subsequent reporting behavior of taxpayers who experienced different types of audits has both research and operational value. Much of the prior literature on the indirect effect of audit has focused on taxpayers who are self-employed (e.g., Beer (2015); DeBacker *et al.* (2015)), finding that these taxpayers increase reporting on measures such as taxable income following an audit, and the effect is more pronounced compared to taxpayers whose income is primarily subject to third-party reporting (DeBacker *et al.* (2015); Kleven *et al.* (2011)). The key point from these studies is that taxpayers – including audited taxpayers – are not a homogenous group, and therefore have different underlying characteristics and may also respond differently to different types of audits. However, little exploration has been done between different types of taxpayers using operational data and selection criteria. Further, the selection mechanism that drives whether or not a taxpayer is audited varies between types of audit.

Additionally, knowing whether and how different types of audits yield different specific indirect effects may help to inform IRS resource allocation decisions. As the Government Accountability Office (GAO) recently pointed out (GAO (2012)), different types of audits yield different direct revenue benefit/cost. The GAO called for greater knowledge of the indirect effect of different audit types in order to understand whether increasing or decreasing audit coverage in various audit categories would result in a long-term increase or decrease in the overall revenue generated

from taxpayers' voluntary reporting compliance. For these reasons, we explore the differential specific indirect effects of three different types of correspondence audits.

However, empirically observing these indirect effects is challenging. Operational audits, unlike research audits such as those conducted under the National Research Program (NRP), are not randomly distributed among the taxpayer population. This fact poses major challenges for causal inference (Kleven *et al.* (2011); Mazzolini *et al.* (2017)). Although we are unable to completely account for such endogeneity in this study, we advance existing knowledge by specifically controlling for the specific operational metrics applied to each return to determine eligibility for audit within each category *and* the priority given to it among all eligible returns in that category. That means that our control group was not drawn from the overall population of unaudited returns, but only from the much smaller sub-population of returns that met all operational eligibility criteria. And, being able to apply the specific method used operationally in each category to prioritize the returns in the eligible pool, we further controlled for this ranking, representing an advance over prior studies that have not had access to such information.

As such, the following are general research objectives that guide this study:

1. Assess whether there is an observable change in taxpayers' individual contributions to IRS revenue, as defined by total tax reporting, in the years subsequent to experiencing a correspondence audit. We do this by comparing audited taxpayers' post-audit tax reporting to the tax reporting of unaudited taxpayers who were eligible during the same Tax Year.
2. Assess whether reporting on other relevant items, specifically on the line items being examined in the different types of correspondence audit, changes in the years after a taxpayer experiences an audit when compared to the reporting of similarly eligible, but ultimately unaudited, taxpayers.
3. Explore potential differences in post-audit reporting behavior across three distinct categories of correspondence audits, each of which is associated with a different underlying population of U.S. taxpayers subject to different audit selection criteria.

Literature Review

Types of 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. Tax evasion may take the form of not filing or misreporting income or other information (such as deductions) on tax returns, and compliance refers to the behaviors of filing tax returns on time, accurately reporting information on tax returns, and paying taxes owed on time (Hallsworth (2014)). 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, an audit) will affect that taxpayer's future probability of compliance, an effect referred to as "specific deterrence" (Slemrod (2016)) or, more generally, as the specific indirect effect. Although an audit may result in immediate funds collected from a noncompliant taxpayer (a direct effect of the audit), that

taxpayer will likely pay taxes for many years to come and therefore the audit may continue to affect taxes paid by that taxpayer in subsequent years. This specific indirect effect is the focus of this study.

Additionally, taxpayers may also have awareness of enforcement activities experienced by other taxpayers that affects their perception of the risk of noncompliance. This secondhand effect of enforcement activities is known as the "general indirect effect"[1] (Plumley (1996); see also Slemrod (2016)). Several studies have found that audit rates at the aggregate level are positively associated with greater tax compliance (Ali *et al.* (2001); Dubin *et al.* (1990); Plumley (1996)). However, field experiments have resulted in mixed findings. There is some evidence that information about the threat of audit for businesses travels through tax preparer networks and corporate relationships (i.e. between parent and subsidiary companies) (Boning *et al.* (2018)). However, studies of the general indirect effect within neighborhoods have not found evidence that neighbors' tax-related experiences spill over to each other (Meiselman (2018)). A full review of evidence for general indirect effect is outside the scope of this study.

Evidence for Specific Indirect Effect

Compliance is, in most cases, impossible to observe because in the absence of a repeat audit, it is difficult to know whether a taxpayer's reporting was accurate. This may be especially true for taxpayers who report self-employment income that is not subject to third-party reporting. As such, most studies of the specific effect examine trends in reporting proxy measures, including income, tax liability, or specific deductions or adjustments. Several themes from this research are relevant to this study: (1) the use of operational versus research audits; (2) the observation of specific effects among the self-employed; and (3) the attenuation of specific indirect effects over time.

A major challenge for the study of indirect effects of enforcement activities is the fact that taxpayers are not usually selected randomly into the "treatment" of being audited. Several countries, including the U.S., conduct randomly assigned research audits, which might be used to circumvent this selection bias problem; however, if taxpayers know that they are audited randomly for research purposes, this may introduce a validity issue insofar as taxpayers may respond to a random audit differently from an operational audit (Slemrod (2016)).

Specific Indirect Effect Among the Self-Employed

Several studies using research program data from the U.S. and other countries suggest evidence for the specific effect on subsequent income reporting, with the strongest effect among the self-employed. In the U.S., a study using NRP data from randomly assigned audits as a "treatment" group along with general taxpayer return information as a "control" group found that being audited increases reported wage income the following year by 1.3 percent on average, but increases reported Schedule C income by 14.2 percent. This effect begins to diminish three years after being audited and mostly disappears after four years (DeBacker *et al.* (2015)).

[1] Other terms for these types of indirect effects: general indirect effect is also referred to as general deterrence or the "spillover effect," and specific indirect effect is also referred to as a "dynamic effect" of audits (Advani *et al.* (2015)).

Further, random audit data from a Danish program has shown that being randomly audited was associated with an increase in income reported the following year, and this increase was largely driven by the self-employed. The results of this study suggest that the self-employed are most likely to be noncompliant but also show the strongest adjustment in reporting one year[2] after an audit (Kleven *et al.* (2011)). Confirming the conclusion about the importance of third-party reporting for an indirect effect,[3] U.K. taxpayers audited at random increased reported tax liability substantially over a four-year period for taxpayers who filed self-assessed[4] income tax returns, which includes individuals with self-employment income and landlords, among others (Advani *et al.* (2015)). Overall, the finding that self-employed taxpayers are more sensitive to the indirect effect of audit for subsequent year reporting suggests that the underlying characteristics of the taxpayer and the return itself are key for understanding how indirect effects work.

Two recent IRS studies examined the impact of audits on future compliance among the self-employed, using operational audit data. In both, audits were not randomly assigned, but rather happened as part of standard operational procedures. Focusing on sole proprietorship compliance, one study found that among taxpayers who all had high DIF scores, audited taxpayers saw decreases in their DIF scores (indicating increased compliance) over the following five years compared to not audited taxpayers; this effect disappeared by the fifth year after audit (Nestor and Beers (2014)). In a second study, researchers used propensity score matching techniques to conduct a quasi-experiment. They found that being audited increased reported Schedule C net profit and taxable income of taxpayers whose previous audits resulted in additional tax liability assessments,[5] and this effect persisted over the next three years. Conversely, taxpayers who were audited previously but the audit did not result in a change in tax liability saw a decline in compliance three years after audit (Beer *et al.* (2015)).

Specific Indirect Effect among Other Populations

In this study, we build on prior work that suggests that differences in population characteristics and also the nature of different categories or types of audits may be differently associated with subsequent year reporting. In addition to the self-employment-focused studies above, a few studies have investigated the specific indirect effect of audits on other populations, such as those taxpayers who report capital gains and losses, list supplemental income, itemize deductions, or claim the Earned Income Tax Credit (EITC) on their returns. Among taxpayers audited randomly

[2] Unlike in the U.S., the Danish audit schedule completes audits in one year (U.S. audits can take anywhere from one to three years after the taxpayer has filed to initiate, and about another year or so after initiation to close). Kleven *et al.* therefore observed income reporting only one year after the audit. They did not test for attenuation in audit effects over time.

[3] This is because, as noted by many researchers, the lack of third-party reporting means that self-employed taxpayers have more room to be noncompliant, since there is no way to cross-reference the information on their returns (DeBacker *et al.* (2015); Erard and Ho (2003); Kleven *et al.* (2011); also discussed in Slemrod (2016)).

[4] In the UK, not all taxpayers have to submit self-assessed tax returns. Those who do need to submit them tend to be individuals with income from self-employment, people with very high incomes, landlords, and people collecting pension income (Advani *et al.* (2015)).

[5] Beer *et al*. (2015) used the outcome of the audit as a proxy for whether or not the taxpayer was assessed as being compliant or noncompliant. That is, if the audit recommended additional tax assessments, the taxpayer was noncompliant (did not report enough tax liability); if the audit resulted in no recommended change, the taxpayer was compliant (reported appropriate tax liability).

in the U.S., there is evidence that Schedule A itemized deductions, adjustments to income, Schedule C income, and Schedule E income are all sensitive to a research audit; in all four cases, taxpayers report more income and fewer deductions after the audit and the effect was persistent for up to six years. Conversely, no evidence was found of Schedule D income changing in response to a research audit. Two studies have shown that Earned Income Tax Credit claiming decreases after experiencing an audit; after a random NRP audit, taxpayers who claimed EITC decrease their future EITC claiming (DeBacker *et al.* (2018)), and taxpayers who were audited operationally for EITC credit validity also reduce EITC claiming in subsequent years, especially within the first year after audit (Guyton *et al.* (2018)).

Other studies have attempted to characterize the specific indirect effect that enforcement actions taken by the IRS have on filing compliance. Given that nonfiling is challenging to observe, research surrounding this topic tends to consider only known nonfilers. Specifically, this includes those who did not appear on a tax return but had income reported to the IRS by a third party, usually through Form W-2, Form 1099-R, or other documents (Datta *et al.* (2015); Guyton *et al.* (2017)). One such study considered the effect of the Automatic Substitute for Return (ASFR) process, a function of the IRS applied to eligible nonfilers who have not responded to prior notices by filing a return. In the study, researchers found evidence of increased timely filing compliance up to four years after the ASFR treatment, with the effect decreasing each year (Datta *et al.* (2015)).

Other Evidence for Specific Indirect Effects

Two additional areas of research provide further evidence for specific indirect effects: laboratory experiments that attempt to replicate the condition of being audited in an artificial setting, and field experiments using enforcement "contacts" as a proxy for audits.

Laboratory experiments have found results that do not correspond to results observed in natural settings. For example, two studies using university students found evidence of a "bomb crater" effect of compliance, in which compliance decreases immediately after an audit, then increases (Kastlunger *et al.* (2009); Maciejovsky *et al.* (2007)).

Enforcement contacts, typically in the form of letters or tax official visits, have been used to study specific indirect effects as well. Studies show that deterrence messages designed to make the threat of audit or other enforcement activity result in an increase in compliance, both immediately and over a period of several years after the fact. In a natural field experiment conducted in Minnesota, a random sample of taxpayers who received a letter alerting them that their returns would be "closely monitored" showed increased payments compared to a control group (Slemrod *et al.* (2001)).

Similar effects have been observed among nonfilers in the U.S. In Detroit, tax "ghosts" (i.e., nonfilers) who received a letter explaining noncompliance penalties were more likely to file back-year returns, remit payments, and report greater tax liability compared to nonfilers who received letters with no penalty message or with non-deterrence messages about civic pride (Meiselman (2018)). Similarly, tax delinquents in three U.S. states were 7 percent more likely to

submit payments within ten weeks after receiving a letter indicating their state's financial penalties for noncompliance compared to a control group (Perez-Truglia and Troiano (2015)).

Research Questions and Hypotheses

In this study, we address the following research questions. For research questions 1 and 2, we separately answer each question for each of the three types of correspondence audit analyzed.

Total Tax Reporting

1. How does tax reported by taxpayers who were audited on any of their returns for Tax Years 2006 through 2012 vary over time after audit compared to the tax reporting of taxpayers who were eligible for the same type of audit, but were not audited?
    a. Hypothesis 1 (H1): We hypothesize that the indirect effect of the audit will have an association with tax reporting, measured in comparison to the reporting of eligible unaudited taxpayers, three to five years after the audit and the effect will subsequently attenuate.

Reporting on Audit-Specific Line Items

2. Is there evidence that audit-specific line item reporting by taxpayers who were audited changes over time compared to the reporting of similar taxpayers who were eligible for the same type of audit, but did not experience an audit?
    a. Hypothesis 2 (H2): We hypothesize that the indirect effect of the audit will have an association with specific line item reporting, measured in comparison to the reporting of eligible unaudited taxpayers, three to five years after the audit and the effect will subsequently attenuate. The audit-specific line items are some Schedule C line items, some Schedule A line items, and some Schedule SE line items, respectively, for the three audit categories.

Exploratory Comparison between Audit Categories

For research questions 3 and 4, statistical testing between categories of audits is not conducted because the underlying populations are different. Our questions are therefore framed as exploratory and we do not present hypotheses.

3. Do the empirical results from our analyses of different categories of audits suggest that the effect of an audit on subsequent tax reporting varies by the category of audit experienced?
4. Do the empirical results from our analyses of different categories of audits suggest that the effect of an audit on subsequent specific line item reporting varies by the category of audit conducted?

Data and Methods

Categories of Correspondence Audits

Correspondence audits, unlike field audits, are conducted via mail and are designed to examine a small set of line items or issues on a taxpayer's return. As such, correspondence audits focus on narrowly defined candidate populations as being "eligible" for a category of correspondence audit.

In this study, we separately compare eligible/not audited and audited taxpayers for three distinct categories of correspondence audit. We selected categories of correspondence audit that were active for the full study period (Tax Years 2006 – 2012), for which we have access to operational eligibility and selection criteria, and for which there was a sufficient volume of audits each year[6]. We control for potential confounding factors by limiting our analysis population only to taxpayers who were part of the candidate population for a given correspondence audit category, as defined by IRS operational procedures. Due to data sensitivity, we cannot further elaborate on the creation of the eligible population.

Audit Category 1: Examines some Schedule C expenses among taxpayers who filed a Schedule C (to report nonfarm business income) and met other category-specific eligibility criteria.

Audit Category 2: Examines some Schedule A deductions among taxpayers who itemized deductions and met other category-specific eligibility criteria.

Audit Category 3: Examines Schedule SE self-employment tax among taxpayers who met certain category-specific eligibility criteria.

Additionally, in order to select which returns to audit from the overall candidate population, examiners for each type of audit rely on different prioritization metrics, typically characteristics of the return. These prioritization metrics are specific to the audit category and cannot be further explained here due to data sensitivity. We treat these prioritization criteria as control variables. As such, we exploit knowledge of operational criteria to help account for confounding factors that inform audit selection.[7]

Data

In this study, we combine data on the three types of correspondence audits described above with return information on the general taxpayer population in the U.S. that met operational eligibility criteria for each type of audit. We use tax return and audit record data for primary Taxpayer Identification Numbers from the IRS's Compliance Data Warehouse (CDW) for Tax Years 2006-2018. In our analyses, we define the "baseline" year as the Tax Year a given taxpayer entered the sample, either because that taxpayer had an audited return for that Tax Year, or because they fell into the sampled eligible-not audited group for that audit type for that Tax Year. In cases where a taxpayer entered the analytical sample multiple times (due to being

[6] We define sufficient volume arbitrarily as having roughly 1,000 cases each Tax Year.

[7] We have access only to IRS operational documents from the most recent one to three Tax Years. As such, we assume that operational criteria stayed relatively stable over time for each correspondence audit type. We cannot know for sure if this assumption is correct.

eligible for the category of audit for multiple years and/or due to being audited multiple years), we handled these taxpayers as follows:

1. For any taxpayers that our queries returned multiple times because they were captured as "eligible" multiple times and were not audited in Tax Year 2006 through Tax Year 2012: we declare the most recent eligibility year as the "baseline" year.
2. For any taxpayers that our queries returned multiple times because they were audited multiple times under the same audit category: we declare the first audit record as the "baseline" year.
3. For any taxpayers that our queries returned as being eligible in one or more years and audited in one or more years: we declare the earliest (or only) audited record as the "baseline" year and consider them solely in the "audited" group.

Audit ("Treatment") Group

To define the audited group, all primary taxpayer identification numbers associated with one of the three types of audits for any Tax Year in the Tax Year 2006through Tax Year 2012 period in the Enforcement Revenue Information System (ERIS) database were identified and retained. For these audited group taxpayers, we collected tax return information from the Form 1040, Schedule A, Schedule C, and Schedule SE for the Tax Year of the baseline year and up to eight Tax Years after (up to TY18). For example, for baseline year 2006, we compiled return data up through Tax Year 2014; for baseline year 2012, we compiled return data up to Tax Year 2018. Eight years after baseline was chosen based on prior literature which suggests that an indirect effect is present from three to five years after audit; this allows for a buffer window at the end to ensure any possible attenuation in effect can be captured.

Eligible, Not Audited ("Control") Group

To define the eligible, not audited group, we applied undisclosed operational filter criteria to return records from the full universe of non-audited taxpayers available in CDW. We restricted the returned records to a random sample of 25,000 taxpayers from the eligible population in each of Tax Year 2006 through Tax Year 2012, as this returned a sufficient sample size for our analysis based upon the known sizes of the audited or "treatment" group. In some Tax Years, there are fewer than 25,000 eligible taxpayers—in this case we selected all eligible taxpayers regardless of the population size. For these eligible group taxpayers, we collected tax return information from the Form 1040, Schedule A, Schedule C, and Schedule SE for the Tax Year of the baseline year and up to eight Tax Years after (up to TY18)Dependent Variables

Total Tax. Our primary dependent variable is total tax as reported on F1040. Total tax is chosen as the dependent variable across audit categories, as the change in tax paid over time most closely represents the "return on investment" that the IRS reaps from any observable specific indirect effect that results from the audit. Total tax, along with all other variables measured in dollars, are all adjusted for inflation to 2018 U.S. Dollars (USD). 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 where the taxpayer has reported zero total tax. The one dollar is added before taking the natural logarithm. If an indirect effect is present, we would expect total tax reporting to increase.

Audit Category 1 Schedule C Items. In our secondary analyses of line items that may have an association with an indirect effect of an audit, we treat some Schedule C line items as the dependent variable for audit category 1 models. We sum these undisclosed line items to create one continuous quantity. Because the sum of these line items is again strongly right skewed, we use the natural logarithm of this sum plus one dollar. Note that in this case, increased reporting of these Schedule C line items should have the effect of decreasing overall tax liability; thus, we would expect a positive indirect effect to be associated with decreased reporting on these line items.

Audit Category 2 Schedule A Items. We next treat some Schedule A line items as the dependent variable for audit category 2 models. We sum these line items to create one continuous quantity. Because the sum of these line items is strongly right-skewed, we use the natural log of this sum plus one dollar. Note that in this case, increasing reporting of these Schedule A deductions should have the effect of decreasing overall tax liability; thus, we would expect a positive indirect effect to be associated with decreased reporting on these line items.

Audit Category 3 Schedule SE Items. Finally, we treat a relevant line item derived on the Schedule SE as the dependent variable for audit category 3. This line item is continuous, measured in dollars. Again, we use the natural logarithm of this line item plus one dollar. Note that in this case, increasing reporting of these Schedule SE line items should have the effect of increasing overall tax liability; thus, we would expect a positive indirect effect to be associated with increased reporting on these line items.

We will now refer to audit category 1, 2, and 3 specific line items as "relevant items."

Independent Variables

Audit-time interaction. The primary variables of interest are audit status and its interaction with time, specified as Tax Years since the baseline year. Audit status is a time-invariant variable for each taxpayer, as they can be considered only as "audited" or "not audited" in our sample. Years after baseline is time-varying, meaning that it takes on a different value for each of a taxpayer's returns to describe the time between that return and the audited or eligible return. We define the baseline year as Year 0, and we fit time 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), adjusted to reflect 2018 U.S. dollars[8]. 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). We derive an urban/not urban (*Urban)* classification using zip code data and Census Bureau definitions[9]. A

[8] Total Positive Income is defined as the sum of wages, salaries, interest, and dividends and does not subtract losses or deductions.

[9] https://www.census.gov/geographies/reference-files/2010/geo/relationship-files.html#par_textimage_470670252

binary wage indicator is derived based on the presence of any non-wage income reported on F1040 (*any wages)*. We adjust for *total exemptions*, and the presence of claiming any *Child Tax Credit*. To account for home ownership, we control for a continuous measure of mortgage interest deductions (*mortgage interest)*. For audit categories 1 and 3, we adjust our estimates for whether or not the taxpayer itemized deductions as indicated by the presence of a Schedule A (*itemized deductions)*. TPI, FS, Urban, any wages, total exemptions, any Child Tax Credit, mortgage interest, and itemized deductions all are treated as time-varying covariates. We also fit *Tax Year* of the return as a categorical variable with possible values Tax Year 2006 through Tax Year 2018. In the models predicting total tax only, we also control for *Priority,* a variable representing the metric used operationally by the audit category in question to rank and select returns for audit. For audit categories 1 and 3, this is measured in 2018 USD with the interpretation that higher priority is more likely to be audited. This variable is distinct for each category of audit and is time-invariant, meaning that it is the taxpayer's assigned priority in the baseline year.

Statistical Analysis

To assess the relationship between audit status and the outcomes of interest over time, a linear mixed effect model is fit for each outcome and audit category. Linear mixed effects models are longitudinal models 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 TIN, which allows each taxpayer 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. Within-taxpayer correlation is modeled with an autoregressive structure, as is common with evenly spaced repeated measures. The model specifications are provided in equations (1) and (2) for the i^{th} taxpayer and j^{th} return (years after baseline). Analyses were conducted with R version 3.4.4, using the modeling package *nlme* (Pinheiro (2019)).

Model 1: Total Tax Reporting Over Time

For each category of audit, we separately estimate model (1) below, in which $\ln(\text{total tax} + 1)_{ij}$ denotes the natural logarithm of total tax in U.S. dollars plus one dollar, adjusted for inflation, for each individual *i* at year *j*. $\beta_{11} audited_i$ is a time-invariant measure of whether or not the taxpayer was audited for the tax return filed at baseline year Models for audit category 2 are not adjusted for whether the taxpayer itemized their deductions since eligibility for this audit necessitates itemizing deductions . γ_{0i} denotes a random effect on TIN.

$$
\begin{aligned}
(1) \qquad & \ln(\text{total tax} + 1)_{ij} \\
& = \beta_0 + \gamma_{0i} + \beta_1 FS_{ij} + \beta_2 TY_{ij} + \beta_3 TPI_{ij} + \beta_4 priority_i + \beta_5 any\ wages_{ij} \\
& + \beta_6 total\ exemptions_{ij} + \beta_7 any\ Child\ Tax\ Credit_{ij} \\
& + \beta_8 itemized\ deductions_{ij} + \beta_9 mortgage\ interest_{ij} + \beta_{10} urban_{ij} \\
& + \beta_{11} audited_i + \beta_{12} year.after.baseline_{ij} + \beta_{13} audited_i \\
& * year.after.baseline_{ij} + \epsilon_{ij}
\end{aligned}
$$

Model 2: Audit-specific Line Items Reporting Over Time

Next, for each category of audit, we separately estimate model (2) below, in which the natural logarithmic transformation of the sum of relevant items + 1 denotes a single or sum of relevant line items for the category of audit in U.S. dollars plus one dollar, adjusted for inflation, for each individual i at year j. $\beta_5 audited_i$ is a time-invariant measure of whether or not the taxpayer was audited for the tax return filed at baseline year. γ_{0i} denotes a random effect on TIN.

$$
\begin{aligned}
(2) \qquad & \ln\left(\sum \text{relevant items} + 1\right)_{ij} \\
& = \beta_0 + \gamma_{0i} + \beta_1 FS_{ij} + \beta_2 TY_{ij} + \beta_3 TPI_{ij} + \beta_6 any\ wages_{ij} \\
& + \beta_7 total\ exemptions_{ij} + \beta_8 any\ Child\ Tax\ Credit_{ij} \\
& + \beta_9 itemized\ deductions_{ij} + \beta_{10} mortgage\ interest_{ij} + \beta_{11} urban_{ij} \\
& + \beta_{12} audited_i + \beta_{13} year.after.baseline_{ij} + \beta_{14} audited_i \\
& * year.after.baseline_{ij} + \epsilon_{ij}
\end{aligned}
$$

Results

Descriptive Statistics

The sample sizes of each audit category and baseline year are shown in Figure 1. Audit categories 1 and 2 have 253,132 and 247,837 unique taxpayers, respectively. For audit category 1, audits were most common in Tax Year 2010 and least common in Tax Year 2011. Similarly, for audit category 2, Tax Year 2011 was a lighter year for audits, while Tax Year 2007 has the highest audit frequency. Audit category 3 is a less common audit for which relatively few taxpayers were eligible but not audited. This category has in total 64,823 audited and not audited taxpayers.

Tables 1-3 summarize return characteristics by audit status for all data in the baseline year by audit category. All values are shown in 2018 USD. For continuous variables, we conducted a Wilcoxon rank-sum test to assess whether the difference between the audited and not-audited groups for each audit category are statistically significant. Categorical variables are assessed for association with audit status using a Chi-Square test. All differences are statistically significant at the $p < 0.05$ level.

For audit category 1, there is evidence to suggest that audited taxpayers at baseline have a statistically significantly higher total tax than the not-audited group. Additionally, the audited taxpayers at baseline appear to have a higher TPI compared to the not audited. In terms of audit priority, which is sensitive, unsurprisingly the audited taxpayers have on average significantly higher priority. However, it is important to note that there is still some overlap between groups: there are audited taxpayers with zero priority and a not-audited taxpayer with an extremely high priority. We cannot access information that would explain why a taxpayer with low priority would be audited and a taxpayer with high priority would not be. It is possible that these audits met some unknown exclusion criteria, the return was selected for a different audit, or something

having to do with the timing of the return filing. The relevant items variable, which is the sum of some Schedule C line items, is also significantly higher for the audited group.

For audit category 2, there is again evidence to suggest that audited taxpayers at baseline have a statistically significantly different value for total tax compared to the not audited taxpayers. However, the audited taxpayers at baseline appear to have a slightly lower TPI compared to the not audited. The not audited group has a slightly higher priority than the audited group, which is further considered in the Discussion section.

Finally, for audit category 3, there is evidence to suggest that audited taxpayers at baseline have a significantly different distribution of characteristics for all variables considered. For example, the audited group appears to have on average higher TPI and total tax.

Timing of Audits

Figure 2 summarizes the time to audit exam start and end by audit category. We assume that exam start date coincides with when the taxpayer is notified that their return is being examined, and thus marks when we might expect to observe a behavioral response to the audit. The distribution of time to exam start in Figure 2 indicates that for most taxpayers and all three audit categories, taxpayers are notified of their audit approximately two to three years after the December of the TY for which they filed the audited return. Almost all taxpayers are aware that they are being audited within four years after the TY of the audited return. This suggests that if an indirect effect is present, it will mostly likely not manifest until two or three years after the TY of the audited return. For example, if a taxpayer is audited for their Tax Year 2008 return, which encompasses taxes paid through December of 2008, they are likely to know about this audit by December of 2011. They will file their Tax Year 2011 return between January and April of 2012, meaning that we can expect this taxpayer to be aware they are being audited and exhibit any potential behavior change in their Tax Year 2011 return (three years after baseline).

Modeling Results: Total Tax

Audit Category 1

Table 4 displays the estimates from the total tax model for audit category 1, which deals with Schedule C line items. Figure 3 shows the predicted changes in total tax over time for the audited and not audited groups based on the estimated coefficients for the audited, years after baseline, and audit*years after baseline interaction variables. There is sufficient evidence to suggest a difference in total tax reporting for the baseline year: on average the audited taxpayers remit 76.3 percent of that of the not audited taxpayers (95 percent confidence interval (CI) 74.6 – 78.0), while holding the control variables constant. One year after baseline, it is estimated that both groups have more similar values of total tax: the audited group paying 64.9 percent (CI 63.4 – 66.4) that of the not audited in baseline's total tax and the not audited group paying 70.9 percent (CI 69.8 – 72.1) that of their own baseline total tax. However, in year two, the audited group's predicted total tax increases sharply to 84.8 percent that of the not audited in baseline (CI 82.8 – 86.8), while the not audited group's estimated total tax decreases in slope. By three years after baseline, both groups show evidence of decreasing total tax over time when adjusting for control variables.

Audit Category 2

For audit category 2, which deals with Schedule A line items, the results of the total tax model are also presented in Table 4. Figure 4 shows the predicted values over time for the audited and not audited groups. In year 0, for taxpayers with the same values of control variables and in the same Tax Year, it is estimated that the audited taxpayer on average has a total tax 2.62 times that of the not audited taxpayer in the same year (CI 2.55– 2.70). While there is evidence that the not audited taxpayer increases its total tax over time, there is also evidence of a significant jump in the audited taxpayers' total tax between two and three years after baseline. Two years after baseline, it is expected that the audited taxpayer has a total tax 4.41 times that of the not audited taxpayer in year 0 (CI 4.28 – 4.54), while the not audited taxpayer is expected to increase its total tax just 2.09 times relative to its baseline tax (CI 2.04 – 2.14). The slope of the audited taxpayers is estimated to decrease beginning three years after baseline, while the not audited estimated total tax is still increasing.

Audit Category 3

Finally, the results of the linear mixed effects model for the log of total tax in audit category 3 is also presented in Table 4 with predicted values plotted in Figure 5. In year 0, for taxpayers with the same values of control variables, and in the same Tax Year, it is estimated that the audited taxpayer on average has a total tax 22.8 percent more than that of the not audited taxpayer in the same year (CI 15.8 – 1.30). After both groups dip one year after baseline, the audited group's estimates increase at two years after baseline while the not audited group remains approximately the same. By three years after baseline, the audited group's estimated total tax is decreasing.

Modeling Results: Audit Category Relevant Line Items

Audit Category 1

Table 5 displays the estimates from the model for audit category 1's sum of relevant items outcome, which deals with Schedule C line items. Figure 6 shows the predicted changes in relevant line items over time for the audited and not audited groups. There is sufficient evidence to suggest a difference in relevant items reporting for the baseline year: on average the audited taxpayers have 309 times more in relevant items (CI 299.95 – 319.16) than that of the not audited taxpayers, while holding the control variables constant. In year 1, the audited taxpayers are estimated to have 22.41 times more in relevant items compared to the not audited group (CI 21.71 – 23.13). However, by year 2, there is insufficient evidence to suggest a difference in relevant items reporting between the audited and not audited groups.

Audit Category 2

The predicted values for the model of relevant Schedule A items for audit category 2 are displayed in Figure 7. On average, it is estimated that in the baseline year the audited group has a sum of relevant items on average 11 percent lower than that of the not audited group (estimate: 0.89, CI 0.86 – 0.91). One year later, the audited taxpayers have a relevant sum 74 percentless than the not audited group in the baseline year (estimate: 0.26, CI (0.25 – 0.26)), while the not audited taxpayers report 70 percent less in relevant items compared to their prior year (CI 0.30 –

0.32). After years two to three, both groups appear to trend towards no longer reporting the relevant items.

Audit Category 3

For audit category 3, regarding Schedule SE items, the audited and not audited groups have relatively similar reporting at baseline assuming the same values of the control variables. The audited group has on average 5 percent higher reporting of relevant items (estimate: 1.05, CI (1.01 – 1.10), p = 0.026). However, the audited group has a marked increase to 5.95 times that of the not audited baseline group by year 1 (CI 5.69 – 6.23), while the not audited group increases less to 3.60 (CI 3.46 – 3.75). By year 2, the audited group peaks in its relevant items reporting to a multiplicative change of 7.92 (CI 7.56 – 8.29). Following that jump, the predicted values begin to have a negative slope and approach the estimates of the control group.

Sensitivity Analysis: Audit Category 1

To address the disparity in baseline characteristics for Audit Category 1 taxpayers, a sensitivity analysis is conducted on taxpayers with similar baseline priority. This is defined as taxpayers with a priority in the baseline year between $5,000 and $8,000 (2018 USD), a range chosen upon inspection of the distribution of priority by audit status. In total, 12,308 taxpayers fall into this range: 4,376 from the audited population and 7,392 from the not audited population.

Total Tax

Table 6 displays the estimates from the total tax model for this sensitivity analysis of Audit Category 1 taxpayers with similar baseline priority. Figure 9 shows the predicted changes in total tax over time for the audited and not audited groups based on the estimated coefficients for the audited, years after baseline, and audit*years after baseline interaction variables. There is sufficient evidence to suggest a difference in total tax reporting for the baseline year: on average the audited taxpayers remit 68.77 percent that in total tax compared to the not audited taxpayers (CI 61.58 - 76.81), while holding the control variables constant. However, in years two and three, the audited group's predicted total tax increases further while the not audited group appears to remain constant in slope. By three years after baseline, both groups show evidence of decreasing total tax over time when adjusting for our control variables.

Relevant Line Items

Table 6 also displays the estimates from the model for audit category 1's sensitivity analysis on the sum of relevant items outcome, which deals with Schedule C line items. Figure 10 shows the predicted changes in line item reporting over time for the audited and not audited groups. There is sufficient evidence to suggest a difference in relevant items reporting for the baseline year: on average the audited taxpayers have 82 percent higher reporting in relevant line items (estimate 1.82 CI 1.56 – 2.13) than that of the not audited taxpayers, while holding the control variables constant. In year 1 both groups decrease in their claiming of relevant line items: the audited taxpayers are estimated to report 16.37 percent that of the not audited group's baseline reporting (CI 14.04 – 19.01) and the not audited reporting on average 11.22 percent that of their reporting the year prior (CI 10.18 – 12.36). However, by year 2, there is insufficient evidence to suggest a difference in relevant items reporting between the audited and not audited groups.

Discussion

In this study, we investigated the indirect effect of experiencing an audit on subsequent total tax reporting and on reporting of other relevant line items for three categories of correspondence audit. We advance prior literature in two ways: 1) by accounting for operational selection criteria and 2) by expanding the focus to include taxpayers who do not report self-employment income, but rather are examined for other types of reporting characteristics. Prior studies that use operational data to construct "treatment" and "control" groups *ex post* have typically relied on DIF scores when considering the likelihood of experiencing an audit (e.g., Beer (2015); Nestor and Beers (2014)); however, in the case of correspondence audits, other criteria are used instead of DIF, and we are able to account for these in this study. As in any study using operational data, we grapple with the challenge that taxpayers are selected into the "treatment" condition of audit based on criteria that is only partially known (Slemrod (2016)), even from within a narrowly defined candidate population.

For all three audit categories, we find evidence suggestive of an indirect effect. In audit category 1, which deals with Schedule C items, there is an increase in predicted total tax for the audited group around one to three years after baseline, followed by an attenuation out to year eight. Considering that most audit category 1 exams will have started three years after baseline, we assume that most of the audited taxpayers have been notified by the peak in reporting observed in Figure 3 at three years after baseline. In this way, our results mirror prior findings from both research audit data on Schedule C filers (DeBacker *et al.* (2015)) and findings using operational data on Schedule C filers (Beer (2015)). Interestingly, we find similar evidence of a specific indirect effect for audit category 2 (Schedule A itemizers) and weak evidence of a specific indirect effect for audit category 3 (self-employment tax). To our knowledge, these specific taxpayer populations have not been explicitly examined in other studies, which have tended to focus on taxpayers who report self-employment income to other taxpayers more generally. Our findings suggest that when other populations of taxpayers who are audited are compared to unaudited taxpayers with similar characteristics (i.e., the "eligible" group), specific indirect effects may emerge more clearly.

Further in alignment with prior studies, our results show evidence of attenuation of specific indirect effects across all three audit categories. We observe peak indirect effects around the time that taxpayers' audits generally start – around year three after the audited return was filed – and we then see convergence between audited and not audited groups starting about five years after audit. It is a notable contribution of this research that this attenuation appears to hold in three separate populations of taxpayers.

Although Table 1 shows differences in the underlying characteristics of the audited and not audited groups, it is important to note that all models are controlling for the audit priority variable. This allows us to account for a degree of selection bias in the "treatment" condition of being audited. Although the IRS could have applied further exclusion criteria of which we are unaware, the priority variable reflects knowledge that is typically unknown to researchers using operational audit data and represents a step in the right direction of accounting for the endogeneity inherent in using non-random audit data, and our future work aims to continue

building on this. Because priority is included in the model as a control variable, interpretation of the estimated coefficients is for taxpayers *with the same audit priority in baseline*. Therefore, the modeling results apply to relatively homogenous taxpayers who are similarly likely to be audited. Our sensitivity analysis on an overlap in Audit Category 1 taxpayers confirms this, in which we see results similar for a subset of audited and not audited taxpayers who are more similar according to baseline priority as we see in the full sample. While the audited and not audited groups in the sensitivity analysis have different baseline estimates for total tax, there is still evidence to suggest that their slopes are significantly different over time. This interaction between audit and time is the primary predictor of interest in this study – the audited and not audited groups may have different baseline values, but we are assessing whether they are parallel in reporting behavior over time.

Our final research questions ask whether the specific indirect effects on total tax reporting and relevant line item reporting vary by category of audit. Because we did not conduct formal hypothesis testing between these disparate populations, our results are more qualitative in nature.

First, in comparing total tax model results for the three audit categories, we see that the trajectories and magnitudes of difference in total tax paid by the audited and not audited groups appear distinct across categories. For audit category 1 (Schedule C), audited taxpayers are estimated to start out at baseline paying less tax than the control group, but the groups' trajectories cross as audited taxpayers increase their total tax reporting, peaking at year three after baseline (when most taxpayers' audits start) and then attenuating. Audit category 1 is the only category in which the not audited group has a consistently negative total tax slope. This may suggest that these audit-eligible taxpayers were reporting higher than usual tax liability in their baseline year, and this tax liability was correlated with the operational criteria that made them eligible for the audit in the first place. However, based on the multiplicative interpretation of the exponentiated model coefficients, the magnitude of the difference between groups appears relatively smaller compared to audit categories 2 and 3. In contrast, the trajectory of the audited group's reporting under audit category 2 (Schedule A) seems to show the most responsiveness among audited taxpayers across all audit categories: these taxpayers are predicted to increase their total tax reporting at years two and three after audit to more than four times the total tax reporting of unaudited taxpayers at the baseline year. Interestingly, there is evidence of attenuation of the specific indirect effect across all categories of audit, where we see the audit group's estimated change in tax approach that of the not audited group in the later years. However, the timing of this appears earlier for audit category 3 (SE tax) than for categories 1 and 2, which could suggest that the latter have longer-lasting effects on tax reporting.

Comparing specific line item models, we see more pronounced differences between audit categories in the reporting trends for line items specific to the type of audit. First, we see that for some audit categories, "baseline" values of relevant items more closely align between the audited and not-audited groups. When adjusting for potential confounders, audit category 3 taxpayers are most alike in their baseline reporting of Schedule SE relevant items, and the audited group increases their reporting after audit more sharply than the unaudited group. This may indicate an "education effect" – that is, audited taxpayers learn about what they should have reported in

order to submit a correct tax return, and they adjust their reporting accordingly. Similarly, audited Schedule A taxpayers appear to decrease their reporting of certain deductions in years one to two more steeply after audit when compared to not audited taxpayers. It is interesting to note that both audited and not audited groups substantially decrease their deduction reporting after the baseline year; this is likely an artifact of the selection criteria that made all of these taxpayers eligible for the audit category at the baseline year when they entered our analytical sample. That is to say, they may have been selected into the audited or eligible group on the basis of having an "unusual" or outlier year for reporting certain items.

In contrast, under audit category 1, the trends in the reporting of certain Schedule C expenses tell a different story. The results shown in Figure 6 portray a heavy selection bias toward taxpayers who report higher values on these Schedule C line items when being designated for audit – one of the challenges of using operational data. By two years after audit, it appears that most of the audited taxpayers are no longer claiming these relevant Schedule C line items. Additional work needs to be done in constructing a more comparable control group for further analysis of line items.

Limitations

We applied operational eligibility criteria to construct a "control" group. In doing so, we operate under the assumption that the categories of audit we analyze here have been relatively stable over time, especially with regard to 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 filters (e.g., those used for Tax Year 2018), but these filters may not necessarily apply to Tax Year 2006 – Tax Year 2012 returns, and we do not have knowledge of the eligibility criteria used in historic years for all audit categories. 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 for each audit category applies to Tax Year 2006 – Tax Year 2012. Further, there appears to be some overlap between the distributions of priority for the audited and not audited groups, as observed in all three audit categories. This could potentially be due to the date the returns were filed and how quickly they were picked up in the correspondence audit cycle. However, discrepancies between priority and audit status could also imply that there exist additional audit selection criteria unknown to us.

Additionally, audited taxpayers have varying notification times, even for audits of returns from the same Tax Year, and results must be interpreted while considering the fact that not every taxpayer is aware of their audit by the time they are preparing their tax return for a subsequent Tax Year. Finally, not all taxpayers have a complete set of returns after the baseline year; this absence is assumed to be Missing at Random (MAR).

Finally, a mixed effects model assumes that the random effects are independent from the residuals. In the presence of unobserved confounders this assumption is not likely to be met, and therefore the estimation could be biased.

Future Research

Our plans for future research include executing analyses comparable to the ones presented here over additional categories of correspondence audit, as well as across other types of audits beyond correspondence. We will also continue to explore whether and how the audit category and underlying differences in population matter in terms of the form that a specific indirect effect takes. This approach has the operational potential of providing new information about which categories of audit have the greatest specific indirect effect on IRS revenue.

We acknowledge that there exist further control variables to be considered in future models, such as those that would better account for tax policy changes. Other data points available to us, such as whether a taxpayer used a tax preparer, might also have some degree of explanatory power in the relationship between audit experience and subsequent tax reporting, and future research should continue to investigate these relationships. Additionally, despite using the best filter criteria to select the control group, there appear to be different underlying characteristics between the audited and not audited groups; thus, an assumption of exchangeability is unlikely to hold here. Ensuring we have comparable control groups for all audit categories is a priority of our research going forward. Given this, we have already arranged for a purely random control group to not be audited among returns filed for a recent Tax Year that meet all of the selection criteria of one of the three categories of audit we featured in this paper. That should allow us to evaluate how much our current results overstate or understate the indirect effect.

References

Advani, Arun, William Elming, and Jonathan Shaw. 2015. "How Long-Lasting Are the Effects of Audits?" *Tax Administration Research Centre* Discussion Paper 011-15. https://tarc.exeter.ac.uk/media/universityofexeter/businessschool/documents/centres/tarc/publications/discussionpapers/How_long_lasting_are_the_effects_of_audits_v3.pdf.

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, and Mohammad Yunus. 2009. "Spatiality and Persistence in U.S. Individual Income Tax Compliance." *National Tax Journal* 62 (1): 101–24.

Beer, Sebastian, Mathias Kasper, Erich Kirchler, and Brian Erard. 2015. "Audit Impact Study." *TAS Research and Related Studies* 2. https://taxpayeradvocate.irs.gov/Media/Default/Documents/2015ARC/ARC15_Volume2_3-AuditImpact.pdf.

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.

Boning, William C., John Guyton, Ronald H. Hodge, II, Joel Slemrod, and Ugo Troiano. 2018. "Heard it Through the Grapevine: Direct and Network Effects of a Tax Enforcement Field Experiment". Working Paper 24305. National Bureau of Economic Research. https://www.nber.org/papers/w24305.pdf.

Datta, Saurabh, Stacy Orlett, and Alex Turk. 2015. "Individual Nonfilers and IRS-Generated Tax Assessments: Revenue and Compliance Impacts of IRS Substitute Assessments When Taxpayers Don't File". Small Business/Self-Employed Division, Internal Revenue Service. https://www.irs.gov/pub/irs-soi/15rescondatta.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). 2015. /paper/Once-Bitten-%2C-Twice-Shy-%3F-The-Lasting-Impact-of-IRS-Debacker-Heim/84175d69cf5c66003f921fa8c31a40d4c7806ba8.

DeBacker, Jason, Bradley T. Heim, Anh Tran, and Alexander Yuskavage. 2018. "The Effects of IRS Audits on EITC Claimants". *National Tax Journal* 71 (3): 451–484. https://doi.org/10.17310/ntj.2018.3.02.

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." https://www.researchgate.net/publication/247913772_The_Influence_of_Tax_Audits_on_Reporting_Behavior.

Erard, Brian, and Chih-Chin Ho. 2003. "Explaining the U.S. Tax Compliance Spectrum." *eJournal of Tax Research* 1 (2): 93–109.

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): 471–75. https://doi.org/10.1257/aer.p20171062.

Guyton, John, Kara Leibel, Day Manoli, Ankur Patel, Mark Payne, and Brenda Schafer. 2018. "Tax Enforcement and Tax Policy: Evidence on Taxpayer Responses to EITC Correspondence Audits". Working Paper 24465. National Bureau of Economic Research. https://www.nber.org/papers/w24465.pdf.

Hallsworth, M. 2014. "The Use of Field Experiments to Increase Tax Compliance." *Oxford Review of Economic Policy* 30 (4): 658–79. https://doi.org/10.1093/oxrep/gru034.

Kastlunger, Barbara, Erich Kirchler, Luigi Mittone, and Julia Pitters. 2009. "Sequences of Audits, Tax Compliance, and Taxpaying Strategies." *Journal of Economic Psychology* 30 (3): 405–18. https://doi.org/10.1016/j.joep.2008.10.004.

Kleven Henrik Jacobsen, Knudsen Martin B., Kreiner Claus Thustrup, Pedersen Søren, and Saez Emmanuel. 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica* 79 (3): 651–92. https://doi.org/10.3982/ECTA9113.

Maciejovsky, Boris, Erich Kirchler, and Herbert Schwarzenberger. 2007. "Misperception of Chance and Loss Repair: On the Dynamics of Tax Compliance." *Journal of Economic Psychology* 28 (6): 678–91. https://doi.org/10.1016/j.joep.2007.02.002.

Mazzolini, Gabriele, Laura Pagani, and Alessandro Santoro. 2017. "The Deterrence Effect of Real-World Operational Tax Audits". Working Paper No. 359. University of Milan Bicocca Department of Economics, Management and Statistics. http://dx.doi.org/10.2139/ssrn.2914374

Meiselman, Ben. 2018. "Ghostbusting in Detroit: Evidence on Nonfilers from a Controlled Field Experiment." *University of Michigan Working Paper*. http://www-personal.umich.edu/~mdbmeis/MeiselmanJMP.pdf.

Moulton, Brent R. 1986. "Random group effects and the precision of regression estimates." *Journal of Econometrics 32 (3) 385-97*. https://doi.org/10.1016/0304-4076(86)90021-7.

Nestor, Mike, and Tom Beers. 2014. "Estimating the Impact of Audits on the Subsequent Reporting Compliance of Small Business Taxpayers: Preliminary Results." *TAS* 2. https://taxpayeradvocate.irs.gov/Media/Default/Documents/2014-Annual-

Report/Estimating-the-Impact-of-Audits-on-the-Subsequent-Reporting-Compliance-of-Small-Business-Taxpayers-Preliminary-Results.pdf.

Perez-Truglia, Ricardo, and Ugo Troiano. 2015. "Shaming Tax Delinquents." Working Paper 21264. National Bureau of Economic Research. https://doi.org/10.3386/w21264.

Pinheiro, José. 2019. 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).

Pomeranz, Dina. 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *The American Economic Review*, no. 8: 2539.

United States Government Accountability Office. 2012. "IRS Could Significantly Increase Revenues by Better Targeting Enforcement Resources." Report GAO-13-151.

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, Marsha Blumenthal, and Charles Christian. 2001. "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota." *Journal of Public Economics* 79 (3): 455–83. https://doi.org/10.1016/S0047-2727(99)00107-3.

Tables

Table 1: Baseline Characteristics for Audit Category 1				
Variable (Unit)	**Audited (N = 123,292)**	**Not Audited (N = 129,840)**	**Total (N = 253,132)**	**p-value**
Total Tax (2018 USD)				
Mean (SD)	51512.88 (397378.4)	11273.45 (33118.53)	30872.71 (279068.48)	<0.0001[1]
TPI (2018 USD)				
Mean	316583.67	96045.56	203462.18	<0.0001[1]
(SD)	(2101042)	(142498.88)	(1473991.57)	
Priority (2018 USD)				
Mean (SD)	31063.95 (1402755.89)	2023.08 (9696.85)	16167.9 (979115.17)	<0.0001[1]
Filing Status (N (%))				
Married Filing Jointly	73668 (59.75%)	81615 (62.86%)	155283 (61.34%)	<0.0001[2]
Single/Other	49624 (40.25%)	48225 (37.14%)	97849 (38.66%)	
Relevant Items (2018 USD)				
Mean (SD)	26625.53 (1215624.72)	1419.59 (3053.31)	13696.55 (848480.5)	<0.0001[1]
Mortgage Interest (2018 USD)				
Mean (SD)	10041.03 (15416.47)	5050.55 (8389.18)	7481.25 (12573.01)	<0.0001[1]
Any wage income (N (%))				
No	8490 (6.89%)	12961 (6.54%)	21451 (8.47%)	<0.0001[2]
Yes	114802 (93.11%)	116879 (88.42%)	231681 (91.53%)	
Any Child Tax Credit (N (%))				
No	97111 (78.77%)	98300 (75.71%)	195411 (77.20%)	<0.0001[2]
Yes	26181 (21.23%)	31540 (24.28%)	57721 (22.8%)	
Itemized deductions (N (%))				
No	40949 (33.21%)	66041 (50.86%)	106990 (42.27%)	<0.0001[2]
Yes	82343 (66.79%)	63799 (49.14%)	146142 (57.73%)	
Urban zip code (N (%))				
No	2694 (2.19%)	2483 (1.91%)	5177 (2.05%)	<0.0001[2]
Yes	120598 (97.81%)	127357 (98.09%)	247955 (97.95%)	
Total Exemptions (N (%))				
0	108 (0.09%)	1703 (1.13%)	1811 (0.72%)	<0.0001[2]
1	31828 (25.82%)	38915 (29.97%)	70743 (27.95%)	
2	40532 (32.87%)	45748 (35.23%)	86280 (34.08%)	
3	20952 (16.99%)	17106 (13.17%)	38058 (15.03%)	
4	18996 (15.41%)	18519 (14.26%)	37515 (14.82%)	
5+	10876 (8.82%)	7849 (6.05%)	18725 (7.4%)	
[1]Wilcoxon rank sum test [2]Chi-Square test				

Table 2: Baseline Characteristics for Audit Category 2				
Variable (Unit)	**Audited (N = 146,337)**	**Not Audited (N = 101,500)**	**Total (N = 247,837)**	**p-value**
Total Tax (2018 USD)				
Mean (SD)	14443.42 (75399.03)	10998.19 (34016.16)	13032.45 (61915.22)	< 0.0001[1]
TPI (2018 USD)				
Mean (SD)	148079.81 (394170.95)	161959.99 (289591.74)	153764.35 (355149.98)	< 0.0001[1]
Priority				
Mean (SD)	0.43 (3.24)	0.68 (18.81)	0.53 (12.30)	< 0.0001[1]
Filing Status (N (%))				
Married Filing Jointly	82211 (56.18%)	61090 (60.19%)	143301 (57.82%)	<0.0001[2]
Single/Other	64126 (43.82%)	40410 (39.81%)	104536 (42.18%)	
Relevant Items (2018 USD)				
Mean	35756.68	46705.85	40240.84	< 0.0001[1]
(SD)	(152083.39)	(481205.66)	(329421.75)	
Mortgage Interest (2018 USD)				
Mean (SD)	7391.78 (12644.42)	7194.94 (15438.33)	7311.17 (13857.24)	< 0.0001[1]
Any wage income (N (%))				
No	16260 (11.11%)	29360 (28.92%)	45620 (18.41%)	<0.0001[2]
Yes	130077 (88.89%)	72140 (71.07%)	202217 (81.59%)	
Any Child Tax Credit (N (%))				
No	119873 (81.92%)	88470 (87.16%)	208343 (84.06%)	<0.0001[2]
Yes	26464 (18.08%)	13030 (12.84%)	39494 (15.94%)	
Urban zip code (N (%))				
No	2728 (1.86%)	2667 (2.63%)	5395 (2.18%)	<0.0001[2]
Yes	143609 (98.14%)	98833 (97.37%)	242442 (97.82%)	
Total Exemptions (N (%))				
0	119 (0.081%)	87 (0.086%)	206 (0.083%)	<0.0001[2]
1	40249 (27.5%)	27298 (26.89%)	67547 (27.25%)	
2	57382 (39.21%)	42225 (41.60%)	99607 (40.19%)	
3	22781 (15.57%)	14223 (14.01%)	37004 (14.93%)	
4	16431 (11.23%)	10944 (10.78%)	27375 (11.05%)	
5+	9375 (6.41%)	6723 (6.62%)	16098 (6.5%)	

[1]Wilcoxon rank sum test [2]Chi-Square test

Table 3: Baseline Characteristics for Audit Category 3				
Variable (Unit)	**Audited (N = 41,849)**	**Not Audited (N = 22,974)**	**Total (N = 64,823)**	**p-value**
Total Tax (2018 USD)				
Mean (SD)	13922.27 (203452.51)	19597.56 (610147.82)	15933.65 (398329.75)	< 0.0001[1]
TPI (2018 USD)				
Mean	120250.99	149923.93	130767.41	< 0.0001[1]
(SD)	(2701267.72)	(2250823)	(2550763.3)	
Priority (2018 USD)				
Mean (SD)	4438.8 (68700.82)	5967.47 (158729.98)	4943.84 (107166.15)	< 0.0001[1]
Filing Status (N (%))				
Married Filing Jointly	15632 (37.35%)	8956 (39.98%)	24588 (37.93%)	<0.0001[2]
Single/Other	26217 (62.65%)	14018 (61.02%)	40235 (62.07%)	
Relevant Items (2018 USD)				
Mean (SD)	21.65 (382.21)	33.71 (616.93)	25.92 (478.78)	< 0.0001[1]
Mortgage Interest (2018 USD)				
Mean (SD)	4156.36 (23719.38)	4544.28 (11014.87)	4293.84 (20155.5)	< 0.0001[1]
Any wage income (N (%))				
No	17433 (41.66%)	12805 (55.74%)	30238 (46.65%)	<0.0001[2]
Yes	24416 (58.34%)	10169 (44.26%)	34585 (53.35%)	
Any Child Tax Credit (N (%))				
No	33654 (80.42%)	17949 (78.13%)	51603 (79.61%)	<0.0001[2]
Yes	8195 (19.58%)	5025 (21.87%)	13220 (20.39%)	
Itemized deductions (N (%))				
No	27132 (64.83%)	14472 (63.00%)	41604 (64.18%)	<0.0001[2]
Yes	14717 (35.17%)	8502 (37.00%)	23219 (35.82%)	
Urban zip code (N (%))				<0.0001[2]
No	998 (2.38%)	721 (3.14%)	1719 (2.65%)	
Yes	40851 (97.62%)	22253 (96.86%)	63104 (97.35%)	
Total Exemptions (N (%))				
0	535 (1.28%)	213 (0.93%)	748 (1.15%)	<0.0001[2]
1	17886 (42.74%)	6484 (27.80%)	24370 (37.59%)	
2	10321 (24.66%)	5038 (21.93%)	15359 (23.69%)	
3	3715 (8.88%)	2184 (9.51%)	5899 (9.1%)	
4	4173 (9.97%)	3839 (16.71%)	8012 (12.36%)	
5+	5219 (12.47%)	5216 (22.70%)	10435 (16.1%)	

[1]Wilcoxon rank sum test [2]Chi-Square test

Table 4. Estimates from linear mixed model predicting the natural log of total tax for all three audit categories. Coefficients are exponentiated and represent a multiplicative change in total tax.

	Audit Category 1		**Audit Category 2**		**Audit Category 3**	
Variable	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*
Audited	0.763 (0.746, 0.780)	0	2.624 (2.553, 2.697)	0	1.228 (1.158, 1.303)	0
Married Filing Jointly	4.660 (4.578, 4.743)	0	5.387 (5.281, 5.494)	0	11.033 (10.609, 11.474)	0
Urban zip code	1.331 (1.288, 1.375	0	1.092 (1.048, 1.138)	0	1.47 (1.37, 1.577)	0
Any wage income	2.946 (2.907, 2.987)	0	3.38 (3.327, 3.434)	0	1.993 (1.944, 2.042)	0
Itemized deductions	2.136 (2.117, 2.156)	0	NA	NA	3.262 (3.179, 3.347)	0
Mortgage interest	1 (1, 1)	0	1 (1, 1)	0	1 (1, 1)	0
Any Child Tax Credit	0.775 (0.765, 0.784)	0	0.975 (0.96, 0.991)	0.002	1.024 (0.993, 1.056)	0.14
Total exemptions						
0	Reference		Reference		Reference	
1	3.127 (2.909, 3.362)	0	3.752 (3.188, 4.415)	0	3.805 (3.384, 4.278)	0
2	1.451 (1.348, 1.563)	0	1.406 (1.194, 1.656)	0	0.949 (0.84, 1.072)	0.40
3	1.089 (1.011, 1.173)	0.026	0.851 (0.723, 1.003)	0.053	0.463 (0.409, 0.525)	0
4	0.851 (0.789, 0.918)	0	0.57 (0.483, 0.671)	0	0.178 (0.157, 0.202)	0
5+	0.538 (0.498, 0.581)	0	0.333 (0.282, 0.393)	0	0.067 (0.059, 0.076)	0
Tax Year						
2006	Reference		Reference		Reference	
2007	0.948 (0.919, 0.979)	0.0009	0.986 (0.95, 1.023)	0.46	1.018 (0.965, 1.075)	0.5054
2008	0.855 (0.829, 0.883)	0	0.601 (0.579, 0.624)	0	0.832 (0.783, 0.884)	0
2009	0.771 (0,746, 0.796)	0	0.421 (0.405, 0.438)	0	0.71 (0.666, 0.756)	0
2010	0.772 (0.746, 0.797)	0	0.47 (0.451, 0.489)	0	0.788 (0.738, 0.841)	0
2011	0.876 (0.845, 0.907)	0	0.511 (0.49, 0.533)	0	0.756 (0.705, 0.811)	0
2012	1.093 (1.052, 1.135)	0	0.607 (0.58, 0.635)	0	0.823 (0.763, 0.888)	0
2013	1.254 (1.204, 1.306)	0	0.647 (0.616, 0.679)	0	0.926 (0.851, 1.007)	0.0734
2014	1.489 (1.425, 1.555)	0	0.761 (0.722, 0.801)	0	1.13 (1.03, 1.239)	0.0094
2015	1.654 (1.578, 1.733)	0	0.735 (0.695, 0.777)	0	1.215 (1.098, 1.344)	0.0002
2016	1.741 (1.655, 1.832)	0	0.687 (0.647, 0.73)	0	1.236 (1.107, 1.38)	0.0002
2017	1.944 (1.841, 2.052)	0	0.806 (0.756, 0.859)	0	1.323 (1.174, 1.491)	0
2018	2.267 (2.138, 2.403)	0	0.997 (0.93, 1.068)	0.93	1.598 (1.401, 1.822)	0
Priority at Baseline	1 (1, 1)	0.47	0.997 (0.996, 0.997)	0	1 (1, 1)	0.024
TPI	1 (1, 1)	0	1 (1, 1)	0	1 (1, 1)	0
Years After Baseline						
Year 0 (baseline)	Reference		Reference		Reference	

Table 4. Estimates from linear mixed model predicting the natural log of total tax for all three audit categories. Coefficients are exponentiated and represent a multiplicative change in total tax.

	Audit Category 1		**Audit Category 2**		**Audit Category 3**	
Variable	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*
Year 1	0.709 (0.698, 0.721)	0	1.719 (1.681, 1.759)	0	0.81 (0.772, 0.849)	0
Year 2	0.631 (0.620, 0.643)	0	2.091 (2.04, 2.143)	0	0.81 (0.769, 0.853)	0
Year 3	0.591 (0.578, 0.603)	0	2.372 (2.308, 2.438)	0	0.822 (0.776, 0.871)	0
Year 4	0.555 (0.542, 0.568)	0	2.482 (2.408, 2.559)	0	0.808 (0.757, 0.862)	0
Year 5	0.508 (0.495, 0.522)	0	2.573 (2.487, 2.663)	0	0.76 (0.707, 0.817)	0
Year 6	0.467 (0.453, 0.482)	0	2.693 (2.593, 2.797)	0	0.734 (0.677, 0.797)	0
Year 7	0.432 (0.417, 0.447)	0	2.752 (2.637, 2.873)	0	0.671 (0.61, 0.738)	0
Year 8	0.404 (0.388, 0.420)	0	2.765 (2.635, 2.901)	0	0.674 (0.602, 0.754)	0
Audited*Years After Baseline						
Audited*Year 1	1.198 (1.171, 1.226)	0	0.647 (0.629, 0.665)	0	1.139 (1.073, 1.208)	0
Audited*Year 2	1.76 (1.720, 1.800)	0	0.803 (0.78, 0.826)	0	1.275 (1.198, 1.357)	0
Audited*Year 3	2.008 (1.962, 2.055)	0	0.731 (0.711, 0.753)	0	1.202 (1.128, 1.282)	0
Audited*Year 4	1.983 (1.938, 2.030)	0	0.664 (0.644, 0.684)	0	1.206 (1.13, 1.287)	0
Audited*Year 5	1.956 (1.911, 2.003)	0	0.6 (0.582, 0.618)	0	1.261 (1.18, 1.347)	0
Audited*Year 6	1.943 (1.897, 1.990)	0	0.54 (0.524, 0.557)	0	1.241 (1.158, 1.33)	0
Audited*Year 7	1.874 (1.827, 1.922)	0	0.49 (0.474, 0.506)	0	1.275 (1.181, 1.378)	0
Audited*Year 8	1.769 (1.723, 1.817)	0	0.453 (0.437, 0.469)	0	1.214 (1.108, 1.33)	0

Table 5. Estimates from linear mixed model predicting the natural log of relevant items for all three audit categories. Coefficients are exponentiated and represent a multiplicative change in relevant items reporting.

	Audit Category 1		**Audit Category 2**		**Audit Category 3**	
Variable	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*
Audited	309.42 (299.99, 319.15)	0	0.886 (0.861, 0.912)	0	1.054 (1.006, 1.105)	0.026
Married Filing Jointly	2.102 (2.051, 2.154)	0	3.398 (3.327, 3.471)	0	1.54 (1.492, 1.589)	0
Urban zip code	0.988 (0.944, 1.034)	0.61	1.069 (1.022, 1.117)	0.0032	1.075 (1.014, 1.139)	0.014
Any wage income	0.439 (0.431, 0.447)	0	1.176 (1.156, 1.196)	0	0.261 (0.256, 0.266)	0
Itemized deductions	0.904 (0.893, 0.916)	0	NA	NA	0.916 (0.897, 0.936)	0
Mortgage interest	1 (1, 1)	0.47	1 (1, 1)	0	1 (1, 1)	0
Any Child Tax Credit	0.832 (0.818, 0.846)	0	0.7 (0.688, 0.712)	0	0.867 (0.845, 0.889)	0
Total Exemptions						
0	Reference		Reference		Reference	
1	2.595 (2.348, 2.869)	0	3.187 (2.669, 3.807)	0	1.523 (1.379, 1.682)	0
2	2.779 (2.509, 3.078)	0	2.124 (1.778, 2.539)	0	1.556 (1.404, 1.724)	0
3	3 (2.705, 3.327)	0	1.847 (1.545, 2.208)	0	1.937 (1.744, 2.151)	0
4	2.961 (2.666, 3.287)	0	1.709 (1.428, 2.044)	0	1.785 (1.606, 1.984)	0
5+	2.732 (2.455, 3.041)	0	1.497 (1.25, 1.794)	0	1.668 (1.5, 1.856)	0
Tax Year						
2006	Reference		Reference		Reference	
2007	0.994 (0.952, 1.038)	0.80	0.9 (0.865, 0.938)	0	0.993 (0.949, 1.04)	0.78
2008	1.087 (1.040, 1.135)	0.0002	0.839 (0.805, 0.875)	0	0.861 (0.818, 0.907)	0
2009	1.092 (1.045, 1.142)	0.0001	0.736 (0.705, 0.768)	0	0.831 (0.788, 0.875)	0
2010	1.118 (1.068, 1.171)	0	0.851 (0.815, 0.889)	0	0.86 (0.815, 0.907)	0
2011	1.186 (1.129, 1.246)	0	0.876 (0.837, 0.917)	0	0.831 (0.786, 0.879)	0
2012	1.252 (1.188, 1.319)	0	0.881 (0.84, 0.924)	0	0.887 (0.836, 0.942)	0.0001
2013	1.406 (1.329, 1.489)	0	0.965 (0.916, 1.016)	0.17	0.837 (0.784, 0.894)	0
2014	1.515 (1.425, 1.611)	0	1.159 (1.097, 1.224)	0	0.845 (0.786, 0.907)	0
2015	1.649 (1.543, 1.761)	0	1.367 (1.288, 1.45)	0	0.86 (0.796, 0.93)	0.0002
2016	1.712 (1.595, 1.838)	0	1.604 (1.506, 1.708)	0	0.862 (0.792, 0.938)	0.0006
2017	1.786 (1.655, 1.926)	0	1.866 (1.745, 1.995)	0	0.849 (0.775, 0.93)	0.0004
2018	1.775 (1.636, 1.927)	0	0.083 (0.077, 0.089)	0	0.783 (0.708, 0.866)	0
TPI	1 (1, 1)	0.54	1 (1, 1)	0	1 (1, 1)	0.67
Years After Baseline						
Year 0 (baseline)	Reference		Reference		Reference	
Year 1	0.432 (0.422, 0.442)	0	0.312 (0.304, 0.319)	0	3.601 (3.459, 3.749)	0

Table 5. Estimates from linear mixed model predicting the natural log of relevant items for all three audit categories. Coefficients are exponentiated and represent a multiplicative change in relevant items reporting.						
	Audit Category 1		**Audit Category 2**		**Audit Category 3**	
Variable	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*
Year 2	0.284 (0.277, 0.292)	0	0.165 (0.161, 0.17)	0	4.237 (4.056, 4.425)	0
Year 3	0.22 (0.214, 0.226)	0	0.106 (0.103, 0.109)	0	4.651 (4.433, 4.879)	0
Year 4	0.182 (0.176, 0.188)	0	0.077 (0.074, 0.079)	0	4.851 (4.601, 5.113)	0
Year 5	0.155 (0.150, 0.161)	0	0.059 (0.056, 0.061)	0	4.794 (4.524, 5.081)	0
Year 6	0.133 (0.127, 0.139)	0	0.05 (0.048, 0.052)	0	4.785 (4.484, 5.107)	0
Year 7	0.119 (0.113, 0.125)	0	0.034 (0.033, 0.036)	0	4.719 (4.376, 5.087)	0
Year 8	0.105 (0.099, 0.110)	0	0.028 (0.026, 0.029)	0	4.99 (4.562, 5.459)	0
Audited*Years After Baseline						
Audited*Year 1	0.168 (0.163, 0.173)	0	0.924 (0.896, 0.953)	0	1.568 (1.49, 1.649)	0
Audited*Year 2	0.038 (0.036, 0.039)	0	0.415 (0.402, 0.428)	0	1.772 (1.68, 1.869)	0
Audited*Year 3	0.018 (0.018, 0.019)	0	0.367 (0.355, 0.379)	0	1.483 (1.405, 1.566)	0
Audited*Year 4	0.015 (0.015, 0.016)	0	0.393 (0.38, 0.405)	0	1.301 (1.231, 1.375)	0
Audited*Year 5	0.014 (0.013, 0.014)	0	0.442 (0.428, 0.457)	0	1.251 (1.182, 1.324)	0
Audited*Year 6	0.013 (0.013, 0.013)	0	0.4 (0.387, 0.413)	0	1.178 (1.111, 1.25)	0
Audited*Year 7	0.012 (0.012, 0.013)	0	0.494 (0.476, 0.511)	0	1.221 (1.143, 1.305)	0
Audited*Year 8	0.012 (0.012, 0.013)	0	0.55 (0.529, 0.571)	0	1.13 (1.045, 1.222)	0.002

Table 6. Estimates from linear mixed models for sensitivity analysis of Audit Category 1. Coefficients are exponentiated and represent a multiplicative change in relevant items reporting.

	Outcome: Total Tax		**Outcome: Relevant Line Items**	
Variable	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*
Audited	0.688 (0.616, 0.768)	0	1.825 (1.564, 2.129)	0
Married Filing Jointly	4.201 (3.872, 4.558)	0	2.512 (2.239, 2.818)	0
Urban zip code	1.14 (0.987, 1.317)	0.075	0.977 (0.797, 1.196)	0.82
Any wage income	2.799 (2.639, 2.97)	0	0.435 (0.4, 0.473)	0
Itemized deductions	1.873 (1.79, 1.959)	0	0.919 (0.863, 0.98)	0.010
Mortgage interest	1 (1, 1)	0	1 (1, 1)	0.0001
Any Child Tax Credit	0.911 (0.861, 0.964)	0.0013	0.865 (0.798, 0.937)	0.0004
Total Exemptions				
0	Reference		Reference	
1	3.813 (2.293, 6.339)	0	1.913 (0.931, 3.931)	0.0777
2	1.735 (1.038, 2.9)	0.0354	1.795 (0.867, 3.715)	0.1151
3	1.266 (0.755, 2.122)	0.3712	2.002 (0.963, 4.162)	0.0632
4	0.855 (0.508, 1.438)	0.5547	2.055 (0.984, 4.291)	0.0552
5+	0.508 (0.3, 0.859)	0.0115	1.714 (0.814, 3.608)	0.1558
Tax Year				
2006	Reference		Reference	
2007	1.171 (0.977, 1.404)	0.0883	0.759 (0.586, 0.982)	0.0355
2008	0.91 (0.759, 1.09)	0.3069	0.764 (0.591, 0.988)	0.04
2009	0.783 (0.651, 0.941)	0.009	0.691 (0.532, 0.896)	0.0053
2010	0.748 (0.62, 0.904)	0.0026	0.888 (0.68, 1.158)	0.3798
2011	0.78 (0.641, 0.949)	0.0129	1.038 (0.788, 1.368)	0.7915
2012	0.864 (0.705, 1.058)	0.1568	1.075 (0.808, 1.429)	0.6197
2013	0.966 (0.778, 1.198)	0.7516	1.335 (0.986, 1.807)	0.0615
2014	1.082 (0.862, 1.358)	0.4986	1.506 (1.095, 2.072)	0.0118
2015	1.189 (0.934, 1.513)	0.1596	1.772 (1.265, 2.482)	0.0009
2016	1.189 (0.922, 1.535)	0.1824	1.877 (1.315, 2.68)	0.0005
2017	1.261 (0.963, 1.652)	0.0913	2.014 (1.383, 2.932)	0.0003
TPI	1 (1, 1)	0	1 (1, 1)	0.81
Priority at baseline	1 (1, 1)	0.27	NA	NA
Year 0 (baseline)	Reference		Reference	
Year 1	0.785 (0.733, 0.841)	0	1.825 (1.564, 2.129)	0
Year 2	0.794 (0.734, 0.857)	0	0.112 (0.102, 0.124)	0
Year 3	0.824 (0.754, 0.901)	0	0.043 (0.039, 0.048)	0
Year 4	0.794 (0.716, 0.88)	0	0.021 (0.019, 0.024)	0
Year 5	0.764 (0.678, 0.86)	0	0.014 (0.012, 0.016)	0

Table 6. Estimates from linear mixed models for sensitivity analysis of Audit Category 1. Coefficients are exponentiated and represent a multiplicative change in relevant items reporting.

	Outcome: Total Tax		**Outcome: Relevant Line Items**	
Variable	*Estimate (95% CI)*	*p-value*	*Estimate (95% CI)*	*p-value*
Year 6	0.745 (0.65, 0.854)	0	0.01 (0.008, 0.012)	0
Year 7	0.719 (0.616, 0.839)	0	0.007 (0.006, 0.009)	0
Year 8	0.692 (0.581, 0.823)	0	0.005 (0.004, 0.006)	0
Audited*Year 1	1.207 (1.067, 1.365)	0.0027	0.799 (0.671, 0.952)	0.012
Audited*Year 2	1.708 (1.511, 1.929)	0	0.331 (0.278, 0.394)	0
Audited*Year 3	1.902 (1.698, 2.131)	0	0.208 (0.177, 0.244)	0
Audited*Year 4	1.867 (1.654, 2.106)	0	0.238 (0.2, 0.282)	0
Audited*Year 5	1.947 (1.724, 2.198)	0	0.261 (0.22, 0.311)	0
Audited*Year 6	1.94 (1.716, 2.195)	0	0.311 (0.261, 0.371)	0
Audited*Year 7	1.975 (1.738, 2.244)	0	0.345 (0.288, 0.414)	0
Audited*Year 8	1.836 (1.564, 2.154)	0	0.37 (0.294, 0.464)	0

Figures

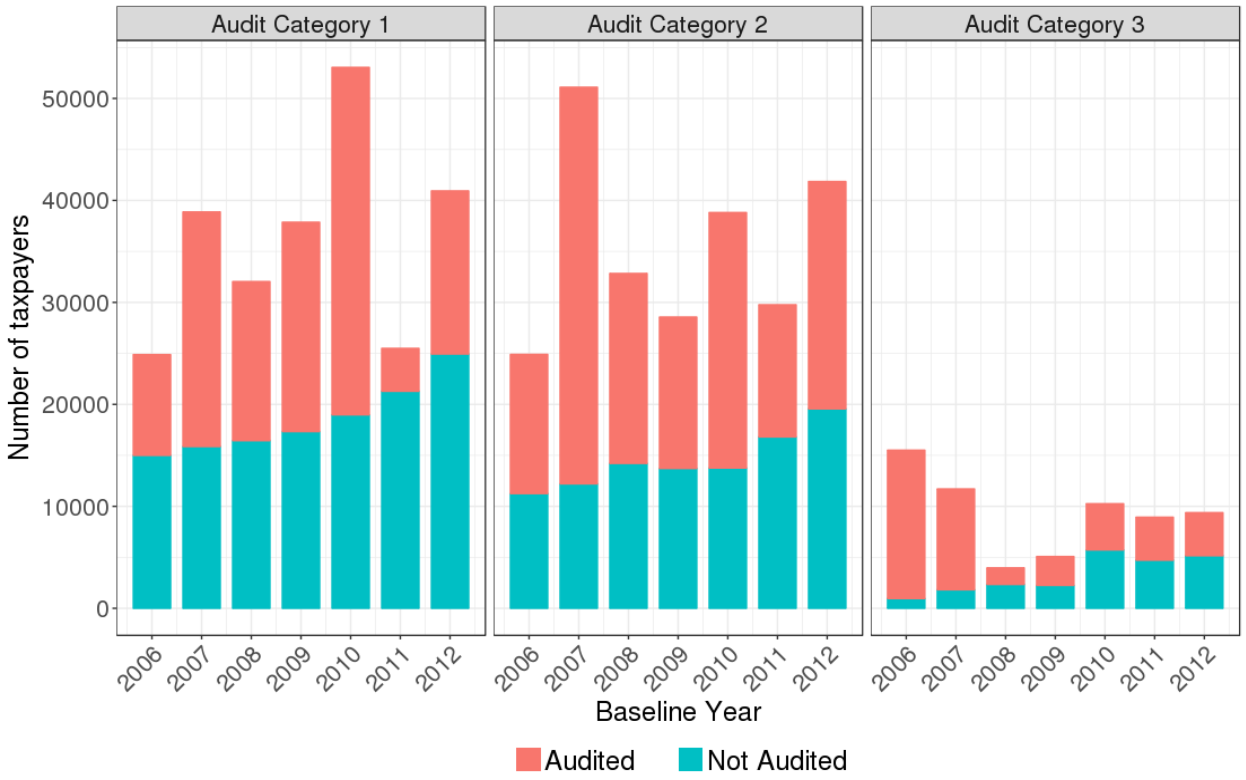

Figure 1. Sample sizes for all baseline years and audit categories. Red bars indicate frequency of audited taxpayers and blue bars indicate frequency of not audited taxpayers.

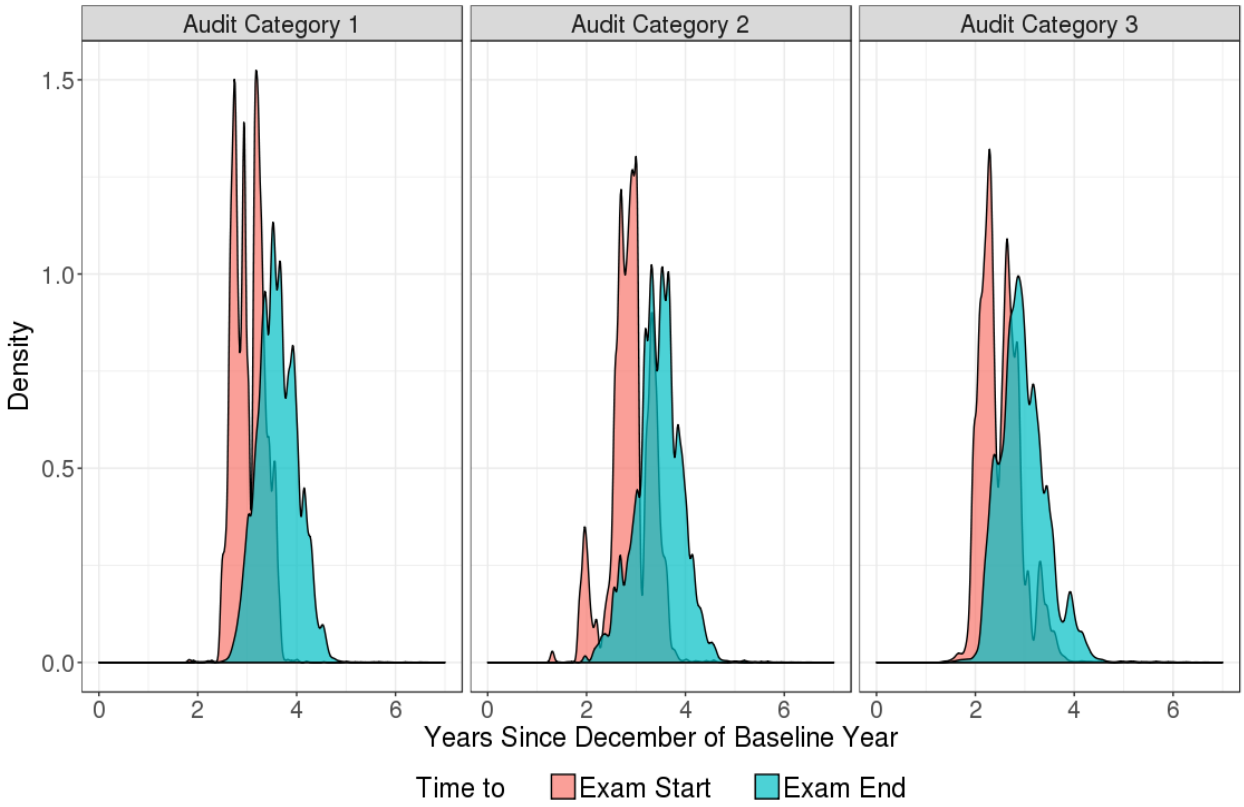

Figure 2. Considering the audited taxpayers only, density plots of the timing of audit exam start and end dates, relative to December of the TY of the audited return.

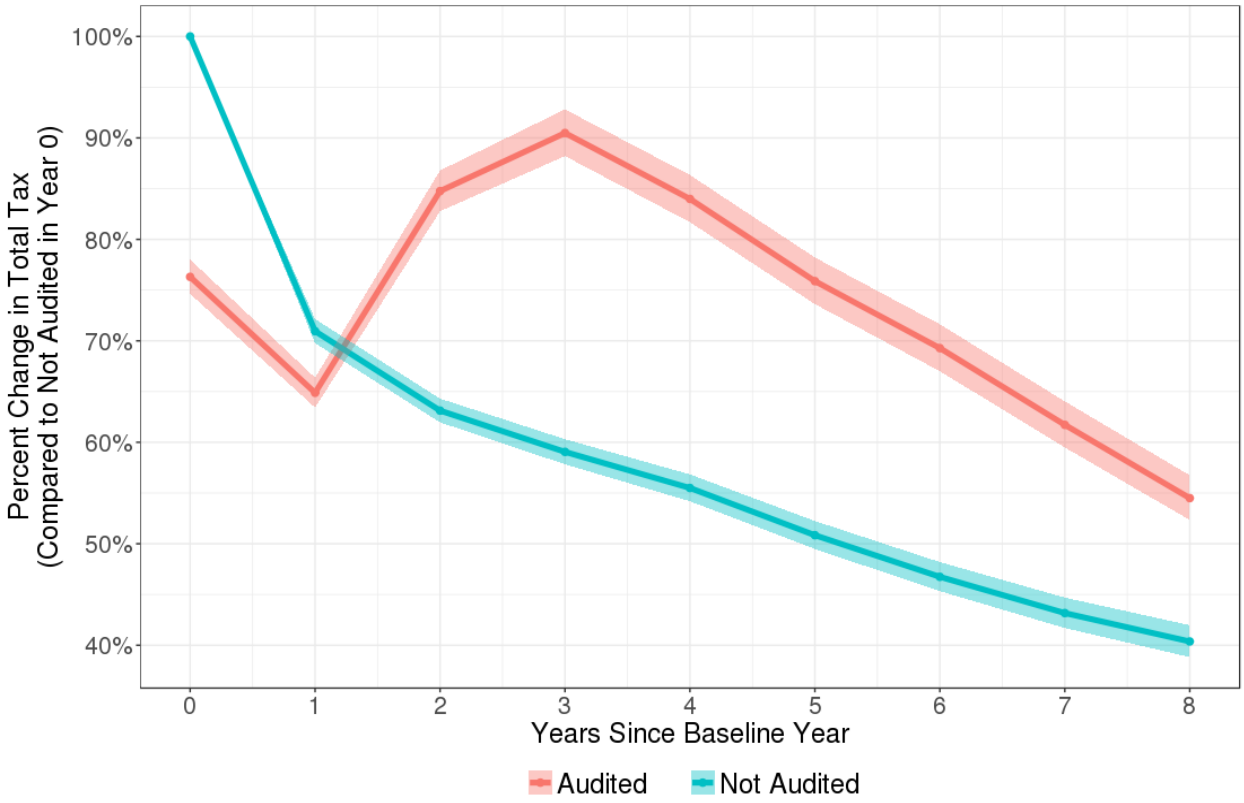

Figure 3. Predicted values for the linear mixed effects model of ln(total tax) in audit category 1. Red dots are estimates for the audited group while blue dots are estimates for the not audited group. Shading represents 95% confidence intervals. Not audited, year 0 is the reference group.

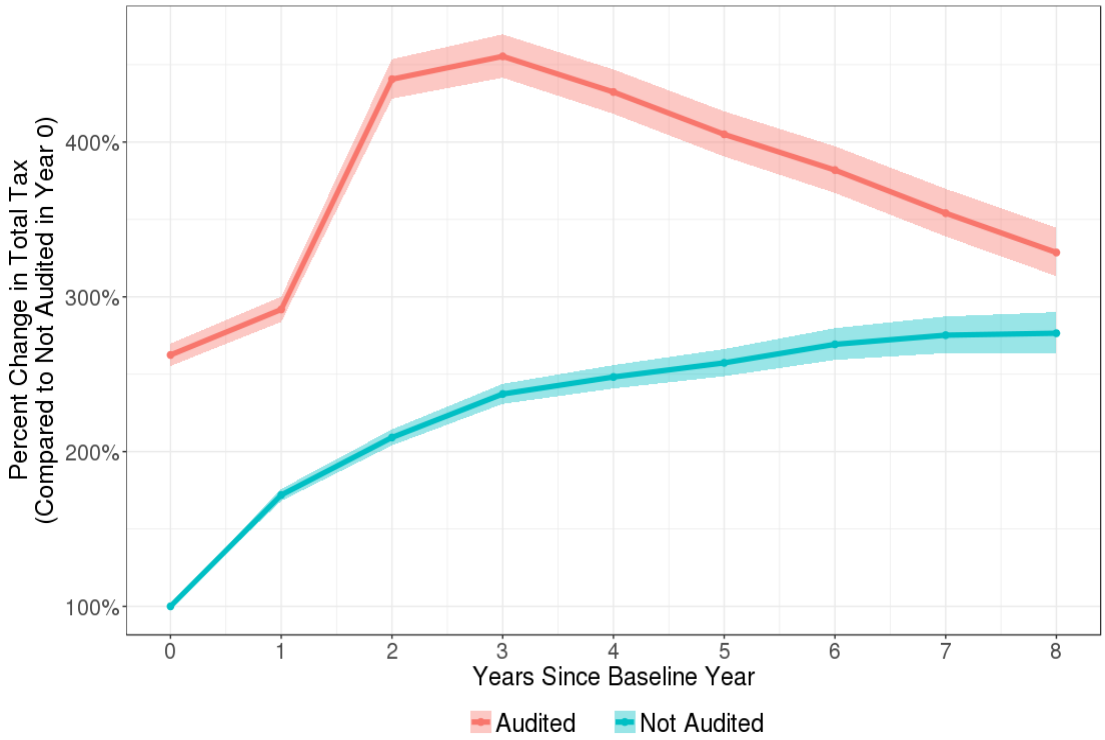

Figure 4. Predicted values for the linear mixed effects model of ln(total tax) in audit category 2. Red dots are estimates for the audited group while blue dots are estimates for the not audited group. Shading represents 95% confidence intervals. Not audited, year 0 is the reference group.

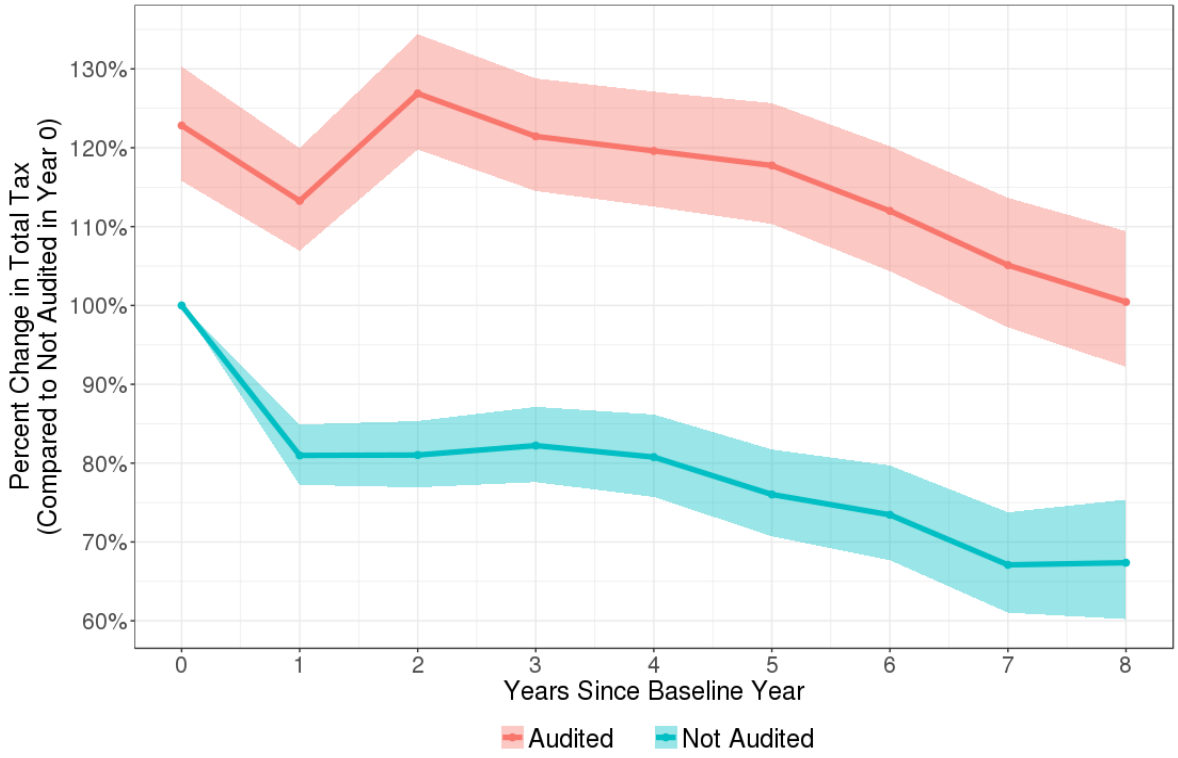

Figure 5. Predicted values for the linear mixed effects model of ln(total tax) in audit category 3. Red dots are

estimates for the audited group while blue dots are estimates for the not audited group. Shading represents 95% confidence intervals. Not audited, year 0 is the reference group.

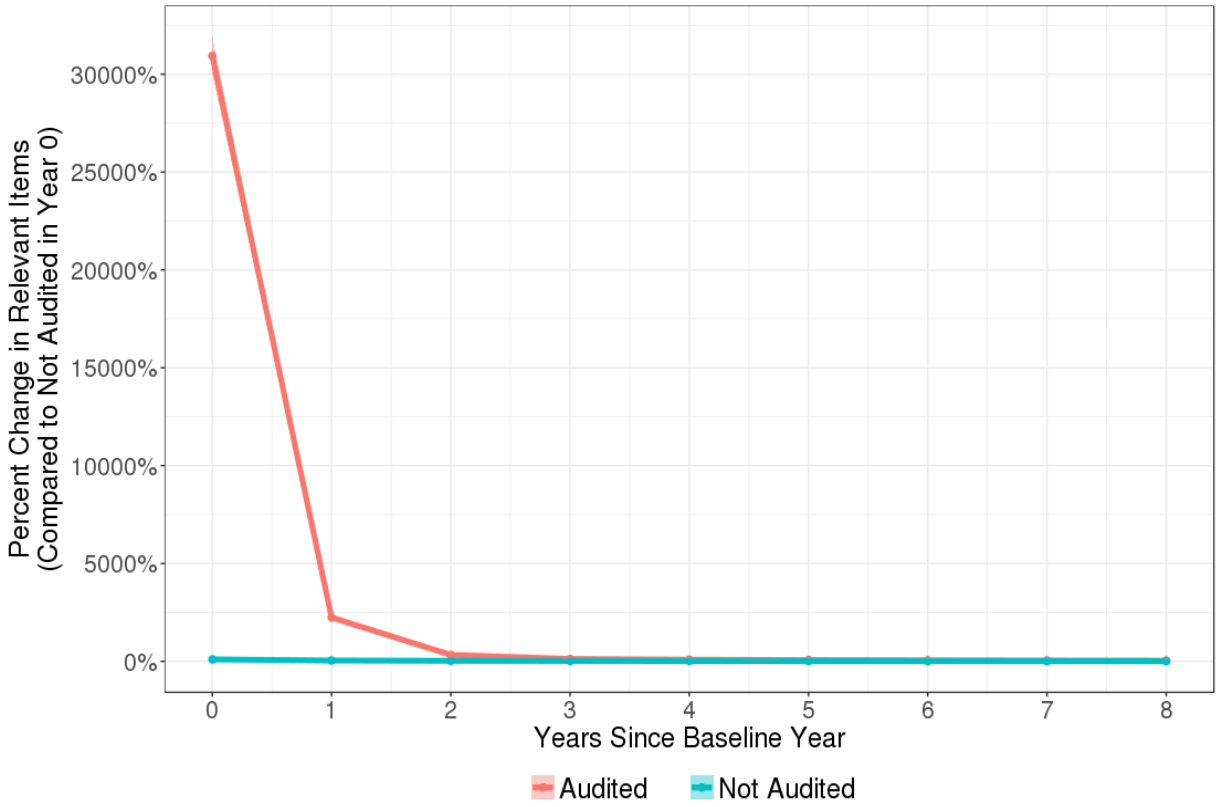

Figure 6. Predicted values for the linear mixed effects model of ln(relevant items) in audit category 1. Red dots are estimates for the audited group while blue dots are estimates for the not audited group. Shading represents 95% confidence intervals. Not audited, year 0 is the reference group.

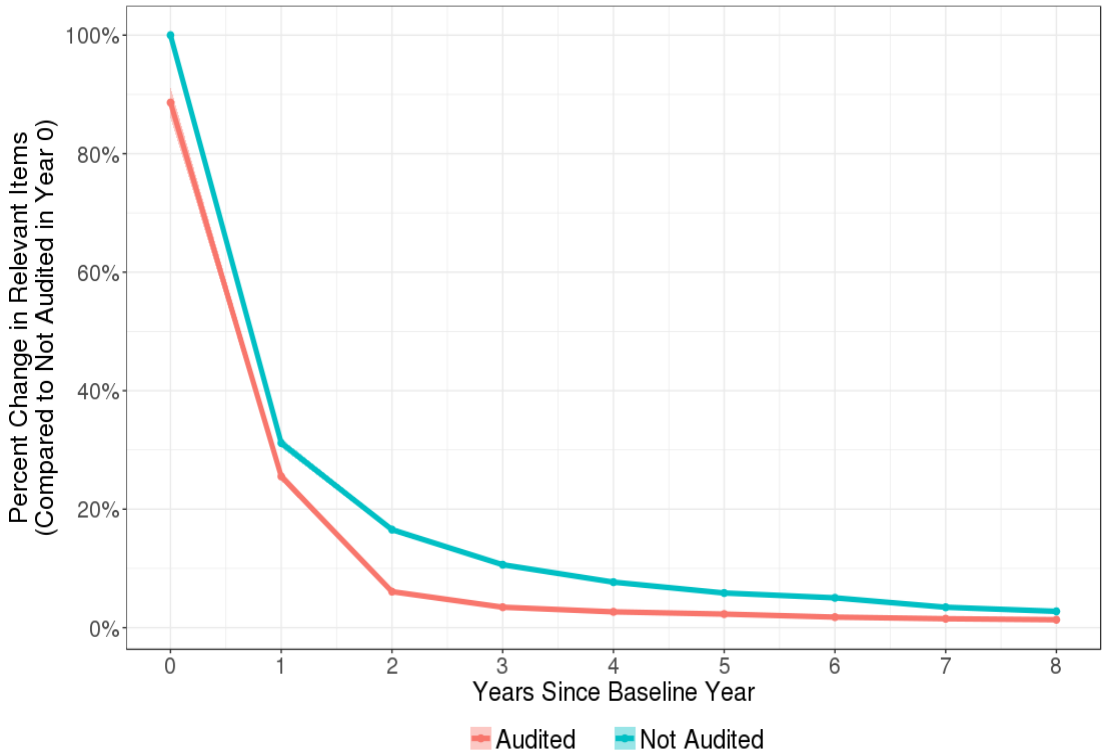

Figure 7. Predicted values for the linear mixed effects model of ln(relevant items) in audit category 2. Red dots are estimates for the audited group while blue dots are estimates for the not audited group. Shading represents 95% confidence intervals. Not audited, year 0 is the reference group.

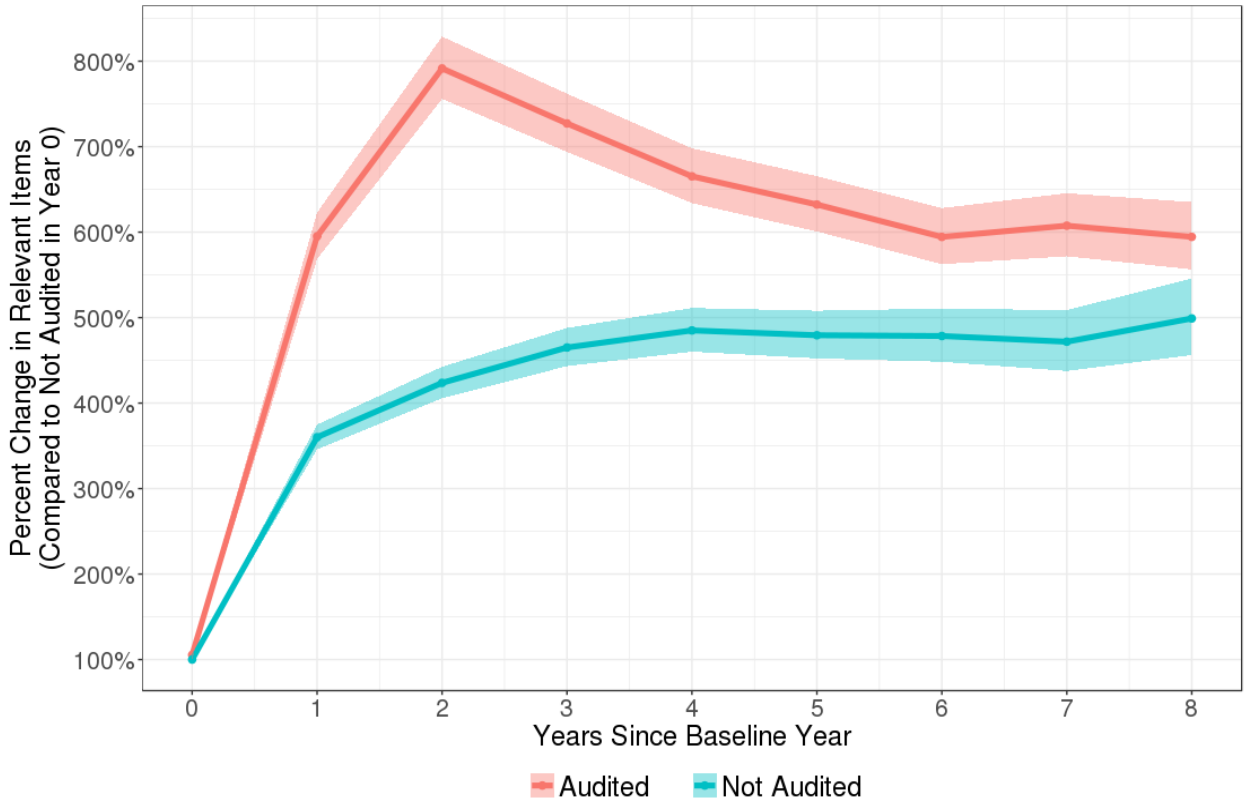

Figure 8. Predicted values for the linear mixed effects model of ln(relevant items) in audit category 3. Red dots are estimates for the audited group while blue dots are estimates for the not audited group. Shading represents 95% confidence intervals. Not audited, year 0 is the reference group.

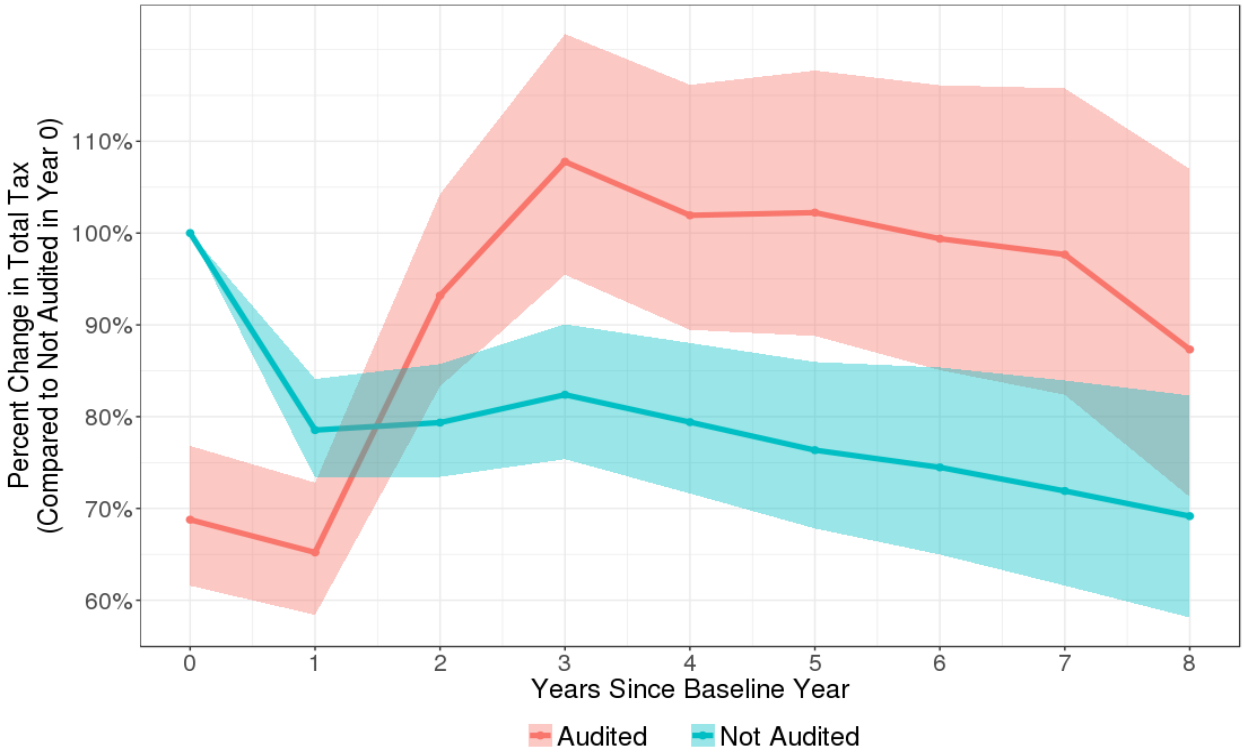

Figure 9. Predicted values for the linear mixed effects model of ln(relevant items) in sensitivity analysis of audit category 1. Red dots are estimates for the audited group while blue dots are estimates for the not audited group. Shading represents 95% confidence intervals. Not audited, year 0 is the reference group.

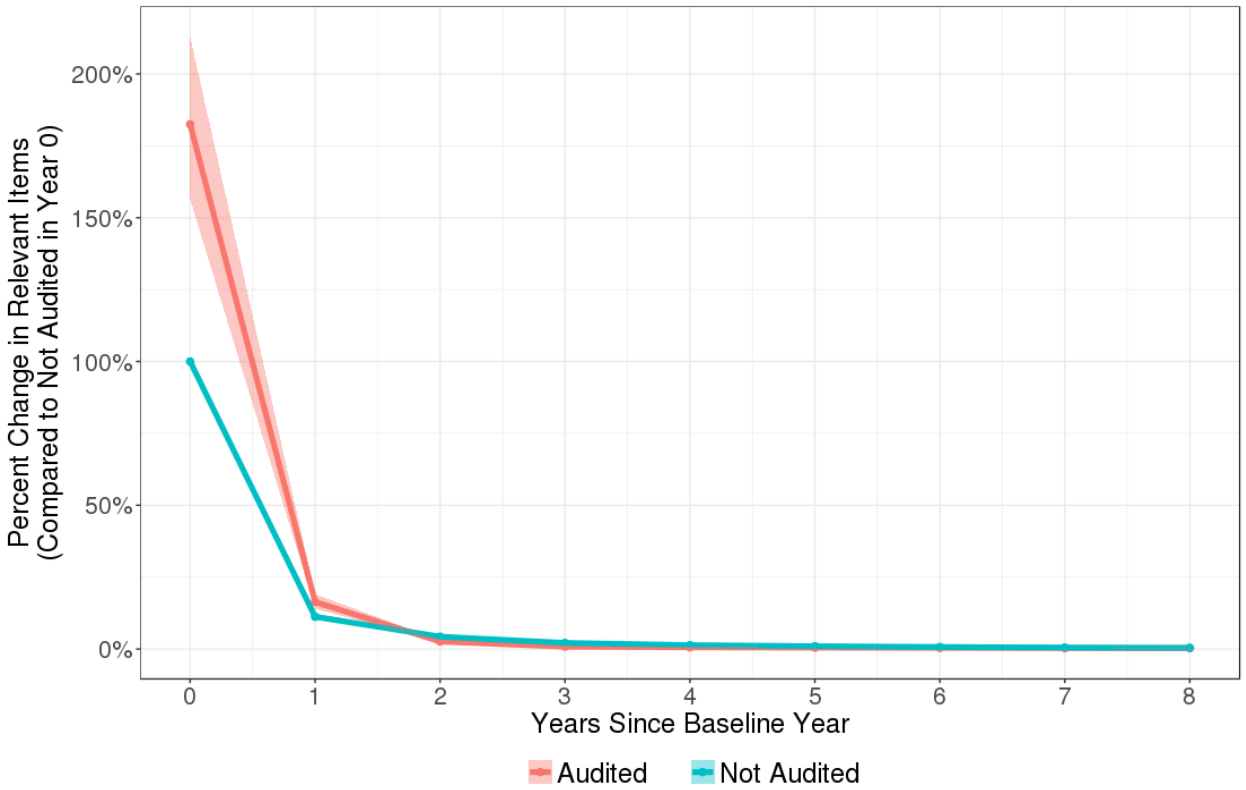

Figure 10. Predicted values for the linear mixed effects model of ln(relevant items) in sensitivity analysis of audit category 1. Red dots are estimates for the audited group while blue dots are estimates for the not audited group. Shading represents 95% confidence intervals. Not audited, year 0 is the reference group.