Volume 2

TAS Research and Related Studies

Audit Impact Study[1]

1 Throughout this study, the term DD refers to difference-in-differences.

EXECUTIVE SUMMARY: AUDIT IMPACT STUDY

SEBASTIAN BEER, MATTHIAS KASPER, ERICH KIRCHLER, AND BRIAN ERARD

1. Introduction

The IRS audits roughly 1.5 percent of all self-employed individual income taxpayers annually. In fiscal year 2014, the direct effect of these audits was over $3 billion in recommended additional tax assessments, although not all of the recommended amount will ultimately be collected (Internal Revenue Service, 2015).[1] Less is known, however, about the indirect long-term effect of audits on subsequent taxpayer reporting behavior. Behavioral changes may either undermine immediate gains in tax collections or further increase the revenue returns of audits. Depending on risk attitudes, norms, moral perceptions, and perhaps most importantly, the subjective appraisal of the audit, enforcement activity has the potential to increase or decrease the willingness to comply with the law and to cooperate with the IRS in the future.

This report evaluates the impact of enforcement activity on the subsequent compliance behavior of nonfarm self-employed taxpayers. Through a statistical comparison of administrative data for a random sample of 2,204 Schedule C filers with under $200,000 in total positive income who were audited subsequent to filing their TY2007 returns with data for a control sample of 4,705 who were not audited, we are able to estimate the short- and long-term impact of audits on tax collections. In our empirical analysis, we distinguish between (seemingly) compliant and (seemingly) noncompliant taxpayers, as the audit response likely differs between these groups. A "direct deterrent effect" (Alm et al., 2009) of additional tax assessments potentially increases the compliance of caught evaders. The response of compliant taxpayers to enforcement activity is ambiguous, however. Audits could be seen as a justified means to enforce the law, increasing the trust in the state and the willingness to comply voluntarily. A coercive experience might have the opposite outcome.

2. Empirical results

Following Gemmel and Ratto (2012) we attempt to distinguish compliant from noncompliant taxpayers on the basis of their audit outcomes. More specifically, we classify taxpayers as compliant if the examination did not result in a recommended additional tax assessment

[1]These figures include both farm and nonfarm business returns; however, returns claiming the Earned Income Credit are excluded as audit coverage statistics for this category do not distinguish between business and non-business returns.

and noncompliant otherwise. This categorization procedure has two important drawbacks. One is that we may only classify audited taxpayers. The second is related to classification errors. Some truly noncompliant taxpayers are likely not detected during an audit and not assessed additional tax. Conversely, some additional tax assessments may be unwarranted and disputed later on. Accordingly, the examination result does not unambiguously signal the subjective inclination to pay taxes voluntarily. We address related concerns below in our discussion of the results.

FIGURE 1

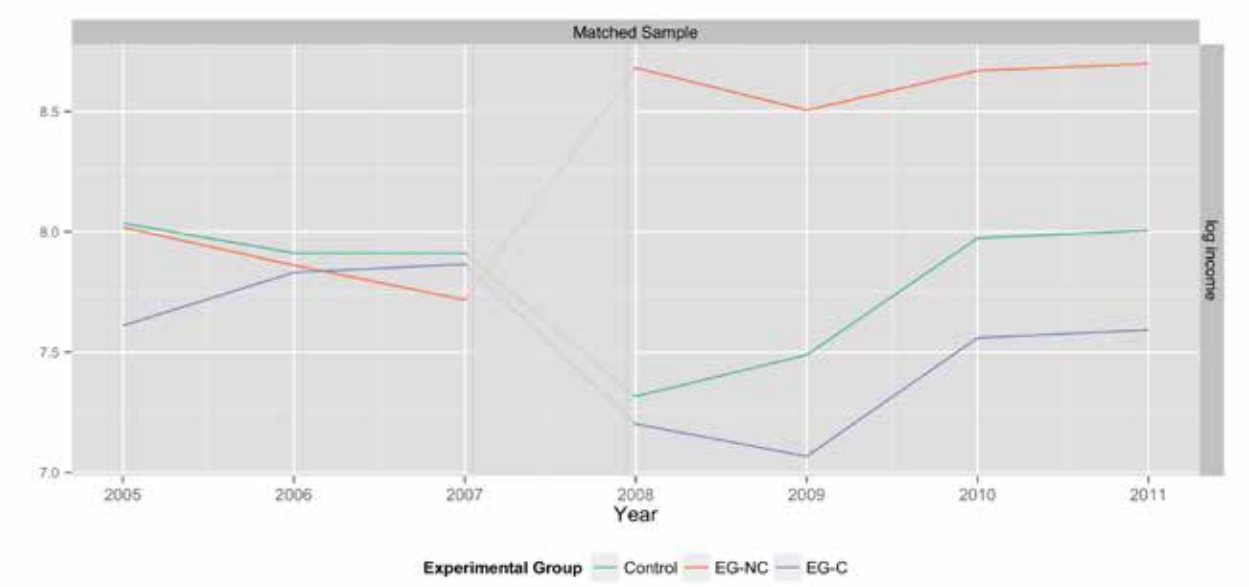


Figure 1 illustrates the impact of audits, conducted between the filing of the TY2007 and TY2008 returns, on the income reporting trends of compliant (blue line) and noncompliant (orange line) self-employed taxpayers. The green line depicts the trend in reported taxable income for an unaudited control group, serving as a benchmark. We rely on a range of non-experimental estimators to refine this comparison and quantify the magnitude of the short- and long-run audit impact. These include the standard difference-in-differences estimator, variants of this method that account for sample selection and attrition,[2], and propensity score matching methods. While propensity score matching overcomes observable differences between our experimental groups, the difference-in-differences approach accounts for unobservable, time-constant effects. It is reassuring that these two alternative approaches yield similar results.

The estimates indicate an enduring effect of audits on taxpayers who receive a positive recommended additional tax assessment. On average, such taxpayers increase their reported

[2]We find that enforcement activity reduces the future likelihood of filing Schedule C by almost 7% among taxpayers who receive a positive recommended additional tax assessment.

taxable income by 250% following an audit. Three years after the audit, the effect is still substantial with an average increase of 120%. Importantly, the results also indicate that audits have a detrimental long-term impact on the reporting behavior of taxpayers who do not experience an additional tax assessment. Three years after having undergone enforcement activity, these taxpayers report around 35% less in taxable income than the control group. The difference is significant at the 1% level. When we employ a less nuanced model that does not distinguish audited taxpayers on the basis of their audit outcome, we find that, on net, reported taxable income for this combined group increases by roughly 20% three years after an audit.

3. Discussion

Our empirical results provide robust evidence that audits have important long-term revenue implications. Three years after an audit, the average small business taxpayer reports around 20% more income.[3] The indirect long-term effect thus clearly adds to the static gain of additional tax assessments. However, by differentiating the response of compliant and noncompliant taxpayers, we find that there is scope for improving the revenue efficiency of audits.

Our more nuanced analysis of the behavioral response to an audit shows that taxpayers who receive a positive additional recommended tax assessment increase their subsequent reporting of taxable income dramatically (+120%), while those who receive no additional tax assessment actually report less (-35%). The positive impact of audits on the former group is likely due to some kind of specific deterrent effect (Alm et al., 2009).

Understanding the observed reduction in reported income among taxpayers in the latter group is probably even more important. There are several plausible explanations for this finding. First, an experience of coercive enforcement activity could reduce tax morale among honest taxpayers, leading to the observed detrimental impact of audits within this group. Second, even if tax morale were unaffected by the examination experience, the audit process might provide currently compliant taxpayers with a "window" on potential opportunities for both legal and illegal tax avoidance. In addition, such taxpayers may infer that the risk of a future examination is low given that no adjustments were made during the recent audit. This newfound awareness of opportunities for reporting and paying lower taxes combined with a low perceived future audit risk could drive some taxpayers to report less income on subsequent returns. A third possibility is that the observed reduction in reported income might be attributable to dishonest taxpayers within this group whose misreporting was not detected during the audit. The experience of having undergone an audit without experiencing

[3]This estimate is substantially larger than that obtained by (DeBacker et al., 2015), perhaps owing to our focus on operational rather than random audits.

any sanction for noncompliance may have emboldened such taxpayers, resulting in even more aggressive future reporting behavior.

Based on the available data, we are unable to pinpoint which of these explanations prevails. The observed reduction in compliance behavior suggests, in any case, that there is scope for improving the efficiency of audits. On the one hand, improved targeting of noncompliant returns and an improved capacity to detect noncompliance would seem likely to improve deterrence among cheaters. On the other hand, a better understanding of the psychological impact of audits on compliant taxpayers may lead to enhanced examination approaches that mitigate the erosion of tax morale and maintain their incentives to comply.

4. Limitations and scope for future work

We conclude by noting a range of limitations of and possible extensions of our current research design. First, we cannot rule out that our estimates are influenced by the economic downturn in 2008. Repeating the analysis for another, less turbulent, timespan would strengthen the credibility of the results. Second, the relatively short time horizon impedes the estimation of a dynamic model, which would allow a more accurate quantification of the decay rate of audit effects. Third, a range of additional analyses, looking at, for instance, the differential impact of alternative audit techniques (such as face-to-face vs. correspondence) or the differential response of low-income and high-income taxpayers, could provide important insights. Fourth, more sophisticated propensity score matching methods would provide further evidence on the robustness of our results and could improve the representativeness of our findings.

References

Alm, J., B. R. Jackson, and M. McKee (2009, April). Getting the Word Out: Enforcement Information Dissemination and Compliance Behavior. *Journal of Public Economics 93*(3-4), 392–402.

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2015, March). Once Bitten, Twice Shy? The Lasting Impact of IRS Audits on Individual Tax Reporting. Technical report.

Gemmel, N. and M. Ratto (2012, March). Behavioral Responses to Taxpayer Audits: Evidence from Random Taxpayer Inquiries. *National Tax Journal 65*(1), 33–58.

Internal Revenue Service (2015, March). *Internal Revenue Service Data Book, 2014*.

AUDIT IMPACT STUDY

1. Introduction

The IRS audits roughly 1.5 percent of all self-employed individual income taxpayers annually. In fiscal year 2014, the direct effect of these audits was over $3 billion in recommended additional tax assessments, although not all of the recommended amount will ultimately be collected (Internal Revenue Service, 2015).[1] Less is known, however, about the indirect long-term effect of audits on subsequent taxpayer reporting behavior. Behavioral changes may either undermine immediate gains in tax collections or further increase the revenue returns of audits. Depending on risk attitudes, norms, moral perceptions, and perhaps most importantly, the subjective appraisal of the audit, enforcement activity has the potential to increase or decrease the willingness of taxpayers to comply with the law and to cooperate with the IRS in the future.

This report evaluates the impact of enforcement activity on the subsequent compliance behavior of nonfarm self-employed taxpayers. Through a statistical comparison of administrative data for a random sample of 2,204 Schedule C filers with under $200,000 in total positive income who were audited subsequent to filing their TY2007 returns with data for a control sample of 4,705 who were not audited, we are able to estimate the short- and long-term impact of audits on tax collections.[2]

In contrast to other recently published studies (e.g., DeBacker et al., 2015; Advani et al., 2015) that have examined the subsequent reporting behavior of taxpayers who were randomly selected for audit, the focus of this study is on taxpayers selected through an ordinary operational audit process. Our focus on operational rather than random audits allows us to identify the average treatment effect on the treated (ATT), rather than the average treatment effect (ATE) in the general population. Operational audits tend to be targeted towards tax returns with a high potential for noncompliance. Given that the response of compliant taxpayers to an audit likely differs from the response of noncompliant taxpayers, the ATT is unlikely to coincide with the ATE. Furthermore, in the random audit studies, taxpayers were aware that they had been chosen at random for a special study, which is unlikely to elicit the same sort of reaction as knowledge of having been targeted through the usual operational audit process.

In our theoretical analysis, we distinguish between compliant and noncompliant taxpayers.[3] A "direct deterrent effect" (Alm et al., 2009) of additional tax assessments potentially increases the compliance of caught evaders. The response of compliant taxpayers to enforcement activity is

This research was conducted for the National Taxpayer Advocate (NTA) by *B. Erard and Associates* under contract TIRNO-14-E-00030. This study was conducted by Sebastian Beer, Matthias Kasper, Erich Kirchler, and Brian Erard with technical support from NTA Technical Advisors Tom Beers and Jeff Wilson. Any opinions expressed in this report are those of the authors and do not necessarily reflect the views of the NTA.

[1]These figures include both farm and nonfarm business returns; however, returns claiming the Earned Income Credit are excluded as audit coverage statistics for this category do not distinguish between business and non-business returns.

[2]Total positive income is computed by summing only the positive reported values for the following income sources (negative reported amounts are treated as zero): wages, interest, dividends, distributions, other income, Schedule C net profit, and Schedule F net profit.

[3]Note that our impact analysis covers only three years. We therefore take the subjective inclination to avoid taxes to be a personality trait (i.e., a time-constant characteristic) in our empirical analysis.

ambiguous, however. While audits could be seen as a justified means to enforce the law, increasing the trust in the state and the willingness to comply voluntarily, a coercive experience might have the opposite outcome.

Empirically, we implement this theoretical distinction between compliant and noncompliant taxpayers by employing information on actual examination results (for a similar approach see Gemmel and Ratto, 2012). More specifically, we classify taxpayers as noncompliant if the examination resulted in an additional recommended tax assessment and as compliant otherwise. This categorization procedure has two drawbacks. One is that we may only classify audited taxpayers. This impedes, for instance, selecting two separate control groups, one for compliant and one for noncompliant taxpayers. The second is that we cannot rule out classification errors among audited taxpayers. Some instances of noncompliance may go undetected during an audit, resulting in a noncompliant taxpayer being classified as compliant. Conversely, some additional recommended tax assessments may be unwarranted and disputed later on. The examination result therefore does not unambiguously signal the subjective inclination to pay taxes voluntarily. To avoid confusion between our theoretical concepts and empirical findings, we will refer to the subsample of audited taxpayers who receive an additional recommended tax assessment as positive-tax-change experimental group "E-PC" and the subsample that does not receive an additional recommended tax assessment as no-tax-change experimental group "E-NC", rather than as "noncompliant" and "compliant", respectively.

To identify the impact of audits on reporting behavior, we rely on several alternative econometric approaches, including a standard difference-in-differences estimator, variants of this method that account for sample selection and attrition, and propensity score matching methods. While propensity score matching should overcome observable differences between our experimental groups, the difference-in-differences approach also accounts for unobservable, time-constant effects. Depending on the data generating process, one or the other method provides consistent estimates of the treatment effect. Our main dependent variables are Schedule C net profit and taxable income. We obtain similar estimates from our alternative estimation approaches when examining the impact of enforcement activity on taxable income. Our estimates are less robust when focusing on Schedule C net profit.

Our empirical results indicate that audits lead to improved reporting compliance among members of the positive-tax-change experimental group (E-PC). Compared with earlier studies, the estimated effect is dramatic. Our preferred specifications are based on the natural logarithm of reported taxable income as they attach less weight to returns in the sample with very high income reports. The findings based on these specifications indicate that, one year after having undergone enforcement activity, taxpayers in group E-PC report approximately 250% more in taxable income than taxpayers in the control group. Three years after the audit, the estimated differential remains quite high at 120%. Looked at another way, these estimates imply that roughly 55% of the income reported among members of the positive-tax-change experimental group on their TY2010 returns is a direct result of their audit experience three years earlier. We find a more substantial response of reported taxable income to an audit than we do for reported Schedule C net profit, suggesting that other components of taxable income are also affected by audits.

Importantly, we also find that audits have a detrimental long-term impact on the reporting behavior of taxpayers in our no-tax-change experimental group (E-NC). Three years after having undergone an audit, taxpayers that were not assessed additional taxes report around 35% less in taxable income than the control group. This difference is significant at the 1% level. Note that although the audit was initiated prior to the filing of the TY2008 return, it generally did not conclude until sometime in 2009 or later, often after the date the TY2008 return was filed. It is therefore not

surprising that we find a weaker response when assessing the short-term impact of audits (i.e., the change in reported income between TY2007 and TY2008).[4]

We also have estimated a more restrictive specification that does not allow the audit impact to vary in accordance with the outcome of the audit. In this specification, all audited taxpayers are treated as a single treatment group, making no distinction between taxpayers with and without a recommended additional tax assessment. The findings for this more restrictive model indicate that audits have an enduring positive effect on income reporting within the combined treatment group: on net, reported taxable income remains 20% higher three years after an audit.

Our results are qualitatively similar when we estimate alternative specifications involving the level of taxable income as the dependent variable rather than the natural logarithm, although the estimated effects are somewhat less dramatic. Among members of our positive-tax-change experimental group (E-PC), reported taxable income under these specifications is estimated to increase by approximately $13,000 (43%) relative to the control group in the year following the audit. Three years later, this differential remains substantial at $8,000 to $9,000. In the case of our no-tax-change experimental group (E-NC), the estimated impact of an audit on reported taxable income remains negative under these specifications, but it is less precisely estimated.

We also have investigated how audits impact one's long-term prospects for remaining self-employed. We find that experiencing an audit that results in an additional recommended tax assessment sharply reduces one's likelihood of filing a Schedule C return in the year following the audit (by approximately 7 percentage points). In contrast, audits that do not result in an additional recommended tax assessment do not significantly impact one's prospects for remaining self-employed.

2. Theoretical considerations

Our objective is to identify the causal impact of audits on reported income. This analysis is complicated by the fact that enforcement activity is likely triggered by both observable and unobservable factors that are not independent of reported income. Examples of such endogenous factors (i.e., variables that are jointly determined with reported income) include the history of reported gross receipts, claimed deductions, and the structure of business expenses. If taxpayers who are audited differ in important ways from unaudited taxpayers, simple comparisons between these groups might not reflect the causal impact of audits.

The literature on modern treatment evaluation, comprehensively summarized by Wooldridge (2010) and Blundell and Costa Dias (2000), provides useful empirical techniques for addressing this issue. The generic problem of this literature is readily applied to our context. Specifically, we characterize taxpayers by three variables: the outcome in the absence of a treatment, Y^0, the outcome in the presence of a treatment, Y^1, and a dummy variable, D, indicating treatment assignment. In our context, the treatment in question is an audit, the outcome variable is a measure of reported income, and the assignment indicator identifies whether a taxpayer was audited prior to making the income

[4]Our short-term findings are consistent with our hypothesis of a differential impact among the no-tax-change and positive-tax-change experimental groups. The initiation of an audit prior to the TY2008 return filing date may have immediately driven noncompliant taxpayers to report more income on their TY2008 returns. On the other hand, the audits ultimately may have prompted compliant taxpayers to engage in more legal tax avoidance (particularly if the audit made them aware of such opportunities). However, since the audit typically did not start until late in 2008 or early in 2009, there would have been limited opportunity to identify and execute legal avoidance strategies for TY2008. Further, to the extent that the audit experience prompted those in the no-tax-change group to become less compliant, this may not have taken root until the completion of the audit, by which time the TY2008 return already would have been filed in most cases.

report. We seek to identify the average treatment effect on the treated, which in our case translates into the impact of audits on audited taxpayers:

$$\tau^{ATT} = \mathbb{E}[Y^1 - Y^0 | D = 1], \tag{1}$$

This measure differs, in general, from the average treatment effect (the expected impact of an audit on a taxpayer who is randomly drawn from the entire population). The two measures coincide only if treatments are randomly assigned, or the impact of a treatment is constant across the entire population (Heckman and Robb, 1986). Both assumptions are unlikely to hold in the present context for two reasons. First, operational audit selection at IRS is guided by sophisticated algorithms (such as the Discriminant Index Function score, or DIF score) that are meant to identify tax returns with a high potential for noncompliance. Second, the response of compliant taxpayers to an audit is likely to differ from the response of noncompliant taxpayers. Under our empirical strategy, we account for the possibility that audits impact compliant and noncompliant taxpayers differently.

We expect that audits enforce the compliance of tax evaders. The impact on a compliant taxpayer, however, is uncertain and could depend on the interaction with the tax administration as well as the taxpayer's motivational posture (Braithwaite et al., 2007). On the one hand, audits may increase a taxpayer's trust in the state, and therefore, serve to reinforce the social norm of voluntary compliance. On the other hand, audits could be perceived as unjustified measures, thereby undermining one's willingness to comply voluntarily.

To account for potential differences among compliant and noncompliant taxpayers in response to an audit, we assume that each taxpayer is characterized by a personal propensity to evade, denoted by e. For simplicity, we assume that this variable only takes on two values. Noncompliant taxpayers are characterized by $e = 1$, while compliant taxpayers are characterized by $e = 0$.

3. Empirical strategy, sample selection, and descriptive statistics

The focus of this report is on nonfarm self-employed taxpayers reporting less than $200,000 in total positive income in TY2007; that is, on taxpayers who were assigned to IRS examination activity classes (EACs) 274 through 277 in that year. After describing our empirical strategy and treatment definition, we present our sample selection criteria, which are aimed at obtaining a relatively homogenous baseline sample of audited and unaudited taxpayers. The fourth subsection discusses some descriptive statistics and explores the trends in our main dependent variables.

3.1. **Empirical strategy.**

Below, we introduce some variants of the difference-in-differences estimation approach that we employ to measure the impact of audits on future taxpayer reporting behavior. A more detailed discussion of our econometric methodology is provided in Appendix B. Our Baseline Difference-in-Differences specification takes the following form:

$$(y_{ik} - y_{i0}) = \alpha + \beta_1 D_i + \beta_2 e_i D_i + \varepsilon_{ik}. \tag{2}$$

The dependent variable is the difference between taxpayer $i's$ reported income (y_{ik}) in post-audit tax year k (either TY2008 or TY2010) and the taxpayer's reported income (y_{i0}) in the pre-audit base year 0 (TY2007). Our main reported income measures are taxable income and Schedule C net profit. We include a constant term (α) in all regressions to account for a common impact of macro-shocks across all taxpayers. The audit group dummy (D_i) takes the value of one for taxpayers in the audit group and zero for those in the control group. The regression disturbance (ε_{ik}) represents the impact of unobserved individual-specific and time-varying factors, which are assumed to be independent of the audit group dummy. The impact of an audit on period k reporting is equal to β_1 for taxpayers in the no-tax-change experimental group (E-NC). Our specification allows for

the possibility that those receiving a positive recommended tax assessment (experimental group E-PC) respond differently to the audit. In particular, the coefficient β_2 of the interaction term $e_i D_i$ represents the difference in the magnitude of the reporting response between experimental groups E-PC and E-NC. The full impact of an audit on taxpayers receiving a positive additional recommended tax assessment is equal to $(\beta_1 + \beta_2)$.

Taking the change in reported income as the dependent variable, rather than the level, removes the potential correlation between the treatment group dummy D_i and unobserved time-constant components in the income process (i.e., individual-specific fixed effects), such as the personal propensity to evade. However, to ensure consistent estimation of the audit impact, two additional conditions need to be satisfied. One is that the experimental groups would have had similar income reporting trends in the absence of any audits (i.e., the common trend assumption).[5] The second is that the audit indicator is not correlated with the regression disturbance (i.e., that there are no unobserved factors, other than the fixed effect that has been differenced out, that influence both the audit selection process and taxpayer reporting behavior). We check the plausibility of the first assumption by investigating the trends in our dependent variables prior to TY2007 graphically. To address the second, we extend the Baseline Difference-in-Differences approach in various ways to control for the role of unobserved factors.

A negative income shock is one important factor that is potentially associated with audit selection. For instance, a taxpayer may be selected for audit as a result of experiencing and reporting an unusually low level of income in a given year. In such a case, a rise in income in subsequent years may not be fully attributable to the impact of the audit: a rebound in income likely would have come about even in its absence (through a phenomenon known as "mean reversion"). Accordingly, the Baseline Difference-in-Differences approach would fail to identify the causal impact of the audit in this case, owing to the correlation between the transitory income shock (captured by the disturbance term) and the audit group dummy in the regression equation. In the treatment effects literature, this problem is referred to as "Ashenfelter's dip" (see Ashenfelter, 1978).

The matching estimator addresses differences between audited and unaudited taxpayers by comparing members of the audit group against a matched control group of unaudited taxpayers with comparable observed characteristics (see, for example, Heckman and Smith, 1999). If the characteristics employed in the matching process include the recent income history of taxpayers, any mean reversion among members of the audit group should also be present among their matched counterparts in the control group (who have experienced similar transitory income shocks as indicated by their lagged income reports). The post-audit difference in income reporting behavior between the audit group and the matched control group will therefore measure the audit impact even when audit selection is influenced by the presence of temporary negative income shocks.

Three related points are worth stressing. First, a regression specification also may be used to consistently estimate the audit impact in this case. Specifically, by incorporating lagged levels of income as additional explanatory variables in our Baseline Difference-in-Differences regression specification, the linear regression model achieves a similar result to the matching estimator: the audit impact is estimated conditional on past income shocks, thereby ensuring that the estimate is not biased in the presence of mean reversion. We refer to this estimation strategy as an *Unrestricted Difference-in-Differences* estimator. Note, however, that the linear regression framework imposes a linear functional form on the covariates. The matching estimator is more flexible as it does not rely on any parametric assumptions regarding functional form.

[5] If this condition is not satisfied, the treatment dummy also captures the difference in trends between the audit group and the control group. Accordingly, the causal impact of audits would be confounded with the differential income reporting trend across the groups.

Second, consistency of the matching estimator in the presence of recent temporary income shocks requires that unobserved time-invariant factors in the income process (i.e., individual-specific fixed effects) do not also influence audit selection. Under the matching process, audited taxpayers are matched with unaudited taxpayers with similar observed characteristics, including their recent past income reports. Thus, while the audited taxpayers and their matched controls are observationally similar, they do differ in that the latter were not selected for audit. If the reason that the matched controls were not selected is that they had different fixed effects, which made auditing them less attractive (e.g., they had a lower personal propensity to evade), one would expect reported income levels across the two experimental groups to differ in subsequent periods even in the absence of any enforcement effect. In particular, the controls would be more likely to have experienced a deeper recent transitory income shock and, therefore, would tend to exhibit a greater level of mean reversion in subsequent periods than the audited taxpayers. The Unrestricted Difference-in-Differences estimator, which includes controls for lagged levels of income, would fail for the same reason.

Third, given that we expect the audit response to vary depending on one's personal propensity to evade, the simple matching estimator will not, in general, produce consistent estimates of the differential impact of audits on those who experience an additional recommended tax assessment and those who do not. The source of this problem is that we are unable to assess whether an unaudited taxpayer would have received an additional recommended tax assessment if that taxpayer had been audited. Consequently, the matched control group does not account for the differences in the income reporting trajectories for these two categories of taxpayers in the absence of an audit. To address this problem, we rely on a Matched Difference-in-Differences estimator, as proposed by Heckman, Ichimura, Smith, and Todd (1998). In particular, we separately compare the change in reported income between period 0 and period k for our matched control group against the change in reported income for experimental group E-PC (i.e, those receiving a positive recommended audit assessment) and for experimental group E-NC (i.e., those receiving no recommended audit assessment). By subtracting the income report in period 0 from the report in period k, we are able to effectively account for any permanent differences in reporting postures (i.e., fixed effects) among the experimental groups (i.e., the differencing operation sweeps away the fixed effects).

To summarize, none of our above estimators are able to simultaneously control for audit selection based on time-varying shocks (such as recent transitory income changes prior to an audit) and audit selection based on individual fixed effects. Our Matched Difference-in-Differences estimator as well as our Unrestricted Difference-in-Differences estimator can address the former potential issue, but not the latter. In contrast, the Baseline Difference-in-Differences estimator can address the latter potential issue, but not the former.[6] We therefore compare results from a range of alternative estimators to assess the potential sources of bias and the robustness of our estimates. Our alternative estimation approaches include some extensions of the aforementioned methodologies to account for sample selection and sample attrition. Overall, we employ six alternative estimation approaches.

Our first two approaches, Baseline DD and Dynamic DD, are robust to unobservable time-constant shocks, such as the propensity to evade.

Baseline DD: The first set of estimates is based on the Baseline Difference-in-Differences approach described by Equation 2. This methodology produces unbiased predictors of the impact of audits on the reporting behavior of taxpayers in both the positive-tax-change

[6]For a detailed discussion of the consistency of alternative estimators in the presence of fixed effects and transitory shocks, refer to Chabe-Ferret (2014).

and the no-tax-change experimental group, so long as audit selection is based only on time-constant variables.

Dynamic DD: To account for other taxpayer characteristics that influence both taxpayer reporting behavior and audit selection, this methodology incorporates lagged changes of reported income and a variety of indicators as additional explanatory variables in the Baseline DD regression specification. Note that recent changes in income are independent of taxpayer fixed effects, thus not introducing bias if audit selection is based on these effects. The set of indicator variables reflects attributes of the tax return, other than reported income, which might increase the likelihood of an audit.[7] The Dynamic DD approach gives unbiased estimates of the audit impact if audit selection is triggered by time-constant variables, recent changes in reported income, or any of the included indicators.

Our third and fourth estimators, Unrestricted DD and Matched DD, control for audit selection based on recent transitory income changes (i.e., Ashenfelter's dip).

Unrestricted DD: Under the Unrestricted DD approach,[8] we substitute lagged changes in income with lagged levels of income (one and two lags). Otherwise, this specification resembles the Dynamic DD approach. By including lagged income levels, we control for recent shocks that potentially drive audit selection. The Unrestricted DD specification returns unbiased estimates of the treatment impact if audit selection is based on any linear function of past income, including its change, or if it is triggered by any of the variables included in the set of indicators. Importantly, the Unrestricted DD specification is not robust to selection on time-constant unobservable variables (i.e., taxpayer fixed effects).

Matched DD: The Matched DD Estimator builds on the assumptions of the Unrestricted DD estimator: it thus provides unbiased predictions of the audit impact if audit selection is based on any linear function of past income or any of the included indicators. We implement this approach by comparing changes in reported income among our two experimental audit groups (E-PC and E-NC) with changes in reported income among our matched control group (see Section 4.2 for details). The Matched DD estimator is less parametric than the Unrestricted DD estimator as it does not assume a specific functional form for the covariates. However, this improved flexibility comes at the expense of a smaller sample size.

Our two remaining estimators, Dynamic DD Plus Sample Selection and Dynamic DD Plus Attrition, control for sample selection on unobservables and sample attrition, respectively. Both specifications build on the Dynamic DD approach.

Dynamic DD Plus Sample Selection: The IRS might rely on certain time-varying variables when deciding which returns to audit that are correlated with the income process but not observable to us. To control for the implied bias, we follow Heckman (1978) by including a synthetic control variable that captures the residual correlation (see Appendix B for details). The Dynamic DD Plus Sample Selection estimator yields consistent estimates of the audit impact if the assumptions of the Dynamic DD approach are satisfied and if certain additional distributional assumptions also hold.

Dynamic DD Plus Attrition: The sixth specification includes another control variable (λ_2) to account for sample attrition. Roughly 8 percent of the baseline sample does not file a Schedule C return after TY2007. If the dropout rate is correlated with the treatment, our estimates might not be valid. The Dynamic DD Plus Attrition estimator yields consistent

[7] See Section 4.1 below for a detailed discussion of these indicators.

[8] For a similar approach, see LaLonde (1986, 1984).

predictions of the audit impact if the assumptions of the Dynamic DD approach are valid and if certain additional distributional assumptions are also satisfied.

3.2. **Definition of treatment.** We assign taxpayers to the treatment group if they were audited prior to filing their TY2008 return but had no audits open or close in the three years preceding their TY2007 return filing date. Our control group consists of taxpayers who had no audits open or close in the three years preceding their TY2008 filing date. The requirement of having been audit-free for three years aims at enhancing the comparability of the two experimental groups and facilitates the assessment of whether income reports among the treatment and control groups follow common trends prior to the audit.

We operationalize this definition by combining three variables: (i.) information on the date an audit started, (ii.) information on a return's recorded transaction date, and (iii.) information on the date a return was posted to the master file. The audit start date is accurately recorded for each taxpayer and each year. The effective filing date, however, is in some cases uncertain. We thus rely on a two-step procedure to infer when the return was filed. In the first step, we assign the recorded transaction date. However, if a return was filed in a timely manner, April 15th is often recorded as a transaction date even if the actual transaction was received well before this date or, sometimes, even after it. In this common situation, we turn to information on the date the return was posted to the master file. This second proxy for the effective filing date is also imperfect: there might be a considerable gap between the date a return is sent to the IRS and the date this return is posted to the master file, especially if the return is sent by mail. To minimize measurement error we drop taxpayers if the recorded audit start date lies within 14 days of the filing date as determined by our two-step procedure.

3.3. **Sample selection.** The preliminary audit data sample drawn for this project included information on a random sample of 6,451 nonfarm self-employed taxpayers (from examination activity classes 274 through 281 in TY2007) who were recorded as having an audit open prior to filing their TY2008 return. Taxpayers were excluded if they had an audit open or close at any time in the three years prior to their approximate TY2007 return filing date.[9]

Also drawn for this project was a preliminary control sample of 11,218 taxpayers who had no audits open or close in the three years preceding their approximate TY2008 return filing date. The control sample was selected so that the distribution of taxpayers across the TY2007 DIF-score ventile categories for each examination activity class (based on the audit data sample) was comparable to that of the audit data sample. Specifically, taxpayers were randomly sampled in such a way that there were approximately 15 control sample taxpayers in each TY2007 examination activity code and DIF-score ventile category for every 10 audited taxpayers in this category.

As discussed above, we have refined our initial audit and control samples to more precisely classify treatments and controls for our empirical analysis. Table 1 summarizes our sample selection process. The number of taxpayers is depicted separately for each experimental group to illustrate the impact of our selection requirements on the sample composition. The preliminary project data sample consists of 17,669 taxpayers. Around 25% of this sample was drawn from EAC class 280 or 281 (nonfarm self-employed taxpayers with over $200,000 in total positive income). We exclude these classes from our analysis, because high income cases often present unique compliance issues. After excluding taxpayers from these classes as well as those that could not be definitively assigned to a treatment or control group on the basis of our two-step procedure for assigning a filing date, there are 12,707 taxpayers remaining in our sample.

[9] Also excluded from the audit data sample as well as the control sample were taxpayers reporting total income of more than $10 million in any tax year between 2005 and 2011.

TABLE 1. Impact of sample selection process

Subsample		Control		Treatment		Total	
Step	Description	Ind.	% Step (x-1)	Ind.	% Step (x-1)	Ind.	% Step (x-1)
0	Initial working sample	11,218	–	6,451	–	17,669	–
1	New definition	8,313	0.74	4,394	0.70	12,707	0.72
2	Schedule C filed 2006 and 2007	6998	0.84	3695	0.84	10693	0.84
3	Chronological filers	5974	0.85	3087	0.84	9061	0.85
4	Not late before 2008	4921	0.82	2425	0.79	7346	0.81
5	TY2005 not audited	4920	1.00	2379	0.98	7299	0.99
6	Outliers	4705	0.96	2204	0.93	6909	0.95

In the second step, we require that all taxpayers filed Schedule C income in both TY2006 and TY2007. This step is necessary to allow matching on the basis of variables for TY2006. The third step eliminates taxpayers who did not file their returns chronologically. Such cases preclude an analysis of long term effects. Furthermore, in order to effectively capture macro-economic trends in our empirical analysis, we require that returns were filed timely. If we included taxpayers who filed their return for TY2007 in, say, TY2012, our constants included in the difference-in-differences regressions would not capture common trends. We increase homogeneity in our treatment group by dropping taxpayers whose returns for TY2005 were audited (subsequent to filing their TY2008 returns). Accordingly, the treatment group consists only of taxpayers who were audited in relation to their TY2006 and/or TY2007 return.[10] Finally, we exclude taxpayers reporting extreme values (the top 2.5% and the bottom 2.5% of the distribution) of our main dependent variables, to ensure that our results are not driven by outliers and are thus representative of the overall sample. The final baseline sample consists of 6,909 taxpayers, including a treatment group of 2,204 taxpayers who were audited prior to filing their TY2008 return and a control group of 4,705 taxpayers who were not audited prior to filing their TY2008 return.

3.4. **Descriptive statistics.** Table 2 presents descriptive statistics separately for the treatment and control groups, with the last column showing the probability of equal means between groups. In order to achieve comparability of DIF scores between the control and treatment groups while protecting the confidentiality of the underlying DIF algorithm, we worked with ventiles of the DIF distribution. Our ventile measure takes values between 1 and 20, with 20 reflecting the most extreme 5% of DIF scores within a given EAC class in our sample. Although the preliminary control group was selected to have approximately the same distribution across TY2007 DIF score ventiles as the treatment group, the latter tended to have higher relative DIF scores in TY2006.

The experimental groups do not differ significantly in terms of taxable income before TY2008.[11] Reported Schedule C net profit is somewhat higher in the treatment group. The ratio of net profit to total gross receipts (profit ratio), on the other hand, is significantly lower in the treatment group. To account for a differential response to audits, we define the binary variable e to take the value of one if an audit resulted in an additional tax assessment. According to this classification, approximately one-half of all taxpayers in the treatment group fall into the additional tax assessment category. The two indicator variables, Schedule C Filed 2008 and Schedule C Filed 2010, take the value one

[10]All of the members of the preliminary audit sample were recorded as undergoing an audit of their TY2007 return prior to the approximate filing date for their TY2008 return. However, some of these taxpayers were later determined to have actually filed their TY2008 return prior to the TY2007 audit. Such taxpayers were retained in the control group if their TY2006 return was audited subsequent to the TY2007 return filing date but prior to the TY2008 return filing date.

[11]We add $1 to the level of taxable income before transforming this variable in order for the natural log transformation to be valid for returns that report zero reported taxable income.

TABLE 2. Descriptive statistics

Subsample	Control group (N=4705)			Treatment group (N=2204)			
Measure	Min.	Mean	Max.	Min.	Mean	Max.	p-value
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DIF ventile 2006	1	8.82	20	1	10.64	20	0.000
DIF ventile 2007	1	10.02	20	1	10.45	20	0.004
Taxable income	0	30,810	290,200	0	30,570	242,400	0.773
Log taxable income	0	7.72	12.25	0	7.76	12.28	0.722
Sch. C Net Profit	-95,900	19,010	194,100	-101,400	19,850	200,200	0.382
Profit ratio	-34,990	-305	24.71	-51,200	-524.3	16.33	0.004
e	0	0	0	0	0.5	1	0.000
Penalty	0	155.4	72,650	0	1,218	97,980	0.000
Schedule C Filed 2008	0	0.94	1	0	0.9	1	0.000
Schedule C Filed 2010	0	0.84	1	0	0.8	1	0.000
Change in log taxable income 2007-2008	-11.44	-0.36	11.65	-11.53	0.11	11.59	0.000
Change in log taxable income 2007-2010	-11.55	0.09	12.07	-11.54	0.23	12.13	0.273
Change in profit ratio 2007-2008	-39,350	-17.54	50,360	-36,230	81.77	39,660	0.184
Change in profit ratio 2007-2010	-47,680	11.47	50,370	-53,420	70.07	50,070	0.496

Table depicts average values between 2006 and 2007 unless a year is mentioned in the variable name.

if a taxpayer filed Schedule C income in the given year. About 94% of the taxpayers in the control group filed Schedule C income in TY2008. In the treatment group, the percentage is significantly lower (90%). Three years after the audit, the share of taxpayers still filing Schedule C is 80% in the treatment group and 84% in the control group.

Finally, the last four rows present the change in two income measures one and three years after the audit, respectively. Control group members experienced a 36% decrease in reported income one year after the audit,[12] likely reflecting the economic downturn in TY2008. In the treatment group, average reported income increased by 11% over this same period. The difference across groups, amounting to 25%, is significant at the 1% level. Three years after the audit, however, the difference in the trends in reported taxable income across groups is no longer statistically significant. The one- and three-year changes in the profit ratio show a similar pattern, although even the one-year change is not significant across groups for this variable, likely due to its high degree of variation within the sample.

3.5. **Trends in income.** Our empirical strategy rests on the assumption that income reporting trends for each the experimental groups would have been similar in the absence of audits. While this assumption is not testable, the similarity of the reporting trends across these groups prior to TY2008 demonstrates the plausibility of this premise. Figure 1 depicts average reported values in the treatment and the control group between TY2005 and TY2011.

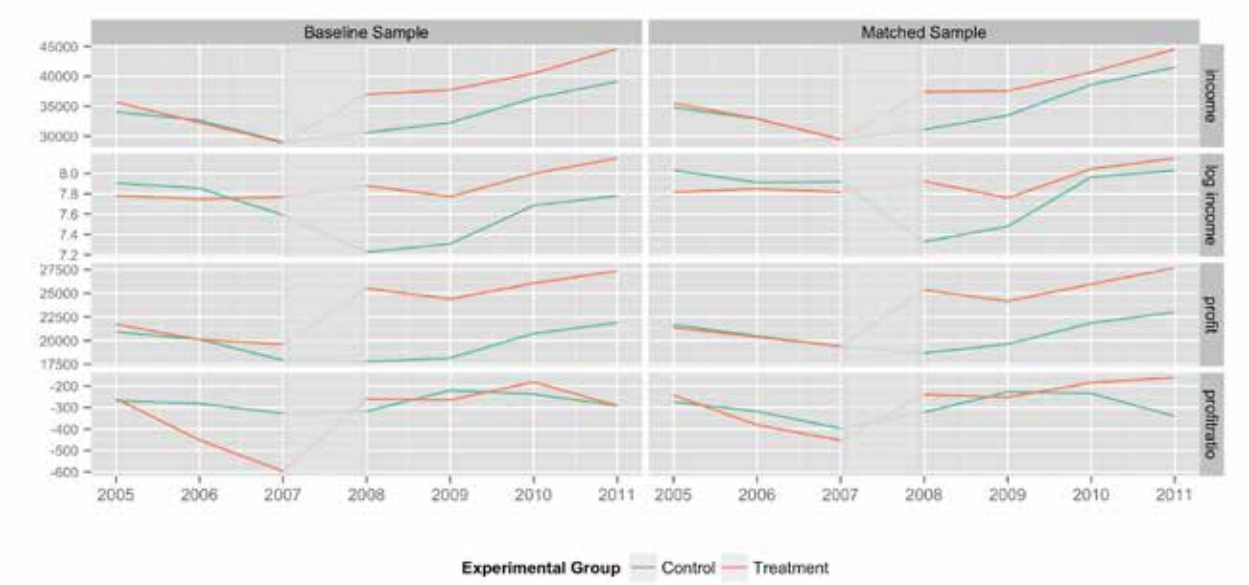
The left panel presents average values of taxable income, the natural logarithm of taxable income, Schedule C net profit, and the ratio of Schedule C net profit to Schedule C gross receipts, our main dependent variables, within the baseline sample. The right panel gives average values for the same variables within in the matched sample (refer to Section 4.2 below for details on our matching procedure). The level of reported taxable income clearly follows a similar trend in the treatment and control groups prior to TY2008 (top panel). The trend in reported Schedule C net profit is also similar across groups (third panel from top). The main focus in our empirical

[12] This follows from noting that $ln(Income_{2008}) - ln(Income_{2007}) = x$ implies $(Income_{2008} - Income_{2007})/Income_{2007} \approx x$ for a small x.

analysis, however, is on the natural logarithm of taxable income and the profit ratio, because these measures are more robust to the presence of outliers. The natural logarithm of taxable income developed comparably in each group between TY2005 and TY2006, supporting the assumption of a common trend. However, relative to the treatment group, the average value of this variable shows a modest decline in the control group between TY2006 and TY2007. This may signify an association between recent changes in reported taxable income and audit selection. Our matched samples, which account for recent income changes, demonstrate that the similarity of the trends in the natural logarithm of reported income improves after accounting for such changes.

When examining trends in the profit ratio (the last panel in Figure 1), the treatment group clearly differs from the control group in the baseline sample. The dip in TY2007 implies that the Baseline Difference-in-Differences estimation approach might not identify the causal impact of audits (Ashenfelter, 1978). The reporting trends prior to TY2008 look much more similar in the matched sample, suggesting that the Matched Difference-in-Differences procedure may hold more promise for this variable.

FIGURE 1. Trends in income



4. Empirical results

This section presents our empirical findings, beginning with our results concerning the determinants of audits (Section 4.1). These results are then used in Section 4.2 to construct a matched control group. The third subsection investigates factors driving the likelihood of reporting Schedule C income in TY2008 and TY2010. In the fourth subsection, we present our main results on the impact of audits.

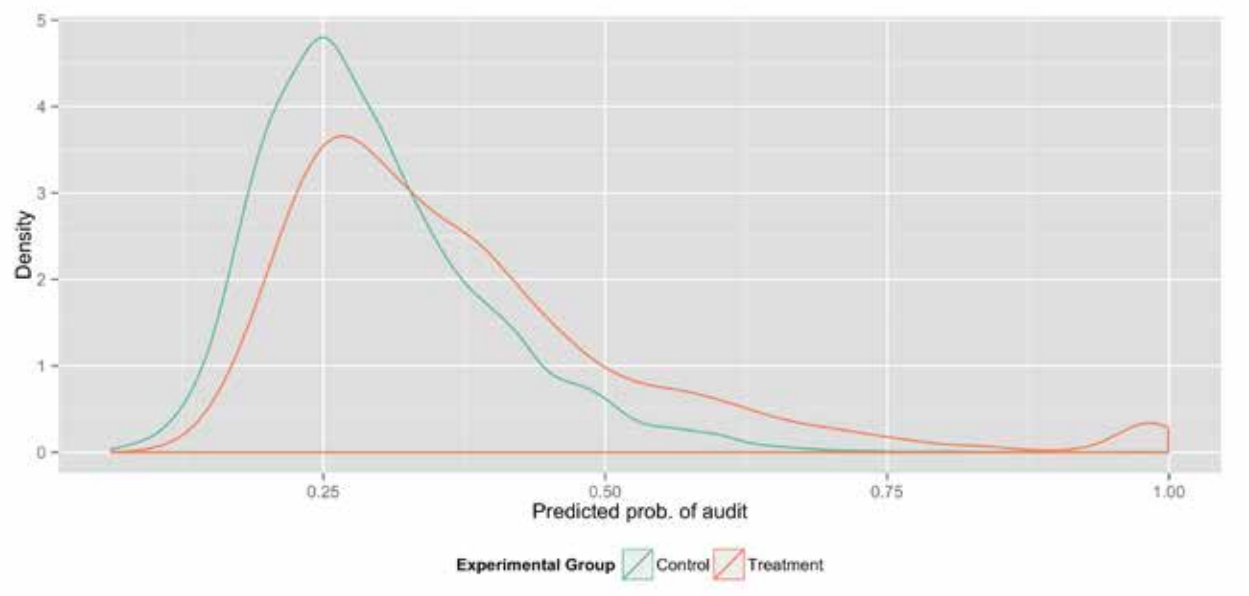
4.1. **The determinants of audits.** To uncover factors driving the relative likelihood of audit selection within our sample, we estimate a binary choice model (probit) with a dummy for treatment

assignment (equal to 1 for audit group members and 0 for controls) as the dependent variable. We incorporate a variety of explanatory variables to explain group assignment.[13]

4.2. **Propensity score matching.** We construct the propensity score (i.e., the estimated probability of assignment to the treatment group for each taxpayer in our sample, conditional on a wide range of explanatory variables) using the results of our probit specification for group assignment. Taxpayers with similar propensity scores are comparable in terms of their observed characteristics. Accordingly, any difference in subsequent reporting behavior between matched audited and unaudited taxpayers should be attributable to the audit impact (see, for example, Smith and Todd, 2005).[14]

Figure 2 illustrates the distribution of propensity scores among the treatment and control group members. Taxpayers in the treatment group are, as expected, more likely to be audited than taxpayers in the control group. The highest estimated probability of treatment assignment is 0.9993 in the treatment group and 0.9741 in the control group. To find an unaudited counterpart for each member of the treatment group, the experimental groups need to be trimmed to the region of common support. The figure shows that there are practically no valid control observations for propensity scores above 0.60. Accordingly, we exclude observations with propensity scores above this threshold (representing approximately 15% of our treatment sample) from the analysis. Consequently, our estimates based on the Matched Difference-in-Differences approach will not necessarily be representative of the impact of audits on taxpayers with very large propensity scores.

FIGURE 2. Estimated probability of treatment assignment



After excluding those observations with propensity scores above 0.6, we match each audited return in our sample to an unaudited control using a nearest neighbor matching algorithm (without

[13]We had access to over 40 indicator variables and selected a subset of these based on prediction quality. Specifically, we executed a stepwise (backward) probit algorithm. Under this algorithm, a probit model was initially estimated using all of the potential explanatory variables. Less significant predictors were then sequentially removed from the model until the Akaike Information Criterion was maximized.

[14]As discussed in Section 3.1, this conclusion also rests on the assumption that audit selection is not attributable to any unobserved factors that are associated with taxpayer reporting behavior.

replacement). This method pairs taxpayers based on the similarity of their propensity scores. Table 3 presents the results of this exercise. Our matched experimental sample consists of 1,980 taxpayer pairs.[15] The table depicts the distribution of propensity scores, separately for the treatment and matched control group, across the deciles of the matched treatment group. The last column presents t-tests comparing mean values between the two groups. As expected, we find more comparable pairs at the bottom of the propensity distribution. The weakest matches are reported in the highest decile covering propensity scores between 0.47 and 0.58.

TABLE 3. Distribution of propensity scores in matched experimental groups

Subsample	Matched control group (N=1980)				Matched treatment group (N=1980)				
Measure	Observations	Min.	Mean	Max.	Observations	Min.	Mean	Max.	p-value
Decile of treated	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	197	0.1095	0.1858	0.2129	198	0.1099	0.186	0.2129	0.95
2	199	0.213	0.2272	0.2393	198	0.2135	0.2273	0.2394	0.91
3	198	0.2394	0.2526	0.2636	198	0.2394	0.2526	0.2636	0.99
4	197	0.2636	0.2752	0.2866	198	0.2636	0.2753	0.2866	0.89
5	198	0.2866	0.3007	0.315	198	0.2869	0.3008	0.3151	0.84
6	198	0.3152	0.3286	0.3447	198	0.3152	0.3288	0.3448	0.83
7	198	0.3449	0.3616	0.3786	198	0.3451	0.3618	0.3793	0.81
8	200	0.3793	0.3969	0.4169	198	0.3794	0.397	0.4172	0.90
9	197	0.4173	0.4431	0.4739	198	0.4174	0.4435	0.4741	0.83
10	196	0.4742	0.5182	0.579	198	0.4743	0.5233	0.5795	0.11

4.3. **The determinants of filing Schedule C income.** We now turn to examining whether audits impact the future likelihood of filing Schedule C. Table 4 below presents the estimation results of a binary choice model (probit) where dummy indicators for the presence of Schedule C earnings in TY2008 and TY2010 serve as the dependent variables.

The first and second specifications in Table 4 examine the probability of filing Schedule C income in TY2008. Controlling for pre-treatment income and DIF scores, we find that audits decrease the likelihood of continued filing of Schedule C in the following year by approximately 7 percentage points among taxpayers in the positive-tax-change experimental group (E-PC). Audits do not have a statistically significant effect on the likelihood of a subsequent Schedule C filing among members of the no-tax-change experimental group (E-NC). Income measures have the anticipated effects on future Schedule C filing behavior: the more profitable a business is (relative to other income sources) the higher the likelihood of filing Schedule C in TY2008.

We add a variety of additional control variables in the second specification. Following the stepwise variable selection strategy described above, we select the subset of all indicator variables that maximizes the model fit. With a magnitude of -12.5%, the largest absolute estimated marginal effect is associated with the moving expense indicator: taxpayers just having moved are significantly less likely to report Schedule C income. Depreciation expenses in TY2007, signaling the presence of valuable assets, increase the likelihood of filing Schedule C in TY2008 by almost 5%. Other pre-audit expenses that increase the likelihood of a subsequent Schedule C filing include expenses on entertainment, legal consulting, wages, and using one's home for business purposes. The other income sources in this specification, with the exception of wage income, have a positive impact on the probability of filing Schedule C income.

[15]Note that only 1,978 observations of the control group are reported in the left column of Table 3 as two observations in the control group have a propensity score above 0.5795.

TABLE 4. Determinants of filing Schedule C past TY2007

Estimation of sample selection (probit estimation)							
Dependent variable	Filed C in 2008		Filed C in 2010				
Model	(1)	(2)	(3)	(4)	Model cont.	(2)	(4)
Explanatory Variables					Explanatory Variables cont.		
Treatment	-0.005	-0.005	-0.003	-0.002	Schedule E indicator 2006	0.025***	-0.001
	(0.009)	(0.008)	(0.013)	(0.013)		(0.008)	(0.018)
Treatment:e	-0.067***	-0.053***	-0.072***	-0.065***	Start up 2006	-0.017*	-0.042**
	(0.015)	(0.013)	(0.019)	(0.018)		(0.009)	(0.017)
DIF vent. 2006	0.004**	0.003*	0.007***	0.005**	Depreciation exp. 2006	-0.01	-0.003
	(0.001)	(0.001)	(0.002)	(0.002)		(0.007)	(0.013)
DIF vent. 2007	0.002	0.000	0.002	-0.001	Meal and enter. exp. 2006	-0.01	-0.014
	(0.001)	(0.001)	(0.002)	(0.002)		(0.006)	(0.012)
DIF vent. squared 2006	-0.000**	-0.000*	-0.000**	-0.000**	Wage indicator 2007	-0.034***	-0.047***
	(0.000)	(0.000)	(0.000)	(0.000)		(0.006)	(0.011)
DIF vent. squared 2007	0.000*	0.000	0.000*	0.000	Capital gains ind. 2007	-0.019**	-0.017
	(0.000)	(0.000)	(0.000)	(0.000)		(0.008)	(0.014)
DIF vent. 2006:DIF vent. 2007	0.000	0.000	-0.000*	0.000	Other gains ind. 2007	-0.026*	-0.028
	(0.000)	(0.000)	(0.000)	(0.000)		(0.013)	(0.021)
Sch C Net Profit 2006	0.000***	0.000***	0.000***	0.000**	Schedule E ind. 2006	-0.027**	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)		(0.011)	(0.018)
Sch C Net Profit 2007	0.000	0.000	0.000***	0.000*	Other income ind. 2007	0.016**	0.005
	(0.000)	(0.000)	(0.000)	(0.000)		(0.006)	(0.014)
Profitratio 2006	0.000***	0.000**	0.000***	0.000***	Moving exp. ind. 2007	-0.125**	-0.199***
	(0.000)	(0.000)	(0.000)	(0.000)		(0.048)	(0.067)
Profitratio 2007	0.000	0.000	0.000**	0.000**	Simple account cont. 2007	0.015	0.009
	(0.000)	(0.000)	(0.000)	(0.000)		(0.009)	(0.019)
Log Taxable Income 2006	-0.002**	0.000	-0.006***	-0.004**	Car truck expenses 2007	0.01	0.027**
	(0.001)	(0.001)	(0.002)	(0.002)		(0.006)	(0.012)
Log Taxable Income 2007	0.000	0.001	-0.001	0.000	Depreciation exp. 2007	0.046***	0.071***
	(0.001)	(0.001)	(0.002)	(0.002)		(0.008)	(0.013)
Taxable Income 2006	-0.000***	-0.000***	0.000	0.000	Legal expenses 2007	0.014**	0.038***
	(0.000)	(0.000)	(0.000)	(0.000)		(0.005)	(0.010)
Taxable Income 2007	0.000**	0.000	0.000*	0.000	Travel expenses 2007	0.010*	0.029**
	(0.000)	(0.000)	(0.000)	(0.000)		(0.006)	(0.011)
Profit/Taxable Income 2006	0.000	0.000	0.000	0.000*	Meal and enter. exp. 2007	0.021***	0.045***
	(0.000)	(0.000)	(0.000)	(0.000)		(0.007)	(0.014)
Profit/Taxable Income 2007	0.000	0.000*	0.000	0.000	Wage expenses 2007	0.016**	0.016
	(0.000)	(0.000)	(0.000)	(0.000)		(0.006)	(0.013)
Interest income ind. 2006		0.028***		0.033***	Business home exp. 2007	0.016***	0.018*
		(0.006)		(0.010)		(0.005)	(0.011)
Capital gains ind. 2006		0.016**		0.028**			
		(0.007)		(0.013)			
Observations	6971	–	6971	–		6971	6971
AIC	3461	–	6060	–		3230	5856

Notes: *,**, and *** indicate significance at the 10%, 5% and 1% level. Robust standard errors (bootstrapped) in parentheses.

The third and fourth specifications investigate determinants of filing Schedule C income in 2010. Three years after having been audited and subjected to an additional recommended tax assessment, the probability of filing Schedule C is estimated to be approximately 7 percentage points lower than if the audit had not transpired. This estimate is quite similar to the estimated impact in TY2008, suggesting that audits may immediately lead some taxpayers with a positive recommended tax change (members of E-PC) to exit self-employment for an extended period of time. The estimated marginal effect of other explanatory variables is similar to their impact in TY2008: taxpayers owning highly profitable businesses that allow spending on investments, entertainment, wages, legal advice and travel, are more likely to report Schedule C income in TY2010. On the negative side,

owning a start up, receiving wage income and reporting moving expenses are the most informative predictors of not filing Schedule C in TY2010.

4.4. **The impact of audits on reported income.** This section presents our main results concerning the impact of audits on reported income. Table 5 depicts the first set of estimation results where we use the change in the natural logarithm of reported taxable income between TY2007 and TY2008 as the dependent variable.

TABLE 5. Short-term impact of audits on taxable income

Dependent variable: log (taxable income 2008) - log (taxable income 2007)						
Estimator	Baseline DD	Dynamic DD	Unrest. DD	Matched DD	Dynamic DD	Dynamic DD
Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.306**	-0.094	-0.021	-0.073	-0.257	-0.174
	(0.138)	(0.131)	(0.123)	(0.159)	(0.337)	(0.370)
Treated:e	1.565***	1.442***	1.352***	1.604***	1.519***	1.394***
	(0.176)	(0.165)	(0.155)	(0.186)	(0.164)	(0.173)
Selection control (λ_1)					0.05	
					(0.225)	
Attrition control (λ_2)						0.890**
						(0.387)
Sum of row 1 and 2	1.258***	1.348***	1.330***	1.531***	1.262***	1.220***
	(0.138)	(0.130)	(0.121)	(0.163)	(0.370)	(0.139)
Observations	6904	6903	6903	3960	6903	6903
Adj. R^2	0.014	0.152	0.262	0.025	0.143	0.143

Notes: *,**, and *** indicate significance at the 10%, 5% and 1% level. Robust standard errors in parentheses.

The Baseline Difference-in-Differences specification in the first column indicates that audits decrease the natural logarithm of reported taxable income by 0.306 among taxpayers that were not assessed additional tax. This translates into a 35.8 percent reduction in the reported level of taxable income.[16] This estimated impact is significant at the 5% level. The interaction of the audit group dummy with the binary variable e, which takes the value of one if an examination resulted in additional recommended tax, shows that audits have a much stronger effect on members of experimental group E-PC. The combined coefficient estimate of 1.258 implies that reported taxable income among those receiving an additional recommended tax assessment increases by approximately 250%.[17]

Although the Baseline Difference-in-Differences specification indicates a significant negative impact of audits on subsequent income reporting behavior when audits result in no additional recommended tax assessment, the estimated size of this impact diminishes and loses its statistical significance in the extended specifications. This suggests that other factors explain both reported income and audit selection.

In the second specification (Dynamic DD), we incorporate the lagged change in reported income and a range of indicators as additional explanatory variables. After accounting for these additional factors, the estimated impact of audits on the reporting behavior of taxpayers who do not receive a recommended additional tax assessment (i.e., members of experimental group E-NC) becomes negligible. The positive effect of audits on taxpayers receiving an additional recommended tax assessment, however, remains sizable and highly significant. The combined coefficient estimate

[16]The percentage change figure is derived as $exp(0.306) - 1 = 0.358$.

[17]The percentage change figure is derived as $exp(1.258) - 1 = 2.50$.

of 1.348 now implies an increase of 285% in taxable income among taxpayers that were assessed additional tax. Put another way, this estimate implies that roughly 75%[18] of the income reported among members of experimental group E-PC in TY2008 is the direct result of enforcement activity. In the third and fourth specification, we allow for selection on time-varying observables. Both sets of estimates indicate no significant impact of audits on income reporting among members of experimental group E-NC but a large and significant positive impact among members of E-PC. The quantitative prediction of the Unrestricted DD approach is smaller in magnitude and closer to the estimates presented in Columns (1) and (2). The Matched Difference-in-Differences estimator indicates that approximately 80% of the income reported in TY2008 among members of experimental group E-PC is a direct result of audits.

We control for selection on unobservables and for attrition in the fifth and sixth columns of Table 5, respectively. The estimated coefficient on the first control function is not significant, indicating the absence of sample selection bias. Including a control for sample attrition (Column (6)), slightly reduces the estimated audit impact on taxpayers receiving a positive recommended audit assessment, while the estimated impact on those receiving no audit assessment remains statistically insignificant.

Table 6 explores the long-term impact of audits on reported income. The dependent variable is now the change in the natural logarithm of reported taxable income between TY2007 and TY2010. Three years after having undergone an audit, we find that members of the positive-tax-change experimental group (E-PC) still report 120% more in taxable income than before the audit. This translates into a direct enforcement effect (i.e., the percentage contribution of an audit to the level of reported income in 2010) of 55%. The estimated audit impact varies only marginally across specifications.

TABLE 6. Long-term impact of audits on taxable income

Dependent variable: log (taxable income 2010) - log (taxable income 2007)						
Estimator	Baseline DD	Dynamic DD	Unrest. DD	Matched DD	Dynamic DD	Dynamic DD
Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.476***	-0.296**	-0.191	-0.405**	-0.064	-0.318**
	(0.156)	(0.147)	(0.132)	(0.180)	(0.418)	(0.146)
Treated:e	1.223***	1.144***	0.985***	1.224***	1.171***	1.064***
	(0.199)	(0.185)	(0.167)	(0.212)	(0.185)	(0.192)
Selection control (λ_1)					-0.172	
					(0.254)	
Attrition control (λ_2)						0.775**
						(0.350)
Sum of row 1 and 2	0.747***	0.849***	0.794***	0.819***	1.107**	0.746***
	(0.157)	(0.145)	(0.131)	(0.186)	(0.418)	(0.152)
Observations	6904	6903	6903	3960	6903	6903
Adj. R^2	0.005	0.163	0.324	0.008	0.138	0.139

Notes: *,**, and *** indicate significance at the 10%, 5% and 1% level. Robust standard errors in parentheses.

Importantly, we now find a significant negative impact of audits among taxpayers in the no-tax-change experimental group (E-NC). Those audited but not assessed additional tax report approximately 35% less income as a result of the enforcement activity. The estimated impact of audits on income reporting among compliant taxpayers is not statistically significant in two of our specifications. The first is the Unrestricted Difference-in-Differences model, which includes lagged income

[18] On average, the income reported in TY2007 is roughly 25%=1/3.85 of its TY2008 value.

levels and various other control variables. Given that the Matched Difference-in-Differences estimator relies on a similar set of assumptions while not imposing a linear functional form, we attach more weight to the latter estimator (which does indicate a significant negative impact of audits). The second is based on the control function model in column (5), which attempts to account for both observable and unobservable differences between the treatment and control groups. The coefficient on the control function is, however, not statistically significant and the inflated standard errors point at a potential problem of multicollinearity in this specification. We thus conclude that, overall, our estimates point to a negative long-term impact of audits on income amounts reported by taxpayers in the no-change-tax group. Our preferred specification (column (6)) suggests that such taxpayers reduce their reported income by 37% three years after having undergone an audit.

We have also estimated a more restrictive model that does not allow the audit impact to vary in accordance with the audit outcome. In this specification, all audited taxpayers are treated as a single treatment group, making no distinction between taxpayers with and without a recommended additional tax assessment. The findings for this more restrictive model indicate that audits have an enduring positive effect on income reporting within the combined treatment group: on net, reported taxable income remains 20% higher three years after an audit.

We present estimates for some alternative specifications involving the change in the level of reported taxable income (rather than the change in its natural logarithm) as the dependent variable in Appendix A (Table 8 and Table 9). One year after the audit, we find that reported income among the positive-tax-change experimental group (E-PC) is increased by around $13,000 (or by about 42 percent); we do not find a statistically significant impact of audits among members of the no-tax-change experimental group (E-NC). Three years after the audit, taxpayers that were assessed additional tax still report around $8,000 to $9,000 more than control group members. The estimated long-run impact on income reporting among members of the no-tax-change experimental group is negative in all specifications. However, the estimated effect is statistically significant only when using the Matched Difference-in-Differences estimator. Thus, while the results based on the change in the level of reported taxable income are qualitatively similar to those based on the change in its natural logarithm, the latter imply a much larger percentage change in income reporting among the positive-tax-change experimental group as a result of the audits.

Figure 3 illustrates these findings. The distance between the positive-tax-change experimental group (orange line) and the control group (green line) widens steeply in TY2008 and shrinks gradually thereafter. The graph also indicates a negative impact of audits on the no-tax-change group (blue line). This effect is most visible in the matched sample when looking at the logarithm of reported income (bottom right panel).

Table 7 presents estimates of the impact of audits on the level of reported Schedule C net profit. We focus on profit levels rather than profit ratios first, because this variable seems to better satisfy the assumptions needed for consistent identification of the treatment impact.

The Baseline Difference-in-Differences specification depicted in Column (1) suggests that members of experimental group E-NC decreased the amount of reported Schedule C net profit by $855 in response to the audit. However, the estimated effect is not statistically significant. Combining this estimate with the coefficient estimate of the interaction, we find that members of experimental group E-PC increased their reported net profit by approximately $10,500 the year after being audited. This latter coefficient is significant at the 1% level.

After controlling for lagged changes in the dependent variable and other additional explanatory factors in the second specification, the estimated audit impact on the positive-tax-change experimental group (E-PC) increases slightly. In the case of the no-tax-change group (E-NC), the estimated change in reported income shifts from a negative to a positive value, but it remains

FIGURE 3. Impact of audits on reported taxable income

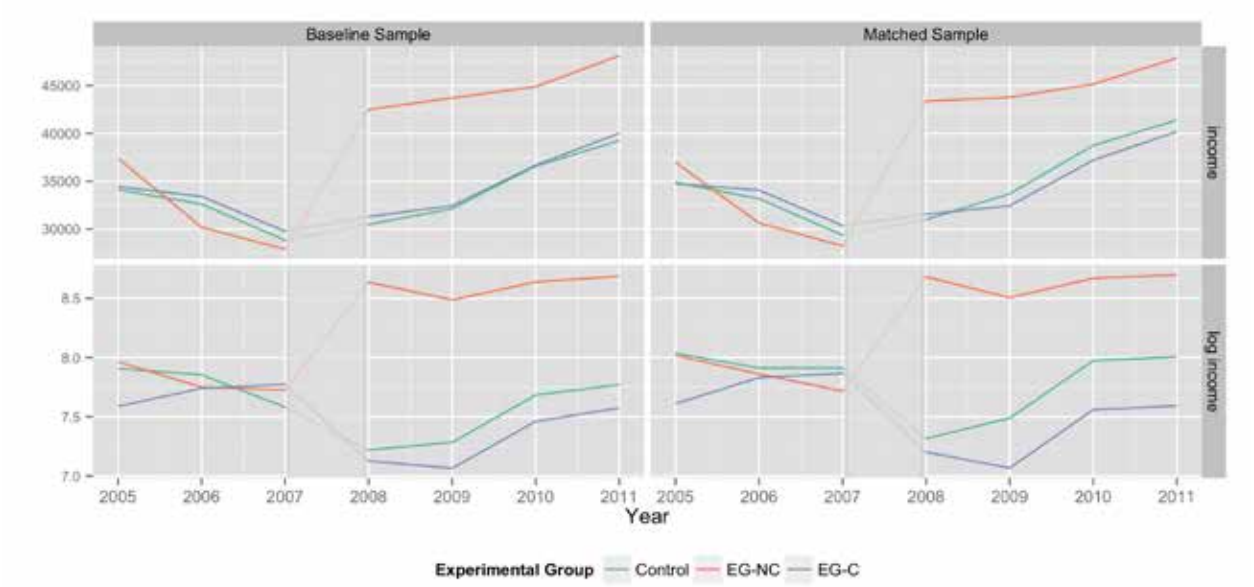


TABLE 7. Short-term impact of audits on Schedule C net profit

Dependent variable: Schedule C net profit 2008 - Schedule C net profit 2007						
Estimator	Baseline DD	Dynamic DD	Unrest. DD	Matched DD	Dynamic DD	Dynamic DD
Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-854.695	496.032	1,121.631	-876.488	-7,347.371***	-152.709
	(943.370)	(940.915)	(919.618)	(1,142.702)	(2,690.039)	(919.323)
Treated:e	11,319.136***	10,439.312***	9,062.906***	13,412.048***	10,924.824***	10,166.602***
	(1,233.504)	(1,211.675)	(1,186.069)	(1,376.111)	(1,200.277)	(1,258.110)
Selection control (λ_1)					4,643.215**	
					(1,640.058)	
Attrition control (λ_2)						6,195.364**
						(3,058.939)
Sum of row 1 and 2	10,464.442***	10,935.345***	10,184.537***	12,535.560***	3,577.453***	10,013.892***
	(986.009)	(964.414)	(943.067)	(1,219.835)	(2,705.077)	(1,023.932)
Observations	6386	6385	6385	3629	6385	6385
$Adj.R^2$	0.018	0.077	0.12	0.033	0.072	0.071

Notes: *,**, and *** indicate significance at the 10%, 5% and 1% level. Robust standard errors in parentheses.

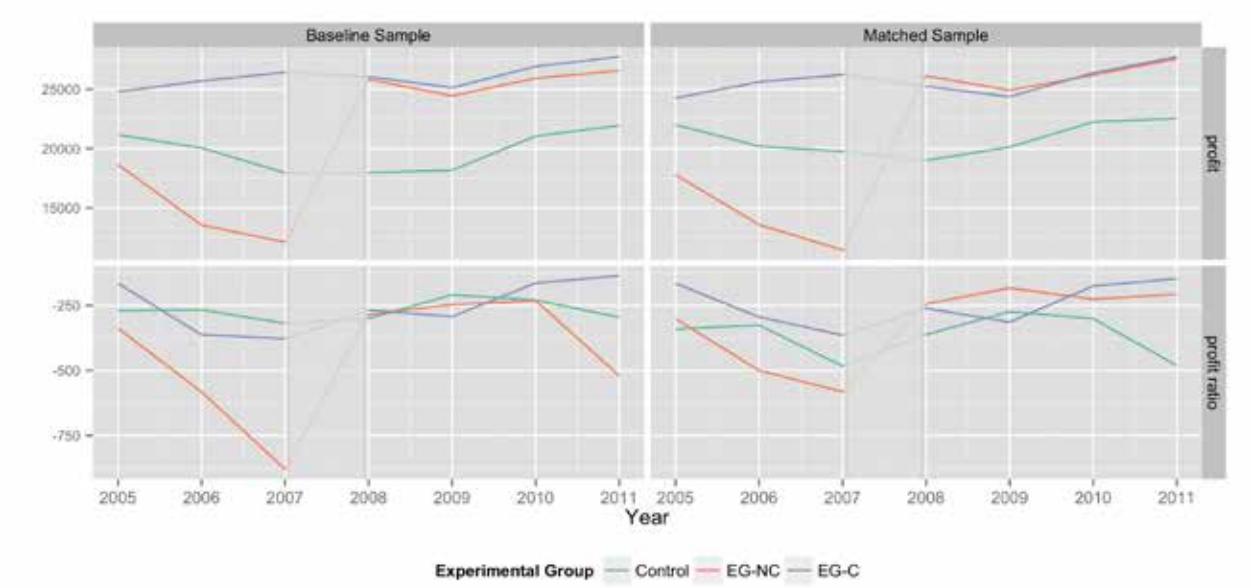
statistically insignificant. Including lagged levels of reported net profit as explanatory variables instead of lagged changes in net profit (Unrestricted DD in Column (3)) only modestly changes the estimated audit impact within the positive-tax-change experimental group, but it doubles the estimated impact on the no-tax-change group. The latter estimate, however, remains statistically insignificant. The Matched Difference-in-Differences estimator points to a slightly larger audit impact on the positive-tax-change group. Similar to the Baseline Difference-in-Differences approach, this approach estimates a modest, but statistically insignificant, impact of audits on net profit reporting within the no-tax-change experimental group. In the fifth specification, we control for potential unobservable factors that are associated with both the income process and the audit rule. The estimated impact of an audit on members of experimental group E-PC decreases to $3,577, while the effect on taxpayers who were not assessed additional tax is now equal to -$7,347 and

is statistically significant. As a robustness check, we have re-estimated this model after including DIF ventiles in TY2006 and TY2007 as additional explanatory variables. The estimated coefficient on the control variable is not significant in this specification and the estimated audit impact on the positive-tax-change and no-tax-change experimental groups is more comparable to the results presented in the other specifications. We therefore suspect that the estimates in Column (5) are biased. The estimates in the last column control for attrition. They confirm the positive impact of audits on taxpayers who were assessed additional tax and the negligible effect on those who were not.

We examine the long-term impact of audits on reported Schedule C net profit in Table 10 in Appendix A. The estimated impact on members of experimental group E-PC remains at approximately $6,500 (and statistically significant) three years after the audit. Similar to the estimation results on the long-term impact of audits on reported taxable income, we find a negative impact of audits on reported Schedule C net profit among members of experimental group E-NC in most of our specifications. The Baseline Difference-in-Differences approach suggests that members of the no-tax-change experimental group decreased their level of reported net profit by $2,400 in 2010. This reduction is statistically significant. However, after controlling for lagged changes in net profit, the estimated audit impact falls to -$1,000 and becomes statistically insignificant. Similar to the above reported results, we find unintuitive estimates when attempting to control for unobservable factors that are associated with both audit selection and profit reporting behavior. Overall, we conclude that audits seem to have a strong impact on reported net profit among those receiving an additional recommended tax assessment and a relatively small impact, possibly negative, on those receiving no additional assessment.

The estimated impact of an audit on the net profit ratio is indeterminate, switching signs depending on the specification used. We do not report these results; instead, we illustrate the impact on different measures of net profit in Figure 4. The impact on members of experimental group E-PC seems to be positive throughout all graphs. The impact on members of experimental group E-NC, however, is not clear. Furthermore, the lower left panel, illustrating the pre-audit period trend in the profit ratio, suggests that taxpayers who were assessed additional tax were selected into treatment due to a negative shock to profitability in the year prior to the audit. The difference-in-differences approach is thus not suitable in this context.

FIGURE 4. Impact of audits on Schedule C net profit



5. Discussion and limitations

Using a range of non-experimental estimators, we find that audits have a significant impact on reported income: taxpayers receiving a positive recommended additional tax assessment increase their reports of taxable income dramatically (+120%), while those receiving no additional assessment respond by reporting less income in future years (-35%). The positive impact on members of experimental group E-PC (i.e., taxpayers that were assessed additional recommended tax) is likely due to some kind of deterrent effect (Alm et al., 2009).

Understanding the observed reduction in reported income among taxpayers in experimental group E-NC (i.e., taxpayers that were not assessed additional tax) is probably even more important. There are several plausible explanations for this finding. First, an experience of coercive enforcement activity could reduce tax morale among honest taxpayers, leading to the observed detrimental impact of audits on compliance among members of experimental group E-NC. Second, even if tax morale were unaffected by the examination experience, the audit process might provide currently compliant taxpayers with a "window" on potential opportunities for both legal and illegal tax avoidance. In addition, such taxpayers may infer that the risk of a future examination is low given that no adjustments were made during the recent audit. This newfound awareness of opportunities for reporting and paying lower taxes combined with a low perceived future audit risk could drive some taxpayers to report less income on subsequent returns. A third possibility is that the observed reduction in reported income might be attributable to dishonest taxpayers within this group whose misreporting was not detected during the audit. The experience of having undergone an audit without experiencing any sanction for noncompliance may have emboldened such taxpayers, resulting in even more aggressive future reporting behavior. Based on the available data, we are unable to pinpoint which of these explanations prevails. The observed reduction in compliance behavior suggests, in any case, that there is scope for improving the efficiency of audits. On the one hand, improved targeting of noncompliant returns and an improved capacity to detect noncompliance would seem likely to improve deterrence among cheaters. On the other hand, a better understanding of the psychological impact of audits on compliant taxpayers may lead to enhanced

examination approaches that mitigate the erosion of tax morale and maintain their incentives to comply.

It is constructive to consider limitations and potential extensions of the above analysis. First, we cannot rule out that our estimates are influenced by the economic downturn in 2008. Repeating the analysis for another, less turbulent, timespan would strengthen the credibility of the results. Second, the relatively short time horizon impedes the estimation of a dynamic model, which would allow a more accurate quantification of the decay rate of audit effects. Third, a range of additional analyses, looking at, for instance, the differential impact of alternative audit techniques (such as face-to-face vs. correspondence) or the differential response of low-income and high-income taxpayers, could provide important insights. Fourth, the existing propensity score matching approach relies on matching without replacement, meaning that a given control group member can be matched to at most one treatment group member. Although this approach can lead to improved precision in estimation, it can also result in increased bias. Intuitively, restricting control group members to no more than one match can adversely impact the quality of the matches to some treatment group members. Therefore, it may be productive to explore whether similar estimation results are achieved when one employs matching with replacement. In addition, the current Matched Difference-in-Differences approach compares the reporting behavior of each subset of the audit group (E-PC and E-NC) to the full set of matched controls. It may be worthwhile to compare the reporting behavior of experimental group E-PC just to the matched controls for this subset of taxpayers and, likewise, the reporting behavior of group E-NC just to the matched controls for this subset of taxpayers. Finally, it may be worth considering estimating the models separately for each examination activity class to see if the results are robust across classes.

References

Advani, A., W. Elming, and J. Shaw (2015). How Long-lasting are the Effects of Audits? Technical report, Discussion Paper No. 011-15, Tax Administration Research Center, University of Exeter.

Alm, J., B. R. Jackson, and M. McKee (2009, April). Getting the Word Out: Enforcement Information Dissemination and Compliance Behavior. *Journal of Public Economics 93*(3-4), 392–402.

Ashenfelter, O. (1978, February). Estimating the Effect of Training Programs on Earnings. *The Review of Economics and Statistics 60*(1), 47–57.

Blundell, R. and M. Costa Dias (2000, December). Evaluation Methods for Non-Experimental Data. *Fiscal Studies 21*(4), 427–468.

Braithwaite, V., K. Murphy, and M. Reinhart (2007, January). Taxation Threat, Motivational Postures, and Responsive Regulation. *Law & Policy 29*(1), 137–158.

Chabe-Ferret, S. (2014, August). Why Does Difference in Difference Matching Work? Technical report.

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2015, March). Once Bitten, Twice Shy? The Lasting Impact of IRS Audits on Individual Tax Reporting. Technical report.

Gemmel, N. and M. Ratto (2012, March). Behavioral Responses to Taxpayer Audits: Evidence from Random Taxpayer Inquiries. *National Tax Journal 65*(1), 33–58.

Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica 46*(4), 931–959.

Heckman, J. J. (1979, January). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153–161.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica 66*(5), 1017–1098.

Heckman, J. J. and R. Robb (1986). Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes. In H. Wainer (Ed.), *Drawing Inferences from Self-Selected Samples*. New York: Springer-Verlag.

Heckman, J. J. and J. A. Smith (1999, July). The Pre-Programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies. *The Economic Journal 109*(457), 313–348.

Internal Revenue Service (2015, March). *Internal Revenue Service Data Book, 2014*.

LaLonde, R. J. (1984). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. Technical report, Industrial Relations Center, Princeton University, Working Paper No. 183.

LaLonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review 76*(4), 604–620.

Rosenbaum, P. R. and D. B. Rubin (1983, April). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*(1), 41–55.

Smith, J. A. and P. E. Todd (2005, March). Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators? *Journal of Econometrics 125*(1-2), 305–353.

Wooldridge, J. M. (2010, October). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Appendix A. Additional tables

Table 8. Short-term impact of audits on taxable income (levels specification)

Dependent variable: taxable income 2008 - taxable income 2007						
Estimator	Baseline DD	Dynamic DD	Unrest. DD	Matched DD	Dynamic DD	Dynamic DD
Treated	109.151	637.189	695.058	-67.036	-3,172.458	118.234
	(903.361)	(911.577)	(902.420)	(1108.523)	(2563.045)	(893.239)
Treated:e	12762.614***	12130.803***	11616.740***	13383.543***	12914.715***	12368.663***
	(1151.681)	(1148.279)	(1139.743)	(1299.849)	(1137.809)	(1198.948)
Selection control (λ_1)					2118.879	
					(1559.561)	
Attrition control (λ_2)						3,887.346
						(2684.980)
Sum of rows 1 and 2	12871.765***	12767.992***	12311.798***	13316.507***	9742.257***	12486.897***
	(906.024)	(899.115)	(892.084)	(1141.892)	(2564.537)	(964.360)
Observations	6904	6903	6903	3960	6903	6903
Adj. R^2	0.029	0.061	0.082	0.037	0.054	0.054

TABLE 9. Long-term impact of audits on taxable income (levels specification)

Dependent variable: taxable income 2010 - taxable income 2007						
Estimator	Baseline DD	Dynamic DD	Unrest. DD	Matched DD	Dynamic DD	Dynamic DD
Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-627.405	-524.02	-360.786	-2481.577*	615.384	-489.323
	(1227.175)	(1235.941)	(1224.295)	(1431.889)	(3488.932)	(1216.307)
Treated:e	9646.843***	9266.843***	8424.823***	9556.641***	9717.567***	9239.317***
	(1564.506)	(1556.867)	(1546.266)	(1679.025)	(1548.837)	(1600.607)
Selection control (λ_1)					-749.637	
					(2122.944)	
Attrition control (λ_2)						3,452.438
						(2921.670)
Sum of rows 1 and 2	9019.439***	8742.823***	8064.037***	7075.064***	10332.951**	8749.994***
	(1230.793)	(1219.045)	(1210.272)	(1474.992)	(3490.962)	(1270.600)
Observations	6904	6903	6903	3960	6903	6903
Adj. R^2	0.008	0.044	0.064	0.008	0.03	0.03

TABLE 10. Long-term impact of audits on Schedule C net profit

Dependent variable: Schedule C net profit 2010 - Schedule C net profit 2007						
Estimator	Baseline DD	Dynamic DD	Unrest. DD	Matched DD	Dynamic DD	Dynamic DD
Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-2444.006**	-1,053.172	129.308	-2,344.188	-9025.484**	-1,588.093
	(1244.386)	(1241.802)	(1176.660)	(1490.196)	(3532.837)	(1214.548)
Treated:e	9606.398***	9027.610***	6496.244***	11298.333***	9341.106***	7607.858***
	(1632.600)	(1605.760)	(1523.607)	(1801.434)	(1593.725)	(1638.046)
Selection control (λ_1)					4690.006**	
					(2147.006)	
Attrition control (λ_2)						14090.140***
						(3201.101)
Sum of rows 1 and 2	7162.391***	7974.439***	6625.552***	8954.144***	315.622	6019.765***
	(1308.172)	(1278.716)	(1212.069)	(1600.375)	(3542.053)	(1320.785)
Observations	5717	5717	5717	3234	5717	5717
Adj.R^2	0.006	0.067	0.164	0.013	0.054	0.057

APPENDIX B. TECHNICAL DETAILS ON ESTIMATION STRATEGY

B.1. **The data generating process.** We assume that reported income is described by the following process in the absence of an audit:

$$Y_{it}^0 = \delta_t + \mu_i + u_{it}, \tag{3}$$

where δ_t is a macro shock affecting all taxpayers i in year t, individual-specific fixed effects are captured by μ_i, and u_{it} is a serially correlated error with zero mean. The unobservable propensity to evade is assumed to be constant over time and is therefore subsumed within μ_i. The impact of enforcement activity is given by

$$\alpha_i = \mathrm{E}[(Y_{it}^1 - Y_{it}^0)|e_i] \quad \text{where} \quad \alpha_i = \begin{cases} \alpha_e & \text{if } e_i = 1 \\ \alpha_{ne} & \text{otherwise,} \end{cases} \tag{4}$$

where Y_{it}^1 represents reported income following an audit, and e_i is a 1/0 indicator for individuals with a high propensity to evade. Reported income within the combined sample may therefore be expressed as:

$$\begin{aligned} y_{it} &= D_{it}Y_{it}^1 + (1 - D_{it})Y_{it}^0 \\ &= \delta_t + \mu_i + \alpha_i D_{it} + u_{it} \\ &= \delta_t + \mu_i + \alpha_{ne} D_{it} + (\alpha_e - \alpha_{ne}) e_i D_{it} + u_{it}, \end{aligned} \tag{5}$$

where D_{it} is equal to 1 if taxpayer i was audited prior to period t and zero otherwise.

B.2. **Difference-in-differences regression approach.** The difference-in-differences estimator exploits the longitudinal structure of panel data to eliminate individual-specific fixed effects. Subtracting reported income in period 0 (the baseline tax reporting period, prior to which no taxpayers have been audited) from reported income in a subsequent reporting period k (prior to which all taxpayers in the treatment group have been audited) yields:

$$\begin{aligned} (y_{ik} - y_{i0}) &= (\delta_k - \delta_0) + \alpha_{ne} D_{ik} + (\alpha_e - \alpha_{ne}) e_i D_{ik} + \varepsilon_{ik} \\ &= \alpha + \beta_1 D_i + \beta_2 e_i D_i + \varepsilon_{ik} \end{aligned} \tag{6}$$

where D_i is a 1/0 dummy for members of the treatment group, $\alpha = (\delta_k - \delta_0)$, $\beta_1 = \alpha_{ne}$, $\beta_2 = (\alpha_e - \alpha_{ne})$, $\varepsilon_{ik} = (\rho^k - 1)u_{i0} + \nu_{ik}$, ρ is the first-order serial correlation coefficient, and ν_{ik} is a white noise disturbance.

The above equation summarizes our Baseline Difference-in-Differences model for quantifying the impact of an audit on taxpayer reporting behavior. In this specification, β_1 represents the impact of an audit on taxpayers with a low propensity to evade, while β_2 captures the differential impact on taxpayers with a high propensity to evade. The total impact for the latter group is computed as the sum $(\beta_1 + \beta_2)$. The interaction term $e_i D_i$ is operationalized in our sample by setting this term equal to 1 in period k for those audit group members who received a positive additional recommended tax assessment (members of experimental group E-PC) and 0 otherwise.

By modeling the change in reported income between period 0 and period k, our baseline methodology effectively controls for unobserved fixed effects that impact the level of reported income and are also potentially associated with audit selection. It will thus return an unbiased estimate of the treatment impact if audit selection is independent of the remaining unobserved components of the differenced equation (i.e., the disturbance term ε_{ik}). As discussed below in the main text, this assumption is unlikely to hold in practice. Generally speaking, bias may arise due to selection on observables or selection on unobservables.

B.3. **Selection on observables.** The Baseline Difference-in-Differences approach fails to produce consistent estimates of the impact of an audit if audit selection depends on taxpayer characteristics that also influence reporting behavior. This problem is straightforward to address, however, if such characteristics are all observable. Specifically, one may consistently estimate the treatment impact by relying on the following "Unrestricted" Difference-in-Differences regression specification:

$$(y_{ik} - y_{i0}) = \alpha + \beta_1 D_i + \beta_2 e_i D_i + \gamma' Z_{ik} + \varepsilon_{ik}, \tag{7}$$

where Z_{ik} is a vector of observable factors that are associated with both income reporting and audit selection.[19] This approach yields unbiased estimates of impact of an audit if: (i) all relevant

[19]The specification is unrestricted in the sense that it allows for more complex dynamics in the income generating and reporting process. The Baseline Difference-in-Differences approach assumes that the change in reported income is independent of lagged reports. This assumption is relaxed in the unrestricted specification by directly including lagged income reports and other factors as additional regressors.

determinants of audits are observable, so that $\mathrm{E}[\varepsilon|e, D, Z] = \mathrm{E}[\varepsilon|e, Z] = 0$; and (ii) the conditional expectation of the error $\mathrm{E}[\varepsilon|e, Z]$ has a linear representation.

The Matched Difference-in-Differences methodology provides an attractive alternative to the Unrestricted Difference-in-Differences approach. The method also assumes that selection is on observables only. It does not, however, assume anything about the functional form of the error term. The matched approach therefore requires fewer parametric assumptions. The validity of this approach rests on the assumption that knowledge about one's audit status does not provide additional information about the income process once all observables are accounted for. More formally, it is required that (see, for example, Heckman et al., 1998):[20]

$$\mathrm{E}[(Y^0_{ik} - Y^0_{i0})|D_i = 1, Z_{ik}] = \mathrm{E}[(Y^0_{ik} - Y^0_{i0})|D_i = 0, Z_{ik}]. \tag{8}$$

This condition implies that the reporting behavior of those taxpayers in the control group who are observationally equivalent to the taxpayers in the audit group serves as a measure of how those in the latter group would have reported in the counterfactual state (i.e., of the subsequent reports that the audited taxpayers would have made had the audits not taken place). A challenge under this approach is to find controls that are observationally similar to audited taxpayers when the conditioning set is large. Rosenbaum and Rubin (1983) have shown that the dimensionality of the matching problem can be reduced by relying on the propensity score, the probability of treatment assignment conditional on the observables. Under the propensity scoring approach, treated subjects are matched to controls with comparable propensity scores. In practice, it is common to have some treatment subjects with propensity scores that are outside the range observed for untreated subjects. In such cases, the treatment subjects with these relatively extreme scores are excluded from the analysis, with the implication that the estimation results may not be representative of the excluded subjects.

The Matched Difference-in-Differences estimators for the impact of an audit for the positive-tax-change and no-tax-change experimental groups in our analysis are constructed as follows:

$$\hat{\alpha}_e = \frac{1}{N_{pc}} \sum_{i \in pc} (y_{ik} - y_{i0}) - \frac{1}{N_c} \sum_{j \in c} (y_{jk} - y_{j0}) \quad \text{and} \quad \hat{\alpha}_{nc} = \frac{1}{N_{nc}} \sum_{i \in nc} (y_{ik} - y_{i0}) - \frac{1}{N_c} \sum_{j \in c} (y_{jk} - y_{j0}), \tag{9}$$

where N_x is the number of taxpayers in group x; x takes the value c for the matched control group (of unaudited taxpayers), pc for the positive-tax-change experimental group, and nc for the no-tax-change experimental group.

B.4. **Selection on unobservables.** The above-sketched techniques build on the assumption that selection is only on observables; i.e., that there are no unobserved factors that influence both the likelihood of an audit and the magnitude of reported income. Heckman (1978) proposed a technique to deal with selection on unobservables. This method attempts to address the lack of observability by imposing stronger distributional assumptions. For simplicity of exposition, we assume that there are no relevant observed determinants of taxpayer reporting behavior in the absence of an audit.[21] The stronger distributional assumption imposed by Heckman's approach is that the regression disturbance (ε_{ik}) in the specification describing income reporting behavior and the disturbance term in a separate (probit) equation describing the probability of an audit prior to the reporting period follow the bivariate normal distribution with correlation coefficient ρ; the standard error

[20] Under the linearity assumptions of the Unrestricted Difference-in-Differences estimator, this condition would be equivalent to the requirement for that model that $\mathrm{E}[\varepsilon|e, D, Z] = \mathrm{E}[\varepsilon|e, Z]$, as we required for the Unrestricted Difference-in-Differences estimator.

[21] To incorporate such factors in the analysis described in this section, one would simply include them as additional explanatory variables in the regression specification.

of the former disturbance is represented by σ_ε and the standard error for the latter disturbance is normalized to one. Conditional on the audit group indicator (D_i) and the set of explanatory variables (Z_i) in the probit equation describing the likelihood of an audit, the expectation of the regression disturbance (ε_{ik}) is then equal to:

$$\text{(10)} \qquad \mathrm{E}[\varepsilon_{ik}|Z_i] = \begin{cases} \rho\sigma_\varepsilon \frac{\phi(\gamma' Z_i)}{\Phi(\gamma' Z_i)} & \text{if } D_i = 1 \\ \\ \rho\sigma_\varepsilon \frac{-\phi(\gamma' Z_i))}{[1-\Phi(\gamma' Z_i)]} & \text{otherwise,} \end{cases}$$

where $\phi(x)$ and $\Phi(x)$ denote values of the probability density function and the cumulative distribution function of the standard normal distribution, respectively, when evaluated at x, and γ is the vector of coefficients in the probit specification for the likelihood of an audit. In our empirical approach we rely on a two-step estimation procedure. In the first stage, we estimate the coefficient vector γ using a probit model for the likelihood of an audit in period 0. We then use this estimate to construct a measure of λ_{id} for each taxpayer in our sample, where λ_{id} is defined as:

$$\text{(11)} \qquad \lambda_{id} = \begin{cases} \frac{\phi(\gamma' Z_i)}{\Phi(\gamma' Z_i)} & \text{if } D_i = 1 \\ \\ \frac{-\phi(\gamma' Z_i))}{[1-\Phi(\gamma' Z_i)]} & \text{otherwise.} \end{cases}$$

In the second stage, we incorporate our measure of λ_{id} as an additional explanatory variable in the Baseline Difference-in-Differences specification:

$$\text{(12)} \qquad (y_{ik} - y_{i0}) = \alpha + \beta_1 D_i + \beta_2 e_i D_i + (\rho\sigma_\varepsilon)\lambda_{id} + \eta_{ik},$$

where the new error component η_{ik} is, by construction, independent of the regressors. In this augmented specification, $(\rho\sigma_\varepsilon)$ represents the coefficient of λ_{id} that is to be estimated.

B.5. **Attrition.** Our descriptive statistics reveal that a significant share of taxpayers do not file a Schedule C return after TY2007. Given that the change in reported income over time services as the basis for our empirical analysis and this change cannot be constructed in the absence of a post-TY2007 return, a portion of our sample is simply neglected by the difference-in-differences approach. To account for such attrition, we employ a similar methodology to that used to control for selection on unobservables. Following Heckman (1979), the conditional expectation of the regression disturbance in our Baseline Difference-in-Differences specification when attrition is present is expressed as:

$$\text{(13)} \qquad \mathrm{E}[\varepsilon_i|F_i = 1, Z_i] = \rho\sigma_\varepsilon \frac{\phi(\gamma' Z_i)}{\Phi(\gamma' Z_i)},$$

where $F_i = 1$ if taxpayer i filed a Schedule C return in period k, Z_i now represents a set of explanatory variables in a probit specification for the likelihood of filing a Schedule C return in this period, and γ is the coefficient vector for this model. As with our approach to modeling selection with unobservables, we follow a two-step estimation procedure to account for attrition. In the first step, we estimate the probability of filing a Schedule C return in period k as a function of Z. We then use the estimated coefficients to construct a measure of $\lambda_{ia} = \frac{\phi(\gamma' Z_i)}{\Phi(\gamma' Z_i)}$. In the second stage, we incorporate this term as an additional explanatory variable in our Baseline Difference-in-Differences regression specification:

$$(y_{ik} - y_{i0}) = \alpha + \beta_1 D_i + \beta_2 e_i D_i + (\rho\sigma_\varepsilon)\lambda_{ia} + \eta_{ik}. \tag{14}$$