# The Dynamic Effects of Tax Audits

Arun Advani* William Elming† Jonathan Shaw‡

January 28, 2021

**Abstract**

We study the effects of audits on long run compliance behaviour, using a random audit program covering more than 53,000 tax returns. We find that audits raise reported tax liabilities for five years after audit, effects are longer lasting for more stable sources of income, and only individuals found to have made errors respond to audit. 60-65% of revenue from audit comes from the change in reporting behaviour. Extending the standard model of rational tax evasion, we show these results are best explained by information revealed by audits constraining future misreporting. Together these imply that more resources should be devoted to audits, audit targeting should account for reporting responses, and audit threat letters miss a key benefit of audit.

JEL codes: D04, H26, H83

Keywords: tax audits, tax revenue, tax reporting decisions, income tax, self assessment, HMRC

*University of Warwick, the Institute for Fiscal Studies (IFS), the Tax Administration Research Center (TARC), and the Centre for Competitive Advantage in the Global Economy (CAGE). A previous version of this paper circulated as 'How long-lasting are the effects of audits?' The authors thank Michael Best, Richard Blundell, Tracey Bowler, Monica Costa Dias, Dave Donaldson, Mirko Draca, James Fenske, Clive Fraser, Claus Kreiner, Costas Meghir, Gareth Myles, Matthew Notowidigdo, Áureo de Paula, Andreas Peichl, Imran Rasul, Chris Roth, Joel Slemrod, and seminar participants at the Tax Systems Conference, Royal Economic Society, Louis-André Gérard-Varet, European Economic Association, Warwick Applied, OFS Empirical Analysis of Tax Compliance, International Institute of Public Finance, Econometric Society European Meetings, and National Tax Administration Conferences, for helpful comments. We also thank Yee Wan Yau and the HMRC Datalab team for assistance with data access.  Correspondence: Dept of Economics, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL. Email: economics@arunadvani.com.

†IFS and TARC at the time of involvement in this work.

‡Financial Conduct Authority.

# 1 Introduction

Audits are a widely used public policy tool for reducing corruption (Bobonis et al., 2016; Avis et al., 2018), improving public service delivery (Zamboni and Litschig, 2018; Lichand et al., 2019; Gerardino et al., 2020), ensuring environmental standards (Duflo et al., 2013, 2018), and improving tax compliance (Kleven et al., 2011; Pomeranz, 2015; Asatryan and Peichl, 2017; Bergolo et al., 2020; among others). But audits are costly, so determining how many to do and how best to allocate them are key policy questions (Slemrod and Yitzhaki, 2002). In tax, the standard approach to setting the number of audits is to compare their costs with the expected missing tax uncovered at audit – the static gain from an audit (Allingham and Sandmo, 1972; Kolm, 1973; Yitzhaki, 1987; Bloomquist, 2013). However, audits may change taxpayer behaviour. A field experiment in Denmark, which followed taxpayers for a year after audit, found an increased reported liability worth 55% of the audit adjustment (Kleven et al., 2011). This suggests that static gains may understate the total gains from audit. However, without a longer horizon it is hard to know by how much, or whether this effect is even reversed in subsequent years, as some lab experiments suggest (Maciejovsky et al., 2007; Kastlunger et al., 2009).

This paper studies the long run effect of tax audits on taxpayer compliance behaviour. We combine confidential administrative data on the universe of UK tax filers over thirteen years with a randomised audit programme. We show three main results. First, audits raise subsequent tax reports, but the effect declines to zero over five to eight years. The aggregate additional revenue after audit is at least 1.5 times the underpayment found at audit, implying substantially more resources should be dedicated to audit than a static comparison would suggest. Second, the revenue gain is longer lasting for more stable income sources. This highlights the importance of dynamics for *targeting* audits, as well as for setting their level. Third, using an event study strategy we show that these effects are driven by individuals who were found to be under-reporting, while there is no response for those found to have reported correctly. These three results are explained by audits providing the tax authority with information about a taxpayer's income at the time of audit. This makes later misreporting more difficult, particularly for stable income sources.

To estimate the long run effect we exploit a random audit programme run by the UK tax authority (HM Revenue and Customs, HMRC). Over 53,000 individual tax filers were unconditionally randomly selected for audit by the programme between 1998/99 and 2008/09, allowing us to address the common concern that audits are typically targeted towards taxpayers believed to be under-reporting. Similar to Denmark (Kleven et al., 2011) and in contrast to the US (Slemrod et al., 2001; DeBacker et al., 2018; Perez-Truglia and Troiano, 2018), taxpayers are not told these audits are random. This is important as taxpayers may respond differently – likely less – to audits they know are random, relative to when they think the tax authority is concerned about something on their return. We combine this audit data with data on the universe of UK self assessment taxpayers – individuals who self-file taxes rather than having all tax collected via withholding – from 1998/99 to 2011/12. This allows us to follow individuals for many years after audit. For our first identification strategy we construct a control group for each year of the programme from individuals who could have been selected for a random audit that year but were not. We then study the difference in reporting behaviour over time.

Our first result is that *dynamic effects* are positive and substantial: taxpayers report higher levels of tax for five to eight years after audit. We see an initial increase, and then steady decline, in total tax reported over time. By eight years after audit there is no difference in average tax paid between audited and unaudited taxpayers, though differences are not statistically significant beyond five years. 60-65% of the total revenue received as a result of audit comes from this change in reporting behaviour. Taking into account this effect, tax authorities should do many more audits: accounting for dynamic effects even random audits provide a return equal to 80% of their cost to the tax authority. Given the recent focus on the value of audits purely as a threat (Slemrod et al., 2001; Fellner et al., 2013; Dwenger et al., 2016; Mascagni, 2018; Bergolo et al., 2020; Lichand et al., 2019), this highlights a benefit of actually performing the audits.

Second, we show that dynamic effects fall to zero slower for more stable income sources. Pension income, which is highly autocorrelated ('stable') in the absence of audit, responds permanently. At the other extreme, the effect on self-employment and dividend income returns to zero by three years after audit. This is important for two reasons. First, it has implications for the targeting of audits. Going after a smaller suspected discrepency on a more stable income source can have high returns, once dynamic effects are included. Reauditing is also more likely to produce additional yield for individuals with less stable income sources. Second, it will be relevant for understanding why people respond to audits, as we describe below. A natural concern in treating this difference causally, and using it to interpret behaviour, is that individuals with different types of income may respond differently. To account for this we also use pairwise comparisons of income sources within individuals who have both sources, and demonstrate that the less stable source still declines more quickly.

Third, we show that audits only change the behaviour of those who are found to have misreported. To do this we use an event study approach. We compare individuals who were audited at some point in our sample and who ultimately all had the same audit outcome, for example were found to be non-compliant. Allowing for individual and calendar time fixed effects, the comparison is essentially between those whose non-compliance has already been uncovered by a random audit and those who will have it uncovered in the future. We find that being audited only changes the behaviour of those who are found to have misreported, and this is true whether or not they received a penalty. Importantly, this tells us that the effect of audits comes not merely from scaring all taxpayers into paying more, but specifically from those who were previously misreporting. It also allows us to rule out audits reducing tax reports, even for those who were found compliant, in contrast with results using alternative identification strategies (Gemmell and Ratto, 2012; Beer et al., 2019).

These results are explained by audits providing the tax authority with information at a point in time, which constrains future misreporting. To see this we extend the canonical model of tax evasion (Allingham and Sandmo, 1972; Yitzhaki, 1987; Kleven et al., 2011), to incorporate (simple) dynamics in the response to audit. This allows us to study the distinct predictions of three different mechanisms that might drive changes in reporting: (1) changes in beliefs about the underlying audit rate or penalty for evasion ('belief updating'); (2) changes in the perceived reaudit risk following audit ('reaudit risk'); and (3) updates to the information held by the tax authority ('information'). Kleven et al. (2011) note that their observed increase in reported tax one year after audit could be explained by some combination of beliefs and reaudit risk, but they cannot disentangle the two. We note that a response to belief updating should be permanent, as taxpayers revise the expected cost of non-compliance (up or down). This is inconsistent with the declining pattern of dynamic effects we see. A response to reaudit risk would decline over time. Whether it took the form of a 'bomb crater' (Mittone, 2006) – that the probability of audit is lower in the years post audit before rising back to baseline – or a worry of higher levels of short term scrutiny, we should see the same effect across all income sources. We see a positive dynamic effect, ruling out 'bomb craters', and see a differential decline across income sources, even within individual, ruling out an effect driven purely by reaudit risk. Instead we propose a third, novel, possibility. As Kleven et al. (2011) note, when taxpayers know the tax authority has access to third party information about some income source, they are much less likely to underreport. Similarly, when the tax authority performs an audit, it gets a snapshot of income at a point in time. Implausibly large deviations in reported income in following years are likely to trigger an audit. As time passes, the snapshot becomes less informative about what current income is likely to be. This is particularly true for less stable income sources.

In this case we should see a decline in dynamic effects over time, with less stable income sources showing a faster decline. We should also only see responses from individuals who were found to have misreported, since no new information about the other taxpayers is revealed to the authority. These are precisely the patterns that are observed.

Our results imply that audits themselves are important, beyond the 'fear' or 'threat' of audit. Much of the recent literature studying the administration of taxes and the policies that can improve taxpayer compliance has focused on 'letter experiments': how different forms and content of information provided to taxpayers can change their behaviour (see Blumenthal et al., 2001; Slemrod et al., 2001, for early work, and Mascagni, 2018; Alm, 2019; Pomeranz and Vila-Belda, 2019; Slemrod, 2019 for recent surveys of this literature). These all aim to change the perceived probability of audit. They have the benefit that they are a very low cost policy for a tax authority, yet show substantial (short term) gains. For example, Bergolo et al. (2020) find, in the context of VAT in Uruguay, that firms do not respond to the actual probability of audit when sent letters informing them of this. Instead firms increase compliance because thinking about the audit scares them into compliance. This raises a question: can high levels of compliance be achieved while reducing the number of audits, by directing more resources towards information campaigns? Our results imply that this is harder than previously thought, as much of the gain from audit is the change in behaviour it promotes. This response is driven by the information received by the authority through actually conducting the audit. Threat letters do not provide this information benefit. To understand any substitutability with audits, more information is needed on the long-term effects of such letters: for how long do threats raise compliance, and can repeated threats continue to maintain high compliance rates?

In contrast, third party information is a more direct substitute for audits. Recent work has shown the importance (and limits) of third party information for improving compliance (Kleven et al., 2011; Pomeranz, 2015; Kleven et al., 2016; Carrillo et al., 2017; Slemrod et al., 2017; Naritomi, 2019). Since this directly reduces the information asymmetry between taxpayer and authority, it will also reduce the information value of audits, which drives the dynamic effects. Conversely, for income sources where third party information can be hard to come by, audits can be a partial alternative to gathering information from other sources. Not only will they improve contemporaneous compliance, but also reduce the scope for future non-compliance. This contrasts with work on firms which finds complementarity between monitoring and enforcement (Almunia and Lopez-Rodriguez, 2018).

We find no evidence of 'backfire' effects, where audits reduce compliance. Worries about backfire effects are common across areas of tax policy (Perez-Truglia and Troiano, 2018). In our context they raise the risk that poorly targeted audits may reduce compliance. Gemmell and Ratto (2012) suggest some reduction in tax reported by individuals who are audited and found compliant, relative to individuals not audited. Similar results are found in the US by Beer et al. (2019), using a matched difference-in-difference approach. Our event study strategy allows for potential differences in unobservable characteristics between compliant and non-compliant individuals, and finds no backfire. The difference in our results, compared to existing work, also suggests that unobservable differences are important in explaining compliance behaviour. Since we find no reduction in overall tax paid, it also suggests that lab experimental evidence of bomb crater effects is not reflected in real-world settings (Maciejovsky et al., 2007; Kastlunger et al., 2009), although we note that not all lab experiments find evidence of such effects (Choo et al., 2013).

Finally, we provide a new theoretical mechanism for why audits have the observed effects. Understanding what motivates compliance is a key question for public policy, and there are rich debates on the extent to which moral versus economic calculations drive behaviour (Alm, 2019). We focus on the narrower question of why *audits* affect compliance, and find information is the key. To do this we use evidence from random audits, both to look at the time path of dynamic effects across income sources and the effects by audit outcome. Though earlier work has (separately) studied both of these issues, we show how they can be used to understand why audits change behaviour.[1] Our results complement those of Bergolo et al. (2020) and Lichand et al. (2019) who find *threat* of audit works through a fear and belief-updating respectively. In contrast receipt of audit works through a change in ability to misreport without being caught, an effect that cannot occur in the absence of actual audit.

The remainder of the paper is organised as follows. Section 2 outlines the policy context and data sources. Section 3 provides evidence on who is non-compliant. Section 4 shows how audits affect reporting behaviour in overall tax, and by different income sources. Section 5 uses an alternative identification strategy to estimate the impact by audit outcome. Section 6 outlines a model of tax evasion with dynamics in the response to audits, to show which mechanisms might rationalised the observed behaviour. Section 7 concludes.

[1] A number of studies consider dynamic effects for one or two years after audit (Long and Schwartz, 1987; Erard, 1992; Tauchen et al., 1993; Kleven et al., 2011; Løyland et al., 2019). Concurrently with this study, DeBacker et al. (2018) have a longer (six year) horizon, and also consider income stability, albeit with US audits where taxpayers are explicitly told they are random, which Slemrod (2019) notes 'would likely trigger different revaluations of how likely a future audit is, and therefore trigger different behavioural changes' (a similar point is made in Kleven et al., 2011). Effects by audit outcome are studied by Gemmell and Ratto (2012) and Beer et al. (2019).

# 2 Context and Data

## 2.1 The UK self assessment tax collection and enforcement system

In this paper, we focus on individuals who file an income tax self assessment return in the UK. Over our sample period (1999-2012) this comprised around 9 million individuals, one-third of all individual income taxpayers in the UK.[2] Income tax is the largest of all UK taxes, consistently contributing a quarter of total government receipts over this period. Most sources of income are subject to income tax, including earnings, retirement pensions, income from property, interest on deposits in bank accounts, dividends, and some welfare benefits. Income tax is levied on an individual basis and operates through a system of allowances and bands. Each individual has a personal allowance, which is deducted from total income. The remainder—taxable income—is then subject to a progressive schedule of tax rates. Table 1 shows the share of individuals in our sample reporting non-zero values for each component of income. When we later study income components separately, we focus on those components where at least five per cent of the population report non-zero values.

Since incomes covered by self assessment tend to be harder to verify, there is a significant risk of non-compliance. As a result, HM Revenue and Customs (HMRC, the UK tax authority) carries out audits each year to deter non-compliance and recover lost revenue. HMRC runs two types of audit, 'targeted' (also called 'operational') and 'random'. Targeted audits are based on perceived risks of non-compliance. Random audits are unconditionally random from the population, and are used to ensure that all self assessment taxpayers face a positive probability of being audited, as well as to collect statistical information about the scale of non-compliance and predictors of non-compliance that can be used to implement targeting.

The timeline for the audit process is as follows. The tax year runs from 6th April to 5th April. Shortly after the end of the tax year, HMRC issues a 'notice to file' to taxpayers who they believe need to submit a tax return. This is based on information that HMRC held shortly before the end of the tax year. Random audit cases are provisionally selected from the population of individuals issued with a notice to file. The deadline by which taxpayers must submit their tax return is 31 January the following calendar year (e.g. 31 January 2008 for the 2006/07 tax year). Once returns have been submitted, HMRC deselects some random audit cases (e.g. due to severe illness or death of the taxpayer). At the same time, targeted audits are selected on the basis of the information provided

[2] Filers include self-employed individuals, those with incomes over £100,000 (lower at the start of the sample period), company directors, landlords, and many pensioners. The remainder have all their income tax collected directly via withholding, so are not required to file. Note that UK tax years run across calendar years—we denote tax years using the later year.

in self assessment returns and other intelligence. Random audits are selected before targeted audits, and individuals cannot be selected for a targeted audit in the same tax year as a random audit. The list of taxpayers to be audited is passed on to local compliance teams who carry out the audits. Up to and including 2006/07, audits had to be opened within a year of the 31 January filing deadline, or a year from the actual date of filing for returns filed late. For tax returns relating to 2007/08 or later, audits had to be opened within a year of the date when the return was filed. Taxpayers subject to an audit are informed when it is opened, but they are not told whether it is a random or targeted audit, in contrast to work done with US random audits (Long and Schwartz, 1987; DeBacker et al., 2018). Even after audit, taxpayers are limited in what they can learn about the audit process, since no details of the programme are made public.[3] Approximately one third of taxpayers on the list passed on to local compliance teams end up not being audited, largely due to resource constraints.[4]

Those who are audited initially receive a letter requesting information to verify what they have reported. If this does not provide all the required information, the taxpayer receives a follow-up phonecall, and ultimately in-person visits until the auditor is satisfied.

Where errors are uncovered, individuals are required to pay the additional tax due, and interest. If non-compliance is deemed to be deliberate, the taxpayer might also face an additional penalty of up to 100% the value of the underpaid tax.

## 2.2 Data sources

We exploit data on income tax self assessment random audits together with information on income tax returns. This combines a number of different HMRC datasets, linked together on the basis of encrypted taxpayer reference number and tax year.

Audit records for tax years 1998/99 to 2008/09 come from CQI (Compliance Quality Initiative), an operational database that records audits of income tax self assessment returns. It includes operational information about the audits, such as start and end dates, and audit outcomes: whether non-compliance was found, and the size of any correction, penalties and interest.

We track individuals before and after the audit using information from tax returns for the years 1998/99 to 2011/12. This comes from two datasets: SA302 and Valid View. The SA302 dataset contains information that is sent out to taxpayers summarising their income and tax liability (the SA302 tax calculation form). It is derived from self assessment returns, which have been put through a tax calculation process. It contains information about total income and tax liability as well as a

[3] Until the publication of this study, even the audit rates were not public information.

[4] We address the implications for identification in Subsection 4.1.

breakdown into different income sources: employment earnings, self employment profits, pensions, and so on. For all of these variables, we uprate to 2012 using the Consumer Prices Index (CPI) to account for inflation, and trim the top 1 per cent to avoid outliers having an undue impact on the results.[5] We supplement these variables with information from Valid View, which provides demographics and filing information (e.g. filing date). Note that we cannot identify actual compliance behaviour after the audit: the number of random audit taxpayers that are re-audited is far too small for it to be possible to focus just on them.

An explicit control group of 'held out' individuals was not constructed at the time of selection for audit. We therefore draw control individuals from the pool of individuals who actually filed a tax return (i.e. those who appear in SA302). This creates some differences in the filing history between those selected for audit and those who we deem as controls. In a given year, first time filers may be issued a notice to file *after* selection for audit has taken place. They may also end up back-filing one or two returns. Since we cannot directly observe the first year in which a notice to file was issued, in our empirical strategy it is necessary for us to control for the length of time each taxpayer has been in self assessment. More details – including tests to demonstrate this ensures samples are balanced – are given in Subsection 4.1 below.

# 3 Tax Evasion in the UK

In this section we first provide some descriptives on the probability and timeline of audits. We then show that there is significant non-compliance among individual self assessment taxpayers, both in the share of taxpayers who are found non-compliant and the share of tax that is misreported. More than one third of self assessment taxpayers are found to be non-compliant, equal to 12 per cent of all income taxpayers.

## 3.1 Audit descriptives

Figure 1 shows the share of individuals per year who face an income tax random audit over the period 1998/99 to 2008/09. On average over the period the probabilities of being audited are 0.04 per cent (four in 10,000) for random audits and 2.8 per cent for targeted audits.

Table A1 provides some summary statistics for lags in, and durations of, the audit process among random audit cases. As described above, up to and including the 2006/07 return, HMRC had to begin an audit within 12 months of the 31 January filing deadline; since then, HMRC has had to

[5] In Appendix C.2 we show our results are robust to alternative levels of trimming.

begin an audit within 12 months of the filing date. The average lag between when the tax return was filed and when the random audit was started is 8.9 months, but 10 per cent have a lag of 14 months or more. The average duration of audits is 5.3 months, but 10 per cent experience a duration of 13 months or more. Taken together, this means that the average time between when a return is filed and when the audit is concluded is 14.3 months but there is a long tail for whom the experience is much more drawn out: for almost 10 per cent it is two years or more. This means that individuals will generally have filed at least one subsequent tax return before the outcome of the audit is clear and some will have filed two tax returns. This will be relevant for interpreting the results in Section 4.

### 3.2 Evidence of non-compliance

We begin by studying the direct results of random audits, using data on 34,630 completed random audits of individual self assesssment taxpayers from 1998/99 to 2008/09.[6] Table 2 summarises the outcomes of these random audits. More than half of all returns are found to be correct, and 11 per cent are found to be incorrect but with no underpayment of tax, but 36 per cent are 'non-compliant' i.e. incorrect and have a tax underpayment.[7] Whilst this is a much higher rate of non-compliance than has been found in other developed country contexts, it should be noted that the self assesssment tax population is a selected subset of all taxpayers. In particular, it covers those for whom simple withholding of income at source is not sufficient to collect the correct tax. This may be either because some income cannot be withheld e.g. property or self-employed income, or because PAYE struggles to assign the correct withholding codes as for people with multiple sources of pension income. Despite this, since self assesssment taxpayers make up a third of all UK taxpayers, this implies an overall non-compliance rate of 8 to 12 per cent among all taxpayers.[8]

Turning to the intensive margin, the average additional tax owed among the non-compliant is £2,314, or 32 per cent of average liabilities. Since just over a third of random audits find evidence of non-compliance, the average additional tax owed from an audit is then £826.[9] However, the distribution is heavily skewed: 60 per cent of non-compliant individuals owe additional tax of £1,000

[6] 53,400 cases were selected for audit over the period, of which 35,630 were implemented.

[7] Incorrect with no underpayment includes those who, for example, owed no taxes since they had legitimate losses, but who had overstated those losses so would owe less in future years. Anecdotally, it also includes some cases where actual overpayments of tax were made, although we cannot separately identify which.

[8] This is a lower bound, since it assumes everyone who should be in self assessment does register, all non-compliance is picked up at audit, and those who do not need to register are also fully compliant. The range from 8 to 12 per cent depends on the assumptions made about the implementation of audits. If among those selected for audit, implementation of audit were random, this would imply a 12 per cent non-compliance rate. On the other hand, if there is perfect compliance among those for whom audits were not implemented, this would imply an 8 per cent rate.

[9] This is the additional tax owed. A further £101 is owed, on average, in penalties. This is highly concentrated, with less than 7% of those audited owing any penalty amount.

or less, whilst four per cent owe more than £10,000. In terms of total revenue, those owing £1,000 or less make up only 9 per cent of the under-reported revenue; the four per cent owing more than £10,000 collectively owe more than 42 per cent of the revenue. Equity concerns around non-compliance are well-known: it is seen as unfair that some are not 'paying their fair share'. But this variation in non-compliance is also important for economic efficiency. Non-compliant individuals previously acted as though there was a lower tax rate. This makes their activities seem relatively more productive than those of compliant individuals, so can lead to resource misallocation.

# 4 Dynamic Impacts of Audits

In this section we establish two main results. First, we show that audits lead to an increase in reported incomes and taxes in subsequent years. Looking at total income and total tax this increase lasts five to eight years after the tax year for which the audit was done. Second, we show variation in this impact by income source. In particular, more autocorrelated income sources (such as pensions) seem to respond permanently to audit. In contrast, income sources which are less autocorrelated, such as self-employment income, more quickly return to baseline. This second result will later help explain *why* we see these dynamic responses. Before describing these results in detail, we first discuss the empirical approach taken. Briefly, we compare individuals selected for random audit with those not selected but who could have been selected. We control for filing history, to account for the way the sample was selected.

## 4.1 Estimation

To understand how audits affect future tax receipts we want to estimate the change in tax paid in the years after audit that is caused by the audit. To recover this we make use of the 'random audits programme' run by the tax authority (HMRC). This programme selects for audit a random sample of taxpayers from the pool of taxpayers known to be required to file for a given tax year. One can therefore compare those selected for audit with others who were not selected but who could have been.

In each audited tax year we select a sample of individuals who were not audited and could have been. We assign them a "placebo audit" for that tax year. We can then compare them over time to individuals actually selected for audit for that year. Our sample therefore consists of individuals who were selected for random audit in some year between 1999 and 2009, and individuals who could have been selected in those same years but were not. Our data on tax returns goes up to 2012.

For every individual selected for audit in a given tax year, we draw six control individuals from the population of those who could have been audited in the same tax year.[10]

In practice a little more than two-thirds of those selected for random audit are actually audited. This is explained by the high workload faced by the compliance teams implementing audits. Additionally, a small fraction of the control group (around two per cent) is also audited. Random audits are selected before targeted audits, and no explicit control group was constructed to 'hold out' some individuals from targeting. To our knowledge, in prior work only (Kleven et al., 2011) have an explicit control group. This explains why they can only study a single year after audit – tax authorities are unwilling to hold off on high value audits for multiple years. Hence we compare those selected for a random audit to a 'business as usual' group, rather than a pure control group. This will tend to reduce the estimated impacts, since individuals in the control group who are most likely to be non-compliant are audited.

In the empirical work to follow, we focus on the local average treatment effect (LATE), instrumenting receipt of audit with selection for random audit. This is the relevant number for a tax authority thinking about simultaneously expanding the size of the random audit programme and the number of auditors. It gives the average impact $h$ years after audit for an additional random audit case that might be worked, against which the cost of the audit would be compared.

One limitation of our data is a slight mismatch between our treated and control samples in terms of their probability of filing in previous years, for reasons relating to the audit timeline and when they were first issued a notice to file, as described in Subsection 2.2. This can be seen in Table A3 which documents (unconditional) sample balance between five and one years before audit, for income and tax totals, income components, and individual characteristics. Overall balancing statistics suggest that the samples are fairly well balanced: the p-value of the likelihood-ratio test of the joint insignificance of all the regressors is 0.181, while the mean and median absolute standardised percentage bias across all outcomes of interest are low at 2.4 per cent and 1.7 per cent respectively.[11] However, the likelihood of being in the sample in previous years ('survival') differs between our treatment and control groups. This difference is consistent with how the treatment and control

[10] In principle the entire population of taxpayers who could have been audited could have been used. However, since the data could be accessed only in a secure facility at the tax office, computational constraints given the available hardware limited the sample size that could be used.

[11] The standardised percentage bias is the difference in the sample means between treated and control groups as a percentage of the square root of the average of the sample variances in the treated and control groups (see Rosenbaum and Rubin, 1985). Rubin's B and R statistics are also well within reasonable thresholds to consider the samples to be balanced, at 10.8 and 0.983 respectively. Rubin's B is the absolute standardised difference of the means of the linear index of the propensity score in the treated and control group. Rubin's R is the ratio of treated to control variances of the propensity score index. Rubin (2001) recommends that B be less than 25 and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced.

groups were selected, so might reflect real differences in the samples. We therefore include controls for presence in the data in the years before audit. Table 3 shows that once we condition on past survival the sample is balanced.

We therefore estimate the following specification:

$$Y_{ihs} = \sum_{h=-5}^{8} \alpha_h \eta_h + \sum_{h=-5}^{8} \beta_h \eta_h D_i + \sum_{s=1999}^{2012} \gamma_s T_s + \sum_{s=-4}^{-1} \delta_s S_{is} + \varepsilon_{ihs} \qquad (1)$$

where $Y_{ihs}$ is the outcome for individual $i$, $h$ years after the tax year selected for audit (with control observations having $h = 0$ for the tax year for which they were drawn as controls), when current calendar year is $s \equiv t+h$. $\eta_h$ are indicators for being $h$ years after the tax year selected for audit; $D_i$ is an indicator for whether the individual is actually audited; $T_s$ is a calendar time indicator for tax year $s$; and $\{S_{i,-1}, \ldots, S_{i,-4}\}$ are indicators for whether the individual was in the data in each of the four years before audit. The error term, $\varepsilon_{ihs}$, is clustered at the individual level. Audit status, $D_i$, is instrumented by (random) selection for audit, $Z_i$. The coefficients of interest are $\beta_h$ $\forall h$. These estimate the impact of the audit on the outcome variable $h$ years after the tax year selected for audit, measured as the difference in the mean outcome for those actually audited and those who would have been audited only if selected for a random audit.

### 4.2 Overall impact of audits

Beyond the direct effects of the audit, described in Section 2, we also see clear evidence of *dynamic* effects. Comparing individuals who were randomly selected for audit with individuals who could have been (but were not) selected, those selected for audit on average report higher levels of tax owed in the years after audit. Figure 2 shows the estimated impact on those who were actually audited (i.e. the LATE). The difference in the share audited between the treated and control group is around 66 percentage points, so the LATE is around 1.5 times the intention to treat estimate.

The impact of an audit peaks two years after the tax year for which the audit is conducted. This is consistent with the fact that many audits are not started until after the following year's tax return has already been submitted.[12] Reported tax among audited taxpayers is significantly greater

[12] In our sample, almost a quarter of audits are not opened for more than 12 months from the date of filing (see Table A1). Additionally, there can be some lag between the tax authority 'taking up' a case for audit and notification being received by the taxpayer. If taxpayers each consistently file at the same time every year, this implies at least one quarter would have filed without knowledge of the audit. More than half will have filed without knowing the result of the audit (Table A1). One could instead set $h = 0$ as the time at which audit begins, but this information is not available for controls, so risks creating bias if the timing of opening audits among individuals selected for audit is non-random.

than among non-audited taxpayers for five years after the audit, and the point estimate appears to decline relatively smoothly, getting close to zero by the eighth tax year after the audited year.

From Figure 2 we can estimate how much revenue audits raise on average by changing the behaviour of audited individuals. Over the five (eight) years after the audited year, the dynamic effects bring in an additional £1,230 (£1,530), 1.5 (1.8) times the direct effect of audit. Although taxpayers in the US are explicitly told that the random audits are random, DeBacker et al. (2018) find a similar ratio between direct and indirect effects of audit. Ex ante one might have expected smaller behavioural effects, since taxpayers are aware that the authority is not acting based on any suspicion of wrongdoing. Our exploration of the mechanism driving these dynamics will explain why, ex post, these effects should be so similar: the dynamics are driven by constraints to misreporting caused by audit, rather than belief-updating or perceived reaudit risk which may both respond to the reasoning behind the audit.

These dynamic effects highlight the policy importance of studying the long term impact of audits: when determining the audit strategy, the revenue raising effects of audits would be grossly understated without considering the impact on future behaviour. This would imply too few audits taking place.

It is important to note that the optimal number of audits will in general not equate the marginal return on audit to the marginal cost of an audit. Audits require real resource costs, while the direct benefits are a transfer of resources from citizens to the state (see Slemrod and Yitzhaki (1987) for a longer discussion of this point). There are likely also indirect benefits in terms of maintaining overall compliance, as well as potentially intrinsic value placed in upholding the rule of law (Cowell, 1990). Additionally, the social cost of audit must incorporate not only the cost to the tax authority, but also the cost to the taxpayer for which accurate figures are difficult to come by (Burgherr, 2020). We therefore do not attempt a full welfare analysis. Instead we merely note that dynamic effects increase the resources that are transfered to the state without increasing the administrative costs of audit. Assuming that a positive weight is placed on such transfers, taking into account dynamic effects increases the number of audits that should be undertaken.

Figure 3 shows that a very similar pattern holds for the impact on *total income* reported. Again there is a clear dynamic effect, peaking two years after the audited year and declining to zero by year eight, though not significantly different from zero by year five. This provides additional support to the previous result for tax, and is not purely by construction, since expenses can often be used to offset income to reduce tax (Carrillo et al., 2017; Slemrod et al., 2017).

### 4.3 Impact by income source

We repeat the previous estimation separately for different income sources, focusing on income sources for which at least five per cent of the sample report non-zero amounts.[13] This will be one way in which we discriminate between different possible explanations for why we see dynamic effects.

Figure 4 shows how the impact of an audit changes over time for the different components of income. Since the magnitudes of these incomes are different, for comparability we rescale them relative to the peak impact for that income source.

We see that, relative to the peak, self-employment income and dividends decline relatively quickly. Three years later point estimates for these are close to zero i.e. reporting is not different to the control group. In contrast, pension income exhibits little decline. Six years later it retains 80% of the impact, and this not statistically different from 100%. This pattern is suggestive of the importance of autocorrelation: income sources which one would expect to be more correlated over time appear to show weaker declines.

Table 4 shows the autocorrelation for each income source. Pension income is highly autocorrelated, since it will typically be an annuity and therefore fixed over time; property income is slightly less stable, since rents may vary more; and at the other extreme self-employment and dividend income are considerably less stable. The relative autocorrelations of income sources line up exactly with their speeds of decline.[14]

There are two caveats to these results. The first is that these measures are noisy, so if confidence intervals were added to Figure 4 for each income source, many would overlap. The second is that individuals with different income sources may have different propensities for non-compliance.

To tackle these concerns, we next use two alternative strategies. First, we compare within individuals who have multiple income sources. This immediately solves the second problem above, since our results will be within individual. It will also lead to ten pairwise comparisons: every unordered pair of the five income sources studied. For each pair, our sample is composed of individuals who had both sources sometime in the three years before audit. We then study the relative fall in reporting of each of these income sources four years after the peak. In each case we expect to find the less autocorrelated source falls fastest.

---

[13] We exclude interest income, since it is both very small and not everyone needs to report this, making it hard to compare. See Table 1 for information on the share of individuals with each income source.

[14] Note that a comparison of pensions versus property income is helpful in distinguishing this effect of autocorrelation compared with the effect of third party information. Both have a high autocorrelation, but pensions was third party reported while property income was not. In Figure 4 we see essentially the same effect for both sources, despite the large difference in third party information. Conversely, comparing property income and dividend income – which like property is also not third party reported but has a low autocorrelation – we see very different effects.

We find this result in eight out of ten cases. If there were no relationship, we should find this to be true in around five of the tests. The probability of this result under the null of no relationship is 5.5%, close to standard significance thresholds. Hence more autocorrelated income sources do seem to decline slower than less autocorrelated ones.

Our second strategy to tackle concern about heterogeneity in who receives different income sources is to reweight individuals based on individual characteristics.[15] This ensures the distribution of observed characteristics is the same across recipients of different incomes. We divide individuals into groups by sex, age band (below 40, 40-65, and above 65—the UK state pension age at which people typically retire), and quartiles of filing history. We then run weighted regressions so that the weighted samples match closely the distribution of these characteristics seen among individuals with self-employment income. We then replicate Figure 4 using the results of the reweighted regression, shown as Figure A2. The results look very similar – the only noticeable effects are that property income appears to decline slightly faster than previously, and dividend income much more quickly.

Our interpretation for this result, which we formalise below, is that audits provide the tax authority with information. Where errors are uncovered, taxpayers file amended returns. Although we do not know, and would not be allowed to reveal, precisely how audit targeting is done, it is clear that 'surprising' deviations from recorded historic reports are part of this. The amended return is therefore creating a new benchmark against which future returns will be compared. Hence, income from highly autocorrelated sources will – once uncovered – be hard to hide again, as deviations from the truth will be easily noticed. In contrast, declines in less autocorrelated income sources are less informative to the authority because they might well be real for an individual taxpayer. Viewed in aggregate, falls and rises should be equally likely, since the control group will account for any trends in the income source. Hence when we observe a decline in aggregate income reports for e.g. dividend income among audited taxpayers, this can be attributed to non-compliance, although we cannot identify which individuals are the ones under-reporting. Since declines are faster for less autocorrelated income sources, this suggests the importance of information provision. This is something we know from other settings to be important (Kleven et al., 2011; Pomeranz, 2015), although the value of audits as a potential source of information about future tax has not previously been recognised.

One caveat to this interpretation is that falls in reporting could alternatively be driven by changes in actual income. For example, those who are audited might sell shares to pay fines, reducing dividend income. Whilst this is possible, it seems unlikely. In cash terms, the peak additional

[15] We thank a referee for this suggestion.

income reported for those who have dividend income is £414. Assuming a high end estimate for the dividend yield of 10%, implies £4140 of undeclared shares. Conservatively assuming also that individuals are on the higher rate of income tax, this implies an additional £135 of tax owed. The absolute maximum penalty for misreporting is 100% of the tax due (on top of paying the tax). So selling all these shares (and hence looking like the control group) would be needed only for an individual who is found to have misreported for at least fifteen years, and receives the maximum fine. While such cases might exist, it seems extreme to assume that this is occuring at the average. Hence we think it is unlikely that the observed pattern represents changes in real behaviour, rather than reporting, though we cannot definitively rule it out.

# 5 Impacts by Audit Outcome

We next consider how dynamic effects vary depending on the outcome of audit. This is important for policy, as it helps distinguish whether merely the process of being audited is enough to impact reported income and tax. We find that those who were found to be correct do not respond, while those for whom errors were found increase reported tax. Being audited per se does not appear to increase reported tax – i.e. there is no change in behaviour among compliant taxpayers – but those found to have underpaid are 18pp more likely to report higher tax owed after audit. We first describe the approach taken to study this question, since our previous control group cannot help us study effects by audit outcome. We then describe the findings highlighted above.

## 5.1 Empirical approach

Since we now wish to study audit impacts separately by audit outcome, we cannot use the earlier identification strategy. In the "placebo audit" group we cannot observe what audit outcomes would have been, so cannot construct separate control groups for each audit outcome. Gemmell and Ratto (2012) study this question by comparing each treatment group to the original control group containing people with a mix of possible outcomes, implicitly assuming that audit outcomes are exogenously assigned. More recently Beer et al. (2019) use a matched difference-in-difference approach, allowing for observable differences in audit outcome.

We take an 'event study' approach to answer this question. Our sample for each regression is the set of observations who are audited and found to have some particular outcome, e.g. found to be compliant. Within that sample, the timing of audit is random – there is nothing systematic that led individuals to be selected in a particular year within the sample. Hence we can compare the outcome for someone audited and found to have a particular status (e.g. to be compliant) with someone who *will be* audited and found to have the same status.

For our variable of interest we now focus on a binary variable measuring whether tax paid increases, rather than on the sizes of increase, as in Pomeranz (2015). In particular, we estimate a linear probability model where the outcome is whether tax paid in year $t$ is larger than in the year before audit. Our interest now is understanding *which* individuals – when split by audit status – respond. This outcome is therefore preferred, since it compares individuals to their own history, and is equally responsive to increases for individuals across the distribution of taxes owed. It is also less sensitive to relatively extreme observations, which is more important in our event study approach because the sample size is now much smaller. Whereas previously we had a treatment group of 53,000 individuals, and could draw a large sample of controls from the non-audit population, now the entire sample is those selected for audit. That sample is then further split into subsamples by audit outcome status, making results more sensitive to outliers and reducing power. Use of a binary variable removes this sensitivity, without limiting our ability to study which groups respond.

In our specification we control for a number of key covariates: sex, age, industry, region and years filing, as well as calendar year fixed effects. Many of these individual characteristics have been shown to be predictive of non-compliance (Advani, forthcoming), so if responsiveness to audit also differs by these characteristics then without such controls we may partly pick up purely a compositional effect.

### 5.2 Results by audit outcome

To assess the reasonableness of the approach, we again begin by studying the estimated impact in the years before audit. The first four rows of Table 5 provides the results for the pre-audit period. It can be seen that all the point estimates are close to zero, providing support for the validity of this approach. A second test of validity can be seen from the 'Not audited' column. This estimates the effect of being selected for audit on individuals who were never actually audited, nor informed that they had been selected. As expected, again the point estimates are very close to zero.

Turning to the columns, three results can be seen. First, those who were audited and found to have made no errors also do not respond. This is important because it tells us that the dynamic response isn't driven by the mere fact of audit. Direct audit effects could happen, for example, if the process of audit were sufficiently unpleasant that taxpayers decided to err upwards when uncertain, in the hope of avoiding further audits. One could also potentially have seen negative direct effects in this group. If some taxpayers were incorrectly found to be compliant, they may learn that the tax authority is less effective at detecting non-compliance than they previously believed, and reduce payments. We find neither of these results: on average those whose returns are found correct do not change their reports. This is in contrast to work by Gemmell and Ratto (2012) and Beer et al. (2019).

Second, those who are found to have errors are more likely to report higher levels of tax in subsequent years. Even four years later they are 13-14pp more likely to report higher tax owed. Hence the long-term effects observed appear to all come from correcting errors made by the taxpayer. Note that even those who made errors but owed no additional tax respond to the audit. This is because the errors made might affect future tax liability. For example, claiming excessively large expenses today might increase the size of a loss on property income that can be carried forward: correcting this increases future tax liabilities. Anecdotally, from speaking to audit officers, in some cases these individuals shift their reports to pay tax in the audit year so that they can smooth out the additional tax liability that they will now face over the coming years.

Third, those who receive a penalty appear to have been driving some of the shape of the dynamics we observed earlier, where we saw a peak two years after the year selected for audit. Whilst those with mistakes but no penalty respond immediately, the response for those with a penalty peaks two years after the year for which the audit is done. This reflects two features of the audit process. Firstly, those who ultimately receive penalties typically take longest to audit, since their under-reporting requires more work to detect. The audit settlement date is thus later. If some taxpayers wait until the audit (and uncertainty about detection) is resolved to respond, this will delay the time until they are observed to respond. Second, taxpayers with mistakes but no penalties will have their *original* return corrected, so an immediate response is observed. On the other hand, those who receive a penalty may not have their return corrected: in most cases they instead file a separate form detailing additional tax, interest, and penalties.

We note that among individual characteristics, the only one which predicts responsiveness overall is sex: women are around 3pp more likely to respond to an audit. This is purely driven by compositional effects. Looking by audit outcome there are no differences in responsiveness by sex.

# 6 Simple Model of Tax Evasion and Audit Response

To help understand the mechanism underlying the observed results, we consider an extended version of the model of rational tax evasion by Allingham and Sandmo (1972), which is based on the Becker (1968) model of crime. In the Allingham and Sandmo (1972) model, individuals receive income and choose how much to report to the authority. Under-reporting has the benefit that individuals end up paying less tax, but the cost that they may be caught and receive a punishment on top of paying the correct tax. The probability of being caught is increasing in the amount of evasion. Kleven et al. (2011) extend this to allow some income to be third-party reported: under-reporting this income is detected with probability 1 so individuals will only evade out of non-third party reported income.

The key innovation of our model is to split non-third party reported income into more versus less stable sources.[16] Incomes from some sources, such as pension annuity income, are very autocorrelated ('stable'), while other sources, such as self-employment income for a sole trader, are much less stable. Autocorrelation captures the extent to which information learned in an audit today is informative about incomes tomorrow. By first extending the model of Kleven et al. (2011) to multiple time periods, and then allowing for differential autocorrelation of income sources we are able to distinguish different possible mechanisms for *why* audits are observed to have long-term effects.

Consider an individual who is audited (for the first time) in year $t$. Being audited may change his reporting for some combination of the following three reasons: (1) beliefs about the underlying audit rate or penalty for evasion ('belief updating'); (2) changes in the perceived reaudit risk following audit ('reaudit risk'); and (3) updates to the information held by the tax authority ('information').[17]

In the first of these mechanisms, there is a change in beliefs about fixed parameters, either audit rate or penalty. Consequently, any response should also be permanent, and common across all income sources. Empirically neither of these is true.

Under the second mechanism the individual perceives a temporary change in the risk of being audited. If he perceives the risk to have risen, he should be more compliant in the short term but as perceived risk returns to baseline reporting should do so also. Conversely if he perceived the risk to have fallen – the so-called 'bomb crater effect' (Mittone, 2006; Maciejovsky et al., 2007; and Kastlunger et al., 2009) – then he should be temporarily less compliant. In both cases the dynamics of this behaviour should be common across income sources. The differential responses across income sources, even within individual, are not consistent with this mechanism.

[16] Full details and formalisation is provided in Appendix B.
[17] A formalisation of the following results is provided in Appendix B.

The final mechanism is that audits provide information, which differentially changes the ability to hide certain sources of income. Performing an audit provides the tax authority with more accurate information on a taxpayer's income at a point in time. In subsequent years, information from the audit will make evasion of more stable income sources easier to detect, but for less stable income sources the effect will rapidly wear off. Hence under this mechanism, the initial impact on reporting behaviour will decline back to baseline, and this decline will be more rapid for income sources which have a lower autocorrelation. This is consistent with our findings, as seen in Figure 4.

## 7 Conclusion

This paper investigated the dynamic effects of audits on income reported in subsequent tax returns. Understanding these effects is important both from the perspective of quantifying the returns to the tax authority from an audit, and for assessing the mechanisms by which audits might influence taxpayer behaviour. To answer this question we exploited a random audit program run by the UK tax authority (HMRC) under which an average of around 4,900 individuals are randomly selected for audit each year. We used data on audits over the period 1998/99 to 2008/09 and tracked responses on tax returns between 1998/99 and 2011/12.

We established three main results. First, we provided evidence of important *dynamic effects*, with the additional tax revenue over the five years post-audit equalling 1.5 times the direct revenue raised by audit. Second, we documented that a return to misreporting occurred more rapidly after audit for income sources which were less autocorrelated. Third, we showed that only those who were found to have made mistakes responded to the audit. Extending the standard model of rational tax evasion, we demonstrated that the observed dynamics are consistent only with audits revealing information to the tax authority, which makes misreporting certain income sources easier to detect for a period after the audit.

Our results have three main policy implications. First, taking dynamic effects into account substantially increases the estimated revenue impact of audits. The direct effect of an audit is (on average) £830, whilst the cumulative dynamic effect over the subsequent five years is £1230, 1.5 times the direct effect. This suggests that the optimal audit rate should be substantially increased relative to the situation where there are no dynamic effects. A back-of-the envelope calculation suggests the cost of an audit to the tax authority is around £2500, so that even random audits are close to breaking even. For targeted audits, including dynamic effects raises the average return from around £6000 to £15000.

Second, the variation in dynamic effects observed across different income components alters the way in which targeted audits should be targeted: audits should focus more on individuals reporting types of income with the largest overall effects, combining immediate and dynamic effects. For example, the peak annual impact on reported self-employment income for each self-employed individual is over £1000, higher than other components. This suggests focusing more on individuals reporting self-employment income. Likewise, although the maximum annual impact on pension income is lower, it is persistent, so there may be more incentive to target individuals believed to be under-reporting pension income. The precise design of any targeting strategy must of course take into account how taxpayers would respond to the strategy, but for the tax authority the first step in designing any targeting strategy must be to know where the revenue is.

Third, there are implications for setting optimal re-auditing strategies. Impacts for reported self-employment income and dividend income die away after about four years, so it might make sense to revisit these individuals around this time. In contrast, the impact on reported pension income seems to persist for at least eight years, implying that there is less need to re-audit these individuals so soon. Again the responses of taxpayers to changes in audit strategy must be considered.

Our findings also highlight the importance of further study of the indirect effect of tax compliance audits. One natural direction for further work would be to understand how the dynamic effects vary in the context of targeted audits, which are focused on individuals deemed likely to be non-compliant. A second avenue for exploration is the spillover effect of audits: does auditing taxpayers change the behaviour of other taxpayers with whom they interact (Boning et al., 2018). A third question is the extent to which cheaper 'threat letters' can be used to maintain consistently high levels of compliance over the long term, in the absence of high audit probabilities. Better understanding these effects is crucial in determining optimal audit policy.

Finally, our results speak to the wider use of audits for public policy, be that to reduce corruption, improve public service delivery, or ensure environmental standards are met. A key lesson is that audits change future behaviour, but *how* that behaviour changes depends on the likelihood of being caught in the future. Unless there are ongoing incentives to improve compliance – such as increased audit risk, increased penalties, or easier verification of misreporting – changes in reporting may be short-lived. However, a key trade-off in public policy contexts is that individuals may be able to substitute entirely out of audited activities, if the strictness of enforcement is too high. This limits the compliance improvements achieved (Tulli, 2019) and may have additional welfare costs as some valuable activities become more expensive (Gerardino et al., 2020) or do not take place (Lichand et al., 2019).

# Bibliography


ADVANI, A. (forthcoming): "Who does and doesn't pay taxes?" *Fiscal Studies.*

ADVANI, A. AND A. SUMMERS (2020): "Capital Gains and UK Inequality," CAGE Working Paper Series, 465.

ALLINGHAM, M. AND A. SANDMO (1972): "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1, 323–338.

ALM, J. (2019): "What Motivates Tax Compliance?" *Journal of Economic Surveys*, 33, 353–388.

ALMUNIA, M. AND D. LOPEZ-RODRIGUEZ (2018): "Under the Radar: The Effects of Monitoring Firms on Tax Compliance," *American Economic Journal: Economic Policy*, 10, 1–38.

ASATRYAN, Z. AND A. PEICHL (2017): "Responses of firms to tax, administrative and accounting rules: Evidence from Armenia," CESifo Working Paper Series, 6754.

AVIS, E., C. FERRAZ, AND F. FINAN (2018): "Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians," *Journal of Political Economy*, 126, 1912–64.

BECKER, G. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 169, 176–177.

BEER, S., M. KASPER, E. KIRCHLER, AND B. ERARD (2019): "Do Audits Deter or Provoke Future Tax Noncompliance? Evidence on Self-employed Taxpayers," International Monetary Fund Working Paper, 19/223.

BERGOLO, M., R. CENI, G. CRUCES, M. GIACCOBASSO, AND R. PEREZ-TRUGLIA (2020): "Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment," NBER Working Paper Series, 23631.

BLOOMQUIST, K. (2013): "Incorporating Indirect Effects in Audit Case Selection: An Agent-Based Approach," *IRS Research Bulletin*, 103–116.

BLUMENTHAL, M., C. CHRISTIAN, AND J. SLEMROD (2001): "Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota," *National Tax Journal*, 54, 125–38.

BOBONIS, G. J., L. R. CÁMARA FUERTES, AND R. SCHWABE (2016): "Monitoring corruptible politicians," *American Economic Review*, 106, 2371–2405.

BONING, W. C., J. GUYTON, R. H. HODGE, J. SLEMROD, AND U. TROIANO (2018): "Heard it through the grapevine: direct and network effects of a tax enforcement field experiment," NBER Working Paper Series, 24799.

BURGHERR, D. (2020): "The costs of administering a wealth tax," Wealth and Policy Working Paper, 126.

CARRILLO, P., D. POMERANZ, AND M. SINGHAL (2017): "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," *American Economic Journal: Applied Economics*, 9, 144–64.

CHOO, L., M. FONSECA, AND G. MYLES (2013): "Lab Experiment to Investigate Tax Compliance: Audit Strategies and Messaging," Research report 308, HM Revenue and Customs.

COWELL, F. A. (1990): *Cheating the government: The economics of evasion*, MIT Press.

DeBacker, J., B. Heim, A. Tran, and A. Yuskavage (2018): "Once Bitten, Twice Shy? The Lasting Impact of IRS Audits on Individual Tax Reporting," *Journal of Law and Economics*, 61, 1–35.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2013): "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India," *The Quarterly Journal of Economics*, 128, 1499–1545.

——— (2018): "The value of regulatory discretion: Estimates from environmental inspections in India," *Econometrica*, 86, 2123–2160.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016): "Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany," *American Economic Journal: Economic Policy*, 8, 203–32.

Erard, B. (1992): "The Influence of Tax Audits on Reporting Behaviour," in *Why People Pay Taxes: Tax Compliance and Enforcement*, ed. by J. Slemrod, University of Michigan Press, chap. 5, 95–114.

Fellner, G., R. Sausgruber, and C. Traxler (2013): "Testing enforcement strategies in the field: Threat, moral appeal and social information," *Journal of the European Economic Association*, 11, 634–660.

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

Gerardino, M. P., S. Litschig, and D. Pomeranz (2020): "Distortion by Audit: Evidence from Public Procurement," NBER Working Paper Series, 23978.

Kastlunger, B., E. Kirchler, L. Mittone, and J. Pitters (2009): "Sequences of Audits, Tax Compliance, and Taxpaying Strategies," *Journal of Economic Psychology*, 30, 405–418.

Kleven, H. J., M. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011): "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark," *Econometrica*, 79, 651–692.

Kleven, H. J., C. T. Kreiner, and E. Saez (2016): "Why can modern governments tax so much? An agency model of firms as fiscal intermediaries," *Economica*, 83, 219–246.

Kolm, S.-C. (1973): "A note on optimum tax evasion," *Journal of Public Economics*, 2, 265–270.

Lichand, G., M. F. M. Lopes, and M. C. Medeiros (2019): "Why Do Audit Threats Work? Information vs. Salience at the Announcement of the Brazilian Anti-Corruption Program," mimeo.

Long, S. and R. Schwartz (1987): "The Impact of IRS Audits on Taxpayer Compliance: A Field Experiment in Specific Deterrence," Paper presented at the Annual Meeting of the Law and Society Association, Washington DC.

Løyland, K., O. Raaum, G. Torsvik, and A. Øvrum (2019): "Compliance effects of risk-based tax audits," 7616, CESifo Working Paper 7616.

Maciejovsky, B., E. Kirchler, and H. Schwarzenberger (2007): "Misperception of Chance and Loss Repair: On the Dynamics of Tax Compliance," *Journal of Economic Psychology*, 28, 678–691.

Mascagni, G. (2018): "From the lab to the field: A review of tax experiments," *Journal of Economic Surveys*, 32, 273–301.

Mittone, L. (2006): "Dynamic Behaviour in Tax Evasion: An Experimental Approach," *Journal of Socio-Economics*, 35, 813–835.

Naritomi, J. (2019): "Consumers as tax auditors," *American Economic Review*, 109, 3031–72.

Perez-Truglia, R. and U. Troiano (2018): "Shaming tax delinquents," *Journal of Public Economics*, 167, 120–137.

Pomeranz, D. (2015): "No taxation without information: Deterrence and self-enforcement in the value added tax," *American Economic Review*, 105, 2539–69.

Pomeranz, D. and J. Vila-Belda (2019): "Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities," *Annual Review of Economics*, 11, 755–781.

Rosenbaum, P. and D. Rubin (1985): "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score," *American Statistician*, 39, 33–38.

Rubin, D. (2001): "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation," *Health Services and Outcomes Research Methodology*, 2, 169–188.

Sandmo, A. (2005): "The Theory of Tax Evasion: A Retrospective View," *National Tax Journal*, 58, 643–663.

Slemrod, J. (2019): "Tax compliance and enforcement," *Journal of Economic Literature*, 57, 904–54.

Slemrod, J., M. Blumenthal, and C. Christian (2001): "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota," *Journal of Public Economics*, 79, 455–483.

Slemrod, J., B. Collins, J. L. Hoopes, D. Reck, and M. Sebastiani (2017): "Does credit-card information reporting improve small-business tax compliance?" *Journal of Public Economics*, 149, 1–19.

Slemrod, J. and S. Yitzhaki (1987): "The Optimal Size of a Tax Collection Agency," *The Scandinavian Journal of Economics*, 89, 183–192.

——— (2002): "Tax Avoidance, Evasion, and Administration," in *Handbook of Public Economics III*, ed. by A. J. Auerbach and M. Feldstein, Amsterdam: Elsevier.

Tauchen, H., A. Witte, and K. Beron (1993): "Tax Compliance: An Investigation Using Individual Taxpayer Compliance Measurement Program (TCMP) data," *Journal of Quantitative Criminology*, 9, 177–202.

Tulli, A. (2019): "Sweeping the Dirt Under the Rug: Measuring Spillovers from an Anti-Corruption Measure," *Job Market Paper*.

Yitzhaki, S. (1987): "On the Excess Burden of Tax Evasion," *Public Finance Quarterly*, 15, 123–137.

Zamboni, Y. and S. Litschig (2018): "Audit risk and rent extraction: Evidence from a randomized evaluation in Brazil," *Journal of Development Economics*, 134, 133–149.

# Tables and Figures

**Table 1: Share of taxpayers with each source of income**

| Income component | Proportion |
|---|---|
| Interest | .587 |
| Employment | .482 |
| Self employment | .375 |
| Dividends | .370 |
| Pensions | .300 |
| Property | .136 |
| Foreign | .048 |
| Trusts and estates | .010 |
| Share schemes | .002 |
| Other | .030 |

***Notes:*** *Annual averages for tax years 1998/99 to 2008/09. Includes observations in year selected for audit or placebo audit.*
***Source:*** *Authors' calculations based on HMRC administrative datasets.*

## Table 2: Random audit outcomes

| | Mean | Std. dev. |
|---|---|---|
| Proportion of audited returns deemed | | |
| Correct | .532 | .499 |
| Incorrect but no underpayment | .111 | .314 |
| Incorrect with underpayment (non-compliant) | .357 | .479 |
| Mean additional tax if non-compliant (£) | 2,314 | 7,758 |
| Distribution of additional tax if non-compliant | | |
| Share £1-100 | .116 | .320 |
| Share £101-1,000 | .483 | .500 |
| Share £1,001-10,000 | .361 | .480 |
| Share £10,001+ | .039 | .194 |
| Observations | 34,630 | |

***Notes:*** *Annual averages for tax years 1998/99 to 2008/09. Includes all individuals with a completed random audit.*
***Source:*** *Authors' calculations based on HMRC administrative datasets.*

## Table 3: Sample balance, conditioning on filing history

| Years after audit | | -5 | -4 | -3 | -2 | -1 |
|---|---|---|---|---|---|---|
| | | | Characteristics | | | |
| Female | Mean | .274 | .276 | .278 | .282 | .287 |
| | Difference | -.005 | -.006 | -.005 | -.006 | -.005 |
| | p-value | .236 | .212 | .292 | .234 | .338 |
| Age | Mean | 49.2 | 49.3 | 49.3 | 49.4 | 49.5 |
| | Difference | .0 | .0 | .1 | .1 | .1 |
| | p-value | .472 | .600 | .188 | .170 | .110 |
| In London or SE | Mean | .333 | .334 | .335 | .333 | .331 |
| | Difference | -.003 | .001* | .003 | .002 | .002 |
| | p-value | .159 | .026 | .015 | .317 | .190 |
| Has tax agent | Mean | .628 | .614 | .603 | .589 | .573 |
| | Difference | -.003 | -.001 | -.001 | .002 | .002 |
| | p-value | .522 | .500 | .376 | .675 | .606 |
| | | | Income and tax totals | | | |
| Total taxable income | Mean | 35,075 | 34,670 | 34,030 | 32,912 | 31,755 |
| | Difference | -2 | 35 | -163 | 71 | 56 |
| | p-value | .979 | .469 | .012 | .280 | .439 |
| Total tax | Mean | 9,646 | 9,539 | 9,321 | 8,979 | 8,635 |
| | Difference | 14 | 12 | -40 | 12 | 15 |
| | p-value | .982 | .288 | .061 | .261 | .887 |
| | | | Income components | | | |
| Employment | Mean | 22,508 | 22,534 | 22,266 | 21,708 | 21,145 |
| | Difference | 11 | -57 | -98 | 112* | 43* |
| | p-value | .758 | .023 | .152 | .049 | .05 |
| Self employment | Mean | 6,546 | 6,379 | 6,200 | 5,950 | 5,581 |
| | Difference | 56 | 38 | -49 | -18 | 29 |
| | p-value | .298 | .435 | .033 | .161 | .684 |
| Interest and dividends | Mean | 4,007 | 3,905 | 3,895 | 3,759 | 3,645 |
| | Difference | -26 | 16 | -27 | 7 | 4 |
| | p-value | .667 | .189 | .235 | .958 | .086 |
| Pensions | Mean | 3,493 | 3,542 | 3,561 | 3,562 | 3,531 |
| | Difference | -23 | -23 | -3 | 4 | 22 |
| | p-value | .806 | .482 | .681 | .463 | .523 |
| Property | Mean | 869 | 844 | 811 | 769 | 726 |
| | Difference | -5 | -6 | 0 | 6 | 0 |
| | p-value | .282 | .209 | .525 | .072 | .518 |
| Foreign | Mean | 194 | 193 | 194 | 181 | 169 |
| | Difference | 1 | -1 | 5 | -4 | 0 |
| | p-value | .627 | .137 | .526 | .117 | .925 |
| Trusts and estates | Mean | 150 | 145 | 145 | 131 | 123 |
| | Difference | 17 | -1 | -6 | 4 | 3 |
| | p-value | .153 | .963 | .204 | .686 | .824 |
| Share schemes | Mean | 91 | 104 | 68 | 62 | 55 |
| | Difference | 8* | -8 | -1 | 9 | 1 |
| | p-value | .019 | .711 | .783 | .683 | .243 |
| Other | Mean | 80 | 75 | 76 | 73 | 71 |
| | Difference | 3 | -3 | 0 | 1 | 2 |
| | p-value | .618 | .184 | .228 | .984 | .194 |

***Notes:*** *'Years after audit' measures time relative to audit, or placebo audit for controls. 'Mean' is the mean outcome in the control (not selected for audit) group across all years. 'Difference' is the coefficient on the treatment dummy in a regression of the outcome on a treatment dummy and dummies for whether the taxpayer filed taxes in each of the four years before audit (or placebo audit for controls). Treatment dummy equals 1 if taxpayer was selected by HMRC for a random audit. p-values are derived from an F-test that coefficients on interactions between treatment and tax year dummies are all zero in a regression of the outcome of interest on tax year dummies, interactions between treatment and tax year dummies, and dummies for whether the taxpayer filed taxes in each of the four years before audit (or placebo audit for controls). This is a stronger test than just testing the coefficient on treatment not interacted. Monetary values are in 2012 prices. Standard errors are clustered by taxpayer. * $p < .05$, ** $p < .01$, *** $p < .001$.*

***Source:*** *Authors' calculations based on HMRC administrative datasets.*

## Table 4: Autocorrelation by income source

| | Corr(t,t-1) | Corr(t,t-2) | Corr(t,t-3) |
|---|---|---|---|
| Pension income | .946 | .904 | .864 |
| Property income | .896 | .836 | .790 |
| Employment income | .862 | .769 | .690 |
| Interest income | .835 | .722 | .640 |
| Self-employment income | .832 | .728 | .644 |
| Dividend income | .813 | .723 | .657 |
| Sample size | 4,506,548 | 4,506,548 | 4,506,548 |

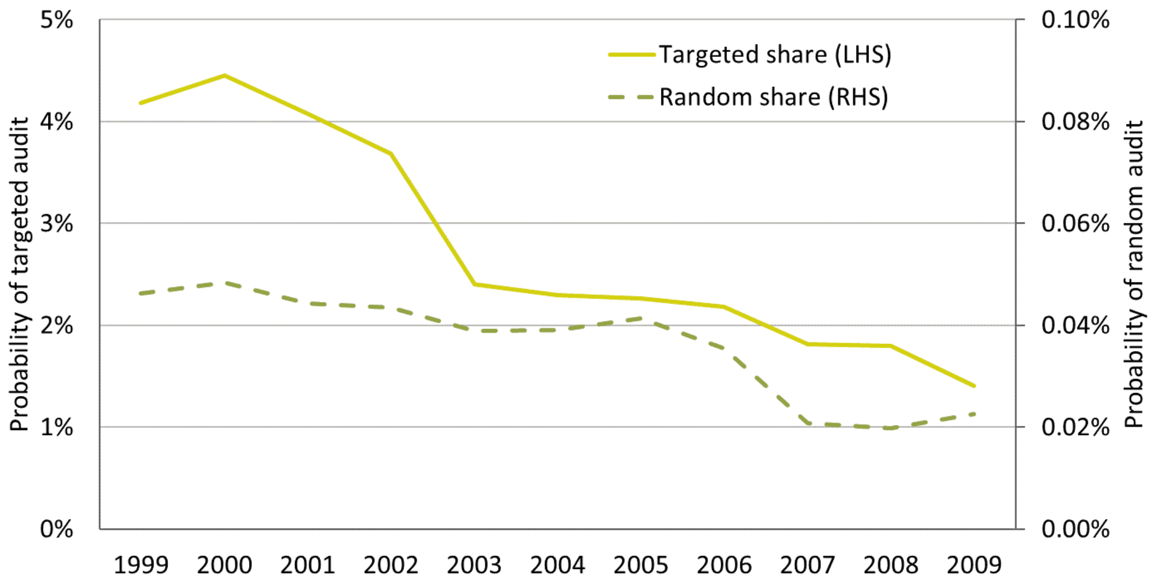
***Notes:*** *Annual averages for years 1998/99 to 2011/12.*
***Source:*** *Calculations based on HMRC administrative datasets.*

## Table 5: Impact by audit outcome

| Years since audit | Overall | Correct | Mistake nonpos | Mistake pos | Pos. yield + penalty | Not audited |
|---|---|---|---|---|---|---|
| -5 | -.006 | -.042* | .048 | .033 | -.014 | -.002 |
| | (.013) | (.018) | (.049) | (.032) | (.072) | (.030) |
| -4 | .007 | -.034 | .068 | .050 | .037 | -.006 |
| | (.014) | (.019) | (.049) | (.033) | (.068) | (.030) |
| -3 | .005 | -.023 | .058 | .039 | .042 | -.016 |
| | (.014) | (.019) | (.050) | (.033) | (.068) | (.030) |
| -2 | .022 | -.005 | .079 | .075* | .032 | -.008 |
| | (.014) | (.019) | (.050) | (.033) | (.068) | (.030) |
| -1 | *Outcome is difference from -1 so zero by construction* | | | | | |
| 0 | .056*** | .016 | .131* | .179*** | .092 | -.014 |
| | (.014) | (.019) | (.051) | (.033) | (.069) | (.030) |
| 1 | .048*** | .012 | .109* | .174*** | .180** | -.037 |
| | (.014) | (.019) | (.051) | (.033) | (.069) | (.030) |
| 2 | .042** | .007 | .135** | .152*** | .207** | -.052 |
| | (.013) | (.020) | (.051) | (.033) | (.069) | (.031) |
| 3 | .030* | -.007 | .135** | .133*** | .171* | -.048 |
| | (.014) | (.020) | (.052) | (.034) | (.069) | (.031) |
| 4 | .031* | -.0024 | .134* | .137*** | .143* | -.045 |
| | (.014) | (.021) | (.052) | (.034) | (.070) | (.031) |
| 5 | .033* | .019 | .160** | .119** | .128 | -.052 |
| | (.016) | (.023) | (.056) | (.037) | (.074) | (.034) |
| N | 124,223 | 46,911 | 9,519 | 25,666 | 6,983 | 35,144 |

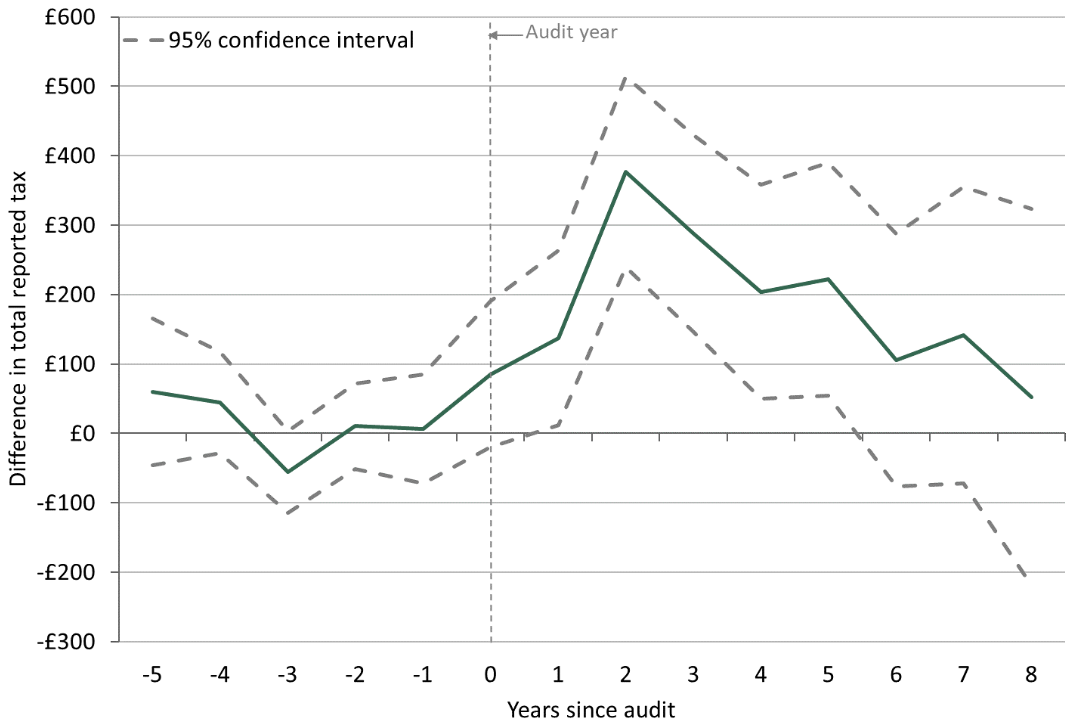
***Notes:*** *The outcome variable is a dummy for whether tax paid is higher in each of the years before/after audit than the year immediately before audit ('-1'). 'Overall' uses the full sample of audited individuals to perform an event study for whether tax paid is higher than in the year before audit. Coefficients from a linear probability model are shown, with standard errors in parentheses. Other columns split the audited sample by audit outcome: tax return found to be correct; tax return found to have a mistake but which doesn't change tax liability (or in a small number of cases reduced liability); tax return found to have a mistake leading to increased tax liability, but no penalty charged (i.e. treated as legitimate error); tax return found to have under-reported liability and a penalty charged (i.e. deemed to be deliberate); tax return selected for audit but no audit actually implemented (placebo test).* * $p < .05$, ** $p < .01$, *** $p < .001$.
***Source:*** *Authors' calculations based on HMRC administrative datasets.*

**Figure 1: Change in the probability of audit over time**



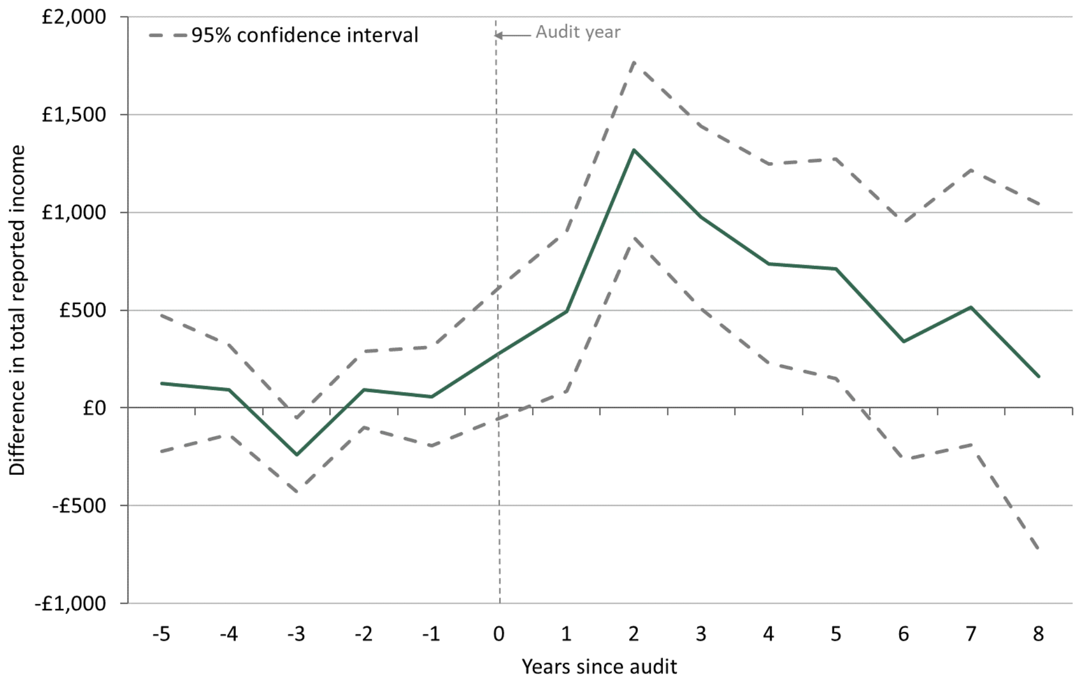

***Notes:*** *Constructed using data on individuals who received an audit of their self assessment tax return for a tax year between 1998/1999 and 2008/2009, and the full sample of self assessment returns for the same period.*
***Source:*** *Calculations based on HMRC administrative datasets.*

**Figure 2: Dynamic effect of audits on total reported tax owed**



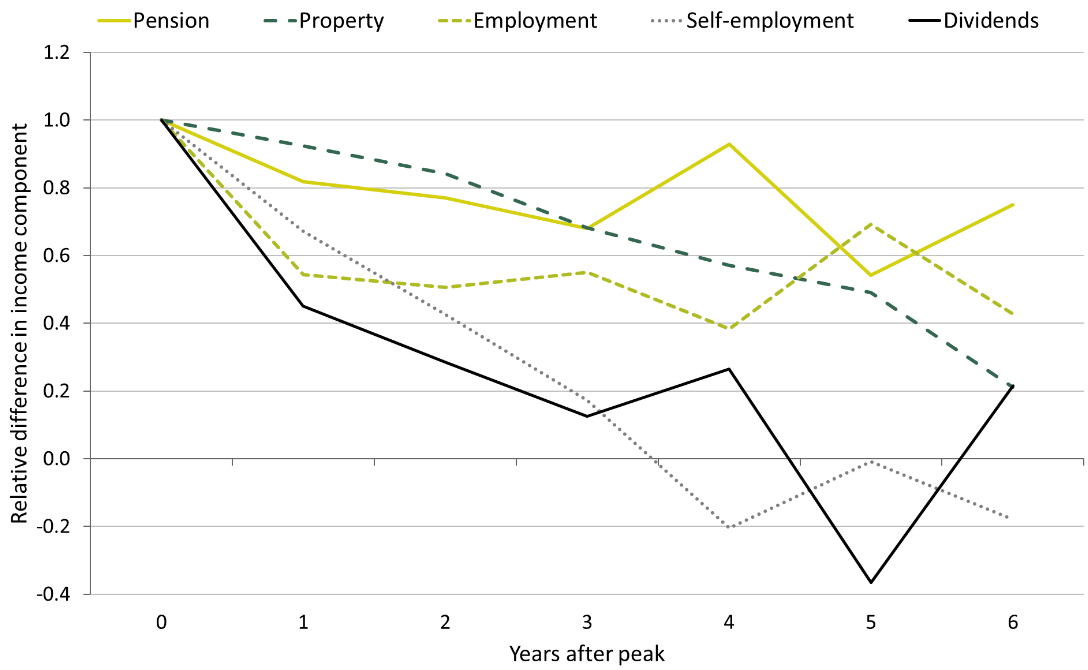

***Notes:*** *Sample includes individuals selected for a random audit between 1998/99 and 2008/09, and control individuals who could have been selected in the same years but were not. It uses tax returns from 1998/99 to 2011/12. The solid line plots the point estimate for the difference in average 'total reported tax' between individuals who were and weren't audited, for different numbers of years after the audit. This comes from a regression of total reported tax on dummies for years since audit (or placebo audit for controls), dummies for years since audit (or placebo audit for controls) interacted with treatment status, tax year dummies, and dummies for whether the taxpayer filed a return in each of the four years before audit, with audit status instrumented by selection for audit. Standard errors are clustered at the individual level.*
***Source:*** *Calculations based on HMRC administrative datasets.*

**Figure 3: Dynamic effect of audits on total reported income**




***Notes:*** *Sample includes individuals selected for a random audit between 1998/99 and 2008/09, and control individuals who could have been selected in the same years but were not. It uses tax returns from 1998/99 to 2011/12. The solid line plots the point estimate for the difference in average 'total reported income' (income from all sources) between individuals who were and weren't audited, for different numbers of years after the audit. This comes from a regression of total reported income on dummies for years since audit (or placebo audit for controls), dummies for years since audit (or placebo audit for controls) interacted with treatment status, tax year dummies, and dummies for whether the taxpayer filed a return in each of the four years before audit, with audit status instrumented by selection for audit. Standard errors are clustered at the individual level.*
***Source:*** *Calculations based on HMRC administrative datasets.*

**Figure 4: Relative dynamics by income source: less autocorrelated sources of income see faster declines**




***Notes:*** *Sample includes individuals selected for a random audit between 1998/99 and 2008/09, and control individuals who could have been selected in the same years but were not. It uses tax returns from 1998/99 to 2011/12. Each line plots the point estimate for the difference in the average of a particular component of income between individuals who were and weren't audited, for different numbers of years after the peak impact for that income source. This comes from a regression of each income component on dummies for years since audit (or placebo audit for controls), dummies for years since audit (or placebo audit for controls) interacted with treatment status, tax year dummies, and dummies for whether the taxpayer filed a return in each of the four years before audit, with audit status instrumented by selection for audit.*
***Source:*** *Calculations based on HMRC administrative datasets.*

# Appendices

## Appendix A Additional Tables and Figures

**Table A1: Random audit lags and durations**

| | Mean | Std. dev. | Median | 75th | 90th |
|---|---|---|---|---|---|
| Lag to audit start (months) | 8.9 | 4.0 | 9 | 11 | 14 |
| Audit duration (months) | 5.3 | 6.6 | 3 | 7 | 13 |
| Total time to audit end (months) | 14.3 | 7.3 | 13 | 17 | 23 |

***Notes:*** *Annual averages for tax years 1998/99 to 2008/09. Includes all individuals with a completed random audit.*
***Source:*** *Authors' calculations based on HMRC administrative datasets.*

**Table A2: Third party reporting arrangements in the UK**

| Income component | Degree of third-party reporting |
|---|---|
| Employment | Complete |
| Interest | Complete |
| Pensions | Partial, via pension provider |
| Self employment | None unless an entertainer, sportsman, or contractor in the construction industry |
| Dividends | None |
| Property | None |

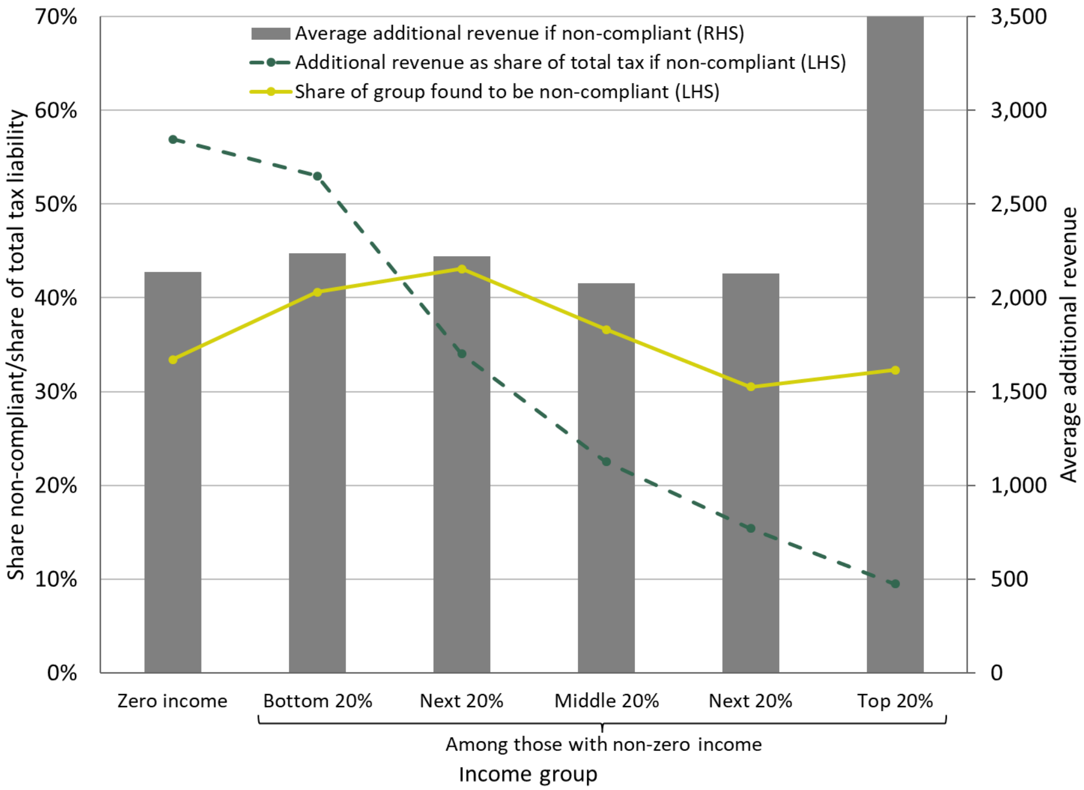
***Source:*** *Personal communication with Tracey Bowler, Tax Law Review Committee.*

## Table A3: Sample balance (unconditional)

| Years after audit | | -5 | -4 | -3 | -2 | -1 |
|---|---|---|---|---|---|---|
| | | Characteristics | | | | |
| Female | Mean | .274 | .276 | .278 | .282 | .287 |
| | Difference | -.006 | -.004 | -.002 | -.001 | -.002 |
| | p-value | .221 | .359 | .606 | .627 | .863 |
| Age | Mean | 49.2 | 49.3 | 49.3 | 49.4 | 49.5 |
| | Difference | .2 | .3 | .3 | .2 | .2 |
| | p-value | .756 | .586 | .390 | .610 | .057 |
| In London or SE | Mean | .333 | .334 | .335 | .333 | .331 |
| | Difference | -.006 | .001* | .004* | .002 | .002 |
| | p-value | .177 | .025 | .011 | .281 | .152 |
| Has tax agent | Mean | .628 | .614 | .603 | .589 | .573 |
| | Difference | .000 | .002 | .001 | .003 | .005 |
| | p-value | .547 | .508 | .405 | .396 | .412 |
| Survives | Mean | .624 | .669 | .728 | .803 | .892 |
| | Difference | .032*** | .039*** | .047*** | .050*** | .050*** |
| | p-value | .000 | .000 | .000 | .000 | .000 |
| | | Income and tax totals | | | | |
| Total taxable income | Mean | 35,075 | 34,670 | 34,030 | 32,912 | 31,755 |
| | Difference | 881 | 492 | 403* | 1,051* | 1,095* |
| | p-value | .374 | .157 | .028 | .012 | .012 |
| Total tax | Mean | 9,646 | 9,539 | 9,321 | 8,979 | 8,635 |
| | Difference | 260 | 63 | 82 | 310 | 337* |
| | p-value | .539 | .303 | .055 | .064 | .027 |
| | | Income components | | | | |
| Employment | Mean | 22,508 | 22,534 | 22,266 | 21,708 | 21,145 |
| | Difference | -31 | -136 | 180* | 909** | 721* |
| | p-value | .162 | .371 | .028 | .006 | .027 |
| Self employment | Mean | 6,546 | 6,379 | 6,200 | 5,950 | 5,581 |
| | Difference | 356 | 328 | 173 | 99 | 200* |
| | p-value | .151 | .174 | .311 | .106 | .025 |
| Interest and dividends | Mean | 4,007 | 3,905 | 3,895 | 3,759 | 3,645 |
| | Difference | -36 | 208 | 18 | 63 | 112 |
| | p-value | .767 | .432 | .700 | .578 | .580 |
| Pensions | Mean | 3,493 | 3,542 | 3,561 | 3,562 | 3,531 |
| | Difference | 176 | 168 | 128 | 148 | 159 |
| | p-value | .425 | .478 | .642 | .307 | .327 |
| Property | Mean | 869 | 844 | 811 | 769 | 726 |
| | Difference | 18 | -2 | 37 | 47 | 31 |
| | p-value | .813 | .952 | .576 | .498 | .134 |
| Foreign | Mean | 194 | 193 | 194 | 181 | 169 |
| | Difference | 23 | -1 | -5 | -6 | 1 |
| | p-value | .759 | .240 | .956 | .317 | .766 |
| Trusts and estates | Mean | 150 | 145 | 145 | 131 | 123 |
| | Difference | 46 | 17 | 8 | 19 | 8 |
| | p-value | .245 | .367 | .290 | .125 | .367 |
| Share schemes | Mean | 91 | 104 | 68 | 62 | 55 |
| | Difference | 17* | 24 | 22** | 11 | -6** |
| | p-value | .043 | .062 | .004 | .132 | .008 |
| Other | Mean | 80 | 75 | 76 | 73 | 71 |
| | Difference | -1 | -4 | 0 | 5 | 6 |
| | p-value | .747 | .675 | .645 | .796 | .167 |

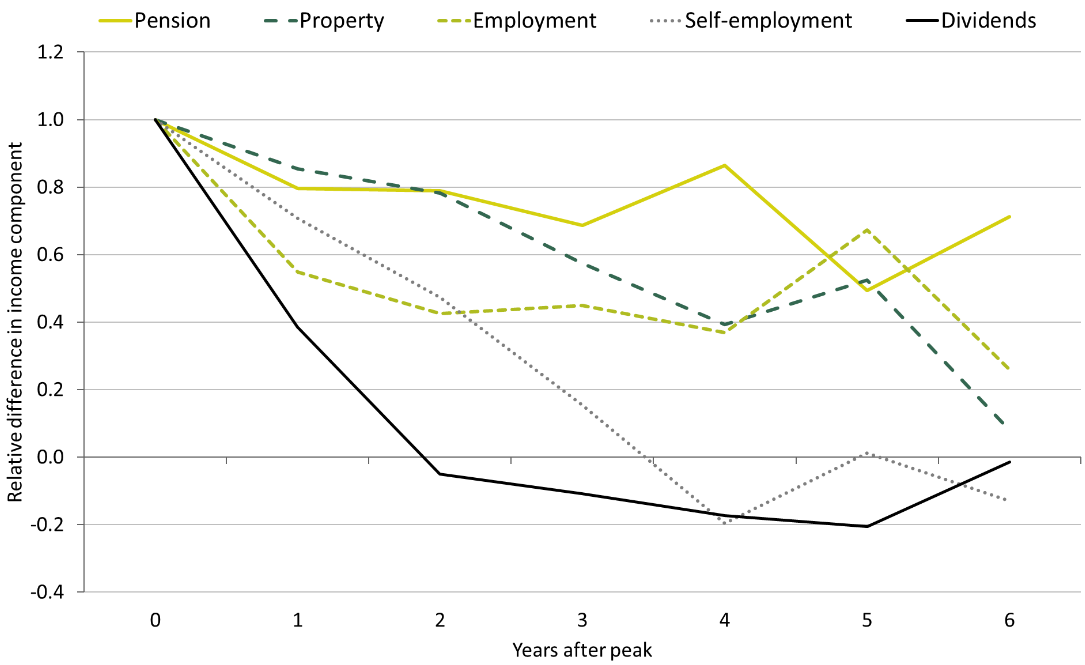
***Notes:*** *'Years after audit' measures time relative to audit, or placebo audit for controls. 'Mean' is the mean outcome in the control (not selected for audit) group across all years. 'Difference' is the coefficient on the treatment dummy in a regression of the outcome on a treatment dummy. Treatment dummy equals 1 if taxpayer was selected by HMRC for a random audit. p-values are derived from an F-test that coefficients on interactions between treatment and tax year dummies are all zero in a regression of the outcome of interest on tax year dummies and interactions between treatment and tax year dummies. This is a stronger test than just testing the coefficient on treatment not interacted. Tests for all outcomes other than 'survives' are conditional on survives = 1. Monetary values are in 2012 prices. Standard errors are clustered by taxpayer.* * $p < .05$, ** $p < .01$, *** $p < .001$.
***Source:*** *Authors' calculations based on HMRC administrative datasets.*

**Figure A1: Non-compliance over the prior year reported income distribution**




***Notes:*** *Constructed using data on individuals who received a random audit of their self assessment tax return for a tax year between 1998/1999 and 2008/2009. Income grouping is done based on previous year's reported income. 16.2% of individuals report having zero income in the previous year. The remaining individuals are divided into five equal sized bins based on their previous income: quintiles conditional on reporting non-zero income. 'Share of group found to be non-compliant' is the share of individual taxpayers who are found to owe additional tax when audited. 'Average additional revenue if non-compliant' is the average total tax in 2012 £ that was not reported among those individuals for whom some tax was not reported (the non-compliant). 'Additional revenue as a share of total tax if non-compliant' is the additional tax owed divided by total tax owed, averaged across individual taxpayers who were non-compliant.*
***Source:*** *Advani (forthcoming).*

## Figure A2: Relative dynamics by income source, after reweighting by characteristics




***Notes:*** *Sample includes individuals selected for a random audit between 1998/99 and 2008/09, and control individuals who could have been selected in the same years but were not. It uses tax returns from 1998/99 to 2011/12. Each line plots the point estimate for the difference in the average of a particular component of income between individuals who were and weren't audited, for different numbers of years after the peak impact for that income source, after reweighting individuals so that the distribution of observed characteristics matches that seen among the self employed. This comes from dividing individuals into groups by sex, age band, and quartile of filing history. Observations are reweighted so that the distribution across these discrete cells is the same as for the self-employed. Point estimates for the treatment effect come from a weighted regression of each income component on dummies for years since audit (or placebo audit for controls), dummies for years since audit (or placebo audit for controls) interacted with treatment status, tax year dummies, and dummies for whether the taxpayer filed a return in each of the four years before audit, with audit status instrumented by selection for audit.*

***Source:*** *Calculations based on HMRC administrative datasets.*

# Appendix B Model details

In this Appendix we formalise the model described in Section 6.

## B.1 Model outline

We consider an extended version of the model by Kleven et al. (2011), which is itself an extension of the model by Allingham and Sandmo (1972). Taxpayers are risk-neutral, and choose how much tax to evade.[18] The probability of detection is endogenous and is increasing in the difference between reported and true income. The key idea underlying this assumption is that other contextual information the tax authority might have, such as where a taxpayer lives, provides some information on the taxpayer's income, so that reports further from the truth are more likely to be investigated (Allingham and Sandmo, 1972; Yitzhaki, 1987; Slemrod and Yitzhaki, 2002; Sandmo, 2005; Kleven et al., 2011). This is consistent with the objective of the tax authority: since not everyone can be audited, larger deviations from the authority's expectation are likely to yield the most revenue.

As noted in Section 6, the key innovation of our model is to note that incomes from some sources are very autocorrelated ('stable'), while other sources are much less stable. Extending the model of Kleven et al. (2011) to multiple time periods, and incorporating differential autocorrelation of income sources, allows us to distinguish different possible mechanisms for dynamic responses. The importance of autocorrelation is that an audit provides a snapshot measure of the current true level of income. For income sources which are relatively stable, this will also be a good predictor of revenue in the future. For those which are less autocorrelated, the predictive power of this snapshot will decline quickly over time. As we discuss below, different mechanisms make distinct predictions about the dynamics of any audit impact *split by autocorrelation of income source*.

Consider a taxpayer with true income $\tilde{y}_s$ in year $s$. This income can be decomposed into three parts: (i) a third party reported component, $\tilde{y}_s^{\mathrm{TPR}}$; (ii) a self-reported permanent component, $\tilde{y}^{\mathrm{perm}}$; and (iii) a self-reported stochastic component, $\tilde{y}_s^{\mathrm{stoch}}$. The distinction between third-party reported and self-reported income can explain why evasion rates appear much lower than would be expected given the empirical probability of audit (Kleven et al., 2011). We distinguish within self-reported income sources those which are fixed over time, and those which are time-varying. This is a sim-

[18] Relaxing risk-neutrality would reduce evasion. It could also introduce a positive correlation between the level of income and evasion, assuming no outside wealth and decreasing absolute risk aversion. This is inconsistent with what we see empirically in our data. Instead we see that there is little variation across the income distribution in the probability of evasion (extensive margin) or the amount of tax evaded in cash terms (intensive margin). The latter result means the *share* of income evaded is falling across the income distribution. See Figure A1 for details.

plification for expositional purposes, but our main results can be generalised to having multiple self-reported income sources with varying degrees of autocorrelation, as in our empirical setting.

In year $s$ the taxpayer reports an income of $y_s$, so evasion is $e_s := \tilde{y}_s - y_s$. Evasion is detected with probability $p(e_s)$, which is increasing in the level of evasion i.e. $p'(e_s) > 0$. This is a composite of both the probability of audit and of the audit successfully detecting evasion.

When evasion is detected, the taxpayer must pay the evaded tax and an additional penalty. The tax is proportional to income, at rate $\tau$, and the penalty is proportional to the tax evaded, with penalty rate $\theta$. The taxpayer's problem is therefore to choose an evasion rate $e_s$ to maximise expected net-of-tax income:[19]

$$\big[1 - p(e_s)\big]\big[(1-\tau)\tilde{y}_s + \tau e_s\big] + p(e_s)\big[(1-\tau)\tilde{y}_s - \theta\tau e_s\big] \tag{B1}$$

Differentiating with respect to evasion, $e_s$, gives the first order necessary condition for an interior optimum:

$$\big[p(e_s) + p'(e_s)e_s\big](1+\theta) = 1 \tag{B2}$$

Analysis of this problem is straightforward, and the literature studying this (from Allingham and Sandmo (1972) onwards) has focused on comparative static predictions with respect to the probability that evasion is detected, the penalty for evasion, and the marginal tax rate.

In this context the main innovation of Kleven et al. (2011) is to note that attempts to evade tax on third party reported income are highly unlikely to succeed. This is because matching of tax returns and third party reports will detect any discrepancies, and auditing these returns will uncover the evasion. On the other hand, evasion of self-reported income is much less likely to be noticed. To capture this intuition, the authors first note that a taxpayer who wishes to evade should evade first on the sources of income which are relatively less likely to be detected, before switching to the more easily detected. This means that any evasion up to $\tilde{y}_s^{\text{self}} = \tilde{y}^{\text{perm}} + \tilde{y}_s^{\text{stoch}}$ should be of self-reported income, and only after this will taxpayers evade by misreporting $\tilde{y}_s^{\text{TPR}}$. Given this optimal structure for any evasion, the probability of detecting evasion is relatively low for $e_s < \tilde{y}_s^{\text{self}}$, but then increases quickly once $e_s$ reaches $\tilde{y}_s^{\text{self}}$. Finally, given these detection probabilities, the optimal strategy for a taxpayer is to evade some amount less than $\tilde{y}_s^{\text{self}}$.

[19] Here we present the taxpayer's problem as a static decision, independent across periods. This simplifies exposition and is equivalent to assuming that the tax authority can neither audit old tax returns when it selects a taxpayer for audit, nor condition future audit probabilities on the outcomes of past audits. The latter assumption is an accurate description of our empirical context.

## B.2 Implications of audit

Our insight builds directly on these ideas. Consider an individual who is audited (for the first time) in year $t$. Until the time of audit his optimal reporting strategy was identical to that described above: he evaded some amount less than the total amount of self-reported income he received, and none of the misreporting related to third party reported income. There are three mechanisms that might explain a change in reporting behaviour: (1) beliefs about the underlying audit rate or penalty for evasion ('belief updating'); (2) changes in the perceived reaudit risk following audit ('reaudit risk'); and (3) updates to the information held by the tax authority ('information').

In the first of these mechanisms, being audited ($D = 1$) changes the perceived audit rate or penalty. Beliefs about the background audit rate or penalty depend on audit status, $p(e, D)$ and $\theta(D)$, and vary with audit status, $p(e, 1) \neq p(e, 0)$ and $\theta(1) \neq \theta(0)$. If this belief is revised upwards, so $\theta(1) > \theta(0)$, then the cost of evasion increases and evasion falls; if it is revised downwards then the opposite occurs. Analagously for audit rate. Note that this does not require any particular assumptions on either the initial beliefs or whether updating is rational. It is simply a statement of the direction in which these beliefs about $p$ and $\theta$ change. Absent any policy changes which shift beliefs about the penalty rate, this change is permanent so any change in behaviour will also be permanent. One might also expect that updating might be different for those who actually receive penalties, compared to those who don't. The effects we see empirically are not permanent, and they do not differ between the non-compliant who do and do not receive penalties.

The second mechanism supposes instead that the perceived reaudit risk varies with time since audit: $p(e_s, h)$ varies with how many years it has been since audit, $h = s - t$. For a taxpayer that has not yet been audited, $h$ will be negative and $p(e_s, h)$ will take the same value for all $h < 0$. If he believes he is initially being monitored more carefully than before, so $p(e, 1) \geq p(e, 0)$, this leads to higher compliance immediately after the audit. Alternatively if he believes he is now unlikely to be audited for some time, the so-called 'bomb crater effect' (Mittone, 2006; Maciejovsky et al., 2007; and Kastlunger et al., 2009), then $p(e, 1) \leq p(e, 0)$, and compliance initially falls. As time since audit increases, the (perceived) effect of having just been audited wears off, so that beliefs about the probability of evasion being detected converge back towards baseline i.e. $|p(e, h+1) - p(e, 0)| \leq |p(e, h) - p(e, 0)|$ for $h > 0$.[20] Again this does not impose anything on where the perceived audit probability comes from, only on how it is updated. The implication is that as the perceived probabilty

[20]Note that, given the risk-neutrality assumption, *permanent* shifts in the level of $p(e, h)$ when $h > 0$ versus $h \leq 0$ are observationally equivalent to a shift in $\theta$. So any permanent shift in perceived audit probability, $p(e, h) - p(e, 0)$ as $h \to \infty$ is observationally equivalent to some shift in the perceived penalty $\theta(1) - \theta(0)$. This equivalence would breakdown with risk aversion, which would allow separate testing of these hypotheses.

of audit converges back to its initial level, the initial impact on reporting behaviour will decline back to baseline. This convergence will be common across incomes from all sources since the probability is common across them all. Again this mechanism does not explain our findings, since we see a differential decline in reporting across income sources, even within individual. This is not consistent with a response driven purely by a differential probability of audit in subsequent years. Instead it can only be explained by a mechanism which leads to a differential shift in reporting behaviour across income sources over time.

The final mechanism by which audits might affect reporting is that they provide information, which differentially changes the ability to hide certain sources of income. Performing an audit provides the tax authority with more accurate information on a taxpayer's income at a point in time. In subsequent years, information from the audit will make evasion of more stable income sources easier to detect. To operationalise this, recall our earlier distinction between the permanent and stochastic components of self-reported income, $\{\tilde{y}^{\text{perm}}, \tilde{y}^{\text{stoch}}_{t+h}\}$. Once the tax authority performs audits a taxpayer, $\tilde{y}^{\text{perm}}$ for that taxpayer is observed. After this, evasion of the permanent component of income is easily detected, analogous to the case with third party reported income. Hence taxpayers should now evade by misreporting $\tilde{y}^{\text{stoch}}_{t+h}$ before any misreporting of $\tilde{y}^{\text{perm}}$ (or $\tilde{y}^{\text{TPR}}_{t+h}$).

Also, if this information is the mechanism by which misreporting can be uncovered, then as the amount of the information about stochastic income is reduced over time, misreporting becomes easier. In particular, information about past income is useful because the tax authority can compare reported income in some period $h$ to their expectation of income given the past observation in the audit year. Such deviations will be more informative if past incomes are a good predictor of current income i.e. the autocorrelation of stochastic income, $\rho(h, \text{stoch}) = \text{Corr}(\tilde{y}^{\text{stoch}}_{t+h}, \tilde{y}^{\text{stoch}}_{t})$, is high. As $h$ increases, $\rho(h, \text{stoch})$ falls, so the 'amount of information' about current stochastic income is lower and so misreporting becomes easier. The case of permanent income is the natural limit of this case, where $\rho(h, \text{perm}) = 1 \; \forall h$. This also makes it clear that if multiple sources of stochastic income were available, misreporting should increase more quickly for those sources which have a lower autocorrelation. Hence under this mechanism, the prediction is that the initial impact on reporting behaviour will decline back to baseline, and this decline will be more rapid for income sources which have a lower autocorrelation. This is consistent with our findings, as seen in Figure 4.

# Appendix C Online Appendix: Additional Robustness Checks

In this Appendix we perform some additional robustness checks.

## C.1 Stratified sampling of controls

In Section 4.1 we describe the procedure for constructing the control sample for the analysis. In particular, since the treated sample was selected unconditionally randomly from those who could have filed, we select six times as many controls as there are treated individuals unconditionally randomly from the population of filers. However, due to the timing of audit selection, we note that this creates some imbalance in filing history between those who were actually treated and the controls we select.

To test whether this imbalance affects our main results, we here use an alternative approach to constructing the control sample. In particular, we perform stratified random sampling using filing history to stratify the population.[21] To do this we partition the population of individuals observed in a given year into groups by the number of years they have previously filed, plus a group combining 'more than four' into a single category.[22] For each treated individual we then select six control individuals with the same filing history.

Table C1 shows the results of the balancing tests for the new sample. Using the stratified random sampling procedure now achieves a match on 'survives' – presence in the data in the years before audit – on which stratification was conditioned, as well as maintaining the balance on other variables.

Table C2 compares the results from using this stratified random sample (Column 2) to the previous results from the unconditional random sample control group (Column 1). Point estimates and standard errors are broadly similar, as is the overall dynamic of the results. The two sets of results are never statistically significantly different from each other, although the new results imply a slightly lower tax take: the dynamic effect over the subsequent five years is £1060 rather than £1230.

[21] We thank a referee for this suggestion.

[22] More precisely, they are assigned the maximum number of years ago they were first observed to file, with first observed *more than* four years ago combined with those first observed four years ago.

## Table C1: Sample balance, using stratified random sampling to construct control sample

| Years after audit | | -5 | -4 | -3 | -2 | -1 |
|---|---|---|---|---|---|---|
| | | | Characteristics | | | |
| Female | Mean | .285 | .277 | .275 | .277 | .277 |
| | Difference | -.006 | -.006 | -.004 | -.005 | -.004 |
| | p-value | .171 | .209 | .282 | .289 | .369 |
| Age | Mean | 49.9 | 49.9 | 49.6 | 49.6 | 49.3 |
| | Difference | .081 | .115 | .072 | .065 | .142 |
| | p-value | .157 | .141 | .132 | .146 | .075 |
| In London or SE | Mean | .320 | .344 | .362 | .281 | .302 |
| | Difference | -.005 | 0 | .003 | .003 | .003 |
| | p-value | .152 | .115 | .11 | .182 | .2 |
| Has tax agent | Mean | .688 | .689 | .692 | .684 | .687 |
| | Difference | -.003 | -.005 | 0 | .003 | .003 |
| | p-value | .493 | .358 | .14 | .134 | .412 |
| Survives | Mean | .647 | .723 | .79 | .866 | .945 |
| | Difference | -.001 | -.003 | -.002 | -.002 | 0 |
| | p-value | .974 | .918 | .997 | .997 | 1 |
| | | | Income and tax totals | | | |
| Total taxable income | Mean | 24,820 | 26,746 | 29,046 | 30,278 | 32,023 |
| | Diff | 77 | -140 | -284 | 492 | 550 |
| | p value | .879 | .675 | .087 | .074 | .268 |
| Total tax | Mean | 6,333 | 6,705 | 7,334 | 7,490 | 7,898 |
| | Diff | 4 | -91 | -148 | 146 | 170 |
| | p value | .925 | .626 | .064 | .186 | .413 |
| | | | Income components | | | |
| Employment | Mean | 15,457 | 16,533 | 17,848 | 18,645 | 19,646 |
| | Diff | -297 | -343 | -264 | 384 | 364 |
| | p value | .169 | .508 | .062 | .052 | .317 |
| Self employment | Mean | 3,663 | 4,088 | 4,478 | 4,636 | 4,774 |
| | Diff | 137 | 116 | 14 | -5 | 70 |
| | p value | .058 | .527 | .47 | .394 | .395 |
| Interest and dividends | Mean | 2,393 | 2,765 | 2,856 | 2,982 | 3,133 |
| | Diff | -67 | 30 | -42 | -44 | 12 |
| | p value | .664 | .927 | .651 | .098 | .499 |
| Pensions | Mean | 1,897 | 2,172 | 2,415 | 2,680 | 2,955 |
| | Diff | 163 | 164 | 143 | 164 | 174 |
| | p value | .124 | .202 | .297 | .122 | .108 |
| Property | Mean | 500 | 510 | 561 | 559 | 570 |
| | Diff | 5 | -5 | 16 | 29 | 31 |
| | p value | .677 | .424 | .768 | .823 | .455 |

***Notes:*** *The control group is selected by stratified random sampling from the population of individuals who could have been selected for audit but were not, and stratification is based on filing history to match the treatment group. 'Years after audit' measures time relative to audit, or placebo audit for controls. 'Mean' is the mean outcome in the control (not selected for audit) group across all years. 'Difference' is the coefficient on the treatment dummy in a regression of the outcome on a treatment dummy. Treatment dummy equals 1 if taxpayer was selected by HMRC for a random audit. p-values are derived from an F-test that coefficients on interactions between treatment and tax year dummies are all zero in a regression of the outcome of interest on tax year dummies, interactions between treatment and tax year dummies. This is a stronger test than just testing the coefficient on treatment not interacted. Differences in control means between this table and Table 3 are driven by no longer conditioning on survives = 1; scaling by survives gives comparable point estimates. Monetary values are in 2012 prices. Standard errors are clustered by taxpayer.* * $p < .05$, ** $p < .01$, *** $p < .001$.

***Source:*** *Authors' calculations based on HMRC administrative datasets.*

## Table C2: Dynamic effect of audit, using different control groups

| Years since audit | Sampling (1) Unconditional | (2) Stratified |
|---|---|---|
| -5 | 60.1 | 38.8 |
| | (54) | (54.7) |
| -4 | 44.7 | 25.8 |
| | (37.3) | (38.4) |
| -3 | -55.4 | -58 |
| | (30.1) | (30.5) |
| -2 | 10.6 | 12.1 |
| | (31.3) | (31.9) |
| -1 | 6.6 | 19.7 |
| | (40) | (40.8) |
| 0 | 85.3 | 105.3 |
| | (53.8) | (54.2) |
| 1 | 137.6* | 139.4* |
| | (64) | (64.1) |
| 2 | 376.8*** | 386.4*** |
| | (69.9) | (69.8) |
| 3 | 288.4*** | 251.9*** |
| | (72.2) | (72.3) |
| 4 | 203.8** | 133.8 |
| | (78.6) | (79.1) |
| 5 | 221.8** | 151.4 |
| | (85.6) | (86.1) |
| 6 | 105.6 | 58.3 |
| | (92.5) | (93.4) |
| 7 | 141.6 | 75.5 |
| | (108.9) | (109.5) |
| 8 | 52.1 | -27.6 |
| | (138.3) | (139.9) |
| N | 2,492,210 | 2,502,906 |

***Notes:*** *Coefficients show the average difference in reported tax between individuals selected for audit, and control individuals, at different points in time relative to audit (or placebo audit for controls). Column (1) uses as controls an unconditional random sample selected from the population of taxpayers who could have been selected for audit but were not. Results come from a regression of reported tax on treatment group selection interacted with time since audit, and dummies for filing history. Column (2) uses as controls a stratified random sample of indivduals who could have been selected for audit but were not, where stratification is done on filing history and sampling rates are selected to match the treatment group. Results come from a regression of reported tax on treatment group selection interacted with time since audit.* * $p < .05$, ** $p < .01$, *** $p < .001$.
***Source:*** *Authors' calculations based on HMRC administrative datasets.*

## C.2 Adjusting trimming

In our main results we show the impact of audits trimming the top 1% of observations, to reduce noise. Since our sample is composed of the population of tax *filers*, and only a third of UK taxpayers had to file in this period (the rest are sufficiently covered by withholding at source), this implies trimming roughly the top 0.3% of incomes. Given the skewness of income distributions – see Advani and Summers (2020) for more information about measurement of UK top incomes using these data – trimming substantially reduces noise. To ensure that it does not also substantially change our results, we show below the results of alternative levels of trimming: no trimming, 0.5% trimming, 1% (the benchmark shown in the main text), and 2.5%.

As can be seen from the results in the table, varying the levels of trimming does not cause much change in the overall shape or magnitudes of these effects. The precise profile of effects changes slightly, though not in a monotone way. The peak is always two years after the year selected for audit. With less trimming than 1% the dynamic effects still decline over the same horizon, though not always as smoothly, and standard errors are larger as expected. With more trimming the same dynamics are seen as at 1%. In each of the cases considered, the results are not statistically significantly different from those in the main specification.

It is important to note again here that the comparison made in our context is between receipt of a random audit and business-as-usual. To the extent that those reporting the highest, and most complex, incomes are more likely to be selected for a targeted audit, the shift in audit probability for these individuals will be small. Hence we should not expect the treatment of these observations in the analysis (exclusion or inclusion) to substantially change the results observed.

## Table C3: Impact of audits on reported tax owed, at different levels of trimming

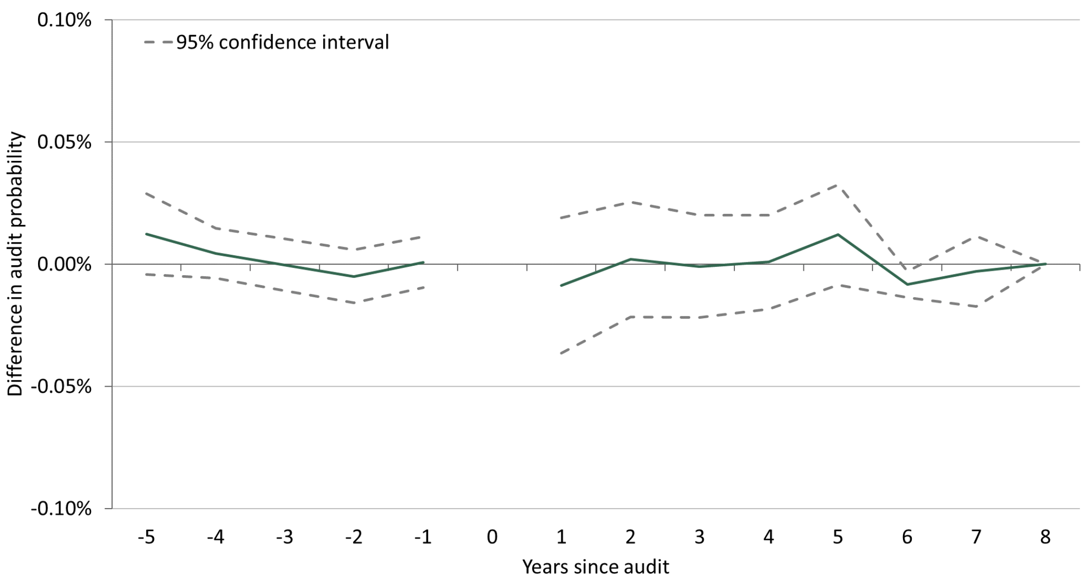
| Years since audit | Level of trimming (1) 0% | (2) 0.5% | (3) 1% | (4) 2.5% |
|---|---|---|---|---|
| -5 | 37.6 | 33.3 | 60.1 | 36.8 |
| | (85.7) | (70.2) | (54.0) | (44.6) |
| -4 | -10.5 | 3.2 | 44.7 | 38.0 |
| | (57.9) | (48.4) | (37.3) | (30.3) |
| -3 | -94.0 | -84.0* | -55.4 | -54.4* |
| | (48.0) | (39.8) | (30.1) | (24.9) |
| -2 | 22.3 | 15.8 | 10.6 | 2.4 |
| | (50.0) | (41.0) | (31.3) | (25.4) |
| -1 | 88.2 | 71.4 | 6.6 | 19.4 |
| | (62.8) | (52.0) | (40.0) | (33.0) |
| 0 | 90.0 | 82.5 | 85.3 | 70.4 |
| | (86.0) | (69.4) | (53.8) | (43.2) |
| 1 | 123.6 | 119.4 | 137.6* | 149.5** |
| | (97.2) | (80.4) | (64.0) | (52.4) |
| 2 | 333.2** | 327.4*** | 376.8*** | 345.8*** |
| | (105.2) | (88.8) | (69.9) | (56.8) |
| 3 | 193.6 | 208.5* | 288.4*** | 208.7*** |
| | (110.5) | (91.3) | (72.2) | (58.2) |
| 4 | 235.8 | 215.9* | 203.8** | 200.7** |
| | (124.4) | (102.6) | (78.6) | (64.8) |
| 5 | 160.1 | 188.4 | 221.8** | 141.5* |
| | (132.4) | (111.8) | (85.6) | (69.7) |
| 6 | 26.9 | 15.3 | 105.6 | 93.8 |
| | (145.5) | (120.5) | (92.5) | (74.9) |
| 7 | 36.8 | 85.4 | 141.6 | 153.6 |
| | (167.0) | (139.4) | (108.9) | (86.6) |
| 8 | 46.3 | 60.3 | 52.1 | 47.7 |
| | (215.2) | (183.1) | (138.3) | (112.9) |
| N | 2,519,604 | 2,509,367 | 2,492,210 | 2,462,362 |

***Notes:*** *Coefficients show the average difference in reported tax between individuals selected for audit, and control individuals, at different points in time relative to audit (or placebo audit for controls). Different columns allow for different levels of trimming: Column (1) uses no trimming, Column (2) trims the top 0.5% of individuals, Column (3) shows our baseline specification of 1% trimming, and Column (4) trims the top 2.5%. Controls are based on an unconditional random sample selected from the population of taxpayers who could have been selected for audit but were not. Results come from a regression of the income variable on treatment group selection interacted with time since audit, and dummies for filing history.* * $p < .05$, ** $p < .01$, *** $p < .001$.
***Source:*** *Authors' calculations based on HMRC administrative datasets.*

## C.3 Comparing risk of audit between treatment and control

As noted in Section 2, the control group continue to face the background risk of receiving an operational audit. One natural question this raises is whether, after audit, there is a change in audit risk for the treatment group relative to the controls. For example, if those selected for a random audit were subsequently at higher risk of reaudit, then this might explain some of the dynamics we observe. Figure C1 shows the difference in audit probability between the treatment and control groups. We see that in the years before selection (or not) for a random audit, the two groups were equally likely to have been selected for an audit i.e. operational audits were equally common for the two groups prior to selection for a random audit. In the audit year, which is excluded from the figure for readability, the point estimate for the difference in audit probability jumps to 98.1% (s.e. 0.1%), indicating that while 100% of the treated group are selected for an audit, around 2% of the controls are also. In the following years the difference is again fractions of a percent, and not significantly different from zero except in year 6 (point estimate 0.01%, p-value 2%). This lack of difference in audit selection probability, except in the year explicitly selected for random audit, implies differences in outcomes are not driven by differences in audit risk.

**Figure C1: Difference in audit probability between treated and controls**




x'' ***Notes:*** *Sample includes individuals selected for a random audit between 1998/99 and 2008/09, and control individuals who could have been selected in the same years but were not. It uses tax returns from 1998/99 to 2011/12. The solid line plots the point estimate for the difference in share of each group given an audit, for different numbers of years after the random audit (excluding zero). This comes from a linear regression of being selected for audit on dummies for years since audit (or placebo audit for controls), dummies for years since audit (or placebo audit for controls) interacted with treatment status, tax year dummies, and dummies for whether the taxpayer filed a return in each of the four years before audit. Standard errors are clustered at the individual level.*
***Source:*** *Calculations based on HMRC administrative datasets.*

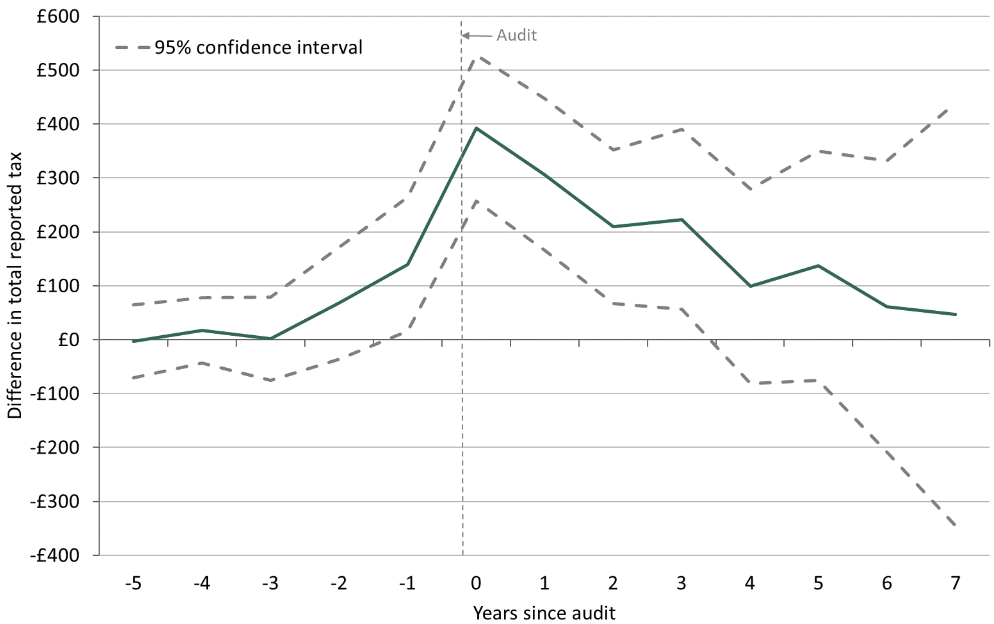
### C.4 Understanding audit timing

As noted in Section 3.1, there are substantial lags between the tax year that is being audited, filing a return for that tax year, and subsequently receiving information that an audit for that year is taking place. One might think that some of the dynamics seen in Figure 2 are therefore driven simply by the timing of audits. However, naively using the time when an audit is actually opened risks creating bias, since the timing of opening among the selected is likely non-random. An alternative therefore is to estimate the timing of audit opening among the treatment group, and then use the mean and distribution of residuals to impute the timing of opening for the controls, with the same distribution of opening timing as for the treatment group.[23] We predict the timing of opening using only the tax year selected for audit. Other available characteristics are likely to be correlated also with outcomes and hence be endogenous. Whilst this might not be problematic if the same correlation held among the controls, since some controls are actually audited (as noted previously they may still receive targeted audits), our results would then become sensitive to correctly matching these controls with the treated individuals who would have been selected for a targeted audit. Further, those randomly audited individuals who would otherwise have been selected for a targeted audit (about 60 per year) would also need to have the same realised characteristics distribution as the audited individuals in the controls. Given the bias that would be caused if this were not to hold, we instead exclude potentially endogenous individual characteristics.

Figure C2 shows the results of this estimation, implementing the same specification as in the main results but using the adjusted timing of $h = 0$ in the tax year in which audit is predicted to be opened. There are three points to note. First, we see that the effect now peaks in year zero, and this peak is similar in magnitude to the peak previously seen in Figure 2. Second, there is some effect now in the year before the audit is opened. This suggests the effects we are observing are driven by reporting rather than real effects: many individuals will not have filed the previous years tax when the audit is opened, since the filing deadline is around 10 months after the end of the tax year. Hence these individuals are responding to an audit in the current tax year by increasing reporting for the previous, unfiled tax year. Third, we again see a slow decline over the subsequent seven years. The dynamics we see are therefore not purely an artefact of audit timing.

[23]We thank a referee for this suggestion.

**Figure C2: Dynamic effect of audits on total reported tax owed, measuring time since audit predicted to have been opened**



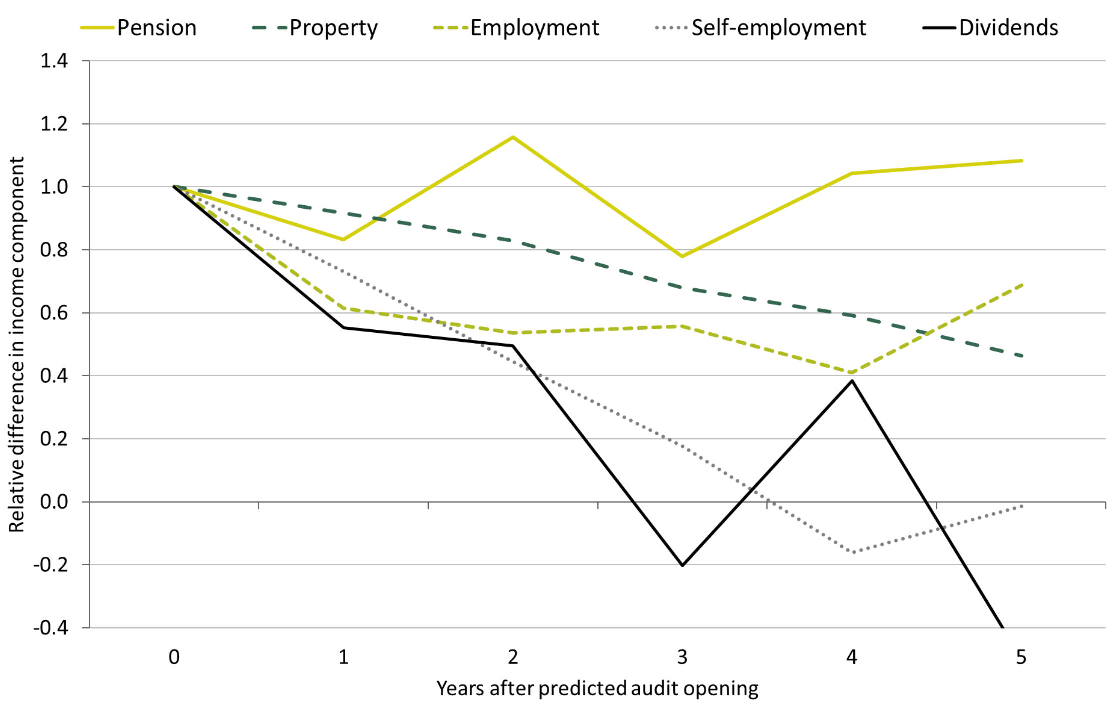

***Notes:*** *Sample includes individuals selected for a random audit between 1998/99 and 2008/09, and control individuals who could have been selected in the same years but were not. It uses tax returns from 1998/99 to 2011/12. The solid line plots the point estimate for the difference in average 'total reported tax' between individuals who were and weren't audited, for different numbers of years after the tax year in which the audit is predicted to have opened. This comes from a regression of total reported tax on dummies for years since predicted audit opening, dummies for years since opening interacted with treatment status, tax year dummies, and dummies for whether the taxpayer filed a return in each of the four years before audit, with audit status instrumented by selection for audit. Standard errors are clustered at the individual level.*
***Source:*** *Calculations based on HMRC administrative datasets.*

## C.5 Understanding audit timing by income source

The previous section studied the dynamics of reported tax owed over the years after *predicted* audit opening, rather than after the tax year in which selection for audit was made (as in the main text). The peak was seen to be in the predicted year of opening. In this section we perform the same exercise, but looking across income sources. Figure C3 repeats Figure 4, but now looking at time since audit was predicted to be opened.

As before the same pattern of results is seen. Relative to the initial impact in the predicted year of opening, pension income – which is highly autocorrelated – does not subsequently appear to decline at all. Property income and employment income are the next most autocorrelated, and decline relatively slowly. Self-employment and dividend income are least autocorrelated and decline most quickly.

**Figure C3: Relative dynamics by income source, over time since predicted audit opening**




***Notes:*** *Sample includes individuals selected for a random audit between 1998/99 and 2008/09, and control individuals who could have been selected in the same years but were not. It uses tax returns from 1998/99 to 2011/12. Each line plots the point estimate for the difference in the average of a particular component of income between individuals who were and weren't audited, for different numbers of years after the tax year in which the audit is predicted to have opened. This comes from a regression of each income component on dummies for years since predicted audit opening, dummies for years since opening interacted with treatment status, tax year dummies, and dummies for whether the taxpayer filed a return in each of the four years before audit, with audit status instrumented by selection for audit.*
***Source:*** *Calculations based on HMRC administrative datasets.*

## C.6 Comparing stable and unstable income sources

In Section 4.3 we note that one concern about the results from comparing across income sources (Figure 4) is that these measures are noisy, so if confidence intervals were added many would overlap. One way we tackled that is to formally test that less autocorrelated sources declined faster than more autocorrelated ones, using pairwise comparisons within individual.

Another approach, which we show here, is to divide income sources up into two categories: stable (employment, property, pensions) and unstable (self-employment, dividends). For each individual, we normalise both stable and unstable income by the value it took in the year selected for audit – this ensures that results between stable and unstable are comparable despite the different absolute amounts of income. We then perform the same analysis as in the main results, comparing stable/unstable income between those randomly selected for treatment, and those randomly not selected. Table C4 shows the results, which can be interpreted as the proportional increase in reported stable/unstable income.

We see that this peaks in year 2, for both stable and unstable sources. The peak is higher for unstable income sources, and declines more rapidly. Fewer individuals report any amount of unstable income, and there is naturally more variation in these income sources, so the standard errors are larger. In this specification, stable income sources are statistically significantly higher among those audited until five years after the year selected for audit. For unstable sources they are never statistically significantly different from zero, even at the peak where the point estimate is larger than for stable income. In the point estimates there is a clear pattern of sharp decline as seen previously.

**Table C4: Impact of audit by stability of income source**

| | Sampling | |
|---|---|---|
| Years since audit | (1) Stable | (2) Unstable |
| -5 | -.007 | .017 |
| | (.005) | (.017) |
| -4 | .004 | .003 |
| | (.003) | (.014) |
| -3 | -.005 | -.006 |
| | (.003) | (.012) |
| -2 | .003 | .004 |
| | (.003) | (.014) |
| -1 | -.001 | -.001 |
| | (.003) | (.017) |
| 0 | .000 | .017 |
| | (.004) | (.015) |
| 1 | .007 | .007 |
| | (.006) | (.024) |
| 2 | .025*** | .032 |
| | (.007) | (.026) |
| 3 | .016* | .015 |
| | (.007) | (.027) |
| 4 | .016* | .009 |
| | (.007) | (.032) |
| 5 | .016* | .010 |
| | (.008) | (.034) |
| 6 | .012 | -.006 |
| | (.009) | (.035) |
| 7 | .013 | .014 |
| | (.010) | (.040) |
| 8 | -.001 | .104 |
| | (.012) | (.061) |
| N | 1,505,826 | 428,526 |

***Notes:*** *Coefficients show the average difference in income between individuals selected for audit, and control individuals, at different points in time relative to audit (or placebo audit for controls). Column (1) uses the sum of all stable income sources – employment, property, and pensions – as the outcome variable. Column (2) uses the sum of all unstable income sources – self-employment and dividends – as the outcome variable. Controls are based on an unconditional random sample selected from the population of taxpayers who could have been selected for audit but were not. Results come from a regression of the income variable on treatment group selection interacted with time since audit, and dummies for filing history.* * $p < .05$, ** $p < .01$, *** $p < .001$.
***Source:*** *Authors' calculations based on HMRC administrative datasets.*