Journal of Public Economics 238 (2024) 105195

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Rounded Up: Using round numbers to identify tax evasion☆

Robert Breunig [a,*], Nathan Deutscher [a], Steven Hamilton [b]

[a] Tax and Transfer Policy Institute, Crawford School of Public Policy, Australian National University, Canberra, ACT 2601, Australia
[b] Department of Economics, George Washington University, Washington, DC 20052, United States of America

ARTICLE INFO

JEL classification:
D9
H20
H24
H26

Keywords:
Bunching
Tax evasion
Tax refunds
Taxpayer behavior
Tax preparer behavior
Loss aversion
Left-digit bias

ABSTRACT

Australian taxpayers display a clear preference for round numbers for end-of-year tax refunds, bunching at positive and salient thresholds such as the tens, hundreds and thousands. Bunching appears to be driven by tax evasion. Data from audited returns shows that bunching is present in returns before audit, but does not persist post-audit. Tax preparers play an important role, being twice as likely to deliver positive round-number refunds as individuals who file their own tax returns. Preparers with greater propensity to bunch deliver larger refunds by lifting deductions and lowering reported income for return items where audits are costly. This highlights how bunching behaviors can help identify tax evasion, including tax preparers who facilitate it and the tax return items which are manipulated.

1. Introduction

In 2018 a typical Australian taxpayer was 84 per cent more likely to be owed a dollar by the government than to owe a dollar, and 70 per cent more likely to be owed $A1000 than $A999. We document the widespread bunching of tax refunds at positive and salient round-number thresholds such as multiples of $A10, $A100 and $A1000. Drawing on nearly three decades of administrative tax data linking individuals, their tax returns and their preparers, coupled with random audit data, we use bunching behavior to draw out broader insights on the role of tax preparers in the tax system and on tax evasion.

Using data from random audits conducted by the Australian Taxation Office, we observe a systematic relationship between bunching and audit outcomes. Amongst individuals whose balance is adjusted downward post-audit, we observe pre-audit bunching but no post-audit bunching. Amongst those whose balance is not adjusted, or adjusted upwards, neither post- nor pre-audit balances show any statistically significant bunching.

Tax preparers play an important role in this bunching behavior. People who use tax preparers are twice as likely to bunch at round number refunds as those who do not use tax preparers. Individuals who receive a refund at or just above one of these thresholds are more likely to remain with their tax preparer in the following year, though are not charged higher fees. The extent of bunching has grown dramatically over time, likely reflecting these dynamics and the advent of electronic returns.

Preparers differ greatly in their propensity to deliver round-number refunds to their clients. Using longitudinal information on taxpayers and exploiting changes in the tax preparer that individuals use, we show that individuals who switch to a preparer with a higher propensity to bunch (“a higher-bunching preparer”) increase their probability of bunching to levels equal to the other clients of that preparer. These taxpayers show no increase in the probability of bunching in the years before switching to the higher-bunching preparer. We compare the propensity across all clients of the new preparer and the old preparer to bunch at a round number.

☆ We thank the Australian Taxation Office for providing us with access to the data which allow us to identify common tax preparers among tax filers and the data on their random audit program. All findings and conclusions are those of the authors and do not represent those of the Australian Government or any of its agencies. We would like to thank Michael Best, Alex Rees-Jones, Arthur Seibold, seminar participants at the Tax and Transfer Policy Institute, and attendees at the National Tax Association annual meetings and the Columbia University tax workshop for useful comments and feedback. We thank Ric Curnow for editing assistance. We thank three anonymous referees and the editor of the *Journal of Public Economics* for comments which have greatly improved the paper. The research plan was approved by the Australian National University Human Ethics Committee, protocol number 2020/538.

* Corresponding author.
E-mail addresses: robert.breunig@anu.edu.au (R. Breunig), nathan.deutscher@anu.edu.au (N. Deutscher), steven_hamilton@gwu.edu (S. Hamilton).

https://doi.org/10.1016/j.jpubeco.2024.105195
Received 5 November 2023; Received in revised form 20 July 2024; Accepted 22 July 2024
Available online 13 August 2024

A move from a preparer that does not bunch to one that bunches at the same rate as observed across our full sample results in a persistent increase in an individual's tax refund of $A52 for bunching around the zero threshold, and $A17 for bunching around the hundred-dollar thresholds. Using these as estimates of personal income tax evasion and mapping to the full population of 15 million taxpayers, implies a fiscal cost of around $A260-780 million a year for the evasion proxied for by bunching.

When taxpayers switch to a high-bunching preparer, we typically observe three changes. First, their reported income falls. The decrease in reported income is driven only by decreases in reported income from businesses, partnerships and trusts. No decrease is reported in salary and wage income. Second, taxpayers claim larger work-related deductions. Third, reflecting lower income and higher deductions, taxpayers receive larger refunds.

In the years following a switch to a high-bunching preparer, we do not see subsequent changes in bunching, reported income, amount of deductions or size of refund claimed over time. A step change occurs when a taxpayer changes to a high-bunching preparer, but no further change.

The effects of high-bunching preparers are robust to a range of standard concerns with event study designs. Results barely change with the inclusion of time-varying controls for location, occupation and income. Further, the effects are relatively symmetric, which suggests they are not driven by learning or uni-directional shocks. They also remain when focusing on those moving amidst large outflows from their preparer – reflecting significant downsizing or closure – which suggests the choice to move does not drive our results.

What conclusions about taxpayers, tax preparers and tax evasion can we draw from these observations?

Taxpayers have preferences for round-number refunds. This is consistent with models of loss aversion and left-digit bias. Loss aversion has previously been studied in relation to taxes (see Rees-Jones (2018) and Engstrom et al. (2015)), but our study makes a novel contribution in documenting bunching in tax refunds consistent with left-digit bias. We present a model which combines loss aversion with left-digit bias and which relates bunching behavior to the cost of reducing one's tax liability. Bunching at round number refunds will reflect preferences, risk aversion and other costs of bunching. While we cannot separate out these factors empirically, electronic filing and tax preparer returns have reduced the cost of bunching. While bunching is rising over time, no evidence from other sources suggests the increase reflects a more general rise in risk-taking behavior with respect to the tax system in Australia.

Our paper makes several novel and important contributions about the behavior of tax preparers. Tax preparers facilitate bunching at round numbers for refunds. Some preparers produce higher rates of round-number refunds. This may reflect a preference of their clients. While it is not associated with higher fees, it is associated with higher client retention. Tax preparers may be engaging in this behavior to keep customers and grow their business.

A great deal of heterogeneity is apparent in the extent to which refunds bunch at positive, salient balances across preparers. The extreme bunching of some preparers leaves open the possibility that some tax preparers may be 'behavioral' in the sense that they derive utility from achieving round number refunds for their clients irrespective of clients' preferences. High-bunching preparers are more likely to charge fees ending in zero, evidence supporting the behavioral hypothesis (as does the more general observation that preparers bunch returns more than individuals).

Our paper's most significant contribution is to demonstrate the relationship between bunching and tax evasion. Tax preparers may achieve round number refunds by working harder and putting in greater effort to achieve legitimate reductions in the taxable income of clients. We do not find evidence for this. Instead, we find the move to a high-bunching preparer has a one-off effect on balances, reported income, bunching propensity and refunds. The effect does not increase over time. This argues against a story where high-bunching preparers work with their clients to better document expenses or to tax plan and shift their affairs over time to take advantage of tax law.

We also see that high-bunching tax preparers have large and precisely estimated impacts on those items in the tax return that are difficult for the tax authority to verify: business-type income and work-related deductions. Salary and wage earnings, which are confirmed through third party reporting, do not change when a taxpayer switches to a high-bunching preparer. The quantities that high-bunching preparers manipulate are costly to verify and exactly those quantities that most concern the Australian Taxation Office when it comes to integrity (Australian Taxation Office, 2021a).

Random-audit data provide strong evidence that bunching is driven by evasion rather than effort. Tax returns which have negative outcomes from audits (where the tax authority determines that more tax is owed after the audit) show bunching pre-audit but no bunching post-audit. Those tax returns that require no negative adjustment after audit show no statistically significant bunching either before or after the audit.

All this suggests the observed refund bunching at positive and salient round numbers is driven by tax evasion activities rather than by additional effort. High-bunching preparers are not more skilful, just more willing to take risks. This has an important policy implication for tax authorities. Audits are costly and better targeting of audits has the potential to reduce the tax gap and lower administration costs. For impact, tax authorities' compliance activities might target both round-number refunds and tax preparers who are 'high-bunching'.[1] Given current patterns of behavior, preparers' tendency to deliver round-number bunching is a useful proxy for preparers' tendency to pursue evasion.

2. Background

The round-number bunching we document reflects a type of left-digit bias. Individuals display left-digit bias in a wide variety of settings. Perhaps most apparent is the long-documented prevalence of prices ending in '9' (Ginzberg, 1936). Similar patterns are also observed in the relationship between a used car's mileage and its price (Lacetera et al., 2012), marathon times (Allen et al., 2017) and financial markets (Heraud and Page, 2024). Perhaps less well understood is the role of other economic actors such as businesses and governments in facilitating these preferences. These actors, in our case tax preparers, may themselves have left-digit bias.

A nascent literature seeks to better understand the behavior of individuals, firms and governments in the presence of 'behavioral' preferences such as loss aversion and left-digit bias.[2] List et al. (2023) and Strulov-Shlain (2023) study pricing in the presence of left-digit bias at Lyft and more general retail settings, respectively, and find that pricing strategies fail to fully adapt to such preferences. Our setting allows us to examine the dynamics of such interactions between preferences and outcomes over nearly three decades. Furthermore, our paper adds to those that highlight how these preferences can be used to draw broader inferences. For example, Dube et al. (2018) provide evidence that employers (rather than employees) drive bunching at round-number hourly wage rates, which they argue is most easily rationalized in a labor market with monopsony power. Reyes (2022)

[1] The benefits of this may of course be temporary if the targeting of such behaviors becomes known and preparers and taxpayers adjust accordingly. The efficacy and longevity of this approach will depend upon how quickly preparers and taxpayers adjust to a new regime.

[2] We occasionally use the term 'behavioral preferences' in this paper to refer to preferences exhibiting loss aversion and left-digit bias, as opposed to more classical preferences which do not. See Heidhues and Kőszegi (2018) for a recent review of behavioral industrial organisation.

also examines such behaviors, arguing that it can proxy for the quality of firm decision making, as firms hiring workers at such wages appear less sophisticated and have worse market outcomes.

We explore bunching around 'behavioral' notches with a view to what we can learn about the broader process of and market for filing tax returns.[3] This is a large market—each year about 15 million Australians file a tax return and, despite significant advances in simplifying tax returns, about two thirds of these use the services of a tax preparer.[4] In the 2018–19 income year the cost of managing individual tax affairs, including tax preparer fees, was around $A2 billion (or over 0.1% of GDP).[5] Decisions made at the point of filing also have a substantial bearing on government revenues.

Past work has highlighted that taxpayers at the point of tax filing act to avoid owing a debt to the tax authority; that is, there is bunching at positive balances. For example, Rees-Jones (2018) quantifies loss aversion in US tax data from 1979–1990, showing that taxpayers facing a liability reduce their liability by $US34 more than those facing a refund. Notably, Rees-Jones (2018) also finds that bunching at zero is slightly more pronounced among tax preparer returns. Exploiting the 'preliminary balance' calculated by Swedish tax authorities prior to tax filing, Engström et al. (2015) show that taxpayers respond to a preliminary deficit by claiming more deductions; similarly (Jones, 2020) shows that Texan homeowners disproportionately appeal property tax assessments that increase their assessed value (and tax liability). Separately, the role of tax preparers has also been explored, with Battaglini et al. (2019) studying their role as information hubs for small business clients and interactions with the audit process. Governments have been found to respond to left-digit bias, for example in Danish municipality tax rates (Olsen, 2013). But the interaction of tax preparers and behavioral preferences remains, to our knowledge, unstudied.

We will show the presence of bunching above zero and above positive and salient thresholds, consistent with loss aversion and left-digit bias. The left-digit bias results in refund balances are novel. We will show the large role played by tax preparers and evidence that bunching, particularly that associated with tax preparers, is linked to tax evasion.

While the fiscal impact of bunching is modest – of the order of a few million dollars a year – theory and our results suggest that it can serve as a proxy for the cost of reducing tax liabilities more broadly where the fiscal costs are orders of magnitude higher. Our results highlight the significant discretion exercised in the final stages of the tax return, and how behavioral notches may shed light on how and where such discretion is exercised.

We present our data and document the bunching behavior in tax refunds in Australia. We examine the correlates of bunching behavior and the important role of tax preparers. We present a simple theoretical framework of tax filing to illustrate the possible behaviors that could drive the observable bunching in the data. We use the random audit data to show that bunching appears to be driven by evasion. We then examine the role of tax preparers in more detail. We first ask why tax preparers might deliver positive, salient balances. We then examine the effect that high-bunching preparers have on tax returns, before concluding in the final section.

3. Data

3.1. Australian Longitudinal Information Files (ALife)

We primarily draw on the Australian Longitudinal Information Files (ALife) produced for research purposes by the Australian Taxation Office (ATO), see Abhayaratna et al. (2021). ALife is a 10 per cent sample of all individual tax returns from the 1991 to 2018 income years.[6] For our research, the ATO has supplemented ALife with a random identifier linking tax returns prepared by the same practice of tax preparers.

The key tax return variable in our analysis is the 'balance assessed'—the amount owing to the ATO after a taxpayer's tax liability is set against tax withheld through the year, and any additional refunds or credits owing. For presentation purposes, we reverse the sign on this variable and round it down to the nearest dollar so that it is the amount owed to the individual. While this variable is generally high quality, there are some instances where the balance assessed in ALife is not consistent with the remainder of the information in the return. When examining bunching, we drop returns where this is the case, which consists of less than 1% of returns from 2001 onwards, less than 7% of returns from 1995–2000 and around a third of returns from 1991–1994; across all years this constitutes around 5.3% of returns.

As noted by Rees-Jones (2018), the tax system can mechanically give rise to bunching at a variety of thresholds. For example, Australia has had a variety of non-refundable tax offsets for low-income individuals. These act to reduce the taxpayer's liability to zero but no further: this results in a mass point at a balance of zero for those who have no taxes withheld through the year. The same mechanism also produces mass points at common refundable offset amounts.[7] Given this mechanical bunching is not of interest, we further restrict attention to taxpayers with a positive net tax liability. Some individuals have a tax liability but no tax withheld. These people automatically have a negative balance (a positive amount of tax owed) and their presence creates a mechanical discontinuity at zero. We also remove them.[8] Together these restrictions drop a further 27.9% of returns. We are not concerned about bunching arising due to a tendency towards round numbers elsewhere—for example, in wage and salary income or in deduction amounts. This is because round numbers in taxable income will not flow through to round numbers in tax liabilities, due to the effect of marginal tax rates.[9]

We draw on a variety of other variables for deductions claimed in the process of filing tax returns. These include deductions for work-related expenses, and for expenses incurred in managing a rental property; these variables are available since 1992 and 1993 respectively. We use the cost of managing tax affairs as a proxy for tax preparer fees for the previous year.[10] This is a noisy proxy, as taxpayers may claim other expenses under this deduction, such as the cost of tax reference materials, tax courses and travel to their tax preparer.[11] Nonetheless, the distribution of these deductions, shown in Appendix Fig. B.1, shows several clear modes as one would expect if they reflected pricing behaviors in the market for tax returns. This variable is available from 2000 onwards.

[3] We call these notches behavioral because they arise from the behaviors of individuals rather than from features of the tax system such as changes in tax rates.

[4] See Australian Taxation Office (2021b), Individual Statistics, Chart 7.

[5] See Australian Taxation Office (2021b), Individual Statistics, Table 6.

[6] Individuals (not households) are the primary unit of taxation in Australia. Australian income years run from 1 July through to 30 June; we will refer to income years by the year in which they end.

[7] For example, the Education Tax Refund, which entitled eligible taxpayers in 2008–09 to a refund of up to $A750 per child for education expenses.

[8] These align with the sample selection choices made in Rees-Jones (2018).

[9] For example, based on the current Australian tax schedule, an extra $A100 deduction would reduce tax liabilities by $A19, $A32.50, $A37 or $A45 at each of the four marginal income tax rates, before considering the application of any additional levies or surcharges.

[10] Tax returns are typically filed (and expenses incurred and hence deductible) in the year following the return year.

[11] See https://www.ato.gov.au/individuals-and-families/income-deductions-offsets-and-records/deductions-you-can-claim[Accessed 7 April 2024].

Table 1
Summary statistics.

	ALife data (1991–2018)			REP data (2016–2020)
	All years	1991	2018	All years
Mean balance ($A)	697	754	497	1,302
SD balance ($A)	25,247	1,714	39,370	7,529
Mean average tax rate (%)	19	20	19	.
Mean age	40	36	42	.
% female	45	41	46	.
% self-employed	7	4	9	.
% tax preparer	72	65	71	77
% paper	14	31	4	.
# returns ('000)	22,525	396	1,050	3.2
# individuals ('000)	1,940	396	1,050	.
# preparers ('000)	48	14	21	.
% 'evading'	.	.	.	70

Note: Summary statistics for our baseline ALife and Random Enquiry Program (REP) samples. The balance for the REP sample is the balance before audit. A taxpayer is defined as 'evading' if their balance after audit is strictly lower than before audit.

3.2. Random Enquiry Program (REP) data

To supplement the main analysis, we draw on a custom data extract from the ATO's 'Random Enquiry Program'. Each year, this program audits the tax returns of a representative random sample of individuals.[12] The results of the audits inform estimates of the 'tax gap'—the difference between tax collected and the tax that would be collected if "every taxpayer was fully compliant with the law"; see Australian Taxation Office (2024). We observe tax return outcomes, including the balance assessed, before and after the audit process. We do not observe all the outcomes required to make the same sample restrictions as in the ALife sample. In particular, we do not have net tax after audit and are unable to remove mechanical bunching at the zero threshold. As such we focus on the hundred-dollar thresholds when analyzing the audit data. Our extract covers the 2016–2020 income years. It includes only those subject to manual review and is thus skewed somewhat towards those with more complicated tax affairs.[13]

3.3. Summary statistics

Table 1 presents summary statistics across our two datasets. The ALife data consist of 22.5 million returns. The average balance assessed is several hundred dollars, though the standard deviation is large. Over the nearly three decades we examine, there has been little change in average tax rates, a modest ageing of the population, and growth in the share of female and self-employed taxpayers.[14] Throughout the period around 70% of taxpayers have used a tax preparer but paper returns have fallen from 31% to 4% of returns.

The REP data are much smaller, consisting of 3,172 returns. Possibly reflecting the skew of the sample to those with more complicated affairs, balances are higher and tax preparer returns more common than in the ALife data. The key advantage of the REP data is that we observe outcomes before and after audit. Around 70% of taxpayers have a lower balance after audit than they did before audit and can, in this sense, be thought of as having 'evaded' taxes. Random audit programs are an important way of learning about such evasion, but are costly to run. As we will show, responses to behavioral kinks and notches provide a less direct, but cheaper way to shed light on how tax returns may be manipulated.

[12] More details on the Random Enquiry Program can be found in the report of the Auditor General, (Australian National Audit Office, 2023).

[13] Individuals with simple affairs, such as income that can be verified from third party sources and no deductions, are not subject to manual review.

[14] Self-employed tax payers are defined by the tax office as having business, partnership or personal services income.

4. Bunching at positive, salient balances

Fig. 1 shows the distribution of the balance at assessment across all 22.5 million tax returns in our sample. Amongst several notable features are both a clear peak at zero and a positive discontinuity. Taxpayers have a tendency to avoid a debt to the tax office on assessment. There are also regularly spaced peaks coincident with hundred and thousand dollar balances. Taxpayers appear to adjust their tax returns to target positive, salient refunds on tax day.

Multiple ways to quantify the behavior emerge from Fig. 1. A simple and intuitive approach is to estimate the discontinuity at the given thresholds using local linear regression. For example, if c_b is the count of returns at integer balance b we can estimate:

$$c_b = \alpha + \beta b + \delta \mathbb{1}[b \geq \tau] + \varepsilon_b \tag{1}$$

in some window around a threshold τ. In this case, the expected discontinuity at the threshold results in a jump in the density of $100\frac{\delta}{\alpha+\beta\tau}$ %. While this is a fairly rigid approach to quantifying behavioral responses, the parametric form and parsimony allow us to readily incorporate covariates (Section 4.2) and much smaller samples, namely clients of individual tax practices (Section 6).[15]

It is helpful to normalize the key variables in this equation so that the δ coefficient has a more direct and consistent interpretation. In particular, we can instead estimate:

$$\tilde{c}_{\tilde{b}} = \alpha + \beta\tilde{b} + \delta(2 \times \mathbb{1}[\tilde{b} > 0] - 1) + \varepsilon_b \tag{2}$$

where we have normalized the count by dividing it by its average over the estimation window ($\tilde{c}_{\tilde{b}} = c_{\tilde{b}}/\bar{c}$), and have re-centered the balance around zero ($\tilde{b} = b - \tau + 1/2$). We will typically use a symmetric estimation window looking $A50 either side of the threshold, in which case the normalized count is the percentage of observations in each integer bin. With no discontinuity at the threshold ($\delta = 0$), this simply estimates a line of best fit that will pass through the average of the normalized count (1) and our re-centered threshold (0). A nonzero discontinuity introduces a symmetric deviation from this, with a value of $1-\delta$ to the left of the threshold and $1+\delta$ to the right. The percentage jump in the count at the threshold is then $100(\frac{1+\delta}{1-\delta}-1)$%, which for small values of δ is approximately 200δ%.

In Fig. 2 we zoom in on the behavior in Fig. 1 and illustrate our approach to quantifying bunching around the zero-, ten-, hundred- and thousand-dollar thresholds. For the last three we select all observations within either an $A5 or $A50 window of the given threshold and then stack our windows. Across all four panels, we see clear bunching at the thresholds. Unlike settings where ability to manipulate is imperfect – such as the marathon times examined in Allen et al. (2017) – we do not see any uptick in the density function below the thresholds; taxpayers do not fall short. This is unsurprising given common tax filing programs allow tax filers to see their calculated balance before they finalize and lodge their return. Taxpayers do, however, sometimes overshoot. This is consistent with taxpayers facing discrete manipulation opportunities, as in the model outlined in Rees-Jones (2018). Fig. 2 also shows in red lines the predicted values from ordinary least squares estimation of Eq. (2). The estimated discontinuities in the normalized count at the thresholds are 0.412 (zero), 0.004 (10s), 0.021 (100s) and 0.073 (1000s). Bunching is most extreme around the zero threshold and for the higher powers of ten.

[15] More standard non-parametric approaches to estimating this bunching are much more challenging to implement in this setting, where the thresholds are many and the windows in which bunching takes place are overlapping. Unlike work aimed at uncovering a specific fundamental parameter, such as the elasticity of taxable income, it is not essential for our approach to precisely measure the excess mass at each threshold, and the more parsimonious approach better lends itself to the exercises of interest in this paper.

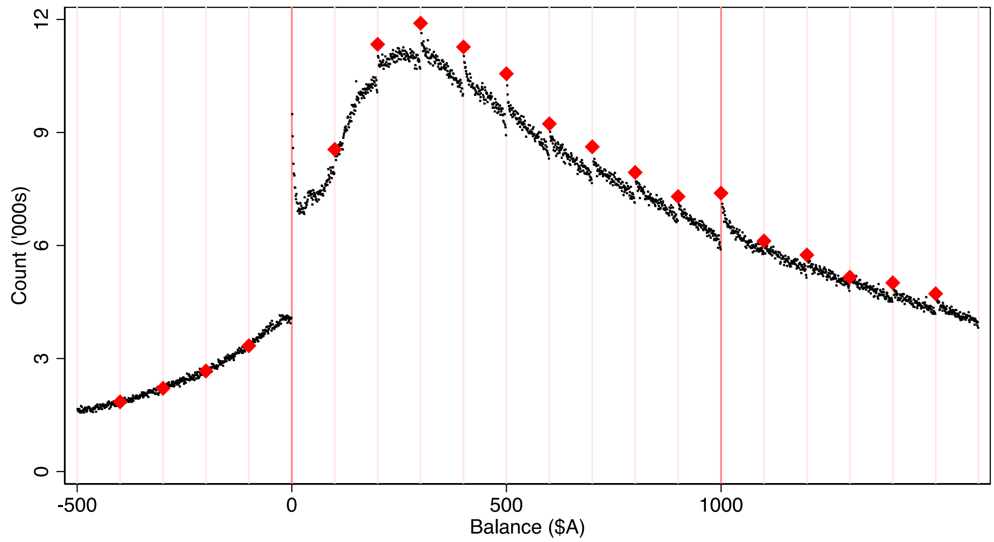

Fig. 1. Distribution of balance of assessment, 1991–2018.
Note: Distribution of the balance of assessment for those with: strictly positive net tax liability and tax withheld; and a balance of assessment consistent with the remainder of the tax return. Counts are for each $A1 bin, with diamonds indicating counts at positive hundred-dollar thresholds. For visual clarity we exclude the count at zero balance of assessment, which is 22,870. Sample consists of 22.5 million tax returns over the 1991–2018 income years.

Finer-grained insights into bunching are presented in Appendix Fig. B.2, where we show how the estimated discontinuities change for specific thresholds. The degree of bunching is generally similar for successive thresholds of the same type, but the exceptions are instructive. Bunching declines as balances increase, particularly where the thresholds have less salience. For example, the discontinuities at hundred-dollar thresholds are largest for balances below $A1000, consistent with classic left-digit bias. However, it also appears that '5s' matter, with larger discontinuities at $A500 thresholds than other $A100 thresholds. Hundred-dollar thresholds matter far less for negative balances, but still significant discontinuities appear at the $A1000 thresholds—consistent with a desire to avoid particularly salient debts. Finally, for those using a tax preparer to file a tax return, their frame of reference for a loss may be the balance net of the tax preparer's fee. As noted earlier, a proxy for this fee can be obtained by the taxpayer's claim for the cost of managing tax affairs in the subsequent year. In Appendix Fig. B.3 we indeed show a modest discontinuity in this variable around zero.

Estimating the number of taxpayers shifting in response to loss aversion and left-digit bias is challenging. The sheer number and frequency of bunching points make it difficult to reliably identify regions outside the bunching window and then estimate a counterfactual density. In Appendix D we show that 0.4–0.6% of taxpayers move in response to the hundred-dollar thresholds (including thousand-dollar thresholds). This relatively small response implies a small fiscal cost. For example, across all the years around 183 million taxpayers were within $A50 of a hundred- or thousand-dollar threshold (everyone with a balance of $50 or greater will be in one of these bunching windows).[16] Applying our preferred relative excess mass estimate of 0.5%, we have 0.9 million taxpayers shifting their balances in response to these thresholds. For an indicative upper bound fiscal cost, assume all the excess mass comes from taxpayers increasing their balances (in reality, some will come from taxpayers reducing their balances). With a $A100 estimation window the average difference in balances between those in the excess mass and missing mass regions will be $A33, which implies a total fiscal cost of $A30 million, or around $A1 million a year. This is a tiny fraction of the total tax take.

While the fiscal cost of bunching itself may be relatively modest, it nonetheless captures information that has wider-ranging implications for tax filing, as shown in the theoretical framework that follows. For the remainder of the paper we explore in more detail what drives bunching at round numbers and what we can learn from these behaviors. We focus on the bunching at zero and at hundred-dollar intervals (combining the hundred and thousand dollar thresholds).

4.1. Theoretical framework

In this section we briefly consider the theoretical predictions from a simple model of tax filing. We also consider the implications of behavioral preferences that reflect the tendency of taxpayers to prefer positive balances, and balances above particularly salient thresholds, such as hundreds and thousands. The aim is to provide some intuition for what determines the degree of bunching at these positive, salient thresholds, which we can subsequently test in our data, but also to relate these behaviors to the broader process of tax filing.

In our setup, an individual taxpayer maximizes their utility $U(b)$ derived from their balance b. Utility is additively separable in benefits and costs such that $U(b) = v(b) - c(b)$ for some value and cost functions $v(.), c(.)$. Given balances are typically small relative to taxpayer income we assume a linear value function, given by:

$$v(b) = v'b. \tag{3}$$

for some $v' > 0$.[17] We also assume a quadratic cost function with $c'' > 0$ that is minimized at some balance B_0. We can think of B_0 as the default balance arising from a return where taxpayers minimize the costs arising from both effort and risk of audit. This would involve reporting all income likely to be reported to the tax authority, but not making any manual claims for deductions or offsets. The parameters defining the utility and cost functions, and the default balance, may all differ between individuals.

Taxpayers with the same observable characteristics should have the same default balance, as they will have had the same amounts withheld and the same tax liability on observable income. But they will differ in their optimal balance due to differences in the marginal benefit,

[16] Multiplying by ten the 18.3 million observed in the ATO's 10% sample. This implicitly assumes that the people whom we drop from our data do not bunch so this number is likely an underestimate.

[17] The typical ratio of the absolute value of balance to taxpayer total income in our sample is 2.8%, among those with strictly positive income.

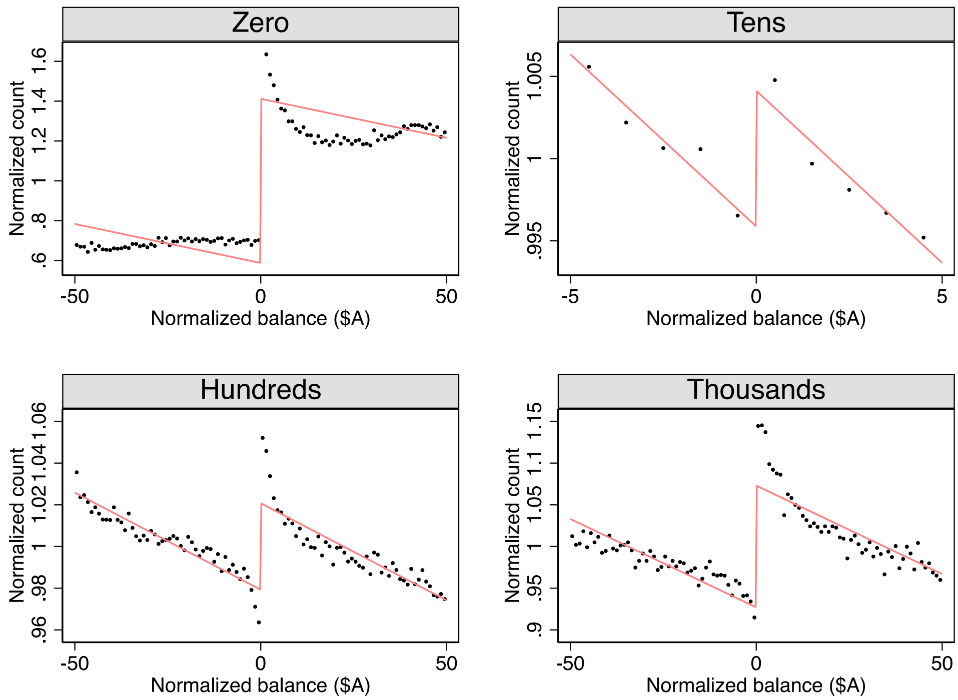

Fig. 2. Distribution of balance of assessment around salient thresholds, 1991–2018.
Note: Distribution of the balance of assessment for those with: strictly positive net tax liability and tax withheld; and a balance of assessment consistent with the remainder of the tax return. Counts are for $A1 bins either side of either zero or multiples of $A10, $A100 or $A1000. The latter are defined to be mutually exclusive—that is we exclude the multiples of higher powers of ten when examining the lower powers. Counts are normalized and a line of best fit, with discontinuity, is estimated as in Eq. (2). For visual clarity we exclude the normalized count at zero balance of assessment, which is 3.94. Sample consists of 22.5 million tax returns over the 1991–2018 income years.

v', they receive from an additional dollar from the tax authority and in the shape of their cost curves. In particular, the optimal balance $B^* := B_0 + b^*$ for an individual taxpayer will satisfy the first-order condition:

$$c'(B_0 + b^*) = v'(B_0 + b^*) \tag{4}$$

which can be expanded using exact Taylor expansions and then simplified as follows:

$$\begin{aligned} & c'(B_0) + c''(B_0)b^* = v'(B_0) + v''(B_0)b^* \\ \Rightarrow \quad & 0 + c''b^* = v' + 0b^* \\ \Rightarrow \quad & b^* = \frac{v'}{c''} \end{aligned} \tag{5}$$

using the fact that B_0 minimizing the cost function implies $c'(B_0) = 0$ while the linear value function implies $v'' = 0$.

Hence, from Eq. (5), the extent to which a taxpayer increases the balance owed to them depends on both the utility they derive from the additional dollars, and also the curvature of the cost curve. With a 'flatter' cost curve the taxpayer makes more claims before they exhaust the possibilities for which the marginal benefit exceeds the marginal cost. The claim-by-claim process which results in taxpayers shifting from default to realized outcomes is rarely observed. Yet, as we will show below, many of the same factors also influence taxpayer responses to loss aversion and left-digit bias, resulting in observable bunching in the distribution of balances.

We consider two models of 'behavioral' preferences, with associated value functions $v_r(b)$. The first model applies to the zero threshold (loss aversion), the second to salient, positive round number refund balances (left-digit bias).

For the zero threshold we consider a model of loss aversion where:

$$v_r(b) = v'b + v'\theta_0 b \mathbb{1}[b < 0]$$

for some $\theta_0 > 0$. With this value function, every additional dollar has value $v'(1 + \theta_0)$ while it reduces a loss, falling to v' once the balance becomes a gain.

For the other salient thresholds we consider a value function:

$$v_r(b) = v'b - v'\left(\theta_{10}\text{mod}(b, 10) + \theta_{100}\text{mod}(b, 100) + \theta_{1000}\text{mod}(b, 1000)\right)$$

for some $\theta_{10}, \theta_{100}, \theta_{1000} > 0$ and $b > 0$. With this second model, the value of additional dollars over 10, 100 and 1000 thresholds are discounted by θ_{10}, θ_{100} and θ_{1000} respectively, with these effects cumulative. This is similar to canonical models of left-digit bias; for example, as examined in the context of price discontinuities with respect to mileage in the used car market (Lacetera et al., 2012). In this case, for simplicity and motivated by later findings, the discount rates are based on fixed powers of ten rather than the highest powers in the decimal expansion, i.e., the left digit. We can combine these into a single value function:

$$\begin{aligned} v_r(b) = {} & v'b + v'\theta_0 b \mathbb{1}[b < 0] \\ & - v'\left(\theta_{10}\text{mod}(b, 10) + \theta_{100}\text{mod}(b, 100) + \theta_{1000}\text{mod}(b, 1000)\right)\mathbb{1}[b > 0] \end{aligned} \tag{6}$$

In Fig. 3 we illustrate this value function, along with a benchmark taxpayer, who does not have behavioral preferences, and two illustrative cost curves. As apparent from the figure, the behavioral value function is characterized by a kink at zero and notches at the salient refunds.

We now consider the implications of strictly positive theta for the equilibrium distribution of balances. The kink introduced by loss aversion will result in taxpayers shifting unambiguously to the right, and possibly bunching at zero (e.g., from A to A' or A''), as each marginal dollar of balance below zero now has more value to them. But left-digit bias may lead to taxpayers shifting to the left (e.g., from B to B' or B''), as each marginal dollar of balance between thresholds now has less value to them, or shifting to the right in response to the notch that accompanies each threshold (e.g., from B to B'''). Shifts to the left may involve bunching at the lower threshold, while shifts to the right can only result in bunching. We can characterize how the extent

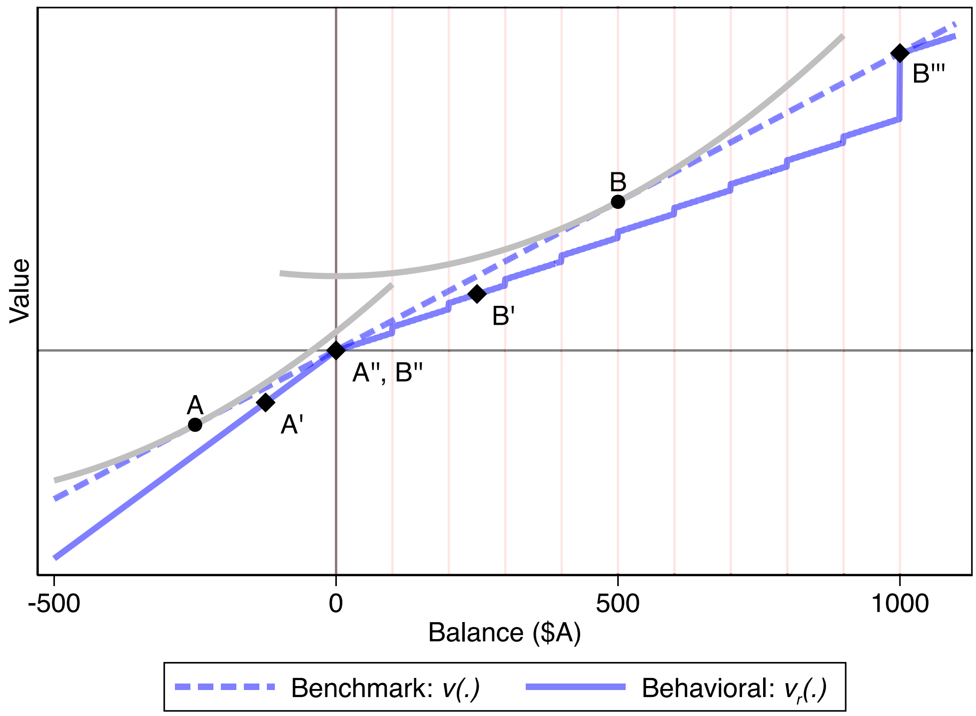

Fig. 3. Value functions with and without behavioral preferences.
Note: Plotted for $v' = 1$, $\theta_0 = 0.4$ and $\theta_{1000} = \theta_{100} = \theta_{10} = 0.2$. The two cost curves are tangent to the benchmark benefit curve at balances of $-\$A250$ and $\$A500$ respectively. The notches at $\$A10$ increments are not visible at this scale.

of bunching at the salient refunds will vary with features of taxpayer preferences.[18]

Proposition 1. *Consider a positive balance which has a largest divisor τ in the set $\{10, 100, 1000\}$ (e.g., for a balance of 200, $\tau = 100$). The mass of taxpayers at this balance is:*

(a) increasing in $\theta_{\tilde{\tau}}$ for $\tilde{\tau} \in \{10, 100, 1000\}$ and $\tilde{\tau} \leq \tau$; and
(b) decreasing in $c''/(v'(1 - \Sigma_{\tilde{\tau}\in\{10,100,1000\},\tilde{\tau}>\tau}\theta_{\tilde{\tau}}))$.

The mass of taxpayers at zero balance is:

(c) increasing in θ_τ for $\tau \in \{0, 10, 100, 1000\}$; and
(d) decreasing in c''/v'.

We leave the proof of this for Appendix A. This result is intuitive from Fig. 3. The more an individual discounts dollars above the thresholds, the more likely they are to bunch at the threshold; further, the more of a premium they place on dollars when they have a negative balance, the more likely they are to bunch at zero ((a) and (c)). Finally, the flatter their cost curve, the more likely they are to bunch (apparent from inspection of Fig. 3). The 'flatness' of the cost curve is determined by the ratio of the second derivative c'' to the benchmark marginal utility of a dollar a balance v', discounted if the threshold is contained within thresholds of higher powers of ten ((b) and (d)). A flat cost curve could come from filing technologies—such as the use of electronic filing or a tax preparer. Another possibility is that c'' comes from the underlying curvature of the utility function. For example, suppose increasing the balance at assessment is purely an exercise in tax evasion that is costless but for a fixed risk of audit p and a penalty that is some large multiple m of the balance claimed B, such that $c(B) = pu(mB)$. In this case a flatter cost curve will reflect less absolute risk aversion. Similar to the (Allingham and Sandmo, 1972) setting, less risk aversion leads to more evasion. Here it is also reflected in more bunching.

These two sources of heterogeneity in bunching differ in important ways in their implications for equilibrium balances. Larger values of θ_{10}, θ_{100} and θ_{1000} will result in more leftward shift between the relevant thresholds, and more bunching at them, but will be bounded in their effect on balances. In contrast, a flatter cost curve will result in not just more bunching but also larger balances in the first instance as shown in Eq. (5). Bunching thus has the potential to capture some of the claim-by-claim behavior (lower declared income and higher claims for deductions) that applies more generally at the point of filing tax returns.

We have abstracted from the role of preparers in our model above, both for tractability and also given the many plausible ways in which they could feature in this setup. As already noted, preparers may play a role in flattening cost curves for clients, through their knowledge but also potentially through a more risk-tolerant approach to tax system compliance. Both these information hub and evasion facilitator roles are envisaged and explored in Battaglini et al. (2019). Given the presence of behavioral preferences, we might expect rational preparers to deliver these positive, salient balances to their clients as part of a profit-maximizing strategy. But it is also possible that there are 'behavioral preparers' that have such preferences themselves. As noted above, these different explanations for heterogeneity in bunching have different implications for the extent to which balances change. In the latter part of this paper we explore heterogeneity in preparer bunching in some detail.

4.2. Correlates of bunching at positive, salient thresholds

We now investigate the correlates of the bunching behaviors identified at the beginning of this section. This serves as a test of some of the competing theoretical explanations for heterogeneity in bunching.

[18] Another approach to explore the implications for the distribution of balances is to simulate outcomes based on assumed distributions of the parameters defining our cost and benefit curves. One such simulation is in Appendix Fig. B.4, which successfully replicates several features of the empirical distribution seen in Fig. 1.

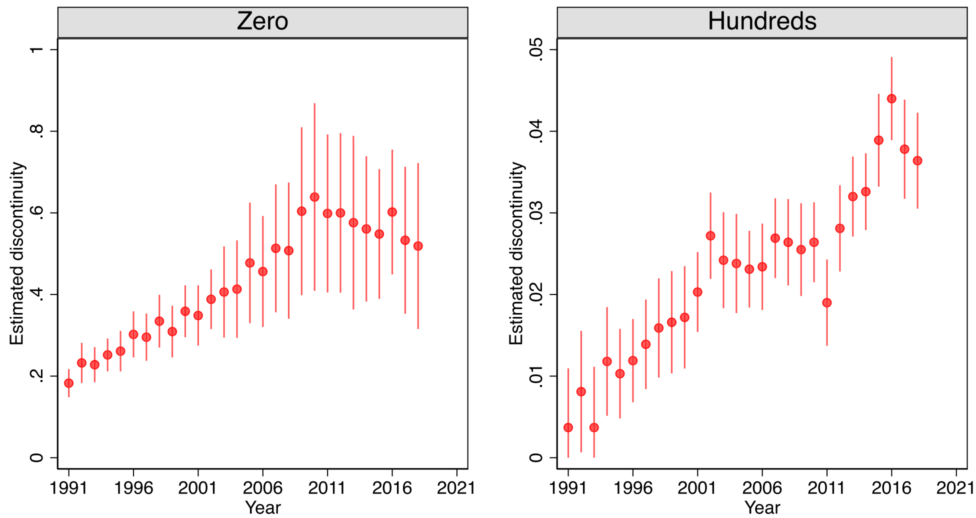

Fig. 4. Estimated discontinuities at zero and hundred-dollar thresholds over time.
Note: Estimated discontinuity δ in the normalized count around the zero and hundred-dollar thresholds over time, with 95% confidence intervals. Based on estimation of Eq. (2) in a window \$A50 either side of the given threshold. For small values of δ an individual is 200δ% more likely to be immediately above the threshold than below it.

In our theoretical framework, differences in bunching propensities can come from either differences in the discount parameters (θ) or differences in the curvature of the cost curves. The latter may reflect both filing technologies, but also fundamental parameters, such as risk aversion.

We begin by noting a dramatic increase in bunching over the past three decades. In Fig. 4 we show the estimated discontinuity at both the zero- and hundred-dollar thresholds over time. Both exhibit a general upward trend, albeit with some plateauing for the zero threshold more recently, and for the hundred-dollar thresholds in the 2000s. In 1991 bunching at zero was modest and bunching at hundred dollar thresholds was negligible, consistent with the relatively modest bunching at zero observed by Rees-Jones (2018) in US tax returns from 1970–1990. By 2018 bunching at zero had more than doubled on this metric, while bunching at hundred-dollar thresholds had emerged.

Such a time trend appears unlikely to be driven by differences in fundamental preferences and risk aversion.[19] To explore both this time trend and heterogeneity in bunching more generally, we could estimate a return-level equivalent of our earlier regressions estimating the discontinuity at various thresholds:

$$100 \times \mathbb{1}[b_i = b_n] = \alpha + \beta b_n + \delta(2 \times \mathbb{1}[b_n > 0] - 1) + \varepsilon_{i,n} \quad (7)$$

where each individual return i contributes 100 observations, subscripted by n, corresponding to balances in our estimation window. This models the event that a return's balance (b_i) is equal to a given balance in the estimation window (b_n) as a linear relationship with a discontinuity at zero. Averaging within each b_n returns us to the earlier Eq. (2). However, this specification allows us to examine the role of individual covariates by interacting them with the constant, slope and discontinuity terms and exploring their interaction with the discontinuity. Multiplying the indicator variable on the left hand side by 100 ensures that the dependent variable averages one, like the normalized count in Eq. (2), and hence that the discontinuities we estimate have the same scaling and can be compared to our earlier results. The downside of this specification is that it is very computationally intensive, as each return contributes 100 observations. Thus, we instead estimate a variant where each return i contributes 4 observations, subscripted by q, corresponding to \$A25 intervals in our estimation window:

$$4 \times \mathbb{1}[b_i = b_q] = \alpha X_i + \beta b_q X_i + \delta(2 \times \mathbb{1}[b_q > 0] - 1)X_i + \varepsilon_{i,q} \quad (8)$$

where the set of covariates X_i includes a constant. This allows us to explore the association between individual-level characteristics and the observed discontinuities at positive and salient balances.

In Tables 2 and 3 we show how the estimated discontinuities δ and their growth over time changes as we expand our set of controls. For ease of computation in the first case we restrict attention to the discontinuities from \$A100 to \$A2500. Beginning with the specifications without any covariates, in columns (1) we see discontinuities of 0.0244 and 0.3754 in the normalized count, in line with those in Fig. 2 and implying increases in the expected count at the thresholds of roughly 5% and 75% respectively. In columns (2) we add a time trend by including years prior to 2018 as a covariate. The constant term now captures the estimated discontinuity in 2018. Consistent with the estimates in Fig. 4, the results imply discontinuities of around 0.0424 and 0.5699 in the normalized count in 2018, that fall away to near-zero and around 0.2 in 1991.

What explains heterogeneity in bunching at positive, salient thresholds? In column (3) we add controls for whether the return was prepared by a tax preparer and lodged electronically, and a variety of demographic controls. Tax preparer returns have significantly larger expected discontinuities—essentially doubling the expected discontinuity at the threshold. This is only modestly attenuated following the addition of location, occupation and individual fixed effects in columns (5), suggesting this is not simply a feature of the types of individuals who use tax preparers or their time-varying observable characteristics. Electronic returns also result in larger discontinuities that are robust to these controls, and also the addition of preparer fixed effects. Tax preparers and electronic returns both facilitate the process of tax filing, and in doing so have the potential to flatten out the cost curve—leading to more bunching, but also potentially higher balances across the board. Demographic factors appear more important in bunching at the hundred-dollar thresholds, with older taxpayers and those working lower-skilled occupations having slightly larger expected discontinuities. These could be either due to these taxpayers being more 'behavioral', that is, more responsive to hundred-dollar thresholds due to larger θ, or having tax affairs with flatter cost curves (e.g., through more manipulation opportunities).

[19] For example, over the four waves of the World Value Survey from the mid 1990s to the late 2010s the proportion of Australians responding that it is never justifiable to cheat on taxes has hovered between 62%–66%; see Haerpfer et al. (2022).

Table 2
Correlates of discontinuities at hundred-dollar thresholds.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.0244***	0.0424***	0.0156***				
	(0.0005)	(0.0010)	(0.0021)				
Years pre-2018		0.0014***	0.0012***	0.0012***	0.0005***	0.0005***	0.0003**
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Preparer			0.0136***	0.0114***	0.0106***		
			(0.0014)	(0.0014)	(0.0020)		
Electronic			0.0108***	0.0119***	0.0117***	0.0152***	0.0128***
			(0.0019)	(0.0019)	(0.0023)	(0.0019)	(0.0023)
Female			−0.0011	−0.0006		−0.0003	
			(0.0010)	(0.0012)		(0.0012)	
Under 25			0.0053***	0.0065***	0.0014	0.0065***	0.0023
			(0.0014)	(0.0014)	(0.0021)	(0.0014)	(0.0022)
Over 65			−0.0005	0.0014	0.0131**	0.0091***	0.0149***
			(0.0035)	(0.0035)	(0.0057)	(0.0035)	(0.0057)
Occupational skill							
2			0.0067***				
			(0.0021)				
3			0.0073***				
			(0.0018)				
4			0.0093***				
			(0.0017)				
5			0.0131***				
			(0.0018)				
Self-employed			−0.0181***	−0.0164***	−0.0169***	−0.0152***	−0.0164***
			(0.0025)	(0.0025)	(0.0032)	(0.0025)	(0.0033)
Fixed effects							
Location				X	X	X	X
Occupation				X	X	X	X
Individual					X		X
Preparer						X	X
R^2	0.0000	0.0001	0.0001	0.0001	0.0844	0.0026	0.0866
Adjusted R^2	0.0000	0.0001	0.0001	0.0001	−0.0109	0.0002	−0.0112
F-stat for additional covariates	1400	547.5	34.46	4.483	0.8842	1.018	0.8847
p-value	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	1.0000
N (million)	56.7	56.7	56.7	56.7	56.7	56.7	56.7

Note: Presents coefficients δ and standard errors from OLS regression estimation of Eq. (8) on the baseline sample. This table examines the hundred-dollar thresholds $\{100, \ldots, 2500\}$. The columns progress through estimating a simple discontinuity (1) to one that varies: over time (2); with return and individual characteristics (3); with location and occupation fixed effects (4); and finally with individual fixed effects (5), preparer fixed effects (6) or both (7). The base case is a man aged 25–64 years old in the highest occupational skill category who self-prepares a paper return. Our location fixed effects use the finest geography available in ALife, `c_sa4_id`, which maps individual resident location to Australian Bureau of Statistics Statistical Area 4 (SA4); large labor markets with populations of typically 300,000 to 500,000 people (around 90 locations). Our occupation fixed effects use the standard occupation variable in ALife, `c_occupation`, which encodes individual occupation using the first edition of the Australian and New Zealand Standard Classification of Occupations (ANZSCO), measured at the two-digit level (around 50 occupations). Occupational skills level range from 5 – skill commensurate with compulsory secondary schooling – through to 1 (the base level) – skill commensurate with bachelor degree or higher. We have separate missing value categories for location and occupation. We also present the F-statistic and associated p-value for the test that the additional covariates in each specification are jointly significant. For specifications (5)–(7) the comparison is relative to specification (4), prior to the addition of individual or preparer fixed effects. *,** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Importantly, the addition of individual fixed effects to the model does not improve our ability to explain bunching beyond what would be expected by chance. The individual fixed effects are not jointly significant (p-value 1.0000) when added to the model with only observable individual and return characteristics (columns 5). In contrast the preparer fixed effects are jointly significant for the hundred-dollar thresholds (p-value 0.0000) (Table 2, column 6). Bunching propensity appears to reflect the choices of tax preparers rather than those of their clients.[20]

The results in Tables 2 and 3 help us to understand the evolution of large discontinuities over time. The time trend falls with the addition of controls, with the most sizeable falls coming from the addition of preparer fixed effects.[21] This suggests a shift over time towards preparers with a greater propensity to bunch. Even so, the time trend remains statistically significant, and hence the propensity of both individuals and preparers to bunch at these thresholds is increasing.

Could increased bunching behavior over time reflect a response to a decreased risk of audit? While the ATO does not provide information on audit intensity, we think that this is unlikely. Technology has made

[20] Another way to illustrate this is to explore persistence in bunching, namely whether being just over the threshold in the prior year is associated with being just over the threshold in the current year. Appendix Table C.1 repeats the specifications in column (3) but with added controls for having a prior-year balance in the window around the given threshold, the continuous value of that balance, and an indicator equal to one if it is above the given threshold. We show the coefficient on the latter variable, which can be interpreted as the effect on the discontinuity at the threshold of bunching in the prior year. We see a strong positive effect in both cases, though the gain in explanatory power as captured by the R^2 is negligible. In both cases adding preparer fixed effects leads to a substantial attenuation of the coefficient. Persistence in bunching behavior largely reflects persistence in preparer, and differences in preparer tendencies to bunch. In the sections to come we return to the role of tax preparers in explaining bunching, and what we can learn from it.

[21] This reflects the fact that the time trends in the more important controls have been relatively modest (Table 1). Tax preparer returns have been relatively stable, rising by 6 percentage points from 65% of returns in 1991 to 71% of returns in 2018. Electronically lodged returns have grown more markedly, by nearly 30 percentage points, from 69% of returns in 1991 to 96% of returns in 2018. Nevertheless, the coefficients in Tables 2 and 3 combined with this increase in electronic returns imply only a modest contribution to the increased bunching propensity.

Table 3
Correlates of discontinuities at zero threshold.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.3754***	0.5699***	0.2880***				
	(0.0025)	(0.0051)	(0.0095)				
Years pre-2018		0.0136***	0.0110***	0.0112***	0.0111***	0.0082***	0.0073***
		(0.0003)	(0.0004)	(0.0004)	(0.0012)	(0.0005)	(0.0014)
Preparer			0.3131***	0.3044***	0.2586***		
			(0.0070)	(0.0071)	(0.0193)		
Electronic			0.0715***	0.0787***	0.0570***	0.1024***	0.0707***
			(0.0083)	(0.0083)	(0.0192)	(0.0087)	(0.0209)
Female			−0.0113**	−0.0050		−0.0077	
			(0.0050)	(0.0055)		(0.0056)	
Under 25			0.0309***	0.0283***	0.0109	0.0360***	0.0111
			(0.0070)	(0.0071)	(0.0231)	(0.0073)	(0.0258)
Over 65			−0.0614***	−0.0575***	0.0522	−0.0300**	0.0792
			(0.0130)	(0.0130)	(0.0481)	(0.0136)	(0.0547)
Occupational skill							
2			−0.0019				
			(0.0113)				
3			0.0247**				
			(0.0096)				
4			0.0079				
			(0.0087)				
5			−0.0071				
			(0.0099)				
Self-employed			0.0299***	0.0310***	−0.0346	0.0085	−0.0447
			(0.0106)	(0.0106)	(0.0347)	(0.0112)	(0.0408)
Fixed effects							
Location				X	X	X	X
Occupation				X	X	X	X
Individual					X		X
Preparer						X	X
R^2	0.0332	0.0340	0.0366	0.0373	0.5059	0.0742	0.5269
Adjusted R^2	0.0332	0.0340	0.0366	0.0371	−0.0926	0.0370	−0.1314
F-stat for additional covariates	39,900	667.4	189.0	3.868	0.7833	0.9971	0.7440
p-value	0.0000	0.0000	0.0000	0.0000	1.0000	0.7281	1.0000
N (million)	2.32	2.32	2.32	2.32	2.32	2.32	2.32

Note: Presents coefficients δ and standard errors from OLS regression estimation of Eq. (8) on the baseline sample. This table examines the zero-dollar threshold. The columns progress through estimating a simple discontinuity (1) to one that varies: over time (2); with return and individual characteristics (3); with location and occupation fixed effects (4); and finally with individual fixed effects (5), preparer fixed effects (6) or both (7). The base case is a man aged 25–64 years old in the highest occupational skill category who self-prepares a paper return. Our location fixed effects use the finest geography available in ALife, `c_sa4_id`, which maps individual resident location to Australian Bureau of Statistics Statistical Area 4 (SA4); large labor markets with populations of typically 300,000 to 500,000 people (around 90 locations). Our occupation fixed effects use the standard occupation variable in ALife, `c_occupation`, which encodes individual occupation using the first edition of the Australian and New Zealand Standard Classification of Occupations (ANZSCO), measured at the two-digit level (around 50 occupations). Occupational skills level range from 5 – skill commensurate with compulsory secondary schooling – through to 1 (the base level) – skill commensurate with bachelor degree or higher. We have separate missing value categories for location and occupation. We also present the F-statistic and associated p-value for the test that the additional covariates in each specification are jointly significant. For specifications (5)–(7) the comparison is relative to specification (4), prior to the addition of individual or preparer fixed effects. *,** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

auditing less costly and more sophisticated over time, so it seems more likely that auditing activity has increased rather than decreased. The literature (Alm et al., 1992; Dhami and al-Nowaihi, 2007) suggests that the returns to evasion are very high and yet most taxpayers do not engage in evasive behavior, in violation of standard expected utility theory.

5. Is bunching driven by effort or evasion?

Taxpayers bunching at these salient thresholds may arrive there in one of two ways. First, they (or their preparers) may simply put more effort into the tax return—making legitimate claims that they would otherwise not have made. For example, this effort could be directed towards better understanding the tax law, or maintaining and consulting records. Second, they (or their preparers) may take on more risk—making claims to which they may not be entitled. We refer to these as the ‘effort’ and ‘evasion’ channels.

We explore these two channels using the audit data from the Random Enquiry Program (REP) described above. Audits can result in the final balance of the tax return staying the same or being adjusted in either a positive or a negative direction. If we observe bunching in tax return balances for all tax returns before but not after audit adjustments, this would suggest that bunching is driven by illegitimate claims and thus a role for evasion. Audits are designed to uncover and unwind the effect of evasion, but we consider it highly unlikely that they would erode differences in effort. For example, in the REP sample, 54% of returns see work-related expense deduction claims fall following audit while less than 3% of returns see such claims rise. For those returns where balances are adjusted downwards (aggregate ‘evaders’), if we again observe bunching pre-audit but no bunching post-audit, this would be suggestive of bunching being driven by evasion. If, further, we cannot reject no bunching pre-audit for those tax returns for which balances are either not adjusted or adjusted upwards (aggregate ‘non-evaders’), this would provide additional evidence that bunching behavior is driven by evasion and not by effort. These three patterns are exactly what we observe in the data.

To test for the presence of these channels we look for discontinuities in the density of the balance before and after audit using the REP data. We use the test developed by Cattaneo et al. (2020) based on local polynomial density estimators and implemented in the Stata command `rddensity` (Cattaneo et al., 2018). We use a local linear estimate of the density function, with a triangular kernel, and bandwidths on either

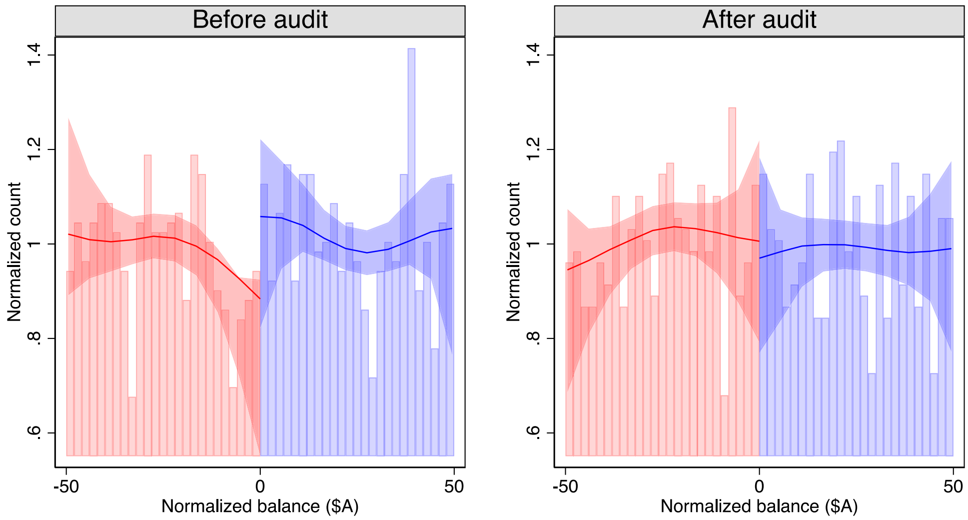

Fig. 5. Distribution of balance of assessment around hundred-dollar thresholds, Random Enquiry Program returns.
Note: Distribution of the balance of assessment for those in the Random Enquiry Program. Counts are normalized, and are for $A1 bins either side of positive multiples of $A100. Sample consists of 2,439 (left panel) and 2,133 (right panel) tax returns over the 2016–2020 income years.

side chosen to minimize the mean squared error of the given density estimator. As noted earlier, our REP data extract does not have all the information required to remove mechanical discontinuities around the zero threshold, so we focus on the hundred-dollar thresholds.

Fig. 5 shows the distribution of the balance of assessment around hundred-dollar thresholds before and after auditing for returns in the REP.[22] The lines and shading indicate the density estimator and bias-corrected confidence intervals. There is modest visual evidence for a discontinuity before audit, but this is not apparent after audit, where the distribution is relatively uniform. The magnitude of the discontinuity is broadly in line with that much more precisely observed in the population data.[23]

In Table 4 we present the results from the formal test for a discontinuity in the density function. Looking at the full sample (Panel A), we can reject the null hypothesis of continuity around positive hundred-dollar thresholds in the before-audit balance at the 5% level (with a p-value of 0.05), but not for the after-audit balance (with a p-value of 0.85). If we restrict the sample to those whose balance is smaller after audit (Panel B) – found to have been evading taxes in aggregate – the discontinuity before audit remains. In contrast, those whose balance stays the same or is larger after audit – an aggregate 'non-evading' sample – do not have a significant discontinuity in balance before or after audit.[24]

This provides strong evidence that the bunching we observe in refund balances is driven more by evasion than by effort. The bunching is caused not by more rigorous application of legitimate deductions but rather by pushing up to and past the boundaries of acceptable (to the tax authorities) income and deduction claims.

6. Why are salient refunds delivered by tax preparers?

The prominent role of tax preparers in delivering positive, salient tax refunds motivates further investigation. What do preparers get out of it? We begin by considering two channels through which preparers may benefit—namely that the clients receiving these refunds may either willingly pay higher fees or be more loyal. Both would have value to profit-maximizing tax preparers.

6.1. Empirical framework

To examine the effect of receiving a positive, salient refund on tax preparer fees and client loyalty we look for evidence of discontinuities in related outcomes around the thresholds in question. We look at two outcomes in particular: the individual's deduction for the cost of managing tax affairs in the following year, as a proxy for the tax preparer fee; and a binary variable taking the value of 100 if the individual was with the same preparer the following year and zero otherwise (the per cent probability of remaining with your tax preparer).

To estimate the discontinuities around the relevant thresholds we estimate the following equation:

$$y_{ijt} = \alpha + \beta_0 b_{ijt} + \sum_{\tau \in T} \beta_\tau \max\{0, b_{ijt} - \tau\} + \gamma \sum_{\tau \in T} \mathbb{1}[b_{ijt} \geq \tau] + \eta X_{ijt} + \varepsilon_{ijt} \quad (9)$$

where T is a set of thresholds in question and the subscripts reflect returns for individual i filed through preparer j in year t. This equation models the outcome of interest y_{ijt} as piecewise linear in the balance b_{ijt}. The slope is allowed to vary beyond each threshold τ (giving slopes coefficients β_0 and β_τ for all $\tau \in T$), but with a fixed discontinuity γ at each threshold. We focus on two cases: the zero threshold ($T = \{0\}$);

[22] For consistency we multiply the density by 100 to express results in terms of the normalized count introduced earlier.

[23] For this comparison we restrict our population tax data to the 2016–2018 income years and consider all positive hundred-dollar thresholds. Then the estimated discontinuities using the parametric approach introduced above are 0.064 (s.e. 0.041) in the REP data and 0.039 (s.e. 0.002) in the equivalent population data. A higher discontinuity in the REP data is expected given its skew towards those with more complicated tax affairs and hence more scope for tax evasion.

[24] One concern here might be the power of our test in this last instance and the fact that the point estimate for the discontinuity is still positive and of meaningful size. Aside from chance, one possibility that a positive, albeit statistically insignificant discontinuity remains is that even where a taxpayer was not evading taxes in aggregate, they may still have been making illegitimate claims. For example, a taxpayer may claim illegitimate deductions even where legitimate deductions have already reduced their net tax liability to zero. If we restrict our 'non-evading in aggregate' sample to also require that total work-related expense deductions stay the same or are larger after audit then the estimated discontinuity is 0.05 (with p-value of 0.39) before audit and −0.04 (with p-value of 0.24) after audit. When we adopt this tighter definition of non-evasion, that excludes those whose work-related expense deductions decrease post-audit, we find a smaller pre-audit discontinuity that is still statistically insignificant. This supports our conclusion that refund bunching is related to evasion.

Table 4
Tests for discontinuity in normalized count at hundred-dollar thresholds.

	Discontinuity	p-value	Bandwidth ($A)		Sample size	
			Left	Right	Left	Right
Panel A: Full sample						
Before audit	0.17	0.05	21.5	21.5	508	562
After audit	−0.04	0.85	21.5	21.5	479	477
Panel B: Sample with smaller post-audit balance ('Evading sample')						
Before audit	0.19	0.06	21.5	21.5	373	417
After audit	−0.03	0.90	39.1	21.5	613	348
Panel C: Sample with same or larger post-audit balance ('Non-evading sample')						
Before audit	0.15	0.35	21.5	21.5	135	145
After audit	−0.03	0.34	21.5	21.5	140	129

Note: Presents the results from tests for a discontinuity in the density of tax return balances at hundred-dollar thresholds. We use tax returns in the ATO's Random Enquiry Program through the 2016 to 2019 income years, and present results for the balance before and after audit, and for the full sample (Panel A) and those for whom the audited balance is smaller (Panel B) or the same or larger (Panel C) than the balance before audit. In all cases, attention is restricted to balances greater than $A50, and converted to dollars either side of the nearest hundred-dollar threshold. We present the discontinuity in point estimates, bias-corrected p-values, and the optimal bandwidths and effective sample sizes either side of the threshold. The test is that developed by Cattaneo et al. (2020) based on local polynomial density estimators and implemented in the Stata command `rddensity` (Cattaneo et al., 2018). We use a local linear estimate of the density function, with a triangular kernel, and bandwidths on either side chosen to minimize the mean squared error of the given density estimator.

and the hundred dollar thresholds up to $A2500 ($T = \{200, \dots, 2500\}$). We estimate this equation for balances within $A100 of the range of thresholds considered. We also allow for a variety of potentially time-varying individual- and preparer-level controls in X_{ijt}, which we will describe alongside the results.

This approach is not a regression discontinuity design: the clear manipulation around positive and salient thresholds that we have identified invalidates such a design. Rather, it is a descriptive exercise about behavior either side of the threshold. If we are willing to go further and assume that with the inclusion of controls the conditional expectation of the error term is zero, then our results will have a causal interpretation.

6.2. Results

In Table 5 we present the estimated discontinuities in our fee proxy (Panel A) and client loyalty (Panel B) at the zero- and hundred-dollar thresholds. As we move from columns (1) to (4) and columns (5) to (8) we add more controls. We begin with year fixed effects; progress to controlling for an individual's tenure with a tax preparer, occupation and location (all interacted with year); then add preparer-year fixed effects; and finally add individual fixed effects. The last two are particularly demanding, with both relying on variation in returns prepared by the same preparer in a given year.

Tax returns just over positive, salient thresholds are not associated with higher fees. Panel A of Table 5 shows no evidence of a positive discontinuity in fees over these thresholds. These are fairly precise null estimates—based on columns (4) and (8) we can reject at the 5% level that the true effect is greater than $A10 around the zero threshold and $A0.50 at the hundred thresholds.

In contrast, tax returns just over positive, salient thresholds are associated with greater loyalty. Panel B of Table 5 shows evidence of positive discontinuities of varying robustness. For the zero threshold, the discontinuity is initially an 0.77 percentage point increase in loyalty; this falls to 0.49 percentage points with the addition of further controls but falls further and loses significance with preparer-year fixed effects. One challenge here is that only a very small number of a preparer's returns will fall within our estimation window in a given year. For the hundred-dollar thresholds we have much larger sample sizes and a more robust increase in loyalty, which is apparent from columns (6)-(8) and ranges from 0.11–0.14 percentage points.

These results are consistent with preparers delivering positive, salient refunds in response to their clients' preferences. The extra work or risk borne to deliver such refunds may be justified by client preferences. It is possible that client satisfaction manifests itself beyond the effects on loyalty we have indicated here. For example, satisfaction may lead to word-of-mouth advertising or allow preparers to make savings elsewhere. Further evidence that preparers are responding to client preferences can be found in the fact that self-prepared returns also exhibit bunching at these thresholds. What is less clear is whether clients are aware that these salient refunds are typically driven through claims that do not survive audit. Sorting on ethical and risk preferences is also quite plausible, and has been seen in the context of financial auditors and their clients (Cook et al., 2020).

Preparers do, however, go well beyond self-preparers in their tendency to settle on positive, salient refunds. Appendix Fig. B.5 shows the (client-weighted) distribution of discontinuities around the zero and hundred dollar thresholds. There is a fat right tail of clients with preparers who are several times more likely than self-preparers to be landing returns that are just over rather than just below the relevant thresholds. While it may be that these tax preparers are rationally maximizing profits, another possibility is that they are themselves 'behavioral' in the sense that they derive utility from delivering particular refunds. Some speculative evidence for this can be found in the distribution of tax preparer fees, which typically end in a '0'. This could reflect preparer preferences for round numbers. Appendix Fig. B.6 shows that high-bunching preparers are more likely to have fees ending in a '0'.

7. How do high-bunching preparers influence tax returns?

We have argued that tax preparers deliver positive, salient refunds partly in response to behavioral preferences held by their clients. However, a preparer's propensity to bunch in response to these preferences will vary with the shape of their underlying cost curve: those with flatter cost curves will bunch more. As shown in Section 4.1, this would suggest that high-bunching preparers should have an impact on returns beyond that required to bunch—they should result in higher claims and balances for their clients more generally.

7.1. Empirical framework

To examine the effect of high-bunching tax preparers on tax returns we use a standard movers design, in this case an event study looking at individuals who move between tax preparers. Recent examples of such designs include their use in the estimation of causal neighborhood effects in the intergenerational mobility literature, by looking at children

Table 5
Effect of a salient refund on fee proxy and loyalty.

	Zero				Hundreds			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Discontinuity in fee proxy ($A)								
	0.26	1.36	0.80	−1.56	−0.15	0.70	0.10	−0.48
	(2.92)	(2.92)	(3.27)	(6.71)	(0.70)	(0.70)	(0.67)	(0.55)
N (million)	0.42	0.42	0.35	0.10	6.50	6.50	6.45	6.22
Panel B: Discontinuity in client loyalty								
	0.77***	0.49**	0.09	−0.34	0.02	0.12**	0.14***	0.11**
	(0.24)	(0.23)	(0.27)	(0.61)	(0.05)	(0.05)	(0.05)	(0.05)
N (million)	0.67	0.67	0.57	0.22	9.28	9.28	9.20	8.94
Fixed effects								
Year	X				X			
Tenure-year		X	X	X		X	X	X
Occupation-year		X	X	X		X	X	X
Location-year		X	X	X		X	X	X
Preparer-year			X	X			X	X
Individual				X				X

Note: Presents coefficient estimates γ and standard errors from OLS regression estimation of Eq. (9) on the baseline sample, restricted to tax preparer returns. This estimates the relationship between the outcome of interest and the balance as a piecewise linear function with a discontinuity at the relevant threshold(s). Columns (1)–(5) examine the zero-dollar threshold (with range $[-100, 99]$), while columns (6)–(10) examine the hundred-dollar thresholds (with range $[100, 2599]$). The columns progress from a specification with only year fixed effects through to also allowing for: tenure-year, where tenure is the number of years filing with the tax preparer, occupation-year and location-year fixed effects; preparer-year fixed effects; and individual fixed effects. See the note to Tables 2 and 3 for more information on the occupation and location covariates. *,** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

who moved neighborhoods while growing up (Chetty and Hendren, 2018; Deutscher, 2020). While the identifying assumptions are strong, we conduct a range of standard tests to support a causal interpretation of our results.

We return to the full ALife sample for the movers design, applying our earlier restrictions only when examining bunching as an outcome. We then restrict attention to individuals who are observed for at least four years with both the old and new preparers. To avoid capturing restructures that involve a change in tax preparer identifier, we exclude moves where the destination preparer receives more than half their clients from, or more than half the clients of, the origin preparer.[25] To abstract from moves associated with labor market entry or exit, we further focus on individuals with nonzero salary and wage income in each of the four years before and after the move. The treatment is D_i, the difference in the bunching discontinuity between the new and the old preparer. We exclude those moves where the difference in bunching propensities is imprecisely estimated, losing about 15% of moves in the process.[26]

We estimate the effect of moving between tax preparers who differ in their bunching propensity through the following equation:

$$y_{it} = \alpha_i + \beta_t + \gamma_l + \delta D_i + \sum_{l \in \{-4,\ldots,3\}, l \neq -1} \zeta_l D_i \mathbb{1}_{t=T-l} + \eta X_{it} + \varepsilon_{it} \tag{10}$$

for individual i in year t and event time l (which is equal to zero in the first year with the new preparer). We examine a variety of outcome variables including the balance, deductions claimed and income reported. Our baseline specification includes individual, year and event-time fixed effects, as well as age fixed effects in the time-varying controls X_{it}.

The coefficients of interest in Eq. (10) are the ζ_l. These can be interpreted as the effect of moving between a preparer that never bunches (with no discontinuity at the threshold) to a preparer where the discontinuity is consistent with always bunching. In some instances, it will also be instructive to scale the estimates such that they capture the effect of moving between preparers that differ by in bunching propensity by amounts more consistent with that seen in the population as a whole.

Eq. (10) allows us to examine the effect of moving to a higher-bunching preparer on various dollar outcomes. To examine the effect on individual bunching propensity, we can return to Eq. (8), and allow the discontinuity to vary with the covariates captured above. Namely, we allow for individual, year, event time, age and treatment variation in the discontinuity in probability of a tax return balance being over a given threshold, and examine the interaction of event time and treatment. When examining the zero threshold we drop individual fixed effects as few people have a balance close to zero in successive years.

The key identifying assumption in these designs is that the difference in bunching propensity between preparers D_i is uncorrelated with other factors affecting tax return outcomes, conditional on our controls. That is, we assume strict exogeneity. This is a strong assumption, as it seems likely that moves between tax preparers at times reflect changes in personal circumstances – such as a change in occupation – that may influence both choice of preparer and return outcomes. Given this we examine the robustness of our results to the inclusion of time-varying controls. Furthermore, we examine how some of our headline results change as we restrict attention to different types of moves: those to or from higher-bunching preparers; and those that are more plausibly exogenous. Similar tests can be found in other work using a movers design (such as Chetty and Hendren, 2018; Deutscher, 2020).[27]

On a final note, recent literature has highlighted potential shortcomings in two-way fixed effects estimation of difference-in-differences and

[25] Leaving these 'false moves' in the data would create even larger bias in the robustness exercise where we examine moves that occur alongside large outflows from particular preparers.

[26] In particular, we drop those moves where the estimated standard error on the treatment is more than 0.06 (when looking at bunching at the hundred-dollar thresholds) or more than 0.30 (when looking at bunching at zero). Appendix Fig. B.7 motivates this by showing the cumulative distribution of these standard errors—these choices exclude the tail of imprecisely estimated treatments.

[27] Another potential concern is the inclusion of the movers in the estimation of the preparer bunching propensity—from which we get our treatment variable. However, estimating the bunching propensities using the 90% of taxpayers not included in the ALife sample does not meaningfully change our results. We see similar, large changes in the balance and work-related expense deductions and the bunching behaviors of movers once again mimic those of the other clients of the destination preparer. This robustness is unsurprising as the movers in our sample are a very small proportion of the full population of taxpayers.

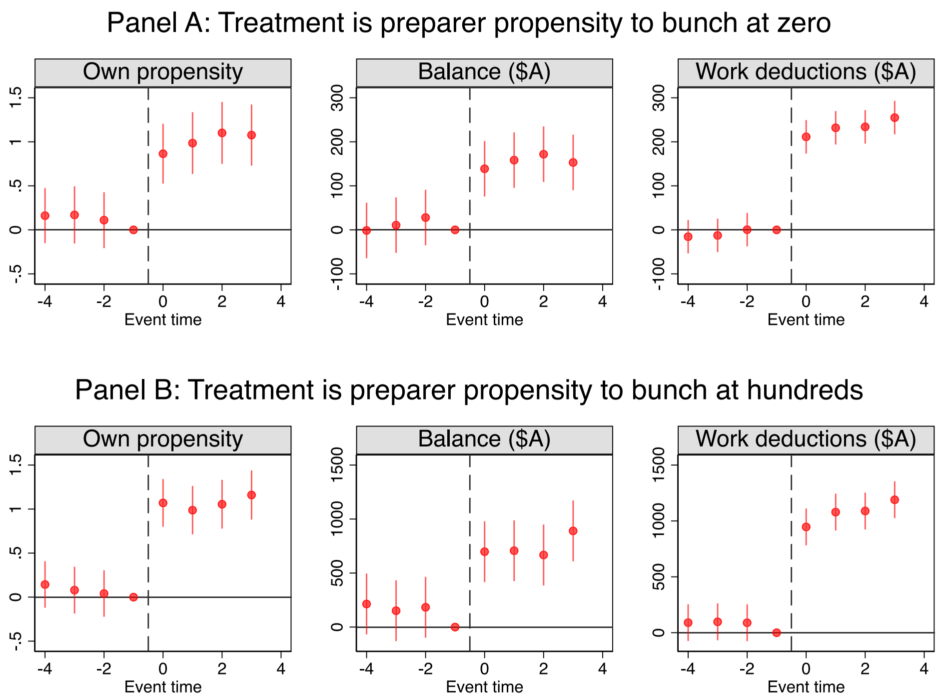

Fig. 6. Effect of moving between preparers differing in propensity to bunch.
Note: Presents coefficient estimates and 95% confidence intervals for ζ_l from Eq. (10), capturing the effect of a one-unit change in preparer bunching propensity at time $l = 0$ on individual bunching propensity, balance at assessment and work-related expense deductions. The first post-event year coefficient and standard error, sample size and R^2 are available in Table 6.

event study designs. The most relevant paper in this instance is Callaway et al. (2021), who consider a difference-in-differences setting with a continuous treatment variable such as this one.[28] They show that with a strong parallel trends assumption, weaker than strict exogeneity, the two-way fixed effects estimator is a weighted average of the average causal responses to treatment but with weights that place more weight on those treatments nearer the mean. In our case, over-weighting the responses to modest differences in the bunching propensity seems less problematic than it otherwise might, as our intent is to establish a link between bunching behaviors and broader tax return outcomes rather than estimate the treatment effect of a particular policy setting *per se*.

7.2. Effects of moving between tax preparers

Fig. 6 shows the effect of a one-unit change in preparer bunching propensity on individual bunching propensity, balance at assessment and the largest group of deductions—work-related expense deductions. The top panel illustrates the effect of moving between preparers that differ in their propensity to bunch at zero, and the bottom panel illustrates the effect of moving between preparers that differ in their propensity to bunch at hundred-dollar thresholds.

The first thing to note is that when the outcome is individual bunching propensity the coefficients after the move are near one. There is no uptick in bunching prior to moving and individuals appear to pick up the full difference in bunching propensity between their preparers. This provides further evidence that most of the differences across preparers are due to preparer behavior rather than the bunching propensities of their clients.

When we turn to look at dollar outcomes, there is a large increase in the balance and total work-related expense deductions, and again no evidence of pre-trends. Further, the increase in balances is much more than that required to generate the increase in bunching. When examining bunching in a $A100 window, an increase in balance of at most $A50 is required to shift individuals from one side of a salient threshold to another. Yet here we observe increases in the balance that are several times larger—an increase of $A698 or $A139 in the first year for bunching around the hundred- and zero-dollar thresholds respectively (Table 6, Panel B). Also notable is a lack of evidence for dynamic effects. This is consistent with the tax preparer influencing decisions made at the point of filing but not through the year. In the first year with the tax preparer, the client will be visiting with the tax year behind them and decisions about income earned, expenses incurred and record keeping already made. If tax preparer advice about the latter factors mattered we might expect to see further rises in the following years, but this upward drift is modest at best. It appears the variation in returns that high-bunching preparers influence is predominantly about what happens when returns are filed.

The dollar amounts in play are also substantial in the context of the bunching we observe in the population as a whole. Moving to a preparer that does not bunch to one that bunches at the same rate as observed across our full sample would result in an increase in the balance owing to the taxpayer of $A17 when considering hundred-dollar thresholds and $A52 when considering the zero threshold.[29] Across a population of 15 million taxpayers, this implies an annual fiscal cost of $A260-780 million, over two orders of magnitude greater than the $1 million mechanical costs of bunching estimated earlier.

For a finer-grained look at the results we turn to Table 6. Given the lack of evidence for dynamic effects, we present the estimate for the first year with the new preparer. As a test for any pre-trends, we also provide the p-value on the Wald test that the coefficients prior to the move are jointly equal to zero. Comfortingly, none of these are below typical thresholds.

[28] In particular, as in their set-up, we have framed all movers as initially untreated, with treatment being the difference in bunching intensity.

[29] Scaling down the $A698 and $A139 referenced earlier based on population discontinuities of 0.0244 and 0.3754 at the hundred- and zero-dollar thresholds respectively (see Tables 2 and 3).

Table 6
Effect of preparer bunching propensity on tax return.

	Zero				Hundreds			
	Coef.	p	R^2	N (million)	Coef.	p	R^2	N (million)
Panel A: Discontinuity in distribution of balances								
At zero	0.86***	0.70	0.06	0.05	1.59**	0.34	0.06	0.06
	(0.17)				(0.73)			
At hundreds	0.06**	0.59	0.12	1.47	1.07***	0.75	0.12	1.81
	(0.03)				(0.14)			
Panel B: Dollar balance, deductions and income								
Balance	139***	0.78	0.34	0.81	698***	0.46	0.34	1.02
	(32)				(144)			
Work-related expenses								
Total	211***	0.77	0.68	0.80	946***	0.61	0.68	1.00
	(19)				(84)			
Car	109***	0.84	0.64	0.80	427***	0.53	0.65	1.00
	(12)				(52)			
Other	57***	0.43	0.66	0.80	335***	0.18	0.66	1.00
	(7)				(30)			
Clothing	16***	0.28	0.61	0.80	80***	1.00	0.61	1.00
	(1)				(5)			
Travel	11***	0.70	0.51	0.80	−12	0.95	0.51	1.00
	(4)				(17)			
Self-education	8***	0.79	0.36	0.80	86***	0.78	0.36	1.00
	(3)				(12)			
Tax affairs	15***	0.17	0.43	0.54	65***	0.63	0.44	0.71
	(3)				(12)			
Gifts	7***	0.85	0.64	0.78	60***	0.91	0.65	0.99
	(2)				(7)			
Rental expenses								
Interest	25	0.15	0.66	0.78	−224	0.15	0.66	0.99
	(35)				(155)			
Other	54***	0.44	0.69	0.78	−3	0.25	0.70	0.99
	(20)				(87)			
Income - total	−45	0.80	0.80	0.81	−2,444***	0.43	0.80	1.02
	(203)				(901)			
Income - wages	143	0.82	0.82	0.81	6	0.85	0.82	1.02
	(164)				(724)			
Income - p'ship/trust	−20	0.85	0.56	0.63	−596**	0.77	0.57	0.82
	(59)				(260)			

Note: Presents coefficient estimate ζ_0 and standard errors from OLS regression of Eq. (8) (bunching outcomes) or Eq. (10) (all other outcomes). These equations are standard event study designs that allow for individual, year, event time and age fixed effects, and where the interaction of the treatment – the difference in bunching propensity between preparers – and event time is the key variable of interest. We show the coefficient for the first year with the new preparer. This can be interpreted as the effect of moving from a preparer that never bunches to one that always bunches. The columns subtitled 'p' present the p-value on the Wald test that the coefficients prior to the move are jointly equal to zero. *,** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

As seen earlier, movers tend to pick up the full difference in bunching propensity, with coefficients of near one indicating that the change in their bunching propensity reflects their new preparer's propensity to bunch. However, preparers that are much more likely to bunch at zero are only modestly more likely to deliver returns over hundred dollar thresholds for their clients. Conversely, while imprecisely estimated, preparers much more likely to bunch at hundred dollar thresholds have an even larger effect on bunching at zero. Bunching at hundreds tends to imply bunching at zero but not vice versa.

More substantively, the parts of the tax return that are most influenced by high-bunching preparers are those where there is perhaps more 'discretion' when filing a tax return. Work-related expenses (WRE) deductions are one such area given the challenges in codifying, understanding and auditing the required connection between such expenses and income earning activities. Random audit programs by the ATO suggest that 44% of the net tax gap (the difference in tax owing and tax paid) among individuals not in business is due to WRE claims (see Australian Taxation Office, 2021a). These audit programs also highlight car, other and clothing claims as the most frequently adjusted WRE claims, which aligns with the large and precisely estimated effects apparent in Table 6.

Turning to income variables, these effects are less precisely estimated, though there does appear to be a fall in total income on moving to preparers more likely to bunch at the hundred dollar thresholds. Wage and salary income is unaffected, perhaps unsurprising given it is subject to third-party reporting and typically pre-filled in electronic tax returns. We do, however, see a fall in reported income from partnerships and trusts, both commonly used by small business entities. Once again, this is consistent with ATO random enquiries, which have suggested that omitted income constitutes 71% of the tax gap for individuals in business.

More generally, the consistency between our quasi-experimental approach and the audit findings again suggests that at least some of this manipulation at the point of tax filing is evasion, rather than the result of more diligent preparers. As noted in Section 4.1, a flat cost curve is consistent with a range of interpretations, which may include more thorough or efficient preparers facing a lower cost of finding additional legitimate claims, but also more risk-tolerant preparers being more willing to make claims that may not hold up in the event of an audit.

7.3. Robustness exercises

As noted earlier, a potential concern with the event study design is that moves between preparers happen for a reason. In this section we explore if and how our results change with the inclusion of covariates, when considering moves to higher or lower bunching preparers, and when looking at more plausibly exogenous moves. We focus on the effects observed for the balance and total WRE claims—the first is what matters for the ultimate fiscal outcome, while the latter is the most precisely estimated effect among specific return items.

Including covariates has a negligible effect on our key headline findings. In Appendix Fig. B.8 we replicate the last two panels of Fig. 6, comparing the baseline results with those where we also include fixed effects for occupation and location, and the natural logarithm of wage and salary income, among our control variables. While the estimated effects are a little lower, the differences between the two series are nearly imperceptible, which suggests observable changes for individuals do not drive the results.

Another approach to assessing the robustness of our results is to check whether they hold for particular subsets of moves. These are fairly demanding tests so to improve power we switch to a specification that replaces the treatment's interaction with event time with its interaction with a simple indicator variable that equals one following the move to the new preparer, namely we estimate:

$$y_{it} = \alpha_i + \beta_t + \gamma_l + \delta D_i + \zeta D_i \mathbb{1}_{t\geq 0} + \eta X_{it} + \varepsilon_{it} \tag{11}$$

where ζ is now the variable of interest.

In Appendix Table C.3 we show that the relationship between the change in preparer bunching propensity and changes in key outcome variables is relatively symmetric: it is not purely driven by moves to higher or lower bunching preparers. For all the threshold and outcome variable combinations we see both positive and negative moves resulting in effects. The results are a little complicated by the loss of power associated with exploiting variation within various bands of the treatment variable (rather than also between these bands). While we cannot reject equality of the treatment effects for the more modest half of positive and negative moves, we can typically reject it across the full range of moves from large negative to large positive moves. This nonlinearity could be consistent with learning effects, whereby individuals going to a high-bunching preparer are more likely to pick up some tax filing behaviors than they are to lose them when leaving such a preparer; it could also reflect differences in the nature of such moves. The comfort from this exercise is that any omitted variable driving both moves and tax return outcomes would need to operate for both moves to and from high-bunching preparers.

As a final exercise, we explore how our results change as we hone in on people leaving their tax preparer at the same time that many other clients are leaving the same preparer. These people are more likely to be leaving because of the retirement of a particular preparer in the practice or practice closure, rather than because of a change in their particular circumstances. This is not dissimilar to studies of the effects of job loss that seek to exploit mass layoffs or firm closure. Appendix Fig. B.9 explores this and plots the resulting coefficients. There is only modest attenuation of the estimated treatment effects as we move from moves that happen amid typical outflows (30% of clients leaving a preparer) to those near closure (100% of clients leaving a preparer). This provides some comfort that moves driven by a purposeful decision to change preparers do not drive our results.

8. Conclusion

In this paper we have shown that Australian taxpayers have a clear preference for round-number tax refunds and that their refunds bunch at positive and salient thresholds. Using random audit data, we have shown that this bunching is driven by tax evasion. Finally, we have explored the role of preparers. Taxpayers who use tax preparers are twice as likely to bunch as those who do not. High-bunching tax preparers assist people to bunch by lowering reported income, increasing deductions and, in the process, generating higher refunds. These tax preparers target income and deductions that are difficult for the tax authorities to audit. The main effect of such tax preparers is on behavior at the point of tax filing and does not influence the future behavior of taxpayers. These observations are consistent with tax preparers helping clients to evade taxes rather than helping them to manage their tax affairs in a legitimate way to reduce their taxable income. Our paper has an important policy implication: tax authorities might want to target round-number refunds and "high-bunching" tax preparers in compliance activity, while these patterns hold true.

Bunching behavior has increased dramatically over time. Individuals who self-prepare are bunching more than in the past as are tax preparers. Bunching also increases over time as more people move to "high-bunching" tax preparers. "High-bunching" preparers do not charge higher fees, but they have stronger client retention. This may provide an incentive for the observed behavior of preparers.

Much of the behavior of preparers appears to be driven by their own characteristics, be it their preferences, beliefs, skills or approach to filing tax returns. In our regression models, individual fixed effects are not statistically significant but preparer fixed effects are, suggesting preparers are more important than individuals in generating the bunching. Individuals who move to a high bunching preparer increase their bunching to look like that of the other clients of the preparer and they do not show an increasing propensity to bunch before they move. Increases in balances and deductions in moves to high-bunching preparers happen even when the moves are driven by mass movements and firm closures. These changes are not purposeful decisions to change preparer at the individual level. Finally, high-bunching preparers are also more likely to charge fees ending in zero which may provide evidence for round-number preferences

Our paper makes several novel contributions. While other papers have shown a preference for positive refunds, we are the first to show evidence for left-digit bias that generates positive refunds at salient amounts such as $A10, $A100 and $A1000. We generate important insights into tax preparer behavior. We show a link between tax evasion and refund bunching. This link is confirmed by the random-audit data and by behavior of tax preparers more consistent with evasion than effort.

The direct cost of bunching is small—about $A1 million per year. However, moves between preparers suggest that going from behaviors associated with not bunching to those associated with the observed level of bunching, scaled up to the level of the population, would result in a fiscal cost of between $A260–780 million a year. This is over two orders of magnitude greater than the direct cost of bunching, and suggests bunching proxies for tax evasion behaviors that are much more significant for government revenues.

Funding

None of the authors received any funding for this work

Declaration of competing interest

The authors have no relevant financial or non-financial interests or other conflict of interests to disclose.

Data availability

The authors do not have permission to share data.

Appendix A. Proof of Proposition 1

Proposition 2. *Consider a positive balance which has a largest divisor τ in the set $\{10, 100, 1000\}$ (e.g., for a balance of 200, $\tau = 100$). The mass of taxpayers at this balance is:*

(a) increasing in $\theta_{\tilde{\tau}}$ for $\tilde{\tau} \in \{10, 100, 1000\}$ and $\tilde{\tau} \leq \tau$; and
(b) decreasing in $c''/(v'(1 - \Sigma_{\tilde{\tau}\in\{10,100,1000\},\tilde{\tau}>\tau}\theta_{\tilde{\tau}}))$.

The mass of taxpayers at zero balance is:

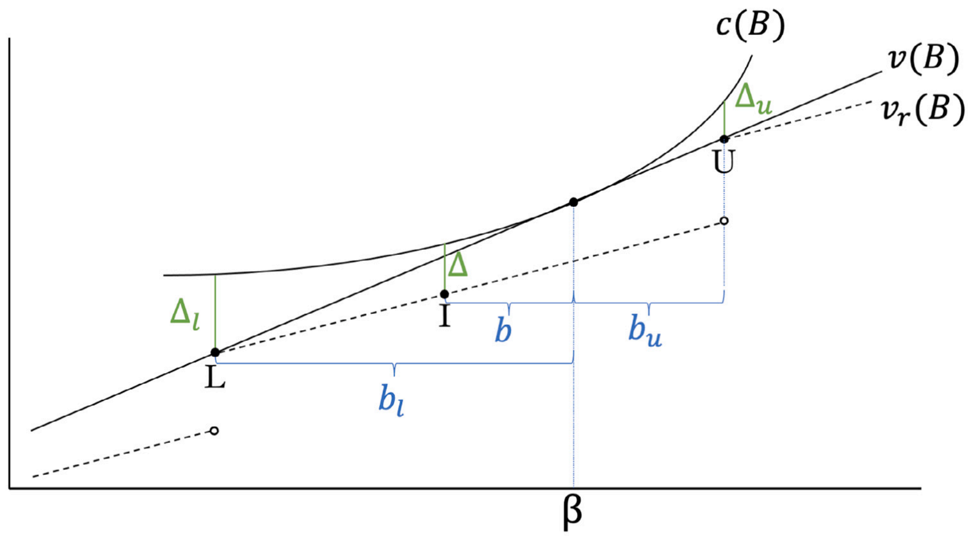

Fig. A.1. Taxpayer cost and benefit curves.
Note: Illustrates the (shifted) taxpayer cost curve $c(B)$, baseline benefits $v(B)$ and behavioral benefits $v_r(B)$ with respect to the balance at assessment B. The optimal balance prior to the introduction of behavioral preferences is β. The new optimum will be one of the leftmost salient threshold L, the rightmost R or an interior solution I.

(c) increasing in θ_τ *for* $\tau \in \{0, 10, 100, 1000\}$
(d) decreasing in c''/v'

Proof. We begin by considering the effect of nonzero θ_{1000} on taxpayers between two successive thousand dollar thresholds τ and $\tau + 1000$. At the baseline equilibrium we have first order condition:

$$c'(\beta) = v'(\beta) = v' \tag{A.1}$$

In Fig. A.1 we illustrate this equilibrium, shifting the cost curve up so that it meets the benefit curve where the tangents are equal.

Note that the distance between this cost curve $c(B)$ and the baseline benefits curve $v(B)$ at any point $\beta + b$ will be:

$$\begin{aligned} c(\beta + b) - v(\beta + b) &= c(\beta) + c'(\beta)b + \tfrac{1}{2}c''(\beta)b^2 - v(\beta) - v'(\beta)b \\ &= (c(\beta) - v(\beta)) + (c'(\beta) - v'(\beta))b + \tfrac{1}{2}c''(\beta)b^2 \\ &= \tfrac{1}{2}c''b^2 \end{aligned} \tag{A.2}$$

We now consider the change to behavioral preferences. There are three possible points where costs minus benefits will be minimized – the left threshold L, the right threshold R or an interior solution I, should it exist, with corresponding minima Δ_L, Δ_I and Δ_R. The interior solution will be characterized by first order condition:

$$\begin{aligned} & c'(\beta + b) = v_r'(\beta + b) \\ \Rightarrow\quad & c'(\beta) + c''(\beta)b = (1 - \theta_{1000})v' \\ \Rightarrow\quad & b = \frac{\theta_{1000}v'}{c''} \end{aligned} \tag{A.3}$$

where we have used the original first order condition (A.1). This solution will exist wherever $b < b_L$.

It follows from Eq. (A.2) that our three possible minima are either:

$$\Delta_L = \tfrac{1}{2}c''b_L^2 \tag{A.4}$$

$$\Delta_R = \tfrac{1}{2}c''b_R^2 \tag{A.5}$$

and

$$\begin{aligned} \Delta_I &= \tfrac{1}{2}c''b^2 + (b_L - b)\theta_{1000}v' \\ &= \tfrac{1}{2}c''\left[b^2 + 2(b_L - b)\frac{\theta_{1000}v'}{c''}\right] \\ &= \tfrac{1}{2}c''\left[b^2 + 2(b_L - b)b\right] \end{aligned} \tag{A.6}$$

Whether Δ_L or Δ_R is smaller is simply a question of whether b_L or b_R is smaller; equivalently, which of the two thresholds β is closest to. A taxpayer originally in $[T, T + 500)$ will only ever bunch to the left, while a taxpayer originally in $(T + 500, T + 1000]$ will only ever bunch to the right. Thus we only need to consider the comparison between a potential interior minima and the minima attained at the thresholds.

First, since the L lies along the same line segment as the interior solution, a taxpayer will only bunch left where the interior solution does not exist, namely $b > b_L$.

Second, a taxpayer will only bunch right if $\Delta_R < \Delta_I$. This will occur when the following expression is strictly negative:

$$\begin{aligned} \Delta_R - \Delta_I &= \frac{1}{2}c'b^2 + (\beta - b)\theta v' \\ &= \frac{1}{2}c''\left[b^2 + 2(\beta - b)\frac{\theta v'}{c''}\right] \\ &= \frac{1}{2}c''\left[b^2 + 2(\beta - b)b\right] \end{aligned}$$

By the quadratic formula it can be shown that this is zero when:

$$b = b_L \pm \sqrt{b_L^2 - b_U^2} \tag{A.7}$$

A taxpayer will bunch right whenever $b \in [b_L - \sqrt{b_L^2 - b_U^2}, b_L + \sqrt{b_L^2 - b_U^2}]$. However, the interior solution does not exist for $b > b_L$. Hence we can expand this domain to $b \in [b_L - \sqrt{b_L^2 - b_U^2}, \infty)$. We thus have four cases depending on the original optimal balance β and b:

$$\begin{aligned} &\beta \in [T, T + 500],\ b \in [0, b_L) && \Rightarrow \text{no bunching} \\ &\beta \in [T, T + 500],\ b \in [b_L, \infty) && \Rightarrow \text{bunching left} \\ &\beta \in [T + 500, T + 1000],\ b \in [0, b_L - \sqrt{(b_L^2 - b_R^2)}) && \Rightarrow \text{no bunching} \\ &\beta \in [T + 500, T + 1000],\ b \in [b_L - \sqrt{(b_L^2 - b_R^2)}, \infty) && \Rightarrow \text{bunching right} \end{aligned}$$

In particular, it follows that bunching is increasing in b. For $\tau = 100$ and $\tau = 10$ the same logic applies after first discounting the marginal utility of an additional dollar of balance by the higher thresholds theta, that is: replacing v' with $v(1 - \theta_{1000})$ or $v(1 - \theta_{1000} - \theta_{100})$ respectively. Eq. (A.3) and analogous expressions for the smaller thresholds then establish the proof. □

Appendix B. Additional charts

See Figs. B.1–B.9.

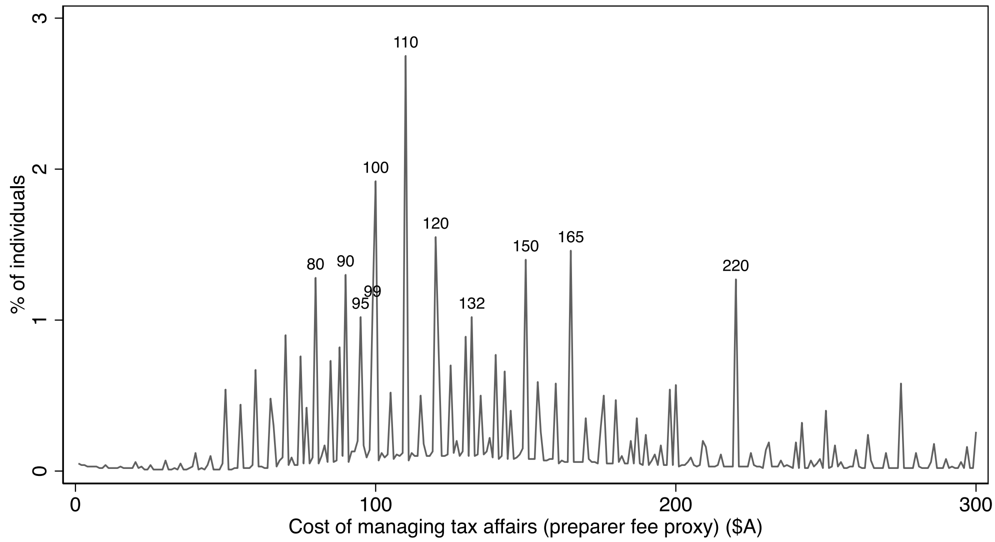

Fig. B.1. Distribution of cost of managing tax affairs, 2000–2018.
Note: Distribution of the deduction for the cost of managing tax affairs, which we use as a proxy for tax preparer fees in the year prior. Based on the full ALife sample, restricted to those with a tax-preparer return and with a non-missing deduction for the cost of managing tax affairs in the following year. Percentages are for each $A1 bin and are based on the full distribution rather than the window shown; bins with more than 1% of the sample are labeled. Around 40% of those with a tax-preparer return do not claim any deduction in the following year. This may reflect either a lost opportunity, or the shifting of fees and deductions to family members on higher marginal tax rates, which is not permitted but may be hard to audit.

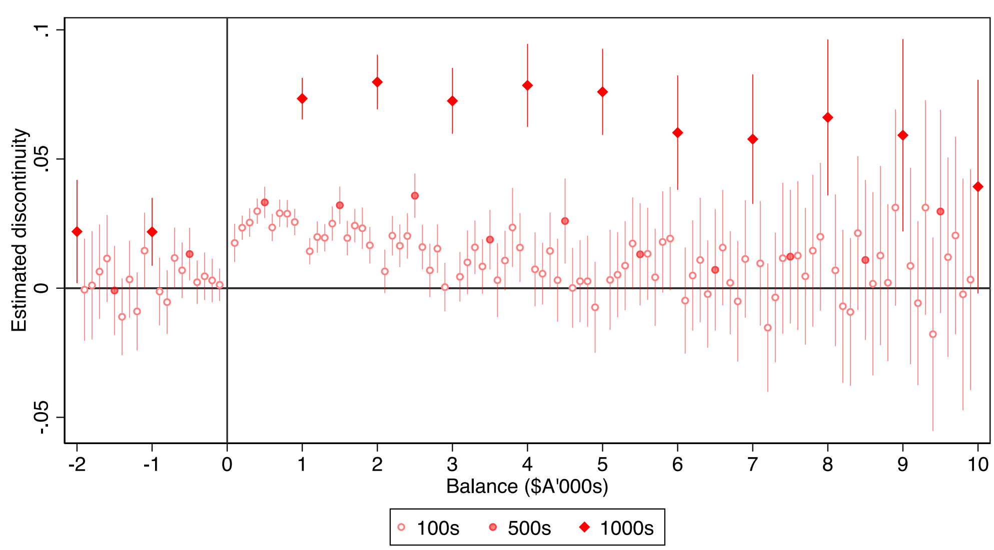

Fig. B.2. Estimated discontinuities at specific thresholds, 1991–2018.
Note: Estimated discontinuity δ in the normalized count around specific $A100 thresholds, with 95% confidence intervals. Based on estimation of Eq. (2) in a window $A50 either side of the given threshold. For small values of δ an individual is 200δ% more likely to be immediately above the threshold than below it.

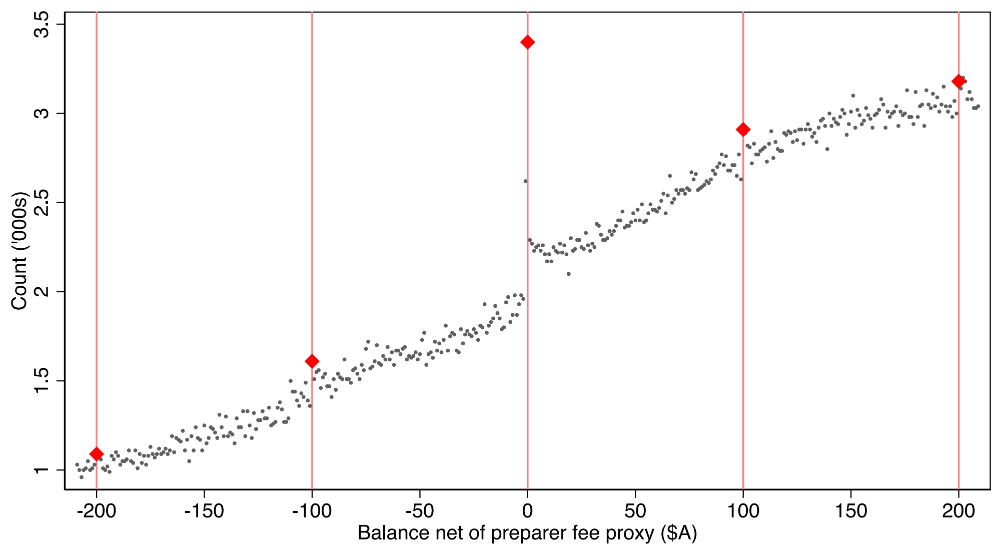

Fig. B.3. Distribution of balance of assessment net of tax preparer fee proxy, 1999–2017.
Note: Distribution of the balance of assessment net of tax preparer fee proxy. Based on the baseline sample of individuals with strictly positive net tax liability and tax withheld and a balance of assessment consistent with the remainder of the tax return, further restricted to those with a tax-preparer return and claiming a nonzero deduction for the cost of managing tax affairs in the following year. Counts are for each $A1 bin. Graph captures 870,000 tax returns over the 1999–2017 income years (since the tax preparer fee proxy is only available from 2000 through to 2018).

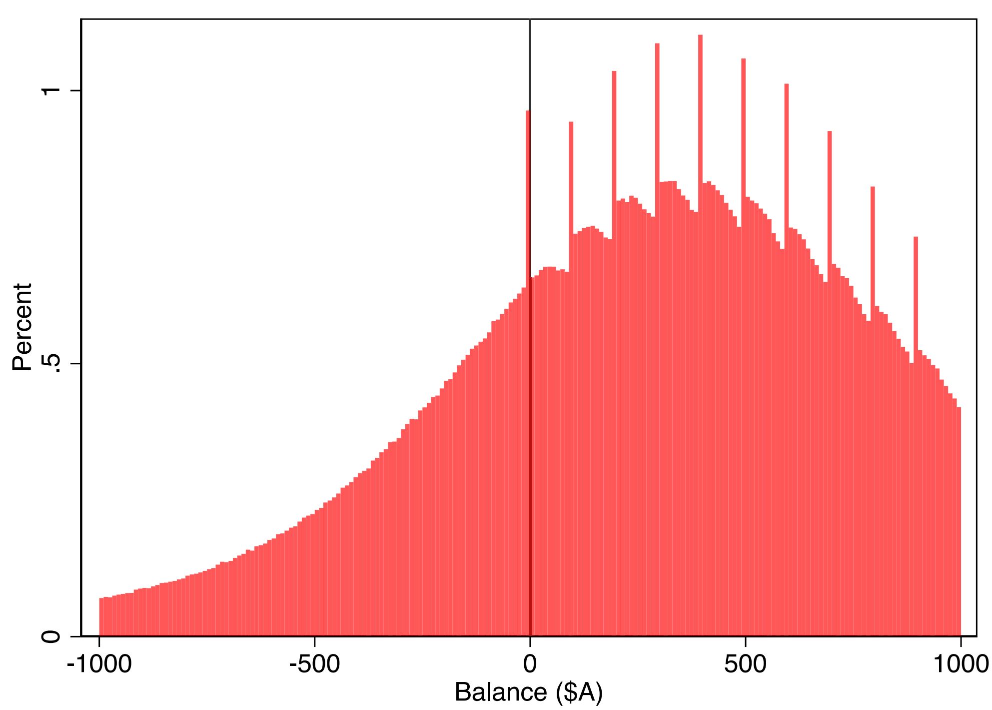

Fig. B.4. Simulated distribution of balances.
Note: Plotted for 10 million observations with $\upsilon = 1$ and a quadratic cost function with linear term normally distributed with mean $-\frac{1}{2}$ and standard deviation 2 and quadratic term uniformly distributed over [0.001,0.011]. Only ten per cent of the population has behavioral preferences—for them, θ_0 is uniformly distributed over [0,0.2] and θ_{100} is uniformly distributed over [0,0.1].

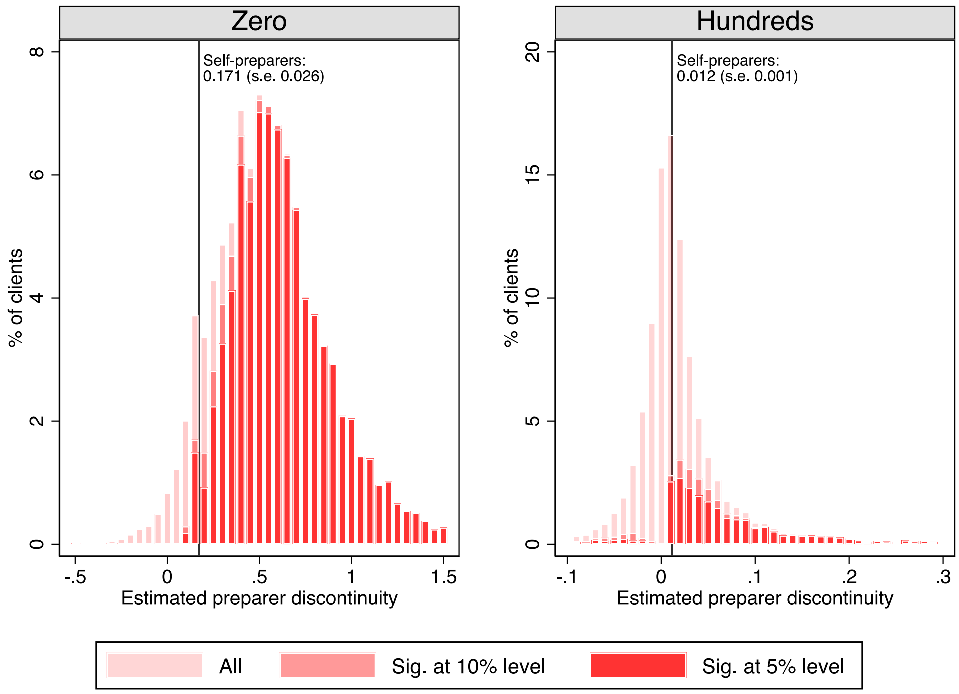

Fig. B.5. Distribution of preparer-specific discontinuities across clients of tax preparers, 1991–2018.
Note: Distribution of the preparer-specific estimates of the discontinuity δ in the normalized count around either the zero or hundred dollar thresholds. Distributions are client-weighted and hence expressed as the percentage of clients of tax preparers falling into given bins. We graph all of the estimated discontinuities and show the fraction that are statistically significant at either the 5% or 10% level. The vertical black line is the estimated discontinuity for self-preparers. It is statistically significant.

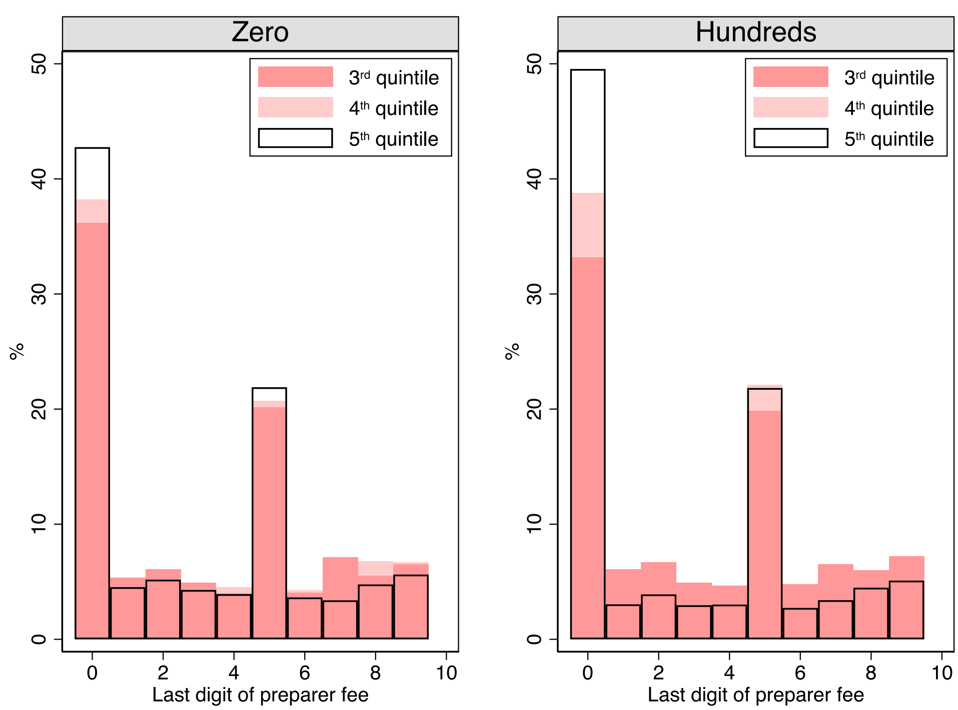

Fig. B.6. Distribution of last-digit of preparer fee proxy by preparer bunching propensity, 1999–2017.
Note: Distribution of the last-digit of the preparer fee proxy by preparer bunching propensity. We focus on the top three quintiles of the preparer bunching distribution.

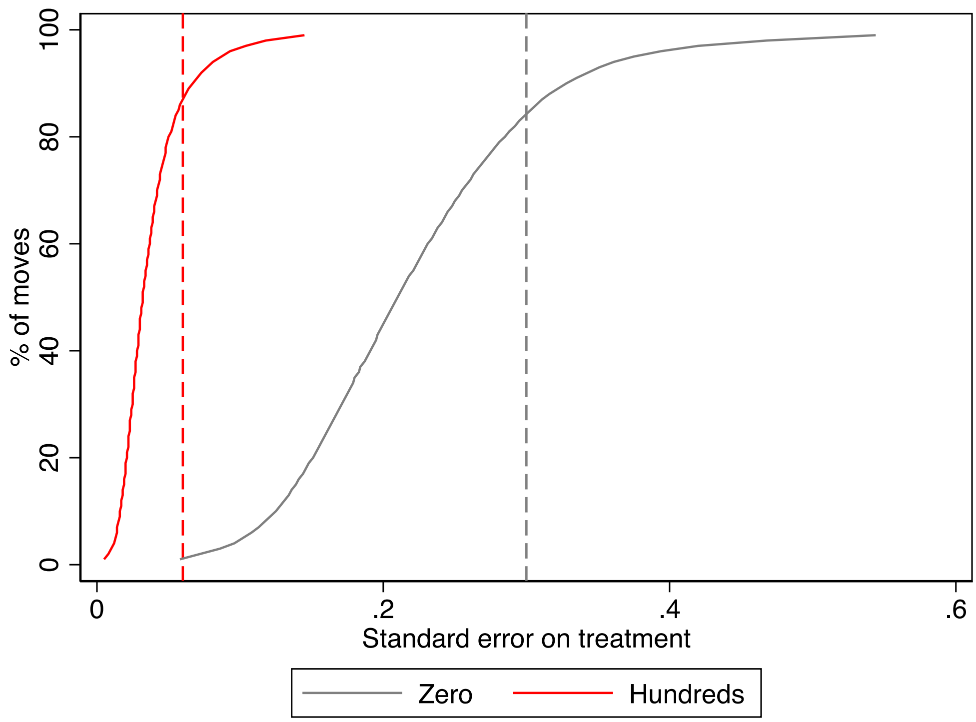

Fig. B.7. Cumulative distribution of standard error on the treatment.
Note: Shows the cumulative distribution of the standard error on the treatment for those in the movers sample. The horizontal lines indicate the cutoffs we use to remove those moves where the difference in bunching propensities is imprecisely estimated.

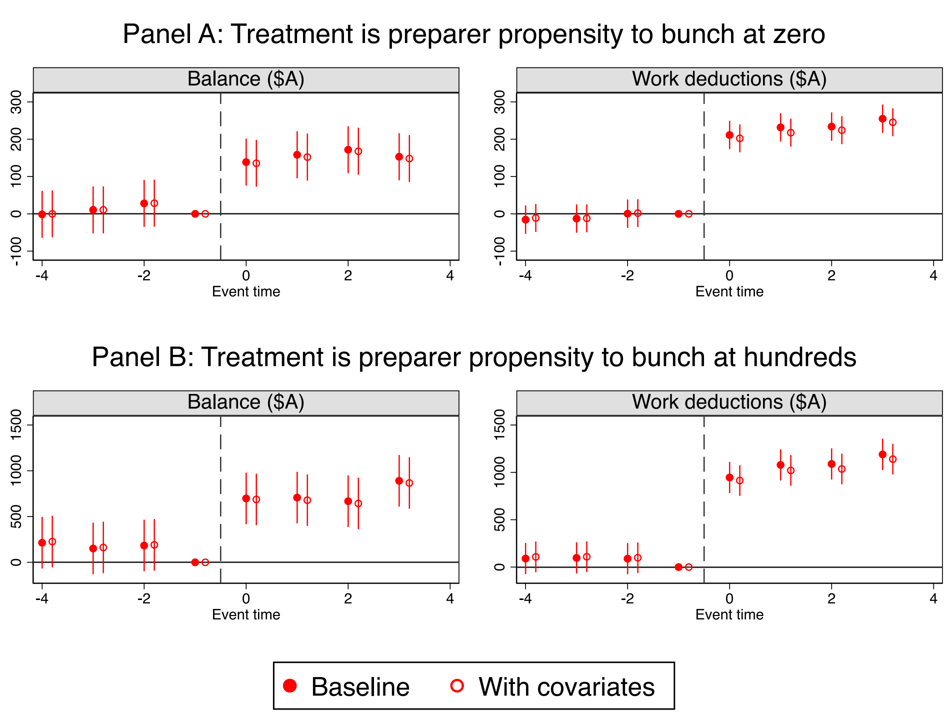

Fig. B.8. Effect of moving to a higher bunching preparer, with and without additional covariates.
Note: Presents coefficient estimates and 95% confidence intervals for ζ from OLS regression estimation of Eq. (11), with and without additional covariates. The additional covariates include fixed effects for occupation and location, and the natural logarithm of wage and salary income. See the note to Table 3 for more information on the occupation and location covariates. The first post-event year coefficient and standard error, sample size and R^2 are available in Table 6 (baseline) and Appendix Table C.2 (with covariates).

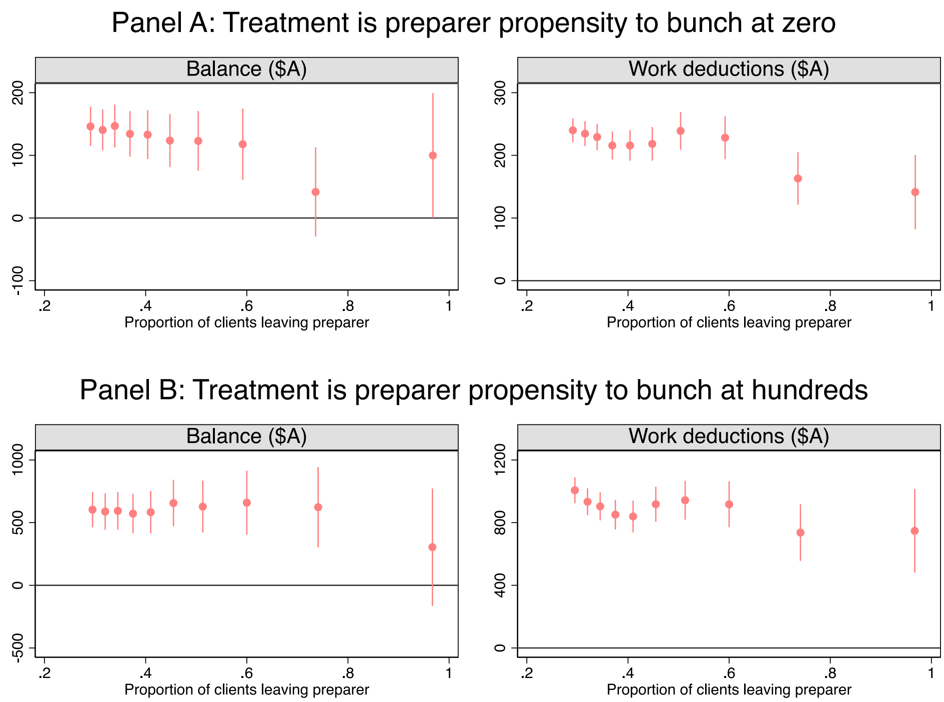

Fig. B.9. Effect of moving to a higher bunching preparer as part of large outflows from original preparer.
Note: Presents coefficient estimates and 95% confidence intervals for ζ from OLS regression estimation of Eq. (11), restricting to moves that were part of increasingly large outflows from the original preparer. We begin by splitting the full sample of movers into deciles based on the proportion of all the original preparers clients leaving in the year of the move. The leftmost point then estimates the effect of moving based on the full sample. The next leftmost considers those in the top 90%, and so on, until we reach moves that are in the top decile based on preparer outflows.

Appendix C. Additional tables

See Tables C.1–C.3.

Table C.1
Correlates of discontinuities at zero and hundred-dollar thresholds—persistence.

	Zero		Hundreds	
	(1)	(2)	(3)	(4)
Prior year balance over threshold	0.8606*** (0.0327)	0.5195*** (0.0335)	0.0110*** (0.0016)	0.0037*** (0.0016)
Fixed effects				
Location	X	X	X	X
Occupation	X	X	X	X
Preparer		X		X
R^2	0.038	0.075	0.000	0.003
N (million)	2.3	2.3	56.7	56.7

Note: Presents coefficients δ and standard errors from OLS regression estimation of Eq. (8) on the baseline sample. Columns (1)–(2) examine the zero-dollar threshold, while columns (3)–(4) examine the hundred-dollar thresholds $\{100,\ldots,2500\}$. Columns (1) and (3) replicate column (3) in Tables 2 and 3 respectively but with added controls for having a prior year balance in the window around the given threshold, the continuous value of that balance, and an indicator for if it is above the given threshold. We show only the coefficient on the last of these. In columns (2) and (4) we add preparer fixed effects. *,** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table C.2
Effect of preparer bunching-propensity on tax return, with additional covariates.

	Zero				Hundreds			
	Coef.	p	R^2	N (million)	Coef.	p	R^2	N (million)
Balance	135*** (32)	0.79	0.34	0.81	687*** (144)	0.40	0.34	1.02
Work-related expenses	202*** (19)	0.83	0.69	0.80	914*** (83)	0.48	0.69	1.00

Note: Presents coefficient estimate ζ_0 and standard errors from OLS regression of Eq. (8) (bunching outcomes) or Eq. (10) (all other outcomes). These equations are standard event study designs that allow for individual, year, event time and age fixed effects, and where the interaction of the treatment – the difference in bunching propensity between preparers – and event time is the key variable of interest. Additional covariates included here are fixed effects for occupation and location, and the natural logarithm of wage and salary income. See the note to Table 2 for more information on the occupation and location covariates. We show the coefficient for the first year with the new preparer. This can be interpreted as the effect of moving from a preparer that never bunches to one that always bunches. *,** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table C.3
Effect of moving to a new tax preparer, by sign and size of change in preparer bunching propensity.

	Zero		Hundreds	
	Balance	Work deductions	Balance	Work deductions
Larger negative	104***	150***	344***	74
	(36)	(22)	(131)	(76)
Smaller negative	127	218***	1100	2,338***
	(131)	(79)	(971)	(568)
Smaller positive	135	222***	−206	1,768***
	(100)	(61)	(842)	(493)
Larger positive	182***	314***	789***	1,636***
	(31)	(19)	(108)	(63)
Equal (p-value)	0.36	0.00	0.02	0.00
Smaller equal (p-value)	0.97	0.97	0.38	0.51
N (million)	0.81	0.80	1.02	1.00

Note: Presents coefficient estimates ζ and standard errors from OLS regression estimation of Eq. (11), where ζ is allowed to vary with the size of the treatment D_i. Namely, we categorize the moves on the basis of D_i into positive and negative moves (those to higher or lower bunching preparers) and also by whether they are above (larger) or below (smaller) the median magnitude of a move within moves of the same sign. We also present p-values on a Wald test of the equality of all the coefficients or the smaller negative and smaller positive coefficients. *,** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Appendix D. Excess mass calculations

This Appendix estimates the proportion of taxpayers moving in response to hundred-dollar thresholds in their tax refunds. We present estimates from the parametric approach used throughout the paper, as well as estimates from a more standard non-parametric approach.

A large empirical literature has developed the theory and associated empirical approaches for quantifying bunching around thresholds in the tax system (Kleven, 2016; Bertanha et al., 2024). These methods have been developed in settings where the focus is typically on precisely estimating responses to a single threshold to uncover a fundamental parameter such as the elasticity of taxable income. The resulting methods are data intensive and do not readily allow for estimation on small samples or a multivariate analysis of what drives bunching intensity.[30] Given this, we use a parametric approach that allows us to readily quantify bunching across the many thresholds of interest and for individual tax preparers, while also permitting a multivariate analysis.

A standard non-parametric approach is more appropriate when estimating the proportion of taxpayers moving in response to the hundred-dollar thresholds. Here the interest is on precisely estimating the response, and the finer details of what drives the response are less of a concern. Following the approach outlined in Kleven (2016), we estimate a high order polynomial with individual fixed effects for all values of the balance in the bunching window W. In particular, we use a 10th order polynomial and estimate the following via ordinary least squares regression:

$$c_b = \sum_{i=0,10} \beta_i \cdot b^i + \sum_{j \in W} \gamma_j \cdot \mathbb{1}_{b=j} + \varepsilon_{it} \tag{D.1}$$

where c_b is the count of returns with dollar balance b. We focus on balances between \$A350 and \$A10,049.[31] Given the visual evidence that missing and excess mass persists some way from the threshold, we examine bunching windows that encompass 40, 45, 48 and 49 observations either side of the hundred-dollar thresholds. An estimate for the excess mass can then be calculated as $\sum_{j\in W^+} \hat{\gamma}_j$ where W^+ is the half of the window that sits above the threshold. We can test whether the excess mass and missing mass are equal by estimating the total mass $\sum_{j\in W} \hat{\gamma}_j$ and examining whether it differs meaningfully from zero. The regressions for the non-parametric approach involve between 7,771 and 9,517 parameters respectively.[32]

A much more parsimonious but less precise estimate can be derived from our parametric approach. Fig. D.1 shows the distribution of balances around hundred-dollar thresholds in our window of interest, with the predicted values from Eq. (2) (with three parameters) as a solid red line. A rough estimate of the relative excess mass can be arrived at by comparing the predicted density to a linear counterfactual density function that coincides with the prediction at its end points—the dashed red line shown. This yields a relative excess mass of $25\delta\%$.[33] By construction the relative excess and missing masses in this approach are equal.

Table D.1
Excess mass and total mass around hundred-dollar thresholds: per cent of taxpayers in window.

	Parametric	Non-parametric			
	(1)	(2)	(3)	(4)	(5)
Excess mass	0.639	0.525	0.552	0.481	0.403
	(0.037)	(0.033)	(0.048)	(0.082)	(0.130)
Total mass		0.102	0.123	−0.003	−0.151
		(0.061)	(0.092)	(0.160)	(0.258)
# observations (million)	14.9	14.9	14.9	14.9	14.9
# bins	100	9700	9700	9700	9700
Half-width of bunching window	NA	40	45	48	49

Note: This table presents estimates of the excess mass and total mass around hundred-dollar thresholds for balances between \$A350 and \$A10,049, and associated robust standard errors. Column (1) applies the parametric approach, while columns (2)–(5) apply a less parsimonious but more traditional non-parametric approach for varying bunching windows.

Table D.1 presents the estimates from the approaches outlined above. The parametric approach suggests 0.6% of taxpayers within the region move in response to hundred-dollar thresholds in their tax refunds. The non-parametric approaches all suggest slightly smaller behavioral responses, ranging between 0.40 and 0.55. However, the estimates are sensitive to the choice of bunching window. Initially, and intuitively, a larger bunching window results in a larger estimate of relative excess mass, as it increases from 0.525 to 0.552. The falling estimates beyond this point may reflect the influence of bunching at \$A50 thresholds, which is apparent in the relatively high count in the first data point in Fig. D.1. This highlights another challenge for which standard approaches are not well suited—the fact that we have many thresholds with potentially overlapping bunching windows.

Based on the pattern of results in Table D.1, our preferred estimate of the proportion of taxpayers moving in response to hundred-dollar thresholds is 0.5%. While a more precise estimate might be possible with further investigation of potential counterfactual distributions and allowance for bunching at \$A50 thresholds, this is beyond the scope of the current paper.

[30] Instead, the drivers of bunching are typically examined through subgroup analysis.

[31] We consider balances of \$A350 or greater to avoid needing to fit the global maximum in the distribution. Doing so results in a fit that has much poorer local performance, which undermines its ability to serve as a credible counterfactual, as needed for bunching estimation. We consider balances up to \$A10,049 to avoid precision issues that arise when fitting a high dimension polynomial in the balance. The chosen window contains 14.9 million returns, or over 80% of the 18.3 million returns that have a balance of \$A50 or greater and are hence within \$A50 of a positive hundred-dollar threshold.

[32] These consist of a constant, ten polynomial terms in the balance and indicator variables for between 80 and 98 observations in each of the 97 bunching windows.

[33] To see this, note that with this normalization, the total mass in a given window is equal to its width W, while the excess mass relative to the counterfactual continuous linear relationship between the two endpoints of the estimated relationship is $\frac{\delta W}{4}$ (the triangle of height δ and base $W/2$).

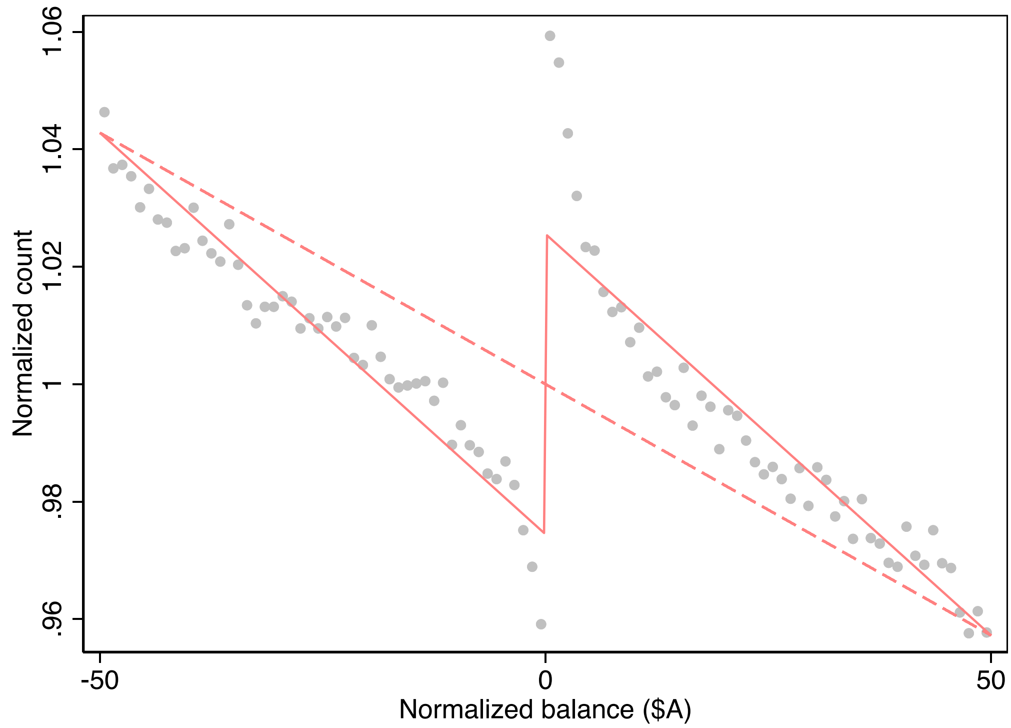

Fig. D.1. Distribution of balance of assessment around hundred-dollar thresholds,1991–2018.
Note: Distribution of the balance of assessment for those with: strictly positive net tax liability and tax withheld; and a balance of assessment consistent with the remainder of the tax return. Further restricted to balances between $A350 and $A10,049 inclusive. Counts are for $A1 bins either side of multiples of $A100. Counts are normalized and a line of best fit, with discontinuity, is estimated as in Eq. (2). Sample consists of 14.9 million tax returns over the 1991–2018 income years.

References

Abhayaratna, Thomas, Carter, Andrew, Johnson, Shane, 2021. The ATO Longitudinal Information Files (ALife): Individuals – A new dataset for public policy research. Tax and Transfer Policy Institute Working Paper 13/2021. Available at https://taxpolicy.crawford.anu.edu.au/publication/ttpi-working-papers/19025/ato-longitudinal-information-files-alife-individuals-new.

Allen, Eric J., Dechow, Patricia M., Pope, Devin G., Wu, George, 2017. Reference-dependent preferences: Evidence from marathon runners. Manage. Sci. 63 (6), 1657–1672.

Allingham, Michael G., Sandmo, Agnar, 1972. Income tax evasion: A theoretical analysis. J. Public Econ. 1 (3–4), 323–338.

Alm, James, McClelland, Gary H., Schulze, William D., 1992. Why do people pay taxes? J. Public Econ. 48 (1), 21–38.

Australian National Audit Office, 2023. Identifying and Reducing the Tax Gap for Individuals Not in Business. https://www.anao.gov.au/work/performance-audit/identifying-and-reducing-the-tax-gap-individuals-not-business. (Accessed: 23 February 2024).

Australian Taxation Office, 2021a. Individuals not in business income tax gap. https://www.ato.gov.au/About-ATO/Research-and-statistics/In-detail/Tax-gap/individuals-not-in-business-income-tax-gap/. (Accessed: 30 September 2021).

Australian Taxation Office, 2021b. Taxation statistics, 2018-19. https://www.ato.gov.au/About-ATO/Research-and-statistics/In-detail/Taxation-statistics/Taxation-statistics-2018-19/. (Accessed: 05 April 2022).

Australian Taxation Office, 2024. Australian tax gaps overview. https://www.ato.gov.au/about-ato/research-and-statistics/in-detail/tax-gap/australian-tax-gaps-overview/overview. (Accessed: 23 February 2024).

Battaglini, Marco, Guiso, Luigi, Lacava, Chiara, Patacchini, Eleonora, 2019. Tax Professionals: Tax-Evasion Facilitators or Information Hubs? National Bureau of Economic Research.

Bertanha, Marinho, Caetano, Carolina, Jales, Hugo, Seegert, Nathan, 2024. Bunching estimation methods. In: Handbook of Labor, Human Resources, and Population Economics. Springer.

Callaway, Brantly, Goodman-Bacon, Andrew, Sant'Anna, Pedro H.C., 2021. Difference-in-differences with a continuous treatment. arXiv preprint arXiv:2107.02637.

Cattaneo, Matias D., Jansson, Michael, Ma, Xinwei, 2018. Manipulation testing based on density discontinuity. Stata J. 18 (1), 234–261.

Cattaneo, Matias D., Jansson, Michael, Ma, Xinwei, 2020. Simple local polynomial density estimators. J. Amer. Statist. Assoc. 115 (531), 1449–1455.

Chetty, Raj, Hendren, Nathaniel, 2018. The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. Q. J. Econ. 133 (3), 1107–1162.

Cook, Jonathan, Kowaleski, Zachary T., Minnis, Michael, Sutherland, Andrew, Zehms, Karla M., 2020. Auditors are known by the companies they keep. J. Account. Econ. 70 (1), 101314.

Deutscher, Nathan, 2020. Place, peers, and the teenage years: Long-run neighborhood effects in Australia. Am. Econ. J.: Appl. Econ. 12 (2), 220–249.

Dhami, Sanjit, al-Nowaihi, Ali, 2007. Why do people pay taxes? Prospect theory versus expected utility theory. J. Econ. Behav. Organ. 64 (1), 171–192.

Dube, Arindrajit, Manning, Alan, Naidu, Suresh, 2018. Monopsony and Employer Mis-Optimization Explain Why Wages Bunch at Round Numbers. National Bureau of Economic Research.

Engström, Per, Nordblom, Katarina, Ohlsson, Henry, Persson, Annika, 2015. Tax compliance and loss aversion. Am. Econ. J.: Econ. Policy 7 (4), 132–164.

Ginzberg, Eli, 1936. Customary prices. Am. Econ. Rev. 26 (2), 296.

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., J., Diez-Medrano, Lagos, M., Norris, P., Ponarin, E., Puranen, B., 2022. World values survey: Round seven - Country-pooled datafile version 5.0. http://dx.doi.org/10.14281/18241.20, Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat.

Heidhues, Paul, Kőszegi, Botond, 2018. Behavioral industrial organization. In: Handbook of Behavioral Economics: Applications and Foundations, vol. 1, pp. 517–612, 1.

Heraud, Florian, Page, Lionel, 2024. Does the left-digit bias affect prices in financial markets? J. Econ. Behav. Organ. 218, 20–29.

Jones, Peter, 2020. Loss aversion and property tax avoidance. Available at SSRN 3511751.

Kleven, Henrik Jacobsen, 2016. Bunching. Annu. Rev. Econ. 8, 435–464.

Lacetera, Nicola, Pope, Devin G., Sydnor, Justin R., 2012. Heuristic thinking and limited attention in the car market. Amer. Econ. Rev. 102 (5), 2206–2236.

List, John A., Muir, Ian, Pope, Devin, Sun, Gregory, 2023. Left-digit bias at Lyft. Rev. Econ. Stud. rdad014.

Olsen, Asmus Leth, 2013. The politics of digits: Evidence of odd taxation. Public Choice 154, 59–73.

Rees-Jones, Alex, 2018. Quantifying loss-averse tax manipulation. Rev. Econ. Stud. 85 (2), 1251–1278.

Reyes, Germán, 2022. Coarse wage-setting and behavioral firms. arXiv preprint arXiv:2206.01114.

Strulov-Shlain, Avner, 2023. More than a penny's worth: Left-digit bias and firm pricing. Rev. Econ. Stud. 90 (5), 2612–2645.