Australian National University

Crawford School of Public Policy

TTPI

Tax and Transfer Policy Institute

Hundreds and thousands: bunching at positive, salient tax balances and the cost of reducing tax liabilities

---

TTPI - Working Paper 12/2022
September 2022

Robert Breunig
Tax and Transfer Policy Institute
Crawford School of Public Policy
Australian National University

Nathan Deutscher
Tax and Transfer Policy Institute
Crawford School of Public Policy
Australian National University

Steven Hamilton
Department of Economics
George Washington University

Abstract

Australian taxpayers display reference-dependent preferences when filing their tax returns - they bunch at positive and salient thresholds. We develop a model of taxpayer behavior to show that bunching heterogeneity reflects both differences in preferences and the rate at which the marginal cost of reducing one's tax liability increases. Consistent with this model, bunching has grown alongside electronically prepared returns over recent decades and the subsequent responses of tax agents. Taxpayers receiving these balances are more likely to stay with their tax agent, but do not pay higher fees. Consistent with having flatter cost curves, `high-bunching' agents deliver larger balances more generally and do so by lifting deductions and lowering reported income for return items where audits are costly.

Keywords: Bunching, tax refunds, taxpayer behavior, tax agent
JEL codes: D03, D90, H21, H24

**We thank the Australian Taxation Office for providing us the data which allow us to identify common tax agents among tax filers. Seminar participants at the Tax and Transfer Policy Institute Friday seminars provided useful comments and feedback. Corresponding author: robert.breunig@anu.edu.au*

Tax and Transfer Policy Institute

Crawford School of Public Policy

College of Asia and the Pacific

+61 2 6125 9318

tax.policy@anu.edu.au

The Australian National University

Canberra ACT 0200 Australia

www.anu.edu.au

The Tax and Transfer Policy Institute (TTPI) is an independent policy institute that was established in 2013 with seed funding from the federal government. It is supported by the Crawford School of Public Policy of the Australian National University.

TTPI contributes to public policy by improving understanding, building the evidence base, and promoting the study, discussion and debate of the economic and social impacts of the tax and transfer system.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

I. Introduction

Having completed their 2018 tax return a typical Australian taxpayer was 84% more likely to be owed a dollar by the government than to owe a dollar, and 70% more likely to be owed AUD1000 than AUD999. We identify bunching at positive, salient balances at the point of assessment, and show that it is consistent with models of loss aversion and reference-dependent preferences. We also use the emergence of this behaviour to shed new light on the role of filing technologies and tax agents, a still understudied area of the tax system. In particular we highlight the potential for this bunching behaviour to act as a proxy for the underlying shape of the curve describing the cost of reducing one's tax liability. A flatter curve – as enabled by electronic and agent-prepared returns – results in both more bunching but also larger balances across the board.

Individuals display reference-dependent preferences in a wide variety of settings, from marathon times (Allen et al. (2017)) to the relationship between a used car's mileage and its price (Lacetera, Pope and Sydnor (2012)). What is perhaps less well understood is how these behaviors emerge and how other economic actors respond to these preferences, or indeed display them themselves. At times the responses to these preferences can be used to draw broader inferences. For example, Dube, Manning and Naidu (2018) provide evidence that employers (rather than employees) drive bunching at round number hourly wage rates, which they argue is most easily rationalised in a labour market with monopsony power.

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. 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 agent.[1] In 2018-19 the cost of managing individual tax affairs, including tax agent fees, was around AUD2 billion (or over 1% of GDP).[2] Decisions made at the point of filing also have the potential to 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 USD34 more than those facing a refund. Notably, Rees-Jones (2018) also finds that bunching at zero is slightly more pronounced among agent-prepared 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. Separately, the role of tax agents has also been explored, with Battaglini et al. (2019) exploring their role as information hubs for small business clients and interactions with the audit process.

[1] See Australian Taxation Office (2021*b*), Individual Statistics, Chart 7.
[2] See Australian Taxation Office (2021*b*), Individual Statistics, Table 6.

In this paper we make three contributions drawing on nearly three decades of Australian tax data linking individual taxpayers to their tax returns and their agents from 1991-2018. First, we document bunching at positive *and* salient balances, namely multiples of AUD10, AUD100 or AUD1000. These behaviours grow markedly over the period we study, with no significant bunching at hundred dollar thresholds in 1991. We develop a model showing that heterogeneity in bunching can reflect differences in the strength of reference-dependent preferences, but also the curvature of the cost function when it comes to claiming a larger balance. We argue that it is unlikely that preferences have changed over this period – rather the appearance of electronic tax returns, coupled with tax agent responses to existing preferences and these new technologies, has resulted in steady growth in bunching. Individuals have sorted to tax agents that are more likely to deliver these refunds, but agents are also increasing in their propensity to bunch.

Our second contribution is to consider the implications of this behaviour – what do agents and individuals get out of this? Agents appear to be responding at least in part to individual taxpayer preferences. Self-prepared electronic returns also display bunching at positive, salient balances, albeit it about half the size of that observed in agent-prepared returns. Further, we show individuals who receive a refund at or just above one of these salient balances are more loyal to their tax agent in the following year. In contrast, there is no effect on fees charged.

There is a great deal of heterogeneity in the extent to which tax agents bunch at positive, salient balances. The extreme bunching of some agents leaves open the possibility of 'behavioural agents' that derive positive, salient balances to satisfy their own preferences (similar to evidence for 'behavioural firms' setting round number hourly wage rates in Dube, Manning and Naidu (2018)). As suggestive evidence for this, we show that agents that are more likely to deliver positive, salient balances are also more likely to charge fees that end in '0' – suggesting they may themselves have a preference for such numbers.

Finally, we used an event study design to show that tax agents differing in their propensity to deliver these balances have a clear causal effect on the outcomes of their clients, over and above that required to deliver the bunching. This provides evidence that bunching can act as a proxy for shape of the cost curve governing the extent to which tax liabilities will be reduced more generally. A move between agents that differ by one standard deviation in their bunching propensity, results in a persistent increase in an individual's tax refund of AUD41 for bunching around the zero threshold, and AUD52 for bunching around the AUD100 thresholds. Mapped to the full population of some 15 million taxpayers this implies a fiscal cost of around AUD600-750 million a year, more than two orders of magnitude greater than the direct cost of bunching itself.

The effects of high-bunching agents 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, where changes may trigger a change both in agent and tax return outcomes. 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 agent – reflecting significant downsizing or closure – which suggests the choice to move does not drive our results.

The largest and most precisely estimated effects for individual items on the tax return are in higher claims for work-related expense deductions, where the cost of auditing claims is high and there is significant room for taxpayer and agent misinterpretation. This is consistent with published results from the Australian Taxation Office (ATO) random enquiry program, where such claims have been identified as those most likely to require adjustment Australian Taxation Office (2021*a*).

Our results show how electronic returns, coupled with the responses of tax agents, has led to the strong growth in bunching at positive, salient thresholds. This reflects agents responding rationally to the preferences of their clients, and being rewarded with greater loyalty; though there is also suggestive evidence for some 'behavioural' agents. While the fiscal impact of bunching is modest – of the order of a few million 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.

Before presenting the data and the empirical phenomena that we explore, we present a simple theoretical framework of tax filing which will help us to think about what drives the observable bunching in the data. We then present the data that we use and the bunching behaviour we observe. We examine the correlates of that behaviour and the important role of tax agents. We then examine why agents might deliver positive, salient balances and the effect that high-bunching agents have on tax returns. We conclude in the final section.

II. Theoretical framework

In this section we briefly consider the theoretical predictions from a simple model of tax filing. We also consider the implications of reference-dependent preferences that reflect the tendency of taxpayers to prefer positive balances, and balances above particularly salient thresholds, such as hundred 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 to also relate these behaviors to the broader process of tax filing.

In our set up, individual taxpayers maximise 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:

$$v(b) = v'b. \tag{1}$$

for some $v' > 0$.[3] We also assume a quadratic cost function with $c'' > 0$ that is minimised at some balance B_0. We can think of B_0 as the default balance arising from a return where taxpayers minimise the costs arising from both effort and risk of audit. This would involves reporting all income likely to be reported to the tax authority, but not making any claims for deductions or offsets.

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

$$\begin{aligned} & & c'(B_0 + b^*) &= v'(B_0 + b^*) \\ & \Rightarrow & c'(B_0) + c''(B_0)b^* &= v'(B_0) \\ (2) \qquad & \Rightarrow & b^* &= \frac{v'}{c''} \end{aligned}$$

Hence the extent to which a taxpayer increases the balance owed to them depends on both the utility they derive from the additional dollars, but 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 realised outcomes is rarely observed. Yet as we will show below, many of the same factors also influence taxpayer responses to reference-dependent preferences, which results in observable bunching in the distribution of balances.

We consider two models of reference-dependent preferences, with associated value functions $v_r(b)$. For the zero threshold we consider a model of loss aversion where:

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

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:

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

for some $\theta_{10}, \theta_{100}, \theta_{1000} > 0$ and $b > 0$. With this 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, Pope and Sydnor (2012)). In this case, for simplicity and motivated by later findings, the discount rates are

[3] 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.

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 reference-dependent value function:

$$(3)\quad v_r(b) = v'b + v'\theta_0 \mathbb{1}[b < 0] \\ - v'\left(\theta_{10}\text{mod}(b, 10) + \theta_{100}\text{mod}(b, 100) + \theta_{1000}\text{mod}(b, 1000)\right)\mathbb{1}[b > 0]$$

In Figure 1 we illustrate this value function, along with a benchmark taxpayer, who does not have reference dependent preferences, and two illustrative cost curves. As apparent from the figure, the reference dependent value function is characteristised by a 'kink' at zero and notches at the salient refunds.

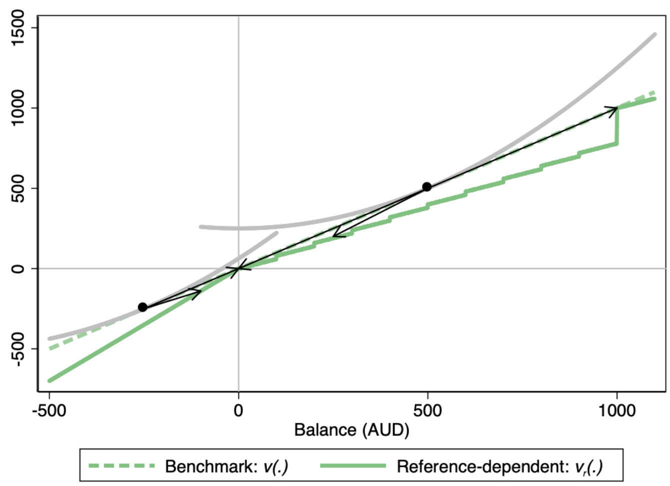

Figure 1. Value functions with and without reference dependent preferences

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 $-AUD250$ and $AUD500$ respectively. The arrows show potential new optima with reference dependent preferences, as described in the text.

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. But the other reference-dependent preferences may lead to taxpayers shifting to the left, as each marginal dollar of balance above a reference point now has less value to them, or shifting to the right in response to the notch that accompanies each reference point. Shifts to the left may involve bunching at the lower reference point, while shifts to the right can only result in bunching. We can characterise how the extent of bunching at the salient reference points will vary with features of taxpayer preferences.[4]

Proposition 1. *Consider a positive balance which has a largest divisor τ in the set $\{10, 100, 1000\}$ (i.e. 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 $\tau \in \{0, 10, 100, 1000\}$

(d) decreasing in c''/v'

We leave the proof of this for Appendix A. This result is fairly intuitive from Figure 1. 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 Figure 1). 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 agent. 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, although here it will be reflected in 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 reference points, 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 equation (2). Bunching thus has the potential to capture some of the claim-by-claim behavior that applies more generally at the point of filing tax returns.

We have abstracted from the role of agents in our model above, both for tractability and also given the many plausible ways in which they could feature in

[4]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 Figure B1, which successfully replicates several features of the empirical distribution that will be seen later.

this set up. As already noted, agents may play a role in flattening cost curves for clients, through their knowledge but also potentially 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 reference-dependent preferences, we might expect rational agents to deliver these positive, salient balances to their clients as part of a profit-maximising strategy. But it is also possible that there are 'behavioral agents' 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 agent bunching in some detail.

III. Data

We draw on the Australian Longitudinal Information Files (ALife) produced for research purposes by the Australian Taxation Office (ATO), see Abhayaratna, Carter and Johnson (2021). ALife is a 10 per cent sample of all individual tax returns from the 1991 to 2018 income years.[5] For our research, the ATO has supplemented ALife with a random identifier linking tax returns prepared by the same practice of tax agents.

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 negate 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.[6] Given this mechanical bunching is not of interest, we further restrict attention to taxpayers with a positive net tax liability. We also restrict to those with positive tax withheld, as otherwise the mass point of taxpayers at zero tax withheld will

[5] 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.

[6] For example, the Education Tax Refund, which entitled eligible taxpayers in 2008-09 to a refund of up to AUD750 a child for education expenses.

produce a mass of individuals with negative balances (since net tax is restricted to be positive) and a mechanical negative discontinuity at zero.[7] Together these restrictions drop a further 27.9% of returns.

We will also draw on a variety of other variables for deductions claimed in the process of filing tax returns. This includes deductions for work-related expenses, and for expenses incurred in managing a rental property; these variables are available from 1992 and 1993 respectively. It also includes deductions for the cost of managing tax affairs, which we will at times use as a proxy for tax agent fees for the previous year.[8] This is a noisy proxy, as taxpayers may forget to claim agent fees or may claim other expenses under this deduction, such as the cost of tax reference materials, tax courses and travel to their tax agent.[9] Nonetheless, the distribution of these deductions, shown in Appendix Figure B2, shows several clear modes as one would expect if they reflected the pricing behaviors in the market for tax returns. This variable is available from 2000 onwards.

IV. Bunching at positive, salient balances

Figure 2 shows the distribution of the balance at assessment across all 22.5 million tax returns in our sample. There are a number of notable features. First, there is both a clear peak at zero and a positive discontinuity. Taxpayers have a tendency to avoid a debt to the tax office on assessment. Beyond this, there are also regularly spaced peaks coincident with hundred and thousand dollar balances. Consistent with the earlier theoretical predictions concerning reference-dependent preferences, taxpayers appear to adjust their tax returns to target positive, salient refunds on tax day.

There are multiple ways to quantify the behaviour apparent in Figure 2. 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:

$$c_b = \alpha + \beta b + \delta \mathbb{1}[b \geq \tau] + \varepsilon_b \tag{4}$$

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 V) and much smaller samples, namely clients of individual tax practices (Section VI).[10]

It is helpful to normalise the key variables in this equation so that the δ co-

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

[8] Given tax returns are typically filed (and expenses incurred and hence deductible) in the year following the return year.

[9] See https://www.ato.gov.au/individuals/income-and-deductions/deductions-you-can-claim/other-deductions/cost-of-managing-tax-affairs/ [Accessed 19 November 2021].

[10] It is more typical to allow the slope to vary either side of the threshold. We prefer this more parsimonious specification for the reasons stated above.

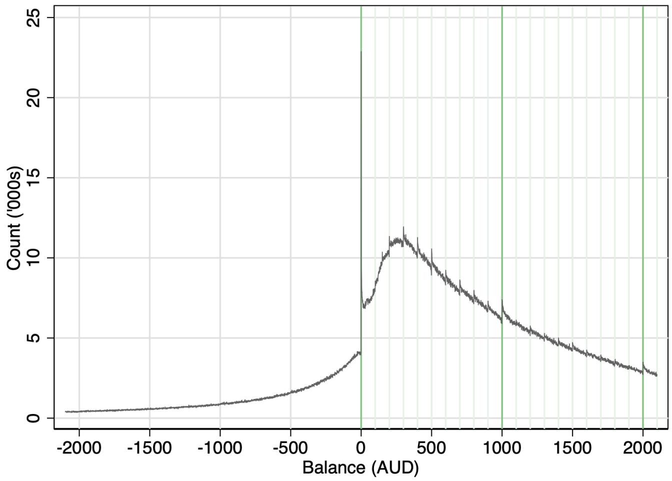

Figure 2. Distribution of balance of assessment, 1991-2018

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 AUD1 bin. Sample consists of 22.5 million tax returns over the 1991-2018 income years.

efficient has a more direct and consistent interpretation. In particular, we can instead estimate:

$$\tilde{c}_{\tilde{b}} = \alpha + \beta\tilde{b} + \delta(2.\mathbb{1}[\tilde{b} > 0] - 1) + \varepsilon_b \tag{5}$$

where we have normalised the count by dividing it by its average over the estimation window ($\tilde{c}_{\tilde{b}} = \frac{c_{\tilde{b}}}{\bar{c}}$), and have re-centred the balance around zero ($\tilde{b} = b - \tau + \frac{1}{2}$). We will typically use a symmetric estimation window looking AUD50 either side of the threshold, in which case the normalised 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 normalised count (1) and our re-centred 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 $200\delta\%$.

In Figure 3 we zoom in on the behavior in Figure 2 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 a AUD5 or AUD50 window of the given threshold and then stack our windows. Consistent with the theoretical predictions in Section II, 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 finalise 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). Figure 3 also shows in green lines the predicted values from ordinary least squares estimation of equation (5). The estimated discontinuities in the normalised 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.

Finer-grained insights into bunching are presented in Appendix Figure B3, 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 might lose some meaning. For example, the discontinuities at hundred dollar thresholds are largest for balances below AUD1000, consistent with classic left-digit bias. However, it also appears that '5s' matter, with larger discontinuities at AUD500 thresholds than other AUD100 thresholds. Reference points matter far less for negative balances, but there are still significant discontinuities at the AUD1000 thresholds – consistent with a desire to avoid particularly salient debts. Finally note that for those using a tax agent to file a tax return, their frame of reference for a loss may be the balance net of the tax agent'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 Figure B4 we show there is indeed a modest discontinuity in this variable around zero.

Estimating the number of taxpayers shifting in response to reference-dependent preferences is challenging. The sheer number and frequency of potential reference points, and the extreme responses to the zero and thousand dollar thresholds, make it difficult to reliably identify regions outside the bunching window to estimate a counterfactual density. Nonetheless, one way to extract a rough estimate of the relative excess mass from the framework above is to compare the predicted density given by the green lines to a counterfactual linear density function that coincides with the prediction at its end points – the dashed green line shown. This yields a relative excess mass of $25\delta\%$.[11] Applied to the discontinuities estimated earlier, this suggests that 10% of taxpayers within AUD50 of zero balance respond to the reference point, versus 1.8% and 0.5% within AUD50 of the thousand and hundred dollar thresholds respectively. Only 0.1% of taxpayers within AUD5 of a ten dollar threshold respond to the reference point.

These relatively small responses imply a small fiscal cost from bunching in

[11] To see this, note that with this normalisation the total mass in a given window is simply equal to its width W, while the 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 $W/2$).

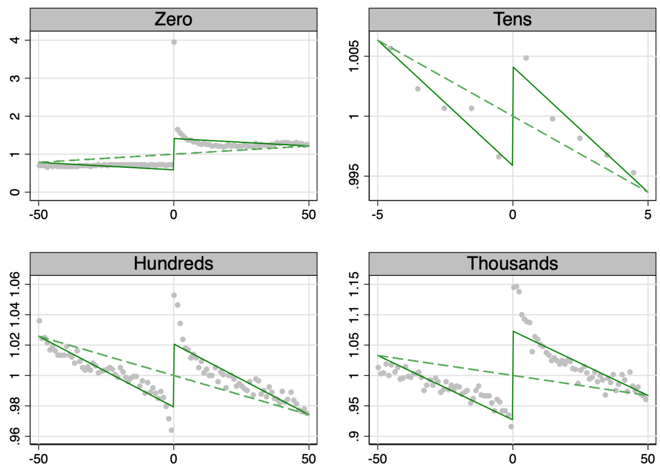

Figure 3. Distribution of balance of assessment around salient thresholds, 1991-2018

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 AUD1 bins either side of either zero or multiples of AUD10, AUD100 or AUD1000. 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 normalised and a line of best fit is estimated as in equation (5). Sample consists of 22.5 million tax returns over the 1991-2018 income years.

response to reference-dependent preferences. For example, across all the years around 183 million taxpayers were within AUD50 of a hundred or thousand dollar threshold.[12] Pooling these thresholds gives an estimated discontinuity of 0.0246 (s.e. 0.0014) and relative excess mass of 0.6% – or 1.1 million taxpayers shifting their balances in response to these reference points. 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 AUD100 estimation window the average difference in balances between those in the excess mass and missing mass regions will be AUD33, which implies a total fiscal cost of AUD37 million, or around AUD1.4 million a year. This is a tiny fraction of the total tax take. A similar train of logic would suggest a fiscal cost of AUD20 million, or around AUD0.76 million a year from moves around the zero threshold. Once again, these estimates are purely indicative. The sign on the average change in balance in response to reference-dependent preferences is actually ambiguous, as suggested by Figure 1. On the other hand, it seems likely that some of those bunching at the zero and thousand dollar thresholds in particular have shifted from outside the AUD100 estimation windows, increasing their balances by well over AUD33.

[12]Multiplying by ten the 18.3 million observed in our 10% sample.

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

V. Correlates of bunching at positive, salient thresholds

We now investigate the correlates of the behaviors identified in the preceding section. This serves as a test of some of the competing theoretical explanations for heterogeneity in bunching. In our earlier 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 that there has been a dramatic increase in bunching over the past three decades. In Figure 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.[13] To explore both this time trend and heterogeneity in bunching more generally, note that there is a return-level equivalent of our earlier regressions estimating the discontinuity at various thresholds. Namely we can estimate:

$$100.\mathbb{1}[b_i = b_n] = \alpha + \beta b_n + \delta(2.\mathbb{1}[b_n > 0] - 1) + \varepsilon_{i,n} \tag{6}$$

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 equation (5). 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. For ease of computation, we aggregate to four bins in our estimation window,

[13] 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% [insert reference].

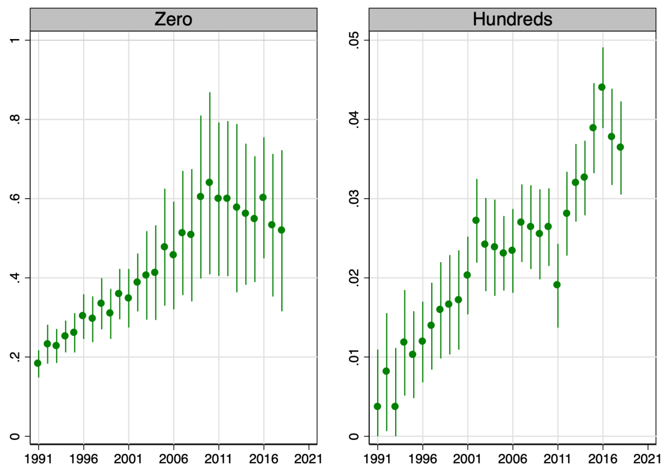

FIGURE 4. ESTIMATED DISCONTINUITIES AT ZERO AND HUNDRED-DOLLAR THRESHOLDS OVER TIME

Note: Estimated discontinuity δ in the normalised count around the zero and hundred-dollar thresholds over time, with 95% confidence intervals. Based on estimation of equation (5) in a window AUD50 either side of the given threshold. For small values of δ an individual is 2δ% more likely to be immediately above the threshold than below it.

estimating the below:

$$4.\mathbb{1}[b_i = b_q] = \alpha X_i + \beta b_q X_i + \delta(2.\mathbb{1}[b_q > 0] - 1)X_i + \varepsilon_{i,q} \tag{7}$$

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 Table 1 we show how the estimated discontinuities δ and their growth over time changes as we expand our set of controls. Columns (1)-(5) present results for the discontinuity around zero, while columns (6)-(10) present results for the discontinuity around hundred dollar thresholds. For ease of computation in the latter case we restrict attention to the discontinuities from AUD100 to AUD2500.[14] Beginning with the specifications without any covariates, in columns (1) and (6) we see discontinuities of 0.375 and 0.0244 in the normalised count, in line with those in Figure 3 and implying increases in the expected count at the thresholds of roughly 75% and 5% respectively. In columns (2) and (7) 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 Figure 4, the results imply discontinuities of around 0.570 and 0.0424 in the normalised count in 2018, that fall away to around 0.2 and near-zero in 1991.

[14]Even with this restriction, the final regression in column (10) takes two days to run in our set up.

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

	Zero					Hundreds				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	0.3754***	0.5699***	0.2767***			0.0244***	0.0424***	0.0145***		
	(0.0025)	(0.0051)	(0.0095)			(0.0005)	(0.0010)	(0.0021)		
Years pre-2018		0.0136***	0.0110***	0.0111***	0.0073***		0.0014***	0.0012***	0.0005***	0.0003**
		(0.0003)	(0.0004)	(0.0012)	(0.0014)		(0.0001)	(0.0001)	(0.0001)	(0.0001)
Agent			0.3131***	0.2586***				0.0136***	0.0106***	
			(0.0070)	(0.0193)				(0.0014)	(0.0020)	
Electronic			0.0715***	0.0570***	0.0707***			0.0108***	0.0117***	0.0128***
			(0.0083)	(0.0192)	(0.0209)			(0.0019)	(0.0023)	(0.0023)
Female			0.0113**					0.0011		
			(0.0050)					(0.0010)		
Under 25			0.0309***	0.0109	0.0111			0.0053***	0.0014	0.0023
			(0.0070)	(0.0231)	(0.0258)			(0.0014)	(0.0021)	(0.0022)
Over 65			-0.0614***	0.0522	0.0792			-0.0005	0.0131**	0.0149***
			(0.0130)	(0.0481)	(0.0547)			(0.0035)	(0.0057)	(0.0057)
Occupational skill										
2			-0.0019					0.0067***		
			(0.0113)					(0.0021)		
3			0.0247**					0.0073***		
			(0.0096)					(0.0018)		
4			0.0079					0.0093***		
			(0.0087)					(0.0017)		
5			-0.0071					0.0131***		
			(0.0099)					(0.0018)		
Self-employed			0.0299***	-0.0346	-0.0447			-0.0181***	-0.0169***	-0.0164***
			(0.0106)	(0.0347)	(0.0408)			(0.0025)	(0.0032)	(0.0033)
Fixed effects										
Location				X	X				X	X
Occupation				X	X				X	X
Individual				X	X				X	X
Agent					X					X
R^2	0.033	0.034	0.037	0.506	0.527	0.000	0.000	0.000	0.084	0.087
N (million)	2.3	2.3	2.3	2.3	2.3	56.7	56.7	56.7	56.7	56.7

Note: Presents coefficients δ and standard errors from OLS regression estimation of equation (7) on the baseline sample. Columns (1)-(5) examine the zero-dollar threshold, while columns (6)-(10) examine the hundred-dollar thresholds $\{100, ..., 2500\}$. The columns progress through estimating a simple discontinuity to one that varies: over time; with return and individual characteristics; with location, occupation and individual fixed effects; and finally agent fixed effects. 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 labour 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.

What explains heterogeneity in bunching at positive, salient thresholds? In columns (3) and (8) we add controls for whether the return was prepared by an agent and lodged electronically, and a variety of demographic controls. Agent-prepared 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 (4) and (9), suggesting this is not simply a feature of the types of individuals who use tax agents or their time-varying observable characteristics. Electronic returns also result in larger discontinuities that are robust to these controls, and also the addition of agent fixed effects. Tax agents 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 reference points due to larger θ_0 and θ_{1000}, or having tax affairs with flatter cost curves (e.g. through more manipulation opportunities).

The results in Table 1 also help understand the evolution of large discontinuities over time. The time trend falls with the addition of controls, with the most sizeable falls with the addition of individual and agent fixed effects.[15] This suggests a shift over time towards individuals or agents with a greater propensity to bunch. Even so the time trend persists, and hence the propensity of specific individuals and agents to bunch at these thresholds has been tending to increase.

Finally, this framework can also be used 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. To examine this, Appendix Table C1 repeats the specifications in columns (3) and (8) in Table 1 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 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 agent fixed effects leads to a substantial attenuation of the coefficient. Persistence in bunching behavior largely reflects persistence in agent, and differences in agent tendencies to bunch. In the sections to come we return to the role of tax agents in explaining bunching, and we can learn from it.

[15]This reflects the fact that the time trends in the more important controls has been relatively modest. Agent prepared 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 Table 1 combined with this increase in electronic returns imply only a modest contribution to the increased bunching propensity.

VI. Why are salient refunds delivered by tax agents?

The prominent role of tax agents in delivering positive, salient tax refunds motivates further investigation. What do agents get out of it? We begin by considering two channels through which agents 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-maximising tax agents.

A. *Empirical framework*

To examine the effect of receiving a positive, salient refund on tax agent 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 agent fee; and a binary variable taking the value of 100 if the individual was with the same agent the following year, and zero otherwise (the percent probability of remaining with your tax agent).

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

$$y_{ijt} = \alpha + \beta_0 b_{ijt} + \sum_{\tau \in T} \beta_\tau \max\{0, b_{ijt} - \tau\} + \gamma \sum_{\tau \in T} \mathbb{1}[b_{ijt} \geq \tau] + \delta X_{ijt} + \varepsilon_{ijt} \tag{8}$$

where T is a set of thresholds in question and the subscripts reflect returns for individual i filed through agent 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 ($T = \{0\}$); and the hundred dollar thresholds up to AUD2500 ($T = \{200, ..., 2500\}$). We estimate this equation for balances within AUD100 of the range of thresholds considered. We also allow for a variety of potentially time-varying individual- and agent-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. It is an interesting 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.

B. *Results*

In Table 2 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 individuals' tenure with a tax agent, occupation and location (all interacted with year); then add agent-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 agent in a given year.

Table 2—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 (AUD)								
	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
Agent-year			X	X			X	X
Individual				X				X

Note: Presents coefficient estimates γ and standard errors from OLS regression estimation of equation (8) on the baseline sample, restricted to agent-prepared 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 agent, occupation-year and location-year fixed effects; agent-year fixed effects; and individual fixed effects. See the note to Table 1 for more information on the occupation and location covariates.

Tax returns just over positive, salient thresholds are not associated with higher fees. Panel A of Table 2 shows no evidence of a positive discontinuity in fees over these thresholds. These are fairly precise null estimates – based on columns (1) and (5) we can reject at the 5% level that the true effect is greater than AUD10 around the zero threshold and AUD0.40 at the hundred thresholds.

In contrast, tax returns just over positive, salient thresholds appear to generate greater loyalty. Panel B of Table 2 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 agent-year fixed effects. One challenge here is that only a very small number of an agents 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 agents delivering positive, salient refunds in response to their clients' preferences. The extra effort to deliver such refunds, or decision to reduce effort once one is secured, is 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 agents to make savings elsewhere. Further evidence that agents are responding to client preferences can be found in the fact that self-prepared returns also exhibit bunching at these thresholds.

Agents do, however, go well beyond self-preparers in their tendency to settle on positive, salient refunds. Appendix Figure B6 shows the (client-weighted) distribution of discontinuities around the zero and hundred dollar thresholds. There is a fat right tail of clients with agents 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 agents are rationally maximising profits, another possibility is that they are themselves 'behavioural' in the sense that they derive utility from delivering particular refunds. Some speculative evidence for this can be found in the distribution of tax agent fees, which typically end in a '0' which is more consistent with satisfying agent preferences for round numbers than responding to their client's preferences. Furthermore, Appendix Figure B6 shows that high-bunching agents are more likely to have fees ending in a '0'.

VII. What effect do high-bunching agents have on tax returns?

We have argued that tax agents deliver positive, salient refunds partly in response to reference-dependent preferences held by their clients. However, an agent'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 II, this would suggest that high-bunching agents should have an impact on returns beyond that required to bunch – they should result in higher claims and balances for their clients more generally.

A. Empirical framework

To examine the effect of high-bunching tax agents on tax returns we use an event study design, looking at individuals who move between tax agents. We restrict attention to individuals who are observed for at least four years with both the old and new agents. To avoid capturing restructures that involve a change in tax agent identifier, we exclude moves where the destination agent receives more than half their clients from, or more than half the clients of, the origin agent.[16] To abstract from moves associated with labour market entry or exit, we further focus

[16]These 'false moves' become more common in robustness exercises where we examine moves that occur alongside large outflows from particular agents.

on individuals with nonzero salary and wage income in each year. The treatment is D_i, the difference in the bunching discontinuity between the new and the old agent. We exclude those moves where the difference in bunching propensities is imprecisely estimated, losing about 15% of moves in the process.[17]

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

$$y_{it} = \alpha_i + \beta_t + \gamma_l + \delta D_i + \sum_{l \in \{-4,\ldots,3\}, l \neq -1} \zeta_l D_i \mathbb{1}_{t=T-l} + \eta X_{it} + \varepsilon_{it} \tag{9}$$

for individual i in year t and event time l (which is equal to zero in the first year with the new agent). 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 equation (9) are the ζ_l. These can be interpreted as the effect of moving between an agent that never bunches (with no discontinuity at the threshold) to an agent 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 agents that differ by one standard deviation in bunching propensity.

Equation (9) allows us to examine the effect of moving to a higher-bunching agent on various dollar outcomes. To examine the effect on individual bunching propensity, we can return to equation (7), 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 are that the difference in bunching propensity between agents 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 agents at times reflect changes in personal circumstances – such as a change in occupation – that may influence both choice of agent 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 agents; and those that are more plausibly exogenous.

On a final note, recent literature has highlighted potential shortcomings in

[17]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 Figure B7 motivates this by showing the cumulative distribution of these standard errors – these choices exclude the tail of imprecisely estimated treatments.

two-way fixed effects estimation of difference-in-differences and event study designs. The most relevant paper in this instance is Callaway, Goodman-Bacon and Sant'Anna (2021), where they consider a difference-in-differences setting with a continuous treatment variable such as this one.[18] 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 this case overweighting 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.

B. Effects of moving between tax agents

Figure 5 shows the effect of a one unit change in agent 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 agents that differ in their propensity to bunch at zero, and the bottom panel illustrates the effect of moving between agents 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 agents. This provides further evidence that most of the differences across agents are due to agent behaviour 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 AUD100 window, an increase in balance of at most AUD100 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. It is also notable that there is little evidence for dynamic effects. This is consistent with the tax agent influencing decisions made at the point of filing but not through the year. In the first year with the tax agent, 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 agent 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 agents influence is predominantly about what happens when returns are filed.

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

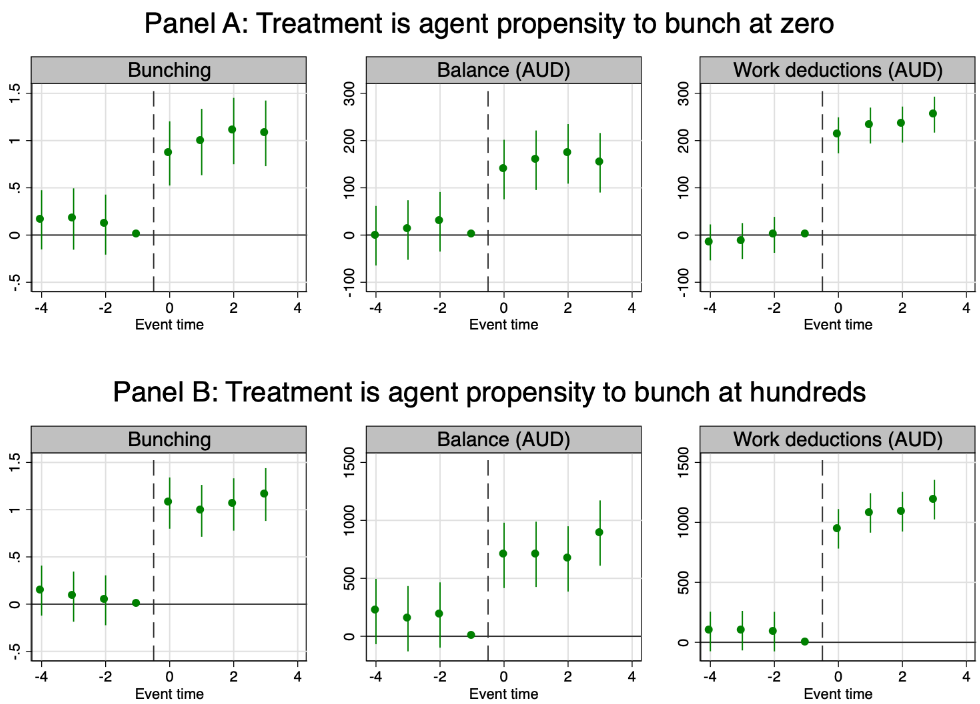

Figure 5. Effect of moving between agents differing in propensity to bunch

Note: Presents coefficient estimates and 95% confidence intervals for ζ_l from equation (9), capturing the effect of a one unit change in agent bunching propensity at time $l = 0$ on individual bunching propensity, balance at assessment and work-related expense deductions.

The dollar amounts in play are also substantial in the context of the variation in agent bunching propensity. Moving to an agent one standard deviation higher in the distribution of bunching propensity would result in an increase in the balance owing to the taxpayer of AUD52 when considering hundred dollar thresholds and AUD41 when considering the zero threshold.[19] Across a population of 15 million taxpayers, this implies an annual fiscal cost of AUD600-750 million, more than two orders of magnitude greater than the mechanical costs of bunching estimated earlier.

For a finer grained look at the results we turn to Table 3. Given the lack of evidence for dynamic effects, we present the estimate for the first year with the new agent. As a test for any pre-trends, we also provide the p-value on the Wald

[19]Based on standard deviations of 0.074 and 0.296 in agent bunching at the hundred and zero dollar thresholds respectively. These standard deviations are disattenuated to adjust for uncertainty in their estimation and are client-weighted.

test that the coefficients prior to the move are jointly equal to zero. Comfortingly, none of these are below typical thresholds.

Table 3—Effect of agent bunching-propensity on tax return

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

Note: Presents coefficient estimate ζ_0 and standard errors from OLS regression of equation (7) (bunching outcomes) or equation (9) (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 agents – and event time is the key variable of interest. We show the coefficient for the first year with the new agent. This can be interpreted as the effect of moving from an agent that never bunches to one that always bunches.

As seen earlier, movers tend to pick up the full difference in bunching propensity, with coefficients of near one on the effect of difference in bunching at a threshold on the discontinuity in the distribution of balances at the same threshold. But the cross effects are mixed. Agents 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, agents 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 agents 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 (2021*a*). 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 3.

Turning to income variables, these effects are less precisely estimated, though there does appear to be a fall in total income on moving to agents 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, common 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 the quasi-experimental approach and audit findings suggests that at least some of this manipulation at the point of tax filing is evasion, rather than the result of more diligent agents. As noted in Section II, a flat cost curve is consistent with a range of interpretations, which may include more thorough or efficient agents facing a lower cost of finding additional legitimate claims, but also more risk-tolerant agents being more willing to make claims that may not hold up in the event of an audit.

C. *Robustness exercises*

As noted earlier, a potential concern with the event study design is that moves between agents 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 agents, 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 negligible effect on our key headline findings. In Appendix Figure B8 we replicate the last two panels of Figure 5, 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 agent, namely we estimate:

$$y_{it} = \alpha_i + \beta_t + \gamma_l + \delta D_i + \zeta D_i \mathbb{1}_{t \geq 0} + \eta X_{it} + \varepsilon_{it} \tag{10}$$

where ζ is now the variable of interest.

In Appendix Table C2 we show that the relationship between the change in agent bunching propensity and changes in key outcome variables is relatively symmetric: it is not purely driven by moves to higher or lower bunching agents. 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 agent are more likely to pick up some tax filing behaviours than they are to lose them when leaving such an agent; 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 agents.

As a final exercise, we explore how our results change as we hone in on people leaving their tax agent amid increasingly large outflows. These people are more likely to be leaving because of the retirement of a particular agent 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 Figure B9 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 an agent) to those near closure (100% of clients leaving an agent). This provides some comfort that moves driven by a purposeful decision to change agents do not drive our results.

VIII. Conclusion

A large literature has established the tendency of individuals to display reference-dependent preferences. Australian taxpayers are no different, being more likely to end up just owed money from the ATO than just owing it, and just above AUD10, AUD100 and AUD1000 thresholds in the amount they are owed. This cements the relevance of such preferences, and in a setting with both high stakes and repeat transactions.

The role of tax agents in delivering positive, salient balances is particularly notable. Agents are more likely to deliver such balances and have been increasing over time in this tendency. While the typical agent-prepared return is more than twice as likely to bunch at the thresholds examined as a self-prepared return, there is a fat right tail of high-bunching agents. Partly this may reflect client preferences, and the resulting benefits to agents such as client loyalty, though we cannot rule out behavioural agents. We don't observe the potential costs borne by clients of such agents, such as whether such behaviours enable higher fees across all their clients or expose their clients to greater risk.

Finally, we show how bunching at these reference points can act as a proxy for other policy-relevant parameters, namely the shape of the curve describing the cost of reducing an individuals tax liability. All else equal, a flatter cost curve will result in more tax deductions being claimed, but also more bunching. The fiscal costs of bunching itself are relatively modest, but using an event study design we show that high-bunching agents have a far larger influence on tax returns. Scaled up to the full population, moves to agents one standard deviation higher in the distribution of bunching propensity would result in a fiscal cost of between AUD600-750 million a year, over two order of magnitude greater than the direct cost of bunching. The return items most influenced by high bunching agents are consistent with those most at risk of evasion – items where audits are costly – and line up with those also flagged in ATO random enquiry programs. This suggests behavioural notches may provide useful quasi-experimental evidence to tax authorities of the where, to what extent and by whom discretion is exercised at the point of tax filing.

Funding

None of the authors received any funding for this work

Conflict of Interest

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

REFERENCES

Abhayaratna, Thomas, Andrew Carter, and Shane Johnson. 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, Patricia M Dechow, Devin G Pope, and George Wu. 2017. "Reference-dependent preferences: Evidence from marathon runners." *Management Science*, 63(6): 1657–1672.

Allingham, Michael G, and Agnar Sandmo. 1972. "Income tax evasion: A theoretical analysis." *Journal of Public Economics*, 1(3-4): 323–338.

Australian Taxation Office. 2021*a*. "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: 2021-09-30.

Australian Taxation Office. 2021*b*. "Taxation Statistics, 2018-19." *https://www.ato.gov.au/About-ATO/Research-and-statistics/In-detail/Taxation-statistics/Taxation-statistics-2018-19/*, Accessed: 2022-04-05.

Battaglini, Marco, Luigi Guiso, Chiara Lacava, and Eleonora Patacchini. 2019. "Tax professionals: tax-evasion facilitators or information hubs?" National Bureau of Economic Research.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2021. "Difference-in-Differences with a Continuous Treatment." *arXiv preprint arXiv:2107.02637*.

Dube, Arindrajit, Alan Manning, and Suresh Naidu. 2018. "Monopsony and employer mis-optimization explain why wages bunch at round numbers." National Bureau of Economic Research.

Engström, Per, Katarina Nordblom, Henry Ohlsson, and Annika Persson. 2015. "Tax compliance and loss aversion." *American Economic Journal: Economic Policy*, 7(4): 132–64.

Lacetera, Nicola, Devin G Pope, and Justin R Sydnor. 2012. "Heuristic thinking and limited attention in the car market." *American Economic Review*, 102(5): 2206–36.

Rees-Jones, Alex. 2018. "Quantifying loss-averse tax manipulation." *The Review of Economic Studies*, 85(2): 1251–1278.

PROOF OF PROPOSITION 1

Proposition 2. *Consider a positive balance which has a largest divisor τ in the set $\{10, 100, 1000\}$ (i.e. 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 $\tau \in \{0, 10, 100, 1000\}$

(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:

$$c'(\beta) = v'(\beta) = v' \tag{A1}$$

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

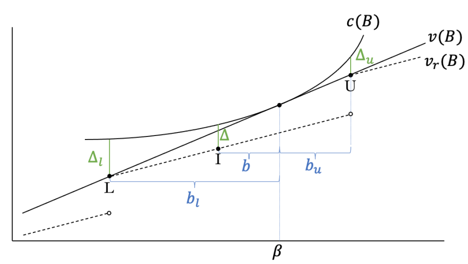

FIGURE A1. TAXPAYER COST AND BENEFIT CURVES

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

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

$$\begin{aligned} c(\beta + b) - v(\beta + b) &= c(\beta) + c'(\beta)b + \tfrac{1}{2}c''(\beta)b^2 - v(\beta) - v'(\beta)b \\ &= (c(\beta) - v(\beta)) + (c'(\beta) - v'(\beta))b + \tfrac{1}{2}c''(\beta)b^2 \\ &= \tfrac{1}{2}c''b^2 \end{aligned} \tag{A2}$$

We now consider the change to reference-dependent preferences. There are three possible points where costs minus benefits will be minimised – 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 characterised by first order condition:

$$\begin{aligned} c'(\beta+b) &= v_r'(\beta+b) \\ \Rightarrow \qquad c'(\beta)+c''(\beta)b &= (1-\theta_{1000})v' \\ \Rightarrow \qquad b &= \frac{\theta_{1000}v'}{c''} \end{aligned} \tag{A3}$$

where we have used the original first order condition (A1). This solution will exist wherever $b < b_L$.

It follows from equation (A2) that our three possible minima are either:

$$\Delta_L = \tfrac{1}{2}c''b_L^2 \tag{A4}$$

$$\Delta_R = \tfrac{1}{2}c''b_R^2 \tag{A5}$$

and

$$\begin{aligned} \Delta_I &= \tfrac{1}{2}c''b^2 + (b_L - b)\theta_{1000}v' \\ &= \tfrac{1}{2}c''\left[b^2 + 2(b_L-b)\frac{\theta_{1000}v'}{c''})\right] \\ &= \tfrac{1}{2}c''\left[b^2 + 2(b_L-b)b)\right] \end{aligned} \tag{A6}$$

Whether Δ_L or Δ_R is smaller 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:

$$b = b_L \pm \sqrt{b_L^2 - b_U^2} \tag{A7}$$

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. Equation $A3$ and analagous expressions for the smaller thresholds then establishes the proof.

□

Additional Charts

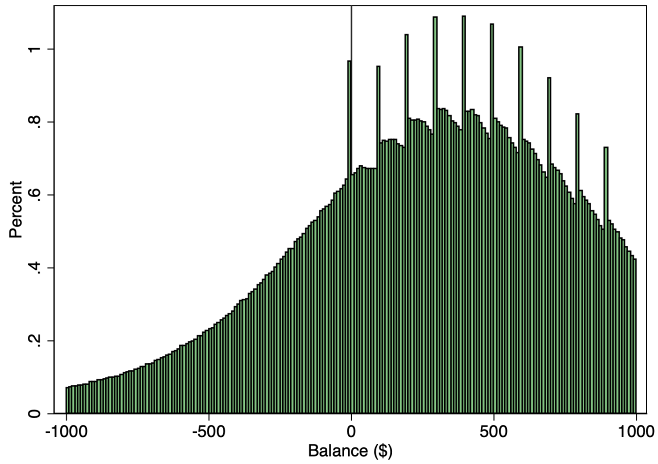

Figure B1. Simulated distribution of balances

Note: Plotted for 10 million observations with $v = 1$ and a quadratic cost function with linear term normally distributed with mean $-frac12$ and standard deviation 2 and quadratic term uniformly distributed over [0.001,0.011]. Only ten per cent of the population has reference dependent preferences – for those with them $theta_0$ is uniformly distributed over [0,0.2] and $theta_{100}$ is uniformly distributed over [0,0.1].

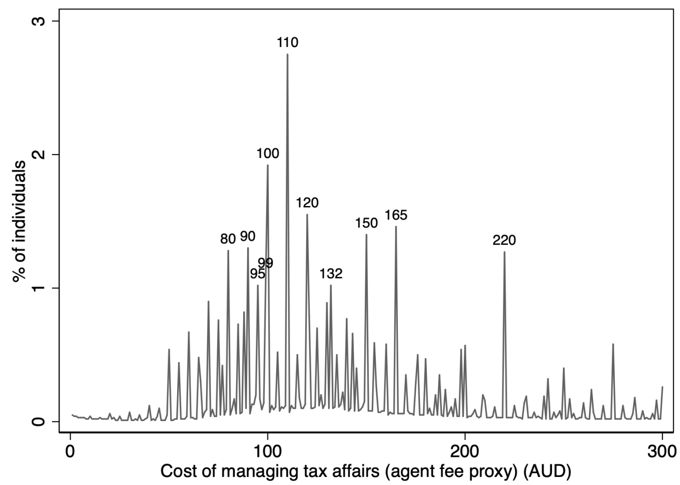

Figure B2. Distribution of cost of managing tax affairs, 2000-2018

Note: Distribution of the deduction for the cost of managing tax affairs, which we use as a proxy for tax agent fees in the year prior. Based on the full ALife sample, restricted to those with an agent-prepared return and with a non-missing deduction for the cost of managing tax affairs in the following year. Percentages are for each AUD1 bin and are based on the full distribution rather than the window shown; bins with more than 1% of the sample are labelled. Around 40% of those with an agent-prepared 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.

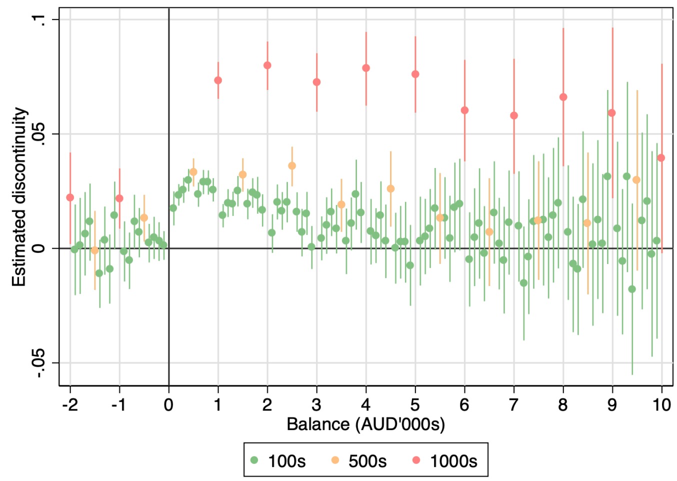

Figure B3. Estimated discontinuities at specific thresholds, 1991-2018

Note: Estimated discontinuity δ in the normalised count around specific AUD100 thresholds, with 95% confidence intervals. Based on estimation of equation (5) in a window AUD50 either side of the given threshold. For small values of δ an individual is 2δ% more likely to be immediately above the threshold than below it.

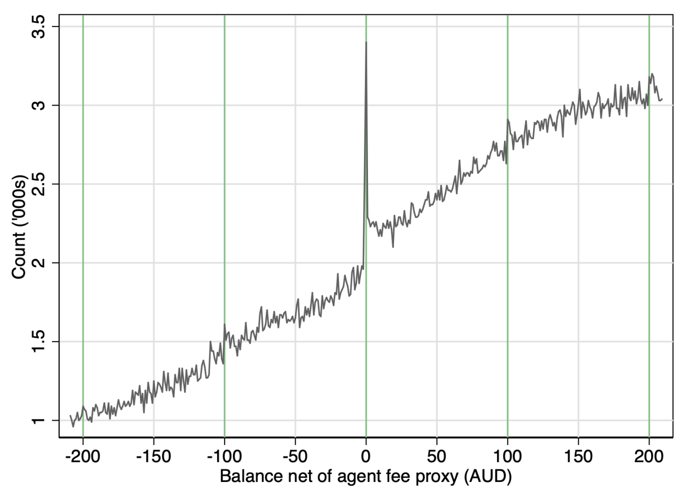

Figure B4. Distribution of balance of assessment net of tax agent fee proxy, 1999-2017

Note: Distribution of the balance of assessment net of tax agent fee proxy. Based on the baseline sample of 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 an agent-prepared return and claiming a nonzero deduction for the cost of managing tax affairs in the following year. Counts are for each AUD1 bin. Graph captures 870,000 tax returns over the 1999-2017 income years (since the tax agent fee proxy is only available from 2000 through to 2018).

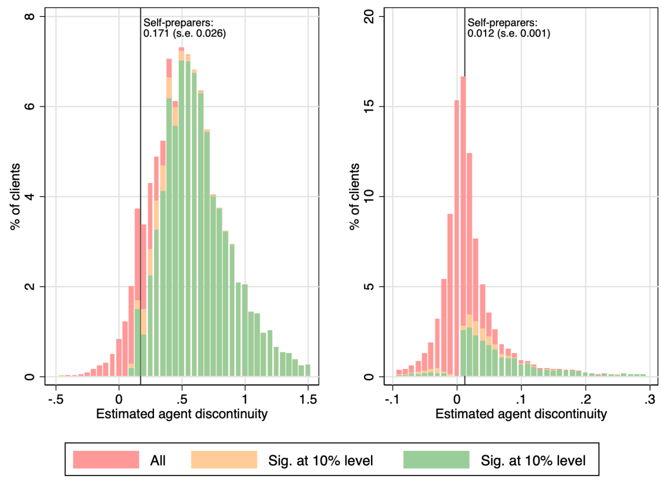

Figure B5. Distribution of agent-specific discontinuities across clients of tax agents, 1991-2018

Note: Distribution of the agent-specific estimates of the discontinuity δ in the normalised count around either the zero or hundred dollar thresholds. Distributions are client-weighted and hence expressed as the percentage of clients of tax agents falling into given bins. We further show whether the estimated discontinuity is statistically significant at either the 5% or 10% level.

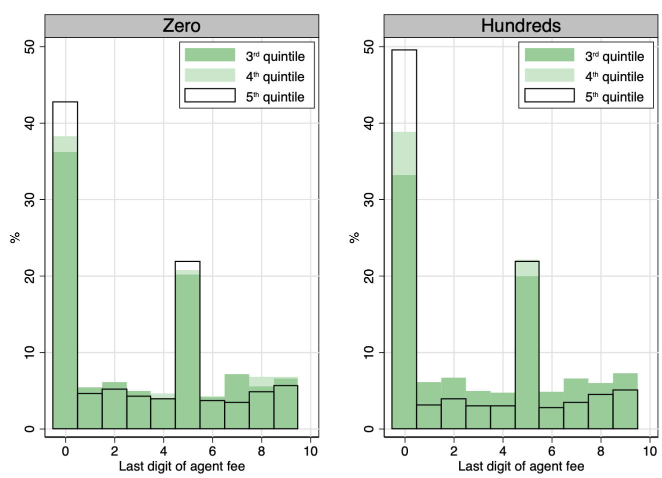

Figure B6. Distribution of last-digit of agent fee proxy by agent bunching propensity, 1999-2017

Note: Distribution of the last-digit of the agent fee proxy by agent bunching propensity...

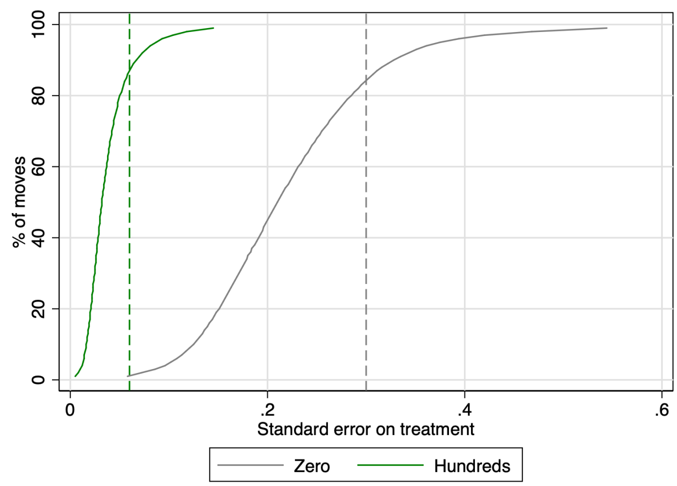

Figure B7. Cumulative distribution of standard error on the treatment

Note: Shows the cumulative distribution of the standard error on the treatment for those in the movers sample.

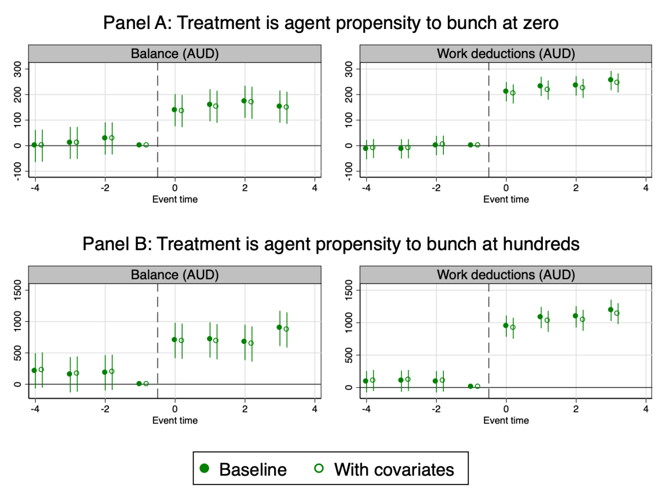

Figure B8. Effect of moving to a higher bunching agent, with and without additional covariates

Note: Presents coefficient estimates and 95% confidence intervals for *zeta* from OLS regression estimation of equation (10), 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 1 for more information on the occupation and location covariates.

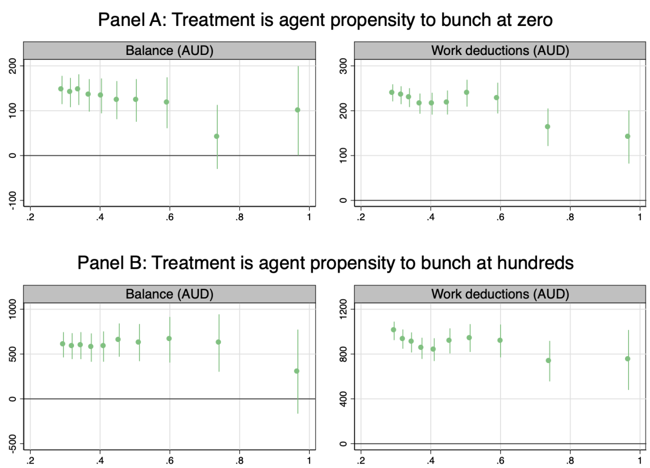

Figure B9. Effect of moving to a higher bunching agent as part of large outflows from original agent

Note: Presents coefficient estimates and 95% confidence intervals for *zeta* from OLS regression estimation of equation (10), restricting to moves that were part of increasingly large outflows from the original agent. We begin by splitting the full sample of movers into deciles based on the proportion of all the original agents 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 agent outflows.

Additional Tables

Table C1—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)	0.0183*** (0.0016)	0.0071*** (0.0016)
Fixed effects				
Location	X	X	X	X
Occupation	X	X	X	X
Agent		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 equation (7) on the baseline sample. Columns (1)-(2) examine the zero-dollar threshold, while columns (3)-(4) examine the hundred-dollar thresholds $\{100, ..., 2500\}$. Columns (1) and (3) replicate columns columns (3) and (8) respectively in Table 1 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 agent fixed effects.

Table C2—Effect of moving to a new tax agent, by sign and size of change in agent 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 equation (10), where ζ is allowed to vary with the size of the treatment D_i. Namely, we categorise the moves on the basis of D_i into positive and negative moves (those to higher or lower bunching agents) and also by whether that are above (larger) or below (smaller) the median magnitude of 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.