Diffuse Bunching with Lumpy Incomes: Theory and Estimation*

Santosh Anagol Allan Davids
Benjamin B. Lockwood Tarun Ramadorai†

July 1, 2022

Abstract

We study the performance of the bunching-based elasticity estimator when income adjustments are lumpy. In our parsimonious model, taxpayers choose their preferred income from random opportunity sets. The model features the standard elasticity of taxable income and a single additional "lumpiness parameter," which can be jointly estimated using maximum likelihood. This model can match key patterns that are inconsistent with the conventional bunching estimator, including diffuse bunching around kinks and notches in the tax schedule and positive mass above tax notches. When incomes are lumpy, simulations demonstrate that the conventional estimator is biased, underestimating the true elasticity by as much as 50%, and incorrectly sized, with the true parameter lying inside bootstrap 95% confidence intervals with less than 10% probability. We apply this method to administrative tax data on small businesses in South Africa, recovering moderate elasticities at higher incomes and large elasticities at low incomes. Firms with paid tax practitioners exhibit sharper bunching, driven primarily by a lower lumpiness parameter rather than by a different income elasticity.

*We wish to acknowledge the National Treasury of South Africa for providing us with access to anonymized tax administrative data. We thank Analytics at Wharton, the Penn Wharton Budget Model, and the Wharton Dean's Research Fund for funding support. The views expressed in this paper are our own, and do not necessarily reflect the views of the National Treasury of South Africa. We are grateful to Wian Boonzaaier, Jacob Mortenson, Alex Rees-Jones, Joel Slemrod, David Thesmar, Andrew Whitten, Eric Zwick, and conference and seminar participants at Economic Research South Africa (ERSA), the European Bank for Reconstruction and Development, Imperial College, LAGV 2021, LMU Munich, CREST, the Toulouse School of Economics, and the South African Revenue Services for helpful comments and to Afras Sial and Laila Voss for excellent research assistance. All errors are our own. This paper subsumes and replaces the working paper titled "Do firms have a preference for paying exactly zero tax?"

†Anagol: Wharton School, University of Pennsylvania. Email: anagol@wharton.upenn.edu. Davids: School of Economics, University of Cape Town. Email: allan.davids@uct.ac.za. Lockwood: Wharton School, University of Pennsylvania and NBER. Email: ben.lockwood@wharton.upenn.edu. Ramadorai: Imperial College London and CEPR. Email: t.ramadorai@imperial.ac.uk

1 Introduction

Income distributions often exhibit excess mass around tax kinks, where the marginal tax rate steps up. Saez (2010) influentially demonstrated how the size of the "bunching mass" around a tax kink can be mapped into an estimate of the elasticity of taxable income, a key parameter in many models of optimal policy and taxation, using a formula derived from a simple model of taxpayer optimization. Kleven and Waseem (2013) extended this approach to handle notches, where the tax level (rather than the marginal rate) steps up.

One concern about this "bunching estimator" approach is that empirical bunching patterns look different from the predictions of the theoretical model used to derive the elasticity formula. Whereas the model predicts an atom of excess mass precisely at the kink point and zero density in a range of dominated incomes above notches, empirically observed bunching is diffuse around kinks and has positive mass at dominated incomes. Such diffusion presents a challenge for empirical estimation, because it obscures the distinction between bunching mass induced by the tax schedule and mass induced by the parameters of the underlying distribution of preferences or abilities. A large body of work has proposed workarounds for isolating the bunching mass component, which is then converted into an elasticity using the usual formula—effectively proceeding as if the estimated diffuse mass were counterfactually located precisely at the kink, while leaving unmodeled the frictions that produce diffusion.[1]

This paper seeks to answer two related questions. First, does this prevailing bunching estimator approach actually recover the true elasticity of taxable income in the presence of frictions that produce diffusion? Second, is useful information discarded when we collapse the rich pattern of income distortions around a kink or notch into a scalar estimate of bunching mass?

We address these questions using a simple model in which diffusion is the result of income "lumpiness": taxpayers select their preferred income from a random set of sparse income opportunities. The interaction of optimizing agents and sparse income opportunity sets leads to detailed and novel predictions about the shape of the bunching mass around kinks and notches. For example, around a kink, the unavailability of the kink point for many agents leads to primarily symmetric diffuse bunching, as agents optimize to get as close to the kink as feasible. Around a notch, the model predicts greater diffuse bunching to the left than right of the notch point, as agents attempt to avoid the dominated region to the right of the notch leads them to choose feasible points in some cases quite far to the left of the notch point.

The lumpiness in income opportunities we study could be generated by the discrete nature of job offers, work shifts, reporting opportunities, or other taxable income determinants.

[1] See Chetty et al. (2011), Mortenson and Whitten (2020), Bosch, Dekker and Strohmaier (2020), Dekker and Schweikert (2021). Kleven (2016) presents an extensive review of the bunching estimation literature.

Although income lumpiness has previously been mentioned as a potential driver of diffuse bunching, it has received little attention relative to other sources of friction such as income uncertainty, imperfect targeting, and adjustment costs. Nevertheless, we show that our model of income lumpiness predicts bunching patterns around kinks and notches that are strikingly similar to those observed in practice, in ways that are not replicated by many other models of frictions.

Using this model of income lumpiness, we find that the answer to the first question, i.e., whether the conventional approach performs well in the presence of frictions, is generally "no." Simulations demonstrate that in the presence of income lumpiness, the conventional bunching-based elasticity estimator underestimates the true elasticity by more than 50% in some specifications, with 95% confidence intervals containing the true elasticity value less than 10% of the time. We present an alternative estimation strategy, based on maximum likelihood methods, to recover the parameters of the model that we propose. This approach succeeds in the same simulations at recovering the true elasticity with accurate confidence intervals.

We find the answer to the second question, i.e., whether the conventional approach discards economically useful information, is "yes." Here, we are motivated by our empirical application which involves understanding the bunching behavior of small businesses in South Africa around three tax kinks. The histogram of taxable incomes around each kink is displayed in Figure 1, where we see evident patterns of diffuse bunching at each bracket threshold.

In addition to estimating the elasticity of taxable income in this setting, our proposed maximum likelihood estimation strategy allows us to glean two additional economic insights that would be unavailable under the conventional approach. First, whereas the bunching patterns in Panels (b) and (c) of Figure 1 look similar to the those usually observed around tax kinks, the pattern in Panel (a)—around the bottom kink—resembles the shape usually observed around a tax *notch* (see, e.g., Kleven and Waseem, 2013) with excess mass to the left of the tax bracket threshold and missing mass to the right. Such behavior could arise from unobserved costs or frictions, whether real or behavioral, and a natural question is how big this "as-if" money-metric notch value is. The conventional approach is not suited to provide an answer, because it requires specifying the notch size (and thus the dominated income region) as an input in order to estimate the elasticity. In contrast, our maximum likelihood estimator can estimate the notch value, which is identified by the *asymmetry* in the diffusion of bunching mass around the threshold; this feature of the data is discarded when the bunching pattern is reduced to a single number under the conventional approach, but exploited as a pattern that naturally results when agents optimize in the face of lumpy income choices.

The second economic insight relates to heterogeneity in bunching responses across different groups of taxpayers. There are notable visual differences in bunching patterns among

firms that employ a paid tax preparer, relative to those that do not: bunching in the former group appears more tightly concentrated around the kink. Despite these observable differences, the conventional approach is unable to discern differences in underlying model parameters between the two groups, producing elasticity estimates that are statistically indistinguishable across the groups. In contrast, our proposed method suggests that the elasticity is in fact statistically higher among firms without paid tax preparers, but such firms also exhibit greater diffusion in their bunching, consistent with a coarser degree of income targeting. This heterogeneity is identified by the differences in the diffusion in the bunching mass for each firm type—information that is discarded by the conventional bunching estimator method.

The idea that diffuse bunching is driven by lumpiness in income opportunities is not new. Indeed, Saez (1999)—the working paper that preceded Saez (2010)—presented simulations demonstrating that income lumpiness process can produce diffusion that qualitatively resembles empirical distributions.[2] Relative to this work, our primary theoretical contribution is a parsimonious model of the income opportunity process that depends on only a single additional "lumpiness parameter," which represents the expected difference between adjacent income opportunity draws. We analytically characterize the continuous income density around a tax kink or notch as a function of the structural income elasticity, the lumpiness parameter and the underlying ability density. Although the underlying density is unobserved, identification can be achieved by assuming that the density is locally smooth in the vicinity of the tax bracket threshold. This assumption resembles the practice—pioneered by Chetty et al. (2011) and often adopted in the bunching estimation literature—of fitting a smooth polynomial to the observed income histogram outside of a visually specified "bunching window" around the tax kink.[3] The strengths and limitations of this identifying assumption are explored by Blomquist et al. (2021).

In addition to producing less biased elasticity estimates and additional insights about economic behavior, our estimation approach has a number of practical advantages relative to the conventional approach. First, it does not require the analyst to specify bounds for a bunching window around the bracket threshold, whether by visual inspection (Chetty et al., 2011) or algorithmically (Dekker and Schweikert, 2021). As a result, our modification reduces the total number of model parameters, and it replaces economically irrelevant parameters (bunching window bounds) with one that is potentially informative (the money-metric lumpiness parameter). Second, estimation results from our model are more robust to misspecification in the

[2] One simulation method from Saez (1999), produced by drawing a random income opportunity set from a uniform distribution around each agent's optimal frictionless income, is similar in spirit to the model we present in Section 2, and to the simulations in recent work by Kosonen and Matikka (2020). Beffy et al. (2019) presents a similarly motivated model in which taxpayers choose between two income opportunities.

[3] The approach proposed here is somewhat more internally consistent, because even in the frictionless model, the ability distribution is more plausibly smooth around the tax bracket threshold than the income distribution outside the bunching window (see Figure 3b).

parametric form of the counterfactual income density. As we show in Section 3, under the conventional approach, when the polynomial is allowed to be more flexible, thin tails of the bunching mass that spill outside of the specified bunching window tend to draw the polynomial "upward" into the bunching mass, biasing down the estimated elasticity. In contrast, because our model endogenously produces diffuse bunching mass, the presence of that mass far from the bracket threshold does not distort the estimated polynomial.

Although this paper focuses on lumpiness in the income process, that is not the only potential source of friction that could produce diffusion in bunching around tax kinks and notches. Others which have received attention include imperfect income targeting due to uncertainty (also simulated in Saez, 1999), dynamic adjustment frictions (Chetty et al., 2011; Chetty, 2012; Gelber, Jones and Sacks, 2020; Mavrokonstantis and Seibold, 2022) and multiple-type models in which some taxpayers are inelastic to the threshold (Kleven and Waseem, 2013). These frictions are not mutually exclusive, and in practice all may play a role in contributing to observed diffusion. We focus exclusively on lumpiness for two reasons. First, these alternative models of frictions have difficulty matching key features of the data observed in our empirical setting (namely, the persistence of diffuse bunching over time, and the pronounced peaks in the income density at the income tax thresholds), suggesting that achieving a closer fit to the data will require some degree of lumpiness. (We describe the inconsistencies between observed bunching patterns and other models of frictions in Appendix A.1.) Second, because the addition of our simple model of lumpiness provides a remarkably close fit to the observed cross-sectional distributions, in the name of parsimony we refrain from including other frictions to be estimated. More generally, we are unaware of any papers that systematically compare the ability of these disparate models to match observed bunching patterns, or papers that examine the ability of the conventional bunching estimator to recover the true elasticity parameter in the presence of frictions, as we do in Section 3.2.

In addition to building on the theoretical and empirical literature on bunching estimators and income frictions, this paper is related to three other areas of the literature. First, our results on the estimation of the tax notch value at statutory kink points relates to the literature on behavioral frictions and misperceptions about the tax code. Rees-Jones (2018) uses bunching behavior around the threshold at which taxpayers face a net refund or balance due in order to quantify their degree of loss aversion. On the question of confusing average and marginal tax rates, Rees-Jones and Taubinsky (2020) experimentally study misperceptions of the income tax code, finding that a substantial share of respondents "irons," misinterpreting an average tax rate as the relevant marginal rate. Outside the domain of taxes, Ito (2014) presents evidence that consumers respond (at the margin) to average rather than marginal electricity prices. Using exogenous variation in worker knowledge about a notch in the Norwegian income tax system,

Kostøl and Myhre (2021) estimate that at least 30% of estimated optimization frictions are due to workers' imperfect knowledge about the tax system.

Second, our empirical application contributes to the literature measuring behavioral responses to taxation in developing economies. These include the subset of papers estimating the elasticity of *corporate* taxable income (e.g., Devereux, Liu and Loretz, 2014), which is a particularly important parameter in emerging market economies, given their greater relative reliance on the corporate income tax base.[4] For examples, see Best et al. (2015) in Pakistan, Bachas and Soto (2021) in Costa Rica, and, as mentioned above, Boonzaaier et al. (2019) in our setting of South Africa.

Third, the method that we propose is more generally applicable to settings outside of taxation where diffuse bunching is prevalent in distributions. Bunching of this type is evident in the distribution of job tenure at retirement age, which has been attributed to the effects of financial incentives in retirement decisions (Manoli and Weber, 2016). Such bunching also shows up in the distributions of the intraday timing of demand for mobile phone services (Grubb and Osborne, 2015), educational test scores for both students (Diamond and Persson, 2016; Dee et al., 2019) and teachers (Brehm, Imberman and Lovenheim, 2017), and realized sale prices for houses (Andersen et al., 2022). In many of these settings, diffuse bunching is attributed to "optimization frictions." Discreteness in the choice variable of the type that we model is a simple way to capture such optimization frictions, though the interpretation of the lumpiness parameter will depend on the setting to which our method is applied.

The rest of the paper proceeds as follows. In Section 2, we present our baseline model of lumpiness, we characterize the resulting observable income density, we and describe the maximum likelihood estimation method. In Section 3, we present simulations to compare our estimation method with the conventional approach. In Section 4, we describe our empirical application to small businesses in South Africa, and present our estimation results. Section 5 concludes.

2 Model

2.1 Baseline bunching model with continuous income choice

Our starting point is the canonical "bunching estimator" presented in Saez (2010), where taxpayers have the following utility function:

[4] Gordon and Li (2009) document that emerging economies rely more on corporate income taxes than developed economies.

$$u(c,z;n) = c - \frac{n}{1+1/e} \cdot \left(\frac{z}{n}\right)^{1+1/e}. \tag{1}$$

In the setting of individual taxpayers, c and z are interpreted as consumption and pre-tax earnings from labor effort, the latter of which is assumed to be observable to the government. Taxpayers are heterogeneous in their income-earning ability indexed by $n \in (0,\infty)$, with distribution $F(n)$ and density $f(n)$. An n-type taxpayer chooses the level of z that maximizes utility in equation (1), subject to their budget constraint:

$$c = z - T(z), \tag{2}$$

where $T(z) = a + tz$ is a (locally) linear income tax.[5] Type n's constrained-optimal level of income $z^*(n)$ satisfies the following first-order condition:

$$z^*(n) = n(1-t)^e. \tag{3}$$

Thus n can also be interpreted as an individual's preferred income under a laissez-faire tax system with $t = 0$. In the standard model with continuous income choice, each type chooses their constrained optimal income given the tax rate t. Equation (3) satisfies $\frac{d\ln z^*(n)}{d\ln(1-t)} = e$, so that e can be interpreted as the elasticity of taxable income with respect to the marginal net-of-tax rate $1-t$.

In the bunching literature, this model is used to estimate the elasticity e based on the behavior of the empirical income density, denoted $h(z)$, around a threshold between two different known linear income tax functions. We employ the following notation to describe the piecewise-linear tax function $T(z)$ around a bracket threshold z^{thr}:

$$T(z) := \begin{cases} T_0(z) = a_0 + t_0 z & \text{if } z \le z^{thr} \\ T_1(z) = a_1 + t_1 z & \text{if } z > z^{thr} \end{cases} \tag{4}$$

The identifying assumption underlying this strategy is that the density of types $f(n)$ is smooth across the range of types with earnings near z^{thr}, implying that observed distortions away from a smooth income density in the vicinity of z^{thr} can be attributed to the change in tax rates at the bracket threshold. In the case of a convex "kink" where the marginal tax rate increases at the bracket threshold ($t_0 < t_1$) but the tax level is continuous ($T_0(z^{thr}) = T_1(z^{thr})$), the standard model that we have described predicts excess mass in the income density at the threshold z^{thr}.

[5] We do not model the choice of whether to file a tax return (i.e., an extensive margin response). See Pollinger (2021) for a bunching model that attempts to estimate both intensive and extensive margins combining both bunching and regression kink design models.

Figure 2 illustrates income choices induced by a progressive tax kink for several selected types of taxpayers, as well as their counterfactual incomes under the linear tax $T_0(z)$. Employing equation (3), taxpayers with n between $z^{thr}/(1-t_0)^e$ and $z^{thr}/(1-t_1)^e$, often called the "marginal non-buncher" and the "marginal buncher", respectively, all choose to earn exactly z^{thr}. They are denoted by types b and c in Figure 2. The marginal buncher's income choice under the kinked schedule is the same as it would be if the linear tax $T_1(z)$ applied everywhere, and thus the marginal buncher's income change ($z_1^*(c)$ vs. $z_0^*(c)$) in response to the difference in tax rates (t_1 vs. t_0) identifies the income elasticity e.

Figure 3 illustrates the observed income density around a tax kink under this model. In Panel (a), red points represent income choices for discrete taxpayer types between the marginal non-buncher and the marginal buncher; the lower panel illustrates the induced probability density function, with bunching types stacking up at the bracket threshold. Types above the marginal buncher reduce their incomes to a new interior optimum, resulting in compression of income choices to the right of z^{thr} relative to the counterfactual choices under the linear tax $T_0(z)$. Panel (b) illustrates this behavior under a continuous type density, which exhibits an atom of mass at the bracket threshold, and a discontinuity in the continuous density around the threshold due to compression.[6] The empirical strategy of Saez (2010) amounts to estimating the "bunching mass" around the bracket threshold and using it to infer the marginal buncher's counterfactual income z_0^*. This process is described in detail in Section 3.2; equation (21) reports the key formula for this inference.

Figure 4 illustrates the observed income density in the presence of *notch*, where the tax level rises discretely at the bracket threshold.[7] As described in Kleven and Waseem (2013), in the presence of a notch, the threshold income level z^{thr} strictly dominates incomes immediately above it, which require greater effort but produce lower net income. This creates a "hole" in the density immediately to the right of the threshold.

In contrast to the model-predicted densities in Figures 3 and 4, empirical income histograms around bracket thresholds features bunching that is diffuse and, in the case of notches, they exhibit positive mass in the dominated income region above the threshold. (See Saez, 2010 and Kleven and Waseem, 2013 for examples.) We now turn to a model of frictions that can produce such patterns.

[6] The density to the right of z^{thr} also shifts leftward due to the kink. Under a uniform type density, this shift does not affect the density; however, if the type density is decreasing in the vicinity of the bracket threshold, then the upward discontinuity in the income density at z^{thr} may be dampened or reversed.

[7] By convention, and in most empirical settings, the threshold z^{thr} is subject to $T_0(z)$ rather than $T_1(z)$ in a notched schedule.

2.2 Bunching with lumpy income choices

We assume that income is earned from discrete—or "lumpy"—opportunities, which come in a continuum of sizes but are encountered only probabilistically. Taxpayers choose their preferred income from the sparse set of income opportunities they encounter.

Such lumpiness could arise for a variety of reasons. A salaried worker might select their employment position from a discrete set of opportunities with different offered incomes. A small business might select jobs from a discrete set of opportunities that it has encountered. An hourly worker might adjust income by working more or fewer shifts, each of which has a delivers a lumpy change in income. Lumpiness might also extend to income reporting responses, rather than real adjustments, as when a firm targets their income by advancing some of their (lumpy) payments into the current tax year, or claiming some (lumpy) expenditures as deductible, as discussed in Rees-Jones (2018), and covered extensively in the accounting literature, e.g., Kothari, Leone and Wasley (2005).

Formally, we assume that each agent faces an *income choice set* comprising a sparse set of potential incomes from which they select the income that maximizes their utility. This modification means that taxpayers who share the same type n but face different income choice sets will make different income choices, thus creating income heterogeneity among n types. Correspondingly, this model implies that agents who ultimately choose a given income have a range of different underlying abilities n.

To operationalize this idea, we assume that within each type n there is a continuum of individuals facing different income choice sets, thus producing a continuous type-conditional incomes density, denoted $g(z|n)$. The observed income density $h(z)$ can then be written as the integral across all types of agents who choose to earn a given income z:

$$h(z) = \int_0^\infty g(z|n) f(n) dn. \tag{5}$$

To characterize the type-conditional income density $g(z|n)$, we must specify a process through which the income choice sets are drawn. We adopt a simple and tractable assumption: each potential income is encountered with a constant probability. To build intuition, consider the case in which all income opportunities come in dollar integer dollar sizes, and any particular dollar amount is encountered with an independent probability of 1%. Then among all taxpayers whose preferred continuous income is \$10,000, only 1% will have that option in their income choice set. Similarly, 1% of these taxpayers will have the option to earn \$9,999, 1% will have the opportunity \$10,001, etc. In this example, this structure implies that among all such taxpayers, the average distance between any two adjacent income opportunities is \$1/0.01 = \$100. Intuitively, any \$1,000 income range will contain, on average, ten income opportunities, with an

expected distance of \$100 between adjacent opportunities.

For our analytic characterizations below, it is convenient to allow for a continuum of job sizes. This can be viewed as a limiting case of discrete models like the one just described. Suppose jobs come in sizes of integer *cents*, rather than dollars, and the probability of encountering an opportunity of any particular dollar-and-cents income level is 0.01% (rather than 1%). Then the average distance between any two adjacent income opportunities remains \$100 = \$0.01/0.0001.[8]

The limiting case of such a model as the grid of job sizes becomes arbitrarily fine is a *Poisson process*, with an arrival rate of $\lambda = 0.01$. Like in the above discrete examples, the average distance between adjacent opportunities ("arrivals", in the typical language of Poisson processes) is $1/\lambda = \$100$. We use μ to denote this average distance, which we call the income "lumpiness parameter." The standard model with continuous adjustment corresponds to the special case in the limit as $\mu \to 0$.

The assumption that income opportunity sets arise from a Poisson process is important for our model and does impose restrictions; as with all models, these restrictions are an oversimplification of reality. That said, we do view this assumption as an attractive modelling choice for three related reasons, which we discuss below.

First, under this model, each agent's income opportunities are drawn with uniform probability in the vicinity of their preferred income choice. In this respect, the assumption aligns with the simulations in Saez (1999) and in Kosonen and Matikka (2020), wherein income choice sets are simulated by drawing from a uniform distribution in a specified neighborhood of agents' preferred incomes $z^*(n)$. By using a Poisson process, we retain the feature of a uniform probability across income opportunity values, while reducing the number of required parameters, because we need not specify bounds on a neighborhood around the preferred income. Moreover, this assumption means that the income process can be modeled independently from the agent's income preferences.

Second, the Poisson process and the parameter μ have clear and transparent interpretations. This can be illustrated by returning to the examples discussed—discussed at the beginning of this subsection—of discrete income choice sets. If the taxpayer is an individual choosing between different positions with annual salaries, this model corresponds to a case where any particular income offer around the individual's preferred income is equally likely, and the dollar-denominated distance between adjacent offers is $\$\mu$. If the taxpayer is a small business taking on jobs throughout the year, our setting corresponds to one in which the business faces

[8] Intuitively, any \$1,000 income range comprises 100,000 dollars-and-cents values, each with probability 0.0001, and will once again contain an average of ten income opportunities, with an expected distance of \$100 between adjacent opportunities.

a discrete set of choices about which final job to take in the present fiscal year, with an average difference of $\$\mu$ between job sizes. In the context of reporting responses, when a business decides which income sources to shift into the present fiscal year, this model corresponds to a setting with a discrete set income sources (e.g., checks to deposit) with an average difference of $\$\mu$ between them. Also in the domain of reporting responses, it can represent a discrete set of auditors prepared to ratify different incomes, with an average difference of $\$\mu$ between them (DeFond and Subramanyam, 1998).

Third, and importantly for our estimation strategy, the Poisson process assumption allows us to tractably characterize the type-conditional income density $g(z|n)$. Specifically, the type-conditional density at a given income z' is equal to the probability that an n-type draws z', multiplied by the probability that z' is optimal for n *conditional on drawing it*, which we denote $\pi(z|n)$. Under a Poisson process, the probability of drawing any particular income, including z', is simply $1/\mu$.

This insight transforms the problem of characterizing $g(z|n)$ into the problem of characterizing $\pi(z|n)$, which is also facilitated by the Poisson process assumption. Let k be a positive integer. Under a Poisson process, the probability of drawing k income opportunities in an interval between any two incomes z_1 and z_2 is given by the Poisson distribution:

$$\frac{\left(\frac{z_2-z_1}{\mu}\right)^k \exp\left[-\frac{z_2-z_1}{\mu}\right]}{k!}. \tag{6}$$

Of particular interest for our application, the probability of drawing *zero* incomes in this interval is therefore $\exp\left[-\frac{z_2-z_1}{\mu}\right]$. We can use this formula to compute $\pi(z|n)$, because the probability of choosing some z' (conditional on having it as an income opportunity) is simply the probability that the agent does not have some other *better* income opportunity. Put differently, $\pi(z'|n)$ is equal to the probability that the n-type agent has zero income opportunity draws from among the set of incomes that give them higher utility than z'.

This calculation is illustrated in Figures 5 and 6 for the case of a locally linear income tax, using an agent of type a. (We will turn to the behavior around tax bracket thresholds below). Panel (a) of Figure 5 illustrates the budget constraint for the agent's constrained optimization problem. Panel (b) illustrates the agent's indirect utility $v(z;a) = u(z-T(z),z;a)$ as a function of income in the vicinity of their optimal income $z^*(a)$. Figure 6 illustrates how this indirect utility function can be used to compute the type-conditional density $g(z|a)$ at a particular income z', by computing the dominating income range. We define the functions $\underline{Z}(z'|a)$ and $\overline{Z}(z'|a)$ to return the lower and upper income values that give an a-type the same utility as income z'. (By construction, when z' lies below a's optimal continuous choice $z^*(a)$, $\underline{Z}(z'|a) = z'$.) Formally, $\underline{Z}(z;n)$ and $\overline{Z}(z;n)$ are defined as the minimum and the maximum, respectively, of the values

of Z that solve the following equation, for a given z':

$$(1-t_0)z' - \frac{n}{1+1/e}\left(\frac{z'}{n}\right)^{1+1/e} = (1-t_0)Z - \frac{n}{1+1/e}\left(\frac{Z}{n}\right)^{1+1/e}. \tag{7}$$

Putting these calculations together, the type-conditional income density is given by the following equation:

$$g(z|n) = \frac{1}{\mu}\exp\left[-\frac{\overline{Z}(z|n) - \underline{Z}(z|n)}{\mu}\right]. \tag{8}$$

When this density is evaluated at n's optimal continuous income choice $z^*(n)$, we get $g(z^*(n)|n) = 1/\mu$, reflecting that if the taxpayer happens to draw $z^*(n)$ as an income opportunity, they will choose it with certainty, and thus the density is simply equal to the probability of drawing $z^*(n)$.

We can combine equations (5) and (8) to characterize the income density $h(z)$ for a given elasticity e, a lumpiness parameter μ, and a specified ability density $f(n)$:

$$h(z) = \int_0^\infty g(z|n)f(n)dn = \frac{1}{\mu}\int_0^\infty \exp\left[\frac{-(\overline{Z}(z|n) - \underline{Z}(z|n))}{\mu}\right] f(n)dn. \tag{9}$$

We discuss the estimation of this model, including the unobserved type density $f(n)$, in Section 2.5 below. First, however, we describe how this model extends to characterize the income density $h(z)$ around a kink or notch in the tax function.

2.3 Diffuse bunching around tax kinks

The above procedure can be easily extended to characterize the income density around a kink in the tax function. Indeed, the only aspect of the preceding calculation that is affected by the kink is the characterization of the dominating income region $\overline{Z}(z'|n) - \underline{Z}(z'|n)$. This requires characterizing agents' indirect utility functions in the presence of a tax kink.

Figure 7 illustrates the construction of indirect utility functions around a tax kink. Panel (a) plots the budget constraint arising from the kinked tax function as a solid line. The figure also extends each linear segment across the bracket threshold as a dashed line, to illustrate the counterfactual budget constraints that would operate if either $T_0(z)$ or $T_1(z)$ applied across all incomes. The optimal continuous income choice for the marginal non-buncher (type b) is displayed, along with the corresponding indifference curve. Panel (b) displays the indirect utility function for the marginal non-buncher, which can be found by retaining the relevant portions

of the indirect utility functions that would arise under each of the linear budget segments:

$$v(z;b) = \begin{cases} v_0(z,b) & \text{if } z \leq z^{thr} \\ v_1(z,b) & \text{if } z > z^{thr} \end{cases} \tag{10}$$

Panels (c) and (d) likewise display the budget constraint and the composite indirect utility function for the marginal buncher, type c. In both cases, the concave kink in the budget constraint produces a corresponding convex kink in agents' indirect utility functions.

Figure 8 demonstrates how these indirect utility functions are translated into type-conditional income densities for the marginal non-buncher (Panel (a)) and the marginal buncher (Panel (b)). Both panels illustrate the calculation of the type-conditional income density at an income z'—which differs across the panels—by identifying the dominating income range for z' as perceived by each type. We then proceed as in the case of a linear tax, already described, by computing the type-conditional income density using equation (8).

The result of this type-conditional density calculation is plotted in the lower portion of each panel. In the case of the marginal non-buncher (Panel a), the effect is to raise the type-conditional density at z', relative to the density $g_0(z|b)$ that would obtain under the linear tax $T_0(z)$, which is plotted for comparison. The reason for this change can be understood from the size of the dominating income range above: some incomes above the threshold z^{thr} that would be preferred to z' under the linear tax $T_0(z)$ are no longer preferred in the presence of a kink (i.e., $\overline{Z}(z';b)$ is lower than in the absence of the kink). As a result, it is less likely that the taxpayer would draw another income that dominates z', and thus the probability that one chooses z' (conditional on drawing z') has increased. Applying this logic to other income choices, we can trace out the shape of the type-conditional density $g(z|b)$ across all incomes. The result is that the kink-induced type-conditional density has left skew relative to the counterfactual $g_0(z|b)$. By the same logic, the kink-induced type-conditional density for type c has right skew.

The type-conditional densities illustrated in Figure 8 cannot be observed in the data, because types are unobservable. Figure 9 illustrates the implications for the observable density of *incomes*. The top portion of each panel shows the optimal continuous choice for agents of types a, b, c, and d in the presence of a kink. The lower portion of Panel (a) illustrates the overlapping type-conditional income densities of each type. (Although we have plotted these densities for only four types, the underlying type distribution is continuous across this range by assumption, so that there is a continuum of types between these four.) The lower portion of Panel (b) shows the observable income density which results from "adding up" the type-conditional densities in Panel (a). The clustering of the type-dependent densities of b and c around the kink produces diffuse excess mass in the income density $h(z)$ around the kink at z^{thr}. If the underlying type

density is uniform, as in this example, the excess mass is approximately symmetric, because the asymmetry in the type-conditional income densities of b and c balances out, as illustrated by Figure 8. A sloped type density will tend to produce asymmetric excess mass, but of a form that can be fully computed for a given type density $f(n)$ and parameters e and μ.

2.4 Asymmetric bunching around tax notches

The model above also extends to the case of notches, with one additional nuance: notches may produce non-monotonicities in the indirect utility function. Figure 10 illustrates the construction of the composite indirect utility functions for the marginal non-buncher and buncher. In each case, the discontinuity in the budget constraints in Panels (a) and (c) produce corresponding discontinuities in the indirect utility functions in Panels (b) and (d). As shown in Panel (d), the marginal buncher has two local maxima in indirect utility: z^{thr}, and the maximal point of $v_1(z;c)$.

To construct the type-conditional income densities, we proceed as before by computing the range of dominating incomes at any given z'; this procedure is illustrated in Figure 11. In the case of the marginal non-buncher (type b), this calculation is straightforward; we need only modify the calculation of $\overline{Z}(z';n)$ to reflect the possibility that the dominating region may be bounded above by z^{thr}, as illustrated in Panel (a). In the case of the marginal buncher (type c), the calculation is complicated by existence of two different income ranges that dominate z', below and above the threshold z^{thr}. To handle such cases, we define the functions $\underline{Z}_1(z;n)$ and $\overline{Z}_1(z;n)$ to return the lower and upper incomes that produce utility equal to $v(z;n)$ under the linear income tax $T_1(z)$, and $\underline{Z}_0(z;n)$ to return the lower income value that produces utility equal to $v(z;n)$ under the linear tax $T_0(z)$. Thus in Panel (b), the left dominating income range is the interval $\left[\underline{Z}_0(z';n), z^{thr}\right] = \left[z', z^{thr}\right]$, and the right dominating income range is the interval $\left[\underline{Z}_1(z';n), \overline{Z}_1(z';n)\right]$. A virtue of the Poisson process governing income opportunities is that the probability of drawing zero opportunities in a dominating range does not depend on the position of the range with respect to the optimal income, nor on its contiguity. The probability depends only the size of the dominating income range(s) in total. Therefore, once these dominating income ranges are determined, they can simply be summed, and the result inserted into the numerator of the bracketed term in equation (8) to compute the type-conditional density.

Figure 12 illustrates the aggregation of the type-conditional income densities to construct the observed income density. As in Figure 9, the top panes of each panel display the income choices of the four types a, b, c, and d. Summing these type-conditional densities (as well as those of all the un-plotted intervening types) produces the observed density $h(z)$, plotted in the lower portion of Panel (b). This figure illustrates that our model of lumpy income choice can

match two features of empirical income distributions that the conventional model with continuous income choice cannot: diffusion in the excess mass left of the threshold, and positive mass in the dominated income region to the right. These features have been the subject of detailed investigation in the literature (Kleven and Waseem, 2013).

2.5 Estimation

We now describe how the parameters of this model can be estimated from empirical data. The empirical strategy is to select the model parameters that maximize the likelihood of observing a given empirical histogram. To do so, we search over the parameter values for the elasticity e and the lumpiness parameter μ. If desired, we can also allow the tax notch size dT to be an estimated parameter, treating it as a revealed feature of taxpayer behavior.

In order to estimate the model, we must impose some parametric structure on the ability density $f(n)$. Like in the rest of the bunching literature, the key identifying assumption is that the ability distribution (and thus the counterfactual income density $h_0(z)$) is, in some sense, smooth in the vicinity of the bracket threshold z^{thr}. Intuitively, this amounts to assuming that the location of the bracket threshold is not selected to occur at an income that happens to coincide with a distortion in the underlying ability distribution. (Blomquist et al. (2021) explore this identification strategy and its limitations at length.)

We operationalize this identification strategy by assuming that the ability density follows a polynomial of order Q, i.e.,

$$f(n;\theta) = \theta_0 + \theta_1 n + \theta_2 n^2 + \dots \tag{11}$$

$$= \sum_{q=0}^{Q} \theta_q n^q \tag{12}$$

for a vector $\theta = \{\theta_0, \theta_1, \dots, \theta_Q\}$.

We then estimate the parameters of the model—e, μ, θ, and (if desired) dT—using maximum likelihood. Letting i index the observations in the data, with X_i denoting each observation's income, our starting point for the likelihood function is

$$L(e, \mu, dT, \theta) = \Pi_i h(X_i = z; e, \mu, dT, \theta). \tag{13}$$

Performing maximum likelihood estimation with this likelihood function will not result in an interior maximum, however, because we have imposed no constraint on the integral of the income density function $h(z; e, \mu, dT, \theta)$. For example, the solver can make (13) arbitrarily high by letting the polynomial intercept θ_0 tend to ∞. To address this, we can normalize the popula-

tion density within a desired range $[z_{min}, z_{max}]$ around the bracket threshold (e.g., the income range reflected in the empirical support of the taxable income distribution). In principle, we could then perform maximum likelihood estimation by computationally searching for the vector (e, μ, θ, dT) that solves the following constrained maximization problem:

$$\max_{e,\mu,\theta,dT} \sum_i \log h(X_i = z; e, \mu, dT, \theta) \quad \text{subject to} \quad \int_{z_{min}}^{z_{max}} h(z; e, \mu, dT, \theta) dz = 1. \tag{14}$$

This estimation can be implemented directly with raw micro data on incomes reported to the tax authority. In many settings, however, privacy or logistical constraints restrict the analyst to operate with a binned histogram of incomes; that is the usual data input in the bunching literature. The approach in equation (14) can be modified for use with binned data using interval censoring, by letting i index bins (rather than observations) and replacing the maximand in equation (14) with $\sum_i H_i \log h(Z_i; e, \mu, \theta, dT)$, where (Z_i, H_i) denote the income and frequency values for each bin i, and $h(Z_i)$ denotes the probability density function from the model-predicted density at bin Z_i. We adopt this modification for our estimations in the simulations and empirical exercises that follow.

Computationally solving the constrained maximization problem in equation (14) presents a challenge. The likelihood function is

$$h(z; e, \mu, dT, \theta) = \int_{-\infty}^{\infty} g(z|n; e, \mu, dT) f(n; \theta) dn. \tag{15}$$

This is difficult because numerically integrating over a large grid of types n is time consuming, and the parameter space is very large when allowing for even a cubic polynomial, which we adopt as our baseline specification.

The problem can be converted into one that is numerically tractable by viewing the selection of the polynomial coefficients θ as an inner problem that is computed conditional on the other parameters, so that we can write the maximum likelihood problem as

$$\max_{e,\mu,dT} \sum_i H_i \log h(Z_i = z; e, \mu, \theta(e, \mu, dT)), \tag{16}$$

with the integration constraint in (14) enforced by appropriate selection of the function $\theta(e, \mu, dT)$. If the inner function $\theta(e, \mu, dT)$ were selected to solve the constrained maximization in equation (14), then this approach would amount to concentrating out the parameter vector θ. For numerical expediency, we instead exploit the structure of the problem in a way that allows us to compute $\theta(e, \mu, dT)$ very quickly using polynomial regression. In effect, we select θ to minimize the sum of squared differences between the observed histogram (normalized to sum to one)

and the predicted income density:

$$\theta(e,\mu,dT) = \min_{\theta} \sum_i \left(\frac{H_i}{\sum_j H_j} - h(Z_i;e,\mu,dT) \right)^2. \tag{17}$$

This problem can be written in regression form as follows, for the case in which $f(n;\theta)$ is cubic, where the θ coefficients are selected to minimize the sum of squared residuals $\sum_i \varepsilon_i^2$:

$$\begin{aligned}
\frac{H_i}{\sum_j H_j} &= h(Z_i;e,\mu,dT) + \varepsilon_i \qquad (18)\\
&= \int_{-\infty}^{\infty} g(Z_i|n;e,\mu,dT) f(n;\theta) dn + \varepsilon_i \\
&= \int_{-\infty}^{\infty} g(Z_i|n;e,\mu,dT) \left(\theta_0 + \theta_1 n + \theta_2 n^2 + \theta_3 n^3\right) dn + \varepsilon_i \\
&= \left[\int_{-\infty}^{\infty} g(Z_i|n;e,\mu,dT) dn\right] \theta_0 + \left[\int_{-\infty}^{\infty} g(Z_i|n;e,\mu,dT) n dn\right] \theta_1 \\
&\quad + \left[\int_{-\infty}^{\infty} g(Z_i|n;e,\mu,dT) n^2 dn\right] \theta_2 + \left[\int_{-\infty}^{\infty} g(Z_i|n;e,\mu,dT) n^3 dn\right] \theta_3 + \varepsilon_i.
\end{aligned}$$

The terms in brackets require only a single numerical computation of the function $g(z|n;e,\mu,dT)$, after which the θ polynomial coefficients can be calculated efficiently using standard matrix inversion. This facilitates rapidly computing equation (16), searching over only the three parameters e, μ, and dT. The integration constraint in equation (14) can be enforced by using a two-step procedure, in which after selecting a provisional θ vector to solve equation (17), we adjust the intercept θ_0 so that the constraint holds exactly.

In spirit, this method resembles the approach—often employed in the conventional bunching literature—of fitting a flexible polynomial to the observed income distribution outside of a selected "bunching window," although two differences should be noted. First, by structurally accounting for the distortion pattern produced by the bracket threshold, we need not select a window around the threshold to "leave out" when computing the best-fit values of θ. Instead, even data near the bracket threshold helps identify θ. This reasoning suggests that this approach may be more robust to choices about the degree of polynomial, since additional flexibility does not attempt to fit the diffuse bunching around the bracket threshold, as may be the case in the conventional approach if some bunching extends outside the excluded bunching window.

Second, this approach assumes that the smooth polynomial structure is a feature of the underlying ability distribution, $f(n)$, rather than of the observed income distribution outside the bunching window. As illustrated by Figure 3, even the continuous adjustment model predicts a discontinuity in income density around the bracket threshold, due to the jump in types and

the condensed mapping from types to income under higher marginal tax rates. By estimating the polynomial coefficients on the type distribution directly, this approach does not impose smoothness across that threshold.

Having implemented the maximum likelihood estimation described by equation (16), we can compute standard errors for our estimates $\hat{e}$, $\hat{\mu}$, and $\hat{dT}$ using the standard maximum likelihood estimator. We compute these using the Matlab *mlecov* function, the standard errors of which are reported in our results below.[9]

3 Simulations

We use simulations to quantitatively examine the issues discussed above. We do so by varying the underlying parameters of the data-generating process (DGP) to generate data, and use this simulated data to explore the performance of the maximum likelihood estimation method of our model as the underlying parameters vary. We also compare the estimates delivered by our model to those from the conventional bunching-based elasticity estimator, which is of particular interest when the underlying parameters of the DGP allow for lumpy income adjustment.

We choose simulation parameters which are of the same order of magnitude as the empirical applications we eventually consider in Section 4. We stipulate a tax schedule with a tax bracket threshold at $z^{thr} = 300{,}000$, at which the marginal tax rate rises from $t_0 = 10\%$ to $t_1 = 20\%$. We set a baseline elasticity of $e_0 = 0.3$ and a lumpiness parameter of $\mu_0 = 10{,}000$, where we use the "0" subscript to denote the true parameters of the data-generating process. We denote estimates of those parameters using the usual convention, i.e., $\hat{e}$ and $\hat{\mu}$. We impose a simple linear underlying ability density, $f(n;\theta) = \theta_0 + \theta_1 n$, with $\theta_0 = 1{,}000$ and $\theta_1 = -50$.

We construct a histogram of simulated income values in two steps. First, we draw N ability values (n_i) from the known ability density $f(n;\theta)$ in the vicinity of the tax bracket threshold.[10] For each ability draw, we simulate a set of income opportunities, from which we choose, for each agent, the highest-utility option.[11]

[9] In our empirical application, we independently verified the robustness of these standard errors using a bootstrapping procedure; this produces very similar results.

[10] Specifically, we draw N values of n_i between a set of bounds $\underline{n}$ and $\overline{n}$, with the probability of drawing any value n equal to the density $f(n;\theta)$. To choose the lower bound $\underline{n}$, we note that due to income lumpiness, the set of agents who actually earn a given z will include types whose preferred continuous income is well below and well above z. Therefore, to simulate the income density near the bounds of an income range $[\underline{z},\overline{z}]$, we must draw from an ability density with preferred continuous incomes well outside that range. We use $z^*(\underline{n}) = \underline{z} - 100{,}000$ and $z^*(\overline{n}) = \overline{z} + 100{,}000$.

[11] To simulate income opportunity sets, we exploit the fact that differences between adjacent elements in a Poisson process are iid draws from an exponential distribution with mean μ. Thus we can construct a random income opportunity set spanning an arbitrarily wide range around a type's preferred income $z^*(n)$ by joining a random set of higher-than-preferred incomes, $\{z^*(n)+\varepsilon_a, z^*(n)+\varepsilon_a+\varepsilon_b, z^*(n)+\varepsilon_a+\varepsilon_b+\varepsilon_c, \ldots\}$ with a random set of lower-than-preferred incomes $\{z^*(n)-\varepsilon_i, z^*(n)-\varepsilon_i-\varepsilon_j, z^*(n)-\varepsilon_i-\varepsilon_j-\varepsilon_k, \ldots\}$, where the ε values are iid draws from an

Figure 13 displays several income histograms from this simulation process, each based on $N = 10$ million simulated observations. Panel (a) plots densities under the baseline parameter values, as well as with lower and higher values of the elasticity e_0. A higher elasticity raises the overall amount of diffuse bunching mass around the kink. Panel (b) holds fixed the elasticity but varies the lumpiness parameter μ_0, which alters the *spread* of the bunching mass around the kink.

The bottom two panels of Figure 13 plot similar simulated histograms with the same parameter values when there is a *notch* in the tax code at $z^{thr} = 300{,}000$; here we impose a notch value of 500. As shown in the baseline (green) histograms, the notch produces asymmetry in the excess bunching mass, which shifts to the left of the bracket threshold. The histogram still has substantial mass even in the income bin immediately to the right of the threshold. Raising the elasticity e_0 again increases the total amount of excess bunching mass, albeit asymmetrically (Panel (a)), while raising the lumpiness parameter creates greater diffusion in the bunching mass and erodes the steep decline in density to the right of the notch. When the lumpiness parameter becomes small, a depression becomes apparent around the dominated region of incomes to the right of the threshold.

Together, these figures illustrate that this income process reproduces two key qualitative features of the income histograms that have been documented in the empirical bunching literature: diffusion in the bunching around kink points, and positive mass in the dominated income region above notches.

3.1 Applying the maximum likelihood estimator

We now explore the ability of the maximum likelihood estimation procedure to recover the true parameters e_0 and μ_0. For this exercise, we simulate many rounds of data from the same underlying data-generating process, and in each case, we apply our estimation procedure to recover joint estimates of $\hat{e}$ and $\hat{\mu}$. We are interested in whether the distribution of these estimates is centered around the true parameter values e_0 and μ_0, and how often the estimated confidence intervals contain the true value.

For the simulations in this exercise, each simulation round uses a much smaller number of simulated observations than in Figure 13, in order to introduce greater sampling variation. We use $N = 100{,}000$, which produces a similar amount of noise to our empirical application in Section 4. One example round of simulated data is plotted by the green data series in Figure

exponential distribution with mean μ. In the context of a kink, where indirect utility functions are concave, only a single element must be drawn in each set, since more distant draws are guaranteed to yield lower utility. For a notch, with non-concave indirect utility functions, a larger number should be drawn, so that each agent's range of income opportunities spans across the local maxima in their indirect utility functions.

14a. The estimated parameters $\hat{e}$ and $\hat{\mu}$ resulting from our maximum likelihood estimation are reported in the upper corner, as well as the 95% confidence interval for each estimate. The orange line plots the income density under these estimated parameter values, which are close to the true values of e_0 and μ_0.

Panel (b) of Figure 14 plots the results from 1,000 such estimation procedures. Each small point plots one set of joint estimates $(\hat{e}, \hat{\mu})$, and the marginal histogram of these estimates is displayed outside each axis. A number of notable features emerge. First, for both $\hat{e}$ and $\hat{\mu}$, the distribution of estimates is centered around the true parameter value. Averaging across simulation rounds, the average value of $\hat{e}$ is 0.307 (measured in 1000s), and the average value of $\hat{\mu}$ is 10.1, close to the true values of $e_0 = 0.3$ and $\mu_0 = 10$.

Second, the spread of both distributions provides an indication of sampling error. In each round of simulated data, the maximum likelihood estimation procedure also provides a standard error estimate, and so a key question is whether this estimate gives an accurate picture of the degree of precision in the estimate. To explore this, we can compare the standard deviation of the distribution of $\hat{e}$ estimates, which is 0.026, to the average *estimate* of the standard error, which is 0.026, indicating that the maximum likelihood estimate of the standard error provides a good sense of the true degree of sampling uncertainty. Accordingly, the estimated confidence intervals provide an accurate sense of the certainty over the estimated elasticity: across the 1000 rounds of simulated estimates, the estimated 95% confidence intervals contained the true e_0 in 95.3% of cases. In the case of μ, the standard deviation of the distribution of $\hat{\mu}$ is 1.121, and the average value of the estimated standard error is 1.099.

A third notable feature of Figure 14b is the upward slope in the cluster of joint estimates. This indicates that when $\hat{e}$ is overestimated due to sampling bias, it is likely that $\hat{\mu}$ is overestimated as well. To explore this phenomenon, Figure 15 plots model-generated income densities for five combinations of (e, μ). The thick solid line plots the baseline density with $e = 0.3$ and $\mu = 10$. The other four lines correspond to the (e, μ) pairs corresponding to the four square-shaped points in Figure 14b.

In Figure 15, the densities corresponding to the points to the northwest and southeast of the baseline are easy to visually distinguish from the baseline, exhibiting substantially lower and higher densities at the kink, respectively. The reason for this pattern can be understood from the simulated densities in Figure 13. A higher elasticity e increases the density at the kink point by raising the total amount of bunching mass (Figure 13a). A *lower* value of the lumpiness parameter also increases the density at the kink point, by concentrating the excess mass more tightly around the kink (Figure 13b). Thus, the parameter combinations to the southeast of the baseline in Figure 14b correspond to densities with substantially higher density around the kink point, like the tallest density displayed in Figure 15. The reverse is true for parameter

combinations to the northwest of the baseline values, where the levels of both parameters (low e and high μ) reinforce each other to push down the density at the kink. In contrast, parameter combinations to the northeast and southwest of the baseline have opposing effects on the density at the kink. They are still distinct, indicating that the model is identified, but their difference is more subtle, involving the density at intermediate points in between the kink point and the bounds of the income window. The pattern of points in Figure 14b corresponds to this visual impression: in the presence of sampling error, it is easier to distinguish—in a statistical sense—between data-generating processes with parameter pairs on the northwest-southeast axis than those on the northeast-southwest axis in Figure 14b.[12] In sum, these points paint a clear picture of the performance of the maximum likelihood estimator when the model is correctly specified. Estimates of the elasticity and the lumpiness parameter appear consistent, in that they are distributed around the true parameters of the data-generating process, and standard errors estimated by maximum likelihood are very close to the standard deviation of the distribution of estimates. They also highlight an important aspect of this model: estimation error in e and μ are likely to have the same sign. This result has important implications for the comparison of this model to the conventional elasticity estimator based on bunching mass.

3.2 Comparison to conventional bunching estimator

We now apply a version of the conventional bunching estimator based on Saez (2010) to estimate the income elasticity in each of the simulated data sets underlying Figure 14b. We use the implementation described in Chetty et al. (2011), which builds on Saez (2010) by estimating a counterfactual using a smoothed polynomial regression.

This estimation procedure involves two steps, first estimating a counterfactual income density based on the income density excluding data points near the kink, and then using the counterfactual density to estimate the excess mass from which the elasticity is recovered. To estimate the counterfactual density, we fit a polynomial of a specified degree to the observed income histogram, excluding the data in a specified window around the kink, using the following specification:

$$C_j = \sum_{i=0}^{q} \beta_i^0 \cdot (Z_j)^i + \sum_{i=R_l}^{R_u} \gamma_i^0 \cdot \mathbf{1}[Z_j = i] + \epsilon_j^0. \tag{19}$$

Here q denotes the order of the polynomial, and R_l and R_u denote the lower and upper bounds of "bunching window" near the kink, which is excluded from the polynomial estimation.[13]

[12] Put differently, the estimator that we propose would find it easier to distinguish between data generated from low e high mu and low mu high e combinations than between low mu low e and high mu high e combinations.

[13] The convention in Chetty et al. (2011) is to set a symmetric bunching window, such that $R_l = -R_u$. We allow for the possibility of an asymmetric bunching window, following the approach in Bosch, Dekker and Strohmaier (2020) which we detail below.

When estimating the polynomial regression, we follow the approach in Chetty et al. (2011) and impose an "integration constraint" such that the total integral of population across the empirical distribution equals the total integral under the counterfactual distribution.

The second step is to compute the excess mass of incomes around the kink relative to this counterfactual density. Using equation (19), we compute the counterfactual mass in each bin within the bunching window, $\hat{C}_j^0$. Subtracting this predicted mass from the observed histogram yields the estimated excess number of individuals who report incomes near the kink relative to this counterfactual distribution:

$$\hat{B} = \sum_{i=R_l}^{R_u} C_j - \hat{C}_j^0 = \sum_{i=R_l}^{R_u} \hat{\gamma}^0. \tag{20}$$

We then map this excess mass estimate to an estimated elasticity using the approximation from Chetty et al. (2011):

$$\hat{e} \approx \frac{\hat{B}}{z^* \cdot h_0(z^*) \cdot \log(\frac{1-t_0}{1-t_1})} \tag{21}$$

Standard errors for $\hat{e}$ are estimated using a bootstrap procedure. We re-sample with replacement from the underlying distribution of firms 1000 times, re-estimating the elasticity each time, and defining the standard error as the standard deviation of the distribution of $\hat{e}$ estimates.

This conventional estimation method relies on three parameter inputs: the lower and upper bounds of the bunching window (R_l and R_u) and the order of the polynomial (q). These are left to the discretion of the researcher to be chosen via "visual inspection". We instead follow the algorithmic approach proposed in Bosch, Dekker and Strohmaier (2020), which allows the polynomial order and the bunching region to be informed by the data itself.[14]

Figure 16a illustrates this approach for the same single round of simulated data as in Figure 14a. The preferred bunching window is bounded by dashed lines, and the estimated elasticity and bootstrap-based 95% confidence interval is displayed in the corner. The estimate is substantially below the true value e_0, and the estimated confidence interval excludes the true value.

Panel (b) of Figure 16 displays the distribution of estimates like the one in Panel (a) across all

[14] This approach proceeds in five steps: (1) Estimate equation (19) with no bunching window—so that the polynomial estimation excludes only the bins adjacent to the kink—for a range of polynomial orders, retaining the specification that minimizes the Bayesian Information Criterion (BIC). (2) Define the lower bound of the bunching window as the leftmost set of two adjacent bins below the threshold where the actual count in each bin exceeds the 95% confidence interval of the predicted bin counts from equation (19), and define the upper bound using an analogous procedure to the right of the kink. (3) Repeat steps (1) and (2), widening the bunching window by one bin above and below the kink each time. Each such iteration produces a candidate set of bounds for a bunching window. (4) From the resulting distributions of candidate bounds, choose the modal lower bound and upper bound to define the preferred bunching window. (5) Using this preferred bunching window, re-estimate the final counterfactual regression with the preferred polynomial order as in Step (1).

1000 rounds of simulated data. The histogram of estimates under the conventional approach is displayed in orange; the histogram of estimates using the maximum likelihood estimation (which reproduces the histogram displayed below the horizontal axis in Figure 14b) is plotted in blue for comparison. The average estimate under the conventional approach is 0.243, which underestimates the true value of the data-generating process by over 20%. The conventional approach also provides a misleading sense of precision in this setting: across the 1000 rounds of simulation, the bootstrap-based 95% confidence intervals contained the true e_0 in less than 10% of the cases.

These results suggest that the presence of lumpiness in the income choice process may lead to downward bias in the conventional bunching estimator. To explore this possibility, we compare the performance of the conventional approach and our proposed approach as income lumpiness becomes a more important force in agents' decisions.

Figure 17 displays results from comparisons like the one in Figure 16b, for a range of lumpiness parameters μ_0. The solid points represent the mean values of the $\hat{e}$ estimates under the conventional approach (red squares) and our proposed method (blue circles) at each value of μ_0. The hollow points display the 95% quantile interval of the distributions, plotting the 2.5th and 97.5th percentiles in the distribution of each estimate. The points plotted for $\mu_0 = 10$ correspond to the distributions displayed in 16b. From this figure, we see that when the lumpiness parameter is very small, approaching the continuous-income-choice model, the mean estimate of $\hat{e}$ under the conventional approach is close to the true value of $e_0 = 0.3$. However, as μ_0 rises, the conventional estimator exhibits substantial bias, underestimating the true parameter by more than 50% at the highest plotted values of μ_0. As in Figure 16b, these estimates also provide a false sense of precision; the 95% quantile intervals remain about the same size as μ_0 rises, and their upper bound falls far below e_0. In contrast, the distribution of the estimated $\hat{e}$ using our maximum likelihood method remains centered around e_0 as lumpiness rises. The 95% quantile interval reflects the rising imprecision in the elasticity estimate as lumpiness increases. This imprecision accurately reflects the increasing difficulty of discerning diffuse bunching mass from underlying features of the smooth ability density when bunching is very diffuse.

Why does income lumpiness cause the conventional bunching estimator to be biased downward? We see two contributing forces with intuitive interpretations. The first relates to the positive correlation in estimation error between $\hat{e}$ and $\hat{\mu}$, as evidenced by the positive slope in the distribution of joint estimates in Figure 14b. This figure indicates that income densities that are generated by jointly higher and jointly lower values of e and μ are more similar to a baseline estimate than are densities in which e and μ move in opposite directions. As a result, if estimation is conditioned on an incorrect value of one of the parameters (heuristically, we can think of the conventional estimator as constraining the mu to always be zero), the best-fitting conditional

estimate of the other parameter will tend to be biased in the same direction—either up or down.

The second reason for the downward bias has to do with the effect of the diffuse bunching mass on the estimation of the polynomial counterfactual. In this model of lumpy income choice, there is a definite window outside of which the bunching mass falls to zero. As a result, some excess bunching mass will "spill over" outside of any particular bunching window—including the preferred bunching window chosen under the conventional approach. This spillover mass will tend to pull up the polynomial fit to the counterfactual in the vicinity of the kink, causing the procedure to underestimate the bunching mass B. This problem is more severe when the polynomial order is higher, because the best polynomial fit has greater flexibility to be distorted in the region of the kink. This phenomenon is illustrated in Figure 18, which applies the conventional approach to the same round of simulated data as in Figure 16a, but conditioning on polynomial orders of 1, 3, or 5. The difference in the estimated polynomial fits is visually subtle, but on inspection, it is clear that the quintic polynomial fit is higher than the linear fit in the bunching window, with the cubic polynomial falling in between. The elasticity estimates, shown in the corner of the plot, demonstrate the effect: the estimate is highest (least biased) under the linear polynomial fit, and becomes smaller (more biased) when the polynomial is more flexible.

For comparison, Panel (b) of Figure 18 also performs our proposed maximum likelihood estimation assuming that the underlying ability density has a polynomial order of either 1, 3, or 5. In this case, the elasticity estimate is not sensitive to the polynomial order. Intuitively, because the diffuse excess mass arises endogenously in the estimation model, it does not distort the estimation of the underlying ability distribution. This exercise illustrates a sense in which our proposed approach is thus robust to misspecification in the shape of the ability density, in a way that the conventional approach is not.

4 Empirical application

We apply our estimation method using administrative data on the distribution of firms around the three prominent tax kinks in the Small Business Corporation tax schedule in South Africa. The bunching patterns at each kink are displayed in Figure 1, and the underlying schedule of marginal tax rates is described below.

4.1 Data and background on small business taxation in South Africa

Like many developing countries, South Africa relies more heavily on corporate income taxes than most developed economies. In 2017, this tax base accounted for 16.2% of total tax revenue

in South Africa, considerably higher than the OECD average of 9.3% and in line with the average share for Africa (18.6%) and Latin America (15.3%).[15] The South Africa corporate income tax consists of a tiered system, with a progressive, kinked tax schedule applying to "Small Business Corporations" (SBCs), and a flat 28% tax applying to other resident companies.[16] Corporate taxable income consists of gross revenues less non-capital expenses and less any incurred losses from previous tax years which can be carried forward.[17] There are no local corporate income taxes in South Africa; businesses pay income tax only at the national level.[18]

We focus on SBCs because their kinked tax schedule is a natural setting for the bunching estimation approach. Businesses can optionally register as an SBC if they meet a set of requirements, the most pertinent being that their annual revenue must be below R20 million (about $1.4 million US).[19] We describe the full set of eligibility requirements, and other details of SBCs, in Appendix A.3. SBCs account for 38% of all formally registered companies, although due to their small size, they account for less than 20% of total tax revenue.

SBCs face a graduated progressive tax schedule. Figure 19 displays the schedule for 2018. The lowest threshold, at which tax rates rise from 0% to 7%, is at R75,750, or about $5,260 US. (In 2018, South African GDP per capita was about $7,000 US.) Below this threshold, firms face no tax liability, although they are still legally required to file a tax return. This threshold moves over time with inflation. The middle and upper thresholds are at R365,000 and R550,000, respectively, and are fixed in nominal terms. At the middle threshold, the marginal tax rate rises from 7% to 21%; at the upper threshold it rises to 28%, so that firms with incomes above this threshold face the same marginal tax rate as non-SBC firms. Appendix Table A1 reports the full SBC tax schedule for each year from 2010 to 2018.

For our analysis, we study the population of SBCs from 2014 to 2018—this is the period over which the three-kink structure illustrated in Figure 19 has been in place.

4.2 Results

Figure 20 shows the empirical histogram of firms around the lower, middle, and upper kinks (in green), and the model-predicted density produced by our maximum likelihood estimation (in orange). Parameter estimates for the income elasticity (e), the lumpiness parameter (μ), and

[15] Data from the OECD is available at `https://stats.oecd.org/Index.aspx?DataSetCode=CTS_REV`.

[16] There are also alternative tax schedules for gold mining companies and micro businesses, neither of which are the focus of this paper.

[17] Corporate dividends are taxed at the shareholder level, at a 15% rate.

[18] See Pieterse, Gavin and Kreuser (2018) for more on the South African corporate income tax data. We plan to analyze personal tax returns in future research.

[19] Throughout the paper, we use an exchange rate of 14.4 South African rand per U.S. dollar, which was the prevailing rate at the end of 2018.

the "as-if" notch value (dT), together with their 95% confidence intervals, are printed in the upper right corner of each plot.

We find income elasticity estimates of 0.27 and 0.23 at the middle and upper kinks, respectively. The estimates are not statistically distinguishable. The elasticity estimated at the lowest kink is substantially higher—in excess of one. This is perhaps not surprising, as the base level of income is much lower at this kink. (In the extreme, as incomes approach zero, any measurable behavioral response would correspond to a high elasticity.)

Figure 20 also reports the estimated lumpiness parameters at each kink. These range from R6,200 to R11,300 ($431 to $785 US). These provide insight into the degree of lumpiness in incomes faced by firms at each tax kink. Between the lower and middle kinks, μ appears to increase with income, as would be the case if the distances between lumpy income opportunities increase with one's total income. However, the estimates are non-monotonic, declining from the middle to the upper kink. This pattern may be explained by heterogeneity in tax practitioner usage, discussed below. We compare these estimates to those derived from the conventional bunching estimator, computed using the method described in Section 3.2. Figure 21 reports the resulting elasticity estimates at each kink and plots the estimated counterfactual density in orange.[20] These elasticity estimates are substantially lower, falling short of the maximum-likelihood-based elasticity estimates by between 30 and 50%. There is no overlap between the 95% confidence intervals produced by the two methods for any of the kinks. These results suggest that the concerns raised in 3.2 about bias and imprecision of the conventional bunching estimator may be economically important in practice.

In addition to providing estimates of the elasticity and income lumpiness parameters, our estimation method uncovers additional insights about the economic behavior of businesses in this setting. As noted in the introduction, although the bunching patterns around the middle and upper thresholds in Figure 20 have the visual features typically associated with a tax kink, the lower kink exhibits strikingly asymmetric bunching and missing mass to the right of the threshold, suggestive of the bunching patterns around a notch. This raises the question of whether businesses are behaving "as if" there is a notch at the lowest threshold. Our estimation method provides a framework for answering this question formally. The parameter value dT reported in each panel represents the model-estimated notch value at each threshold. The notch value estimated at the lowest kink is R350, or about $24 US, and is highly statistically significant, suggesting that the model strongly rejects pure kink behavior. Interestingly, the estimated notch value at the middle kink is similar in magnitude and also highly significant. This result highlights that conditional on a given notch value, the degree of visual asymmetry is heavily

[20] These estimates conform closely with Boonzaaier et al. (2019), which uses the conventional bunching estimator approach to estimate income elasticities at each of these kinks.

mediated by the income elasticity—an insight consistent with the simulations displayed in Figure 13c. This suggests that the behavioral tendency to treat a kink as though it were a notch is not isolated to the lowest kink, where the tax liability changes from zero to positive, but rather it may be a more general phenomenon. As such, it suggests that the source of this "as-if" notch value is unlikely to be driven by a behavioral aversion to paying a positive tax liability. Although identifying the source of "as-if" notch behavior is beyond the scope of this paper, such behavior would be consistent with a subset of taxpayers mistaking the discontinuity in marginal tax rates for a discontinuity in *average* tax rates, or other frictions which produce a perceived discrete cost when one's income surpasses each kink. The estimated notch value at the upper kink is small in magnitude—equal to about $3 US—suggesting that taxpayer behavior at that threshold is not meaningfully different from that expected around a pure kink.

We additionally explore heterogeneity in bunching behavior across firms. Specifically, we compare businesses that use a registered tax practitioner to those that do not. Figure 22 displays plots analogous to 20, partitioning businesses on tax practitioner usage. The raw histograms exhibit notably more pronounced bunching among firms with tax practitioners. Table 1 reports these results along with the estimates for the aggregate population in Figure 20. Differences in bunching behavior does not appear to be driven by a consistent difference in the income elasticity between businesses that do and do not use tax practitioners. Although the income elasticity is higher among firms that use tax practitioners at the lowest kink, the reverse is true at the middle kink, and at the upper kink, the elasticities are not statistically distinguishable. However, a clear and consistent difference can be discerned in the lumpiness parameters within each group. At every kink, income lumpiness appears to be substantially smaller among firms who use tax practitioners. This would be consistent with such firms fine-tuning their incomes more precisely in response to tax incentives.[21]

Tax practitioner usage also predicts the "as-if" notch value: firms with paid tax preparers treat a statutory kink less like a notch than other firms. If "as-if" notch behavior is driven by average-versus-marginal tax rate confusion of the nature described above, this result would be consistent with tax-practitioner-using firms exhibiting less such confusion. This result may help explain the small size of the estimated notch value at the upper kink in the aggregate sample (Figure 20c), which appears to be driven primarily by businesses with paid tax practitioners.

This heterogeneity in behavior across tax practitioner usage, although visually apparent in Figure 22, is undetected by the conventional bunching estimator approach. Table 2 reports the income elasticities estimated by the conventional approach at each kink of the three kinks.

[21] These estimates may also help explain the non-monotonicity in μ across incomes in Figure 20. After conditioning on tax preparer usage, the apparent decline in μ from the middle kink to the upper kink shrinks considerably, and confidence intervals for these estimates overlap, suggesting that differences may not be economically meaningful.

Naturally, the conventional bunching estimator does not provide insights into differences in income lumpiness or the "as-if" notch value, since these parameters are not estimated by the model. However, it may also fail to detect differences in the parameter it does measure—the income elasticity—which are evident under the maximum likelihood approach. The behavior around the middle kink illustrates this phenomenon. Comparing Panels (b) and (e) in Figure 22, firms with tax preparers exhibit a substantially—and statistically significantly—lower income elasticity than those without tax preparers (0.26 vs. 0.50). But as reported in Table 2, the conventional bunching estimator returns statistically indistinguishable elasticity estimates (0.13 vs. 0.11). In words, in the presence of income lumpiness, the conventional approach may misinterpret heterogeneity in lumpiness or "as-if" notch behavior in a way that masks real differences in the elasticity parameter of interest.

5 Conclusion

This paper studies the performance of the bunching estimator for estimating the elasticity of taxable income in the presence of lumpy income adjustments. A simple model of income lumpiness produces key patterns observed in empirical bunching settings, such as diffuse bunching at tax kinks and positive mass above tax notches, which are inconsistent with the frictionless model underpinning the conventional bunching elasticity formula, and it produces bunching patterns that fit observed data better than other common models of frictions. Simulations suggest that when incomes are lumpy, conventional bunching elasticity estimates are biased downward, falling short of the true value by more than 50% in our highest-lumpiness specifications with overly precise confidence intervals.

We address these issues by proposing an alternative estimation method, which we apply to learn about behavioral responses to three prominent tax kinks among small business corporations in South Africa. Our model uses fewer parameters than the conventional model, it produces bunching patterns that better match observed data. It recovers an elasticity estimate that is not biased in the presence of lumpiness. And it provides additional insights by also recovering an economically meaningful lumpiness parameter and, if desired, an estimate of the "as-if" notch value, which indicates whether taxpayers treat a tax kink as though it is a notch. We estimate income elasticities at the middle and upper tax kinks between 0.2 and 0.3, and high elasticities in excess of one at the lowest kink. Our estimation method also uncovers additional insights unavailable under the conventional approach: we find strong evidence that taxpayers treat the lowest tax kink like a notch, and we estimate substantially lower income lumpiness among firms paid tax preparers, consistent with finer income targeting among that group.

References

Andersen, Steffen, Cristian Badarinza, Lu Liu, Julie Marx and Tarun Ramadorai. 2022. "Reference Dependence in the Housing Market." *American Economic Review, forthcoming* .

Bachas, Pierre and Mauricio Soto. 2021. "Corporate Taxation under Weak Enforcement." *American Economic Journal: Economic Policy* 13(4):36–71.

Beffy, Magali, Richard Blundell, Antoine Bozio, Guy Laroque and Maxime Tô. 2019. "Labour Supply and Taxation with Restricted Choices." *Journal of Econometrics* 211(1):16–46.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn and Mazhar Waseem. 2015. "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan." *Journal of Political Economy* 123(6):1311–1355.

Blomquist, Sören, Whitney K. Newey, Anil Kumar and Che-Yuan Liang. 2021. "On Bunching and Identification of the Taxable Income Elasticity." *Journal of Political Economy* 129(8):2320–2343.

Boonzaaier, Wian, Jarkko Harju, Tuomas Matikka and Jukka Pirttilä. 2019. "How Do Small Firms Respond to Tax Schedule Discontinuities? Evidence from South African Tax Registers." *International Tax and Public Finance* 26(5):1104–1136.

Bosch, Nicole, Vincent Dekker and Kristina Strohmaier. 2020. "A Data-Driven Procedure to Determine the Bunching Window: An Application to the Netherlands." *International Tax and Public Finance* 27(4):951–979.

Brehm, Margaret, Scott A Imberman and Michael F Lovenheim. 2017. "Achievement Effects of Individual Performance Incentives in a Teacher Merit Pay Tournament." *Labour Economics* 44:133–150.

Chetty, Raj. 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80(3):969–1018.

Chetty, Raj, John N. Friedman, Tore Olsen and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *The Quarterly Journal of Economics* 126(2):749–804.

Dee, Thomas S, Will Dobbie, Brian A Jacob and Jonah Rockoff. 2019. "The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations." *American Economic Journal: Applied Economics* 11(3):382–423.

DeFond, Mark L. and K.R. Subramanyam. 1998. "Auditor Changes and Discretionary Accruals." *Journal of Accounting and Economics* 25(1):35–67.

Dekker, Vincent and Karsten Schweikert. 2021. "A Comparison of Different Data-driven Procedures to Determine the Bunching Window." *Public Finance Review* 49(2):262–293.

Devereux, Michael P., Li Liu and Simon Loretz. 2014. "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records." *American Economic Journal: Economic Policy* 6(2):19–53.

Diamond, Rebecca and Petra Persson. 2016. "The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests." *Working Paper no. 22207, National Bureau of Economic Research* .

Gelber, Alexander M., Damon Jones and Daniel W. Sacks. 2020. "Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test." *American Economic Journal: Applied Economics* 12(1):1–31.

Gordon, Roger and Wei Li. 2009. "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation." *Journal of Public Economics* 93(7-8):855–866.

Grubb, Michael D and Matthew Osborne. 2015. "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock." *American Economic Review* 105(1):234–271.

Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review* 104(2):537–563.

Kleven, Henrik J. and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *The Quarterly Journal of Economics* 128(2):669–723.

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics* 8:435–464.

Kosonen, Tuomas and Tuomas Matikka. 2020. "Discrete Labor Supply: Empirical Evidence and Implications." *Working Paper no. 132, VATT Institute for Economic Research* .

Kostøl, Andreas R. and Andreas S. Myhre. 2021. "Labor Supply Responses to Learning the Tax and Benefit Schedule." *American Economic Review* 111(11):3733–3766.

Kothari, S.P., Andrew J. Leone and Charles E. Wasley. 2005. "Performance Matched Discretionary Accrual Measures." *Journal of Accounting and Economics* 39(1):163–197.

Manoli, Day and Andrea Weber. 2016. "Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions." *American Economic Journal: Economic Policy* 8(4):160–182.

Mavrokonstantis, Panos and Arthur Seibold. 2022. "Bunching and Adjustment Costs: Evidence from Cypriot Tax Reforms." *Working Paper no. 9773, CESifo* .

Mortenson, Jacob A. and Andrew Whitten. 2016. "How Sensitive Are Taxpayers to Marginal Tax Rates? Evidence from Income Bunching in the United States." *Working Paper* .

Mortenson, Jacob A. and Andrew Whitten. 2020. "Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule." *American Economic Journal: Economic Policy* 12(3):402–432.

Pieterse, Duncan, Elizabeth Gavin and C. Friedrich Kreuser. 2018. "Introduction to the South African Revenue Service and National Treasury Firm-Level Panel." *South African Journal of Economics* 86:6–39.

Pollinger, Stefan. 2021. "Kinks Know More: Policy Evaluation Beyond Bunching with an Application to Solar Subsidies." *Working Paper* .

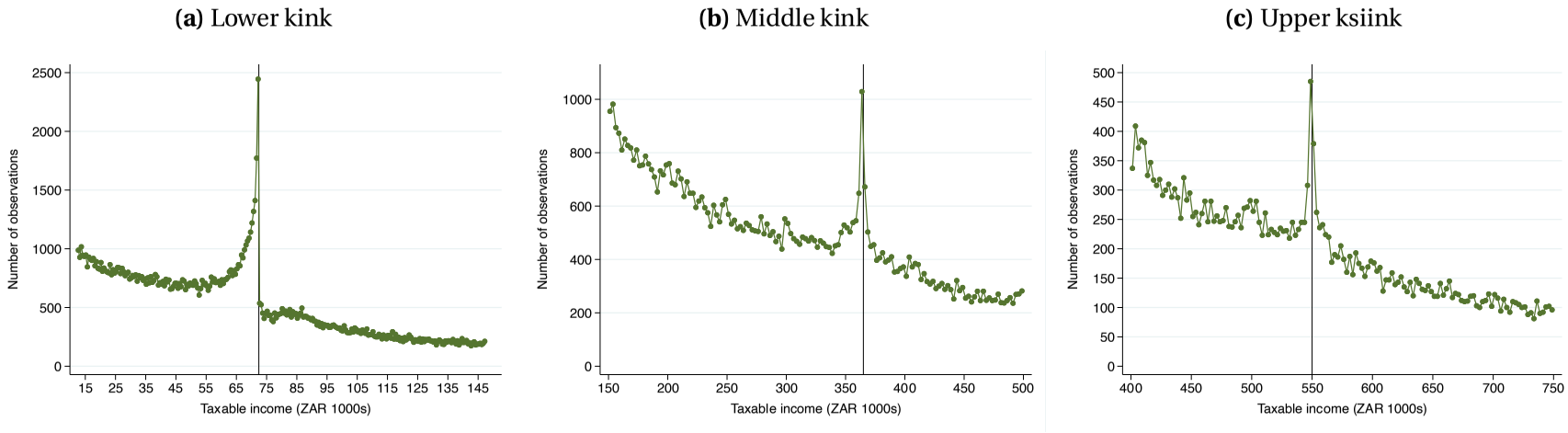
Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *The Review of Economic Studies* 85(2):1251–1278.

Rees-Jones, Alex and Dmitry Taubinsky. 2020. "Measuring "Schmeduling"." *The Review of Economic Studies* 87(5):2399–2438.

Saez, Emmanuel. 1999. "Do Taxpayers Bunch at Kink Points?" *Working Paper no. 7366, National Bureau of Economic Research* .

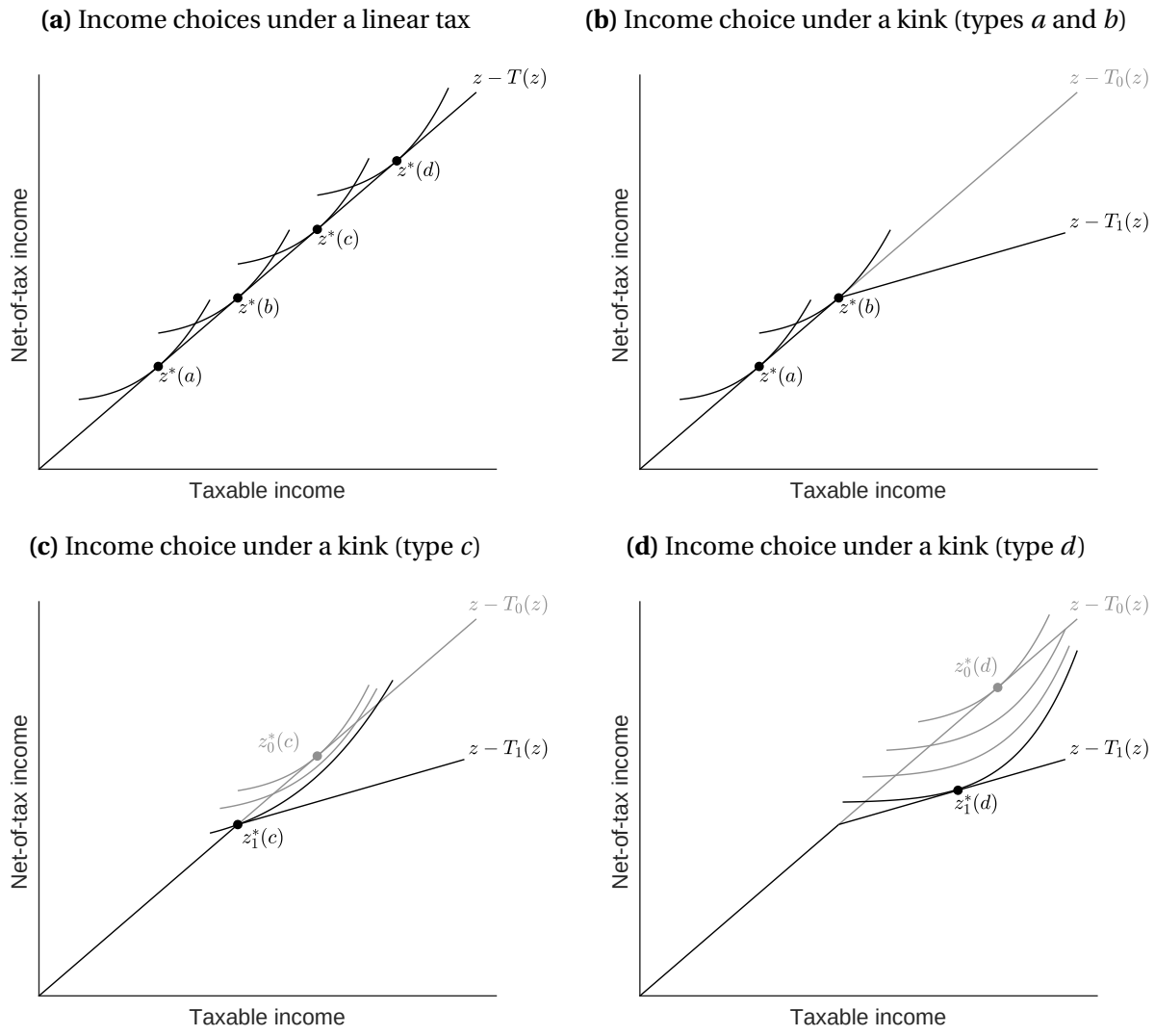
Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2(3):180–212.

Figure 1: Bunching in the income distribution of South African Small Business Corporations



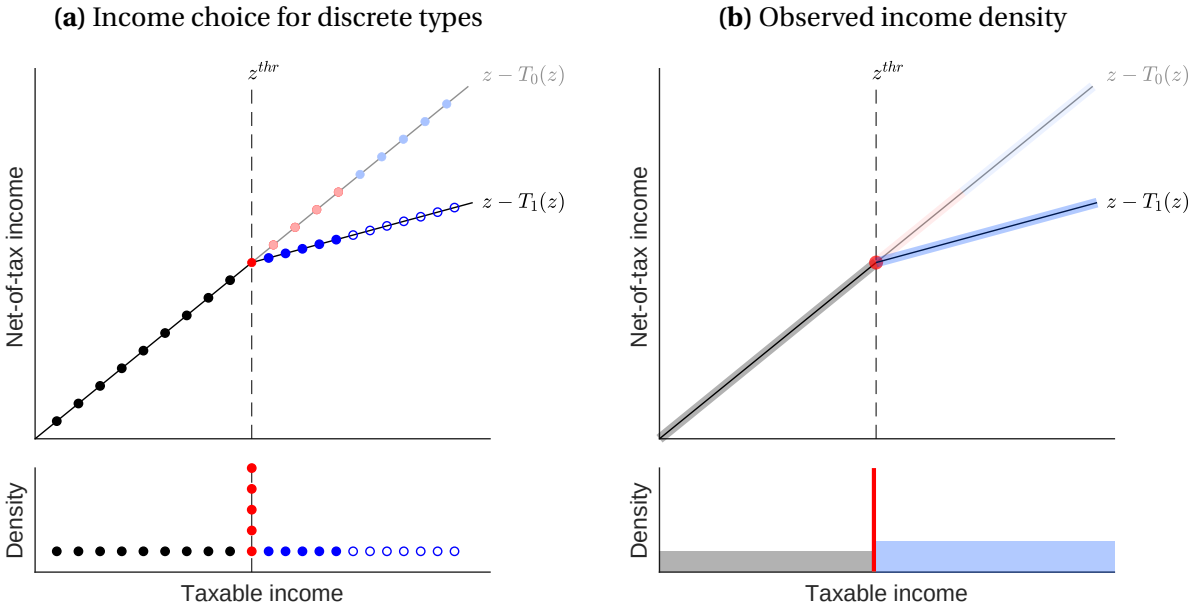
The green points plot the empirical histogram of firms with different earnings in the data. The sample consists of all South African small business corporations that filed tax returns in 2014–2018. The lowest bracket threshold adjusts over time for inflation. In this and all similar figures, we shift the histogram around the lower threshold within each year so that each year's threshold aligns at the same point; the horizontal axis reflects the lower bracket threshold in 2018.

Figure 2: Conventional (continuous-choice) bunching model with a progressive tax kink



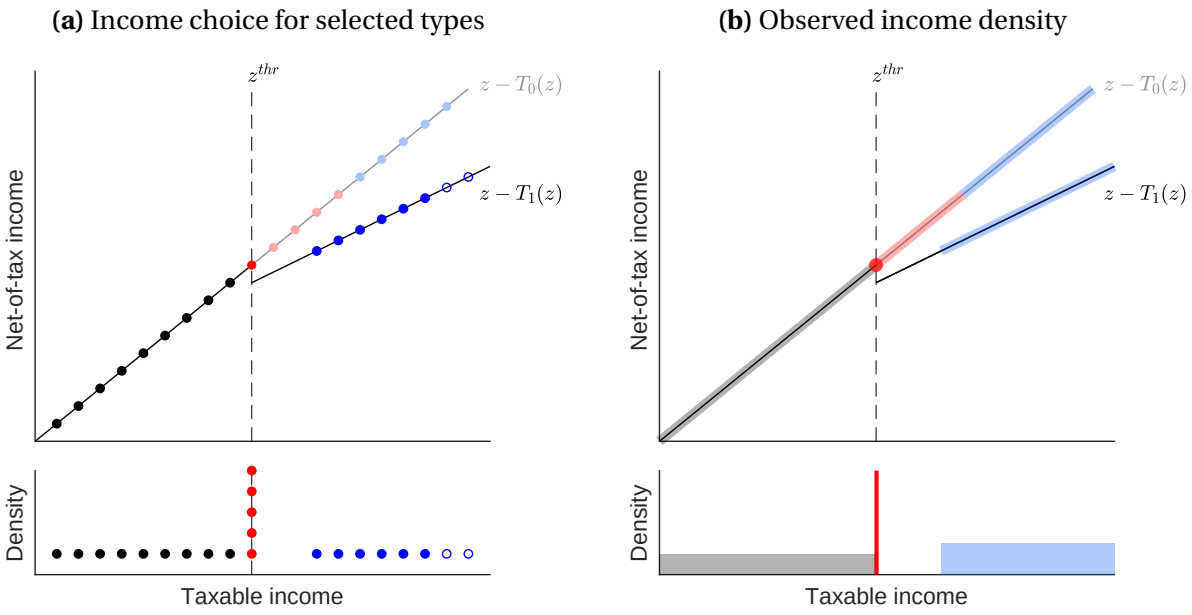
This figure illustrates the income choice around a tax kink under the conventional model with continuous income choices. Panel (a) illustrates the optimal choice of income, z^*, for four selected types of taxpayers under a linear income tax. Panels (b), (c), and (d) illustrate the optimal choice for each type under a kinked income tax, where the tax changes from the $T_0(z)$ to $T_1(z)$ at the threshold z^{thr}. Incomes z_0^* and z_1^* denote the optimal choice under the linear tax $T_0(z)$ and $T_1(z)$, respectively.

Figure 3: Income density around a kink under the conventional model



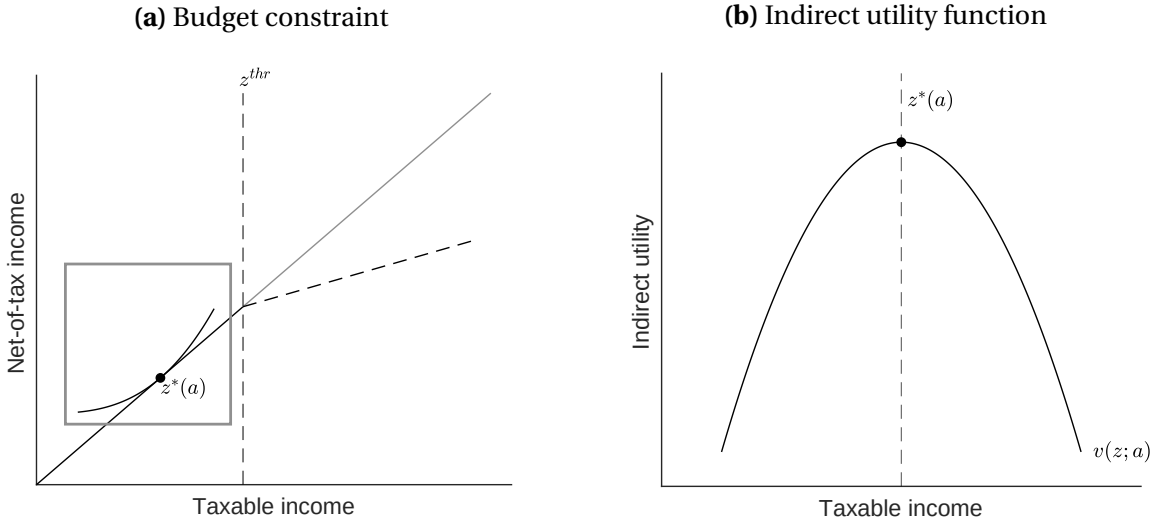
Panel (a) plots the income choices for discrete types in the presence of a kink, for a uniform type distribution. Black points denote types who choose the same income under the linear tax $T_0(z)$ and the kinked tax schedule. Red points denote types who bunch at the bracket threshold z^{thr} under the kinked tax schedule; their counterfactual income choices under $T_0(z)$ are plotted in light red for reference. Blue points denote types who choose incomes above the threshold under the kink. Hollow blue points denote agents whose counterfactual incomes under $T_0(z)$ lie outside this plotted range of incomes. The lower pane displays the observed probability density function from these choices. Panel (b) translates to the case of continuous types, which exhibits an atom of mass at the threshold z^{thr} and a jump in the density around that threshold, due to the compression of types in response to the higher marginal tax rate above the kink.

Figure 4: Income density around a notch under the conventional model



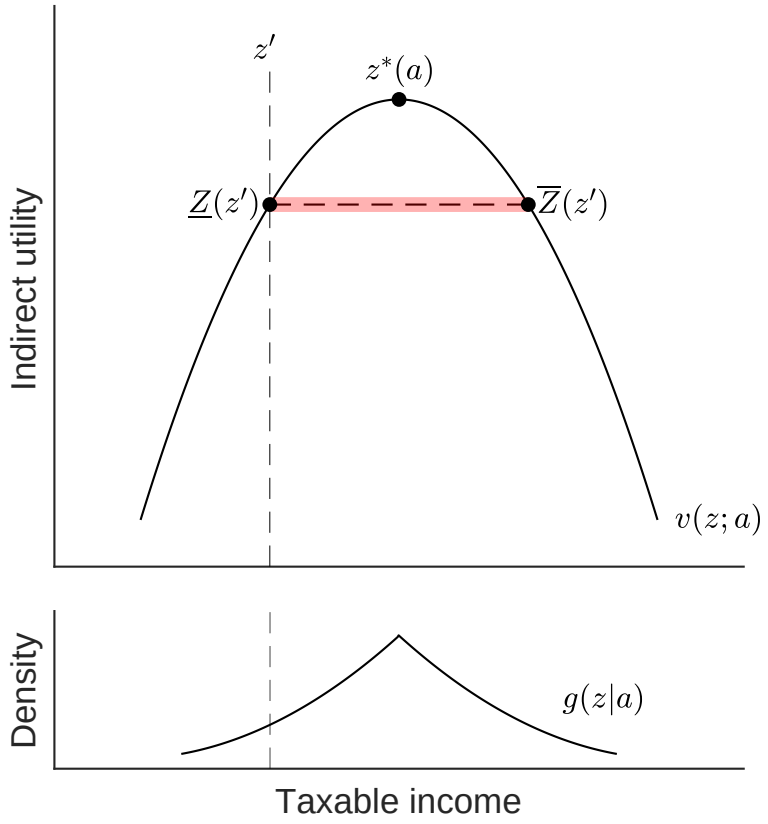
This figure illustrates the observed income density around a notch under the conventional model with continuous income choice.

Figure 5: Utility from income choices under a linear tax (type a)



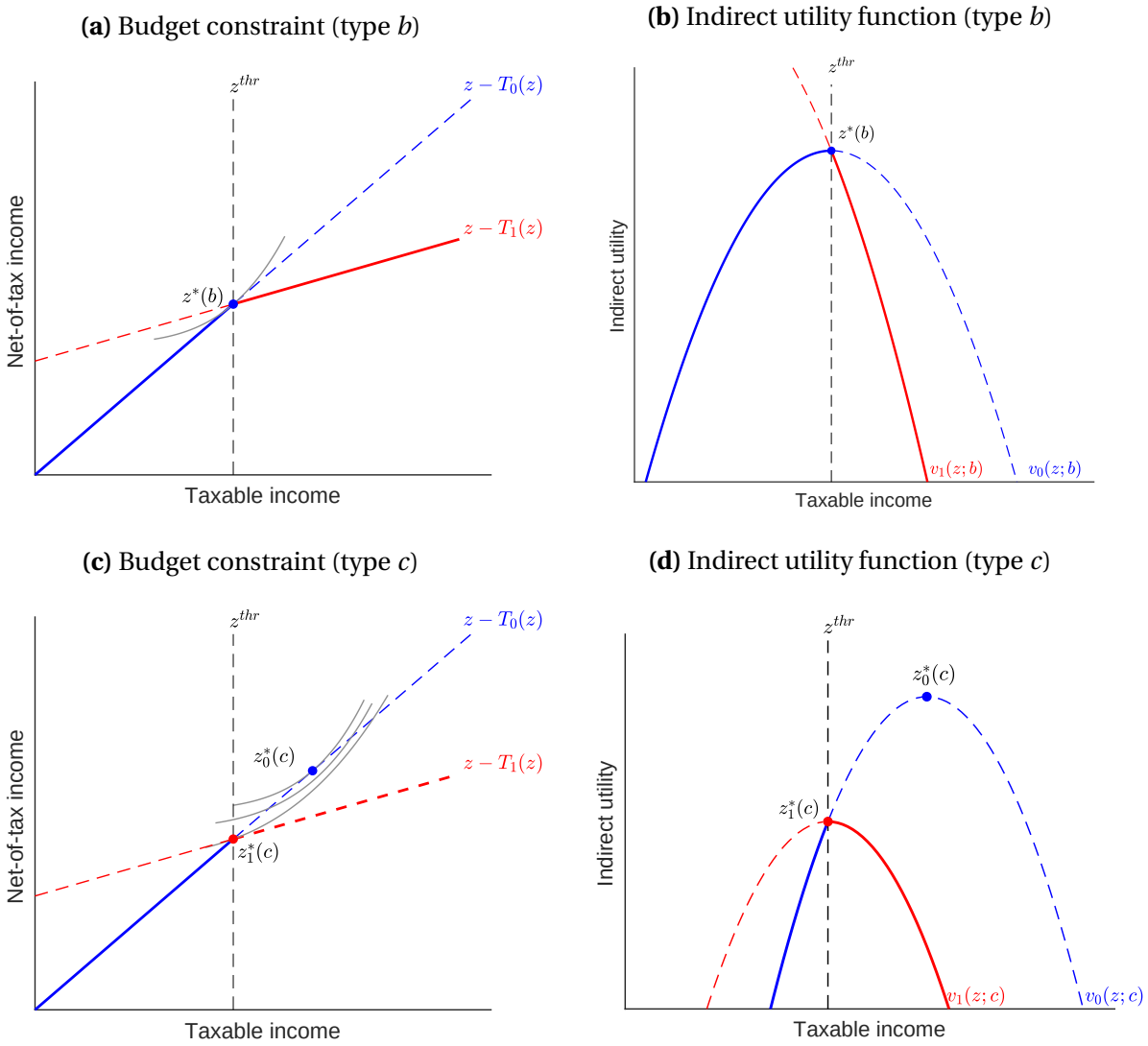
Panel (a) shows the optimal choice of continuous income for an agent type a. Panel (b) plots this type's indirect utility function $v(z;a) \equiv u(z - T(z), z; a)$ in the neighborhood of this optimal choice.

Figure 6: Type-conditional income density under a linear tax (type a)



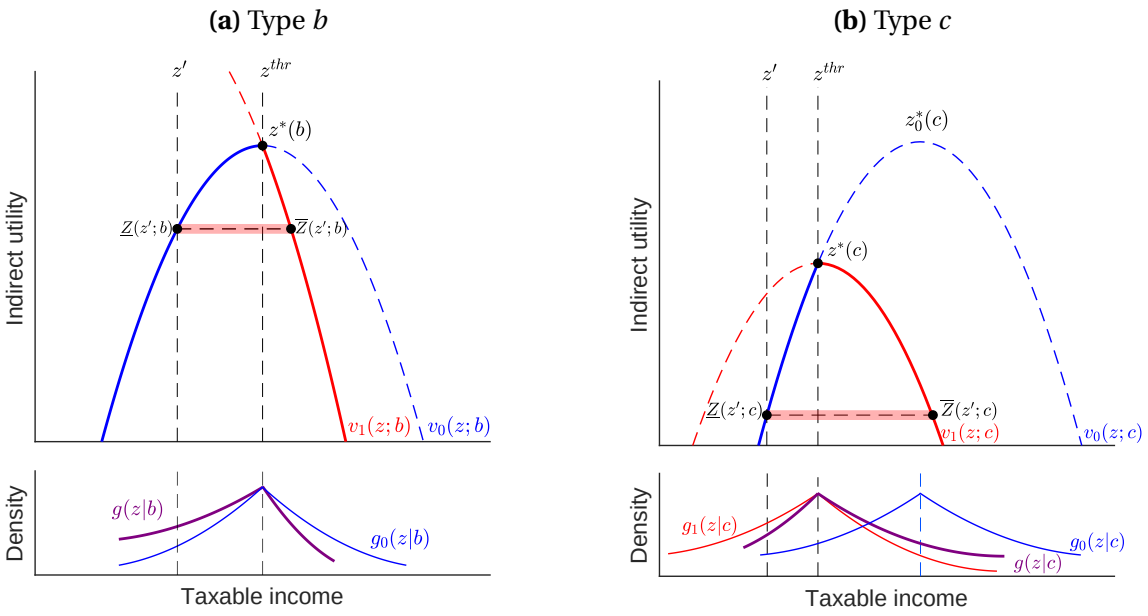
This figure illustrates the calculation of the type-conditional income density among a-type agents at a particular income level z', under a locally linear income tax. The top portion reproduces the indirect utility function from Figure 5b. There is a continuum of a-type agents facing this indirect utility, each of whom draw a sparse set of income opportunities in the vicinity of their optimal choice $z^*(a)$. An agent who has z' in their income opportunity set will select this income iff they do not have some other income opportunity that yields higher utility, i.e., iff they do not have an income opportunity in the "dominating income range" between $\underline{Z}(z')$ and $\overline{Z}(z')$ in the figure above. The probability of having zero income choices in this interval is given by the Poisson distribution, and is equal to $\pi(z'|a) = \exp\left[\frac{-(\overline{Z}(z'|a)-\underline{Z}(z'|a))}{\mu}\right]$. The type-conditional density $g(z'|a)$ is equal to this conditional probability multiplied by the probability of drawing z', which is $1/\mu$.

Figure 7: Utility from income choices around a tax kink



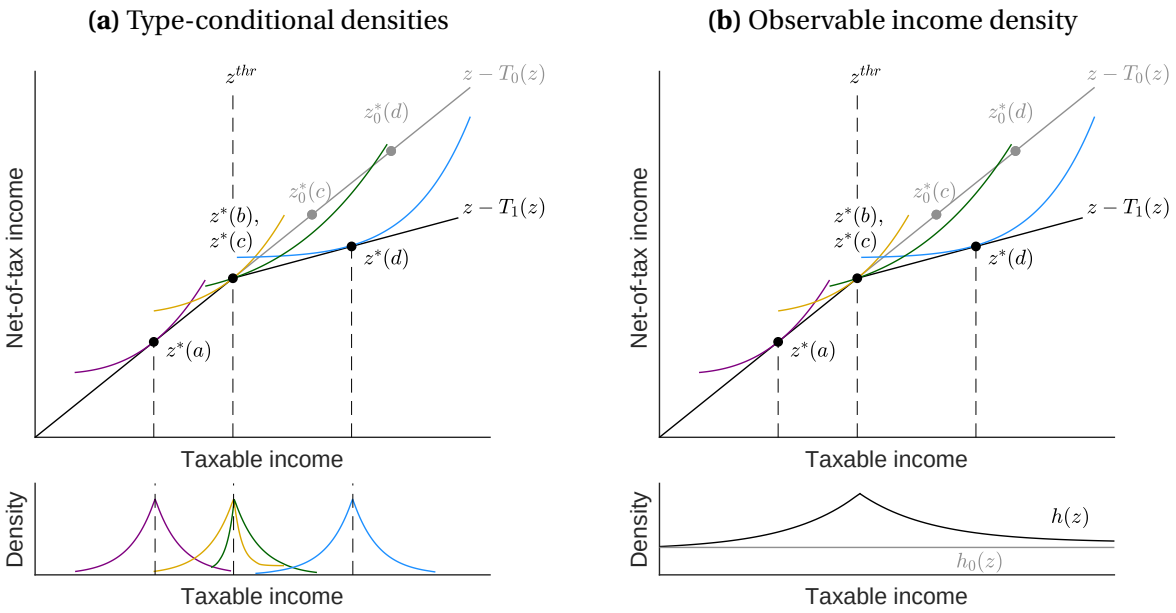
Panels (a) and (b) illustrate the construction of the indirect utility function around a progressive tax kink for the marginal non-buncher. Panel (a) shows the taxpayer's budget constraint, plotted as a solid line, where $T_0(z)$ and $T_1(z)$ are the linear income taxes below and above the bracket threshold z^{thr}, respectively. Panel (b) plots the indirect utility functions $v_0(z;b)$ and $v_1(z;b)$ which would obtain if the linear tax functions $T_0(z)$ or $T_1(z)$ applied across all incomes. Type b's indirect utility function under the kinked tax schedule, plotted as a solid line, is given by $v_0(z;b)$ below z^{thr} and $v_1(z;b)$ above z^{thr}. Panels (c) and (d) show analogous illustrations for the marginal buncher, type c. This taxpayer's optimal continuous incomes under the linear taxes $T_0(z)$ and $T_1(z)$ are denoted $z_0^*(c)$ and $z_1^*(c)$.

Figure 8: Type-conditional income density around a kink



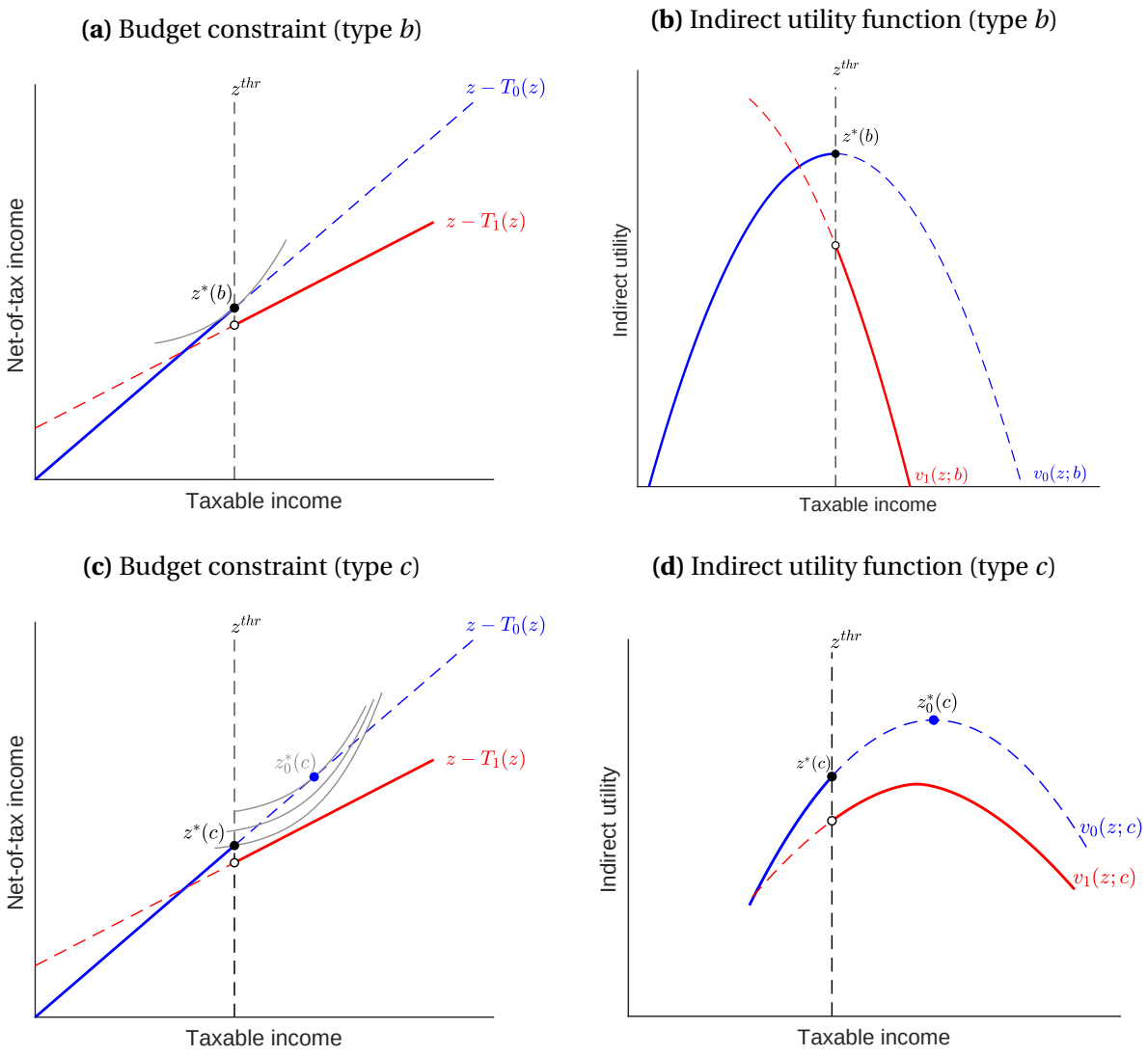
This figure illustrates how the indirect utility functions from Figure 7 are used to compute the type-conditional income densities. Panels (a) and (b) show the calculations for the marginal non-buncher and the marginal buncher (types b and c), respectively. Each panel illustrates the calculation of the type-conditional income density $g(z|n)$ at a (different) income z'. We first identify the range of incomes that dominate z' for each taxpayer, corresponding to the horizontal dashed line, and we proceed as in Figure 6. The type-conditional densities are plotted in purple. For reference, the type-conditional density under the counterfactual linear taxes $T_0(z)$ and $T_1(z)$ are plotted in blue and (in Panel (b)) in red, respectively.

Figure 9: Aggregating type-conditional densities into observable income density (kink)



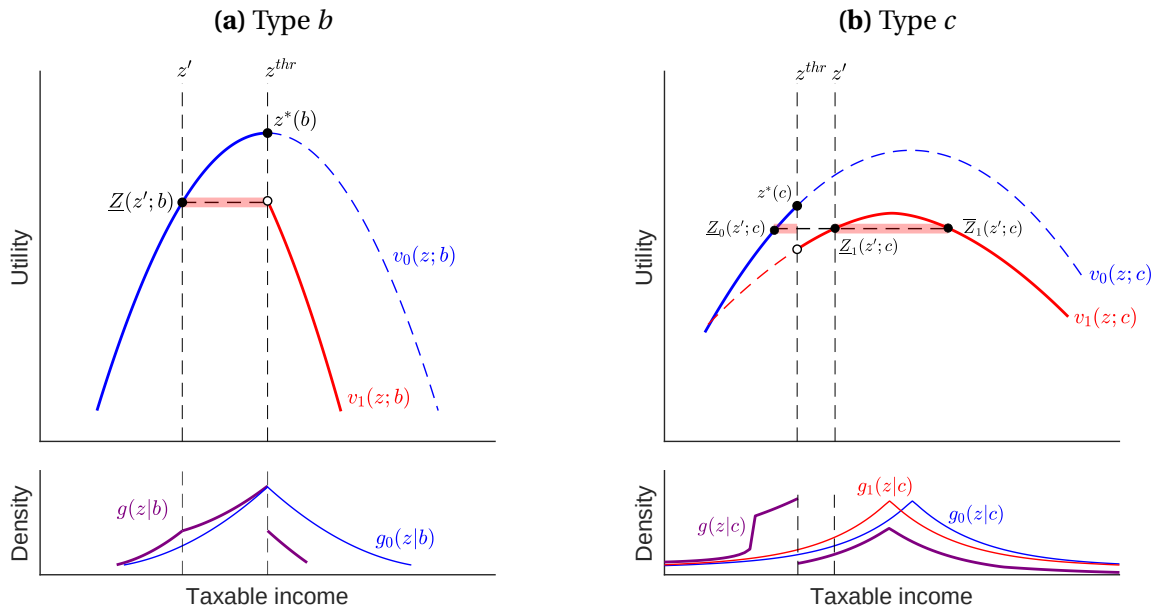
The top portion of each panel shows the optimal continuous income choice for agents of types a, b, c, and d in the presence of a kink. The lower portion of Panel (a) illustrates the overlapping type-conditional income densities of each type. The lower portion of Panel (b) sums the type-conditional densities across all types to get the observable income density $h(z)$. The counterfactual income density $h_0(z)$, which would apply under the linear tax function $T_0(z)$, is plotted in gray for reference.

Figure 10: Utility from income choices around a tax notch



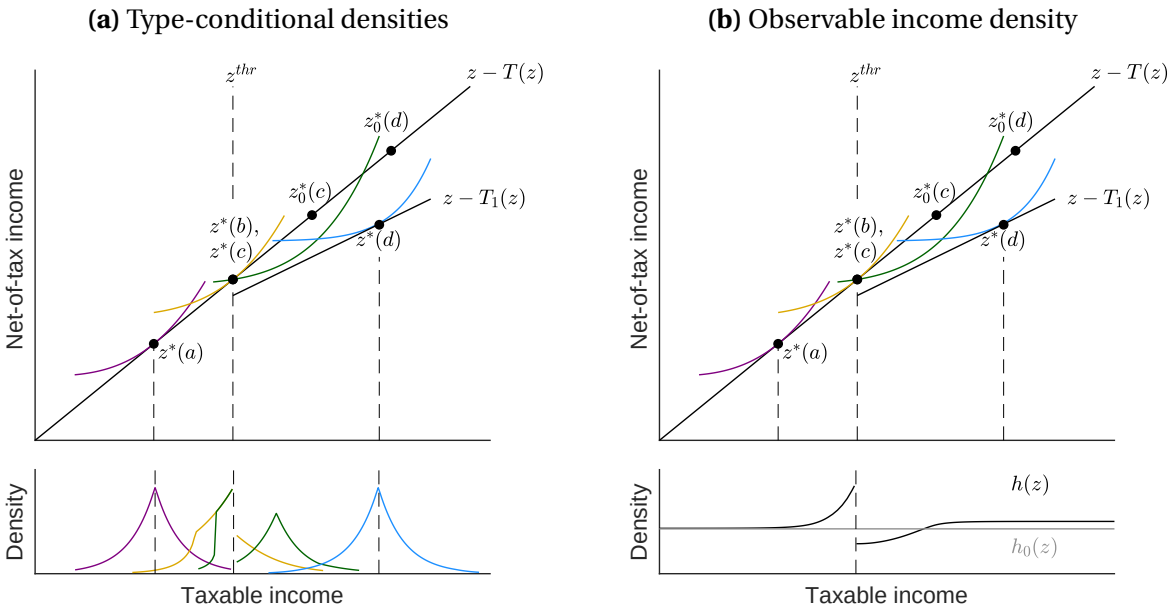
This figure is analogous to Figure 7, but in the presence of a notch, which produces a discontinuity in the indirect utility function (Panels (b) and (d)). In the case of type c, the notch produces a non-monotonic indirect utility function with two local maxima (Panel (d)).

Figure 11: Type-conditional income density around a notch



This figure is analogous to Figure 8, but in the presence of a notch. As shown in Panel (b), when the indirect utility function has multiple local maxima, the dominating income range may be a disjoint set, in which case the type-conditional density is multimodal.

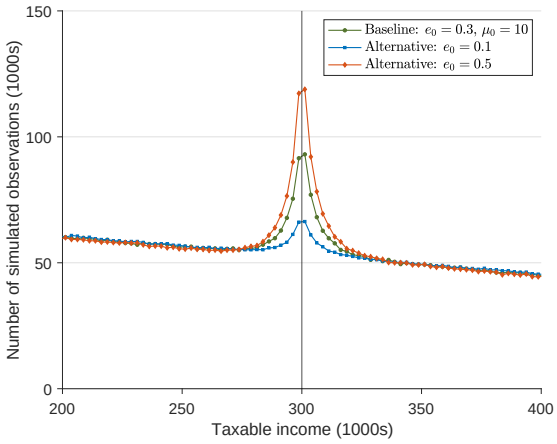
Figure 12: Aggregating type-conditional densities into observable income density (notch)



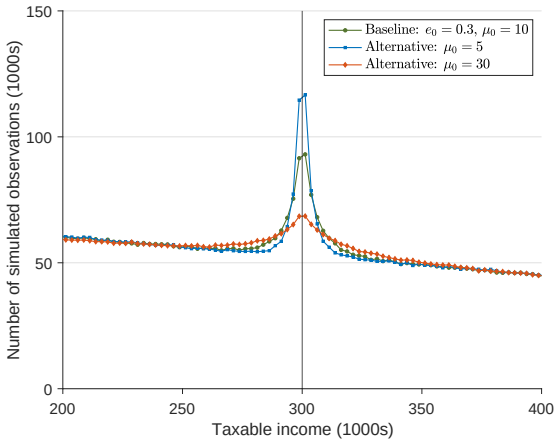
This figure is analogous to Figure 9, but in the presence of a notch. The asymmetric excess mass to the left of the threshold z^{thr} accumulates across types, producing asymmetry in the observed income density $h(z)$.

Figure 13: Simulated income densities around a bracket threshold for various parameter values

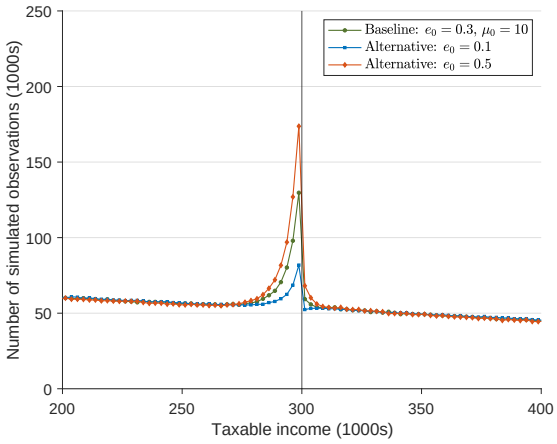
(a) Different elasticities (kink)



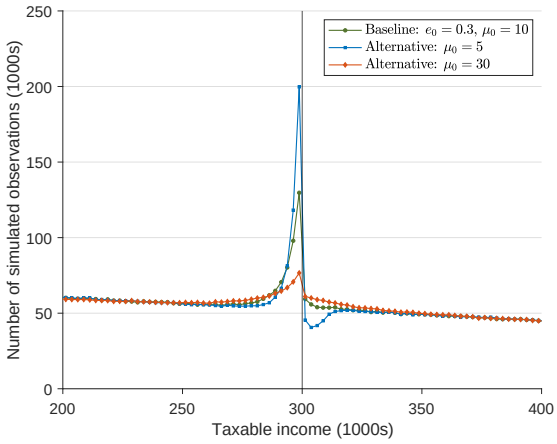
(b) Different lumpiness parameters (kink)



(c) Different elasticities (notch)



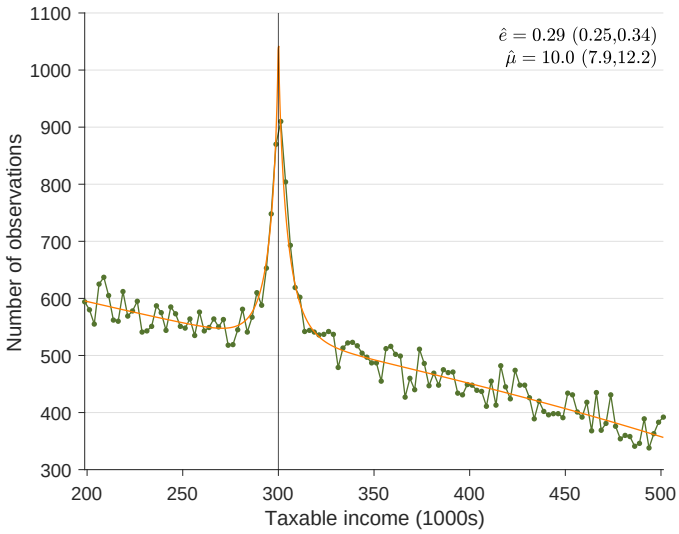
(d) Different lumpiness parameters (notch)



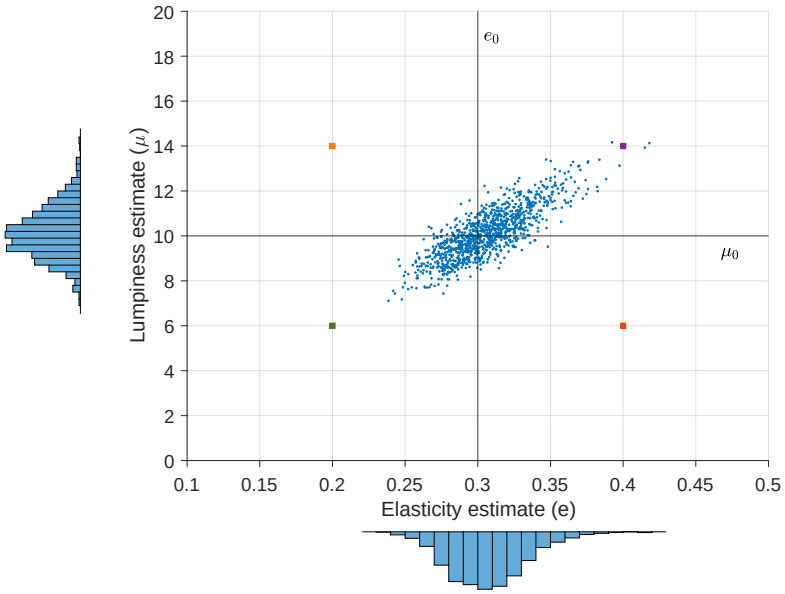
This figure plots income histograms from simulated data sets. For each simulation, we draw agents from an ability distribution with a linear density. We assume agents have a homogeneous income elasticity, e_0, and for each agent we then draw a sparse set of income opportunities in the vicinity of their preferred income from a Poisson process with a specified lumpiness parameter, μ_0. Each agent chooses the income opportunity that delivers the highest utility. We bin the resulting set of incomes to construct the income histograms displayed above. Panels (a) and (b) display simulated income histograms around a progressive tax kink for different values of e_0 and μ_0, respectively. Panels (c) and (d) display analogous histograms around a tax notch. In each case, the marginal tax rate rises from 10% to 20% at \$300,000, and for the simulations in in Panels (c) and (d), the level of tax liability increases by \$500.

Figure 14: Parameter estimates from simulated data

(a) Estimates for a single simulation round

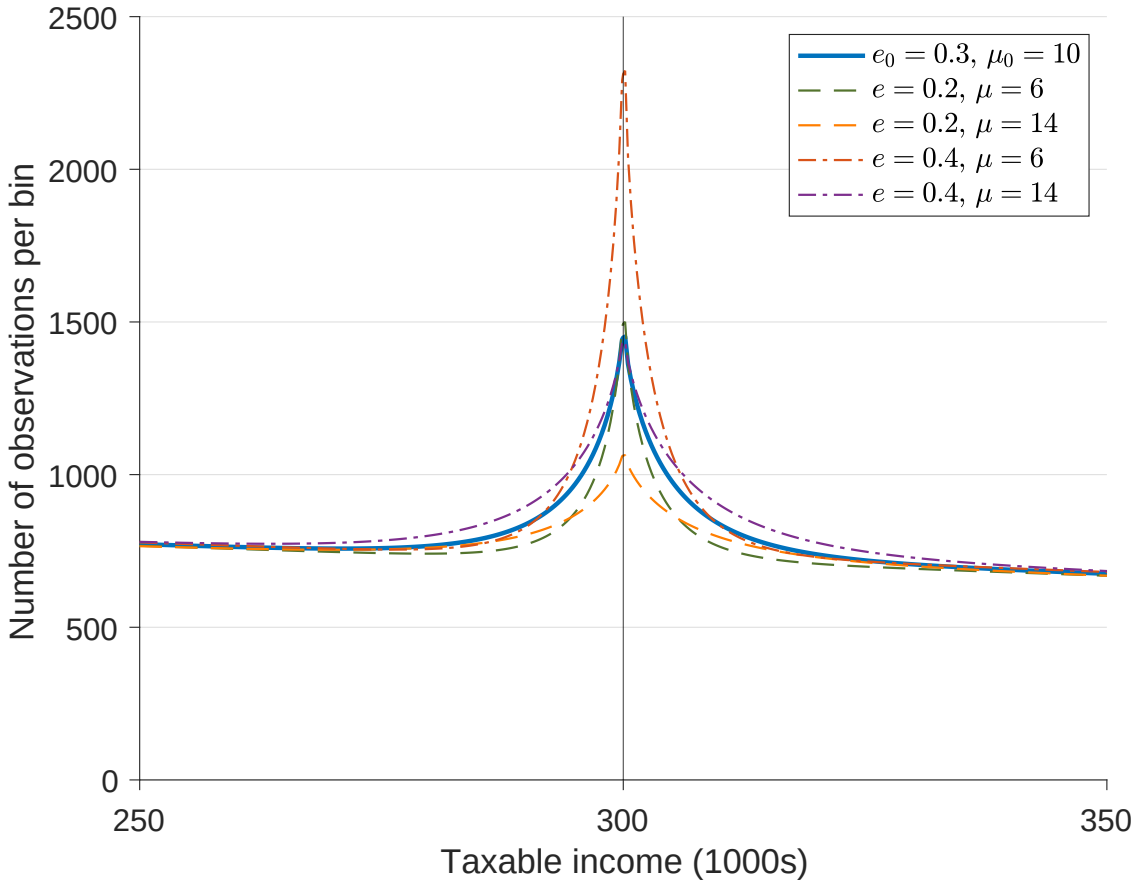


(b) Joint distribution of $\hat{e}$ and $\hat{\mu}$ estimates



This figure displays the results of applying the maximum likelihood estimation described in Section 2.5 to simulated data. These simulations are constructed as in Figure 13, with $e_0 = 0.3$ and $\mu_0 = 10$, but using a smaller number of drawn observations ($N = 100{,}000$) to produce a level of sampling noise similar to that in our empirical application in Section 4. Panel (a) plots the income histogram for one round of simulated data (in green), and the model-predicted density from the maximum likelihood estimates $\hat{e}$ and $\hat{\mu}$, whose values are reported with 95% confidence intervals in the upper corner. In Panel (b), each small point plots the combination of parameter estimates $(\hat{e}, \hat{\mu})$ from a round of simulated data like that in Panel (a). Marginal histograms of the estimates are plotted for each axis. The larger points in Panel (b) correspond to four combinations of e and μ whose densities are plotted in Figure 15 for illustrative purposes.

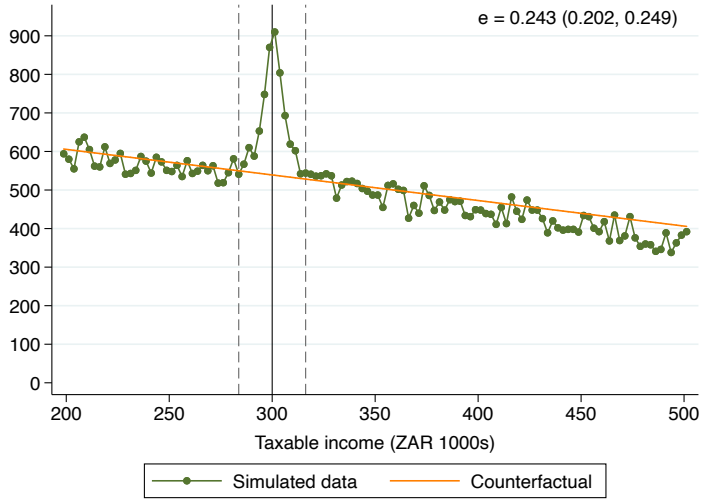
Figure 15: Income densities for different combinations of e and μ



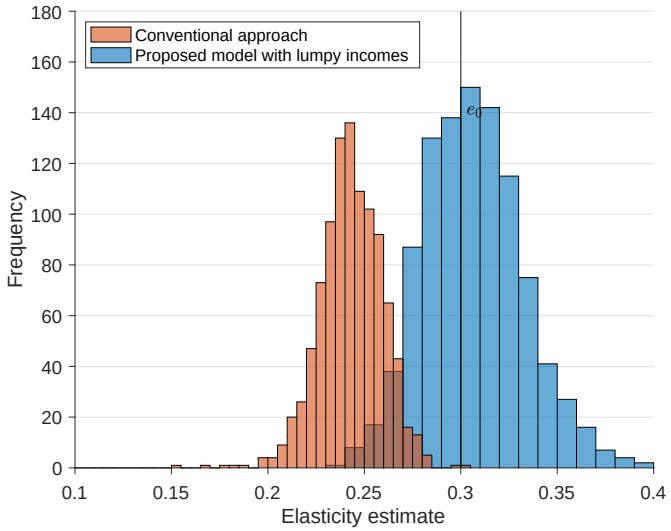
This figure plots the model-generated income density under the true parameters of the data-generating process, $e_0 = 0.3$ and $\mu_0 = \$10,000$, as well as under the four different combinations corresponding to the colored points in Figure 14b.

Figure 16: Elasticity estimates using the conventional approach

(a) Conventional estimate for a single simulation round

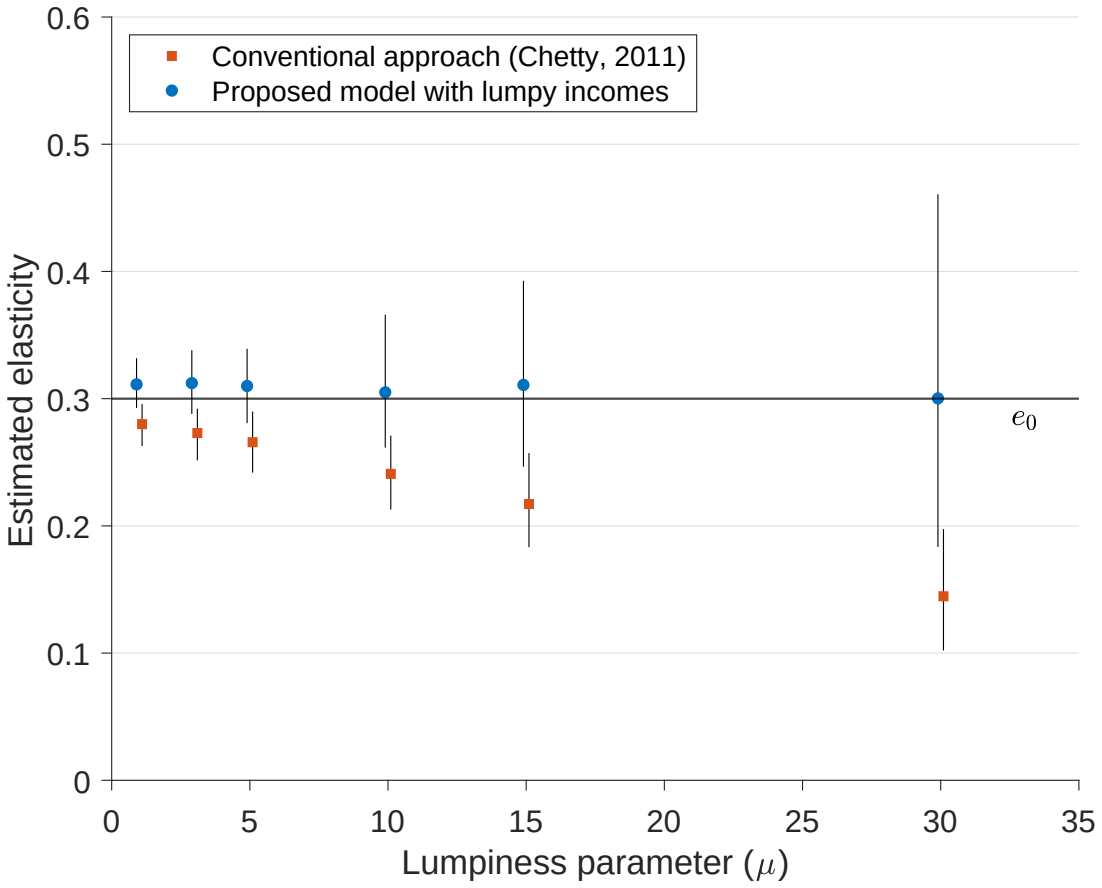


(b) Distribution of elasticity estimates under each approach



Panel (a) illustrates the conventional bunching estimator, applied to the same simulated data as in Figure 14a, resulting in an elasticity estimate well below the true value $e_0 = 0.3$. Panel (b) plots the histogram of elasticity estimates under the conventional approach (orange), and the set of estimates from the maximum likelihood method allowing for income lumpiness (blue). The vertical line at e_0 locates the true parameter value used to construct the simulated data sets.

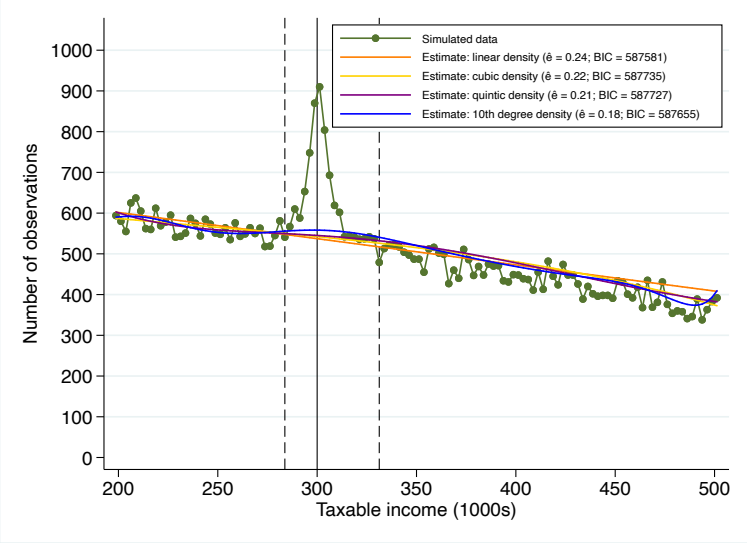
Figure 17: Elasticity estimates using the conventional approach and our approach, for different lumpiness parameters



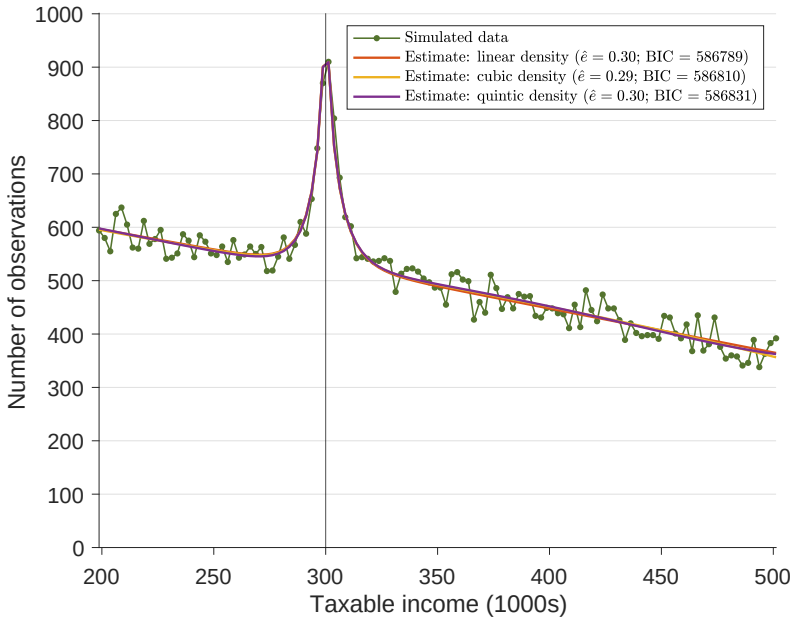
To construct this figure, we simulate 1000 rounds of data using a constant elasticity $e_0 = 0.3$ at each value of the lumpiness parameter μ_0 shown in the plots. We then estimate the elasticity $\hat{e}$ using our estimation approach and the conventional bunching estimator. The series of solid blue circles plots the average value of $\hat{e}$ using our approach across each value of μ_0; the hollow blue circles plot the 2.5th and 97.5th percentiles of our $\hat{e}$ estimates. The red squares plot the analogous series for the conventional bunching estimator.

Figure 18: Estimated elasticities assuming different polynomial degrees

(a) Continuous income choice

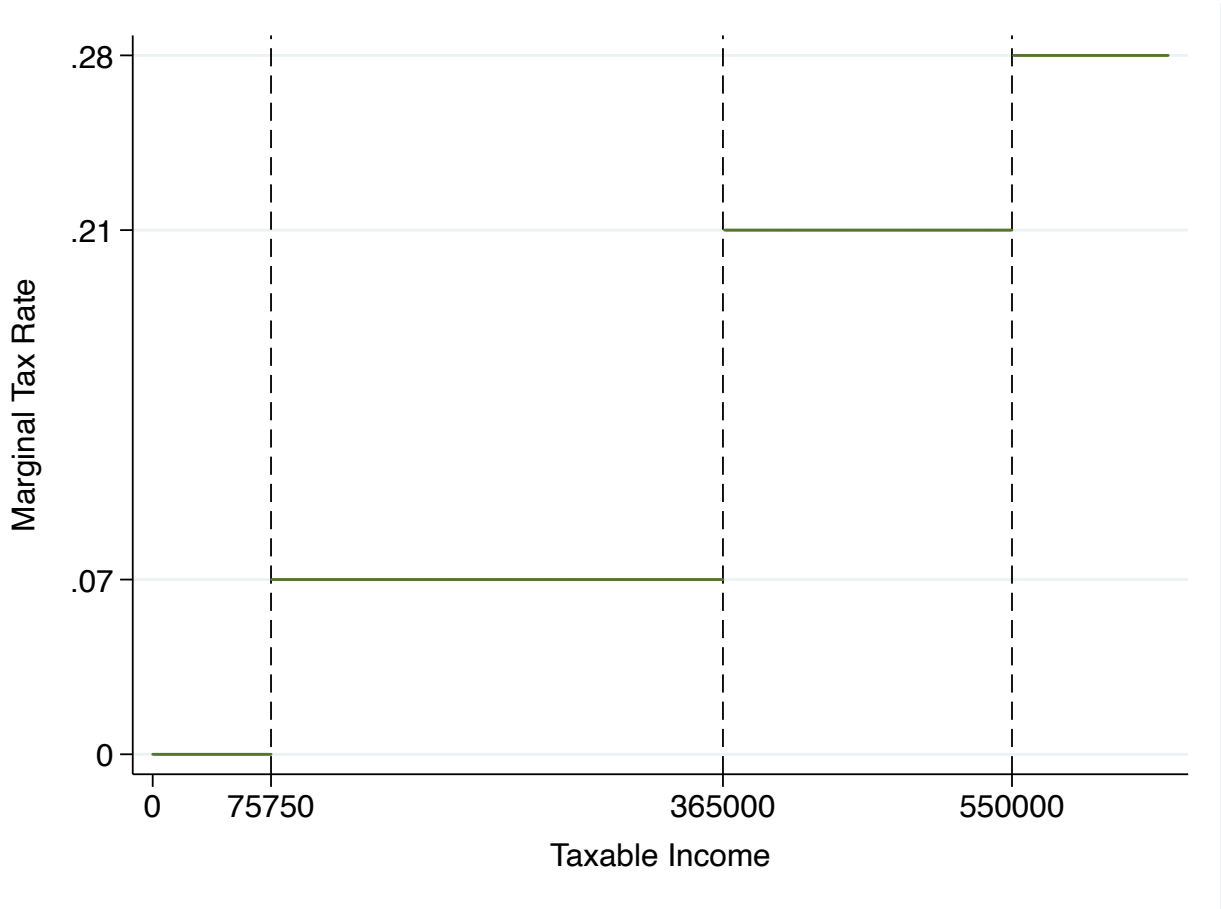


(b) Lumpy income choice



This figure reports the estimated elasticity for one round of simulated data, plotted in green, using both the conventional approach with continuous income choice (Panel a) and our estimation method with lumpy income choice (Panel b), assuming either a linear, cubic, or quintic polynomials for the ability density. The true ability density of the data-generating process is linear, with a true elasticity value of $e_0 = 0.3$.

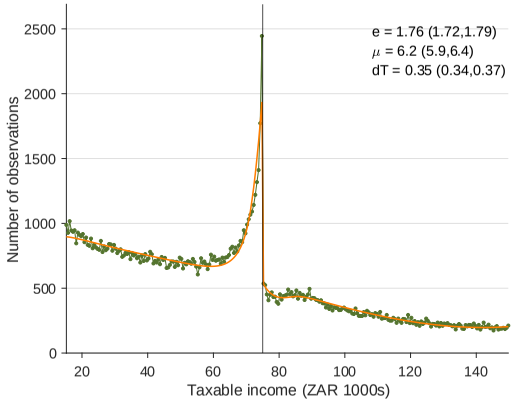
Figure 19: Tax Schedule for Small Business Corporations in 2018



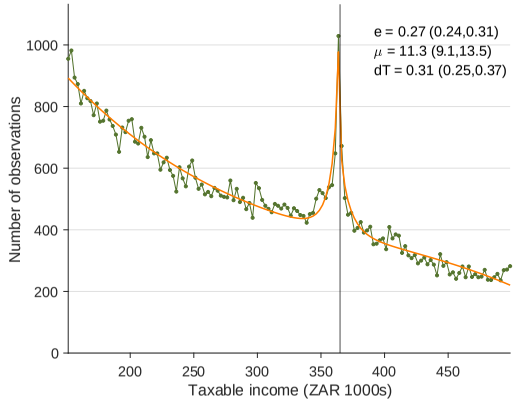
This figure shows the marginal tax rate schedule for small business corporations in South Africa in 2018. The horizontal axis is measured in South African rand (ZAR), and vertical dashed lines indicate bracket thresholds where marginal tax rates change.

Figure 20: Parameter Estimates and Predicted Histograms at each Bracket Threshold

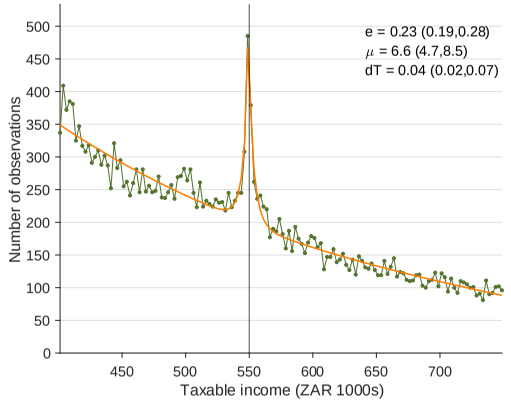
(a) Lower kink



(b) Middle kink

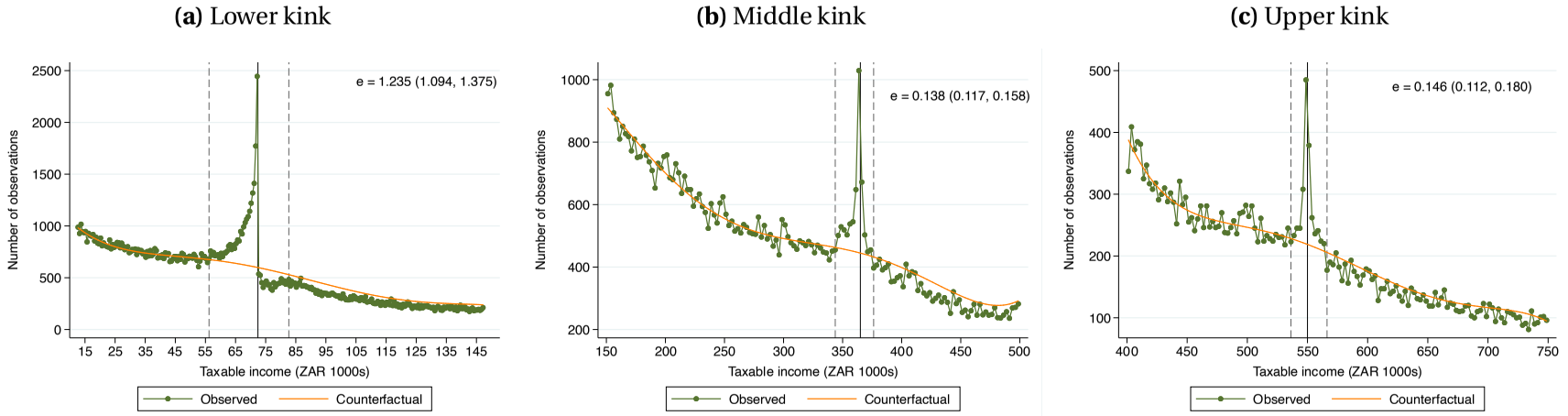


(c) Upper kink



The green points plot the empirical histogram of firms with different earnings in the data. The orange line plots the predicted density generated by the maximum likelihood estimates of the model parameters e (elasticity of taxable income), μ (average distance between income opportunities in ZAR 1000s), and dT (the estimated "as-if" discrete change in tax liability at the bracket threshold, in ZAR 1000s). Numbers in parentheses indicate the 95% confidence interval on parameter estimate.

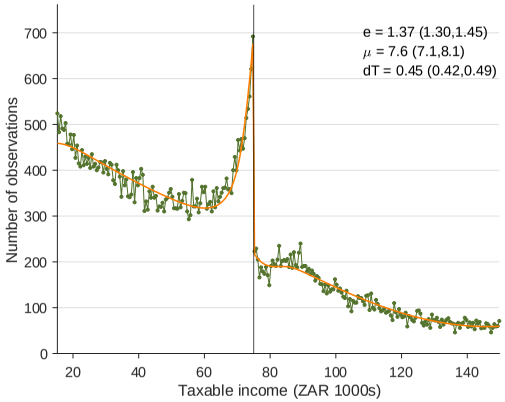
Figure 21: Elasticity estimates using conventional method (Chetty et al., 2011)



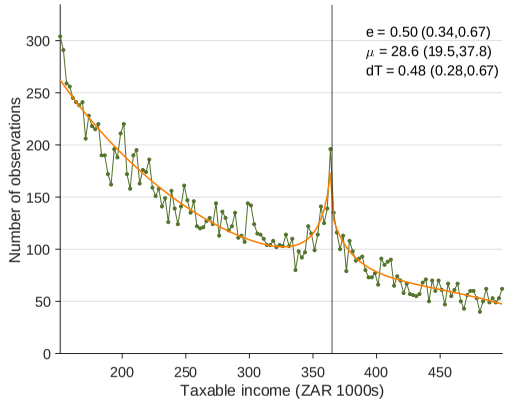
The green points plot the empirical histogram of firms with different earnings in the data. The orange line plots the predicted density generated by fitting a polynomial counterfactual to the histogram following the approach of Chetty et al. (2011). We choose the order of the polynomial and the bunching region (indicated by dashed lines) using the approach described in Section 3.2. Numbers in parentheses indicate the 95% confidence interval on parameter estimates generated by a bootstrap method where we re-sample with replacement from the underlying distribution of firms and re-estimate the model.

Figure 22: Heterogeneity by tax practitioner usage

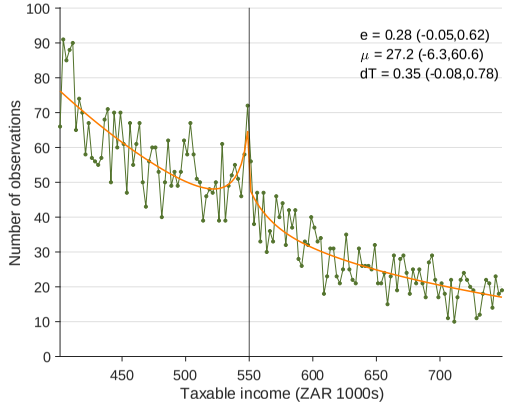
(a) No tax practitioner, lower kink



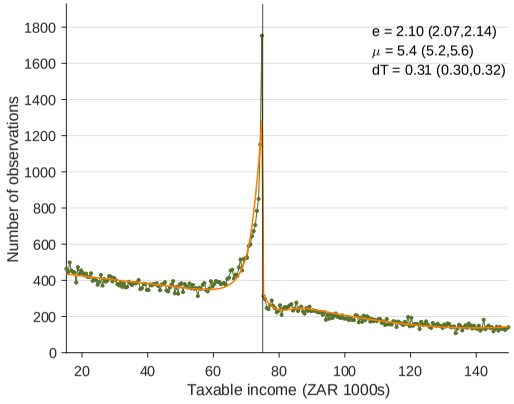
(b) No tax practitioner, middle kink



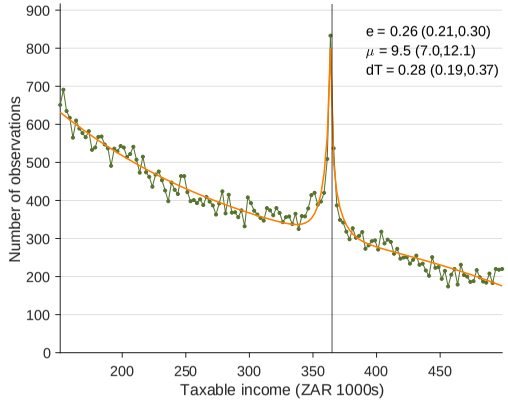
(c) No tax practitioner, upper kink



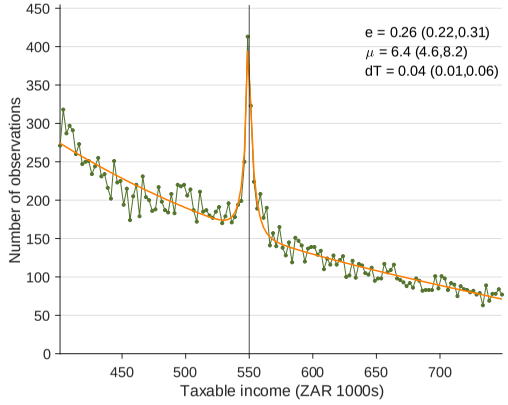
(d) Uses tax practitioner, lower kink



(e) Uses tax practitioner, middle kink



(f) Uses tax practitioner, upper kink



The format of these plots is the same as in Figure 20, but the sample is split into firms that do and do not use professional tax practitioners to prepare their tax returns.

Table 1: Primary parameter estimates

(a) Elasticity of taxable income ($\hat{e}$)

	Lower	Middle	Upper
Full population	1.76 (1.72, 1.79)	0.27 (0.24, 0.31)	0.23 (0.19, 0.28)
Without tax practitioner	1.37 (1.30, 1.45)	0.50 (0.34, 0.67)	0.28 (-0.05, 0.62)
With tax practitioner	2.10 (2.07, 2.14)	0.26 (0.21, 0.30)	0.26 (0.22, 0.31)

(b) Lumpiness parameter ($\hat{\mu}$), in ZAR 1000s

	Lower	Middle	Upper
Full population	6.2 (5.9, 6.4)	11.3 (9.1, 13.5)	6.6 (4.7, 8.5)
Without tax practitioner	7.6 (7.1, 8.1)	28.6 (19.5, 37.8)	27.2 (-6.3, 60.6)
With tax practitioner	5.4 (5.2, 5.6)	9.5 (7.0, 12.1)	6.4 (4.6, 8.2)

(c) As-if notch value (dT), in ZAR 1000s

	Lower	Middle	Upper
Full population	0.35 (0.34, 0.37)	0.31 (0.25, 0.37)	0.04 (0.02, 0.07)
Without tax practitioner	0.45 (0.42, 0.49)	0.48 (0.28, 0.67)	0.35 (-0.08, 0.78)
With tax practitioner	0.31 (0.30, 0.32)	0.28 (0.19, 0.37)	0.04 (0.01, 0.06)

This table reports our maximum likelihood estimates of the elasticity of taxable income (e), the average distance between income adjustment opportunities (μ) and the revealed preference ("as-if") value of the change in tax liability at each bracket threshold. The values of μ and dT are measured in ZAR 1000s. Results are reported separately for the aggregate population, and for the subset of firms who do and do use paid tax practitioners to prepare their tax returns.

Table 2: Comparison of results to conventional bunching estimator

	Our methodology			Chetty et al. (2011)		
	Lower kink	Middle kink	Upper kink	Lower kink	Middle kink	Upper kink
Full population	1.76 (1.72, 1.79)	0.27 (0.24, 0.31)	0.23 (0.19, 0.28)	1.23 (1.14, 1.33)	0.14 (0.12, 0.16)	0.15 (0.11, 0.18)
Without tax practitioner	1.37 (1.30, 1.45)	0.50 (0.34, 0.67)	0.28 (-0.05, 0.62)	0.76 (0.66, 0.87)	0.11 (0.08, 0.15)	0.10 (0.06, 0.15)
With tax practitioner	2.10 (2.07, 2.14)	0.26 (0.21, 0.30)	0.26 (0.22, 0.31)	1.51 (1.37, 1.65)	0.13 (0.11, 0.15)	0.14 (0.11, 0.17)

This table reports our maximum likelihood estimates of the elasticity of taxable income (e) and the same elasticity estimated using the conventional approach of Chetty et al. (2011). Confidence intervals are reported in parentheses. Results are reported separately for the aggregate population, and for the subset of firms who do and do use paid tax practitioners to prepare their tax returns. For more details on the implementation and estimation of the conventional method, see section 3.2.

A Appendix

A.1 Bunching patterns produced by other models of frictions

As noted in the introduction, the possibility that diffuse bunching is driven by lumpiness in income opportunities was introduced in Saez (1999), the working paper that preceded Saez (2010). However, a number of other income adjustment frictions have also been proposed as potential contributors to bunching diffusion. Several such models are reviewed in Kleven (2016). This appendix discusses the ability of these alternative models to produce the bunching patterns observed in settings like Saez (2010), Kleven and Waseem (2013), and our empirical evidence from South Africa in Section 4.

One class of models involves adjustment or attention costs. For example, Chetty et al. (2011) presents a model in which taxpayers draw an initial job offer from a distribution and can reoptimize to their continuous optimum by paying a random utility cost. In the context of notches, the model proposed in Kleven and Waseem (2013) involves a subset of taxpayers who are unresponsive to a tax notch, either because of adjustment or attention costs, in order to rationalize observed positive mass in range of dominated incomes above a notch. Although these models can explain why some taxpayers whose ideal income is at the kink point do not in fact bunch there, they do not naturally produce the patterns of diffusion like those displayed in Figure 1. The reason is simple: in these models, taxpayers who *do* pay the adjustment or attention cost reoptimize as in the continuous model, and thus all such taxpayers bunch precisely at the kink point.

We therefore focus on an alternative friction that has received substantial attention—that of *uncertainty* or *imperfect targeting* in the income process. Saez (1999) also proposed a simple model of this sort of friction, which posits that agents face uncertainty about the relationship between their actions and ultimate income realizations, and as a result they can only imperfectly target their optimal income choices. This appendix explores the bunching patterns produced by that model.

We reproduce the simulation exercise in Saez (1999) by simulating a data set of taxable income around a kink and a notch by varying the underlying parameters of the data-generating process. We stipulate a tax schedule with a tax bracket threshold at $z^{thr} = 4{,}000$, at which the marginal tax rate rises from $t_0 = 0\%$ to $t_1 = 10\%$. We set a baseline elasticity of $e = 0.3$ and also assume a uniform distribution of abilities around the kink point. We retain the same parameters for the simulations around a notch, but allow for notch value, $dT = 80$.

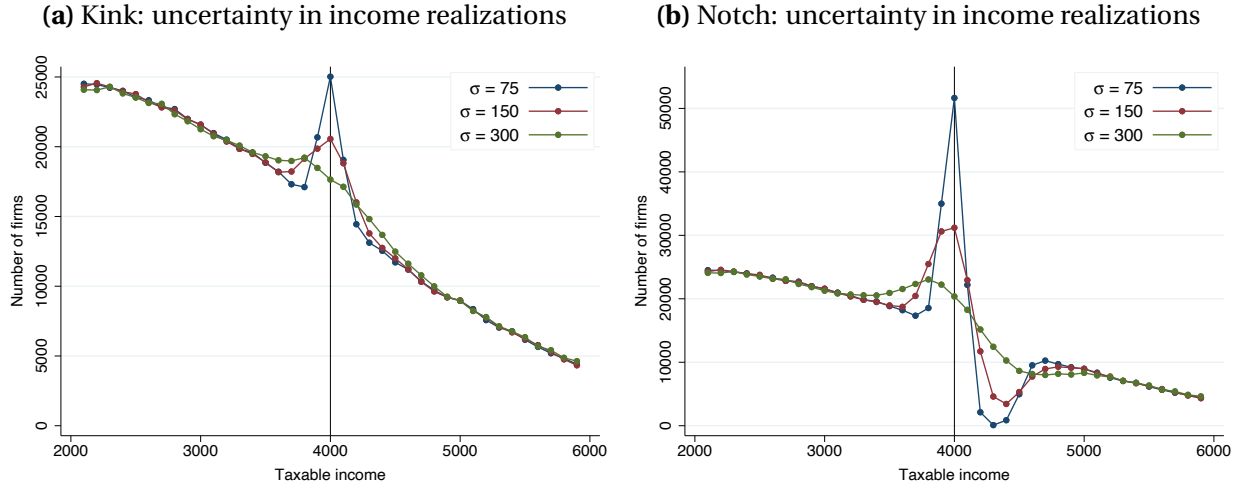
Following Saez (1999), we consider a model in which individuals select their target income, z_{targ}, as in the continuous model, but this target is then altered by a random shock $\varepsilon \sim N(0, \sigma^2)$, so that their realized income is $z = z_{targ} + \varepsilon$. The effective level of income, z, for a given value

of σ thus lies in the interval $(z_{targ} - 1.96 * \sigma, z_{targ} + 1.96 * \sigma)$ with 95% probability. As in Saez (1999), we assume taxpayers are naive about the shock and select their continuous optimal income $z^*(n)$ as their target. For a given ability density $f(n)$, income elasticity e, and income uncertainty variance σ^2, this income process can be simulated by drawing a vector of abilities n as described in Section 3, computing each n's income target $z^*(n)$, and then adding to each income target a random shock $\varepsilon \sim N(0, \sigma^2)$.

Figure A1 plots income histograms resulting from simulations with varying degrees of income uncertainty, $\sigma = 75$, 150, and 300. Panel (a) reports a histogram around a pure kink, while Panel (b) plots the histogram around a notch. Consistent with the results in Saez (1999), this model produces diffusion at kink points. Larger values of the income shock standard deviation σ produce more diffusion in bunching.

However, as illustrated in Panel (b), in the case of a notch, the uncertainty model predicts a pattern of bunching that looks different from that produced by the income lumpiness model. Under income uncertainty, the diffuse mass spills over to the right of the threshold, because half of the bunching agents—those who target their preferred income exactly at the bracket threshold—realize incomes above the threshold due to the uncertainty shock. This bunching shape contrasts with the pattern observed at the notches in Kleven and Waseem (2013), which exhibit diffusion to the left of the threshold and a discontinuous drop to the right, and which resemble the "notch-like" pattern of bunching at the lower kink in this paper.

Figure A1: Bunching patterns around kinks and notches under the uncertainty model



This figure presents simulated income histograms of bunching patterns under the uncertainty model. Panel (a) shows bunching around a kink in the tax schedule, and Panel (b) shows bunching around a notch. We set $t_0 = 0\%$, $t_1 = 0.1\%$, $e = 0.3$ and $z^{thr} = 4{,}000$. In the simulations around a notch, we set $dT = 80$. Plot results for three values of the standard deviation of the income uncertainty shock: $\sigma = 75$, 150, and 300.

A.2 Alternative specifications for the conventional approach

In Section 3.2, we compare the elasticities produced under our approach to the elasticity estimates produced under the approach developed in Chetty et al. (2011), one of the most widely used conventional bunching estimators. This approach involves fitting a flexible polynomial to the observed data, excluding the observations in the bunching region, and uses this to construct a single counterfactual which represents the counterfactual distribution that would occur if the lower tax rate below the kink threshold also applied above the threshold. As we discuss in the main text, this approach imposes an "integration constraint" such that the total integral of population across the empirical distribution equals the total integral under the counterfactual distribution. The integration constraint makes the assumption that all of the bunching mass comes from the income distribution in the underlying histogram and rules out the possibility that any mass shifts beyond the region depicted in the histogram. Given the counterfactual of Chetty et al. (2011) assumes that the lower tax rate below the kink applies above the kink, this means that the entirety of the bunching mass gets reallocated above the kink into the income bins depicted in the histogram. The practical implication of this is that the counterfactual density is shifted upward in order to ensure that the total integral of population across the empirical distribution equals the total integral under the counterfactual distribution

In this section, we compare our estimator to other conventional bunching estimators that differ in how they construct counterfactuals, namely Saez (2010) and Mortenson and Whitten (2016), the working paper that preceded Mortenson and Whitten (2020). We illustrate the differences between these approaches in Figure A2a. The approach developed in Saez (2010) constructs two linear counterfactuals on either side of the kink with the assumption that the densities are uniformly distributed on either side of the threshold. In order to construct the counterfactual, the approach takes the average value of the densities that occur outside of the bunching window and extrapolates that density through to the kink threshold. This is done on either side of the kink resulting in two counterfactuals. An alternative approach is developed in Mortenson and Whitten (2016) who construct a piecewise linear counterfactual on either side of the kink, in a similar vein to Saez (2010), but allow for that counterfactual to take into account the slope of the observed densities on either side of the kink. Finally, we also consider an implementation of the approach in Chetty et al. (2011) where we do not impose the "integration constraint." This allows for the possibility that the bunching mass may be reallocated to income bins beyond the region depicted in the histogram, which would cause the total integral under the counterfactual distribution to be smaller than the total integral of population across the empirical distribution. The practical implication of this is that the counterfactual distribution is shifted downward relative to the approach which imposes the integration constraint, as is depicted in Figure A2a.

Next, we compare the elasticities produced under these four approaches to our estimates for varying values of the lumpiness parameter. We report these results in Figure A2b. Imposing the integration constraint in the Chetty et al. (2011) approach produces lower elasticities than when the constraint is not imposed. Intuitively, by imposing the constraint, the counterfactual density is shifted upward, which causes the estimate of bunching to fall, leading to a lower elasticity. The Mortenson and Whitten (2016) elasticities are very similar to the Chetty et al. (2011) elasticities without the integration constraint. In that sense, the specification is nearly identical to the counterfactual specification in Mortenson and Whitten (2016), apart from the fact that the latter approach estimates two counterfactuals on either side of the threshold, thereby allowing for a different slope on the counterfactual on either side of the kink threshold. The visual similarity between these counterfactuals is evident in Figure A2a. Out of all of the conventional approaches, the Saez (2010) approach produces the largest elasticities. The reason for this becomes evident when considering the counterfactuals produced in Figure A2a. Given the empirical distribution slopes downwards, by assuming uniformly distributed densities, the Saez (2010) approach produces a counterfactual that is significantly lower than the other counterfactuals in the bunching region above the kink, leading to a higher measure of bunching, and a higher estimated elasticity.

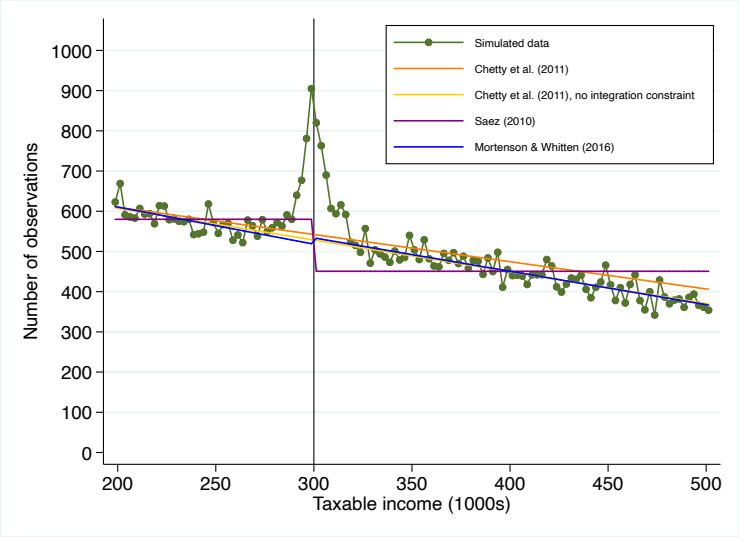
For smaller values of the lumpiness parameter, only Mortenson and Whitten (2016) and the Chetty et al. (2011) approach without the integration constraint can recover the true elasticity. However, for large values of the lumpiness parameter, not even these approaches are able to recover the true elasticity and exhibit a significant downward bias, whereas our approach can consistently recover the elasticity, irrespective of the extent of lumpiness in the observed empirical distribution.

A.3 Details of South African small business corporations

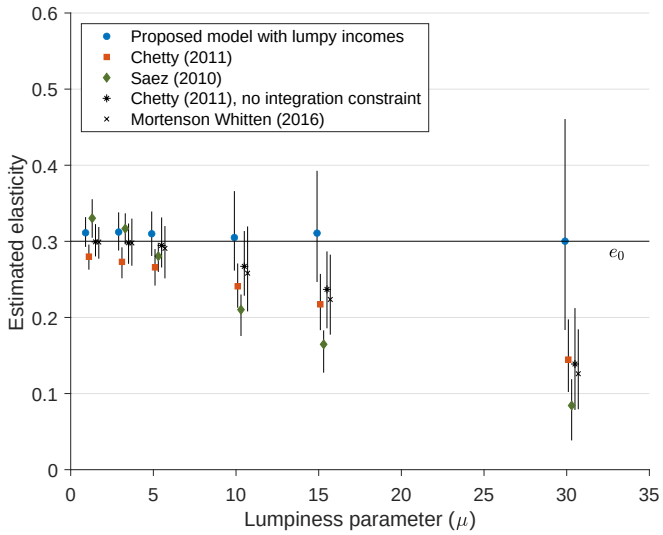
In South Africa, small business corporations (SBCs) face a progressive schedule of marginal tax rates that are lower than those applied to other firms. Table A1 reports the full schedule of SBC tax rates in each year from 2010 to 2018. In addition to qualifying for lower tax rates, SBCs are also eligible for an accelerated depreciation allowance, and they are granted more generous deductible allowances for movable assets.

Figure A2: Counterfactuals and elasticity estimates using various conventional approaches and our approach, for varying lumpiness parameters

(a) Alternative approaches to constructing counterfactuals



(b) Comparing the elasticity estimates



In Panel (a), we illustrate the counterfactuals produced under four different conventional bunching approaches to estimating elasticities for a simulated dataset where $\mu = 10$. In Panel (b) we simulate 100 rounds of data using a constant elasticity $e_0 = 0.3$ at each value of the lumpiness parameter μ_0 shown in the plots. We then estimate the elasticity $\hat{e}$ using our estimation approach and four conventional bunching estimators. The vertical lines indicate the 95% confidence intervals for the $\hat{e}$ estimates. For the conventional methods, we adapt the automated bunching window approach in Bosch, Dekker and Strohmaier (2020) in order to account for each method's approach to constructing a counterfactual distribution.

Table A1: Small Business Corporation Tax Schedule, 2010–2018

Tax Year	Taxable income	Marginal tax rate
2010	0 - 54,200	0%
	54,200 - 300,000	10%
	Above 300,000	28%
2011	0 - 57,000	0%
	57,000 - 300,000	10%
	Above 300,000	28%
2012	0 - 59,570	0%
	59,570 - 300,000	10%
	Above 300,000	28%
2013	0 - 63,556	0%
	63,556 - 350,000	7%
	Above 350,000	28%
2014	0 - 67,111	0%
	67,111 - 365,000	7%
	365,001 - 550,000	21%
	Above 550,000	28%
2015	0 - 70,700	0%
	70,700 - 365,000	7%
	365,001 - 550,000	21%
	Above 550,000	28%
2016	0 - 73,650	0%
	73,651 - 365,000	7%
	365,001 - 550,000	21%
	Above 550,000	28%
2017	0 - 75,000	0%
	75,001 - 365,000	7%
	365,001 - 550,000	21%
	Above 550,000	28%
2018	0 - 75,750	0%
	75,751 - 365,000	7%
	365,001 - 550,000	21%
	Above 550,000	28%

This table indicates the small business corporation (SBC) graduated income tax system for the tax years 2010–2018. Tax years run from April 1 to March 31.

Businesses are eligible to register as an SBC if they meet each of the following requirements:[22]

- The business is a close corporation, co-operative, private company or personal liability company.[23]
- All shareholders are natural persons (i.e, individuals and not companies or other legal structures) during the year of assessment.
- No shareholders may hold any shares or hold any interest in any other company, subject to certain exemptions. Some of these exemptions include listed companies, collective investment schemes and venture capital companies.
- The gross income of the company must not exceed R20 million for the year of assessment.
- The company may not be a personal service provider.[24]
- Investment income and income from rendering personal services may account for a maximum of 20% of all receipts, accruals and capital gains.[25]

SBCs account for approximately 26% to 31% of the total number of corporate tax filings between 2010 and 2018. Table A2 compares SBCs to other types of businesses in South Africa. It reports summary statistics for three groups of firms: non-SBCs, SBCs and size-matched non-SBCs, where the latter group consists of non-SBC businesses with revenues below the R20 million SBC eligibility threshold. Size-matched non-SBCs therefore comprise two types of firms: (i) firms who are eligible to register as an SBC but do not, either intentionally or because they are are unaware of the SBC program, and (ii) firms who are eligible to register as an SBC under the gross revenue requirement, but who do not meet one of the other requirements listed above. We are unable to distinguish between these two types of firms in our data. While SBCs account for 38% of all companies, size matched non-SBCs account for over half of all companies we observe. This discrepancy can be accounted for by the fact that since firms are not taxed when making losses and the number of loss-making firms greatly outnumbers the number of profit-making firms, many SBC eligible firms do not register given they make a loss and as such there

[22] More information on these requirements can be found in an interpretation note provided by SARS at `https://www.sars.gov.za/wp-content/uploads/Legal/Notes/LAPD-IntR-IN-2018-08-Arc-08-IN9-Issue-6-Small-Business-Corporations.pdf`.

[23] A close corporation is a firm that was required to have 10 or less owners. After 2019, new companies could no longer incorporate as close corporations, but previously registered close corporations could maintain this form.

[24] Personal service providers refer to companies that have less than 3 employees and where more than 80% of the company income is derived from one client.

[25] Personal services refer to any company services performed personally by any person who holds an interest in that company when that company employs less than 3 employees. In this scenario, the tax authority deems the income being generated to be a function of the personal skill of that individual and not the company.

is no incentive to SBC status. The size of the SBC sector is therefore a subset of the eligible SBC population, which would be closer in size to the total number of all small- and medium-sized enterprises (SMMEs), which stands at over 90% of all formally registered companies.[26]

[26]This only takes into account formally registered firms. Given South Africa's large informal economy, the true number of SMMEs will be even larger.

Table A2: Summary statistics for businesses filing corporate income tax returns, 2014–2018

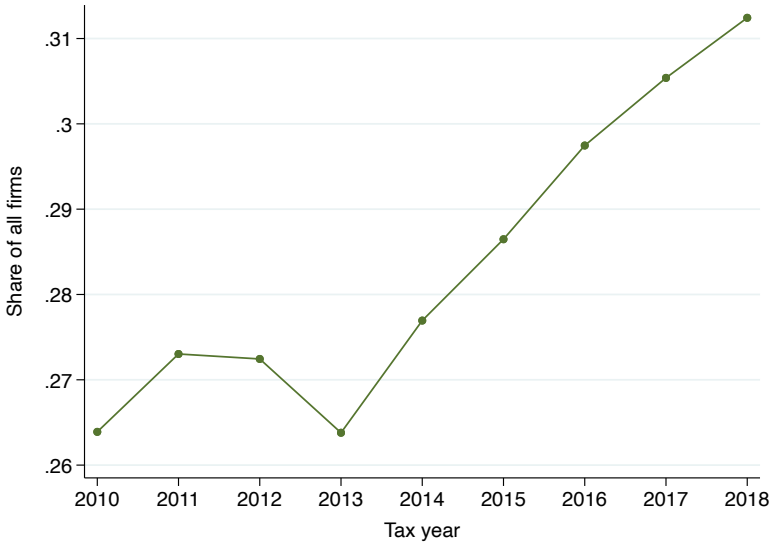
Company Type	Non-SBC	Non-SBC Size Matched	SBC
Turnover (in R'000)	121,249.1	3,028.22	2,628.6
	(521,791.0)	(4,275.89)	(3,531.4)
Expenses (in R'000)	33,713.74	1,684.13	1,244.78
	(281,924.3)	(2,632.84)	(1,785.2)
Assets (in R'000)	73,928.56	3,422.88	1,264.86
	(658,585.4)	(16,459.2)	(2,344.89)
Liabilities (in R'000)	51,659.0	2,231.0	673.0
	(510,496.1)	(13,190.15)	(1,708.29)
Inventory (in R'000)	10,782.82	284.19	173.72
	(69,005.84)	(2,017.28)	(651.83)
Cash (in R'000)	7,185.93	312.86	199.56
	(47,614.75)	(1,606.54)	(630.77)
Net profit (in R'000)	5,280.92	92.31	125.68
	(31,990.6)	(1,066.98)	(502.82)
Number of employees	90.75	4.97	3.93
	(579.44)	(19.11)	(11.69)
Number of salaried directors	2.26	1.47	1.32
	(5.64)	(0.83)	(0.62)
Taxable income (in R'000)	-462.12	-346.43	6.39
	(150,598.8)	(3,541.95)	(753.68)
Tax liability (in R'000)	1,510.64	49.41	30.56
	(6,659.3)	(176.17)	(119.75)
% of firms with a salaried director	35.18%	13.70%	17.34%
% of firms with a tax practitioner	73.09%	71.12%	64.16%
Number of unique tax returns	137,872	653,755	457,198
Share of Tax Revenue	81.82%	12.69%	5.49%
Number of unique companies	41,289	238,830	172,440
Share of companies	9.12%	52.77%	38.10%

This table reports summary statistics for corporate income tax returns in South Africa between 2014 and 2018 for 3 groups of firms: "Non-SBCs", "Size Matched Non-SBCs" and "SBCs." "Size Matched Non-SBCs" represent "Non-SBCs" with revenues below R20 million, the "SBC" eligibility threshold. "Size Matched Non-SBCs" and "Non-SBCs" are mutually exclusive categories. Standard deviations are shown in parentheses.

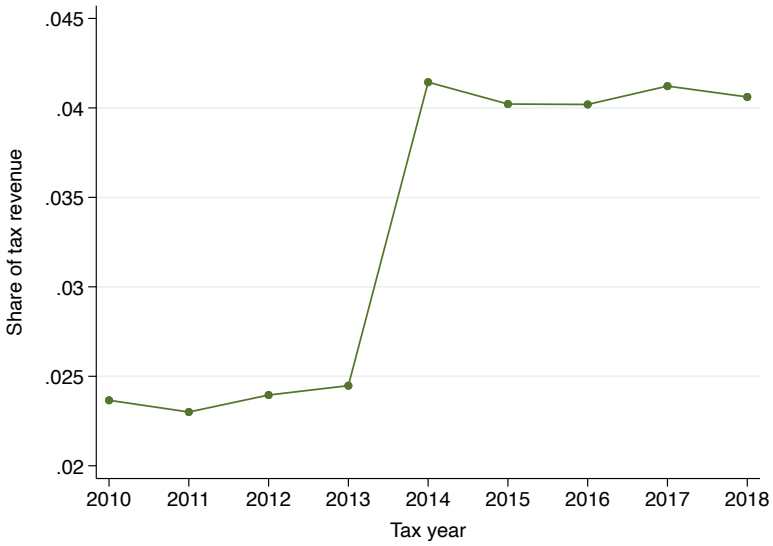
The share of SBCs has risen over time. Figure A3a shows a clear upward trend since 2013, with SBCs accounting for over 31% of all tax filings in 2018. While large in number, the contribution of SBC tax revenue to total corporate income tax revenue, as shown in Figure A3b, is more modest, with SBCs accounting for 3% of overall corporate tax revenue on average during our sample period. This share has however increased since 2014, rising from around 2.5% to 4%. The jump in share of tax revenue from SBCs between 2013 and 2014 coincides with the year in which the gross income requirement for SBC eligibility was increased from R14 million to R20 million; this allowed a larger number of companies to register as SBCs and therefore increased the fraction of corporate tax revenue originating from SBCs.

Figure A3: Small Business Corporation (SBC) prevalence and contribution to corporate income tax revenue

(a) SBCs as a share of all firms



(b) SBCs contribution to total corporate income tax revenue



Panel (a) shows the share of SBC tax filings relative to all corporate tax filings between tax years 2010 and 2018. Panel (b) shows the share of corporate tax revenue contributed by SBC's as a percentage of total tax revenue between tax years 2010 and 2018. In 2014, the income ceiling for SBCs was raised from R14 million to R20 million, generating a substantial increase in their contribution to total revenues. To calculate tax revenue, we sum up the tax liability of firms. We do not observe whether a payment was made and as a result, the figure should be viewed as indicative of taxes owed.