# BETTER BUNCHING, NICER NOTCHING

Marinho Bertanha* Andrew H. McCallum† Nathan Seegert ‡

First draft: August 5, 2017
This draft: August 14, 2020

**Abstract**

We study the bunching identification strategy for an elasticity parameter that summarizes agents' response to changes in slope (kink) or intercept (notch) of a schedule of incentives. A notch identifies the elasticity but a kink does not, when the distribution of agents is fully flexible. We propose new non-parametric and semi-parametric identification assumptions on the distribution of agents that are weaker than assumptions currently made in the literature. We revisit the original empirical application of the bunching estimator and find that our weaker identification assumptions result in meaningfully different estimates. We provide the Stata package `bunching` to implement our procedures.

**JEL:** C14, H24, J20

**Keywords:** partial identification, censored regression, bunching, notching

*Corresponding author. Department of Economics, University of Notre Dame, 3060 Jenkins Nanovic Halls, Notre Dame IN 46556. Email: mbertanha@nd.edu. Website: www.nd.edu/∼mbertanh.

†Trade and Quantitative Studies Section, International Finance, Board of Governors of the Federal Reserve System. Email: andrew.h.mccallum@frb.gov. Website: www.andrewhmccallum.com.

‡Eccles School of Business, University of Utah. Email: nathan.seegert@eccles.utah.edu. Website: www.nathanseegert.com.

## 1 Introduction

Estimating agents' responses to incentives is a central objective in economics and many other social sciences. A continuous distribution of agents that face a piecewise-linear schedule of incentives results in a distribution of responses with mass points located where the slope or intercept of the schedule changes. For example, a progressive schedule of marginal income tax rates induces a mass of heterogeneous individuals to report the same income at the level where marginal rates increase. Many studies in economics use mass points in the response distribution to recover primitive parameters that govern agents' responses to incentives.

Pioneering work by Saez (2010), Chetty, Friedman, Olsen, and Pistaferri (2011), and Kleven and Waseem (2013) develop bunching estimators to use mass points in response distributions to recover primitive parameters. These estimators are widely applied in economics and rely on the idea that a mass point is larger, the more responsive agents are to incentives. The size of the mass point, however, also depends on the unobserved distribution of agents' heterogeneity. Current methods are only able to map the size of mass points to primitive parameters because they make specific assumptions about the unobserved distribution.

This paper places bunching estimators on a statistical foundation and makes three contributions on the identification of a primitive parameter that summarizes agents' responses to incentives. First, we clarify how the mapping of observed variables to an elasticity parameter depends on assumptions about the unobserved distribution of heterogeneity. The elasticity parameter captures the log percentage change of a response to a log percentage change in an incentive. A change in the intercept of the incentive schedule admits non-parametric point identification of the elasticity but a change in slope does not. Second, we examine the assumptions made by current bunching methods and propose weaker assumptions for partial and point identification of the elasticity. Third, we revisit the original empirical application of the bunching estimator, which is in the economics literature

that examines the largest means-tested cash transfer program in the United States —the Earned Income Tax Credit (EITC). Our weaker assumptions about the unobserved distribution of heterogeneity result in meaningful changes in estimates of individual responses to taxes.

Our first contribution is to clarify the importance of assumptions about unobserved heterogeneity for the identification of the elasticity. Many existing estimates are based on an agent optimization problem with a piece-wise linear constraint that has one change in slope or intercept. Slope changes in the constraint are often referred to as ''kinks'' while intercept changes are often called ''notches.'' We generalize the constraint of the agent's problem to a schedule with multiple changes in intercepts and slopes because agents typically encounter a combination of both kinks and notches.

We highlight three insights about identification with kinks and notches assuming a non-parametric family of distributions for unobserved heterogeneity that have continuous probability density functions (PDFs). First, if the constraint has at least one notch, it is possible to point identify the elasticity. Identification comes from using the empty interval in the support of the observed distribution that is created by agents' responses to a notch. Second, point identification is impossible if the incentive schedule only contains kinks. Identification is impossible because there always exists an unobserved distribution that reconciles any elasticity with the observed distribution of responses. Third, inference methods designed for one kink can be applied in cases with multiple kinks at each kink separately, as long as there are no notches preceding the kink under study. This is because the range of heterogeneous agents that bunch at a kink is the same regardless of whether it is the first kink in the schedule or if it is followed by another kink. In contrast, the range of agents that bunch at a kink changes if that kink is preceded by a notch. Thus, methods designed for one kink could be invalidated by a preceding notch, but our new identification strategies can handle both kinks and notches simultaneously.

Our second contribution is to propose three novel identification strategies for the

elasticity if the incentive schedule has kinks but no notches. Each of these strategies relies on weaker assumptions than those implicit in current implementations of the bunching estimator. Our first strategy identifies upper and lower bounds on the elasticity —partially identifies the elasticity —by making a mild shape restriction on the non-parametric family of heterogeneity distributions. The other two strategies point identify the elasticity using covariates and semi-parametric restrictions on the distribution of heterogeneity.

The first strategy partially identifies the elasticity by assuming a bound on the slope magnitude of the heterogeneity PDF, that is, Lipschitz continuity. Intuition for identification of the elasticity in this setting is as follows. We observe the mass of agents who bunch, which equals the area under the heterogeneity PDF inside an interval. The length of this bunching interval depends on the unknown elasticity. The maximum slope magnitude of the PDF implies upper and lower bounds for all possible PDF values inside the bunching interval that are consistent with the observed bunching mass. This translates into lower and upper bounds, respectively, on the size of the bunching interval, which corresponds to lower and upper bounds on the elasticity. These bounds allow researchers to examine the magnitude of the impossibility result in their empirical context. Depending on the data, it might take an unreasonably high slope magnitude on the heterogeneity PDF to produce bounds that include all possible elasticity values. In other settings, the difference between upper and lower bounds may be economically large even for small slope magnitudes.

The next two strategies rely on the fact that bunching can be rewritten as a censored regression model with a middle censoring point. We stress that while these strategies necessarily add structure to point identify the elasticity, they do not require fully parametric assumptions, such as normality, on the unconditional distribution of heterogeneity.

The second strategy identifies the elasticity by estimating a maximum likelihood mid-censored model, using data truncated to a window local to the kink. The likelihood function assumes that the unobserved distribution conditional on covariates is parametric, but we demonstrate that correct specification of the conditional distribution is not necessary

for consistency, as long as the unconditional distribution is correctly specified. For example, conditional normality yields a mid-censored Tobit model, which has a globally concave likelihood and is easy to implement. Nevertheless, consistency only requires that the unobserved distribution is a semi-parametric mixture of normals, and conditional normality is not necessary. Truncating the sample around the kink point improves the fit of the model and further weakens these distribution assumptions.

The third strategy restricts a quantile of the unobserved distribution, conditional on covariates, and point identification follows existing theory for censored quantile regressions (Powell, 1986; Chernozhukov and Hong, 2002; Chernozhukov, Fernández-Val, and Kowalski, 2015).

Both of the two semi-parametric methods are censored regression models that incorporate covariates. These approaches extend bunching estimators to control for observable heterogeneity for the first time. Observable individual characteristics generally account for substantial variation across agents and leave less heterogeneity unobserved. This fact suggests that identification strategies that utilize covariates should be preferred over identifying assumptions that only restrict the shape of the unobserved distribution without covariates.

Our third contribution is to illustrate the empirical relevance of our methods by revisiting Saez (2010)'s original influential application of bunching in the distribution of U.S. income caused by kinks in the EITC schedule. That approach implicitly assumes the unobserved PDF of agents that bunch is linear and uses a trapezoidal approximation to compute the bunching mass. This assumption fits poorly when the true density is non-linear or the interval of agents that bunch is large. We compare elasticity estimates based on our identification assumptions with estimates based on the trapezoidal approximation using annual samples of U.S. federal tax returns from the Internal Revenue Service (IRS).

Our partial identification method indicates that households adjust their reported income in response to marginal tax rates by a considerable amount. Placing a conservative limit on

the slope magnitude, the lower bound for the elasticity is 0.34 —that is, a one percent increase in the marginal tax rate results in a reduction in reported income of at least 0.34 percent. This estimate contrasts with the estimate of 0.43 using the trapezoidal approximation. The difference in these estimates matters. For example, Saez (2001) shows that the optimal top marginal tax rate for an economy with an elasticity of 0.34 is 13 percentage points higher than when the elasticity is 0.43.

The truncated Tobit model with covariates fits well the observed distribution of income making our semi-parametric consistency result operative. Elasticity estimates from this model differ substantially from estimates based on the trapezoidal approximation for some categories of U.S. taxpayers. For example, we estimate an an elasticity of 0.72 versus a trapezoidal estimate of 1.10 for married and self-employed individuals. This large difference highlights the sensitivity of estimates to functional form assumptions, as well as the need for methods that rely on weaker assumptions.

Our three new methods provide a suite of ways to recover elasticities from bunching behavior. Each method differs in the assumptions they make about the unobserved distribution to achieve identification. There is no way to determine which assumption is correct because the unobserved distribution is not fully identified. Nevertheless, estimates that are stable across many methods indicate that different identifying assumptions do not play a major role in the construction of those estimates. On the contrary, estimates that are sensitive to different assumptions are dependent on the validity of those assumptions. Therefore, we recommend that researchers examine the sensitivity of elasticity estimates across all available methods as a matter of routine.

Bunching estimators are widely applied in settings including fuel economy regulations (Sallee and Slemrod, 2012), electricity demand (Ito, 2014), real estate taxes (Kopczuk and Munroe, 2015), labor regulations (Garicano, Lelarge, and Van Reenan, 2016), prescription drug insurance (Einav, Finkelstein, and Schrimpf, 2017), marathon finishing times (Allen, Dechow, Pope, and Wu, 2017), attribute-based regulations (Ito and Sallee, 2018), education

(Dee, Dobbie, Jacob, and Rockoff, 2019; Caetano, Caetano, and Nielsen, 2020b), minimum wage (Jales, 2018; Cengiz, Dube, Lindner, and Zipperer, 2019), and air-pollution data manipulation (Ghanem, Shen, and Zhang, 2019), among others. Variation in the size of the mass point across groups of individuals has also been used as a first stage in a two stage approach to control for endogeneity (Chetty, Friedman, and Saez, 2013; Caetano, 2015; Grossman and Khalil, 2019).[1] An additional complication in many applications arises when the bunching mass is spread over a range instead of being a mass point. Blomquist, Kumar, Liang, and Newey (2019) provide a discussion about the potential sources for this complication and Cattaneo, Jansson, Ma, and Slemrod (2018) propose a filtering method to resolve it. Kleven (2016) reviews the many applications and branches of the bunching literature and Jales and Yu (2017) relates bunching to regression discontinuity design (RDD).

In the context of kinks, Blomquist and Newey (2017) were the first to prove the impossibility of point identification and the possibility of partial identification —and earlier Blomquist, Kumar, Liang, and Newey (2015) provide the outline for those proofs. We derive partial identification bounds by assuming the PDF has a bounded slope, whereas Blomquist and Newey (2017) assume the PDF of heterogeneity is monotone. We developed our impossibility and partial identification results independently of theirs. Our partial identification approach has three valuable features: closed-form solutions, observed bunching always implies a positive elasticity, and nesting of the original bunching estimator. Blomquist and Newey (2018) explain that a notch can identify the elasticity and a formal proof of identification appears contemporaneously in an earlier version of our paper (Bertanha, McCallum, and Seegert, 2018). To the best of our knowledge, ours is the first paper to demonstrate point identification using censored regression models, covariates, and semi-parametric assumptions on the distribution of heterogeneity. More generally, the

[1]Econometric approaches using bunching for causal identification include Khalil and Yildiz (2017), Caetano and Maheshri (2018), Caetano, Kinsler, and Teng (2019), and Caetano, Caetano, and Nielsen (2020a).

theory demonstrating that a kink fails to point identify the elasticity relates to the literature on impossible inference reviewed by Bertanha and Moreira (2020).

The paper proceeds with an utility maximization model subject to a piecewise-linear budget constraint in Section 2. Section 3 investigates the identification of the elasticity in the case of kinks and notches. We propose the three identification strategies for the elasticity in Section 4 and illustrate these methods empirically in the context of the EITC in Section 5. Section 6 concludes. Appendix A contains all proofs, and supplemental Appendix B collects auxiliary results and examples. Finally, we developed the Stata command `bunching` that implements our procedures.[2]

## 2 Utility Maximization Subject to Piecewise-Linear Constraints

Firms' and individuals' optimization problems often face piecewise-linear constraints. The nature of constraints is dictated by differential tax rates, insurance reimbursement rates, or contract bonuses. A budget set is fully characterized by a sequence of intercepts and slopes that change at known points. A change in the intercept is referred to as a notch, and a change in the slope is referred to as a kink.

### 2.1 Model Setup

We start with the labor supply characterization employed by the vast majority of the literature, which follows the seminal work of Saez (2010) and Kleven and Waseem (2013). Agents maximize an iso-elastic quasi-linear utility function and choose consumption and labor subject to a piecewise-linear budget set. Well-known models that fit into this category include those of Burtless and Hausman (1978), Best and Kleven (2018), Einav, Finkelstein, and Schrimpf (2017), among others. For ease of exposition, we focus on budget sets with one

[2]The Stata package is available at the Statistical Software Components (SSC) online repository. Type `ssc install bunching` in Stata to install the package. The package is also available for download from the website of the authors.

kink or one notch in the main text. In supplemental Appendices B.1 and B.2, we generalize the literature to any combination of kinks and notches. Section 3 below briefly discusses new insights for the identification of the elasticity that arise in the problem with multiple kinks and notches.

Consider a population of agents that are heterogeneous with respect to a scalar variable $N^*$, referred to as ability. Ability is distributed according to a continuous probability density function (PDF) $f_{N^*}$, with support $(0, \infty)$, and a cumulative distribution function (CDF) $F_{N^*}$. Agents know their $N^*$, but the econometrician does not observe the distribution of $N^*$.

Agents maximize utility by jointly choosing a composite consumption good $C$ and labor supply $L$. Utility is increasing in $C$ and decreasing in $L$. These variables are constrained by a budget set, where the agent may consume all of its labor income net of taxes plus an exogenous endowment $I_0$. For simplicity, we assume the price of labor and consumption are equal to one, such that taxable labor income $Y$ is equal to $L$.

In the budget constraint with a kink, the tax rate increases from $t_0$ to $t_1$ as income increases above the kink value $K$. The budget constraint has a notch when the agent is charged a lump-sum tax of $\Delta > 0$ as income crosses $K$. Agent type $N^*$ maximizes utility $U(C, Y; N^*)$ as follows,

$$\max_{C,Y} \quad C - \frac{N^*}{1 + 1/\varepsilon} \left( \frac{Y}{N^*} \right)^{1+\frac{1}{\varepsilon}} \tag{1}$$

$$s.t.$$

$$C = \mathbb{I}\{Y \leq K\}[I_0 + (1 - t_0)Y] + \mathbb{I}\{Y > K\} \left[ I_1 + (1 - t_1)(Y - K) \right], \tag{2}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function; the budget line has intercept $I_0$ and slope $1 - t_0$ if $Y \leq K$, but intercept $I_1 = I_0 + K(1 - t_0) - \Delta$ with slope $1 - t_1$ if $Y > K$; and $\varepsilon$ is the elasticity of income $Y$ with respect to one minus the tax rate when the solution is interior. In the case of a kink, $\Delta = 0$, and the budget frontier is continuous; otherwise, in the case of a notch, it has a jump discontinuity of size $\Delta$ at $Y = K$. The solution is always on the

budget frontier in Equation 2.

### 2.2 Model Solution

The solution for $Y$ in Problem 1 is well known in the literature, when $K$ is a kink (Saez, 2010) and when $K$ is a notch (Kleven and Waseem, 2013):

$$Y = \begin{cases} N^*(1-t_0)^{\varepsilon} & \text{, if } \quad 0 < N^* < \underline{N} \\ K & \text{, if } \quad \underline{N} \leq N^* \leq \overline{N} \\ N^*(1-t_1)^{\varepsilon} & \text{, if } \quad \overline{N} < N^*, \end{cases} \tag{3}$$

where the expressions for the thresholds $\underline{N}$ and $\overline{N}$ are given below.

In the case of a kink, $\underline{N} = K(1-t_0)^{-\varepsilon}$, and $\overline{N} = K(1-t_1)^{-\varepsilon}$. The budget frontier is continuous, but its slope suddenly decreases at $Y = K$. For values of $N^*$ inside the bunching interval $[\underline{N}, \overline{N}]$, the agent's indifference curve is never tangent to the budget frontier, and we have the non-interior solution $Y = K$. For values of $N^*$ outside of the bunching interval, the indifference curve is always tangent to some point on the budget frontier.

In the case of a notch, the solution is interior for $N^* < \underline{N} = K(1-t_0)^{-\varepsilon}$, but there are no tangent indifference curves for $N^* \in [K(1-t_0)^{-\varepsilon}, K(1-t_1)^{-\varepsilon}]$, just as in the case of a kink. Although tangency occurs for $N^* > K(1-t_1)^{-\varepsilon}$, some of the resulting utility levels are lower than the utility at the notch point. The budget frontier with a jump-down discontinuity at $Y = K$ has an interval of income values $(K, Y^I]$ that no agent ever chooses. The value $Y^I > K$ corresponds to the interior solution of the agent with $N^* = N^I$; that is, the smallest $N^*$ such that the agent's utility is equal to the utility of the agent choosing $Y = K$. Thus $\overline{N} = N^I$, and the solution is at $Y = K$ for $N^* \in [\underline{N}, \overline{N}]$. As the ability $N^*$ increases above $N^I$, the utility gets larger than the utility at $K$, and again there is an interior solution. Supplemental Appendix B.2 has a formal definition of $N^I$ in Equation B.3.

To make the solution more tractable, we take the natural logarithm of all variables.

Define $y = \log(Y)$, $n^* = \log(N^*)$, $k = \log(K)$, $s_0 = \log(1 - t_0)$, and $s_1 = \log(1 - t_1)$.

$$y = \begin{cases} n^* + \varepsilon s_0 & \text{, if } \quad n^* < \underline{n} \\ k & \text{, if } \quad \underline{n} \leq n^* \leq \overline{n} \\ n^* + \varepsilon s_1 & \text{, if } \quad \overline{n} < n^*. \end{cases} \tag{4}$$

As ability $n^*$ increases, the optimal choice of $y$ increases, except when $n^*$ falls inside the bunching interval $[\underline{n}, \overline{n}]$, in which $y$ remains constant and equal to $k$.

### 2.3 Bunching and the Counterfactual Distribution of Income

The solution in the previous section expresses income as a function of the model parameters and $n^*$. For given values of $(t_0, t_1, k, \varepsilon)$, the continuously distributed $n^*$ maps into a mixed continuous-discrete distribution for $y$. The model predicts bunching in the distribution of $y$ at a kink or notch point (i.e. $\mathbb{P}(y = k) > 0$), but a continuous distribution of $y$ otherwise. The amount of bunching depends on the elasticity $\varepsilon$ and the unobserved distribution $n^*$,

$$B \equiv \mathbb{P}\left(y = k\right) = \mathbb{P}\left(\underline{n} \leq n^* \leq \overline{n}\right) = \int_{\underline{n}}^{\overline{n}} f_{n^*}\left(u\right) \, du = F_{n^*}\left(\overline{n}\right) - F_{n^*}\left(\underline{n}\right), \tag{5}$$

where the length of the interval $[\underline{n}, \overline{n}]$ varies with $\varepsilon$.

The literature typically defines $B$ in terms of the counterfactual distribution of income in the scenario without any kinks or notches. Let counterfactual income be $y_0$ in such case. The solution to Problem 1 is simply $y_0 = n^* + \varepsilon s_0$ for every value of $n^*$. The variable $y_0$ has continuous PDF $f_{y_0}$ and CDF $F_{y_0}$. The bunching mass is derived as

$$B = \int_{k}^{k+\Delta y} f_{y_0}\left(u\right) \, du = F_{y_0}\left(k + \Delta y\right) - F_{y_0}\left(k\right), \tag{6}$$

where $\Delta y = \varepsilon(s_0 - s_1)$. Figure 1, Panels a and b, illustrates the distributions of $y$ and $y_0$, and how they relate to each other, to $B$, and to $f_{n^*}$.

Saez (2010)'s insight is that the mass of agents bunching $B$ is increasing in the elasticity $\varepsilon$ for a given distribution of $y_0$. Stated another way, the more agents shift income to the kink-point $k$, the more sensitive they are to changes in tax rates. All current bunching and notching estimators use this insight to identify the elasticity. First, the researcher obtains an estimate of the counterfactual distribution of $y_0$ and the bunching mass $B$. Plugging these into Equation 6 allows us to solve for an estimate of the elasticity.

The treatment of the problem thus far abstracts from the existence of optimization and friction errors in the solution of Problem 1. In reality, instead of $y$, researchers typically observe the distribution of $\widetilde{y} = y + e$, where $e$ is a random variable accounting for optimization frictions. In this case, the distribution of $\widetilde{y}$ has the bunching mass distributed over a range around the kink point, as opposed to being right at the kink.

In a recent survey article, Kleven (2016) summarizes an identification strategy commonly used in the literature to estimate the distribution of $y_0$. The ''polynomial strategy'' was first proposed by Chetty et al. (2011) (Equations 14-15 and Figures 3-4), and it consists of fitting a flexible polynomial to an estimate of the PDF of $y$. The polynomial regression excludes observations that lie in a range around the kink point. The researcher chooses the range based on the support of the distribution of friction errors. The polynomial fit is then extrapolated to this excluded region as a way of predicting $f_{y_0}$. The procedure is widely used in the bunching literature; see, for example, Figure 6 by Bastani and Selin (2014) , Figure 1 by Devereux, Liu, and Loretz (2014), and Figure 4 by Best and Kleven (2018). In supplemental Appendix B.3, we give more details in the context of a simple example, where $n^*$ is uniformly distributed. The example shows that such an identification strategy fails to recover both $B$ and $f_{y_0}$, even when the proposed polynomial fit is perfect.

The strategy fails for two reasons. First, the distribution $y$ is observed with error, and a proper deconvolution method must be used to retrieve the distribution of $y$, given the distribution of $\widetilde{y}$. Second, even when the distribution of $y$ is known, it is not possible to obtain the distribution of $y_0$ inside the integration domain of Equation 6. Although $y = y_0$

when $n^* < \underline{n}$, we have that $y = k$, while $y_0 = n^* + \varepsilon s_0$ when $n^* \in [\underline{n}, \overline{n}]$ (Figures 1a and 1b). The shape of the distribution of $y_0$ is unidentified when $n^*$ falls in the bunching interval.

The rest of this paper focuses on the second problem of the identification strategy, namely the problem of identifying the elasticity, $\varepsilon$, using the distribution of $y$ instead of the distribution of $\widetilde{y} = y + e$. In fact, our methods apply to the many examples of bunching that do not have friction errors, for example, Figure 4 by Glogowsky (2018) and Figure 1 by Goncalves and Mello (2018). The study of identification in the presence of optimization frictions is deferred to future research. In work in progress, Cattaneo et al. (2018) study identification of the distribution of $y$ given the distribution of $\widetilde{y}$ plus minimal assumptions on the distribution of $e$.

## 3 Identification

The general solution to Problem 1 with multiple kinks and notches in supplemental Appendix B.2 brings new insights to the identification of the elasticity, when compared to the particular solution in the case of one kink or notch. First, in a budget set with multiple kinks but no notches, the general solution is simply a combination of solutions local to each kink. The bunching intervals of consecutive kinks do not overlap (Equation B.4). As a result, inference methods for the elasticity that are valid in the case of one kink may still be used locally to each kink.

Second, a notch at $k$ creates an empty interval in the support of the distribution of $y$ right after $k$. Such an empty interval may or may not contain the next tax change point $k' > k$, depending on the value of $\varepsilon$. For example, eligibility for Medicaid benefits in the United States creates a sizeable notch that may overshadow the next tax change in the budget set of some individuals. In this case, inference methods that focus on kinks without accounting for other notches may produce misleading conclusions about the elasticity.

The rest of this section investigates identification with one notch or one kink. We show that identification is possible with one notch without any restriction on the distribution of

$n^*$. On the other hand, the identification in case of a kink is impossible, unless the researcher imposes restrictions on the distribution of $n^*$.

### 3.1 Identification with at Least One Notch

In the problem without optimization error, the existence of one notch produces additional identifying information in the observed distribution of income. However, the identification strategy is different from previous studies, which solely rely on Equation 6.

Even when $N^*$ has full support $(0, \infty)$, there exists an empty interval in the distribution of $Y$, to right of the notch.[3] Following the solution in Equation 3, the empty interval is $(K, Y^I]$, where $Y^I = N^I(1-t_1)^\varepsilon$, and $N^I$ is defined above. Once $Y^I$ is identified from the support of the distribution of $Y$, we numerically solve for $\varepsilon$ that satisfies the indifference condition Equation 7 below.

**Theorem 1.** *Suppose the support of $N^*$ is equal to $(0, \infty)$, that $K$ is a notch, and that the upper limit of the empty interval in the support of $Y$ to the right of $K$ is equal to $Y^I$. Then the indifference condition that defines $Y^I$ is equivalent to*

$$Y^I + \varepsilon K \left(\frac{K}{Y^I}\right)^{\frac{1}{\varepsilon}} = (1+\varepsilon)\left(\frac{C + I_1 + K(1-t_1)}{1-t_1}\right), \tag{7}$$

*where $C$ is the consumption value on the budget frontier at the notch point. Moreover, there exists an unique $\varepsilon$ that solves Equation 7 as a function of $Y^I$, $K$, $C$, $I_1$, $t_1$. Therefore the elasticity is identified.*

A proof for this theorem is in Appendix A.1 and all our other proofs are in Appendix A.

### 3.2 Lack of Identification With One Kink

Although bunching is increasing in the elasticity for a fixed distribution of $y_0$ or $n^*$, it is also true that, for a fixed elasticity, bunching increases as $f_{n^*}$ becomes more concentrated

[3] In this subsection, it is analytically simpler to work with the solution in levels rather than in logs.

between $\underline{n}$ and $\overline{n}$. If all we know about $f_{n^*}$ is that it is continuous with full support and that its integral over $[\underline{n}, \overline{n}]$ equals $B$, then there is no way to identify both the elasticity and $f_{n^*}$ using only Equation 5; equivalently, there is no way to identify both the elasticity and the distribution of $y_0$ using only Equation 6. Intuitively, identification using only (5) or (6) is impossible because each uses one equation to solve for two unknowns. This is shown by Blomquist et al. (2015) and Blomquist and Newey (2017). We present the impossibility result in this section as a building block to our novel identification strategies in the next sections.

Formally, the data and model comprise five objects: 1) the CDF of earnings $F_y$, 2) the kink point $k$, 3) the slopes of the piecewise-linear constraint $s_0$ and $s_1$; 4) the CDF of the latent variable $F_{n^*}$, and 5) the elasticity $\varepsilon$. Equation 4 is a mapping $T$ that takes objects (2)–(5) and maps them into the CDF of optimal incomes across agents:
$F_y = T\left(k, s_0, s_1, F_{n^*}, \varepsilon\right)$.

The researcher observes objects (1)–(3), but does not observe the last two, $F_{n^*}$ and $\varepsilon$. The problem of identification consists of inverting the mapping $T$ such that the unobserved $\varepsilon$ is a function that only depends on the first three objects $(F_y, k, s_0, s_1)$, regardless of what $F_{n^*}$ may be. We denote the class of admissible distributions of $n^*$ as $\mathcal{F}_{n^*}$. If the class $\mathcal{F}_{n^*}$ contains all possible continuous distributions of $n^*$, then identification of $\varepsilon$ is impossible.

**Lemma 1.** *Let $\mathcal{F}_{n^*}$ be the class of all CDFs $F_{n^*}$ that have continuous PDFs $f_{n^*}$ with support $(-\infty, \infty)$. Let $\mathcal{F}_y$ be the class of all CDFs $F_y$ that are mixed continuous-discrete with one mass point at $k$, and continuous PDF $f_y$ otherwise. Take $F_y$, $k$, $s_0$, and $s_1$ as givens. For every elasticity $\varepsilon \in (0, \infty)$, there exists $F_{n^*,\varepsilon} \in \mathcal{F}_{n^*}$ such that $F_y = T\left(k, s_0, s_1, F_{n^*,\varepsilon}, \varepsilon\right)$. Therefore it is impossible to point-identify $\varepsilon$.*

Figure 1 provides intuition for the proof of Lemma 1. It illustrates that the observable PDF $f_y$ in Figure 1a is generated by applying Equation 4 to two different combinations of latent variable distributions and elasticities, $f_{n^*,\varepsilon}$ and $f_{n^*,\varepsilon'}$ in Figures 1c and 1d, respectively. Lemma 1 clarifies that current bunching methods are either implicitly restricting $\mathcal{F}_{n^*}$ or simply inconsistent for the true elasticity.

A direct consequence of Lemma 1 is that it is impossible to test restrictions on $\mathcal{F}_{n^*}$. Below we consider a couple of examples of identifying restrictions from the literature.

**Example 1.** *Saez (2010) implicitly restricts $\mathcal{F}_{n^*}$ when using a trapezoidal approximation to solve the integral in Equation 6, in levels rather than in logs (Saez's Equation 4 on page 186). That is,*

$$B = \int_K^{K+\Delta Y} f_{Y_0}(u)\ du \cong \left(\frac{f_{Y_0}(K+\Delta Y) + f_{Y_0}(K)}{2}\right)\Delta Y, \tag{8}$$

*where $\Delta Y = K\left[((1-t_0)/(1-t_1))^{\varepsilon} - 1\right]$. A sufficient condition for the approximation to be true is to assume $f_{Y_0}(u)$ is an affine function of $u$ for values of $u \in [K, K+\Delta Y]$. Given that $Y_0 = N^*(1-t_0)^{\varepsilon}$, the PDF $f_{N^*}(u) = f_{Y_0}(u(1-t_0)^{\varepsilon})(1-t_0)^{\varepsilon}$ is restricted to be an affine function of $u$ inside the interval $[K(1-t_0)^{-\varepsilon}, K(1-t_1)^{-\varepsilon}]$. This is equivalent to restricting $f_{n^*}$ to have an exponential shape within $[k-\varepsilon s_0, k-\varepsilon s_1]$.*

*The rest of Saez's identification strategy uses the fact that $f_{Y_0}(K) = f_Y(K^-)$, and $f_{Y_0}(K+\Delta Y) = f_Y(K^+)((1-t_1)/(1-t_0))^{\varepsilon}$, where $f_Y$ is the PDF of the continuous portion of the distribution of $Y$, and $f_Y(K^{\pm})$ denotes side limits $\lim_{Y\to K^{\pm}} f_Y(K^{\pm})$. Substituting these into Equation 8,*

$$B \cong \frac{1}{2}\left(f_Y(K^+)\left(\frac{1-t_1}{1-t_0}\right)^{\varepsilon} + f_Y(K^-)\right)K\left[\left(\frac{1-t_0}{1-t_1}\right)^{\varepsilon} - 1\right], \tag{9}$$

*which is Equation 5 by Saez (2010). It is then possible to solve implicitly for $\varepsilon$ as a function of the side limits of $f_Y$, the tax rates, the kink point, and the bunching mass.*

One may argue that the affine assumption is a good approximation to any potentially non-linear density $f_{N^*}$, if the bunching interval $[K(1-t_0)^{-\varepsilon}, K(1-t_1)^{-\varepsilon}]$ is small. The problem with this argument is that the size of the interval is itself a function of the elasticity. It is impossible to state that the interval is small and the linear approximation is a good one without a priori knowledge of the elasticity.

**Example 2.** *The derivation by Chetty et al. (2011) of Equation 6 on page 761 assumes that the PDF $f_{Y_0}$ is constant inside the bunching interval $[K, K + \Delta Y]$. This is equivalent to assuming that $N^*$ is uniformly distributed in that region and thus restricts the class $\mathcal{F}_{n^*}$. For some scalar $a$, assume $F_{Y_0}(u) = a + f_{Y_0}(K)u$ for $u \in [K, K + \Delta Y]$, so that the PDF of $Y_0$ is constant and equal to $f_{Y_0}(K)$ in the bunching interval. Then,*

$$\begin{aligned} B =& \int_K^{K+\Delta Y} f_{Y_0}(u)\, du = F_{Y_0}(K + \Delta Y) - F_{Y_0}(K) \\ =& f_{Y_0}(K)\Delta Y = f_{Y_0}(K)K\left[\left(\frac{1-t_0}{1-t_1}\right)^{\varepsilon} - 1\right] \\ \cong& f_{Y_0}(K)K\varepsilon \ln\left(\frac{1-t_0}{1-t_1}\right) \\ \varepsilon \cong& \frac{B/f_{Y_0}(K)}{K \ln\left(\frac{1-t_0}{1-t_1}\right)}, \end{aligned} \tag{10}$$

*where the second to last approximate equality uses $[(1-t_0)/(1-t_1)]^{\varepsilon} - 1$ $\cong \ln[(1-t_0)/(1-t_1)]^{\varepsilon}$ for small tax changes; and the last approximate equality is Equation 6 by Chetty et al. (2011). The rest of their identification procedure relies on the polynomial strategy to obtain $B$ and $f_{Y_0}(K)$, as described in the supplemental Appendix B.3.*

*The constant PDF assumption on $f_{Y_0}$ is more restrictive than the affine PDF assumption that justifies Saez's trapezoidal approximation. The trapezoidal approximation allows for $f_{Y_0}$ to have a non-zero slope in the bunching interval, whereas the constant PDF assumption does not.*

There are more flexible restrictions one could impose on $\mathcal{F}_{n^*}$. For example, one could say $n^*$ follows a distribution inside a parametric family of distributions.

**Example 3.** *In general, let $\mathcal{F}_{n^*} = \{G_{n^*}(n;\theta)\ ,\ \theta \in \Theta\}$, where $G_{n^*}$ are CDFs indexed by a $p \times 1$ vector of parameters $\theta$ in a parameter space $\Theta$. Identification of the elasticity requires that the bunching mass and the shape of the distribution of $y$ around the kink point are sufficient to identify $\theta$ and $\varepsilon$. That is, the family of distributions $\mathcal{F}_{n^*}$ is such that, for any feasible choice of $(k, s_0, s_1, \varepsilon, \theta)$, there exists an unique solution $(\bar{\varepsilon}, \bar{\theta}) = (\varepsilon, \theta)$ to the following*

*system of equations:*

$$G_{n^*}\left(k-\varepsilon s_1;\theta\right)-G_{n^*}\left(k-\varepsilon s_0;\theta\right)=G_{n^*}\left(k-\bar{\varepsilon}s_1;\bar{\theta}\right)-G_{n^*}\left(k-\bar{\varepsilon}s_0;\bar{\theta}\right) \tag{11}$$

$$G_{n^*}(u-\varepsilon s_0;\theta)=G_{n^*}(u-\bar{\varepsilon}s_0;\bar{\theta}) \quad \textit{for } \forall u<k \tag{12}$$

$$G_{n^*}(u-\varepsilon s_1;\theta)=G_{n^*}(u-\bar{\varepsilon}s_1;\bar{\theta}) \quad \textit{for } \forall u>k. \tag{13}$$

*For example, the family of normal distributions with unknown mean and variance satisfies these conditions (see supplemental Appendix B.4). Identification is also possible in families with more than just two parameters. The objects on the left-hand side (LHS) of the three equations above, evaluated at the true* $(\varepsilon,\theta)$*, are identified from the data. Thus, if* $\mathcal{F}_{n^*}$ *satisfies* (11)-(13)*, then the elasticity and* $F_{n^*}$ *are identified.*

## 4 Solutions

The rest of the paper focuses on methods that identify the elasticity in the kink case. We present three types of identification assumptions on the distribution of ability, from less restrictive to more restrictive. We start with a non-parametric shape restriction that bounds the slope magnitude of $f_{n^*}$, which leads to partial identification of $\varepsilon$. Next, we connect bunching to the literature on censored regressions, where $n^*$ is the regression error. It becomes natural to use covariates to explain $n^*$, and we propose two types of semi-parametric restrictions on the distribution of $n^*$ that point-identify the elasticity. The first restricts the distribution of $n^*$, conditional on covariates; and the second restricts a quantile of the distribution of $n^*$, conditional on covariates. In general, more data variation and structure are needed to provide any information about the elasticity.

### 4.1 Non-parametric Bounds

Our partial identification approach relies on restricting the class $\mathcal{F}_{n^*}$ to PDFs, $f_{n^*}$, that are Lipschitz continuous with constant $M\in(0,\infty)$. In other words, the slope magnitude of

any $f_{n^*} \in \mathcal{F}_{n^*}$ is bounded by $M$. The following theorem gives the partially identified set for $\varepsilon$ as a function of identified quantities and the maximum slope magnitude $M$.

**Theorem 2.** *Assume $\mathcal{F}_{n^*}$ contains all distributions with PDF $f_{n^*}$ that are Lipschitz continuous with constant $M \in (0, \infty)$. Then the elasticity $\varepsilon \in \Upsilon$, where*

$$\Upsilon = \begin{cases} \emptyset & , \text{ if } B < \frac{\left|f_y(k^+) - f_y(k^-)\right| \left[f_y(k^+) + f_y(k^-)\right]}{2M} \\ [\underline{\varepsilon}, \overline{\varepsilon}] & , \text{ if } \frac{\left|f_y(k^+) - f_y(k^-)\right| \left[f_y(k^+) + f_y(k^-)\right]}{2M} \leq B < \frac{f_y(k^+)^2 + f_y(k^-)^2}{2M} \\ [\underline{\varepsilon}, \infty) & , \text{ if } \frac{f_y(k^+)^2 + f_y(k^-)^2}{2M} \leq B \end{cases},$$

*where $\emptyset$ is the empty set, and*

$$\underline{\varepsilon} = \frac{2\left[f_y(k^+)^2/2 + f_y(k^-)^2/2 + M\ B\right]^{1/2} - \left(f_y(k^+) + f_y(k^-)\right)}{M(s_0 - s_1)}$$
$$\overline{\varepsilon} = \frac{-2\left[f_y(k^+)^2/2 + f_y(k^-)^2/2 - M\ B\right]^{1/2} + \left(f_y(k^+) + f_y(k^-)\right)}{M(s_0 - s_1)}.$$

Figures 1c and 1d provide the intuition behind the derivation of the bounds in $\Upsilon$. For a fixed value of $\varepsilon$, the length of the interval $[\underline{n}, \overline{n}]$ is fixed. If the magnitude of the derivative of $f_{n^*}$ is bounded by $M$, we obtain maximum and minimum areas under $f_{n^*}$ over $[\underline{n}, \overline{n}]$. We repeat this exercise for every value of $\varepsilon$ to get a range of possible areas associated with each $\varepsilon$. Given the probability of bunching $B$ is the area under the true $f_{n^*}$ over $[\underline{n}, \overline{n}]$, the partially identified set has all values of $\varepsilon$ whose range of possible areas contains $B$. The partially identified set is empty if $M$ is not big enough to allow for the existence of a continuous function $f_{n^*}$ which connects $f_y(k^-) = f_{n^*}(k - \varepsilon s_0)$ to $f_y(k^+) = f_{n^*}(k - \varepsilon s_1)$. The partially identified set is unbounded if $M$ is large enough to allow $f_{n^*}$ to be zero inside the interval $[\underline{n}, \overline{n}]$.

The expression for the partially identified set depends on the value of $M$ and the researcher must specify this value to compute the bounds. The uniform approximation in Example 2 says that $f_{n^*}$ has zero slope inside the bunching interval, that is, $M = 0$. The trapezoidal approximation in Example 1 implicitly chooses $M = m_0$ such that $m_0$ is the

smallest value of $M$ for which we have bounds that are well defined. Formally, $m_0$ solves $B = |f_y(k^+) - f_y(k^-)| \ [f_y(k^+) + f_y(k^-)] / 2m_0$, which makes $\underline{\varepsilon} = \overline{\varepsilon}$ and point-identifies $\varepsilon$. Thus the exercise of computing bounds necessarily involves assumptions weaker than the uniform and trapezoidal approximations.

Lemma 1 makes clear that it is impossible to identify, and thus estimate, the value of $M$. A useful starting point for the magnitude of $M$ comes from the maximum slope magnitude of the continuous part of $f_y$, say $m_1$. The PDF $f_y$ is identified and is the shifted PDF of $n^*$. Thus, the maximum slope of $f_{n^*}$ outside of the bunching interval is identified and equal to $m_1$. If we assume that that the slope of $f_{n^*}$ inside the bunching interval is never bigger than outside, then $M = m_1$.

As a rule of thumb, we recommend researchers to plot the bounds in Theorem 2 as a function of $M$ for a range of values that includes $m_0$, $m_1$, and possibly bigger values, e.g., up to $2m_1$. Theorem 2 is important to quantify the magnitude of the impossibility problem presented in Lemma 1. If the bounds plotted for a range of $M$ values admit elasticities that are too different in economic terms, then the identifying assumptions play a critical role in determining the elasticity. We give full details and implement this sensitivity analysis in the empirical section using our `bunching` Stata package (Section 5).[4]

While we assume the PDF has bounded slope, Blomquist and Newey (2017) partially identify the elasticity by assuming the PDF of heterogeneity is monotone. Our approach has three valuable properties. The first is that the bounds of our partially-identified set have closed form solutions. Second, an observed mass point implies a positive elasticity even for large values of the slope $M$, which is in line with the theoretical prediction that agents respond to a change in incentives. Third, it nests and is easily comparable to the original bunching estimator based on the trapezoidal approximation.

[4]It is important to clarify that the problem of choosing $M$ is different than the typical problem of choosing a tuning parameter, e.g., a bandwidth or polynomial order in non-parametric estimation. The value of $M$ represents a choice of functional form assumption, while in non-parametric estimation, you typically choose the tuning parameter to achieve desirable properties of the estimator for a given functional form assumption.

We end this subsection with the case of a budget set with several kinks $k_j$, $j = 1, \ldots, J$, but no notches. One may ask whether the existence of several kinks helps identify the elasticity. As noted above, the bunching intervals do not overlap across kinks, that is, $\overline{N}_j = K_j(1 - t_{j-1})^{-\varepsilon} < K_j(1 - t_j)^{-\varepsilon} = \underline{N}_j$. Lemma 1 applies to each kink, and multiple kinks do not necessarily point-identify $\varepsilon$, because the distribution of $n^*$ may be very different across different bunching intervals.

Multiple kinks do help with the identification of $\varepsilon$, as long as the researcher restricts the slope of $f_{n^*}$ and believes the model in Equation 1 applies to all individuals. This arises from the fact that every individual is assumed to have the same elasticity parameter $\varepsilon$, and that the bounds of Theorem 2 vary in length as $B_j$, $f_y(k_j^{\pm})$, $s_j$ vary across cutoffs $j = 1, \ldots, J$. The partially identified set is narrowed down by the intersection of bounds specific to each one of the multiple kinks.

**Corollary 1.** *Assume the conditions of Theorem 2 for each kink $k_j$, $j = 1, \ldots, J$. Then the elasticity $\varepsilon \in \bigcap_{j=1}^{J} \Upsilon_j$, where $\Upsilon_j$ is the partially identified set of Theorem 2 applied to kink $k_j$.*

### 4.2 Semi-parametric Identification with Covariates

Identification with kinks is impossible when the distribution of ability $n^*$ belongs to the non-parametric class of all continuous distributions. Parametric functional form assumptions identify the elasticity, but identification relies on fitting such functional form to non-bunching individuals and extrapolating the functional form to bunching individuals.

This section considers alternative identification assumptions that rely on the existence of additional covariates in the dataset. There is strong empirical evidence suggesting that ability is well explained by individual characteristics, such as age, demographics, filing status, etc. For example, the ability distribution of young workers may have a very different mean and variance, compared to that of older workers. Extrapolations based on covariates that predict $n^*$ are much more reasonable than extrapolations solely based on the shape of the

PDF of $n^*$. The key assumption is that covariates that help explain the distribution of $n^*$ for non-bunching individuals also help explain the distribution of $n^*$ for bunching individuals.

We start by connecting bunching to censored regression models. This allows us to relate to the vast econometrics literature in this area. Consider again the data generating process given by Equation 4. The model for $y$ is a mid-censored model, where the error term is $n^*$, the intercept to the left of the kink is $\varepsilon s_0$, the intercept to the right of the kink is $\varepsilon s_1$, and the censoring point is $k$. The main difference between (4) and a typical censored regression model is that the latter has the censoring point at either the minimum or maximum of the distribution of $y$ (see Equation 15 in the next subsection). Identification, estimation, and inference in these models have been widely studied in econometrics since Tobin (1958).

There are many advantages of framing the estimation of $\varepsilon$ as estimation of a censored model. Surveys of censoring models and their applications are provided by Maddala (1983), Amemiya (1984), Dhrymes (1986), Long (1997), DeMaris (2005), and Greene (2005). There are straightforward extensions that account for optimizing frictions. Moreover, censored models are easily estimated with a number of different techniques that are available in many computer packages. Most importantly, it becomes extremely practical to add covariates as explanatory factors for the distribution of $n^*$.

Assume the researcher has access to a vector of covariates $X \in \mathbb{R}^{1\times(d+1)}$, where $X$ contains an intercept variable and the distribution of $X$ is unrestricted. We build on censoring models with covariates to identify the elasticity by imposing two types of semi-parametric assumptions on the distribution of $n^*$.

The first type of assumption states that the distribution of $n^*$ is a mixture of normal distributions averaged over the distribution of covariates. This assumption does not imply conditional normality of $n^*$ given $X$ but it is implied by conditional normality of $n^*$. Although the Tobit likelihood assumes normality of the unobserved distribution conditional on covariates, we demonstrate that the Tobit estimator remains consistent under the semi-parametric class of normal mixtures. In addition, the researcher may estimate a

truncated Tobit model on data in a small neighborhood of the kink point, which requires even weaker distribution assumptions for consistency. This robustness property remains true if we replace the normal distribution by another parametric distribution to form the semi-parametric mixture. For example, the maximum likelihood estimator for the elasticity that assumes that $n^*$ conditional on $X$ is exponential remains consistent when the unconditional distribution of $n^*$ is a mixture of exponentials averaged over $X$, whether or not the distribution of $n^*$ conditional on $X$ is exponential. In the rest of this section, we keep this first type of assumption in terms of normals for ease of exposition and practical reasons: the Tobit likelihood is globally concave and software to estimate Tobit models is ubiquitous.

The second type of assumption imposes a parametric functional form on a quantile of the conditional distribution of $n^*$, given $X$. Sufficient variation in covariates yields point-identification of the elasticity, which is consistently estimated by mid-censored quantile regressions.

### 4.2.1 Tobit Regression

The first type of semi-parametric assumption is formally stated in Lemma 2 below. In the meantime, we construct the Tobit estimator by assuming that there exists unique $(\beta, \sigma) \in \mathbb{R}^{1\times(d+1)} \times \mathbb{R}_+$, such that

$$F_{n^*|X}(n, x) = \Phi\left(\frac{n - x\beta}{\sigma}\right), \tag{14}$$

where $F_{n^*|X}$ denotes the CDF of $n^*$ conditional on $X$, and $\Phi(\cdot)$ is the CDF of a standard normal distribution. Assumption 14 does not restrict the distribution of $X$; thus the unconditional CDF of $n^*$ lives in a semi-parametric class and needs not to be normal. The more variation in covariates one has, the richer is this class of distributions.

The elasticity parameter $\varepsilon$ is consistently estimated using a mid-censored Tobit regression. Define the error term $U = n^* - X\beta$, the latent variables $y_0^* = \varepsilon s_0 + X\beta + U$, and

$y_1^* = \varepsilon s_1 + X\beta + U$, where $y_1^* < y_0^*$, since $\varepsilon > 0$ and $s_0 > s_1$. Then $y$ follows a mid-censored Tobit model

$$y = \begin{cases} y_0^* & \text{, if } \quad y_1^* < y_0^* < k \\ k & \text{, if } \quad y_1^* \leq k \leq y_0^* \\ y_1^* & \text{, if } \quad k < y_1^* < y_0^* \end{cases} = \min\{y_0^*;\ \max\{k;\ y_1^*\}\}. \tag{15}$$

This is different from the classic Tobit model, where the censoring point is either at the minimum or at the maximum of the distribution of $y$. A possible estimation strategy is to adapt the two-step Heckit estimator to our setting (Heckman, 1976, 1979). In the first step, estimate a binary outcome for bunching and not bunching individuals including covariates. In the second step, regress income of not bunching individuals on covariates and the equivalents of the inverse Mills ratio. Another extremely practical way of estimating this mid-censored Tobit model is to estimate two classic Tobit models. To see that, construct the variables $y_0 = \min\{y, k\}$ and $y_1 = \max\{k, y\}$. It turns out that $y_0$ follows a right-censored Tobit with intercept $\varepsilon s_0 + \beta_0$, slope coefficients $\beta_1, \ldots, \beta_d$, where $\beta = (\beta_0, \beta_1, \ldots, \beta_d)$. Similarly, $y_1$ follows a left-censored Tobit with intercept $\varepsilon s_1 + \beta_0$, and slope coefficients $\beta_1, \ldots, \beta_d$. Thus, the elasticity is consistently estimated by the difference of both intercepts $(\varepsilon s_1 + \beta_0) - (\varepsilon s_0 + \beta_0)$ divided by $(s_1 - s_0)$. Despite its practicality, this estimation strategy does not constrain the slope coefficients and variances to be equal on both sides of the cutoff, which translates into loss of efficiency. The mid-censored Tobit likelihood naturally takes these constraints into account and provides the most efficient estimates. It is therefore our preferred implementation.

Let $(y_i, X_i)$, $i = 1, \ldots, n$ be an iid sample of observations. The maximum likelihood estimator (MLE) for $(\varepsilon, \beta, \sigma)$ is constructed by maximizing the log-likelihood function of the

sample of $y_i$s, conditional on $X_i$s.

$$
\begin{aligned}
&L(y_1, \ldots, y_n | X_1, \ldots, X_n; \varepsilon, \beta, \sigma) \\
&\qquad = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{y_i < k\} \log \left[ \frac{1}{\sigma} \phi \left( \frac{y_i - \varepsilon s_0 - X_i \beta}{\sigma} \right) \right] \\
&\qquad\quad + \mathbb{I}\{y_i = k\} \log \left[ \Phi \left( \frac{k - \varepsilon s_1 - X_i \beta}{\sigma} \right) - \Phi \left( \frac{k - \varepsilon s_0 - X_i \beta}{\sigma} \right) \right] \\
&\qquad\quad + \mathbb{I}\{y_i > k\} \log \left[ \frac{1}{\sigma} \phi \left( \frac{y_i - \varepsilon s_1 - X_i \beta}{\sigma} \right) \right] \\
&\qquad \equiv \frac{1}{n} \sum_{i=1}^{n} \ell_i(\varepsilon, \beta, \sigma). \qquad (16)
\end{aligned}
$$

Regardless of the true distribution $F_{n^*|X}$, the MLE is consistent for the parameter that maximizes the population average of the log-likelihood function, that is, $(\bar{\varepsilon}, \bar{\beta}, \bar{\sigma}) \in \arg\max \mathbb{E}[l_i(\varepsilon, \beta, \sigma)]$. Standard textbook analyses of Tobit models demonstrate uniqueness of $(\bar{\varepsilon}, \bar{\beta}, \bar{\sigma})$ as solution to the maximization problem.[5] We say the elasticity is identified by a mid-censored Tobit when the true parameter $\varepsilon$ coincides with $\bar{\varepsilon}$. We show that the normality Assumption 14 is not necessary for identification of $\varepsilon$.

**Lemma 2.** *Let $G_{n^*}(n; \beta, \sigma, F_X) = \mathbb{E}\left[\Phi\left(\frac{n - X\beta}{\sigma}\right)\right]$, where the expectation is taken over the distribution of $X$ with CDF $F_X$. Assume the true distribution of $n^*$ belongs to the semi-parametric family*

$$
\mathcal{F}_{n^*} = \{G_{n^*}(n; \beta, \sigma, F_X), (\beta, \sigma, F_X) \in \mathbb{R}^{1 \times (d+1)} \times \mathbb{R}_+ \times \mathcal{F}_X\}, \qquad (17)
$$

*where $\mathcal{F}_X$ is the class of all CDFs of $X$. Suppose $\mathcal{F}_{n^*}$ satisfies (11)–(13) for the true $F_X$. Define $G_y(y; \varepsilon, \beta, \sigma, F_X)$ to be the unconditional CDF of $y$ obtained by transforming $G_{n^*}(n; \beta, \sigma, F_X)$, according to Equation 4 and a given value of $\varepsilon$. Let $F_y(y)$ be the true CDF of $y$. If $G_y(y; \bar{\varepsilon}, \bar{\beta}, \bar{\sigma}, F_X) = F_y(y)$, then $\bar{\varepsilon}$ equals the true elasticity, regardless of Assumption 14. This remains true if we replace the normal distribution by another parametric*

[5] For example, see Hayashi (2000), Section 8.3.

*distribution to form the semi-parametric mixture in* (17).

If the Tobit best-fit distribution of $y$ matches the true distribution of $y$, Lemma 2 guarantees that the elasticity estimated by the Tobit is consistent for the true elasticity, regardless of whether $F_{n^*|X}$ is normal. Essentially, Lemma 2 requires the unconditional distribution of $n^*$ to be a mixture of normal distributions, where the average is taken across the distribution of covariates. Standard quasi-MLE asymptotic inference procedures apply here. Namely, the MLE $(\hat{\varepsilon}, \hat{\beta}, \hat{\sigma})$ obtained from (16) and centered at $(\bar{\varepsilon}, \bar{\beta}, \bar{\sigma})$ is asymptotically normal, with zero mean and the usual variance-covariance matrix in the "sandwich form."

One of the features of bunching estimators is the reliance on data local to the kink point. With the mid-censored Tobit model, the researcher may also restrict the sample to observations of $y$ lying in a small neighborhood of $k$ and estimate a truncated Tobit.[6] The truncated Tobit is an attractive estimation strategy, because consistency of $\hat{\varepsilon}$ relies on a much weaker version of Assumption 14. Moreover, the smaller the truncation window, the easier it is to fit the unconditional distribution of $y$ with a Tobit, and the stronger is the robustness result of Lemma 2.

As a matter of routine, we recommend researchers estimate a truncated Tobit model for various window sizes around the kink point and examine two things: first, the plot of the estimated elasticity as a function of the size of the truncation window; second, the plot of the best-fit Tobit distribution of $y$ compared to the histogram of $y$ for various sizes of truncation windows. The distribution fit tends to improve as the size of the window decreases. The better the fit, the more likely the conditions of Lemma 2 are met, and the closer is the elasticity to the truth. We illustrate this exercise with simulated data below and with real data in Section 5.

[6] The truncated Tobit model has a log-likelihood that is slightly different than (16). Instead of the log-likelihood of $y|X$, it maximizes the log-likelihood of $y|X, k-\delta < y < k+\delta$ for $\delta > 0$, which has a truncated normal distribution.

Consider the following simulation experiment. Let $U_1$ and $U_2$ be Bernoulli with probability of success $\sqrt{2}/2$, and $U_3$ be normal with mean 1.59 and variance $0.7^2$, where all three variables are independent. Let the covariates be $X_1 = U_1$ and $X_2 = U_1 U_2$, and ability be $n^* = \sqrt{2}X_1 + 2X_2 + U_3$. This model was chosen to match moments of the real data in Section 5. Generate an iid sample with 500,000 observations of $X_1$, $X_2$, and $y$, according to Equation 4 with $\varepsilon = 1$. As in the EITC example in Section 5, the kink point is $k = 2.1494$ (i.e., log of 8.580), with $t_0 = -0.34$ and $t_1 = 0$.

The first exercise estimates a mid-censored Tobit that is correctly specified with both covariates $X_1$ and $X_2$. We start with the full sample of simulated data and produce estimates for truncation windows that are symmetric around the kink point and shrink in size. For example, Figures 2a and 2b show the histogram of simulated data for $y$, and the best-fit Tobit distributions for two truncation sizes, 100% and 40%. Although $f_{n^*|X_1,X_2}$ is normal, it is clear from the figures that $f_y$ is not a censored normal and therefore $f_n^*$ is not normal. Figure 2c displays the elasticity estimate as a function of the percentage of data used in each truncated estimation. The elasticity estimate is stable over all truncation windows, because the model is correctly specified. The Tobit fits the distribution of $y$ perfectly for all truncation windows, and the estimated elasticity is approximately equal to the truth.

The second exercise estimates a misspecified model using the same simulated data. Specifically, we drop $X_2$ out of the model. In this case, $f_{n^*|X_1}$ does not have a normal distribution, and Assumption 14 is not satisfied. Estimation using all of the data does not fit the distribution of $y$ (Figure 2d). On the other hand, Figure 2e demonstrates that the truncated Tobit matches the distribution of $y$ perfectly for windows that use 40% of the data or less. In line with Lemma 2, elasticity estimates converge to the truth, as the truncation window decreases below 40%.

### 4.2.2 Censored Quantile Regressions

Another type of semi-parametric assumption on the ability distribution consists of

restricting a quantile of the distribution of $n^*$, conditional on $X$. Namely, for $\tau \in (0,1)$, we assume that there exists an unique $\beta(\tau) \in \mathbb{R}^{1\times(d+1)}$ such that

$$Q_\tau\left(n^* \mid X\right) = X\beta(\tau), \tag{18}$$

where $Q_\tau$ denotes the $\tau$-th quantile of a distribution. A common choice in applied work is $\tau = 1/2$ or the median regression. The restriction in (18) may be a flexible one if one includes transformations of $X$ on the right-hand side, e.g., polynomials and interaction terms.

Equation 15 leads to $y = \min\{\varepsilon s_0 + n^*;\ \max\{k;\ \varepsilon s_1 + n^*\}\}$, which is an increasing and continuous function of $n^*$. The quantile of an increasing and continuous function of $n^*$ is equal to that same function evaluated at the quantile of $n^*$. Using Assumption 18,

$$Q_\tau\left(y \mid X\right) = \min\{\varepsilon s_0 + X\beta(\tau);\ \max\{k;\ \varepsilon s_1 + X\beta(\tau)\}\}. \tag{19}$$

For those observations such that $X'\beta(\tau) < k - \varepsilon s_0$ or $X'\beta(\tau) > k - \varepsilon s_1$, the quantile $Q_\tau\left(y \mid X\right)$ varies linearly with $X$; otherwise, it is constant and equal to $k$. Intuitively, if there is enough variation in $X$ for uncensored observations, then the slope coefficients and the intercepts are identified. This leads to identification of $\varepsilon$.

**Lemma 3.** *Define* $\tilde{X} = [X,\ \mathbb{I}\{Q_\tau\left(y \mid X\right) > k\}]$, *a random vector in* $\mathbb{R}^{1\times(d+2)}$. *Assume* $\mathbb{E}\left[\mathbb{I}\left\{Q_\tau\left(y \mid X\right) \neq k\right\}\tilde{X}'\tilde{X}\right]$ *has full rank and that Assumption 18 holds. Then* $\varepsilon$ *is identified.*

In the absence of covariates or restrictions on $Q_\tau\left(y \mid X\right)$, the rank condition is never satisfied. This confirms the impossibility demonstrated in Lemma 1. For example, suppose the researcher has two dummy variables, $W_1$ and $W_2$. An unrestricted $Q_\tau\left(y \mid W_1, W_2\right)$ contains four parameters, because the conditional quantile takes at most four possible values. In the best case scenario for identification, these four values are all different from $k$. In terms of Lemma 3, $d = 3$, and $X = [1, W_1, W_2, W_1W_2]$ is $1 \times 4$. The matrix

$\mathbb{E}\left[\mathbb{I}\left\{Q_\tau\left(y \mid X\right) \neq k\right\}\tilde{X}'\tilde{X}\right]$ is $5 \times 5$ but has rank equal to 4 at most. Thus $Q_\tau\left(y \mid X\right)$ must be restricted to fewer parameters for identification to be possible.

Theoretical work on estimation and inference of parameters in censored quantile regression (CQR) models dates back to the 1980s (Powell (1984, 1986)). Recent advances include the computationally attractive three-step estimator by Chernozhukov and Hong (2002), and CQR with endogeneity by Chernozhukov et al. (2015). In the simpler case of $Q_\tau\left(y \mid X\right) = X\beta(\tau)$, Koenker and Bassett (1978) show that a consistent estimator for $\beta(\tau)$ is obtained by the solution to the problem

$$\min_{b \in \mathbb{R}^{d+1}} \sum_{i=1}^{n} \left[\rho_\tau\left(y_i - X_i b\right)\right], \tag{20}$$

where $(y_i, X_i)$ $i = 1, \ldots, n$ is an iid sample and $\rho_\tau\left(u\right) = \left(\tau - 1\left(u \leq 0\right)\right)u$ is the so-called "check function." In our case, the parametric conditional quantile function $Q_\tau\left(y \mid X\right)$ is given in Equation 19. The slope and intercept coefficients are estimated by

$$(\hat{b}(\tau), \hat{\delta}(\tau)) = \arg \min_{b \in \mathbb{R}^d, \delta \in \mathbb{R}} \sum_{i=1}^{n} \left[\rho_\tau\left(y_i - \min\{X_i'b;\ \max\{k;\ X_i'b + \delta\}\}\right)\right], \tag{21}$$

where $\hat{b}(\tau)$ is consistent for $\beta(\tau) + [\varepsilon s_0,\ 0,\ \ldots,\ 0]'$, and $\hat{\delta}(\tau)$ is consistent for $\varepsilon(s_1 - s_0)$. Therefore the elasticity is consistently estimated by $\hat{\varepsilon} = \hat{\delta}/(s_1 - s_0)$, and it is asymptotically normal.

The optimization problem in Equation 21 is computationally difficult. For the left (or right) censored case, Chernozhukov and Hong (2002) proposed a fast and practical estimator that consists of three steps. Our case of middle censoring requires a straightforward modification of their method. We delineate practical steps to obtain $\hat{\varepsilon}$ and its standard error using CQR in Section B.5 of the supplemental appendix.

## 5 Application to EITC

We demonstrate and compare our new methods using bunching behavior created by kinks in the earned income tax credit (EITC). Each method differs in the assumptions they make about the unobserved distribution to achieve identification. There is no way to determine which assumption is correct because the unobserved distribution is not fully identified. Nevertheless, estimates that are stable across many methods indicate that different identifying assumptions do not play a major role in the construction of those estimates. On the contrary, estimates that are sensitive to different assumptions are dependent on the validity of those assumptions. Patel, Seegert, and Smith (2016) provide an empirical illustration of this sensitivity.

First, we use our non-parametric bounds to provide initial information about how sensitive the elasticity estimate is to different shapes of the underlying ability distribution. When the bounds are tight, then the shape of the underlying distribution is not critical. But when the bounds are wide, then the shape is critical. In this case, reducing the range of possible elasticities requires either stronger restrictions on the shape of the ability distribution or additional data on determinants of ability.

Second, we combine observed determinants of ability with our semi-parametric approach to point identify the elasticity. We compare the resulting best-fit Tobit income distribution to the observed distribution for alternative samples that range from using all observations to using only data local to the kink. When the best-fit Tobit distribution coincides with the observed distribution, the estimated elasticity is consistent (Lemma 2). Furthermore, if the Tobit elasticity is within narrow non-parametric bounds, then the identifying assumptions are inconsequential; if within wide bounds, then the identifying assumptions are not contradictory and the covariates provide point identification. In contrast, if the Tobit elasticity is outside of the bounds, then the elasticity estimate is not robust to the two alternative identifying assumptions. Finally, when the best-fit Tobit distribution does not

coincide with the observed distribution, the determinants of ability used for estimation are uninformative or the semi-parametric assumption is inappropriate.

We recommend that researchers examine the sensitivity of elasticity estimates across all available methods as a matter of routine. We illustrate these steps in the context of the EITC in the rest of this section.

### 5.1 Data

We use data from the Individual Public Use Tax Files, constructed by the IRS. The annual cross-section for each year 1995 to 2004 includes sampling weights which allow interpretation of any estimates as being based on the population of U.S. income tax returns. This data was initially used by Saez (2010) to demonstrate how to use bunching to estimate an elasticity.[7]

The income distribution for individuals with one child demonstrates clear bunching around the $8,580 kink (year 2008 dollars) in the EITC schedule. Because the marginal tax rate increases from −34 percent to 0 percent at $8,580, individuals have strong incentives to report less income than if the tax rate had remained −34 percent above the kink point.

Observed bunching in the distribution of income suggests that people do respond to changes in tax rates. To effectively set tax rates, however, it is imperative to quantify this response precisely. Small variation in elasticity estimates imply large differences in optimal tax rates. For example, variation in the elasticity of taxable income between 0.1 and 0.2 implies an optimal top marginal tax rate between 82% or 69%.[8] As demonstrated above,

[7]We replicate some of the estimates in Saez (2010) using publicly available code from the website of the *American Economic Journal: Economic Policy* and report them in supplemental Appendix B.6.

[8]This example comes from Saez (2001). In particular, Equation 9 states $\bar{\tau} = (1-g)/(1-g+\varepsilon^u\varepsilon^c(a-1))$, where $g$ is defined as the value the government has for the marginal consumption of high income earners (often set to 0), $a$ is the Pareto parameter (with baseline value of 2), and $\varepsilon^c$ and $\varepsilon^u$ are the compensated and uncompensated elasticities of taxable income. For the calculation in the text, we utilize $\varepsilon^u = \varepsilon^c$, a Pareto parameter of 2, and a $g$ value of 0.1.

identifying the elasticity requires information on the amount of bunching and the income distribution. The following sections show how different methods leverage different types of variation to identify the elasticity.

The methods of this paper are designed for data without friction errors or sharp bunching. For examples of sharp bunching data, see Figure 4 by Glogowsky (2018) and Figure 1 by Goncalves and Mello (2018). Nevertheless, the IRS data do have friction errors as the excess mass due to bunching is visibly dispersed in a small interval near the kink (for example, Figure 5 by Weber (2016)). Therefore, to apply our procedures to the IRS data, we first need to filter reported income out of friction error. A proper deconvolution theory must be developed to tackle this problem, but it is beyond the scope of this paper. For now, we simply need a practical way of removing friction error before applying the different bunching estimators, so that they may be properly compared.

Following the intuition of Chetty et al. (2011), we fit a seventh-order polynomial to the empirical CDF of reported income with friction errors $\tilde{y}$. As does Saez (2010), we exclude observations that lie within \$1,500 of the kink and allow an intercept change at the kink. The extrapolation of the fitted polynomial to the excluded region results in a CDF with a jump discontinuity at the kink. This is an estimate for the CDF of income without friction error, that is, $F_y(y)$. The size of the discontinuity equals the bunching mass. We then rely on the fact that $y = F_y\left(F_{\tilde{y}}^{-1}(\tilde{y})\right)$ and use the estimated CDFs to transform $\tilde{y}$ into $y$.

Our filtering procedure is different from the polynomial strategy discussed in Section 2.3. We simply aim at removing the friction error from the sample, while the the polynomial strategy of Example 2 aims to remove friction error and recover the counterfactual distribution of income, which requires much stronger restrictions according to Lemma 1. Our filtering procedure works well in cases in which 1) the researcher has a good prior on the support of the friction error distribution (\$1,500 in this case), 2) the friction error affects bunching individuals more than non-bunching individuals, or 3) the variance of the friction

error is small. A more general filtering method is deferred to future work.[9]

### 5.2 Estimates Across Methods

Table 1 reports estimates of the elasticity of taxable income using a classic bunching method, non-parametric bounds, and Tobit models with covariates. Each of these estimates relies on a different set of assumptions to identify the elasticity of taxable income, and together they provide insights into which assumptions are most defensible in the context of the EITC.

Column 1 reports our estimates of the elasticity of taxable income using a trapezoidal approximation (Example 1).[10] This method assumes the unobserved PDF is linear in the bunching region, which prior literature believed to approximate non-linear distributions well. In practice, the appropriateness of this approximation depends on the true distribution and length of the bunching region, which are both unobserved. Linearity may be inappropriate if the distribution is sufficiently non-linear or the bunching region is wide. Column 1 demonstrates substantial heterogeneity in estimates across different subsamples. In particular, the elasticity estimate is 0.426 for the all filers sample, 0.854 for self-employed individuals, 1.102 for self-employed married individuals, and 0.784 for self-employed not married individuals.

The guidelines for implementation of our non-parametric bounds in Section 4.1 utilizes a range of values for $M$ that includes the maximum slope magnitude of $f_y$. We reiterate that $M$ is unidentified and that the slope of $f_y$ provides a starting point. The `bunching` Stata package consistently estimates the maximum slope of $f_y$ by taking the maximum slope in the histogram of $y$ across all consecutive bins. We find that the slope is never bigger than 0.5 across our subsamples. For a more conservative view, we report our non-parametric

[9] Supplemental Appendix B.6 recomputes our estimates using the filtering procedure employed by Saez (2010).

[10] We estimate the PDF of the variables in logs rather than in levels, which simplifies the elasticity formula based on the trapezoidal approximation in Example 1.

bounds using $M = 0.5$ and $M = 1$ in Table 1, Columns 2 and 3, and plot bounds for $M$ up to 2 in Figure 3. The vertical lines in these figures designate the minimum and maximum slope, such that both the upper and lower bounds are finite numbers. The first line is the smallest slope that allows a continuous PDF to be consistent with both the bunching mass and observed income distribution. At the minimum slope, both lower and upper bounds are equal to the estimate based on the trapezoidal approximation, reported in column 1.

As $M$ increases, the set of possible PDF shapes in the bunching region becomes richer. The second line is the maximum slope before the set of possible distributions allows for a PDF that touches zero in the bunching interval. In that case, the bunching mass remains constant for arbitrarily large $\varepsilon$, and the upper bound is infinity (Theorem 2).

A large range between lower and upper bounds in Figure 3 suggests the estimates change substantially with the shape of the unobserved distribution. For example, the bounds are uninformative for the self-employed married sample, even for small values of $M$. This indicates that the data will not provide precise information on the elasticity unless the researcher imposes further functional form restrictions on the distribution of $n^*$. In contrast, we learn the most in the case of all filers and self-employed not married, where the bounds are narrower than in other subsamples for $M = 0.5$. The lower bound is always defined for larger choices of $M$, which gives partial information on the elasticity without the need of being precise with the choice of $M$. For the exceedingly high value of $M = 2$, the lower bound is about 0.25 for all filers and 0.5 for the other three subsamples.

Columns 4–7 report our estimates of the Tobit model using the full sample and truncated samples at 75%, 50%, and 25% of the data. Figures 4–7 complement these estimates by graphing the actual distribution and the implied distribution from the Tobit estimates at different levels of truncation in panels a through e. These estimates incorporate covariates, including indicator variables for married, tax preparer used, real estate interest deduction, employment status, contributed to charity, tax form used, and tax filing status. The fit of the Tobit model generally improves as we truncate the sample closer to the kink, which

implies that the semi-parametric assumption of mixed normals is more reasonable locally than globally. The minimum truncation necessary for a reasonable fit varies by subsample. For example, for self-employed not married, the fit seems reasonable using 80% or less of the data, but for all filers, the fit only becomes reasonable at around 20%. It is interesting to observe that the Tobit with covariates fit the distribution better in narrower cuts of the data than for all filers.

Panel f in Figures 4–7 graph the elasticity estimate as a function of the percentage of data used. The estimates tend to plateau as the distribution fits improve. For example, in the self-employed not married sample depicted in Figure 7, the estimates are all around 0.75, using less than 80% of the data. It is worth pointing out that truncated samples with less than 20% of the data lead to numerical issues, such as perfect collinearity of covariates and lack of convergence in the likelihood maximization. This leads to imprecise estimates, as indicated by an upward bend in left extremity of the curves depicted in panel f of Figures 4–7.

### 5.3 Comparisons Across Methods

Comparisons across methods provides insights into the reasonableness of different assumptions used to estimate the elasticity. The trapezoidal approximation is always within the bounds, because its estimate is based on a linear interpolation of the PDF in the bunching region. The slope of such line equals the minimum slope for which the bounds are defined. In contrast, the Tobit model using 100% of the data is often below the lower bounds, but the Tobit distribution fails to fit the observed distribution of income globally. Truncated Tobit estimates generally enter the bounds as the truncation window decreases and, as a result, the fit of the Tobit distribution improves. For the all filers sample, an M larger than 1 is needed for the bounds to cover the Tobit estimate truncated at 25%. This reiterates our previous discussion that the Tobit fit for all filers is poor until we use 20% or less of the data.

Consider self-employed married and self-employed not married filers. Figures 6 and 7

demonstrate that the bunching mass is larger for self-employed not married individuals than for self-employed married individuals. This difference in bunching mass might lead a researcher to conjecture that the elasticity is larger for self-employed not married individuals. Whether this conjecture is true depends, however, on differences in the underlying distribution of heterogeneity. Estimates based on the trapezoidal approximation contradict that conjecture. The estimates in column 1 of Table 1 are larger for self-employed married individuals than self-employed not married. The global distributional assumption of a mixture of normals averaged over covariates produces a larger elasticity for self-employed not married (column 4 in Table 1). However, the credibility of these Tobit estimates is questioned by the poor fit shown in panel a of Figures 6 and 7. Truncating the sample obtains a better fit, and we find that the elasticities are approximately the same for married and not married. The disagreement across methods for these subsamples indicates that assumptions on the distribution of heterogeneity are critical to obtain informative elasticity estimates.

## 6 Conclusion

We show how to use bunching from piecewise-linear budget constraints to identify elasticities, under conditions weaker than those used in the literature on kinks and notches. The key theoretical point is that bunching is determined by the elasticity parameter and the shape of an unobserved distribution. Additional assumptions or data are needed to identify the elasticity.

We propose a suite of estimation techniques that allow researchers to tailor their estimation to different assumptions and data variation. These include non-parametric bounds and semi-parametric censored models with covariates. The non-parametric bounds are the least restrictive method and also nest estimators from the previous literature.

These techniques have wide applicability, because piecewise-linear budget constraints are common across fields, from public finance and labor, to industrial organization and

accounting. Our estimation strategies also provide a foundation for future advances in techniques that will account for different empirical hurdles. Of particular interest are extensions that consider optimization and friction errors, extensive margin responses, and panel data methods.

## 7 Acknowledgements

The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Board or the Federal Reserve System. We would like to thank Matias Cattaneo, Bill Evans, Roger Gordon, Jim Hines, Dan Hungerman, Michael Jansson, Henrik Kleven, Brian Knight, Erzo Luttmer, Byron Lutz, Dayanand Manoli, Magne Mogstad, Whitney Newey, Andreas Peichl, Emmanual Saez, Dan Silverman, and Joel Slemrod for valuable comments and discussions. The paper also benefited from feedback received from seminar participants at the UCSD Workshop on Bunching Estimators, Econometric Society, International Association for Applied Econometrics, International Institute of Public Finance, National Tax Association, Dartmouth College, Federal Reserve Board, and University of Michigan. Jessica C. Liu, Michael A. Navarrete, and Alexis M. Payne provided excellent research assistance. All remaining errors are our own. Bertanha acknowledges financial support received while visiting the Kenneth C. Griffin Department of Economics, University of Chicago.

## References


Allen, E. J., P. M. Dechow, D. G. Pope, and G. Wu (2017, June). Reference-Dependent Preferences: Evidence from Marathon Runners. *Management Science 63*(6), 1657--1672.

Amemiya, T. (1984). Tobit Models: A Survey. *Journal of Econometrics 24*(1-2), 3--61.

Bastani, S. and H. Selin (2014). Bunching and Non-bunching at Kink Points of the Swedish Tax Schedule. *Journal of Public Economics 109*, 36--49.

Bertanha, M., A. H. McCallum, and N. Seegert (2018, March). Better Bunching, Nicer Notching. Working Paper 3144539, SSRN.

Bertanha, M. and M. J. Moreira (2020). Impossible inference in econometrics: Theory and applications. *Journal of Econometrics*.

Best, M. C. and H. J. Kleven (2018). Housing Market Responses to Transaction Taxes: Evidence From Notches and Stimulus in the UK. *Review of Economic Studies 85*(1), 157--193.

Blomquist, S., A. Kumar, C.-Y. Liang, and W. Newey (2015, May). Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income. Working Paper 21/15, Cemmap.

Blomquist, S., A. Kumar, C.-Y. Liang, and W. Newey (2019, October). On Bunching and Identification of the Taxable Income Elasticity. Working Paper 53/19, Cemmap.

Blomquist, S. and W. Newey (2017, September). The Bunching Estimator Cannot Identify the Taxable Income Elasticity. Working Paper 40/17, Cemmap.

Blomquist, S. and W. Newey (2018, March). The Kink and Notch Bunching Estimators Cannot Identify the Taxable Income Elasticity. Working Paper 2018:4, Uppsala Universitet.

Burtless, G. and J. A. Hausman (1978). The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment. *Journal of Political Economy 86*(6), 1103--1130.

Caetano, C. (2015). A Test of Exogeneity Without Instrumental Variables in Models With Bunching. *Econometrica 83*(4), 1581--1600.

Caetano, C., G. Caetano, and E. Nielsen (2020a). Correcting endogeneity bias in models with bunching. Technical report, Working Paper.

Caetano, C., G. Caetano, and E. R. Nielsen (2020b). Should children do more enrichment activities? leveraging bunching to correct for endogeneity. Technical Report 2020-036, Board of Governors of the Federal.

Caetano, G., J. Kinsler, and H. Teng (2019). Towards causal estimates of children's time allocation on skill development. *Journal of Applied Econometrics 34*(4), 588--605.

Caetano, G. and V. Maheshri (2018). Identifying dynamic spillovers of crime with a causal approach to model selection. *Quantitative Economics 9*(1), 343--394.

Cattaneo, M., M. Jansson, X. Ma, and J. Slemrod (2018, March). Bunching Designs: Estimation and Inference. Working paper, UCSD Bunching Workshop.

Cattaneo, M. D., M. Jansson, and X. Ma (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association 0*(0), 1--7.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019, August). The Effect of Minimum Wages on Low-wage Jobs. *Quarterly Journal of Economics 134*(3), 1405--1454.

Chernozhukov, V., I. Fernández-Val, and A. E. Kowalski (2015). Quantile Regression with Censoring and Endogeneity. *Journal of Econometrics 186*(1), 201--221.

Chernozhukov, V. and H. Hong (2002). Three-step Censored Quantile Regression and Extramarital Affairs. *Journal of the American Statistical Association 97*(459), 872--882.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *Quarterly Journal of Economics 126*(2), 749--804.

Chetty, R., J. N. Friedman, and E. Saez (2013, December). Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings. *American Economic Review 103*(7), 2683--2721.

Dee, T. S., W. Dobbie, B. A. Jacob, and J. Rockoff (2019, July). The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations. *American Economic Journal: Applied Economics 11*(3), 382--423.

DeMaris, A. (2005). Truncated and Censored Regression Models. In *Regression with Social Data: Modeling Continuous and Limited Response Variables*, Chapter 9, pp. 314--347. John Wiley & Sons, Ltd.

Devereux, M. P., L. Liu, and S. Loretz (2014). The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records. *American Economic Journal: Economic Policy 6*(2), 19--53.

Dhrymes, P. J. (1986). Limited Dependent Variables. In Z. Griliches and M. D. Intriligator (Eds.), *The Handbook of Econometrics*, Volume 3 of *6*, Chapter 27, pp. 1567--1631. North Holland.

Einav, L., A. Finkelstein, and P. Schrimpf (2017). Bunching at the Kink: Implications for Spending Responses to Health Insurance Contracts. *Journal of Public Economics 146*, 27--40.

Garicano, L., C. Lelarge, and J. Van Reenan (2016, November). Firm Size Distortions and the Productivity Distribution: Evidence from France. *American Economic Review 106*(11), 3439--3479.

Ghanem, D., S. Shen, and J. Zhang (2019, January). A Censored Maximum Likelihood Approach to Quantifying Manipulation in China's Air Pollution Data. Working paper, University of California - Davis.

Glogowsky, U. (2018). Behavioral Responses to Wealth Transfer Taxation: Bunching Evidence from Germany. Working Paper 3111993, SSRN.

Goncalves, F. and S. Mello (2018). A Few Bad Apples? Racial Bias in Policing. Working paper, University of California - Los Angeles.

Greene, W. H. (2005). Censored Data and Truncated Distributions. In T. Mills and K. Patterson (Eds.), *Palgrave Handbook of Econometrics*, Volume 1 of *5*, Chapter 20, pp. 695--736. London: Palgrave Macmillan.

Grossman, D. and U. Khalil (2019). Neighborhood Networks and Program Participation. *Journal of Health Economics 70*(forthcoming), 102257.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Heckman, J. J. (1976, January). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, NBER Chapters, pp. 475--492. National Bureau of Economic Research, Inc.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153--161.

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

Ito, K. and J. M. Sallee (2018, May). The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel Economy Standards. *Review of Economics and Statistics 100*(2), 319--336.

Jales, H. (2018). Estimating the effects of the minimum wage in a developing country: A density discontinuity design approach. *Journal of Applied Econometrics 33*(1), 29--51.

Jales, H. and Z. Yu (2017, January). Identification and estimation using a density discontinuity approach. In M. D. Cattaneo and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory and Applications*, Volume 38, pp. 29--72. Emerald Publishing Limited.

Khalil, U. and N. Yildiz (2017). A test of the selection-on-observables assumption using a discontinuously distributed covariate. Technical report, working paper.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*, 435--464.

Kleven, H. J. and M. Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics 128*(2), 669--723.

Koenker, R. and G. Bassett (1978). Regression Quantiles. *Econometrica 46*(1), 33--50.

Kopczuk, W. and D. Munroe (2015). Mansion Tax: The Effect of Transfer Taxes on the Residential Real Estate Market. *American Economic Journal: Economic Policy 7*(2), 214--57.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables* (2 ed.). SAGE Publications.

Maddala, G. S. (1983). *Limited-dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs. Cambridge University Press.

Patel, E., N. Seegert, and M. G. Smith (2016). At a Loss: The Real and Reporting Elasticity of Corporate Taxable Income. Working Paper 2608166, SSRN.

Powell, J. L. (1984). Least Absolute Deviations Estimation for the Censored Regression Model. *Journal of Econometrics 25*(3), 303--325.

Powell, J. L. (1986). Censored Regression Quantiles. *Journal of Econometrics 32*(1), 143--155.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies 68*(1), 205--229.

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

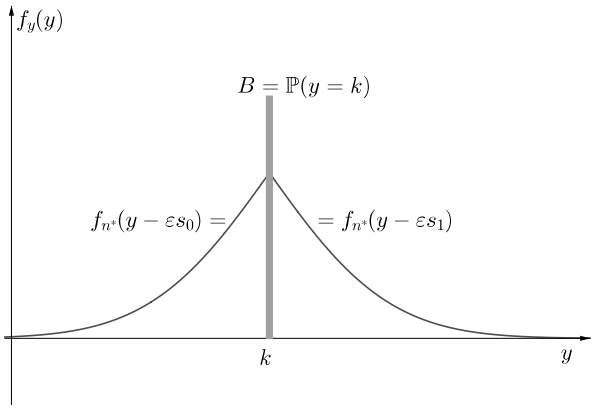
Sallee, J. M. and J. Slemrod (2012). Car Notches: Strategic Automaker Responses to Fuel Economy Policy. *Journal of Public Economics 96*(11), 981--999.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica 26*(1), 24--36.

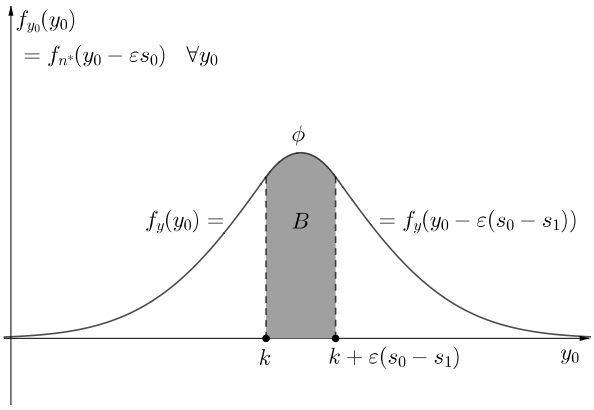
Weber, C. (2016). Does the Earned Income Tax Credit Reduce Saving by Low-Income Households? *National Tax Journal 69*(1), 41--76.

Figure 1: Identification of the Elasticity in the Case of a Kink

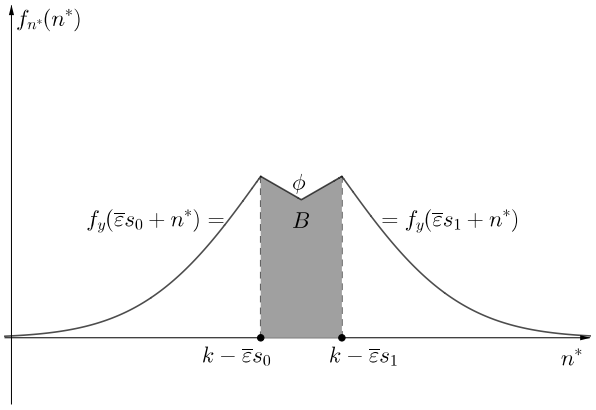
(a) Distribution of Observed Income



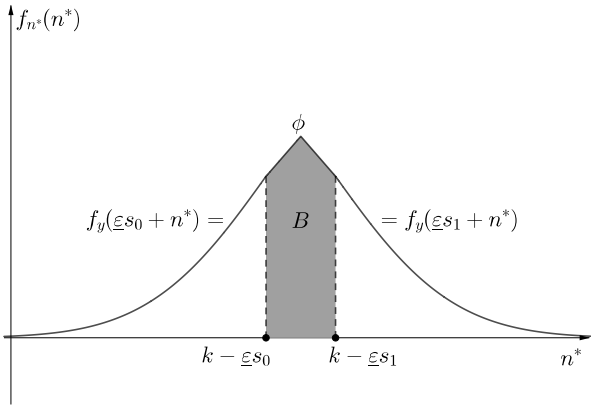

(b) Counterfactual Distribution of Income in the Absence of Kink




(c) Distribution of Ability Consistent with Observed Income and Higher Elasticity

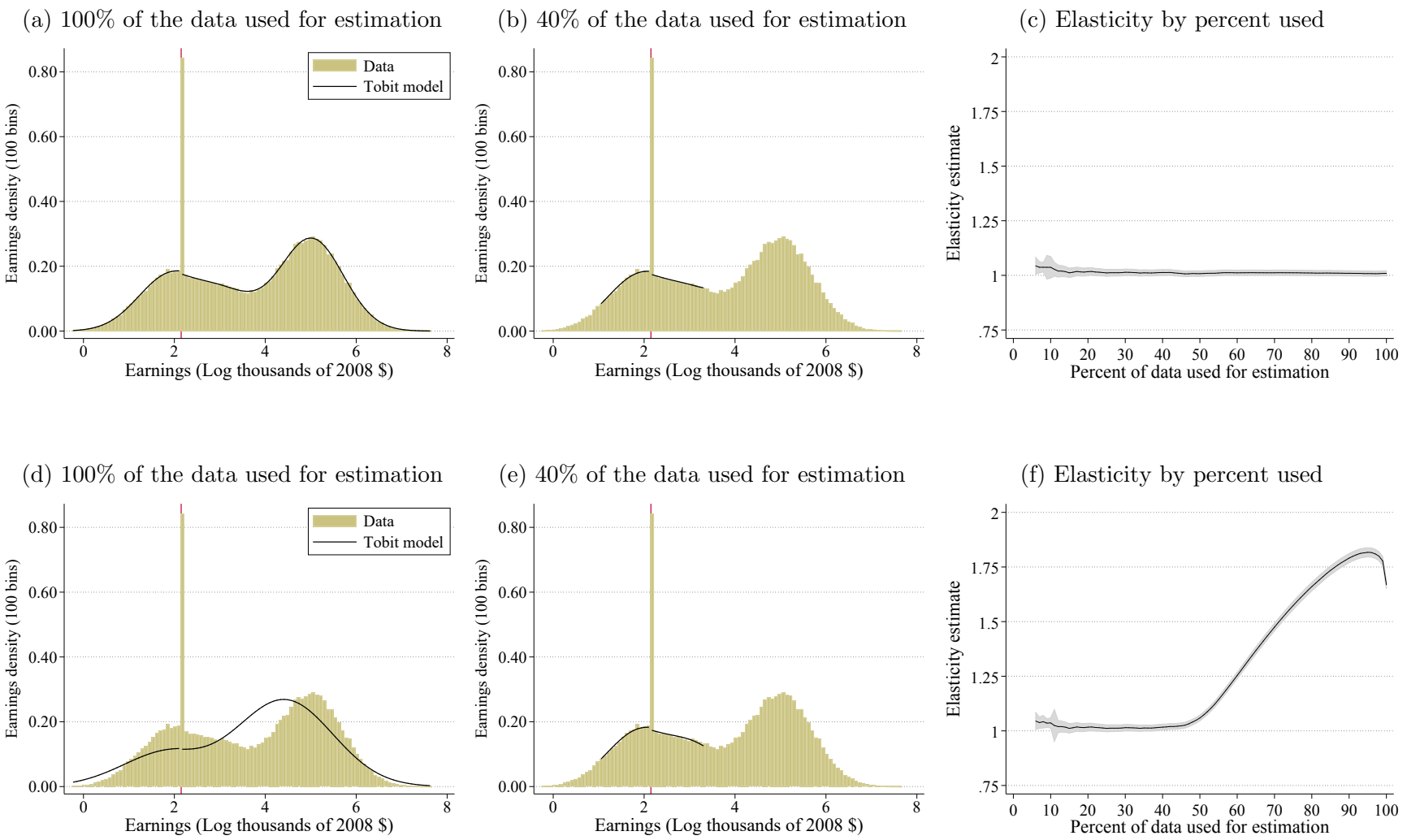



(d) Distribution of Ability Consistent with Observed Income and Lower Elasticity




*Notes:* Panel 1a plots an example of PDF of $y$. The continuous portions are equal to the PDF of ability $n^*$ shifted by $\varepsilon s_0$ for $y < k$, and by $\varepsilon s_1$ for $y > k$, respectively. The shaded area represents a discrete mass point with probability $B = \mathbb{P}(y = k)$, that is, the probability of bunching. Panel 1b shows the counterfactual PDF of $y_0$, that is, the distribution of income if tax rates did not change at the kink. The PDF of $y_0$ is continuous, and equals the PDF of $n^*$ shifted by $\varepsilon s_0$. It is also equal to the PDF of $y$ before the kink, and to the shifted PDF of $y$ after the kink. However, the distribution of $y$ does not reveal the shape of the PDF of $y_0$ in the bunching region (i.e. $\phi$). The shaded area under $\phi$ integrates to the probability of bunching $B$. The last two panels (Panels 1c-1d) display two different distributions of $n^*$ that generate the same distribution of income $y$ (Panel 1a) with two different elasticities, $\underline{\varepsilon} < \bar{\varepsilon}$, according to Equation 4. The PDF of $n^*$ outside of the bunching region is equal to the PDF of $y$ shifted by $\varepsilon s_0$, if $n^* < k - \varepsilon s_0$; or shifted by $\varepsilon s_1$, if $n^* > k - \varepsilon s_1$. Aside from $B$, the distribution of income does not contain any information about the shape of $\phi$ in the PDF of $n^*$. If we assume $f_{n^*}$ is Lipschitz continuous with known constant, it is possible to derive upper and lower bounds for $\phi$, which correspond, respectively, to lower and upper bounds on the elasticity (Theorem 2).

Figure 2: Robustness of Tobit Estimates to Lack of Normality




*Notes:* The simulation experiment illustrates the robustness of Tobit estimates to deviations from the normality assumption (Assumption 14). The experiment generates 500,000 observations of $y$, and two covariates $(X_1, X_2)$, assuming $\varepsilon = 1$, and $n^*|X_1, X_2 \sim \text{Normal}(\beta_0 + \beta_1 X_1 + \beta_2 X_2, \sigma^2)$ (see details in Section 4.2.1). As in the EITC case, the kink point is $k = 2.1494$, with $t_0 = -0.34$ and $t_1 = 0$. The first exercise estimates a mid-censored Tobit that is correctly specified with both covariates $X_1$ and $X_2$. Panels (a) and (b) show the histogram of simulated data for $y$, and the best-fit Tobit distributions for two truncation sizes, 100% and 40% of the sample used. Panel (c) displays the elasticity estimate as a function of the percentage of data used in each truncated estimation, along with 95% confidence bands. The second exercise drops $X_2$ and estimates a misspecified model. Panels (d)-(f) are analogous to Panels (a)-(c), except that they use the estimates from the misspecified Tobit model, where $n^*|X_1$ is not normal. The estimation truncated at 40% fits the distribution of $y$, and the elasticity converges to the true value (Lemma 2).

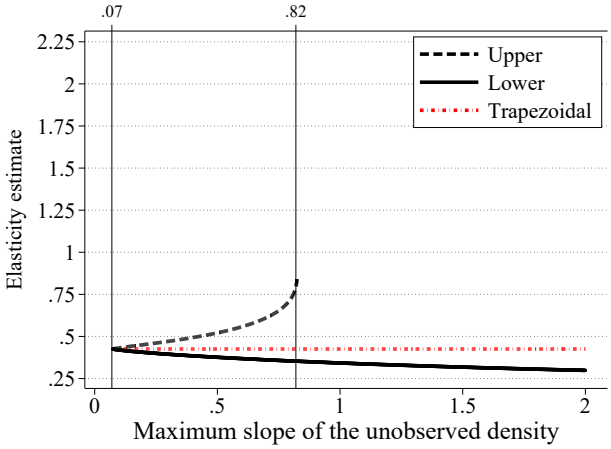
Table 1: Estimates Using U.S. Tax Returns 1995--2004

| Statistical Model | (1) Trapezoidal Approximation | (2) Theorem 2 Bounds M = 0.5 | (3) Theorem 2 Bounds M = 1 | (4) Tobit Full Sample | (5) Tobit Trunc. 75% | (6) Tobit Trunc. 50% | (7) Tobit Trunc. 25% | (8) Sample details | |
|---|---|---|---|---|---|---|---|---|---|
| *All* | | | | | | | | Obs. | 189.1m |
| Elasticity ($\varepsilon$) | 0.426 | $[0.376, 0.521]$ | $[0.342, \infty]$ | 0.195 | 0.280 | 0.291 | 0.326 | Avg. | \$54.1k |
| | (0.0289) | | | (0.0001) | (0.0002) | (0.0002) | (0.0002) | Std. | \$131.1k |
| *Self-employed* | | | | | | | | Obs. | 33.5m |
| Elasticity ($\varepsilon$) | 0.854 | $[0.721, 1.294]$ | $[0.639, \infty]$ | 0.603 | 0.790 | 0.787 | 0.796 | Avg. | \$61.8k |
| | (0.0885) | | | (0.0006) | (0.0008) | (0.0008) | (0.0009) | Std. | \$168.2k |
| *Self-employed, married* | | | | | | | | Obs. | 24.0m |
| Elasticity ($\varepsilon$) | 1.102 | $[0.718, \infty]$ | $[0.587, \infty]$ | 0.373 | 0.586 | 0.692 | 0.722 | Avg. | \$75.0k |
| | (0.3081) | | | (0.0006) | (0.0010) | (0.0012) | (0.0013) | Std. | \$185.6k |
| *Self-employed, not married* | | | | | | | | Obs. | 9.6m |
| Elasticity ($\varepsilon$) | 0.784 | $[0.741, 0.843]$ | $[0.692, 0.974]$ | 0.894 | 0.749 | 0.713 | 0.753 | Avg. | \$28.7k |
| | (0.1024) | | | (0.0010) | (0.0009) | (0.0009) | (0.0014) | Std. | \$106.3k |

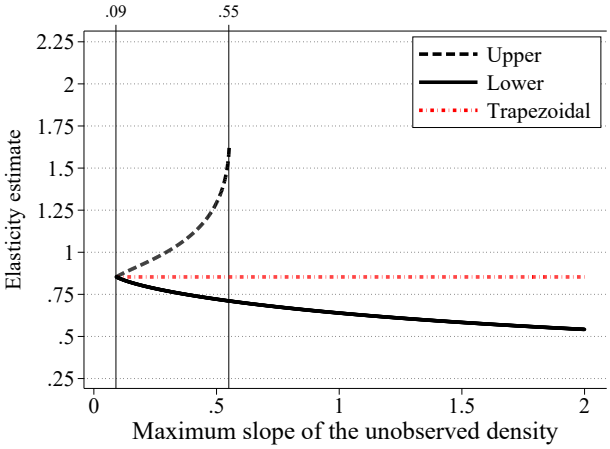
*Notes:* The table shows estimates of the elasticity for four different subsamples of the IRS data, and using three different approaches discussed in the paper. The first approach (column 1) uses the trapezoidal approximation to point-identify the elasticity (Example 1). We obtained non-parametric estimates of the side limits of $f_y$ at the kink using the method of Cattaneo, Jansson, and Ma (2019). The estimate for the bunching mass equals the sample proportion of $y$ observations that equals the kink point (see discussion on friction errors in Section 5.1). We obtained standard errors using 100 bootstrap iterations. The second approach (columns 2 and 3) uses the same estimates of the bunching mass and side limits to compute partially identified sets for the elasticity (Theorem 2). Upper and lower bounds are calculated for two choices of M, that is, the maximum slope of the PDF of the unobserved heterogeneity $n^*$. Column 4 has Tobit MLE estimates of the elasticity that utilizes the full sample of data, along with robust standard errors. Columns 5 through 7 report truncated Tobit MLE estimates. As we move from column 5 to column 7, we restrict the estimation sample to shrinking symmetric windows around the kink that utilizes 75% to 25% of the data. The set of covariates that enters the Tobit estimation is kept constant across different truncation windows. It includes dummy variables such as marital and employment status, year effects, types of deductions or social security benefits received, and whether the filer used a tax prep software.

Figure 3: Partial Identification Bounds for the Elasticity

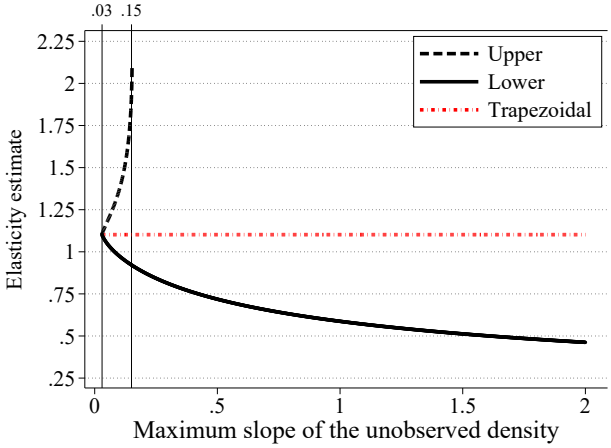
(a) All Filers



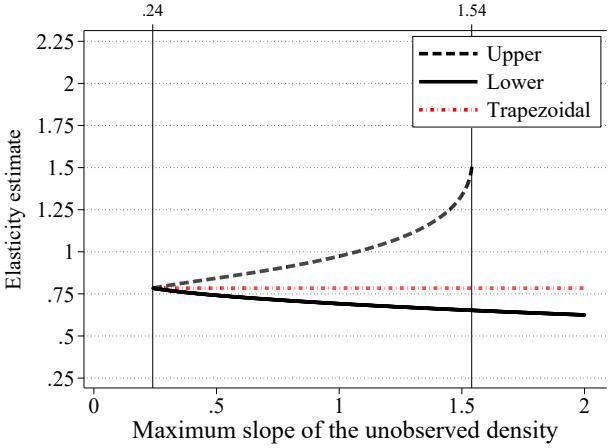

(b) Self-Employed Filers




(c) Self-employed and Married Filers

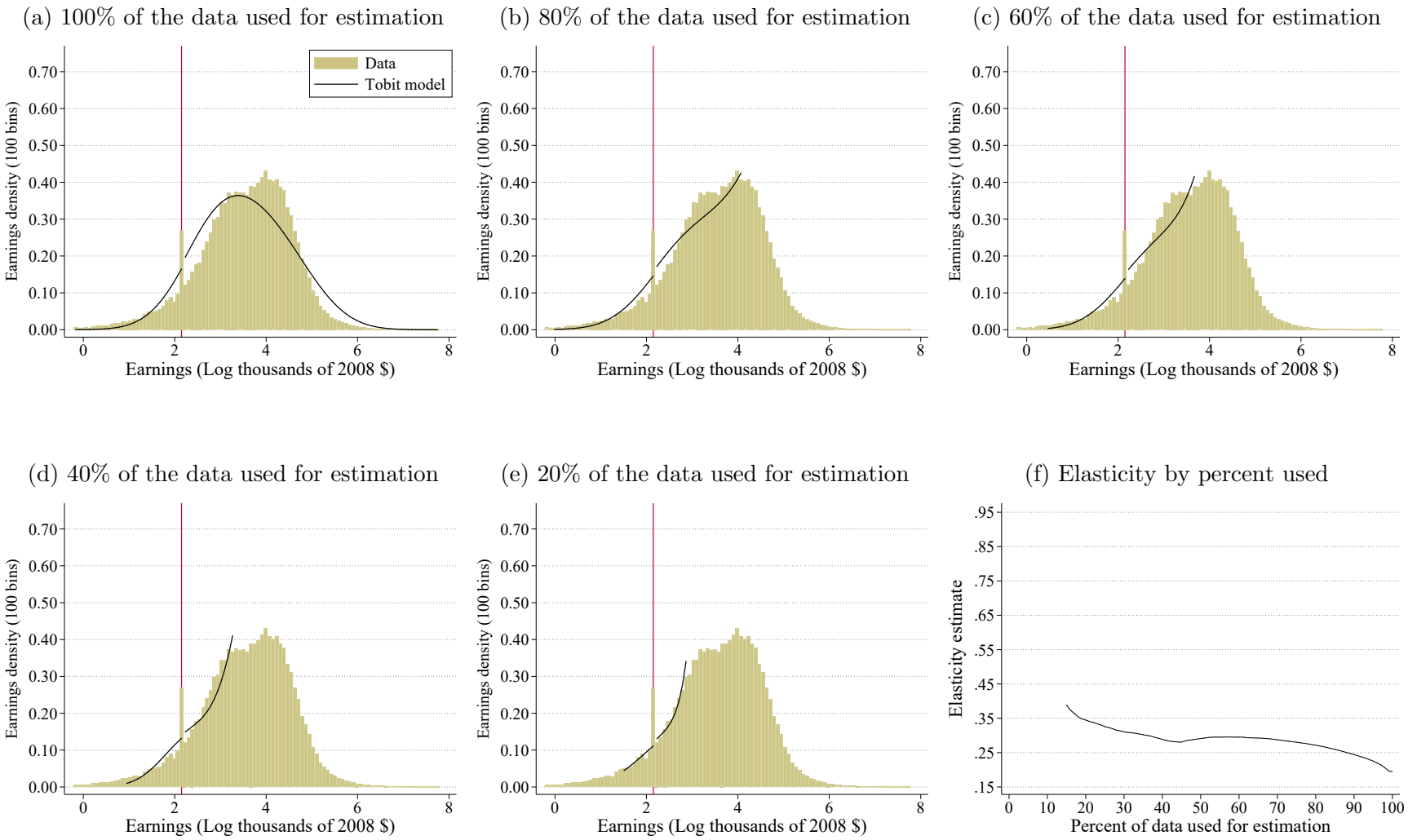



(d) Self-employed and Not Married Filers



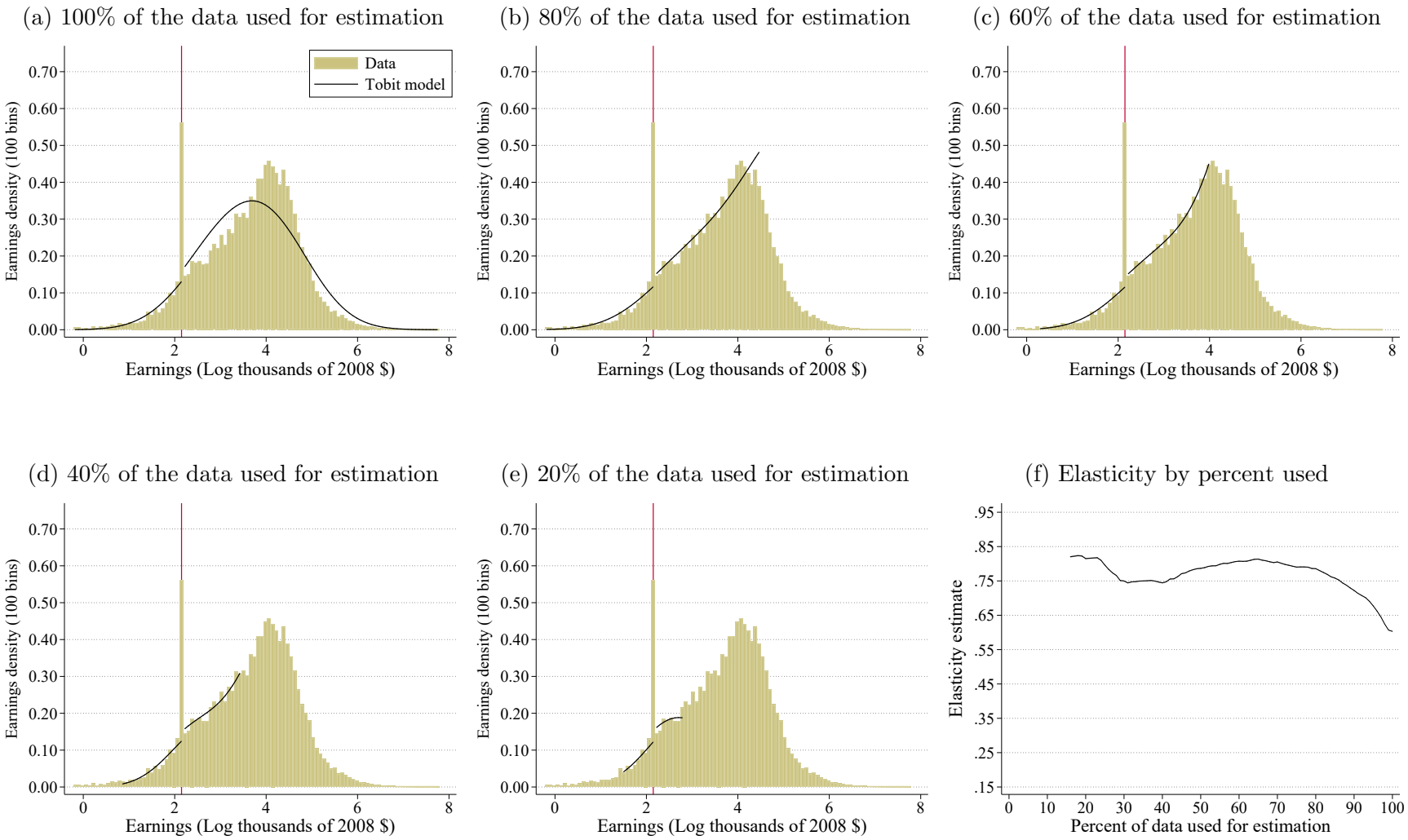

*Notes:* Panels a through d display partially identified sets for the elasticity for all filers with one child, and three other subsamples defined by employment and marital status. The y-axis has elasticity values between lower and upper bounds given various choices of $M$ on the x-axis, that is, the maximum slope magnitude of the PDF of the unobserved heterogeneity $n^*$ (Theorem 2). Each panel has two vertical lines. The line on the left corresponds to the smallest choice of $M$ for which the bounds are defined. At the smallest $M$, upper and lower bounds are equal to the elasticity estimate based on the trapezoidal approximation (Example 1). The vertical line on the right corresponds to the largest choice of $M$ for which the upper bound is finite. Higher slopes allow for the possibility of PDFs that are zero in the bunching window. As a result, we may have a finite bunching mass for any arbitrarily large elasticity.

Figure 4: Truncated Tobit - All Filers



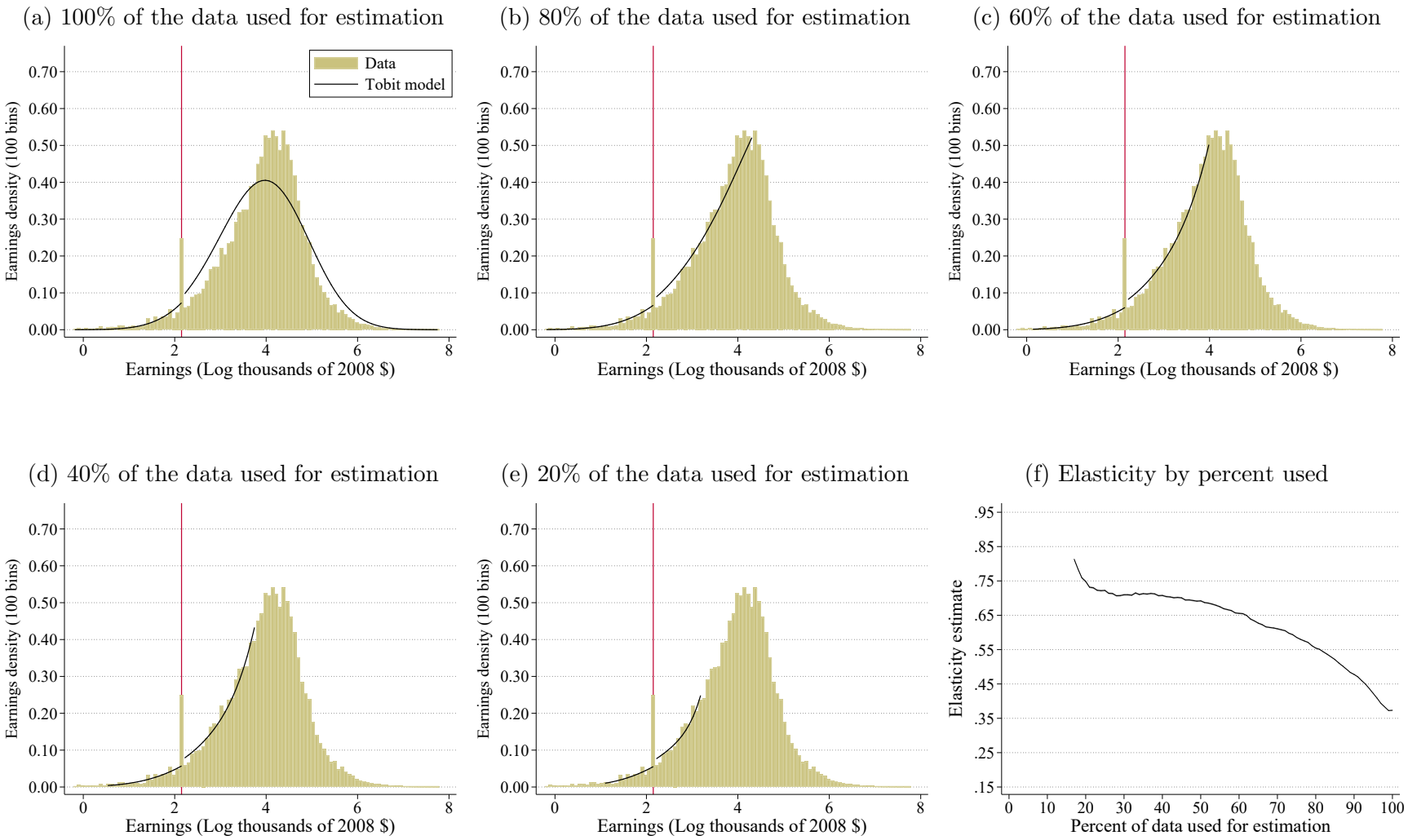

*Notes:* the figure displays best-fit Tobit distributions and elasticity estimates for various choices of a symmetric truncation window around the kink point. Estimation uses the following dummy variables as covariates: marital and employment status, year effects, types of deductions or social security benefits received, and whether the filer used a tax prep software. The set of included covariates is kept constant across different truncation windows. Panels a through e show the histogram of income for all filers (bars), along with the best-fit Tobit PDF for each truncation window (line). The best-fit PDF is constructed using the truncated Tobit likelihood averaged over covariate values in the sample. Panel f displays the Tobit elasticity estimate as a function of the percentage of data used in estimation.

Figure 5: Truncated Tobit - Self-employed Filers



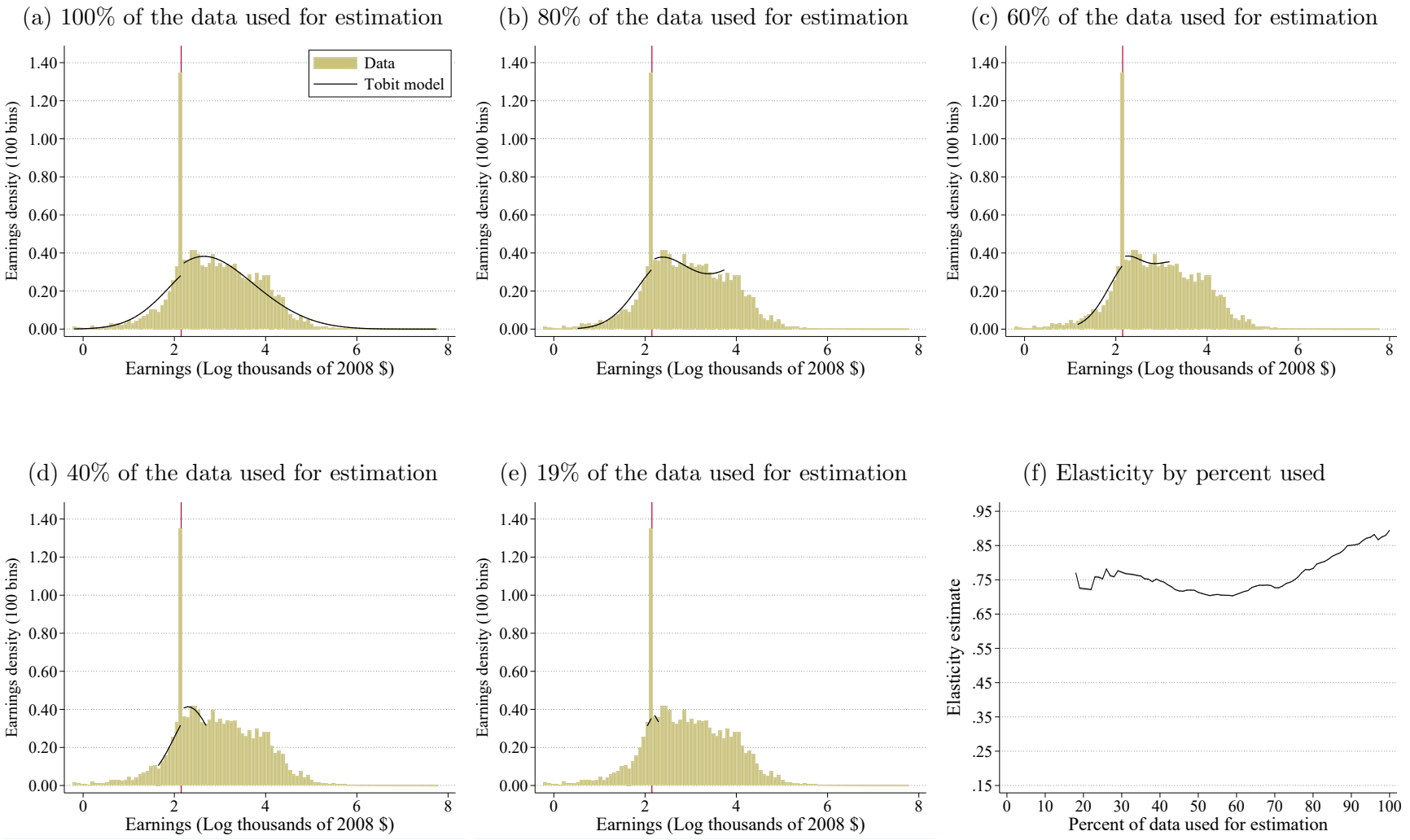

*Notes:* the figure displays best-fit Tobit distributions and elasticity estimates for various choices of a symmetric truncation window around the kink point. Estimation uses the following dummy variables as covariates: marital status, year effects, types of deductions or social security benefits received, and whether the filer used a tax prep software. The set of included covariates is kept constant across different truncation windows. Panels a through e show the histogram of income for self-employed filers (bars), along with the best-fit Tobit PDF for each truncation window (line). The best-fit PDF is constructed using the truncated Tobit likelihood averaged over covariate values in the sample. Panel f displays the Tobit elasticity estimate as a function of the percentage of data used in estimation.

Figure 6: Truncated Tobit - Self-employed and Married Filers




*Notes:* the figure displays best-fit Tobit distributions and elasticity estimates for various choices of a symmetric truncation window around the kink point. Estimation uses the following dummy variables as covariates: year effects, types of deductions or social security benefits received, and whether the filer used a tax prep software. The set of included covariates is kept constant across different truncation windows. Panels a through e show the histogram of income for self-employed and married filers (bars), along with the best-fit Tobit PDF for each truncation window (line). The best-fit PDF is constructed using the truncated Tobit likelihood averaged over covariate values in the sample. Panel f displays the Tobit elasticity estimate as a function of the percentage of data used in estimation.

Figure 7: Truncated Tobit - Self-employed and Not Married Filers




*Notes:* the figure displays best-fit Tobit distributions and elasticity estimates for various choices of a symmetric truncation window around the kink point. Estimation uses the following dummy variables as covariates: year effects, types of deductions or social security benefits received, and whether the filer used a tax prep software. The set of included covariates is kept constant across different truncation windows. Panels a through e show the histogram of income for self-employed and not married filers (bars), along with the best-fit Tobit PDF for each truncation window (line). The best-fit PDF is constructed using the truncated Tobit likelihood averaged over covariate values in the sample. Panel f displays the Tobit elasticity estimate as a function of the percentage of data used in estimation.

# A Appendix

## A.1 Identification with a Notch - Proof of Theorem 1

We present the proof of Theorem 1 in the more general case of multiple tax changes with at least one notch (Sections B.1 and B.2 in the supplemental appendix). Let $p \in \{1, \ldots, L\}$ be the index of the smallest notch $K_p$. As explained in the text, the presence of a notch may remove the next tax change $K_{p+1}$ from the solution to the utility maximization problem with multiple kinks and notches (Lemma B.1 in the supplemental appendix). Let $q \in \{p+1, \ldots, L\}$ be the index of the next tax change that appears in the solution. Following the proof of Lemma B.1, the distribution of $Y$ does not have any mass in the interval $(K_p\ ;\ Y_p^I]$ where $Y_p^I = N_p^I(1-t_{q-1})^\varepsilon$, and $N_p^I$ is defined as part of the solution in Equation B.4 in the supplemental appendix. The econometrician observes the value of $Y_p^I$, which is between $K_{q-1}$ and $K_q$. The goal is to solve for $\varepsilon$ using this information.

The proof of Lemma B.1 says $N_p^I$ satisfies the equation below.

$$N_p^I(1-t_{q-1})^{1+\varepsilon} + \varepsilon\left(N_p^I\right)^{-1/\varepsilon}(K_p)^{\frac{1+\varepsilon}{\varepsilon}} = (1+\varepsilon)\left[C_p - I_{q-1} + K_{q-1}(1-t_{q-1})\right]$$

Use the fact that $Y_p^I = N_p^I(1-t_{q-1})^\varepsilon$ and $\left(Y_p^I\right)^{-\frac{1}{\varepsilon}}(1-t_{q-1}) = \left(N_p^I\right)^{-\frac{1}{\varepsilon}}$ and substitute these in the equation above to get

$$\begin{aligned} Y_p^I(1-t_{q-1}) + \varepsilon\left(Y_p^I\right)^{-\frac{1}{\varepsilon}}(1-t_{q-1})(K_p)^{\frac{1+\varepsilon}{\varepsilon}} &= (1+\varepsilon)\left[C_p - I_{q-1} + K_{q-1}(1-t_{q-1})\right] \\ Y_p^I + \varepsilon K_p\left(\frac{K_p}{Y_p^I}\right)^{\frac{1}{\varepsilon}} &= (1+\varepsilon)\left(\frac{C_p - I_{q-1} + K_{q-1}(1-t_{q-1})}{1-t_{q-1}}\right) \end{aligned} \tag{A.1}$$

The elasticity $\varepsilon$ is identified if there exists an unique solution for $\varepsilon$ in Equation A.1 as function of $Y_p^I$, $K_p$, $C_p$, $I_{q-1}$, $t_{q-1}$. We know a solution exists, and we show it must be unique. Consider the left-hand and right-hand sides of (A.1) as functions of $\varepsilon$. The solution occurs at the value of $\varepsilon$ where both of these functions intersect. Uniqueness is equivalent to single-crossing of these functions.

The function on the right-hand side (RHS) of (A.1) has positive intercept equal to $[C_p - I_{q-1} + K_{q-1}(1-t_{q-1})]/(1-t_{q-1})$. The function on the left-hand side (LHS) has intercept equal to $Y_p^I$ because $\varepsilon K_p\left(\frac{K_p}{Y_p^I}\right)^{\frac{1}{\varepsilon}}$ converges to zero as $\varepsilon \downarrow 0$. The intercept of the LHS is strictly bigger than the intercept of the RHS:

$$\begin{aligned} Y_p^I &\gtrless \frac{C_p - I_{q-1} + K_{q-1}(1-t_{q-1})}{1-t_{q-1}} \\ I_{q-1} + (Y_p^I - K_{q-1})(1-t_{q-1}) &\gtrless C_p \\ C_p^I &\gtrless C_p \end{aligned}$$

where $C_p^I$ is the consumption value on the budget frontier when income is equal to $Y_p^I$ which is strictly greater than $C_p$. In fact, the consumer is indifferent between $(C_p^I, Y_p^I)$ and $(C_p, Y_p)$ where $Y_p^I > Y_p$. Since utility is strictly decreasing in $Y$ and increasing in $C$, we must have $C_p^I > C_p$. Therefore, $Y_p^I > \left[C_p - I_{q-1} + K_{q-1}(1-t_{q-1})\right]/(1-t_{q-1})$.

The function on the RHS of (A.1) has positive slope equal to $[C_p - I_{q-1} + K_{q-1}(1-t_{q-1})]/(1-t_{q-1})$. The function on the LHS has strictly positive derivative for any positive $\varepsilon$,

$$\frac{\partial}{\partial \varepsilon} LHS = K_p \left(\frac{K_p}{Y_p^I}\right)^{\frac{1}{\varepsilon}} \left[1 - \frac{1}{\varepsilon} \ln\left(\frac{K_p}{Y_p^I}\right)\right]$$

which is strictly positive because $K_p > 0$, $\left(\frac{K_p}{Y_p^I}\right)^{\frac{1}{\varepsilon}} \in (0,1)$, and $-\frac{1}{\varepsilon}\ln\left(\frac{K_p}{Y_p^I}\right) > 0$. The derivative is strictly increasing with $\varepsilon$,

$$\frac{\partial^2}{\partial \varepsilon^2} LHS = K_p \left(\frac{K_p}{Y_p^I}\right)^{\frac{1}{\varepsilon}} \frac{1}{\varepsilon^3} \left[\ln\left(\frac{K_p}{Y_p^I}\right)\right]^2$$

which is also strictly positive. The limit of $\frac{\partial}{\partial \varepsilon} LHS$ as $\varepsilon \to \infty$ is equal to $K_p$. Therefore, the slope of the LHS is positive, strictly increasing but always less than $K_p$. Next, we show that $K_p$ is strictly less than the constant slope of the RHS.

$$K_p \gtrless \frac{C_p - I_{q-1} + K_{q-1}(1-t_{q-1})}{1-t_{q-1}}$$
$$I_{q-1} + (K_p - K_{q-1})(1-t_{q-1}) \gtrless C_p.$$

The value $C_p^* = I_{q-1} + (K_p - K_{q-1})(1-t_{q-1})$ is what consumption would be if income were equal to $K_p$ and the budget segment between $K_{q-1}$ and $K_q$ were extrapolated back to $K_p$. We know that the indifference curve touches this budget segment at one point $(C_p^I, Y_p^I)$, and every other point on the extrapolated budget segment has strictly lower utility. We also know that $(C_p, K_p)$ is on such indifference curve, so that $(C_p, K_p)$ is strictly preferred to $(C_p^*, K_p)$. Therefore, $C_p^* < C_p$, and $K_p < [C_p - I_{q-1} + K_{q-1}(1-t_{q-1})]/(1-t_{q-1})$, and the slope of the function on the LHS of (A.1) is always less than the slope of the function of the RHS.

In summary, the intercept of the function on the LHS of (A.1) is greater than the intercept of the function on the RHS. Both functions are strictly increasing: the one on the RHS has constant slope, and the one on the LHS has increasing slope that is smaller than the slope of the RHS function. Therefore, the intersection of these two functions is unique. □

### A.2 Impossibility of Non-parametric Identification of the Elasticity - Proof of Lemma 1

Consider the case of one kink $k$, with one tax change, $s_0$ to $s_1$. It suffices to show that for every $\varepsilon > 0$, there exists $F_{n^*,\varepsilon} \in \mathcal{F}_{n^*}$ such that $F_y = T(k, s_0, s_1, F_{n^*,\varepsilon}, \varepsilon)$ for fixed $F_y$, $k$, $s_0$, and $s_1$. To show the existence of such an $F_{n^*,\varepsilon}$, fix arbitrary $\varepsilon > 0$ and then construct $F_{n^*,\varepsilon}$ as follows:

1. First, define a continuous function $\phi : [k - \varepsilon s_0, k - \varepsilon s_1] \to \mathbb{R}_{++}$ such that: (a) $\phi(k - \varepsilon s_0) = \lim_{u \uparrow k} f_y(u)$; (b) $\phi(k - \varepsilon s_1) = \lim_{u \downarrow k} f_y(u)$; and (c)

$\int \phi(u)\ du = F_y(k) - \lim_{u \uparrow k} F_y(u)$.

2. Second, compute the CDF $F_{n^*,\varepsilon}$ by integrating the following PDF:

$$f_{n^*,\varepsilon}(v) = \begin{cases} f_y(\varepsilon s_0 + v), & v \in (-\infty, k - \varepsilon s_0) \\ \phi(v), & v \in [k - \varepsilon s_0, k - \varepsilon s_1] \\ f_y(\varepsilon s_1 + v) & v \in (k - \varepsilon s_1, +\infty). \end{cases}$$

□

### A.3 Partial Identification with Non-parametric Restrictions - Proof of Theorem 2

First, let's fix $\varepsilon > 0$. We look at all possible PDFs in $\mathcal{F}_{n^*}$ and compute the maximum and minimum integrals over the interval $[\underline{n}, \overline{n}]$. The length of this interval is $\varepsilon(s_0 - s_1)$. Thus, without loss of generality, we restrict our attention to $f_{n^*}$ over the interval $[0, \varepsilon(s_0 - s_1)]$ such that:

(i) $f_{n^*}$ is continuous, and it connects the point $(0, f_y(k^-))$ to $(\varepsilon(s_0 - s_1), f_y(k^+))$ in the (x,y) plane;

(ii) the absolute value of the slope of $f_{n^*}$ is bounded by M.

First, start with $f_{n^*}$ being a line. The magnitude of the slope is $\frac{|f_y(k^+) - f_y(k^-)|}{\varepsilon(s_0 - s_1)}$. Suppose this magnitude is bigger than $M$. Then, any $f_{n^*}$ satisfying (i) will have a slope magnitude higher than $M$ at some point. Therefore, we need to look at $\varepsilon \geq \varepsilon_1$ where $\varepsilon_1 = \frac{|f_y(k^+) - f_y(k^-)|}{M(s_0 - s_1)}$.

For fixed $\varepsilon \geq \varepsilon_1$, the slope of the line will be less or equal to $M$. The maximum possible area is attained when the function has the shape of a hat with two line segments that attain the maximum slope. The first line segment starts at $(0, f_y(k^-))$ and has slope $+M$; the second line segment ends at $(\varepsilon(s_0 - s_1), f_y(k^+))$ and has slope $-M$. Call this function $\overline{f}_{n^*}$. These lines intersect at $x^*$ where

$$x^* = \frac{f_y(k^+) - f_y(k^-) + M\varepsilon(s_0 - s_1)}{2M}.$$

Note that $x^*$ is always such $0 \leq x^* \leq \varepsilon(s_0 - s_1)$ because $\varepsilon \geq \varepsilon_1$. Note that it is impossible to find another $f_{n^*}$ that satisfies (i), it is greater than $\overline{f}_{n^*}$, and that has slope magnitude less or equal than $M$. The maximum area is

$$\begin{aligned} \overline{A}(\varepsilon) &= \int_0^{\varepsilon(s_0 - s_1)} \overline{f}_{n^*}(v)\ dv \\ &= (1/4M)\left[M^2\varepsilon^2 s_0^2 - 2M^2\varepsilon^2 s_0 s_1 + M^2\varepsilon^2 s_1^2 + 2M\varepsilon f_y(k^-)\right. \\ &\left. s_0 - 2M\varepsilon f_y(k^-)s_1 + 2M\varepsilon f_y(k^+)s_0 - 2M\varepsilon f_y(k^+)s_1 - f_y(k^-)^2 + 2f_y(k^-)f_y(k^+) - f_y(k^+)^2)\right] \end{aligned}$$

The function $\overline{A}(\varepsilon)$ is strictly increasing with respect to $\varepsilon$ over $\varepsilon \geq \varepsilon_1$. In fact, the derivative is $((s_0 - s_1)(f_y(k^-) + f_y(k^+) + M\varepsilon(s_0 - s_1))/2$ which is strictly positive.

The minimum possible area is attained when the function has the shape of an inverted hat whose lines attain the maximum slope. That is, a combination of two line segments.

One that starts $(0, f_y(k^-))$ and has slope $-M$, and another that ends at $(\varepsilon(s_0 - s_1), f_y(k^+))$ and has slope $+M$. Differently the hat function, the intersection $(x^{**}, y^{**})$ of this inverted hat function may or may not be above the x-axis. That is, $y^{**}$ may be negative, but $f_{n^*}$ is always positive. In that case, we simply set the function to zero in the region where it would be negative. Call this function $\underline{f}_{n^*}$.

The intersection occurs at

$$x^{**} = \frac{f_y(k^-) - f_y(k^+) + M\varepsilon(s_0 - s_1)}{2M}.$$

Note that $x^{**}$ is always such $x^{**} \geq 0$ because $\varepsilon \geq \varepsilon_1$. The y-value of the intersection is

$$y^{**} = \frac{f_y(k^-) + f_y(k^+) - M\varepsilon(s_0 - s_1)}{2M}.$$

and this is positive as long as $\varepsilon \leq \varepsilon_2$ where $\varepsilon_2 = \frac{|f_y(k^+)+f_y(k^-)|}{M(s_0-s_1)}$. Note also that $\varepsilon_1 < \varepsilon_2$.

For $\varepsilon_1 \leq \varepsilon \leq \varepsilon_2$, the minimum area is

$$\begin{aligned}\underline{A}(\varepsilon) &= \int_0^{\varepsilon(s_0-s_1)} \underline{f}_{n^*}(v)\, dv \\ &= (-1/4M)\left[M^2\varepsilon^2 s_0^2 - 2M^2\varepsilon^2 s_0 s_1 + M^2\varepsilon^2 s_1^2 - 2M\varepsilon f_y(k^-)s_0\right. \\ &\left. +2M\varepsilon f_y(k^-)s_1 - 2M\varepsilon f_y(k^+)s_0 + 2M\varepsilon f_y(k^+)s_1 - f_y(k^-)^2 + 2f_y(k^-)f_y(k^+) - f_y(k^+)^2\right]\end{aligned}$$

The function $\underline{A}(\varepsilon)$ is strictly increasing with respect to $\varepsilon$ over $\varepsilon_1 \leq \varepsilon < \varepsilon_2$. In fact, the derivative is $((s_0 - s_1) * (f_y(k^-) + f_y(k^+) - M\varepsilon(s_0 - s_1)))/2$ which is strictly positive once we take into account $\varepsilon < \varepsilon_2$. The function $\underline{A}(\varepsilon)$ is constant with respect to $\varepsilon$ over $\varepsilon \geq \varepsilon_2$.

Therefore, we have characterized the maximum and minimum areas $\underline{A}(\varepsilon)$ and $\overline{A}(\varepsilon)$ for any given $\varepsilon$. These areas are undefined if $\varepsilon < \varepsilon_1$, they are equal if $\varepsilon = \varepsilon_1$, they are strictly increasing wrt $\varepsilon$ and $\underline{A}(\varepsilon) \leq \overline{A}(\varepsilon)$ for $\varepsilon \in (\varepsilon_1, \varepsilon_2)$. For $\varepsilon \geq \varepsilon_2$, $\overline{A}(\varepsilon)$ continues to grow wrt $\varepsilon$ but $\underline{A}(\varepsilon)$ stays constant at $\underline{A}(\varepsilon_2)$. The expression for $\underline{A}(\varepsilon_2)$ is $(f_y(k^-)^2 + f_y(k^+)^2)/2M$. Finally, we define the partially identified set.

**Case I:** If $B < \underline{A}(\varepsilon_1) = \overline{A}(\varepsilon_1)$, there does not exist any function $f_{n^*}$ consistent with any elasticity $\varepsilon$, so the set is empty. The expression for $\underline{A}(\varepsilon_1) = \overline{A}(\varepsilon_1)$ is $(|f_y(k^-) - f_y(k^+)|(f_y(k^-) + f_y(k^+)))/(2M)$.

**Case II:** Suppose $B \geq \underline{A}(\varepsilon_1)$ and $B < \underline{A}(\varepsilon_2)$. There is an interval range for $\varepsilon$ such that for any $\varepsilon$ in this interval there exists a function $f_{n^*}$ whose integral equals $B$. The minimum possible elasticity solves $\overline{A}(\underline{\varepsilon}) = B$. That gives

$$\underline{\varepsilon} = \frac{2\left[f_y(k^+)^2/2 + f_y(k^-)^2/2 + M\ B\right]^{1/2} - (f_y(k^+) + f_y(k^-))}{M(s_0 - s_1)}.$$

The maximum possible elasticity solves $\underline{A}(\overline{\varepsilon}) = B$. That gives

$$\overline{\varepsilon} = \frac{-2\left[f_y(k^+)^2/2 + f_y(k^-)^2/2 - M\ B\right]^{1/2} + (f_y(k^+) + f_y(k^-))}{M(s_0 - s_1)}$$

**Case III:** Suppose $B \geq \underline{A}(\varepsilon_2)$. It is still possible to find a minimum elasticity that solves $\overline{A}(\underline{\varepsilon}) = B$. However, for any elasticity $\varepsilon \geq \underline{\varepsilon}$ we have $\underline{A}(\varepsilon) \leq B$, so $\overline{\varepsilon}$ is infinity.
□

### A.4 Tobit Regression - Proof of Identification Lemma 2

By Assumption 17, there exists true values $(\varepsilon, \beta, \sigma)$ such that

$$\begin{aligned} B = F_y(k^+) - F_y(k^-) =& G_{n^*}(k - \varepsilon s_1; \beta, \sigma, F_X) - G_{n^*}(k - \varepsilon s_0; \beta, \sigma, F_X) \\ F_y(u) =& G_{n^*}(u - \varepsilon s_0; \beta, \sigma, F_X) \quad \text{for } \forall u < k \\ F_y(u) =& G_{n^*}(u - \varepsilon s_1; \beta, \sigma, F_X) \quad \text{for } \forall u > k, \end{aligned}$$

where $F_X$ is the true CDF of $X$. The MLE estimator is consistent for $(\bar{\varepsilon}, \bar{\beta}, \bar{\sigma})$, and we have that $G_y(y; \bar{\varepsilon}, \bar{\beta}, \bar{\sigma}, F_X) = F_y(y) \ \ \forall y$. Thus,

$$\begin{aligned} &G_{n^*}(k - \bar{\varepsilon} s_1; \bar{\beta}, \bar{\sigma}, F_X) - G_{n^*}(k - \bar{\varepsilon} s_0; \bar{\beta}, \bar{\sigma}, F_X) \\ &\qquad = G_{n^*}\left(k - \varepsilon s_1; \beta, \sigma, F_X\right) - G_{n^*}\left(k - \varepsilon s_0; \beta, \sigma, F_X\right) \\ &G_{n^*}(u - \bar{\varepsilon} s_0; \bar{\beta}, \bar{\sigma}, F_X) = G_{n^*}(u - \varepsilon s_0; \beta, \sigma, F_X) \quad \text{for } \forall u < k \\ &G_{n^*}(u - \bar{\varepsilon} s_1; \bar{\beta}, \bar{\sigma}, F_X) = G_{n^*}(u - \varepsilon s_1; \beta, \sigma, F_X) \quad \text{for } \forall u > k \end{aligned}$$

The parametric family created by $G_{n^*}(n; \beta, \sigma, F_X)$, with $F_X$ fixed at the truth, satisfies (11)-(13) by assumption. Therefore, the equations above solve uniquely with $(\bar{\varepsilon}, \bar{\beta}, \bar{\sigma}) = (\varepsilon, \beta, \sigma)$. □

### A.5 Censored Quantile Regression - Proof of Identification Lemma 3

Call $D = \mathbb{I}\{Q_\tau(y \mid X) \neq k\}$. Let $\beta(\tau) = [\beta_0(\tau), \beta_1(\tau), \ldots, \beta_d(\tau)]'$. Define $\tilde{\beta}(\tau) = [\beta_0(\tau) + \varepsilon s_0, \beta_1(\tau), \ldots, \beta_d(\tau), \varepsilon(s_1 - s_0)]'$. Multiplying Equation 19 by $D$ yields

$$DQ_\tau(y \mid X) = D\tilde{X}\tilde{\beta}(\tau).$$

Pre-multiplying it by $\tilde{X}'$ and taking expectations leads to

$$\begin{aligned} D\tilde{X}'Q_\tau(y \mid X) &= D\tilde{X}'\tilde{X}\tilde{\beta}(\tau) \\ \mathbb{E}\left[D\tilde{X}'Q_\tau(y \mid X)\right] &= \mathbb{E}\left[D\tilde{X}'\tilde{X}\tilde{\beta}(\tau)\right] \\ \tilde{\beta}(\tau) &= \mathbb{E}\left[D\tilde{X}'\tilde{X}\right]^{-1}\mathbb{E}\left[D\tilde{X}'Q_\tau(y \mid X)\right]. \end{aligned} \tag{A.2}$$

An infinite amount of data identifies the joint distribution of $(y, X)$. This identifies the function $Q_\tau(y \mid X = x)$ for every $x$ in the support of $X$, and the joint distribution of $(y, Q_\tau(y \mid X), X, \tilde{X}, D)$. Therefore, $\tilde{\beta}(\tau)$ is identified by Equation A.2. Finally, $\varepsilon = \tilde{\beta}_{d+1}(\tau)/(s_1 - s_0)$. □

# "BETTER BUNCHING, NICER NOTCHING"

Marinho Bertanha, Andrew McCallum, Nathan Seegert

## B Supplemental Appendix for Online Publication

### B.1 General Utility Maximization Problem with Multiple Kinks and Notches

To generalize the objective function in Equation 1, we update the budget set to have $J$ different tax regimes that change at cutoff points $0 < K_1 < \ldots < K_J$ on pre-tax labor income $Y$. Each tax regime has income tax $t_j$ such that $0 \leq t_0 \leq t_1 \leq \ldots \leq t_J < 1$. There are two possible tax changes. A change in tax rate is a kink. A lump-sum tax change is called a notch. Agent type $N^*$ maximizes utility $U(C, Y; N^*)$ as follows

$$\max_{C,Y} \quad C - \frac{N^*}{1+1/\varepsilon}\left(\frac{Y}{N^*}\right)^{1+\frac{1}{\varepsilon}} \tag{B.1}$$

$$s.t. \quad C = \sum_{j=0}^{J} \mathbb{I}\{K_j < Y \leq K_{j+1}\}\left[I_j + (1-t_j)(Y - K_j)\right], \tag{B.2}$$

where $K_0 = 0$, $K_{J+1} = \infty$, $\mathbb{I}\{\cdot\}$ is the indicator function, the solution is always on the budget frontier (Equation B.2), and we assume the agent resolves indifference by choosing the smallest value of $Y$. The elasticity of income $Y$ with respect to $(1-t_j)$ is equal to $\varepsilon$ when the solution is interior.

The budget frontier is continuous except when there is a notch. The limit of the budget frontier when $Y \downarrow K_j$ is equal to $I_j$, but equal to $I_{j-1} + (1-t_{j-1})(K_j - K_{j-1})$ when $Y \uparrow K_j$. The size of the jump discontinuity at a notch location $K_j$ is equal to $I_j - I_{j-1} - (1-t_{j-1})(K_j - K_{j-1})$. The intercepts $I_j$ and $I_{j-1}$ are assumed to be such that jump discontinuities at notches are negative.

### B.2 General Solution with Multiple Kinks and Notches

Lemma B.1 below provides a general solution to Problem B.1 with any combination of kinks and notches.

**Lemma B.1.** *Define $\mathcal{N} = \cup_{j=0}^{J}\left(K_j(1-t_j)^{-\varepsilon}; K_{j+1}(1-t_j)^{-\varepsilon}\right]$ as the set of $N^*$ values for which the indifference curves are tangent to the budget frontier. The function $Y^* : \mathcal{N} \to \mathbb{R}$, $Y^*(N^*) = \sum_{j=0}^{J} \mathbb{I}\{K_j(1-t_j)^{-\varepsilon} < N^* \leq K_{j+1}(1-t_j)^{-\varepsilon}\} N^*(1-t_j)^{\varepsilon}$, maps $N^*$ values to the $Y$ values corresponding to such tangency points. Similarly, $C^*(N^*)$ is consumption on the budget frontier (Equation B.2) when $Y = Y^*(N^*)$. Let $C_j$ be the value of $C^*(N^*)$ whenever $Y^*(N^*) = K_j$, $j = 1, \ldots, J$. For a notch-point $K_j$, define the value of $N_j^I$ to be that of the first indifference curve tangent to the budget frontier on the right of $Y = K_j$, such that the utility level is equal to the utility of the notch-point $K_j$,*

$$N_j^I = \min\left\{N^* \in \mathcal{N} \ : \ U(C_j, K_j) = U(C^*(N^*), Y^*(N^*))\right\}. \tag{B.3}$$

*In the case of a kink, the bunching interval is defined as* $[\underline{N}_j, \overline{N}_j]$*, where* $\underline{N}_j = K_j(1-t_{j-1})^{-\varepsilon}$*, and* $\overline{N}_j = K_j(1-t_j)^{-\varepsilon}$*. In the case of a notch, the expression for* $\underline{N}_j$ *equals that of the kink case, but* $\overline{N}_j$ *changes to* $N_j^I$*.*

*Note that the bunching intervals of two consecutive kinks do not overlap, that is,* $K_j(1-t_j)^{-\varepsilon} < K_{j+1}(1-t_j)^{-\varepsilon}$*. The same is not true for a kink or a notch* $K_{j+1}$ *that comes right after a notch* $K_j$*, because* $N_j^I$ *may be greater than* $K_{j+1}(1-t_j)^{-\varepsilon}$ *depending on* $\varepsilon$*. In this case,* $Y = K_{j+1}$ *does not appear in the solution. To account for that, construct a subsequence* $\{j_l\}_{l=1}^{L}$ *of* $\{1,\ldots,J\}$ *such that: (i)* $j_1 = 1$*; and (ii) for* $l \geq 2$*, set* $j_l$ *to be the smallest* $j$ *such that* $\underline{N}_j > \overline{N}_{j_{l-1}}$*. Then, the solution to the maximization problem in (B.1) is given by*

$$Y = \begin{cases} N^*(1-t_{j_1-1})^{\varepsilon} & , \text{ if } \ 0 < N^* < \underline{N}_{j_1} \\ K_{j_1} & , \text{ if } \ \underline{N}_{j_1} \leq N^* \leq \overline{N}_{j_1} \\ N^*(1-t_{j_2-1})^{\varepsilon} & , \text{ if } \ \overline{N}_{j_1} < N^* < \underline{N}_{j_2} \\ \vdots & \\ N^*(1-t_{j_L-1})^{\varepsilon} & , \text{ if } \ \overline{N}_{j_{L-1}} < N^* < \underline{N}_{j_L} \\ K_{j_L} & , \text{ if } \ \underline{N}_{j_L} \leq N^* \leq \overline{N}_{j_L} \\ N^*(1-t_J)^{\varepsilon} & , \text{ if } \ \overline{N}_{j_L} < N^* < \infty. \end{cases} \tag{B.4}$$

**Proof.** For every $N^* > 0$, there exists an unique solution on the budget frontier. If the consumer is indifferent between two solutions, we assume the consumer takes the solution with less $Y$. The proof is by induction over $\bar{J} = 0, 1, \ldots, J$. Denote the budget frontier $BF^{\bar{J}}$ by

$$C = \sum_{j=0}^{\bar{J}} \mathbb{I}\{\bar{K}_j < Y \leq \bar{K}_{j+1}\} \left[I_j + (1-t_j)\left(Y - \bar{K}_j\right)\right].$$

where $\bar{K}_j = K_j$ for $j = 0, 1, \ldots, \bar{J}$ and $\bar{K}_{\bar{J}+1} = \infty$.

As we change the budget frontier from $BF^{\bar{J}}$ to $BF^{\bar{J}+1}$, $K_{\bar{J}+1}$ takes a finite value strictly greater than $K_{\bar{J}}$, and $K_{\bar{J}+2}$ is set to $\infty$. If the solution to Problem B.1 with budget frontier $BF^{\bar{J}}$ is such that $Y < K_{\bar{J}+1} < \infty$, then this is also the solution to Problem B.1 with budget frontier $BF^{\bar{J}+1}$. In fact, points on $BF^{\bar{J}}$ dominate points on $BF^{\bar{J}+1}$, and they coincide for $Y < K_{\bar{J}+1}$.

**Part I:** $\bar{J} = 0$*, solve Problem B.1 with budget* $BF^0$*.*

This is a standard consumer maximization problem where the optimal choice for $Y$ occurs at the point the indifference curve is tangent to $BF^0$. Therefore, for $N^* > 0$, $Y = N^*(1-t_0)^{\varepsilon}$.

**Part II:** $\bar{J} = 1$*, solve Problem B.1 with budget* $BF^1$*.*

The budget frontier $BF^1$ has two segments $BF_0^1$ for $0 < Y \leq K_1$, and $BF_1^1$ for $K_1 < Y$. If $N^* < K_1(1-t_0)^{-\varepsilon}$, then the solution of Part I, $Y = N^*(1-t_0)^{\varepsilon} < K_1$, is also the solution in Part II. It remains to find the solution for $N^* \geq K_1(1-t_0)^{-\varepsilon}$. These solutions must lie on $BF^1$ for $Y \geq K_1$ because they strictly dominate those that lie to the left of $K_1$.

*Case I : Suppose* $K_1$ *is a kink.*

Assume $N^*$ is such that $K_1(1-t_0)^{-\varepsilon} \leq N^* \leq K_1(1-t_1)^{-\varepsilon}$. If the solution is interior to $BF_1^1$, then it must be at a tangent point in which case $Y = N^*(1-t_1)^{\varepsilon}$. However,

$Y = N^*(1-t_1)^{\varepsilon} \leq K_1$, a contradiction because this $Y$ falls outside of the interior of $BF_1^1$. Therefore, if $N^*$ is such that $\underline{N}_1 = K_1(1-t_0)^{-\varepsilon} \leq N^* \leq K_1(1-t_1)^{-\varepsilon} = \overline{N}_1$, then the solution is $Y = K_1$. Suppose $N^* > \overline{N}_1$. Then, the solution is in the interior of $BF_1^1$, and it is equal to $Y = N^*(1-t_1)^{\varepsilon}$.

*Case II : Suppose $K_1$ is a notch.*

There is a jump-down discontinuity in $BF^1$ at $K_1$, and $BF^1$ is continuous from the left. Consider the point $(C, Y) = (C_1, K_1)$ on $BF_0^1$. Define $Y^D$ to be the value of $Y$ such that the corresponding $C$ value on $BF_1^1$ is equal to $C_1$. The jump-down discontinuity creates a strictly dominated region on $BF_1^1$ because the utility of $(C_1, K_1)$ is strictly greater than the utility of any solution with $Y \in (K_1, Y^D)$. Indifference between $K_1$ and $Y^D$ is resolved towards $K_1$ by assumption. Therefore, we cannot have solutions to Problem B.1 with budget $BF^1$ such that $Y \in (K_1, Y^D]$.

Define the point $\widetilde{N}_1^I$ as being the solution of Problem B.1 with budget $BF^1$ (instead of $BF$). This is the smallest $N^*$ for which Problem B.1 with budget $BF_1^1$ has solution with utility equal to $U(C_1, K_1)$.

First, a solution $\widetilde{N}_1^I$ exists. To see that, note that for small $N^*$, the tangent point $Y = N^*(1-t_1)^{\varepsilon}$ along $BF_1^1$ falls in the dominated region $Y \in (K_1, Y^D]$, and the utility is less than $U(C_1, K_1)$; on the other hand, the utility at this tangent point increases with $N^*$, and it eventually equals $U(C_1, K_1)$. The solution is such that $\widetilde{N}_1^I \geq Y^D(1-t_1)^{-\varepsilon} > K_1(1-t_1)^{-\varepsilon}$.

Second, the solution $\widetilde{N}_1^I$ is unique. To see that, solve for $N^*$ in the equation below.

$$U(C_1, K_1) = U\left(I_1 + N^*(1-t_1)^{\varepsilon+1} - K_1(1-t_1)\ ,\ N^*(1-t_1)^{\varepsilon}\right)$$

where $C = I_1 + N^*(1-t_1)^{\varepsilon+1} - K_1(1-t_1)$ is consumption on $BF_1^1$ when $Y = N^*(1-t_1)^{\varepsilon}$. Evaluating and rearranging the equality gives

$$N^*(1-t_1)^{1+\varepsilon} + \varepsilon(N^*)^{-1/\varepsilon}(K_1)^{\frac{1+\varepsilon}{\varepsilon}} = (1+\varepsilon)\left[C_1 - I_1 + K_1(1-t_1)\right]$$

The solution is unique because the derivative of the right-hand side is strictly positive given $N^* > K_1(1-t_1)^{-\varepsilon}$. Note that $\widetilde{N}_1^I$ is the unique solution to Problem B.1 when the budget is $BF^1$.

Call $\widetilde{Y}_1^I = \widetilde{N}_1^I(1-t_1)^{\varepsilon}$. Suppose there is a solution to Problem B.1 with budget $BF^1$ such that $Y^D < Y \leq \widetilde{Y}_1^I$. This solution is interior to budget $BF_1^1$, so we must have $Y = N^*(1-t_1)^{\varepsilon}$ for some $N^*$. But such a solution cannot be a solution to Problem B.1 with budget $BF^1$ because $Y \leq \widetilde{Y}_1^I$ and so dominated by $(C_1, K_1)$. Therefore, we cannot have solutions to Problem B.1 with budget $BF^1$ such that $Y \in (K_1, \widetilde{Y}^I]$.

It remains to characterize the solution when $N^*$ is such that $K_1(1-t_0)^{-\varepsilon} \leq N^*$. If $N^*$ is such that $\underline{N}_1 = K_1(1-t_0)^{-\varepsilon} \leq N^* \leq \widetilde{Y}^I(1-t_1)^{-\varepsilon} = \overline{N}_1$, the solution cannot be in the interior of $BF_0^1$ since $Y = N^*(1-t_0)^{\varepsilon} \geq K_1$; it cannot be in $(K_1, \widetilde{Y}^I]$ either. Assume it is in the interior of $BF_1^1$ with $Y > \widetilde{Y}_I$. Since it is interior, it satisfies $Y = N^*(1-t_1)^{\varepsilon}$, but $N^* \leq \widetilde{Y}^I(1-t_1)^{-\varepsilon}$ which makes $Y \leq \widetilde{Y}^I$, a contradiction. Therefore, the solution to Problem B.1 with budget $BF^1$ when $N^* \in [\underline{N}_1; \overline{N}_1]$ is $Y = K_1$. Finally, suppose $N^* > \overline{N}_1$. Then, the solution is in the interior of $BF_1^1$, and it is equal to $Y = N^*(1-t_1)^{\varepsilon}$.

**Part III:** *Assume the solution of Problem B.1 with budget $BF^{\bar{J}}$ and $1 \leq \bar{J} < J$ is as*

*in Equation B.4 with $\bar{J}$. Show that (B.4) with $\bar{J}+1$ solves Problem B.1 with budget $BF^{\bar{J}+1}$.*

Consider Problem B.1 with budget $BF^{\bar{J}}$ and solution B.4 with $L$ being $\bar{L}$. If $N^*$ is such that $Y < K_{\bar{J}+1} < \infty$, then $Y$ also solves Problem B.1 with budget $BF^{\bar{J}+1}$. Therefore, the solution to Problem B.1 with budget $BF^{\bar{J}+1}$ or budget $BF^{\bar{J}}$ coincide for those values of $N^*$. Note also that, if $K_j$ is a notch and $j < j_{\bar{L}}$, then the value of $\overline{N}_j$ (defined in (B.3)) does not change when the budget changes from $BF^{\bar{J}}$ to $BF^{\bar{J}+1}$. If $K_{j_{\bar{L}}}$ is a notch, then the value $\overline{N}_{j_{\bar{L}}}$ may change (case IV below). In what follows, consider the last two budget segments of $BF^{\bar{J}+1}$: $BF_{\bar{J}}^{\bar{J}+1}$ and $BF_{\bar{J}+1}^{\bar{J}+1}$.

*Case I : $K_{j_{\bar{L}}}$ is a kink, $K_{\bar{J}+1}$ is a kink*

In this case, $j_{\bar{L}+1} = \bar{J}+1$ because $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} > \overline{N}_{j_{\bar{L}}}$, so that $\bar{J}+1$ is the smallest $j$ such that $\underline{N}_j > \overline{N}_{j_{\bar{L}}}$. It is also true that $j_{\bar{L}} = \bar{J}$. To see that, note that consecutive intervals $[\underline{N}_j, \overline{N}_j]$ never overlap for kinks because $\overline{N}_j = K_j(1-t_j)^{-\varepsilon} < K_{j+1}(1-t_j)^{-\varepsilon} = \underline{N}_{j+1}$. The upper limit of a kink interval $j$ is strictly smaller than the lower limit of a notch interval $j+1$. However, the upper limit of a notch interval $j$ may be bigger than the lower limit of the next interval $j+1$. Suppose $j_{\bar{L}} = \bar{J}$ were not true, that is, $j_{\bar{L}} < \bar{J}$. Then, any $j$ such that $j_{\bar{L}} < j \leq \bar{J}$ is not in the subsequence $\{j_l\}$ because $K_{j_{\bar{L}}}$ is a notch, and its interval overlaps with the $j$ interval. But this is a contradiction with $K_{j_{\bar{L}}}$ being a kink point.

If $N^* < K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$, then the solution B.4 with budget $BF^{\bar{J}}$ is $Y < K_{\bar{J}+1}$, and $Y$ also solves Problem B.1 with budget $BF^{\bar{J}+1}$ for that same value of $N^*$. It remains to characterize the solution when $N^* \geq K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$

Assume $N^*$ is such that $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} \leq N^* \leq K_{\bar{J}+1}(1-t_{\bar{J}+1})^{-\varepsilon} = \overline{N}_{\bar{J}+1}$. As seen in Part II, Case I, the solution cannot be interior to $BF_{\bar{J}+1}^{\bar{J}+1}$. The solution must be at $K_{\bar{J}+1}$. Assume $N^* > \overline{N}_{\bar{J}+1}$. Then, the solution is interior to $BF_{\bar{J}+1}^{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$.

*Case II : $K_{j_{\bar{L}}}$ is a kink, $K_{\bar{J}+1}$ is a notch*

As seen in Part III, Case I, $j_{\bar{L}} = \bar{J}$. We also have $j_{\bar{L}+1} = \bar{J}+1$ because the $j$ interval $[\underline{N}_j, \overline{N}_j]$ of a kink does not overlap with the $j+1$ interval of a notch.

If $N^* < K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$, then the solution B.4 with budget $BF^{\bar{J}}$ is $Y < K_{\bar{J}+1}$, and $Y$ also solves Problem B.1 with budget $BF^{\bar{J}+1}$ for that same value of $N^*$. It remains to characterize the solution when $N^* \geq K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$

Assume $N^*$ is such that $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} \leq N^* \leq \overline{N}_{\bar{J}+1}$, where $\overline{N}_{\bar{J}+1}$ is the solution of Problem B.3 when the budget is $BF^{\bar{J}+1}$. As seen in Part II, Case II, the solution $Y$ cannot be in $(K_{\bar{J}+1}, \overline{N}_{\bar{J}+1}(1-t_{\bar{J}+1})^{\varepsilon}]$ or in the interior of $BF_{\bar{J}+1}^{\bar{J}+1}$. Therefore, the solution is $Y = K_{\bar{J}+1}$. Assume $N^* > \overline{N}_{\bar{J}+1}$. Then, the solution is interior to $BF_{\bar{J}+1}^{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$.

*Case III : $K_{j_{\bar{L}}}$ is a notch, $\overline{N}_{j_{\bar{L}}} < \underline{N}_{\bar{J}+1}$*

For the notch $K_{j_{\bar{L}}}$, the solution $\overline{N}_{j_{\bar{L}}}$ to Problem B.3 when the budget is $BF^{\bar{J}}$ does not change when the budget becomes $BF^{\bar{J}+1}$ precisely because $\overline{N}_{j_{\bar{L}}} < \underline{N}_{\bar{J}+1}$. In this case, $j_{\bar{L}+1} = \bar{J}+1$. For $N^*$ such that $\overline{N}_{j_{\bar{L}}} < N^* < K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$, the solution B.4 with budget

$BF^{\bar{J}}$ is $Y < K_{\bar{J}+1}$, and $Y$ also solves Problem B.1 with budget $BF^{\bar{J}+1}$ for that same value of $N^*$. It remains to characterize the solution when $N^* \geq K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon}$.

Assume $K_{\bar{J}+1}$ is a kink, and that $N^*$ is such that $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} \leq N^* \leq K_{\bar{J}+1}(1-t_{\bar{J}+1})^{-\varepsilon} = \overline{N}_{\bar{J}+1}$. As seen in Part II, Case I, the solution cannot be interior to $BF^{\bar{J}+1}_{\bar{J}+1}$. The solution must be at $K_{\bar{J}+1}$. Assume $N^* > \overline{N}_{\bar{J}+1}$. Then, the solution is interior to $BF^{\bar{J}+1}_{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$.

Assume $K_{\bar{J}+1}$ is a notch, and that $N^*$ is such that $\underline{N}_{\bar{J}+1} = K_{\bar{J}+1}(1-t_{\bar{J}})^{-\varepsilon} \leq N^* \leq \overline{N}_{\bar{J}+1}$, where $\overline{N}_{\bar{J}+1}$ is the solution of Problem B.3 when the budget is $BF^{\bar{J}+1}$. As seen in Part II, Case II, the solution $Y$ cannot be in $(K_{\bar{J}+1}, \overline{N}_{\bar{J}+1}(1-t_{\bar{J}+1})^{\varepsilon}]$ or in the interior of $BF^{\bar{J}+1}_{\bar{J}+1}$. Therefore, the solution is $Y = K_{\bar{J}+1}$. Assume $N^* > \overline{N}_{\bar{J}+1}$. Then, the solution is interior to $BF^{\bar{J}+1}_{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$.

*Case IV : $K_{j_{\bar{L}}}$ is a notch, $\overline{N}_{j_{\bar{L}}} \geq \underline{N}_{\bar{J}+1}$*

The indifference value for $Y$ at $\overline{N}_{j_{\bar{L}}}$ is $Y^I_{j_{\bar{L}}} = \overline{N}_{j_{\bar{L}}}(1-t_{\bar{J}})^{\varepsilon} \geq \underline{N}_{\bar{J}+1}(1-t_{\bar{J}})^{\varepsilon} = K_{\bar{J}+1}$. If $\overline{N}_{j_{\bar{L}}} = \underline{N}_{\bar{J}+1}$, the solution to Problem B.3 when the budget is $BF^{\bar{J}}$ remains unchanged when the budget becomes $BF^{\bar{J}+1}$. If $\overline{N}_{j_{\bar{L}}} > \underline{N}_{\bar{J}+1}$, then $Y^I_{j_{\bar{L}}} > K_{\bar{J}+1}$, and the solution to Problem B.3 when the budget is $BF^{\bar{J}}$ changes when the budget becomes $BF^{\bar{J}+1}$. The value of $\overline{N}_{j_{\bar{L}}}$ increases such that the new indifference point satisfies $Y^I_{j_{\bar{L}}} = \overline{N}_{j_{\bar{L}}}(1-t_{\bar{J}+1})^{\varepsilon}$.

There does not exist a $j$ such that $\underline{N}_j > \overline{N}_{j_{\bar{L}}}$ because $K_{\bar{J}+1}$ is the last tax-change point available and $\underline{N}_{\bar{J}+1} \leq \overline{N}_{j_{\bar{L}}}$. Therefore, when constructing the solution of Problem B.1 with budget $BF^{\bar{J}+1}$, the last term in the subsequence $\{j_l\}$ remains $j_{\bar{L}}$.

The point $K_{j_{\bar{L}}}$ is a notch, so Part II, Case II says that for $N^*$ such that $\underline{N}_{j_{\bar{L}}} = K_{j_{\bar{L}}}(1-t_{j_{\bar{L}}-1})^{-\varepsilon} \leq N^* \leq \overline{N}_{j_{\bar{L}}}$, the solution $Y$ cannot be in $(K_{j_{\bar{L}}}, \overline{N}_{j_{\bar{L}}}(1-t_{\bar{J}+1})^{\varepsilon}]$ or in the interior of $BF^{\bar{J}+1}_{\bar{J}+1}$. Therefore, the solution is $Y = K_{j_{\bar{L}}}$. Assume $N^* > \overline{N}_{j_{\bar{L}}}$. Then, the solution is interior to $BF^{\bar{J}+1}_{\bar{J}+1}$, and it equals to $Y = N^*(1-t_{\bar{J}+1})^{\varepsilon}$. □

### B.3 Friction Errors and Failure of the "Polynomial Strategy"

This section presents a counterexample that illustrates the failure of a common identification strategy used in applied work to estimate the elasticity using kinks. For a review, see Kleven (2016).

First, we set the parameters of the model. The true values are: $\varepsilon = 1.5$ (elasticity); $t_0 = .2$ and $t_1 = 0.3$ (before and after tax rates); kink-point $k = 0$. The bunching interval is $[\underline{n}, \overline{n}] = [0.335, .535]$. The distribution of the ability variable is assumed uniform, $n^* \sim U[-.565; 1.435]$; that is, the support is centered at 0.435 and has length equal to 2. The probability of bunching, or bunching mass $B$, is equal to 10% in this example. The friction error $e$ is also assumed uniformly distributed $e \sim U[-0.5; 0.5]$. The value of labor income observed by the researcher is $\widetilde{y} = y + e$, where $y$ is a function of $n^*$, $\varepsilon$, $t_0$, and $t_1$, as described in Equation 4.

In the counterfactual scenario of no tax change, we have $\underline{n} = \overline{n}$, and the counterfactual income with friction error is denoted $\widetilde{y}_0$. The counterfactual income without friction error is $y_0$. Figure B.1a depicts the PDF of $\widetilde{y}$ and $\widetilde{y}_0$.

A common identification strategy used in applied work is to fit a polynomial to the PDF of $\widetilde{y}$ excluding observations in the neighborhood of the kink $k = 0$, that corresponds to the support of the measurement error (i.e. $[-0.5; 0.5]$). The estimated bunching mass is the area between the PDF of $\widetilde{y}$ and the polynomial fit extrapolated to the excluded neighborhood around the kink. Figure B.1b illustrates the procedure. The figure shows that such strategy fails to identify the true bunching mass, even when the polynomial fit of 7th order is perfect, and we assume the researcher knows the support of $e$.

The last part of the estimation strategy uses the extrapolated polynomial to predict the counterfactual PDF of $y_0$. Following Equation 6, identification of $\varepsilon$ requires the counterfactual PDF of $y_0$, without measurement error. Figure B.1c shows that the polynomial strategy fails to retrieve the PDF of $y_0$. The PDF predicted by the polynomial regression does not integrate to one, and thus it is not a PDF. If we divide the polynomial-based PDF in Figures B.1b and B.1c by its integral, the PDF shifts up in the graphs. The re-normalized PDF still misses the true $f_{y_0}$, and the underestimation of $B$ is larger than before.

The polynomial strategy fails for two reasons:

1. The PDF of $\widetilde{y}$ is not simply the PDF of $y$ plus the PDF of $e$ (Figure B.1a), but the convolution between the two PDFs. While $y_0$ and $e$ have uniform distributions, with a flat PDF, their convolution does not have a flat PDF. As a result, extrapolating the polynomial to find the bunching mass and to predict the PDF of $y_0$ is misleading;

2. The counterfactual distribution required for identification of the elasticity is the PDF of $y_0$, and not the PDF of $\widetilde{y}_0$ (Equation 6). Moreover, even if friction errors were not a problem, it is not possible to use the distribution of $y$ to back out the distribution of $y_0$ for values of $y_0$ inside $[k, k + (s_0 - s_1)\varepsilon]$. The shape of the distribution of $y_0$ is unidentified when $n^*$ falls in the bunching interval (Figure 1).

### B.4 Parametric Gaussian Family Identifies the Elasticity

We demonstrate how to verify conditions (11) - (13) in the parametric Gaussian case. Suppose the distribution of $n^*$ follows a normal distribution with unknown mean $\mu$ and unknown variance $\sigma^2$, such that $F_{n^*}(n) = G_{n^*}(n; \mu, \sigma^2) = \Phi\left(\frac{n-\mu}{\sigma}\right)$ where $\Phi$ denotes the standard normal CDF.

Take $(k, s_0, s_1, \varepsilon, \mu, \sigma^2)$ arbitrary. The goal is to show that $\bar{\varepsilon} = \varepsilon$, $\bar{\mu} = \mu$, and $\bar{\sigma}^2 = \sigma^2$ are the only solutions to the equalities below:

$$\Phi\left(\frac{k - \varepsilon s_1 - \mu}{\sigma}\right) - \Phi\left(\frac{k - \varepsilon s_0 - \mu}{\sigma}\right) = \Phi\left(\frac{k - \bar{\varepsilon} s_1 - \bar{\mu}}{\bar{\sigma}}\right) - \Phi\left(\frac{k - \bar{\varepsilon} s_0 - \bar{\mu}}{\bar{\sigma}}\right) \tag{B.5}$$

$$\Phi\left(\frac{u - \varepsilon s_0 - \mu}{\sigma}\right) = \Phi\left(\frac{u - \bar{\varepsilon} s_0 - \bar{\mu}}{\bar{\sigma}}\right) \quad \text{for } \forall u < k \tag{B.6}$$

$$\Phi\left(\frac{u - \varepsilon s_1 - \mu}{\sigma}\right) = \Phi\left(\frac{u - \bar{\varepsilon} s_1 - \bar{\mu}}{\bar{\sigma}}\right) \quad \text{for } \forall u > k \tag{B.7}$$

Take (B.6), and apply $\Phi^{-1}(\cdot)$ to both sides.

$$\frac{u - \varepsilon s_0 - \mu}{\sigma} = \frac{u - \bar{\varepsilon} s_0 - \bar{\mu}}{\bar{\sigma}}, \quad \forall u < k.$$

These are two lines that must have the same slope, $1/\sigma = 1/\bar{\sigma}$, and the same intercept $(\varepsilon s_0 + \mu)/\sigma = (\bar{\varepsilon} s_0 + \bar{\mu})/\bar{\sigma}$. These imply that $\bar{\sigma} = \sigma$, and $\bar{\varepsilon} s_0 + \bar{\mu} = \varepsilon s_0 + \mu$.

Similarly, (B.7) implies that $\bar{\varepsilon} s_1 + \bar{\mu} = \varepsilon s_1 + \mu$. Subtracting this last equation from the previous one gives $\bar{\varepsilon}(s_1 - s_0) = \varepsilon(s_0 - s_1)$, which yields $\bar{\varepsilon} = \varepsilon$. Finally, $\varepsilon s_1 + \bar{\mu} = \varepsilon s_1 + \mu$ gives $\bar{\mu} = \mu$.

□

### B.5 Implementation of Censored Quantile Regressions

The optimization problem in Equation 21 is computationally difficult. For the left (or right) censored case, Chernozhukov and Hong (2002) proposed a fast and practical estimator that consists of three steps. First, you fit a flexible Probit model that explains the probability of no censoring; then, you select observations whose values of $X$ lead to a predicted probability of no censoring that is greater than $1 - \tau$. Second, you fit a quantile regression of $y$ on $X$ using the selected observations in the first step; then, you select observations whose values of $X$ lead to a predicted quantile that is greater than $k$. Third, repeat the second step using the observations selected at the end of the second step. Chernozhukov and Hong (2002) demonstrate consistency and asymptotic normality of their three-step estimator. Moreover, they show that the standard errors computed by the quantile regression in the third step are valid.

Our case of middle censoring requires a straightforward modification of the method proposed by Chernozhukov and Hong (2002). Inspired by their algorithm, we propose the following implementation steps.

1. Create dummies $\delta_i^- = \mathbb{I}\{y_i < k\}$ (not censored, left of $k$) and $\delta_i^+ = \mathbb{I}\{y_i > k\}$ (not censored, right of $k$). Fit two Probit models to estimate $\mathbb{P}[\delta_i^+|X_i] = \Phi(X_i g^+)$ and $\mathbb{P}[\delta_i^-|X_i] = \Phi(X_i g^-)$, where $\Phi$ denotes the cdf of a standard normal distribution, and $g^\pm$ are vectors of parameters. You may use powers and interactions of $X_i$ to make this stage as flexible as possible. Select two subsamples as follows. Compute the 10th quantile of the empirical distribution of $\Phi(X_i\hat{g}^+) - (1-\tau)$ conditional on $\Phi(X_i\hat{g}^+) > 1 - \tau$. Let $\kappa_0^+(\tau)$ be the 10th quantile of that distribution. The first subsample is $J_0^+(\tau) = \{i : \Phi(X_i\hat{g}^+) > 1 - \tau + \kappa_0^+(\tau)\}$. The second subsample is $J_0^-(\tau) = \{i : \Phi(X_i\hat{g}^-) > \tau + \kappa_0^-(\tau)\}$, where $\kappa_0^-(\tau)$ is the 10th quantile of the empirical distribution of $\Phi(X_i\hat{g}^-) - \tau$ conditional on $\Phi(X_i\hat{g}^-) > \tau$. Create a dummy $W_i^0 = \mathbb{I}\{i \in J_0^+(\tau)\}$.

2. Fit the quantile regression model $Q_\tau(y_i|X_i, W_i^0) = X_i b(\tau) + W_i^0 \delta(\tau)$ using observations in $J_0^-(\tau) \cup J_0^+(\tau)$. Use the estimates of this quantile regression, that is $\hat{b}^0(\tau)$ and $\hat{\delta}^0(\tau)$, to create two subsamples as follows. The first subsample is $J_1^+(\tau) = \{i : X_i\hat{b}^0(\tau) + \hat{\delta}^0(\tau) > k + \kappa_1^+(\tau)\}$, where $\kappa_1^+(\tau)$ is the 3rd quantile of the empirical distribution of $X_i\hat{b}^0(\tau) + \hat{\delta}^0(\tau) - k$ conditional on $X_i\hat{b}^0(\tau) + \hat{\delta}^0(\tau) > k$. The second subsample is $J_1^-(\tau) = \{i : X_i\hat{b}^0(\tau) < k + \kappa_1^-(\tau)\}$, where $\kappa_1^-(\tau)$ is the 97th

quantile of the empirical distribution of $X_i\hat{b}^0(\tau) - k$ conditional on $X_i\hat{b}^0(\tau) < k$. Create a dummy $W_i^1 = \mathbb{I}\{i \in J_1^+(\tau)\}$.

3. Fit the quantile regression model $Q_\tau(y_i|X_i, W_i^1) = X_i b(\tau) + W_i^1\delta(\tau)$ using observations in $J_1^-(\tau) \cup J_1^+(\tau)$ to obtain estimates $\hat{b}^1(\tau)$ and $\hat{\delta}^1(\tau)$. The elasticity estimator is $\hat{\varepsilon} = \hat{\delta}^1(\tau)/(s_1 - s_0)$.

### B.6 Estimates with the Filtering Method of Saez (2010)

In this section, we recompute the estimates of Table 1 using a different filtering method. Specifically, we employ the procedure used by Saez (2010) to obtain the bunching mass and the side limits of the distribution of income without friction error $Y$. The procedure implicitly defines a way to estimate the unobserved distribution of $Y$ given the observed distribution of income with friction error $\tilde{Y}$. We refer the reader to Figure 2 by Saez (2010).

The first step is to construct a histogram-based estimate of the PDF $f_{\tilde{Y}}$, and then average $f_{\tilde{Y}}$ for $\tilde{Y} \in [K - 2\delta, K - \delta] \cup [K + \delta, K + 2\delta]$, where $K = 8,580$ is the kink point, and $\delta = 1,500$ defines the excluded region. Call that average $\bar{f}$. The bunching mass is estimated by the area between two curves, $f_{\tilde{Y}}$ and $\bar{f}$. The continuous portion of $f_Y$ equals $f_{\tilde{Y}}$, except for the excluded region $[K - \delta, K + \delta]$, where $f_Y$ equals $\bar{f}$. We obtain the CDFs $F_Y$ and $F_{\tilde{Y}}$ from their PDF estimates. Finally, we rely on $Y = F_Y\left(F_{\tilde{Y}}^{-1}(\tilde{Y})\right)$ to transform $\tilde{Y}$ into $Y$.

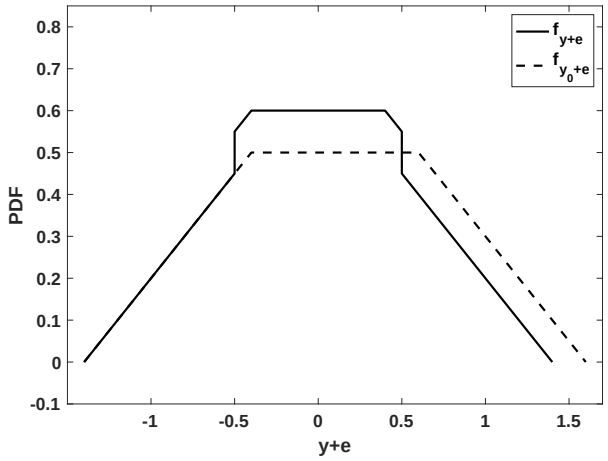
Table B.1: Estimates Using U.S. Tax Returns 1995--2004

| Statistical Model | (1) Saez (2010) | (2) Theorem 2 Bounds M = 0.5 | (3) Theorem 2 Bounds M = 1 | (4) Tobit Full Sample | (5) Tobit Trunc. 75% | (6) Tobit Trunc. 50% | (7) Tobit Trunc. 25% | (8) Sample details | |
|---|---|---|---|---|---|---|---|---|---|
| *All* | | | | | | | | Obs. | 189.1m |
| Elasticity ($\varepsilon$) | 0.235 | $[0.225, 0.249]$ | $[0.210, 0.282]$ | 0.124 | 0.177 | 0.182 | 0.199 | Avg. | \$54.1k |
| | (0.0311) | | | (0.0002) | (0.0001) | (0.0001) | (0.0002) | Std. | \$131.1k |
| *Self-employed* | | | | | | | | Obs. | 33.5m |
| Elasticity ($\varepsilon$) | 0.933 | $[0.765, 1.298]$ | $[0.663, \infty]$ | 0.617 | 0.809 | 0.805 | 0.822 | Avg. | \$61.8k |
| | (0.0759) | | | (0.0006) | (0.0008) | (0.0008) | (0.0009) | Std. | \$168.2k |
| *Self-employed, married* | | | | | | | | Obs. | 24.0m |
| Elasticity ($\varepsilon$) | 0.391 | $[0.328, 0.382]$ | $[0.285, \infty]$ | 0.187 | 0.286 | 0.330 | 0.331 | Avg. | \$75.0k |
| | (0.0823) | | | (0.0004) | (0.0007) | (0.0008) | (0.0008) | Std. | \$185.6k |
| *Self-employed, not married* | | | | | | | | Obs. | 9.6m |
| Elasticity ($\varepsilon$) | 1.260 | $[1.130, 1.508]$ | $[1.008, \infty]$ | 1.145 | 0.991 | 1.003 | $1.270^{\dagger}$ | Avg. | \$28.7k |
| | (0.1193) | | | (0.0012) | (0.0011) | (0.0012) | (0.0018) | Std. | \$106.3k |

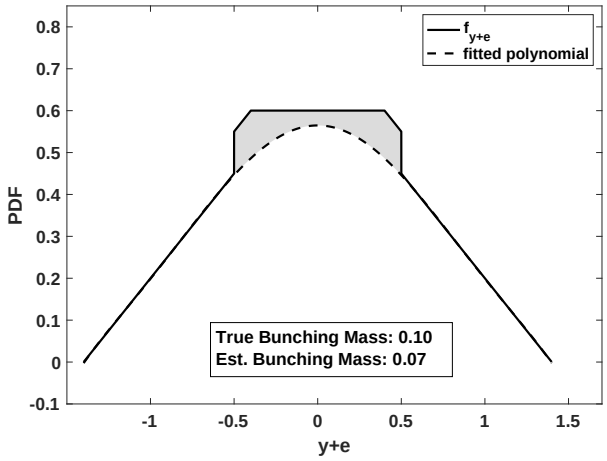
*Notes:* The table shows estimates of the elasticity for four different subsamples of the IRS data, and using three different approaches discussed in the paper. The first approach (column 1) uses the trapezoidal approximation to point-identify the elasticity (Example 1). Estimates and standard errors were computed using the publicly available code by Saez (2010) at the website of the American Economic Journal, Economic Policy. The second approach (columns 2 and 3) computes partially identified sets for the elasticity (Theorem 2), using non-parametric estimates of the side limits of $f_y$ at the kink, and the bunching mass. Side limits were estimated using the method of Cattaneo et al. (2019). The estimate for the bunching mass equals the sample proportion of $y$ observations that equals the kink point (see discussion in Section B.6 on friction errors). Upper and lower bounds are calculated for two choices of M, that is, the maximum slope of the PDF of the unobserved heterogeneity $n^*$. Column 4 has Tobit MLE estimates of the elasticity that utilizes the full sample of data, along with robust standard errors. Columns 5 through 7 report truncated Tobit MLE estimates. As we move from column 5 to column 7, we restrict the estimation sample to shrinking symmetric windows around the kink that utilizes 75% to 25% of the data. The set of covariates that enters the Tobit estimation is kept constant across different truncation windows. It includes dummy variables such as marital and employment status, year effects, types of deductions or social security benefits received, and whether the filer used a tax prep software. †There are too few observations for the maximum likelihood estimator to converge when using 25% of the sample. This estimate uses 27% instead.

Figure B.1: Counterexample where "Polynomial Strategy" Fails

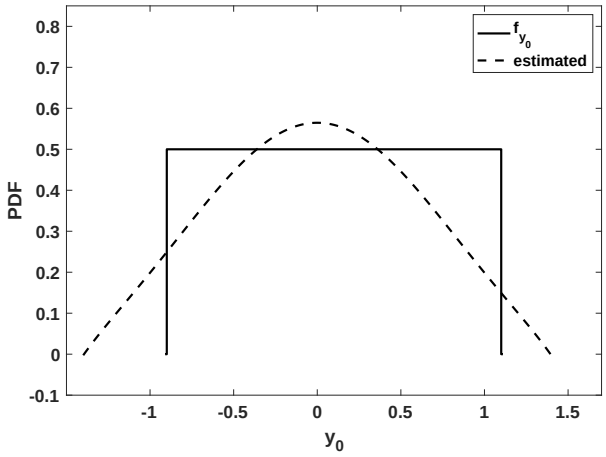
(a) Distribution of Income with Friction Error




(b) Estimation of Bunching Mass




(c) Counterfactual Distribution of Income without Friction Error




*Notes:* The population model of this example has $\varepsilon = 1.5$, $t_0 = .2$, and $t_1 = 0.3$ at kink $k = 0$. The distribution of ability is assumed uniform, $n^* \sim U[-.565; 1.435]$. The probability of bunching is equal to 10%, and the distribution of the friction error is $e \sim U[-0.5; 0.5]$. The researcher observes $\widetilde{y} = y + e$, where $y$ is a function of $n^*$, $\varepsilon$, $t_0$, and $t_1$, as described in Equation 4. Figure B.1a displays the PDF of $\widetilde{y}$ and $\widetilde{y}_0$. Figure B.1b displays the fitted 7th-order polynomial to the PDF of $\widetilde{y}$ using observations in $(-\infty, -0.5) \cup (0.5, \infty)$. The bunching mass is estimated by the integral of the difference between $f_{\widetilde{y}}$ and the fitted polynomial, inside the excluded region. The polynomial strategy understimates the true bunching mass, and does not retrieve the PDF of $y_0$ (Figure B.1c).