

BETTER BUNCHING, NICER NOTCHING

Marinho Bertanha*

Andrew H. McCallum[†]

Nathan Seegert [‡]

First draft: August 5, 2017

This draft: June 12, 2023

Abstract

This paper studies the bunching identification strategy for an elasticity parameter that summarizes agents' responses to changes in slope (kink) or intercept (notch) of a schedule of incentives. We show that current bunching methods may be very sensitive to implicit assumptions in the literature about unobserved individual heterogeneity. We overcome this sensitivity concern with new non- and semi-parametric estimators. Our estimators allow researchers to show how bunching elasticities depend on different identifying assumptions and when elasticities are robust to them. We follow the literature and derive our methods in the context of the iso-elastic utility model and an income tax schedule that creates a piece-wise linear budget constraint. We demonstrate bunching behavior provides robust estimates for self-employed and not-married taxpayers in the context of the U.S. Earned Income Tax Credit. In contrast, estimates for self-employed and married taxpayers depend on specific identifying assumptions, which highlight the value of our approach. 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.

[†]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. Piecewise-linear schedules provide identifying variation in incentives to estimate these responses. 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 nonparametric 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 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.”¹ The literature began with an iso-elastic utility model and an income tax schedule that creates a piece-wise linear constraint. We demonstrate our methods in this context and note that our results extend to other contexts with piece-wise linear constraints.

¹We 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. The general problem and solution are in Sections B.1 and B.2 of the supplement. For ease of exposition, we keep the problem with one kink or notch in the main text (Section 2).

We highlight two insights about identification with kinks and notches assuming a nonparametric 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.

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 nonparametric 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 that the unconditional distribution implied by the Tobit model matches that; 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. In addition, covariates generally allow for more precise estimates.

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 among self-employed married individuals is 0.48—that is, a one percent increase in the marginal tax rate results in a reduction in reported income of at least 0.48 percent. This estimate contrasts with the estimate of 0.77 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 0.48 is 48%, while the optimal tax rate for an elasticity of 0.77 is 37%—a difference of 11 percentage points.

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 elasticity of 0.55 versus a trapezoidal estimate of 0.77 for self-employed and married 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; Goff, 2022), 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, 2020), minimum wage (Jales, 2018; Cengiz, Dube, Lindner, and Zipperer, 2019), charitable giving (Hungerman and Ottoni-Wilhelm, 2021), and air-pollution data manipulation (Ghanem, Shen, and Zhang, 2019), among others. Kleven (2016) and Bertanha, Caetano, Jales, and Seegert (2023) provide a recent review of the many applications and branches of the bunching literature; Jales and Yu (2017) relates bunching to regression discontinuity design (RDD).²

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 in the iso-elastic quasi-linear utility model—and an earlier paper provides intuition for the impossibility result (Blomquist, Kumar, Liang, and Newey, 2015). 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 partial identification result independently of theirs. Our partial identification approach has three valuable features that make it novel: 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 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. The Stata package is presented by Bertanha, McCallum, Payne, and Seegert (2022) and available for download from the website of the authors or the Statistical Software Components (SSC) online

²Variation 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). 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.

repository.³

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. For ease of exposition, we focus on budget sets with one kink or one notch in the main text and generalize to multiple kinks and notches in the supplemental appendix, Sections B.1 and B.2

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,

$$\begin{aligned} \max_{C, Y} \quad & C - \frac{N^*}{1 + 1/\varepsilon} \left(\frac{Y}{N^*} \right)^{1 + \frac{1}{\varepsilon}} \\ \text{s.t.} \quad & \end{aligned} \tag{1}$$

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

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 ε 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, $\Delta = 0$,

³Type `ssc install bunching` in Stata to install the package.

and the budget frontier is continuous; otherwise, in the case of a notch, it has a jump discontinuity of size Δ 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):

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

where the expressions for the thresholds \underline{N} and \overline{N} are given below. Section B.2 in the supplement presents the proof of this result and the more general solution in the case of multiple kinks and notches. We provide a graphical analysis of these solutions in Section B.8 in the supplement.

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. Section B.2 in the supplemental appendix 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 $y = \log(Y)$, $n^* = \log(N^*)$, $\underline{n} = \log(\underline{N})$, $\overline{n} = \log(\overline{N})$, $k = \log(K)$, $s_0 = \log(1 - t_0)$, and $s_1 = \log(1 - t_1)$.

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

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 ε and the unobserved distribution n^* ,

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

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

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

$$B = \int_k^{k+\Delta y} f_{y_0}(u) du = F_{y_0}(k + \Delta y) - F_{y_0}(k), \quad (6)$$

where $\Delta y = \varepsilon(s_0 - s_1)$. Figure 1, Panels a and b, illustrate 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 ε for a given distribution of y_0 . In other words, 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.

In reality, instead of y , researchers typically observe the distribution of $\tilde{y} = y + e$, where e is a random variable accounting for optimization and friction errors. We focus on the identification problem associated with identifying the counterfactual distribution of y_0 in the absence of errors e . In work in progress, Cattaneo et al. (2018) show how to solve this problem with a deconvolution method, which is necessary to recover the distribution in the absence of these errors. Common strategies such as the polynomial strategy, first proposed by Chetty et al. (2011), fails to solve this issue. We provide a simple counterexample in the supplemental appendix Section B.3 where the polynomial strategy fails to recover the true distribution of y . A practical solution that works under limited assumptions is provided by Bertanha et al. (2022) and implemented in Section 5.

3 Identification

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,

identification in case of a kink is impossible, unless the researcher imposes restrictions on the distribution of n^* . The general solution to Problem 1 with multiple kinks and notches is found in Section B.2 of the supplement. That section discusses interesting insights to the identification of the elasticity that arise in that context.

3.1 Identification from Gaps in the Distribution

We show that identification in the case of a notch is possible using the additional information from the gap in the distribution that is not used in previous studies. Specifically, the gap in the distribution of Y 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 ε that satisfies the indifference condition in 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*

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

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

This and all other proofs are given in Appendix A.⁴ Theorem 1 and its proof consider the case of a notch from a lump-sum tax, $\Delta > 0$ in Equation 2. A minor change to that proof shows the elasticity is also nonparametrically identified in the case of a notch from a lump-sum subsidy, $\Delta < 0$. Another case where the elasticity is nonparametrically identified is that of a concave kink, that is, a kink arising from a decrease in tax rates, $t_0 > t_1$. This is formally demonstrated in Section A.5 of the appendix and is the only part of the paper where we refer to a kink caused by $t_0 > t_1$. When the rest of the paper refers to a kink, we mean a kink generated by $t_0 < t_1$.

The identification for these three cases (notch with $\Delta > 0$, notch with $\Delta < 0$, and kink with $t_0 > t_1$) does not depend on the bunching mass, but instead on the size of the region with missing mass (size of the gaps in the distribution). In all the cases we consider, the distribution of income cannot have optimization frictions, which we discuss in detail in Section B.3 of the supplement.

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 between \underline{n}

⁴Similar arguments were given by Blomquist and Newey (2018) around the same time this result appeared in an earlier version of our paper, Bertanha et al. (2018).

and \bar{n} . If all we know about f_{n^*} is that it is continuous with full support and that its integral over $[\underline{n}, \bar{n}]$ 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 was first shown in Theorem 1 by Blomquist and Newey (2017), although the idea was first presented by Blomquist et al. (2015). We discuss the impossibility result in this section and present our novel identification strategies in the next sections.

Figure 1 provides intuition behind this impossibility result. 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. In fact, for any value of the elasticity $\varepsilon > 0$, there exists a continuous PDF of n^* that justifies the observed distribution f_y according to Equation 4. The model assumptions imply no restrictions on ε over $(0, \infty)$. Theorem 1 by Blomquist and Newey (2017) clarifies that current bunching methods are either implicitly restricting \mathcal{F}_{n^*} or simply inconsistent for the true elasticity. Section B.4 in the supplement details the implicit restrictions on \mathcal{F}_{n^*} made by the original bunching methods, namely, the affine PDF assumption of Saez (2010) and the uniform PDF assumption of Chetty et al. (2011). A direct consequence of the impossibility result is that restrictions on \mathcal{F}_{n^*} are untestable. One may argue that the affine or uniform assumption is a good approximation to any potentially non-linear density f_{n^*} if the bunching interval $[k - \varepsilon s_0, k - \varepsilon s_1]$ 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.

We conclude this section by stating a sufficient condition on how flexible \mathcal{F}_{n^*} may be for point identification of ε to be possible.

Assumption 1. *Let \mathcal{F}_{n^*} be a set of all possible CDFs of n^* that are continuously differentiable. For any $F_{n^*} \in \mathcal{F}_{n^*}$, consider the possible values of $e \geq 0$ and $G_{n^*} \in \mathcal{F}_{n^*}$ that satisfy the following system of equations:*

$$G_{n^*}(u - \varepsilon s_0) = F_{n^*}(u - \varepsilon s_0) \text{ for } \forall u < k, \quad (8)$$

$$G_{n^*}(u - \varepsilon s_1) = F_{n^*}(u - \varepsilon s_1) \text{ for } \forall u \geq k. \quad (9)$$

The set of distributions \mathcal{F}_{n^} is restricted to be such that the only values of $e \geq 0$ and $G_{n^*} \in \mathcal{F}_{n^*}$ that satisfy Equations 8–9 are $e = \varepsilon$ and $G_{n^*} = F_{n^*}$.*

Intuitively, Assumption 1 restricts \mathcal{F}_{n^*} in such a way that knowledge of the tails of an unknown CDF in \mathcal{F}_{n^*} is enough to reconstruct that entire CDF; and no other CDF in \mathcal{F}_{n^*} has the same tails. The assumption is easily verified in parametric families, e.g., $\mathcal{F}_{n^*} = \{G_{n^*}(n; \theta) \mid \theta \in \Theta\}$, for some set of parameters $\Theta \subseteq \mathbb{R}^p$. In Section B.4, we show how to verify Assumption 1 using an example of the Gaussian family.

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 nonparametric shape restriction that bounds the slope magnitude of f_{n^*} , which leads to partial identification of ε . 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 Nonparametric Bounds

Our partial identification approach relies on restricting the class \mathcal{F}_{n^*} to distributions with PDFs, f_{n^*} , that are Lipschitz continuous with constant $M \in (0, \infty)$. In other words, the slope magnitude of any f_{n^*} in this class is bounded by M . The following theorem gives the partially identified set for ε 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*

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

where \emptyset is the empty set, and

$$\begin{aligned} \underline{\varepsilon} &= \frac{2 [f_y(k^+)^2/2 + f_y(k^-)^2/2 + M B]^{1/2} - (f_y(k^+) + f_y(k^-))}{M(s_0 - s_1)} \\ \bar{\varepsilon} &= \frac{-2 [f_y(k^+)^2/2 + f_y(k^-)^2/2 - M B]^{1/2} + (f_y(k^+) + f_y(k^-))}{M(s_0 - s_1)}. \end{aligned}$$

Figures 1c and 1d provide the intuition behind the derivation of the bounds in Υ . For a fixed value of ε , the length of the interval $[\underline{n}, \bar{n}]$ 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 $[\underline{n}, \bar{n}]$. We repeat this exercise for every value of ε to get a range of possible areas associated with each ε . Given the probability of bunching B is the area under the true f_{n^*} over $[\underline{n}, \bar{n}]$, the partially identified set has all values of ε 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 $[\underline{n}, \bar{n}]$.

The uniform approximation made by one of the original estimators (Example 2 in Section B.4 of the supplement) says that f_{n^*} has zero slope inside the bunching interval, that is, $M = 0$. The trapezoidal approximation (Example 1 in that same section) 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 $\underline{\varepsilon} = \bar{\varepsilon}$ and point-identifies ε . Thus the exercise of computing bounds necessarily involves assumptions weaker than the uniform and trapezoidal approximations.

Smoothness assumptions such as slope restrictions are now common in the partial identification literature. For example, Kim, Kwon, Kwon, and Lee (2018) study partial identification of average treatment effects under smoothness conditions on the treatment response function; Rambachan and Roth (2023) derive bounds on treatment effects in difference-in-difference designs under smoothness conditions on the class of deviations of the parallel trend assumption. A common issue in this literature is that the researcher must choose the smoothness assumption. In the case of Theorem 2, the researcher must specify the value of M . The impossibility of identifying the elasticity without structure on \mathcal{F}_{n^*} (Section 3.2) implies that it is impossible to identify, and thus estimate, the value of M .

We recommend researchers to conduct a sensitivity analysis by plotting the bounds in Theorem 2 as a function of M , for a range of values of M that is considered reasonable given the empirical context. From above, we know that $M = m_0$ yields point identification. A useful reference for 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 the slope of f_{n^*} inside the bunching interval is never bigger than outside, then $M = m_1$. Thus, a rule of thumb for the range of values of M is to start at m_0 and go up to at least m_1 . One may also estimate the worst case PDFs of Figures 1c and 1d to evaluate the visual effect of M on the shape of the latent distribution. Sensitivity analysis of this kind are not new in the partial identification literature. The idea is to report what can be learned under a sequence of progressively weaker assumptions. For examples, we refer the reader to Kim et al. (2018) and Rambachan and Roth (2023).

Theorem 2 is important to quantify the magnitude of the impossibility problem of identification using kinks. 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).⁵

We developed Theorem 2 independently of Blomquist and Newey (2017), who were the first to present a partial identification result for ε . While we assume the PDF has bounded slope, Blomquist and Newey (2017) partially identify the elasticity by assuming the PDF of heterogeneity

⁵It 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 nonparametric estimation. The value of M represents a choice of functional form assumption, while in nonparametric estimation, you typically choose the tuning parameter to achieve desirable properties of the estimator for a given functional form assumption.

is monotone. Our approach has three valuable properties that make it novel. 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.

We end this subsection with the case of a budget set with several kinks k_j , $j = 1, \dots, J$, 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, $\bar{N}_j = K_j(1 - t_{j-1})^{-\varepsilon} < K_j(1 - t_j)^{-\varepsilon} = \underline{N}_j$. Multiple kinks do not necessarily point-identify ε , because the distribution of n^* may be very different across different bunching intervals.

Multiple kinks do help with the identification of ε , 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 ε , and that the bounds of Theorem 2 vary in length as B_j , $f_y(k_j^\pm)$, s_j vary across cutoffs $j = 1, \dots, J$. 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 , $j = 1, \dots, J$. Then the elasticity $\varepsilon \in \bigcap_{j=1}^J \Upsilon_j$, where Υ_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 nonparametric 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 εs_0 , the intercept to the right of the kink is εs_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 10 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 ε 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 one intercept variable and d slope variables, $\mathbb{E}[X'X]$ has full rank, but otherwise 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 certain 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 model assumes normality of the unobserved distribution conditional on covariates, we demonstrate that the Tobit estimator remains consistent under our semi-parametric class of normal mixtures—as long as the unconditional distribution for y implied by the Tobit model matches the true distribution of y . 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. Our main motivation to study the robustness of Tobit to lack of normality comes from practical reasons: Tobit is extremely popular and easy to implement due to its likelihood function being globally concave and software being 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 assumption is formally stated in Assumption 2 below. In the meantime, we construct the Tobit estimator by simply assuming that $F_{n^*|X}(n, x) = \Phi\left(\frac{n - x\beta}{\sigma}\right)$ for $\beta \in \mathbb{R}^{(d+1) \times 1}$ and $\sigma > 0$, where $F_{n^*|X}$ denotes the true CDF of n^* conditional on X , and $\Phi(\cdot)$ is the CDF of a standard normal distribution. Again, X contains one intercept variable and d slope variables, and $\mathbb{E}[X'X]$ has full rank. Our goal is to first relate bunching to Tobit, which is the most popular censoring model. Conditional normality is assumed for ease of exposition in defining the mid-censored Tobit estimator; it will be relaxed in Assumption 2 below.

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 $s_0 > s_1$. Using Equation 4, we see that y follows a mid-censored Tobit model:

$$y = \begin{cases} y_0^* & , \text{ if } y_1^* < y_0^* < k \\ k & , \text{ if } y_1^* \leq k \leq y_0^* \\ y_1^* & , \text{ if } k < y_1^* < y_0^* \end{cases} = \min\{y_0^*; \max\{k; y_1^*\}\}. \quad (10)$$

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, we estimate a binary outcome for bunching and not bunching individuals including covariates. In the second step, we regress income of not bunching individuals on covariates and the equivalents of the inverse Mills ratio. It is useful to relate the mid-censored Tobit model to two classic Tobit models (left- and right-censored). 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 β_1, \dots, β_d , where $\beta = (\beta_0, \beta_1, \dots, \beta_d)$. Similarly, y_1 follows a left-censored Tobit with intercept $\varepsilon s_1 + \beta_0$, and slope coefficients β_1, \dots, β_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 $(s_1 - s_0)$. Although readily implementable in most statistical packages, this estimation strategy does not constrain the slope coefficients and variances to be equal on both sides of the kink, 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) , $i = 1, \dots, n$, be an iid sample of observations. The maximum likelihood estimator (MLE) for the true parameters $(\varepsilon, \beta, \sigma)$ is constructed by maximizing the log-likelihood function with respect to (e, b, s) given the sample data,

$$\begin{aligned}
L(y_1, \dots, y_n | X_1, \dots, X_n; e, b, s) &= \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{y_i < k\} \log \left[\frac{1}{s} \phi \left(\frac{y_i - \varepsilon s_0 - X_i b}{s} \right) \right] \\
&\quad + \mathbb{I}\{y_i = k\} \log \left[\Phi \left(\frac{k - \varepsilon s_1 - X_i b}{s} \right) - \Phi \left(\frac{k - \varepsilon s_0 - X_i b}{s} \right) \right] \\
&\quad + \mathbb{I}\{y_i > k\} \log \left[\frac{1}{s} \phi \left(\frac{y_i - \varepsilon s_1 - X_i b}{s} \right) \right] \\
&\equiv \frac{1}{n} \sum_{i=1}^n \ell_i(e, b, s).
\end{aligned} \tag{11}$$

Regardless of what the true distribution $F_{n^*|X}$ is, the MLE based on (11) is consistent for the parameters that maximize the population average of the log-likelihood function, that is, $(e^*, b^*, s^*) = \arg \max_{e, b, s} \mathbb{E}[\ell_i(e, b, s)]$, where the solution is unique (see, e.g., Hayashi (2000), Section 8.3.). We say the elasticity is identified by a mid-censored Tobit when $e^* = \varepsilon$. This occurs when $F_{n^*|X}(n, x) = \Phi \left(\frac{n-x\beta}{\sigma} \right)$, but we would like to relax the normality assumption and still have $e^* = \varepsilon$. We pursue this exercise in the next paragraphs. The Tobit estimator is extremely practical to implement, and we believe it is worth investigating a set of assumptions on $F_{n^*|X}$ that are weaker than normality and yet sufficient for consistency. We start by describing these conditions on the set of possible distributions $\mathcal{F}_{n^*|X}$.

Assumption 2. *The set of true conditional CDFs of n^* given X is denoted $\mathcal{F}_{n^*|X}$. We assume*

$\mathcal{F}_{n^*|X}$ is such that:

- (i) for every $F_{n^*|X} \in \mathcal{F}_{n^*|X}$, the unconditional CDF satisfies $F_{n^*}(n) = \mathbb{E}[F_{n^*|X}(n, X)] = \mathbb{E}\left[\Phi\left(\frac{n-Xb}{s}\right)\right]$ for some $b \in \mathbb{R}^{(d+1) \times 1}$ and $s > 0$. The set of all unconditional CDFs $\mathcal{F}_{n^*} = \{F_{n^*}(n) = \mathbb{E}[F_{n^*|X}(n, X)] \text{ for } F_{n^*|X} \in \mathcal{F}_{n^*|X}\}$ satisfies Assumption 1;
- (ii) recall that $\underline{n} = k - \varepsilon s_0$, $\bar{n} = k - \varepsilon s_1$, $B = \mathbb{E}[F_{n^*|X}(\bar{n}, X) - F_{n^*|X}(\underline{n}, X)]$; define $D = \mathbb{I}\{n^* \geq \bar{n}\}$ and $B_N(X, \delta, \theta, s) = \Phi((\bar{n} - \delta - X\theta)/s) - \Phi((\underline{n} - X\theta)/s)$; for every $F_{n^*|X} \in \mathcal{F}_{n^*|X}$, we have $\mathbb{E}[F_{n^*|X}(n, X)] = \mathbb{E}\left[\Phi\left(\frac{n-X\beta}{\sigma}\right)\right]$ for β and σ that satisfy

$$(0, \beta, \sigma) = \arg \min_{\delta, \theta, s} \frac{1-B}{2} \left\{ \log(s^2) + \frac{1}{s^2} \mathbb{E}\left[(n^* - D\delta - X\theta)^2 | n^* \notin [\underline{n}, \bar{n}]\right] \right\} - B \mathbb{E}[\log(B_N(X, \delta, \theta, s)) | n^* \in [\underline{n}, \bar{n}]]. \quad (12)$$

Assumption 2(i) says that if we take any conditional distribution of n^* given X from the set $\mathcal{F}_{n^*|X}$ and integrate it over X we obtain a marginal distribution of n^* which is a mixture of normals. The mixture is averaged over the distribution of X and the more variation in covariates one has, the richer the set \mathcal{F}_{n^*} is. Assumption 2(i) also assumes the set of mixtures of normals satisfies Assumption 1, so it is sufficient for point identification of the elasticity. To see that, note that each mixture of normals in \mathcal{F}_{n^*} is characterized by $b \in \mathbb{R}^{(d+1) \times 1}$ and $s > 0$, that is, $G_{n^*}^{b,s}(n) = \mathbb{E}\left[\Phi\left(\frac{n-Xb}{s}\right)\right]$. That along with a value of the elasticity $e \geq 0$ imply a distribution of y ,

$$G_y^{e,b,s}(y) = \begin{cases} G_{n^*}^{b,s}(y - es_0) & , \text{ if } y < k, \\ G_{n^*}^{b,s}(y - es_1) & , \text{ if } y \geq k. \end{cases}$$

If we search for a mixture of normals in \mathcal{F}_{n^*} and an elasticity value that matches the observed distribution of y , we will solve uniquely for the true elasticity and a distribution $G_{n^*} \in \mathcal{F}_{n^*}$ (Assumption 1). Uniqueness of G_{n^*} does not necessarily mean that G_{n^*} is indexed by a unique set of parameter values (b, s) . The assumption still allows for cases where different values of parameters (b, s) lead to the same CDF, G_{n^*} . For example, let $X = [1, W]$ and (n^*, W) be distributed as a standard bivariate normal with correlation ρ . The CDF of n^* conditional on X equals $\Phi(n - \rho W)$. The unconditional CDF of n^* is equal to $\Phi(n)$, which is the same as $\mathbb{E}[\Phi(n - \rho W)]$ or $\mathbb{E}[\Phi(n)]$. Thus, $\Phi(n)$ belongs to \mathcal{F}_{n^*} with $s = 1$ and either $b = (0, \rho)$ or $b' = (0, 0)$. Coming back to Assumption 2, part (i) says that $G_y^{e,b,s} = F_y$ for some value of (e, b, s) where e is always equal to ε .

The mid-censored Tobit estimator produces a value for the elasticity e^* and picks the mixture of normals characterized by (b^*, s^*) ; both of these imply a distribution of y : $G_y^* = G_y^{e^*, b^*, s^*}$. The Tobit best-fit distribution G_y^* may or may not match the observed distribution of y , even if Assumption 2(i) is true; in case it does match, the equality $F_y = G_y^*$ implies $e^* = \varepsilon$ by virtue of Assumption 2(i). In general, we may not have $F_y = G_y^*$ because the Tobit MLE does not necessarily minimize the distance between F_y and $G_y^{e,b,s}$ for a choice of parameters (e, b, s) . In fact, the Tobit MLE minimizes the Kullback-Leibler divergence between two conditional distributions of y given X averaged over

X , where the two conditional distributions are $F_{y|X}$ and $G_{y|X}^{e,b,s}$ for a choice of parameters (e, b, s) . These are two different minimization problems in general. Assumption 2(ii) states a condition on $F_{n^*|X}$ that guarantees that minimizing the average Kullback-Leibler divergence is equivalent to minimizing the distance between F_y and $G_y^{e,b,s}$.

To interpret Assumption 2(ii), note that the objective function of the minimization problem in (12) is a known function of (δ, θ, s) once we fix $F_{n^*|X} \in \mathcal{F}_{n^*|X}$ plus knowledge of the distribution of X , which is identified. That objective function is the average of two terms weighted by the bunching mass B and $1 - B$. The term on the LHS is the objective function of the least-squares problem conditional on $n^* \notin [n, \bar{n}]$, and the term on the RHS is an average of the log of the bunching mass conditional on X in case the distribution of n^* given X were normal. Thus, the objective function is a “penalized” least-squares problem. To gain further insight, assume $\varepsilon \approx 0$, so that $B \approx 0$ and $\underline{n} \approx \bar{n}$. In this case, (12) is approximately

$$(0, \beta, \sigma) = \arg \min_{\delta, \theta, s} \left\{ \log(s^2) + \frac{1}{s^2} \mathbb{E} \left[(n^* - D\delta - X\theta)^2 \right] \right\}.$$

This is equivalent to saying that the population regression of n^* on (D, X) must produce $(0, \beta)$ as coefficients, and that the variance of the regression error must be σ^2 . This translates to linear restrictions on the first two moments of the joint distribution of $(n^*, \mathbb{I}\{n^* \geq \bar{n}\}, X)$, but that distribution is otherwise unrestricted. In particular, a Gaussian distribution $F_{n^*|X}(n, X) = \Phi\left(\frac{n - Xb}{s}\right)$ belongs to the set $\mathcal{F}_{n^*|X}$, but not every element of $\mathcal{F}_{n^*|X}$ is Gaussian. We give examples of $F_{n^*|X}$ that satisfy Assumption 2 and are not Gaussian in the simulation experiments at the end of this section (Figures 2–3) and in Section B.7 of the supplemental appendix. Lemma 1 below shows that the mid-censored Tobit MLE identifies the elasticity under Assumption 2, and the proof is in Section A.3 of Appendix A.

Lemma 1. *Let F_y be the CDF of the observed distribution of y and G_y^* be the CDF of the Tobit best-fit distribution for y as defined above. Suppose Assumption 2(i) holds. If $G_y^* = F_y$, then $e^* = \varepsilon$. Moreover, suppose Assumption 2(ii) holds. Then, $G_y^* = F_y$ and $e^* = \varepsilon$.*

If the Tobit best-fit distribution of y matches the true distribution of y , Lemma 1 guarantees that the elasticity estimated by the Tobit is consistent for the true elasticity, regardless of whether $F_{n^*|X}$ is normal. We provide an example of this in the second simulation experiment at the end of this section (Figure 3), as well as in Section B.7 of the supplemental appendix (Figure B.2). Standard quasi-MLE asymptotic inference procedures apply here. Namely, the MLE $(\hat{\varepsilon}, \hat{\beta}, \hat{\sigma})$ obtained from (11) and centered at (e^*, b^*, s^*) 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

⁶The truncated Tobit model has log-likelihood that is slightly different from (11). Instead of the log-likelihood of

estimation strategy, because consistency of $\hat{\varepsilon}$ has weaker requirements in terms of Assumption 2 and the $G_y^* = F_y$ condition. Assumption 2 only needs to hold for the distribution of n^* conditional on n^* being in a small interval containing $[\underline{n}, \bar{n}]$. 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 1.

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 1 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.

To end this section, we carry out two simulation experiments to illustrate the robustness property of our Tobit estimator to lack of normality. In both experiments, we use the Skewed Generalized Error Distribution (SGED) with parameters μ , σ , k , and λ (see Theodossiou (2000), Section 5A). We denote the PDF $f_{n^*}(n)$ of a SGED as $SGED(n; \mu, \sigma, k, \lambda)$. Parameters $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ equal the mean and standard deviation of the distribution, respectively. The parameter $k \in \mathbb{R}_+$ relates to kurtosis, while $\lambda \in (-1, 1)$ regulates skewness. SGED nests the Laplace distribution ($\lambda = 0$, $k = 1$), the normal distribution ($\lambda = 0$, $k = 2$), and the uniform distribution ($\lambda = 0$, $k \rightarrow \infty$) as special cases. The data generating processes of these experiments have some empirical features that resemble those of the EITC data in Section 5, e.g., location of the kink and range of the distribution.

Experiment 1 has two goals. First, the experiment shows that a mixture of normals can be almost any distribution as long as the distribution of X is rich enough. The unconditional distribution of n^* does not need to be “locally normal” or “locally symmetric” at the kink. The second goal of Experiment 1 is to show that truncation without covariates may require a much smaller truncation window to fit the distribution of y compared to truncation with covariates. The key parameters of Equation 4 are $\varepsilon = 1$, $k = 2.0794$, $s_0 = 0.2624$, and $s_1 = -0.1054$. Figure 2a displays the distribution of n^* , which is a mixture of two SGEDs: $f_{n^*}(n) = (1/2)SGED(n; 1.6, 0.75, 4, -.5) + (1/2)SGED(n; 6, 0.75, 1, .5)$. The random variable X is a scalar and $\beta = 1$. We specify $F_{n^*|X}(n^*) = \Phi((n - X)/0.0717)$ and solve numerically for the distribution of X that satisfies $F_{n^*}(n) = \mathbb{E}[\Phi((n - X)/0.0717)]$ (Figure 2b). We generate 50,000 observations of (y, X) according to this model, and the histogram of y is displayed in Figure 2c. Although $f_{n^*|X}$ is normal, it is clear from the figures that f_{n^*} is very far from Gaussian, even within smaller truncation windows.

We estimate two different Tobit models using data from Experiment 1. The first model is correctly specified with covariate X . 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.

$y|X$, we maximize the log-likelihood of $y|X$, $k - \delta < y < k + \delta$ for $\delta > 0$, which has a truncated normal distribution censored at k .

For example, Figures 2c–2e show the histogram of simulated data for y , and the best-fit Tobit distributions for three truncation sizes, 100%, 60%, and 20%. Figure 2f displays the elasticity estimate as a function of the percentage of data used in each truncated estimation. As expected, 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 model we estimate with data from Experiment 1 omits the covariate X . The Tobit model is misspecified because f_{n^*} is not normal. As expected, estimation using all of the data does not fit the distribution of y (Figure 2g). As a general rule, the smaller the truncation window, the better the fit to the distribution of y (Figures 2g–2i), although a perfect fit is not always guaranteed for reasonable sample sizes. It is only in the last feasible truncation window of 20% that the fit becomes reasonable and the elasticity estimate reaches the true value (Figure 2j). This experiment shows the importance of covariates for the fit of the distribution and that truncation does not always yield the same fit that the inclusion of relevant covariates does.

We move to Experiment 2, which again has two goals. First, not only normality of n^* is not required for consistency; conditional normality of $F_{n^*|X}$ is not required either. Second, consistency of the Tobit elasticity and perfect fit of the distribution of y do not require truncation in models without conditional normality (i.e., models with $F_{n^*|X}$ misspecified). Figure 3a plots the PDF of n^* , which is approximately the mixture of two SGEDs, $(1/2)SGED(n; 1, 0.75, 4, -.5) + (1/2)SGED(n; 6, 0.75, 1, .5)$. The distribution of scalar X is discrete with 20 mass points (Figure 3b) and is chosen such that $F_{n^*}(n) = \mathbb{E}[\Phi((n - X)/0.1919)]$ approximates the CDF of the mixture of two SGEDs. We solved numerically for non-normal conditional distributions of n^* given X that satisfy Equation 12. Figure 3d displays the true PDFs $f_{n^*|X=x}$ in black and the normal PDFs $g_{n^*|X=x}$ assumed by the Tobit in gray, for all values of x . We clearly see that $f_{n^*|X}$ is not normal. We then generate 50,000 observations of (y, X) and fit our Tobit model with the covariate X to the entire sample. Despite the lack of conditional normality and truncation, the Tobit model fits the distribution of y (Figure 3c) and estimates the elasticity at $\hat{\varepsilon} = 1.0083$ (S.E. 0.0073). Section B.7 in the supplement repeats this experiment for the case n^* has uniform distribution.

This section gives a practical rationale for using the truncated Tobit model but other more flexible model assumptions, such as index models or the semi-parametric Tobit model of Chen, Tamer, and Torgovitsky (2011) are also possible. For example, and in terms of our notation, Section 7.3 of Chen et al. (2011) assumes $n^* = X\beta + \sigma(X)\nu$, with G being the CDF of the distribution of ν conditional on X . Both σ and G are unknown but smooth functions, and ε and β are partially identified. The more structure one imposes on σ and G , the narrower the partially identified set gets, eventually getting to point identification as in our case. Chen et al. (2011) then construct confidence regions for ε and β that are valid regardless of point or partial identification using a sieve MLE bootstrap. Chen, Christensen, and Tamer (2018) provide a computationally attractive alternative to the sieve MLE bootstrap that is based on Monte Carlo simulations. Overall these methods constitute helpful approaches to performing sensitivity analysis of model assumptions and

suggest a rich area for future work.

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 $\beta(\tau) \in \mathbb{R}^{1 \times (d+1)}$ such that

$$Q_\tau(n^* | X) = X\beta(\tau), \quad (13)$$

where Q_τ denotes the τ -th quantile of a distribution. A common choice in applied work is $\tau = 1/2$ or the median regression. The restriction in (13) may be a flexible one if one includes transformations of X on the right-hand side, e.g., polynomials and interaction terms.

Equation 10 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 Equation 13,

$$Q_\tau(y | X) = \min\{\varepsilon s_0 + X\beta(\tau); \max\{k; \varepsilon s_1 + X\beta(\tau)\}\}. \quad (14)$$

For those observations such that $X\beta(\tau) < k - \varepsilon s_0$ or $X\beta(\tau) > k - \varepsilon s_1$, the quantile $Q_\tau(y | X)$ 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 ε .

Lemma 2. *Define $\tilde{X} = [X, \mathbb{I}\{Q_\tau(y | X) > k\}]$, a random vector in $\mathbb{R}^{1 \times (d+2)}$. Assume $\mathbb{E}[\mathbb{I}\{Q_\tau(y | X) \neq k\} \tilde{X}' \tilde{X}]$ has full rank and that Equation 13 holds. Then ε is point identified.*

The quantile method does not contradict the impossibility of point identification discussed in Section 3.2, that is, we *do* need restrictions on the distribution of n^* conditional on X to point identify the elasticity. These restrictions are Equation 13 and the rank condition of Lemma 2. To see that, note that in the absence of covariates, the rank condition is never satisfied. When we have covariates, we are not free to specify $Q_\tau(n^* | X)$ as flexible as desired. For a fixed distribution of X , the rank condition eventually fails as we increase the flexibility of $Q_\tau(n^* | X)$.

We illustrate this fact with a simple example. Suppose the researcher has two dummy variables, W_1 and W_2 , and wants to be fully flexible. An unrestricted $Q_\tau(n^* | W_1, W_2)$ contains four parameters, because the conditional quantile function takes at most four different values. That is, $Q_\tau(n^* | W_1, W_2) = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \beta_3 W_1 W_2$. In terms of Lemma 2, $X = [1, W_1, W_2, W_1 W_2]$ is 1×4 , \tilde{X} is 1×5 , which implies $d = 3$. In the best case scenario for identification, four values of $Q_\tau(n^* | W_1, W_2)$ translate to four values of $Q_\tau(y | W_1, W_2)$ that are all different from k , so that $\mathbb{I}\{Q_\tau(y | X) \neq k\} = 1$. The matrix $\mathbb{E}[\mathbb{I}\{Q_\tau(y | X) \neq k\} \tilde{X}' \tilde{X}] = \mathbb{E}[\tilde{X}' \tilde{X}]$ is 5×5 but has rank equal to 4 at most. Thus, $Q_\tau(n^* | X)$ must be restricted to fewer parameters for point identification to be possible.

As seen in the example with dummies, the amount of variation in covariates trades off with the degree of flexibility in specifying the parametric functional form for $Q_\tau(n^* | X)$. The more variation researchers have (e.g., continuous vs. discrete X), the more flexible the parametric functional form may be, and the less damaging misspecification errors will be. Unfortunately, our data in Section 5 only have binary covariates, which severely limits our ability to obtain sensible estimates using the quantile identification method. An interesting extension to Lemma 2 relates to the work by Hong and Tamer (2003) and Khan and Tamer (2009), which connects censored quantile regressions to moment inequalities and partial identification. Although increasing the flexibility in the specification of $Q_\tau(n^* | X)$ may violate the rank condition of Lemma 2, researchers may still use those methods to partially identify the elasticity. For inference, researchers may use the simulation-based confidence regions of Chen et al. (2018), which are valid regardless of point or partial identification.

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 Hong and Tamer (2003), Khan and Tamer (2009), and Chernozhukov et al. (2015). In the simpler case of $Q_\tau(y | X) = X\beta(\tau)$, Koenker and Bassett (1978) show that a consistent estimator for $\beta(\tau)$ is obtained by the solution to the problem

$$\min_{b \in \mathbb{R}^{d+1}} \sum_{i=1}^n [\rho_\tau(y_i - X_i b)], \quad (15)$$

where (y_i, X_i) $i = 1, \dots, n$ is an iid sample and $\rho_\tau(u) = (\tau - 1(u \leq 0))u$ is the so-called “check function.” In our case, the parametric conditional quantile function $Q_\tau(y | X)$ is given in Equation 14. The slope and intercept coefficients are estimated by

$$(\hat{b}(\tau), \hat{\delta}(\tau)) = \arg \min_{b \in \mathbb{R}^d, \delta \in \mathbb{R}} \sum_{i=1}^n [\rho_\tau(y_i - \min\{X_i' b; \max\{k; X_i' b + \delta\}\})], \quad (16)$$

where $\hat{b}(\tau)$ is consistent for $\beta(\tau) + [\varepsilon s_0, 0, \dots, 0]'$, 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 is asymptotically normal.

The optimization problem in Equation 16 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 nonparametric 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 1). Furthermore, if the Tobit elasticity is within narrow nonparametric 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.

The income distribution for individuals with one child demonstrates clear bunching around the \$8,580 kink (year 2008 dollars) in the EITC schedule (Figure 5). Because the marginal tax rate increases from -34 percent to 0 percent at \$8,580, individuals have strong incentives to report more income up to 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 (Saez, 2001).

For example, a true elasticity of 0.2 implies the optimal top marginal tax should be 69%. If instead, the true elasticity is 0.3, the optimal top marginal tax should be 60%.⁷ As demonstrated above, 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), Figure 1 by Goncalves and Mello (2018), and Figure 8 by Goff (2022). 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.⁸ 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 and Section B.3 of the supplement. We simply aim at removing the friction error from the sample, while the polynomial strategy aims to remove friction error and recover the counterfactual distribution of income, which requires much stronger restrictions according to the discussion in Section 3.2. Our filtering procedure consistently estimates the CDF of y under the following conditions: (i) frictions only affect bunching individuals additively; (ii) friction error is independent of unobserved heterogeneity n^* and has known support containing zero, i.e., $[-1500, +1500]$; (iii) the CDF of y is represented by a seventh order polynomial with an intercept change at the kink. Section B.9 of the supplement has the proof of this claim. 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,

⁷This 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 ε^c and ε^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.

⁸There are several works in progress investigating frictions. See, for example, Aronsson, Jenderny, and Lanot (2018); Alvero and Xiao (2020); Cattaneo et al. (2018), and McCallum and Navarrete (2022).

⁹Section B.6 of the supplement recomputes our estimates using the filtering procedure employed by Saez (2010).

nonparametric 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 0.32 for the all filers sample, 0.81 for self-employed individuals, 0.77 for self-employed married individuals, and 0.84 for self-employed not married individuals.

The guidelines for implementation of our nonparametric 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 nonparametric 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 4. 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 ε , and the upper bound is infinity (Theorem 2).

A large range between lower and upper bounds in Figure 4 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 at least 0.4 for the other three subsamples.

¹⁰We 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.

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 5–8 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 a list of indicator variable covariates for whether a tax return filer: filed as married, used a tax preparer, claimed a real estate interest deduction, claimed mortgage interest as a business expense, received unemployment compensation, contributed to charity, received social security benefits, took the educator expenses deduction, paid into an Individual Retirement Account (IRA) for the primary filer, claimed a student loan interest deduction, claimed exemption for children living away from home, received pension income, capital gains, farm profit or loss, a positive tax credit, positive wages, owed positive income tax before tax credits but zero after, took a self-employed health insurance deduction, paid into a Keogh retirement plan, received income from a business or farm, filed a 1040 form instead of simpler form, claimed secondary taxpayer exemption, and paid into an IRA for a secondary filer. We include year dummies variables with 1995 as the excluded year.

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 60% 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 5–8 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 8, the estimates are all around 0.75, using less than 60% 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 5–8.¹¹

5.3 Comparisons Across Methods

Comparisons across methods provide 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

¹¹We note that the filtering procedure that removes optimizing errors dictates the shape of the distribution close to the kink. Researchers should keep in mind, as part of their robustness checks, that different filtering methods may yield different shapes of the distribution close to the kink, which may in turn lead to different elasticity estimates.

the truncation window decreases and, as a result, the fit of the Tobit distribution improves. For the all filers sample, an M larger than 0.5 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 7 and 8 demonstrate that the bunching mass is approximately 6 times 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 much 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 in column 1 of Table 1 indicate a higher elasticity for self-employed not married individuals and so do the lower bounds of our partial identification method and Tobit estimates. They disagree in the magnitudes. The trapezoidal estimate in column 1 says the elasticity for self-employed not married individuals is 9% higher compared to self-employed married individuals; conservative lower bounds in column 3 say it is 53% higher while the truncated Tobit estimates in column 7 point to a 46% difference. 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 allows researchers to tailor their estimation to different assumptions and data variation. These include nonparametric bounds and semi-parametric censored models with covariates. The nonparametric 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, Marcelo Moreira, Whitney

Newey, Andreas Peichl, Emmanuel 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

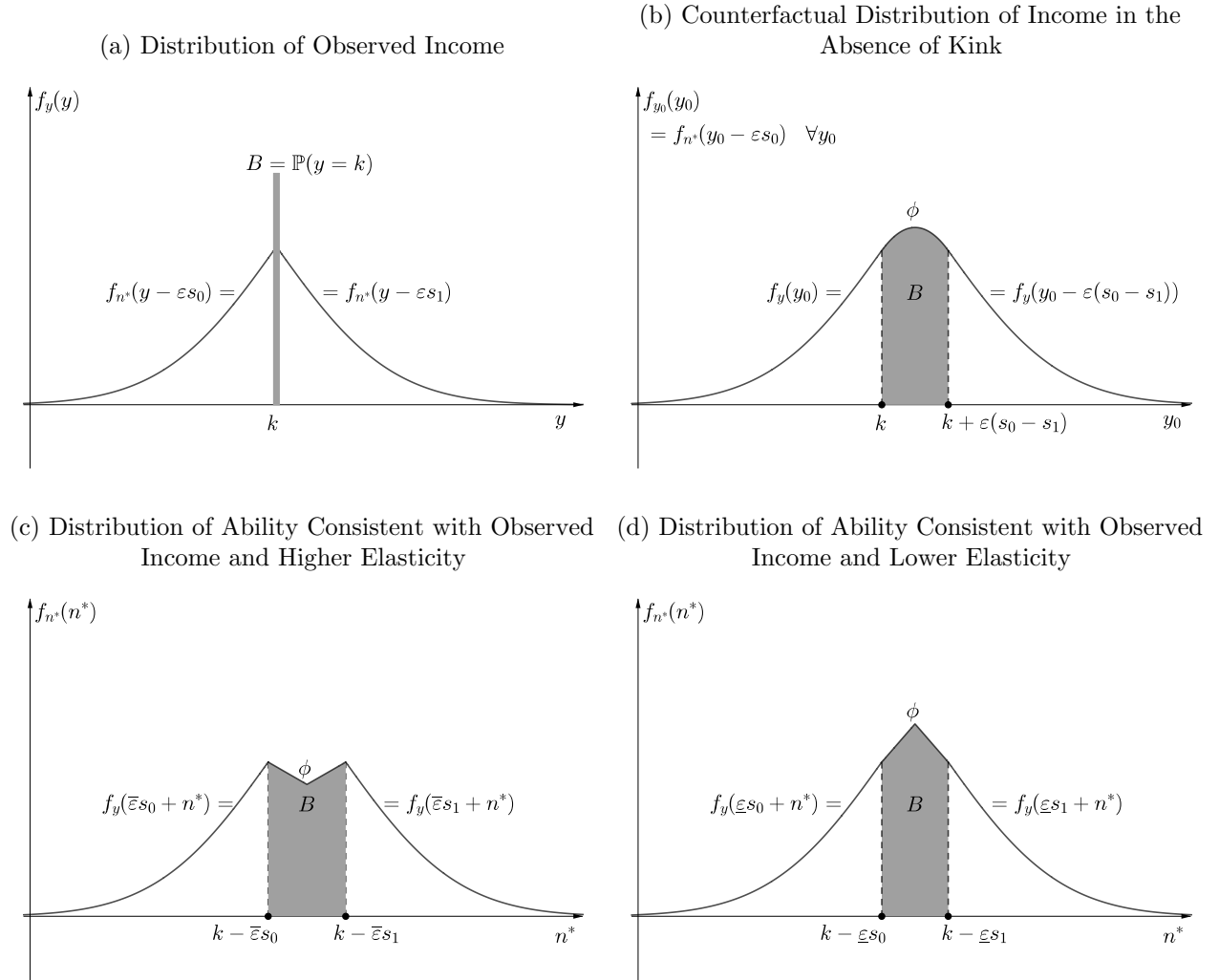
- Allen, E. J., P. M. Dechow, D. G. Pope, and G. Wu (2017, June). Reference-Dependent Preferences: Evidence from Marathon Runners. *Management Science* 63(6), 1657--1672.
- Alvero, A. and K. Xiao (2020). Fuzzy bunching. Working Paper 3611447, SSRN.
- Amemiya, T. (1984). Tobit Models: A Survey. *Journal of Econometrics* 24(1-2), 3--61.
- Aronsson, T., K. Jenderny, and G. Lanot (2018, June). Alternative parametric bunching estimators of the ETI. Umeå Economic Studies 956, Umeå University, Department of Economics.
- 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., C. Caetano, H. Jales, and N. Seegert (2023). Bunching estimation methods. In K. F. Zimmermann (Ed.), *Handbook of Labor, Human Resources, and Population Economics (forthcoming)*. Springer.
- Bertanha, M., A. H. McCallum, A. Payne, and N. Seegert (2022). Bunching estimation of elasticities using stata. *The Stata Journal* 22(3), 597--624.
- Bertanha, M., A. H. McCallum, and N. Seegert (2018, March). Better Bunching, Nicer Notching. Working paper 3144539, SSRN, presented at the 2018 UCSD Bunching Estimator Workshop.
- Bertanha, M. and M. J. Moreira (2020). Impossible inference in econometrics: Theory and applications. *Journal of Econometrics* 218(2), 247--270.
- 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 12/15, 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.

- 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. R. Nielsen (2020). Should children do more enrichment activities? leveraging bunching to correct for endogeneity. Technical Report 2020-036, Board of Governors of the Federal.
- 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.
- Chen, X., T. M. Christensen, and E. Tamer (2018). Monte Carlo Confidence Sets for Identified Sets. *Econometrica* 86(6), 1965--2018.
- Chen, X., E. T. Tamer, and A. Torgovitsky (2011, November). Sensitivity Analysis in Semiparametric Likelihood Models. Discussion Paper 1836, Cowles Foundation.
- 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.
- Goff, L. (2022). Treatment effects in bunching designs: The impact of the federal overtime rule on hours. Technical report, University of Calgary.
- 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.
- Hong, H. and E. Tamer (2003). Inference in Censored Models with Endogenous Regressors. *Econometrica* 71(3), 905–932.
- Hungerman, D. M. and M. Ottoni-Wilhelm (2021). Impure impact giving: Theory and evidence. *Journal of Political Economy* 129(5), 1553–1614.
- 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.
- Khan, S. and E. Tamer (2009). Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities. *Journal of Econometrics* 152(2), 104–119.

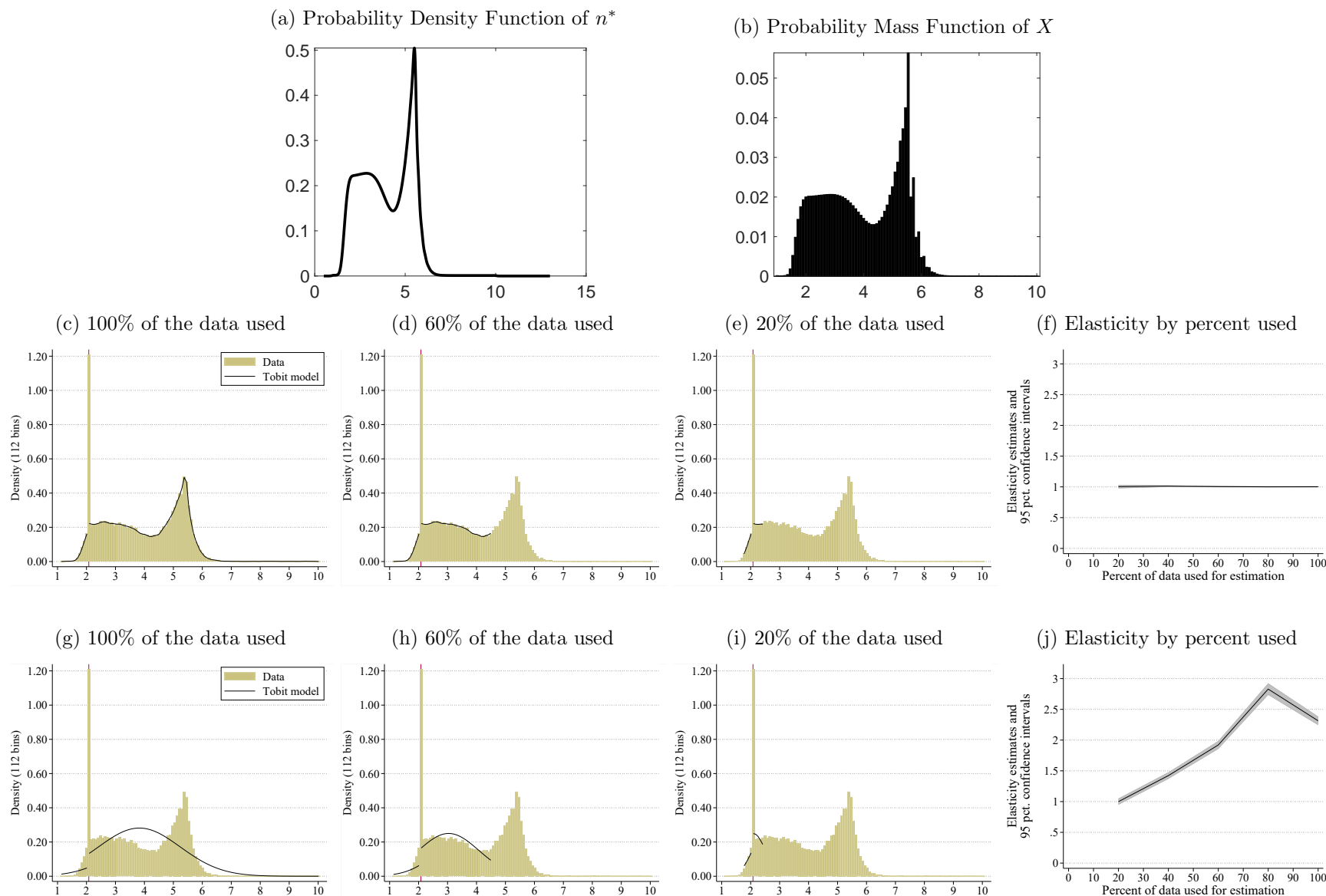
- Kim, W., K. Kwon, S. Kwon, and S. Lee (2018). The identification power of smoothness assumptions in models with counterfactual outcomes. *Quantitative Economics* 9(2), 617--642.
- Kleven, H. J. (2016). Bunching. *Annual Review of Economics* 8, 435--464.
- Kleven, H. J. and M. Waseem (2013). Using Notches 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.
- McCallum, A. H. and M. A. Navarrete (2022). Why Don't Taxpayers Bunch at Kink Points? mimeo, available on request, Board of Governors of the Federal Reserve System (U.S.).
- 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.
- Rambachan, A. and J. Roth (2023, 02). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*. rdad018.
- 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.
- Theodossiou, P. (2000, May). Skewed generalized error distribution of financial assets and option pricing. Working paper 219679, SSRN.
- 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



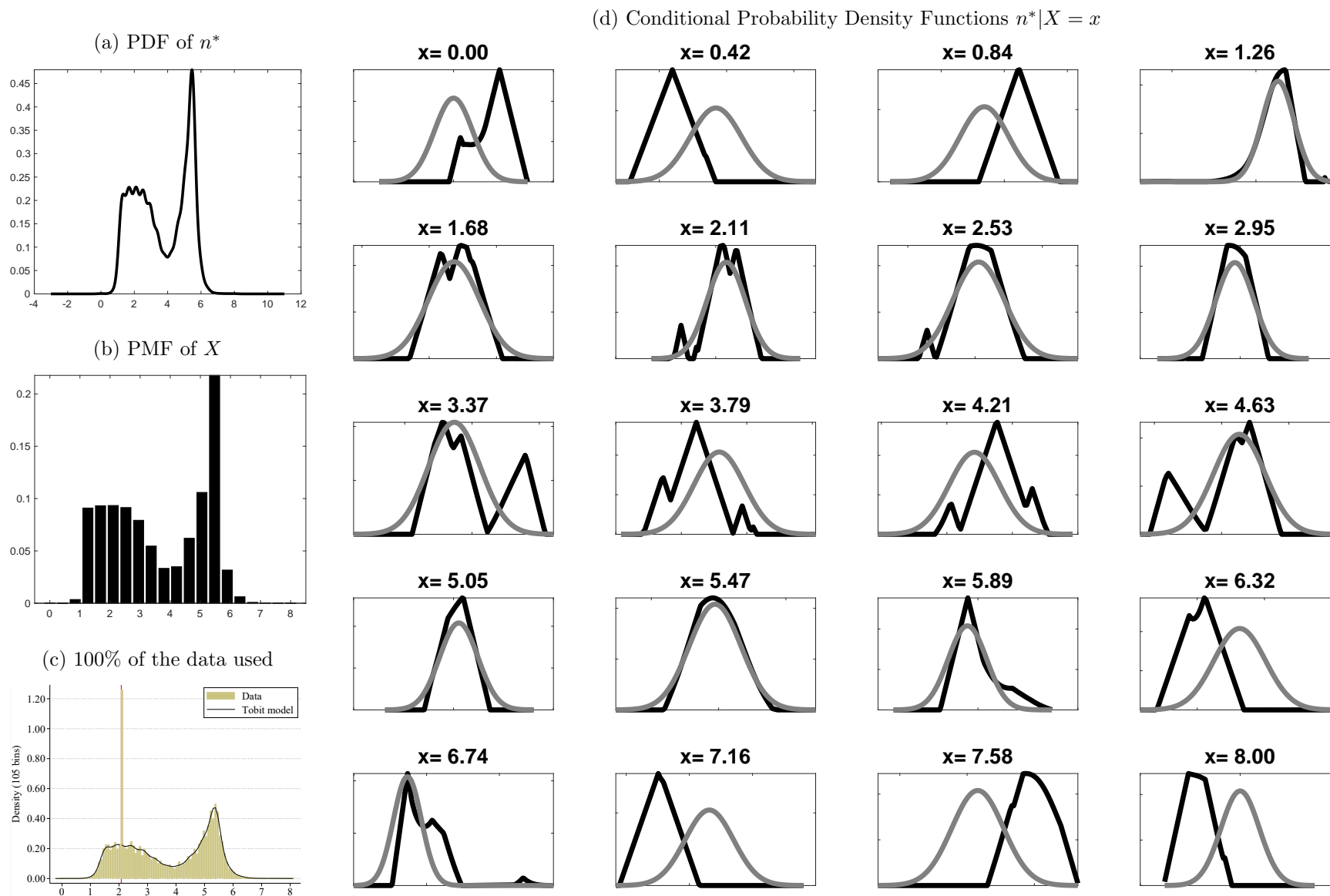
Notes: Panel 1a plots an example of PDF of y . The continuous portions are equal to the PDF of ability n^* shifted by εs_0 for $y < k$, and by εs_1 for $y > k$, respectively. The shaded area represents a discrete mass point with probability $B = \mathbb{P}(y = k)$, that is, the probability of bunching. 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 εs_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., ϕ). The shaded area under ϕ 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 εs_0 , if $n^* < k - \varepsilon s_0$; or shifted by εs_1 , if $n^* > k - \varepsilon s_1$. Aside from B , the distribution of income does not contain any information about the shape of ϕ 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 ϕ , which correspond, respectively, to lower and upper bounds on the elasticity (Theorem 2).

Figure 2: Robustness of Tobit Estimates to Lack of Normality—Experiment 1



Notes: This simulation experiment illustrates that the mid-censored Tobit model is able to fit non-normal distributions of n^* and retrieve the right elasticity using covariates and truncation. We generate 50,000 observations of y and a scalar X following Experiment 1 detailed in Section 4.2.1. The variable n^* is distributed as a mixture of two Skewed Generalized Error Distributions (Panel a), $n^*|X$ is Gaussian, the kink point is at $k = 2.0794$, and $\varepsilon = 1$. Panels c–f correspond to estimates from a Tobit model that is correctly specified with covariate X . Panels c–e show the histogram of simulated data for y and the best-fit Tobit distributions for three truncation sizes. Panel f displays the elasticity estimate as a function of the percentage of data used in each truncated estimation, along with 95% confidence intervals. Similarly, Panels g–j correspond to estimates from a Tobit model that is incorrectly specified without covariate X .

Figure 3: Robustness of Tobit Estimates to Lack of Normality—Experiment 2



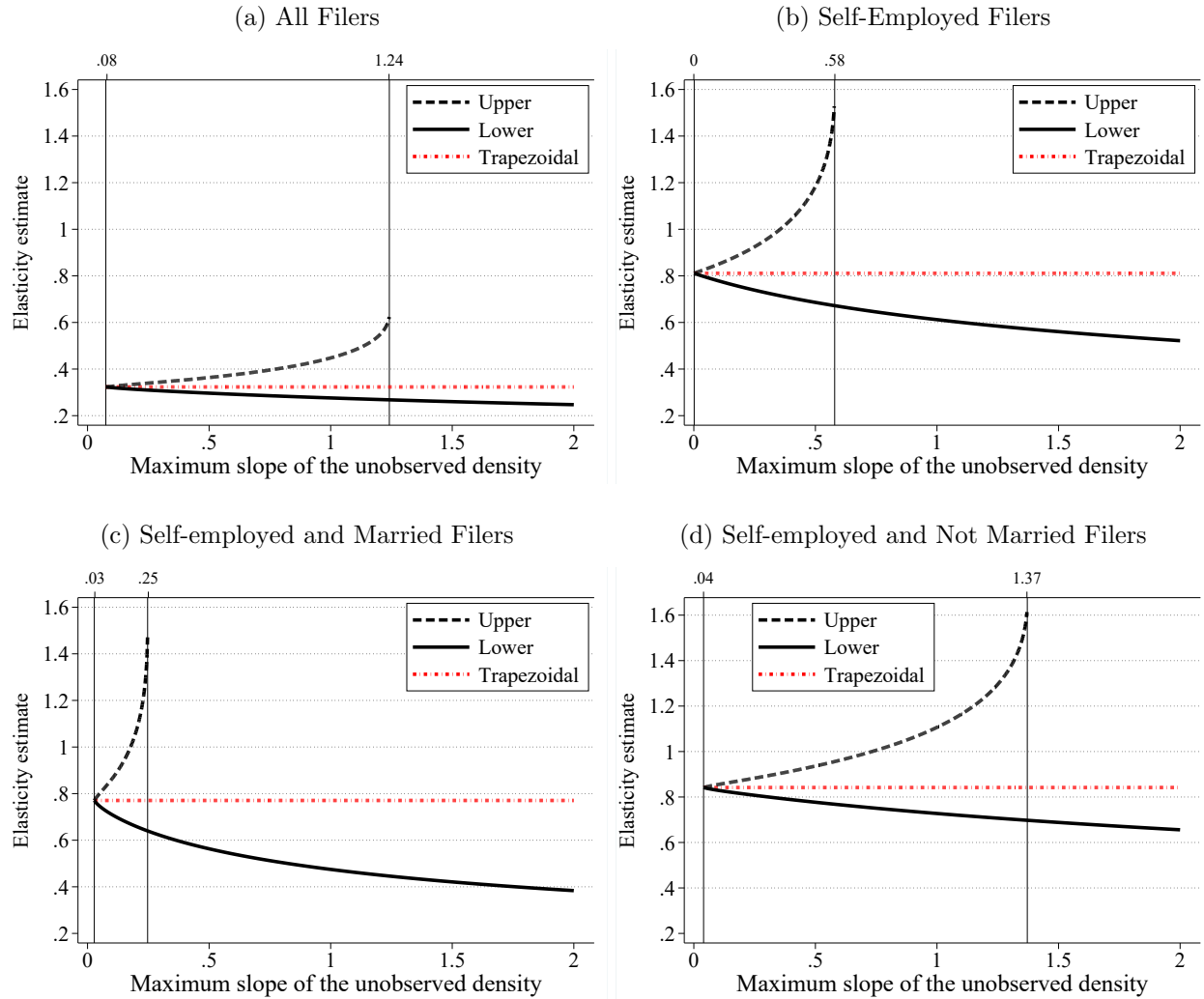
Notes: This simulation experiment illustrates that the mid-censored Tobit model is able to fit non-normal distributions of n^* and retrieve the right elasticity even when the conditional distribution $n^*|X$ is not Gaussian and there is no truncation. We generate 50,000 observations of y and a scalar X following Experiment 2 detailed in Section 4.2.1. The variable n^* is approximately a mixture of two Skewed Generalized Error Distributions (Panel a), the distribution of X is discrete (Panel b), the kink point is at $k = 2.0794$, and $\varepsilon = 1$. Panel c shows the histogram of simulated data for y and the best-fit Tobit distribution using covariate X and no truncation. Panel d displays the true conditional PDFs of $n^*|X$ in black along with the Gaussian PDFs in gray that are assumed by the Tobit model. The elasticity is estimated at $\hat{\varepsilon} = 1.0083$ (S.E. 0.0073).

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
<i>All</i>								Obs. 188.3m
Elasticity (ε)	0.323	[0.297, 0.364]	[0.276, 0.448]	0.163 (0.0001)	0.242 (0.0002)	0.250 (0.0002)	0.278 (0.0002)	Avg. \$53.5k Std. \$64.6k
<i>Self-employed</i>								Obs. 33.4m
Elasticity (ε)	0.811	[0.686, 1.183]	[0.612, ∞]	0.568 (0.0005)	0.752 (0.0007)	0.746 (0.0008)	0.750 (0.0008)	Avg. \$60.7k Std. \$77.2k
<i>Self-employed, married</i>								Obs. 23.9m
Elasticity (ε)	0.770	[0.563, ∞]	[0.475, ∞]	0.304 (0.0006)	0.464 (0.0009)	0.540 (0.0010)	0.554 (0.0010)	Avg. \$73.6k Std. \$84.4k
<i>Self-employed, not married</i>								Obs. 9.5m
Elasticity (ε)	0.842	[0.777, 0.936]	[0.728, 1.107]	0.875 (0.0010)	0.774 (0.0009)	0.745 (0.0010)	0.807 (0.0015)	Avg. \$28.3k Std. \$39.6k

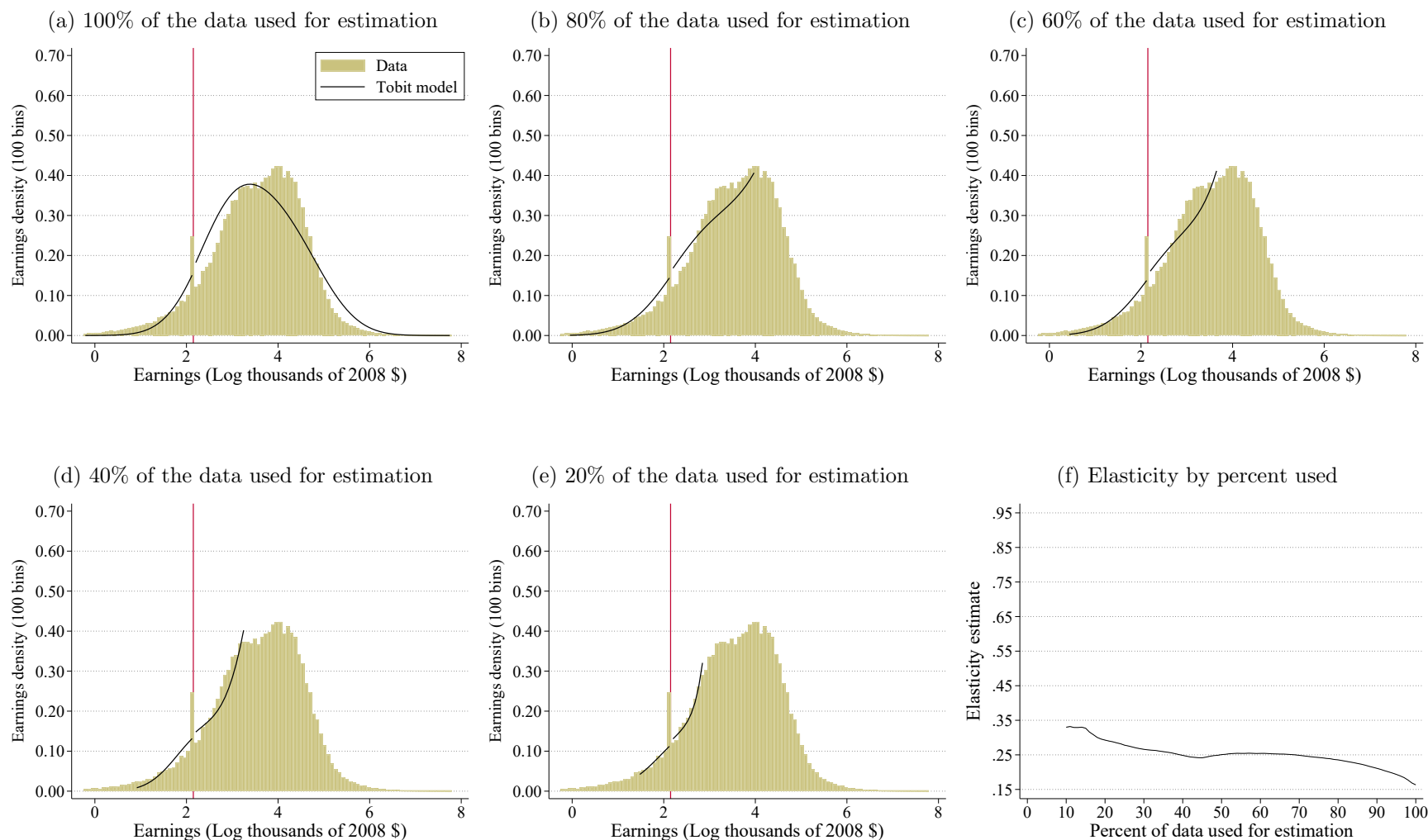
Notes: The table shows estimates of the elasticity for four different subsamples of the IRS data, and using three different approaches discussed in the paper. The first approach (column 1) uses the trapezoidal approximation to point-identify the elasticity (Example 1). We obtained non-parametric estimates 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). 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 in parentheses. 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 and are listed in Section 5.2.

Figure 4: Partial Identification Bounds for the Elasticity



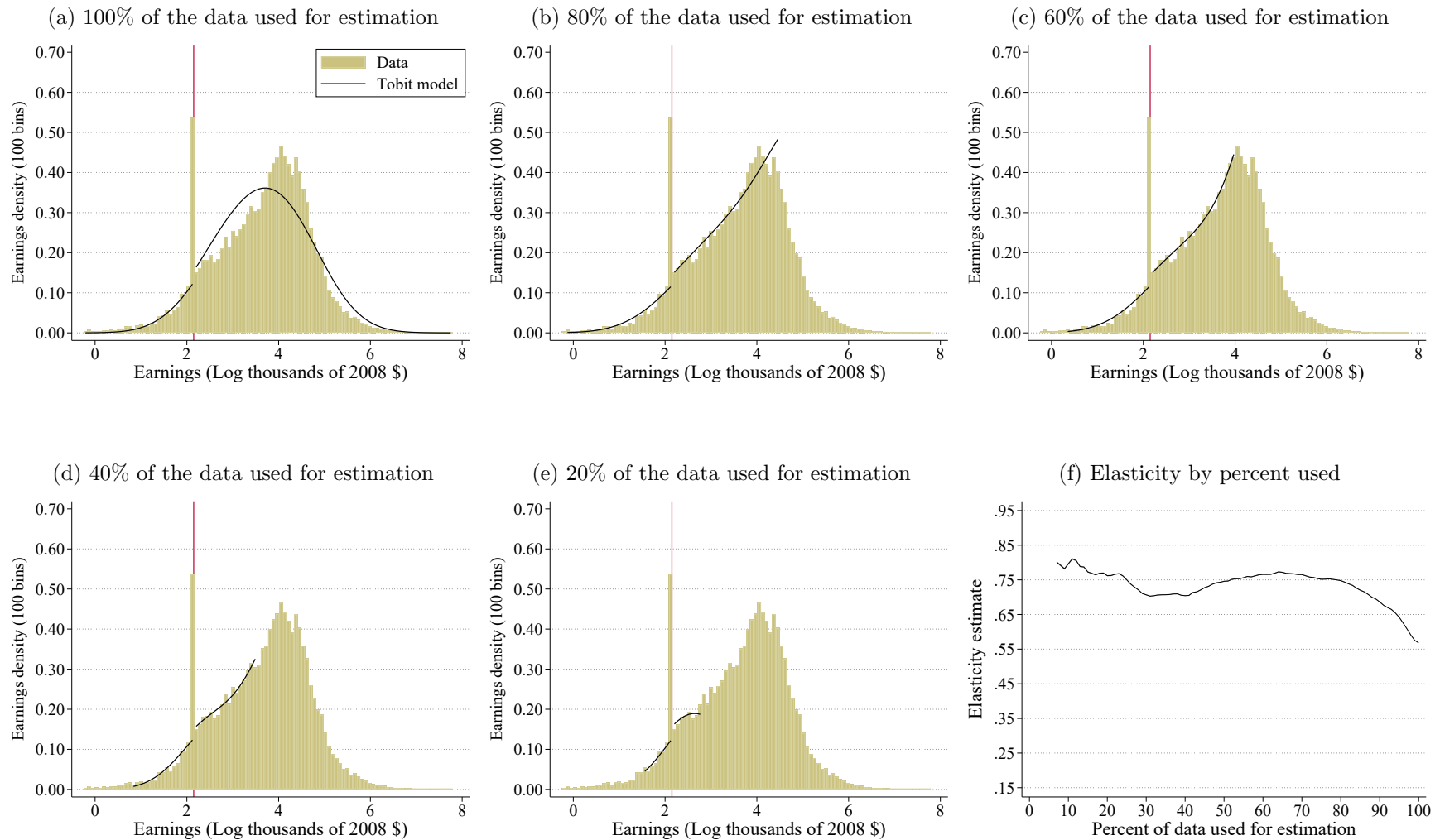
Notes: Panels a through d display partially identified sets for the elasticity for all filers with one child, and three other subsamples defined by employment and marital status. The y-axis has elasticity values between lower and upper bounds given various choices of M on the x-axis, that is, the maximum 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 5: Truncated Tobit - All Filers



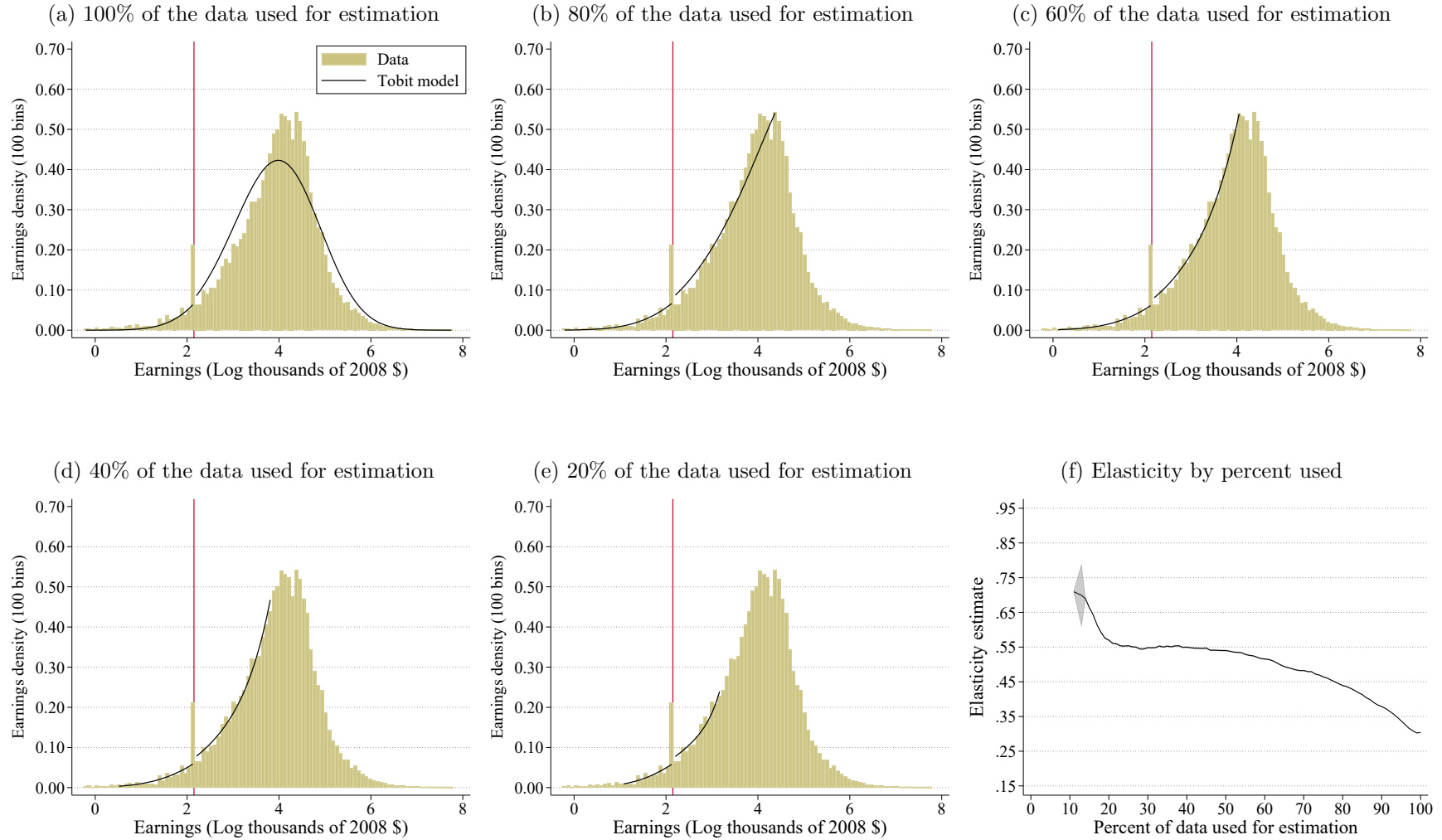
Notes: the figure displays best-fit Tobit distributions and elasticity estimates for various choices of a symmetric truncation window around the kink point. The set of covariates that enters the Tobit estimation is kept constant across different truncation windows and are listed in Section 5.2. Panels 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 6: Truncated Tobit - Self-employed Filers



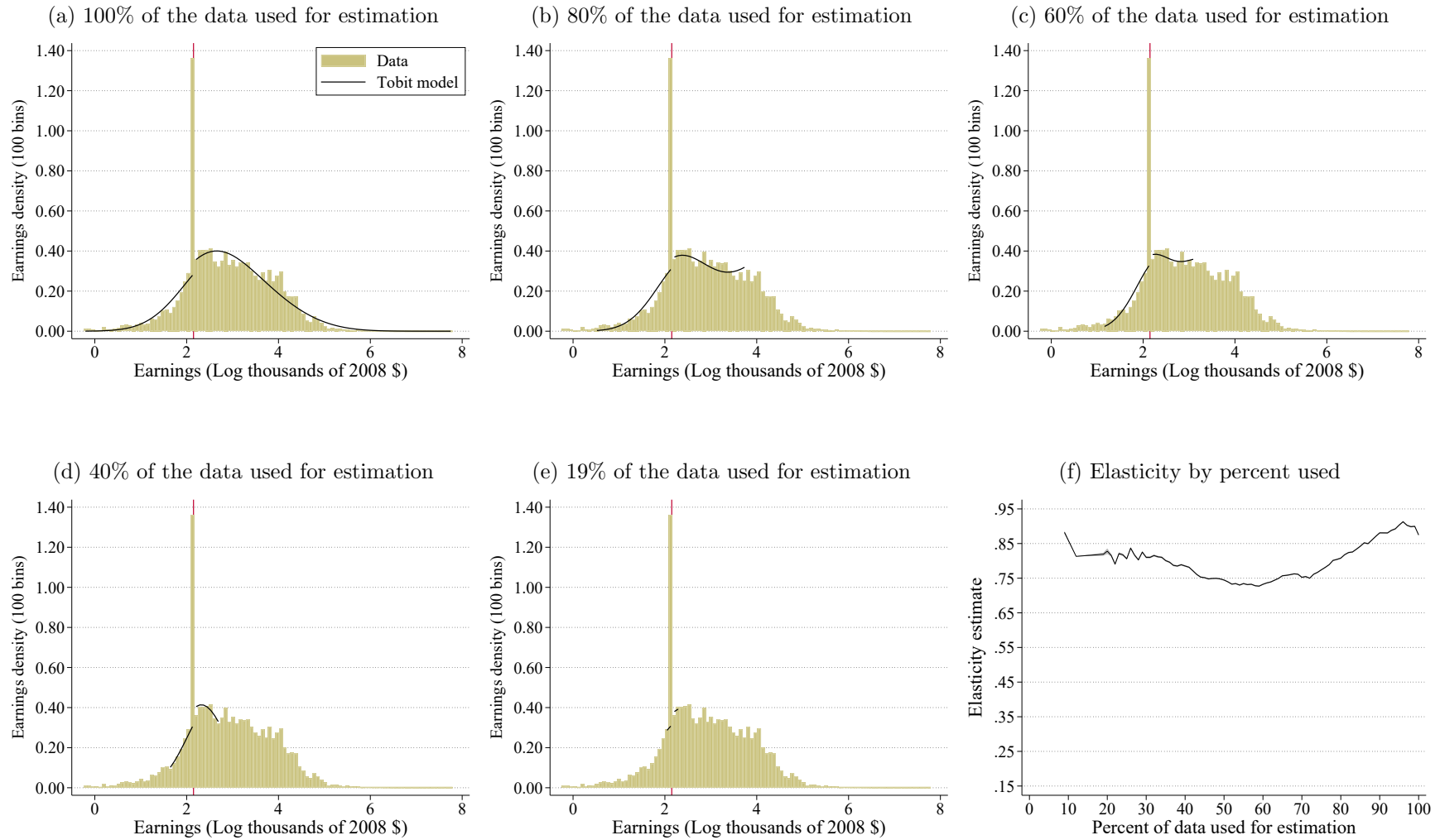
Notes: the figure displays best-fit Tobit distributions and elasticity estimates for various choices of a symmetric truncation window around the kink point. The set of covariates that enters the Tobit estimation is kept constant across different truncation windows and are listed in Section 5.2. Panels 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 7: 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. The set of covariates that enters the Tobit estimation is kept constant across different truncation windows and are listed in Section 5.2. 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 8: 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. The set of covariates that enters the Tobit estimation is kept constant across different truncation windows and are listed in Section 5.2. 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 $p \in \{1, \dots, L\}$ 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 $q \in \{p+1, \dots, L\}$ 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 ε using this information.

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

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

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

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

The elasticity ε is identified if there exists a unique solution for ε 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 ε . The solution occurs at the value of ε 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:

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

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, $Y_p^I > [C_p - I_{q-1} + K_{q-1}(1 - t_{q-1})]/(1 - t_{q-1})$.

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 ε ,

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

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 ε ,

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

which is also strictly positive. The limit of $\frac{\partial}{\partial \varepsilon} LHS$ as $\varepsilon \rightarrow \infty$ 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.

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

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

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 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 $[\underline{n}, \bar{n}]$. 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 \bar{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 \bar{f}_{n^*} , and that has slope magnitude less or equal than M . The maximum area is

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

The function $\bar{A}(\varepsilon)$ is strictly increasing with respect to ε 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

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

The function $\underline{A}(\varepsilon)$ is strictly increasing with respect to ε 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 ε over $\varepsilon \geq \varepsilon_2$.

Therefore, we have characterized the maximum and minimum areas $\underline{A}(\varepsilon)$ and $\bar{A}(\varepsilon)$ for any given ε . These areas are undefined if $\varepsilon < \varepsilon_1$, they are equal if $\varepsilon = \varepsilon_1$, they are strictly increasing wrt ε and $\underline{A}(\varepsilon) \leq \bar{A}(\varepsilon)$ for $\varepsilon \in (\varepsilon_1, \varepsilon_2)$. For $\varepsilon \geq \varepsilon_2$, $\bar{A}(\varepsilon)$ continues to grow wrt ε 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 $B < \underline{A}(\varepsilon_1) = \bar{A}(\varepsilon_1)$, there does not exist any function f_{n^*} consistent with any elasticity

ε , 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 ε such that for any ε 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

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

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

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

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.3 Tobit Regression - Proof of Identification Lemma 1

Part I:

The probability limit of the Tobit MLE is (e^*, b^*, s^*) and (b^*, s^*) correspond to a unconditional distribution of n^* : $G_{n^*}^*(n) = \mathbb{E}[\Phi((n - Xb^*)/s^*)]$, where $G_{n^*}^* \in \mathcal{F}_{n^*}$. The condition $F_y(y) = G_y^*(y)$ is equivalent to:

$$\begin{aligned} G_y^*(y) &= G_{n^*}^*(y - e^*s_0) = F_{n^*}(y - \varepsilon s_0) = F_y(y) \quad \text{for } \forall y < k, \\ G_y^*(y) &= G_{n^*}^*(y - e^*s_1) = F_{n^*}(y - \varepsilon s_1) = F_y(y) \quad \text{for } \forall y \geq k. \end{aligned}$$

Assumption 2(i) says that \mathcal{F}_{n^*} satisfies Assumption 1. Assumption 1 says that the only values of $e \geq 0$ and $G_{n^*} \in \mathcal{F}_{n^*}$ that solve the equations

$$\begin{aligned} G_{n^*}(y - es_0) &= F_{n^*}(y - \varepsilon s_0) = F_y(y) \quad \text{for } \forall y < k, \\ G_{n^*}(y - es_1) &= F_{n^*}(y - \varepsilon s_1) = F_y(y) \quad \text{for } \forall y \geq k, \end{aligned}$$

are $e = \varepsilon$ and $G_{n^*} = F_{n^*}$. Therefore, $e^* = \varepsilon$ and $G_{n^*}^* = F_{n^*}$.

Part II:

Start with a distribution $F_{n^*|X} \in \mathcal{F}_{n^*|X}$. By Assumption 2(ii), we have that $\mathbb{E}[F_{n^*|X}(n, X)] = \mathbb{E}[\Phi((n - X\beta)/\sigma)]$ and $(0, \beta, \sigma)$ is the solution to the minimization problem in (12). We show below that this implies that $(e^*, b^*, s^*) = (\varepsilon, \beta, \sigma)$, where (e^*, b^*, s^*) is the probability limit of the Tobit MLE.

We show that the minimization problem in Equation 12 maps to the mid-censored Tobit

maximization problem.

$$\begin{aligned}
& \min_{\delta, \theta, s} \frac{1-B}{2} \left\{ \log(s^2) + \frac{1}{s^2} \mathbb{E} \left[(n^* - D\delta - X\theta)^2 \mid n^* \notin [\underline{n}, \bar{n}] \right] \right\} \\
& \quad - B \mathbb{E} [\log(B_N(X, \delta, \theta, s)) \mid n^* \in [\underline{n}, \bar{n}]], \\
& \iff \\
& \max_{\delta, \theta, s} - \frac{1-B}{2} \left\{ \log(s^2) + \frac{1}{s^2} \mathbb{E} \left[(n^* - D\delta - X\theta)^2 \mid n^* \notin [\underline{n}, \bar{n}] \right] \right\} \\
& \quad + B \mathbb{E} [\log(B_N(X, \delta, \theta, s)) \mid n^* \in [\underline{n}, \bar{n}]] \\
& \iff \\
& \max_{\delta, \theta, s} (1-B) \left\{ \mathbb{E} \left[\log \left(\frac{1}{s} \phi \left(\frac{n^* - D\delta - X\theta}{s} \right) \right) \mid n^* \notin [\underline{n}, \bar{n}] \right] \right\} \\
& \quad + B \mathbb{E} [\log(B_N(X, \delta, \theta, s)) \mid n^* \in [\underline{n}, \bar{n}]] \\
& \iff \\
& \max_{\delta, \theta, s} \mathbb{E} \left\{ \mathbb{I}\{n^* < \underline{n}\} \log \left(\frac{1}{s} \phi \left(\frac{n^* - X\theta}{s} \right) \right) \right. \\
& \quad + \mathbb{I}\{n^* > \bar{n}\} \log \left(\frac{1}{s} \phi \left(\frac{n^* - \delta - X\theta}{s} \right) \right) \\
& \quad \left. + \mathbb{I}\{\underline{n} \leq n^* \leq \bar{n}\} \log \left(\Phi \left(\frac{k - \varepsilon s_1 - \delta - X\theta}{s} \right) - \Phi \left(\frac{k - \varepsilon s_0 - X\theta}{s} \right) \right) \right\}.
\end{aligned} \tag{A.2}$$

Re-parametrize the maximization problem using the one-to-one mapping:

$$\begin{pmatrix} e \\ b \\ s \end{pmatrix} = \begin{pmatrix} e \\ b_0 \\ b_1 \\ \vdots \\ b_d \\ s \end{pmatrix} \rightarrow \begin{pmatrix} (e - \varepsilon)(s_1 - s_0) \\ b_0 + s_0(e - \varepsilon) \\ b_1 \\ \vdots \\ b_d \\ s \end{pmatrix} = \begin{pmatrix} \delta \\ \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \\ s \end{pmatrix} = \begin{pmatrix} \delta \\ \theta \\ s \end{pmatrix},$$

where we use that $b = (b_0, b_1, \dots, b_d)$ and $\theta = (\theta_0, \theta_1, \dots, \theta_d)$. The parameters b_0 and θ_0 correspond to the intercept variable in X . The maximization problem becomes:

$$\begin{aligned}
& \max_{e, b, s} \mathbb{E} \left\{ \mathbb{I}\{n^* < \underline{n}\} \log \left(\frac{1}{s} \phi \left(\frac{n^* - s_0(e - \varepsilon) - Xb}{s} \right) \right) \right. \\
& \quad + \mathbb{I}\{n^* > \bar{n}\} \log \left(\frac{1}{s} \phi \left(\frac{n^* - s_1(e - \varepsilon) - Xb}{s} \right) \right) \\
& \quad \left. + \mathbb{I}\{\underline{n} \leq n^* \leq \bar{n}\} \log \left(\Phi \left(\frac{k - \varepsilon s_1 - s_1(e - \varepsilon) - Xb}{s} \right) - \Phi \left(\frac{k - \varepsilon s_0 - s_0(e - \varepsilon) - Xb}{s} \right) \right) \right\}.
\end{aligned}$$

Use that $y = n^* - \varepsilon s_0$ if $n^* < \underline{n} \Leftrightarrow y < k$, $y = n^* - \varepsilon s_1$ if $n^* > \bar{n} \Leftrightarrow y > k$, and $\underline{n} \leq n^* \leq \bar{n} \Leftrightarrow y = k$

to get

$$\begin{aligned} \max_{e,b,s} \mathbb{E} \left\{ \mathbb{I}\{y < k\} \log \left(\frac{1}{s} \phi \left(\frac{y - es_0 - Xb}{s} \right) \right) \right. \\ \left. + \mathbb{I}\{y > k\} \log \left(\frac{1}{s} \phi \left(\frac{y - es_1 - Xb}{s} \right) \right) \right. \\ \left. + \mathbb{I}\{y = k\} \log \left(\Phi \left(\frac{k - es_1 - Xb}{s} \right) - \Phi \left(\frac{k - es_0 - Xb}{s} \right) \right) \right\}, \end{aligned} \quad (\text{A.3})$$

which is the mid-censored Tobit likelihood maximization problem.

By the equivalence of the optimization problems and the one-to-one re-parametrization mapping, we have that the solution of (A.2) maps to the solution of (A.3). That is, $(\delta^*, \theta^*, s^*) = (0, \beta, \sigma)$ maps to $(e^*, b^*, s^*) = (\varepsilon, \beta, \sigma)$.

Therefore, $G_{n^*}^*(n) = \mathbb{E}[\Phi((n - Xb^*)/s^*)] = \mathbb{E}[\Phi((n - X\beta)/\sigma)] = F_{n^*}(n)$ and $G_y^*(y) = F_y(y)$.

□

A.4 Censored Quantile Regression - Proof of Identification Lemma 2

Call $D = \mathbb{I}\{Q_\tau(y | X) \neq k\}$. Let $\beta(\tau) = [\beta_0(\tau), \beta_1(\tau), \dots, \beta_d(\tau)]'$. Define $\tilde{\beta}(\tau) = [\beta_0(\tau) + \varepsilon s_0, \beta_1(\tau), \dots, \beta_d(\tau), \varepsilon(s_1 - s_0)]'$. Multiplying Equation 14 by D yields

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

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

$$\begin{aligned} D\tilde{X}'Q_\tau(y | X) &= D\tilde{X}'\tilde{X}\tilde{\beta}(\tau) \\ \mathbb{E}[D\tilde{X}'Q_\tau(y | X)] &= \mathbb{E}[D\tilde{X}'\tilde{X}\tilde{\beta}(\tau)] \\ \tilde{\beta}(\tau) &= \mathbb{E}[D\tilde{X}'\tilde{X}]^{-1} \mathbb{E}[D\tilde{X}'Q_\tau(y | X)]. \end{aligned} \quad (\text{A.4})$$

An infinite amount of data identifies the joint distribution of (y, X) . This identifies the function $Q_\tau(y | X = x)$ for every x in the support of X , and the joint distribution of $(y, Q_\tau(y | X), X, \tilde{X}, D)$. Therefore, $\tilde{\beta}(\tau)$ is identified by Equation A.4. Finally, $\varepsilon = \tilde{\beta}_{d+1}(\tau)/(s_1 - s_0)$. □

A.5 Concave Kinks

The discussion in the main text focuses on “convex kinks”, where the tax rate increases at the kink point (the tax schedule is convex). They are simply referred to as kinks. A possible extension is to consider “concave kinks”, that is, when the tax rate decreases at the kink point (the tax schedule is concave). The budget line has a kink at $Y = K$,

$$C = \mathbb{I}\{Y \leq K\}[I_0 + (1 - t_0)Y] + \mathbb{I}\{Y > K\}[I_1 + (1 - t_1)(Y - K)]. \quad (\text{A.5})$$

where $t_0 > t_1$ and $I_1 = I_0 + (1 - t_0)K$ ensures continuity.

The shape of an agent's indifference curve depends on the agent's type N^* . As N^* increases, the point of tangency of the highest indifference curve shifts to the right along the budget line; we start with just one tangency point at $Y < K$, go to two tangency points $Y < K$ and $Y' > K$ for one value of N^* , and finally to one tangency point $Y' > K$. The threshold level \underline{N} is that marginal

value of N^* between the first and second regimes. The convention we adopt is to resolve the agent's indifference in the second regime towards the smaller value of Y , that is, less labor supply.

The solution for Y has two cases in terms of N^* , whereas the convex kink case in the main text had three cases (Equation 3). Optimal income depends on the value of agent heterogeneity with respect to the threshold value \underline{N} ,

$$Y = \begin{cases} N^*(1 - t_0)^\varepsilon & , \text{ if } 0 < N^* \leq \underline{N} \\ N^*(1 - t_1)^\varepsilon & , \text{ if } \underline{N} < N^*. \end{cases} \quad (\text{A.6})$$

We find \underline{N} by solving an indifference condition. Let $\underline{Y} = \underline{N}(1 - t_0)^\varepsilon$ be biggest tangency point before the kink, and let $\bar{Y} = \underline{N}(1 - t_1)^\varepsilon$ be the smallest tangency point after the kink. The indifference condition requires the utility at \underline{Y} to be the same as the utility at \bar{Y} ,

$$\begin{aligned} I_0 + (1 - t_0)\underline{Y} - \frac{\underline{N}}{1 + 1/\varepsilon} \left(\frac{\underline{Y}}{\underline{N}} \right)^{1+1/\varepsilon} \\ = I_1 + (1 - t_1)\bar{Y} - \frac{\underline{N}}{1 + 1/\varepsilon} \left(\frac{\bar{Y}}{\underline{N}} \right)^{1+1/\varepsilon}, \end{aligned} \quad (\text{A.7})$$

where $I_0 + (1 - t_0)\underline{Y}$ is consumption when income equals $\underline{Y} < K$, and $I_1 + (1 - t_1)\bar{Y}$ is consumption when income equals $\bar{Y} > K$. We obtain an expression for \underline{N} by plugging in $\underline{Y} = \underline{N}(1 - t_0)^\varepsilon$ and $\bar{Y} = \underline{N}(1 - t_1)^\varepsilon$ in Equation A.7:

$$\underline{N} = \frac{(1 + \varepsilon)K [(1 - t_0) - (1 - t_1)]}{[(1 - t_0)^{\varepsilon+1} - (1 - t_1)^{\varepsilon+1}]} \quad (\text{A.8})$$

A continuous distribution of N^* implies a distribution of Y that has a gap $[\underline{Y}, \bar{Y}]$ of zero mass around the kink point but is continuous otherwise. The limits of this gap are: $\underline{Y} = \underline{N}(1 - t_0)^\varepsilon$ and $\bar{Y} = \underline{N}(1 - t_1)^\varepsilon$, where $\bar{Y} > \underline{Y}$ because $t_0 > t_1$. The researcher observes the distribution of Y and thus knows those two limits. Taking the difference of logs of \bar{Y} and \underline{Y} , we obtain: $\bar{y} - \underline{y} = \varepsilon(s_1 - s_0)$, where lower-case variables denote the log of upper-case variables, and $s_j = \log(1 - t_j)$, $j = 1, 2$. Therefore, the elasticity is identified,

$$\varepsilon = \frac{\bar{y} - \underline{y}}{s_1 - s_0}.$$

Without having to impose much structure on the distribution of N^* , it is possible to non-parametrically identify the elasticity, unlike the case of a convex kink (Theorem 1 by Blomquist and Newey (2017)).