ON BUNCHING AND IDENTIFICATION OF THE TAXABLE INCOME ELASTICITY

Sören Blomquist
Whitney Newey
Anil Kumar
Che-Yuan Liang

On Bunching and Identification of the Taxable Income Elasticity
Sören Blomquist, Whitney Newey, Anil Kumar, and Che-Yuan Liang
NBER Working Paper No. 24136
December 2017, Revised November 2019
JEL No. C14,H20

ABSTRACT

The taxable income elasticity is a key parameter for predicting the effect of tax reform or designing an income tax. Bunching at kinks and notches in a single budget set have been used to estimate the taxable income elasticity. We show that when the heterogeneity distribution is unrestricted the amount of bunching at a kink or a notch is not informative about the size of the taxable income elasticity, and neither is the entire distribution of taxable income for a convex budget set. Kinks do provide information about the size of the elasticity when a priori restrictions are placed on the heterogeneity distribution. They can identify the elasticity when the heterogeneity distribution is specified across the kink and provide bounds under restrictions on the heterogeneity distribution. We also show that variation in budget sets can identify the taxable income elasticity when the distribution of preferences is unrestricted and stable across budget sets. For nonparametric utility with general heterogeneity we show that kinks only provide elasticity information about individuals at the kink and we give bounds analogous to those for isoelastic utility. Identification becomes more difficult with optimization errors We show in examples how results are affected by optimization errors.

Sören Blomquist
Nationalekonomiska Institutionen
Uppsala Universitet
Box 513
751 20 UPPSALA, Sweden
soren.blomquist@nek.uu.se

Whitney Newey
Department of Economics, E52-424
MIT
50 Memorial Drive
Cambridge, MA 02142
and NBER
wnewey@mit.edu

Anil Kumar
Federal Reserve Bank of Dallas
Anil.Kumar@dal.frb.org

Che-Yuan Liang
Department of Economics
Box 513, 751 20 Uppsala
Sweden
che-yuan.liang@nek.uu.se

1 Introduction

The taxable income elasticity is a key parameter when predicting the effect of tax reform or designing an income tax. A large literature has developed over several decades which attempts to estimate this elasticity. However, due to a large variation in results between different empirical studies there is still some controversy over the size of the elasticity. A common way to estimate the taxable income elasticity has been to use variation in budget sets, often from data for several tax systems at different points in time. More recently kinks and notches for a single budget set have been used to estimate the taxable income elasticity.

We show that when the heterogeneity distribution is unrestricted the amount of bunching at a kink or a notch is not informative about the size of the taxable income elasticity. We also show that the entire distribution of taxable income for a convex budget set is not informative about the size of the elasticity. The problem is that a kink or notch probability may be large or small because of the size of the elasticity or because more or fewer individuals like to have taxable income around the kink or notch. Intuitively, for a single budget set, variation in the tax rate only occurs with variation in preferences. The conjoining of variation in the tax rate and preferences makes it impossible to nonparametrically distinguish the taxable income elasticity from heterogeneity with a single budget set. Small kinks do not solve this problem. Whether the kink probability is large or small it is possible to match the distribution of taxable income to any elasticity by specifying the distribution of preferences in a certain way.

This lack of identification can also be understood as failure of the order condition for identification, that there be as many distinct equations relating reduced form and structural parameters as there are structural parameters. There is one equation giving the kink probability as a function of two structural "parameters," the elasticity and the heterogeneity distribution. One equation is not enough to identify two structural parameters. Similarly there is one equation at each taxable income value that relates the distribution of taxable income to the distribution of heterogeneity and the taxable income elasticity is an additional structural parameter that cannot be separately identified from any of these equations or from all of them together.

Kinks do provide information about the size of the elasticity when a priori restrictions are placed on the heterogeneity distribution. The elasticity can be identified when the heterogeneity distribution is completely specified across the kink. We also give bounds on the elasticity based on bounds on the heterogeneity density. These bounds can be viewed as measures of sensitivity of the taxable income elasticity to assumptions about the heterogeneity distribution. We find in an empirical example using data like Saez (2010) that the taxable income elasticity can be quite sensitive to the heterogeneity assumptions.

[1]

The lack of identification from one budget set motivates a search for conditions that will identify the taxable income elasticity when the distribution of heterogeneity is unrestricted. Variation in the budget set is one potential source of identification that is analogous to price variation in demand analysis. Even with such variation it is not possible to achieve identification just using kinks. We do find that identification may be achieved using the entire distribution of taxable income. If the heterogeneity distribution is stable across two budget sets and at least one individual type chooses taxable income with differing tax rates for the two budget sets then the taxable income elasticity is identified for an isoelastic utility function. This is an intuitive condition for identification, that the marginal tax rate varies for some individual type. We also find that any difference in the two budget sets suffices to restrict the set of elasticities that are consistent with the data.

Similar identification problems occur for bunching at notches, where there is a discontinuity in the budget set. Like kinks, the probability of locating at a notch will not identify the taxable income elasticity because there is only one equation (for the notch probability) and two structural parameters (the elasticity and the heterogeneity distribution). Unlike kinks, the taxable income elasticity may be identified from the distribution of taxable income for a single budget set. With a notch there is a "gap" where no individual type with isoelastic utility would choose to be. The size of this gap is an identified, one-to-one function of the taxable income elasticity and so gives identification. This source of identification does depend on the form of the isoelastic utility. Also, in practice such gaps are not observed in the data and so cannot be used to identify the taxable income elasticity.

Nonidentification for the isoelastic utility clearly implies nonidentification for nonparametric utility. Nevertheless a formula for the kink probability for nonparametric utility with general heterogeneity does clarify what can be learned from kinks. We find that the kink probability is a weighted integral of compensated elasticities, allowing for income effects, of individual types that would choose to locate at the kink for a linear budget set with tax rate in between the rates to the right and left of the kink.

In data it is often observed that there is little or no bunching at kinks. This feature of the data has been accounted for by allowing departures from utility maximization, referred to as optimization errors. Hausman's (1981) specification included an additive disturbance to account for the lack of bunching. Saez (2010) considered a bunching interval that included the kink in its interior. Chetty (2012) considered optimization errors determined by the utility losses from deviating from the optimum. Optimization errors make identification more difficult because there are more things to identify from the same data. We give examples showing that optimization errors can have large effects on bunching estimators. Recently Cattaneo et al.

(2018) give results on identification of tax effects from bunching when there are optimization errors and the heterogeneity density has a known functional form across the kink.

In this paper we set aside the issue of statistical inference and focus on identification. This allows us to clarify fundamental issues of what can be learned about taxable income elasticities from data. The bounds we give can be estimated from data and we do so. It is straightforward to derive confidence intervals based on these bound as in Chernozhukov, Hong, and Tamer (2007) or Imbens and Manski (2004). To avoid additional notation and detail we omit these derivations.

Bunching estimators of the taxable income elasticity were developed and extended in influential work by Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013).[1] The Saez (2010) estimator can be interpreted as combining density values at the edges of the bunching interval with assuming that the density is linear across the kink to estimate the elasticity. Chetty et al. (2011) assume that the density is a polynomial near the kink. These results impose a known heterogeneity distribution across the kink. Imposing a known distribution of unobserved heterogeneity is unusual in the literature on identifying the effects of changing the slope of a budget set.

The rest of this paper is organized as follows. In the remainder of this Section we give a brief literature review. Section 2 lays out the model of individual behavior we consider and shows nonidentification from a single budget set. Section 3 gives bounds on the taxable income elasticity based on a single budget set and bounds on the density of heterogeneity. Section 4 applies these results to data like that used in Saez (2010). Section 5 shows how variation in budget sets helps identify the taxable income elasticity. Section 6 considers nonidentification and identification from notches. Section 7 gives extensions to nonparametric utility with general heterogeneity and considers the role of optimization errors. Section 8 gives brief conclusions. Proofs and some additional results are given in the Appendix.

Nonparametric nonidentification of compensated tax effects from a kink was shown and bounds provided in Blomquist et al. (2015). There it was shown how the amount of bunching at a kink was determined by both compensated elasticities and heterogeneity. These results are now incorporated in Section 7 of this paper. McCallum and Seegert (2017) give identification results when covariates are present and Blomquist and Newey (2017) showed that for a parametric, isoelastic utility function the distribution of taxable income for one budget set provides no information about the elasticity, provided bounds, and showed identification from budget set variation. Those results are incorporated in Sections 2, 3, and 5 of this paper. Bertanha,

---

[1] Bastani and Selin (2014), Gelber et al. (2017), Marx (2012), Le Maire and Schjering (2013) and Seim (2015) are a few of the recent papers that apply the bunching method.

McCallum. and Seegert (2017) consider nonidentification results and give bounds based on variability of the heterogeneity pdf. Bertanha, McCallum, and Seegert (2018) and Blomquist and Newey (2018) independently showed that a single budget set with a notch could identify the isoelastic taxable income elasticity.

Blomquist and Newey (2002) used variation in budget sets to nonparametrically estimate the average labor supply effect of the Swedish tax reform of 1990-1991 for scalar heterogeneity, with optimization errors. Blomquist et al. (2015) showed that these results are valid with general heterogeneity and demonstrated how to impose all the restrictions of utility maximization in estimating average labor supply. Manski (2014), and Kline and Tartari (2016) nonparametrically identified and estimated bounds on important effects, without allowing for optimization errors. Einav et al. (2017) provided recent empirical evidence on the sensitivity of policy effects to kink modeling assumptions for the elderly in Medicare Part D, where there is substantial bunching around the famous "donut hole." Van Soest (1995), Keane and Moffitt (1998), Blundell and Shephard (2012), and Manski (2014) have considered labor supply when hours are restricted to a finite set.

Before the development of bunching methods, following the seminal work by Feldstein (1995), the elasticity of taxable income was typically estimated using difference-in-differences. Saez et al. (2011) provided a review of this literature. Whereas earlier studies mostly found elasticities between one and three, the subsequent literature found elasticities between zero and one, with the benchmark estimate of approximately 0.4 by Gruber and Saez (2002) being frequently cited. Blomquist and Selin (2010), Weber (2014), Burns and Ziliak (2017), and Kumar and Liang (2017) provided recent estimated elasticities between 0.2 and 1.0.

2   Nonidentification from a Kink

We consider individuals with preferences defined over after-tax income c (value of consumption) and before tax income y (cost of effort). After and before tax income are related by $c = B(y)$, where $B(y) = y - T(y)$ for taxes $T(y)$. The utility function of an individual will be

$$U(c, y, \eta),$$

where η represents individual heterogeneity.

Figure 1 illustrates a budget set that has two linear segments with slopes (net of tax rates) $\rho_1 > \rho_2$ and a kink at K. An individual with preferences of type η will choose the point on the budget set where their utility is highest. Different individuals may have different η and so choose different taxable incomes. In this utility maximization model the variation in taxable income for one budget set comes from variation in preferences. Heterogeneity of preferences

[4]

is necessary in order to have a distribution of taxable income for a single budget set. If all individuals had the same preferences we would only observe one taxable income choice and no inference about preferences could be drawn from that.

An important preference specification is the isoelastic utility function considered by Saez (2010),

$$U\left(c, y, \eta\right) = c - \frac{\eta}{1 + \frac{1}{\beta}} \left(\frac{y}{\eta}\right)^{1+1/\beta}, \eta > 0, \beta > 0, \tag{2.1}$$

where η is a scalar. Maximizing this utility function subject to a linear budget constraint $B(y) = \rho y + R$ with slope (net of tax rate) ρ and intercept (nonlabor income) R gives the taxable income function for a linear budget constraint

$$Y(\rho, \beta, \eta) = \rho^\beta \eta.$$

The taxable income elasticity $\partial \ln Y(\rho, \beta, \eta)/\partial \ln \rho = \beta$ is constant for this specification and there is no income effect of changing R. The variable η is a scalar that represents unobserved individual heterogeneity in preferences with each η corresponding to a type of individual. We note that $Y(\rho, \beta, \eta)$ is increasing in η and ρ.

Bunching estimators estimate the taxable income elasticity from the proportion P_K of individuals at K. We follow Saez (2010) as we describe the general idea behind this approach, but omit some details that are not important for our analysis. Saez (2010) considers a counterfactual, hypothetical change in a budget constraint. In Figure 2 we consider individuals maximizing their utility for a linear budget set with slope ρ_1. Suppose next that a kink is introduced, and the slope of the budget constraint after the kink is $\rho_2 < \rho_1$. Suppose that individuals who would have been in the interval $(K, K + \Delta A]$ along the first segment now choose the kink point. We refer to the individual who would have chosen $K + \Delta A$ when there is no kink as the marginal buncher. In Figure 2 we have drawn two indifference curves for the marginal buncher. Before the (hypothetical) change in the budget constraint, the individual had a tangency on the extended segment at $K + \Delta A$. After the change in the budget constraint the individual has a tangency on the second segment at K. The discrete (e.g. arc) taxable income elasticity of the marginal buncher is

$$e = \frac{\Delta A/K}{(\rho_1 - \rho_2)/\rho_1} \tag{2.2}$$

With a kink in place we cannot observe incomes at the individual level on the extended first segment that goes beyond K and has slope ρ_1, so that we do not know ΔA. From the data we identify the proportion P_K of individuals located at the kink. Then we have

$$P_K = \int_K^{K+\Delta A} f_1(y) dy, \tag{2.3}$$

[5]

where $f_1(y)$ is the density of taxable income along the extended first segment. If $f_1(y)$ were identified we could identify ΔA from this equation. The problem is $f_1(y)$ is not identified, because it is the density for those grouped at the kink. This means that there are two structural parameters, the ΔA and the density $f_1(y)$, but only one equation involving the reduced form parameter. It is impossible to identify two structural parameters from one equation. An order condition of having as many identifying equations as structural parameters is not satisfied.

This nonidentification problem does not go away as the bunching probability becomes smaller. No matter how small P_K is there are still two unknown structural parameter ΔA and f_1. For every positive P_K there will be multiple values of ΔA and f_1 such that equation (2.3) is satisfied.

We can see nonidentification even more clearly for the isoelastic utility function where the choice for a linear budget set is $\rho^\beta \eta$. As η increases from zero the choice of taxable income will move along the first segment of the budget set. The highest value η_ℓ giving a tangency solution on the first segment satisfies $K = \eta_\ell \rho_1^\beta$. As η increases beyond η_ℓ each individual will choose the kink until η equals the lowest value η_u giving a tangency solution on the second segment, which satisfies $K = \eta_u \rho_2^\beta$. Thus the set of η where an individual will choose to be at the kink is $[\eta_\ell, \eta_u] = [K\rho_1^{-\beta}, K\rho_2^{-\beta}]$, which we refer to as the bunching window. Therefore the kink probability satisfies

$$P_K = \Pr(Y = K) = \int_{\eta_\ell}^{\eta_u} \phi(t)\, dt = \int_{K\rho_1^{-\beta}}^{K\rho_2^{-\beta}} \phi(t)\, dt, \tag{2.4}$$

where $\phi(\eta)$ is the probability density function (pdf) of η. Here we can clearly see the problem with trying to identify β from the kink. The kink probability P_K is one "reduced form" object that is identified from the data. There are two "structural parameters" that appear in this equation, the taxable income elasticity β and the pdf $\phi(\eta)$ of η, but only one equation (2.4) relating structural parameters to reduced form parameters. It is true that as β increases the right side of this equation increases. However, it is also true that for a given taxable income elasticity, the larger the mass of the preference distribution located in the bunching window $[\eta_\ell, \eta_u]$, the larger the bunching will be. It is not possible to separate those two effects using one kink equation, and so the taxable income elasticity is not identified from a kink.

Using more information about the distribution of taxable income than the kink probability does not help to identify the taxable income elasticity. The next result shows that for any distribution of taxable income with positive kink probability and any $\beta > 0$ there is a distribution of η that generates the distribution of taxable income.

THEOREM 1: *Suppose that the CDF $F(y)$ of taxable income y is continuously differentiable of order $D > 0$ to the right and to the left at K, with pdf bounded away from zero in a*

neighborhood of K, and $P_K = \Pr(Y = K) > 0$. Then for any β there exists a CDF $\Phi(\eta)$ of η such that the CDF of taxable income obtained by maximizing the utility function in equation (2.1) equals $F(y)$, and $\Phi(\eta)$ is continuously differentiable of order D.

Theorem 1 shows that for any possible taxable income elasticity we can find a heterogeneity distribution such that the CDF of taxable income for the model coincides with that for the data. Furthermore, we can do this with a heterogeneity CDF that is differentiable to the same order as the taxable income CDF. Thus we find that the entire distribution of taxable income for one budget set with one kink has no information about the size of the taxable income elasticity when the distribution of heterogeneity is unrestricted. The same result can be shown for any continuous, piecewise linear budget frontier $B(y)$ with nondecreasing marginal tax rates and each kink having positive probability.

Figure 3 illustrates the non-identification result in Theorem 1. In Panel A we present a pdf for taxable income when utility is isoelastic, the budget set is piecewise linear with one kink at $K = 20,000$, $\rho_1 = 1$, $\rho_2 = .84$, and $\beta = 0.4$. Above the kink the distribution of taxable income is Gaussian and calibrated to the histogram of observed taxable income in Figure 6A in Saez (2010), having the same mode of 40,000 USD and the same quantile at 60,000 USD. We also assumed that the distribution of η is Gaussian before the kink. Because our purpose here is just to illustrate nonidentification we did not attempt to find a heterogeneity distribution below the kink that matches well the taxable income distribution there.

In Panel B we graph the density of three heterogeneity pdf's, all of which produce the same distribution of taxable income, but with three different values of the taxable income elasticity at $\beta = .2$, $\beta = .4$ or $\beta = .8$. Below the kink the heterogeneity pdf's are the same because $\rho_1 = 1$. To highlight the differences between these pdf's we choose the density to be constant in the bunching window, equal to the value which makes the integral over the bunching window equal to P_K. We see that choosing $\beta = .2$ well below the true value $\beta = .4$ leads to a heterogeneity pdf that is large over the bunching window, while choosing $\beta = .8$ well above the true value leads to a heterogeneity pdf that is low over the bunching window. This pattern occurs because the length of the bunching window $[\eta_\ell, \eta_u] = [K\rho_1^{-\beta}, K\rho_2^{-\beta}]$ is monotonic increasing in β. In order to make the integral of the heterogeneity pdf over the bunching window be equal to P_K for each β we must have a large heterogeneity pdf across the window when β is small and a small heterogeneity pdf when β is large.

The taxable income elasticity is not identified no matter how close ρ_2 is to ρ_1. It is true that to fit the distribution of taxable income as β varies will require more extreme (larger or smaller) heterogeneity pdf's in the bunching window when ρ_2 is very close to ρ_1. Nevertheless, such heterogeneity pdf's are allowed when there is no a priori information about the heterogeneity

pdf. In Section 3 we will consider what can be learned if there are known bounds on the heterogeneity pdf's.

Showing Theorem 1 relies entirely upon nonidentification from a kink. For any particular β there is only one pdf of heterogeneity that is consistent with the taxable income distribution away from the kink. This occurs because taxable income $Y = \rho^\beta \eta$ is a scalar multiple of heterogeneity η away from the kink, where ρ is the slope of the budget frontier on a segment away from the kink. Therefore for any possible β we only have freedom to choose the heterogeneity pdf over the kink, i.e. in the bunching window $[\eta_\ell, \eta_u]$. Over the bunching window any pdf $\phi(t)$ such that equation (2.4) holds will do. To show Theorem 1 we choose such a pdf with D derivatives with D^{th} derivative equal to that of taxable income at η_ℓ and η_u. In this way for any $\beta > 0$ we can find a heterogeneity pdf such that the implied distribution for taxable income is the actual distribution of taxable income.

A kink having positive probability does identify that $\beta > 0$ when η is continuously distributed. If $\beta = 0$ then $Y = \eta$ so Y is continuously distributed by η being continuously distributed and hence $P_K = \Pr(Y = K) = 0$, so $P_K > 0$ implies $\beta > 0$. For the isoelastic utility model a positive kink probability means that we know that individuals respond to incentives, i.e. to changes in the tax rate. However Theorem 1 shows that the size of β is not identified from a single tax schedule. Since the size of the taxable income elasticity is the key parameter determining important policy questions, such as predicting the effect of tax reform or designing an income tax, Theorem 1 shows that the distribution of taxable income for a single tax schedule is not informative about these policy questions when the distribution of individual heterogeneity is unrestricted.

Saez (2010) and Chetty et al. (2011) do estimate the taxable income elasticity from a kink. By the order condition for identification we know that to identify β from the kink nothing else, other than β, must be unknown. In particular, all the information about the density of η across the kink must come from somewhere else. The Saez (2010) estimator can be obtained by assuming that the density $\phi(\eta)$ is linear over the bunching window $[\eta_\ell, \eta_u]$ and is continuous from the left at η_ℓ and from the right at η_u. To demonstrate, let $f^-(K)$ and $f^+(K)$ denote the limit of the density of taxable income at the kink K from the left and from the right, respectively. Accounting for the Jacobian of the transformation $y = \eta \rho_1^\beta$ we have $\phi(\eta_\ell) = f^-(K)\rho_1^\beta$ and

$\phi(\eta_u) = f^+(K)\rho_2^\beta$. Assuming that $\phi(\eta)$ is linear on the bunching interval we then have

$$
\begin{aligned}
P_K &= \int_{\eta_\ell}^{\eta_u} \phi(t)\, dt = \frac{1}{2}\left[\phi(\eta_\ell) + \phi(\eta_u)\right](\eta_u - \eta_\ell) \qquad (2.5)\\
&= \frac{1}{2}\left[f^-(K)\rho_1^\beta + f^+(K)\rho_2^\beta\right]\left(K\rho_2^{-\beta} - K\rho_1^{-\beta}\right)\\
&= \frac{K}{2}\left[f^-(K) + f^+(K)(\rho_1/\rho_2)^{-\beta}\right]\left[(\rho_1/\rho_2)^\beta - 1\right].
\end{aligned}
$$

This is the formula for β found in equation (5) of Saez (2010).

Here we see that the Saez (2010) formula of β can be obtained by imposing linearity of the heterogeneity density over the bunching interval $[\eta_\ell, \eta_u]$. More generally the formula assumes the trapezoid approximation

$$
\int_{\eta_\ell}^{\eta_u} \phi(t)\, dt = \frac{1}{2}\left[\phi(\eta_\ell) + \phi(\eta_u)\right](\eta_u - \eta_\ell),
$$

and will be valid under this condition. We could obtain other formulas for β by making other assumptions about $\phi(\eta)$ on $[\eta_\ell, \eta_u]$, e.g. those of Chetty et al. (2011). The elasticity that is implied by a distribution of taxable income will generally vary as $\int_{\eta_\ell}^{\eta_u} \phi(t)\, dt$ varies with the choice of functional form of $\phi(\eta)$ in the bunching window.

3 Elasticity Bounds for a Single Budget Set

From Panel B of Figure 3 we see that the heterogeneity pdf will need to be large when the true β is small or small when the true β is large. This pattern suggests that prior knowledge of bounds on the heterogeneity pdf could lead to bounds on β. In this Section we derive such bounds. We specifically consider bounds based on upper and lower bounds on the heterogeneity pdf. The bounds we give will also be satisfied if the heterogeneity pdf is monotonic on the bunching interval. We could also consider how other kinds of information could help us bound the taxable income elasticity but for simplicity we focus on upper and lower bounds on the pdf.

In constructing the elasticity bounds we follow the literature and consider a range of taxable incomes around the kink rather than just the kink itself. We do this because in taxable income data there is often not bunching at kinks but the interval probabilities can still provide information about the taxable income elasticity, as further explained in Section 7. Let y_1 and y_2 denote lower and upper endpoints for a taxable income interval that includes the kink. Let $\eta_1 = y_1\rho_1^{-\beta}$ and $\eta_2 = y_2\rho_2^{-\beta}$ denote corresponding lower and upper endpoints, and

$$
f^-(y_1) = \lim_{y \to y_1, y < y_1} f(y), \quad f^+(y_2) = \lim_{y \to y_2, y > y_2} f(y).
$$

Consider the two functions

$$
D^-(\beta) = f^-(y_1)\left[y_2\left(\frac{\rho_1}{\rho_2}\right)^\beta - y_1\right], \quad D^+(\beta) = f^+(y_2)\left[y_2 - y_1\left(\frac{\rho_2}{\rho_1}\right)^\beta\right].
$$

[9]

We have the following result:

THEOREM 2: *If there are positive scalars $\bar{\sigma} \geq 1$ and $\sigma \leq 1$ such that for $\eta \in [\eta_1, \eta_2]$,*

$$\sigma \min\{\phi(\eta_1), \phi(\eta_2)\} \leq \phi(\eta) \leq \bar{\sigma} \max\{\phi(\eta_1), \phi(\eta_2)\} \tag{3.6}$$

then the taxable income elasticity β satisfies

$$\sigma \min\left\{D^-(\beta), D^+(\beta)\right\} \leq \Pr(y_1 \leq Y \leq y_2) \leq \bar{\sigma} \max\{D^-(\beta), D^+(\beta)\}. \tag{3.7}$$

If $\phi(\eta)$ is monotonic then these bounds hold for $\sigma = \bar{\sigma} = 1$. If $\Pr(y_1 \leq Y \leq y_2) < \min\{D^-(0), D^+(0)\}$ then there is no β satisfying this equation. Otherwise the set of all nonnegative β satisfying this equation is a subset of $[0, \infty)$. In addition these bounds are sharp, meaning that if $\sigma \min\left\{D^-(\beta), D^+(\beta)\right\} > \Pr(y_1 \leq y \leq y_2)$ or $\bar{\sigma} \max\left\{D^-(\beta), D^+(\beta)\right\} < \Pr(y_1 \leq Y \leq y_2)$ there is no $\phi(\eta)$ satisfying the constraint of equation (3.6) such that $\Pr(y_1 \leq Y \leq y_2) = \int_{\eta_1}^{\eta_2} \phi(t)dt$.

The bounds given here are based on the a priori restriction in equation (3.6) for the heterogeneity pdf. For $\sigma = \bar{\sigma} = 1$ this inequality imposes the restriction that the pdf of η across the kink is bounded between the maximum and minimum of the pdf at the endpoints. A monotonic $\phi(\eta)$ would satisfy this condition and so would any other $\phi(\eta)$ that is bounded between that maximum and minimum. A restriction that the pdf is bounded above and below by a known constant could be imposed by choosing σ and $\bar{\sigma}$, when $\phi(\eta_1)$ and $\phi(\eta_2)$ are known.

Intuitively, upper and lower bounds on the heterogeneity pdf provide information about the elasticity because the length of the bunching window increases monotonically in β. The upper bound on $\phi(\eta)$ rules out values of β that make the bunching window so small that it is impossible to match $\Pr(y_1 \leq Y \leq y_2)$. Similarly, the lower bound on $\phi(\eta)$ rules out values of β that make the bunching window too large to match $\Pr(y_1 \leq Y \leq y_2)$.

The bounds will tend to be wider when $f^+(y_2)$ is further from $f^-(y_1)$. Smaller differences in $f^+(y_2)$ and $f^-(y_1)$ are sometimes evident on the right side of the mode of the taxable income distribution. Of course if one thought that $\bar{\sigma}$ and σ were further apart on the right side of the mode then the bounds could be just as wide on the right side of the mode.

A known relationship between the pdf of η outside the kink and across the kink could give equation (3.6). For example if a nonparametric estimator of the pdf of taxable income is monotonically increasing or decreasing on both sides of the kink and it was known that monotonicity in the heterogeneity pdf outside the kink implied monotonicity across the kink then (3.6) would hold. In this way information about the taxable income distribution away

from the kink could be used to provide information about the heterogeneity pdf across the kink. Note though that the availability of such information could depend on where the kink is located. If the kink is located so that the pdf of taxable income is increasing below the kink and decreasing above the kink it might be unreasonable to think that the heterogeneity pdf is monotonic on $[\eta_1, \eta_2]$.

These bounds are different than in much of the partial identification literature in economics. Many economic parameter bounds make use of economic behavior. The bounds here are based on knowledge about the heterogeneity pdf which is not restricted by economics.

One could think of these bounds as a sensitivity check on how the results are affected by allowing variation in the heterogeneity pdf in the bunching window. This could make them difficult to interpret because different individuals might have different ideas about upper and lower bounds on the heterogeneity pdf. Also, these bounds are correct only if there are no optimization errors outside the bunching interval, as further discussed in Section 7.

To estimate the bounds we can plug in nonparametric estimators $\hat{f}^-(y_1)$ and $\hat{f}^+(y_2)$ to obtain

$$
\begin{aligned}
\hat{D}^-(\beta) &= \hat{f}^-(y_1)\left[y_2\left(\rho_1/\rho_2\right)^{\beta} - y_1\right], \\
\hat{D}^+(\beta) &= \hat{f}^+(y_2)\left[y_2 - y_1\left(\rho_2/\rho_1\right)^{\beta}\right].
\end{aligned}
$$

Estimated bounds for β are $\hat{\beta}_{\ell}$ and $\hat{\beta}_u$ that solve

$$
\begin{aligned}
\bar{\sigma}\max\left\{\hat{D}^-\left(\hat{\beta}_{\ell}\right), \hat{D}^+\left(\hat{\beta}_{\ell}\right)\right\} &= \Pr(y_1 \widehat{\leq Y} \leq y_2), \\
\underline{\sigma}\min\left\{\hat{D}^-\left(\hat{\beta}_u\right), \hat{D}^+\left(\hat{\beta}_u\right)\right\} &= \Pr(y_1 \widehat{\leq Y} \leq y_2).
\end{aligned}
$$

4   An Application to US Data

In this Section we reanalyze the federal income tax application in Saez (2010, Figure 6A) for married tax filer for the years 1960 to 1969. Like Saez (2010), we use Individual Public Use Microdata files from 1960 to 1969 released by the Statistics of Income Division of the Internal Revenue Service. We access the data through the NBER Unix System. The 1960-1969 data currently existing on NBER servers differ slightly from those used by Saez (2010), but yield estimates similar to Saez (2010). These files are stratified random samples of the entire tax filing population of the United States with oversampling of high-income individuals. Like Saez (2010), years 1961, 1963, and 1965 are excluded from analysis, as public use files for these years are not available. The sample size ranges from approximately 86,000 to 100,000 in various years between 1960 and 1969. The public-use files come with some well-known limitations, e.g., individual identifiers are removed, some variables are blurred to prevent public disclosure, and demographic

information is missing. Our definition of taxable income is identical to Saez (2010)—Adjusted Gross Income (AGI) net of exemptions and deductions. Following Saez (2010), we focus on tax returns with taxable income (in 2008 dollars) between -20,000 and 65,000. All bunching estimation is based on binned data, with $100 wide bins, constructed using population weights.

In Figure 4 we reproduced the taxable income distribution but shifted the domain upwards by 20,000 to avoid negative values, for which the elasticity is not defined. At the first 20,000 kink that Saez focused on, the marginal tax rate went up from 0 to 16 percent (across-year average) for married tax filer for the years 1960 to 1969. There are additional smaller kinks after 45,000. See Saez (2010) for a detailed description of the institutional setting. We think this application is important for several reasons. First, it concerns a large part of the population. Second, the elasticity estimates that Saez (2010) found for this application were less sensitive across specifications compared to other applications in which it was harder to plausibly quantify the bunching probability with any precision. Third, similar low income tax elasticities for broad population groups have been found in other countries (e.g., Chetty et al., 2011; Bastani and Selin, 2014).

It is interesting to observe that in Figure 4 the heights of the histograms are substantially different from each other even for neighboring intervals. This feature of the data suggests that bounds which allow for substantial variance in the density over the bunching window might be appropriate.

In practice, bunching methods proceed in two steps: 1) Extrapolate the pdf of taxable income $f(y)$ outside specified bunching limits $y_1 < y_2$ into the bunching interval $[y_1, y_2]$. This gives estimated densities from the left and right, $\hat{f}^-(K)$ and $\hat{f}^+(K)$ of the kink. Also it gives an estimator $\hat{P}_K$ of the bunching probability as the excess mass, which is the difference between the integral of observed and extrapolated pdfs inside the bunching bands; 2) Use the estimates $\hat{f}^-(K)$, $\hat{f}^+(K)$, and $\hat{P}_K$ with a formula that relates these parameters to the elasticity β under certain functional form assumptions. We have thus far focused on the second step assuming known values for $f^-(K)$, $f^+(K)$, and P_K, but estimation as described in the first step is needed in an empirical application.

For step 1), Saez (2010) used a procedure that extrapolated the density inside the left (of the kink) bunching band using the average density in an adjacent (equal sized) outer left bunching band, and analogously to the right of the kink. For step 2), he used the trapezoid formula in equation (2.5) assuming that $\phi(\eta)$ is linear in η inside $[\eta_\ell, \eta_u]$. The preferred estimate in Saez (2010) uses $y_1 = K - 1500$ and $y_2 = K + 1500$ and also tries intervals where 1500 is replaced by 1000, 2000, 3000, and 4000. Chetty et al. (2011) fitted a high-order polynomial using the entire taxable income distribution outside the bunching bands and extrapolated inside $[y_1, y_2]$.

[12]

In step 2) they used a small kink elasticity formula from Saez (2010).

In Table 1, we report elasticity estimates using the methods in Saez (2010) and in Chetty et al. (2011) varying the bunching bandwidth between 1000 and 4000. In Table 1 the Saez (2010) estimate .157 for the bunching bandwidth of 2000 is quite close to the .170 that Saez (2010) obtained for bunching bandwidth 1500. We also provide estimates of the bounds on β obtained by setting $\sigma = \bar{\sigma} = 1$ and assuming equation (3.6) holds, which includes the case where the heterogeneity pdf is monotonic. Table 1 shows quite large differences across the methods. Furthermore, the estimated bounds are quite wide and the width increases with $y_2 - y_1$.

Table 1: Comparing Saez, Chetty et al., and bounds for Saez data

Bunch band	Saez	Chetty et al.	Lower bound	Upper bound
1000	.088	.156	.036	.144
2000	.157	.247	0	.385
3000	.335	.378	0	.908
4000	.390	.454	0	1.516

The difference in density estimates explains the width of the bounds in Table 1. One could construct tighter bounds by putting more restrictions on the heterogeneity pdf. However, all such bounds are based entirely on prior information when there is only a single budget set. As we have discussed, the data provides no information on the heterogeneity density for individuals at the kink.

5 Identification from Budget Set Variation

Given the identification difficulties for bunching it seems important to consider what will identify the taxable income elasticity. We know from Section 3 that identification requires more information than the distribution of taxable income for a single budget set when the distribution of heterogeneity is unrestricted. In this section we show identification can be achieved with budget set variation and stable preferences. We specifically consider how much budget set variation suffices for identification.

Stable preferences means that budget sets are statistically independent of preference. Independence can be a reasonable assumption in some settings, such as the Swedish tax reform analyzed in Blomquist and Newey (2002), where there is little a priori concern about correlation between the timing of reform and tastes for work. Independence would be more problematic across budget sets within a cross-section. For instance, in US data where budget sets are affected by housing choices buget sets and preferences seem likely to be dependent. Also, correlation of nonlabor income and preferences could be a source of budget set endogeneity. In some cases control functions can be used to control for dependence between budget sets and preferences,

as in Blomquist et al. (2015). For simplicity we focus here on what can be learned from two budget sets with stable preferences.

The elasticity cannot generally be identified only from kinks, even for multiple budget sets. Intuitively, each kink only gives one equation and each kink probability depends on both β and $\phi(\eta)$. Because $\phi(\eta)$ is an unknown function the integral over the heterogeneity distribution will generally be distinct for each kink, so that there are still more unknown parameters than equations, and the order condition is not satisfied. When there are known bounds on the heterogeneity pdf multiple kinks could help shrink the identified set, since the true β must satisfy the inequalities of Section 3 for each kink.

The elasticity β can be identified from the entire distribution of taxable income for just two budget sets. An order condition again provides insight. If there are two budget sets, the data identifies CDF's of taxable income, one corresponding to each budget set. For isoelastic utility there is one unknown CDF (of η), and one unknown parameter, the taxable income elasticity β. Two CDF's may be more than enough to identify one CDF and one parameter. In fact, the taxable income elasticity may be overidentified with strong restrictions being imposed on the distribution of taxable income across the two budget sets.

We continue to work with budget sets that are piecewise linear with one kink and a declining net of tax rate. To describe the identification results let $F(y)$ and $\tilde{F}(y)$ be the distributions of taxable income for two budget sets. Also, let $\rho(y)$ and $\tilde{\rho}(y)$ be the respective slopes on the budget sets at any y that is not a kink and be equal to the slope from the right at the kink (i.e. the slope of the upper segment). Here $\rho(y)$ and $\tilde{\rho}(y)$ are the marginal net of tax rates for an increase in y at each taxable income for the respective budget sets. The following result gives a sufficient condition for identification of β.

THEOREM 3: *Suppose that i) taxable income is chosen by maximizing isoelastic utility; ii) the CDF of η is the same for both budget sets; iii) the CDF of η is continuous and strictly monotonic increasing on $(0, \infty)$. Then if there exists y and $\tilde{y}$ such that $F(y) = \tilde{F}(\tilde{y})$ and $\rho(y) \neq \tilde{\rho}(\tilde{y})$ then*

$$\beta = \frac{\ln(\frac{\tilde{y}}{y})}{\ln(\frac{\tilde{\rho}(\tilde{y})}{\rho(y)})}.$$

Here we see that β is identified from any pair of taxable incomes y and $\tilde{y}$ with the same value of the CDF for the two budget sets but a different marginal tax rate. To interpret this result, recall from Section 2 that on any linear segment with slope ρ the taxable income choice $\eta\rho^\beta$ is strictly increasing in η. Consequently, identical quantiles of F and $\tilde{F}$ correspond to identical quantiles of the distribution of η, which correspond to identical individual types η. Therefore

[14]

the condition $F(y) = \tilde{F}(\tilde{y})$ is a statement that y and $\tilde{y}$ correspond to the same individual type. Since $\rho(y)$ is the marginal net of tax rate, the sufficient condition for identification in Theorem 3, that $F(y) = \tilde{F}(\tilde{y})$ and $\rho(y) \neq \tilde{\rho}(\tilde{y})$, is that there is an individual type η that faces different marginal tax rates at the values of taxable income chosen for the respective budget sets. This is an intuitive condition for identification of the taxable income elasticity, that some individual type has different marginal tax rates at the choice for each of the two budget sets.

The fact that β will be identified from any y and $\tilde{y}$ satisfying the conditions of Theorem 3 can impose many testable restrictions. Each such y and $\tilde{y}$ will lead to a value for β. The restrictions could be tested by comparing corresponding different such values of β.

A simple sufficient condition for identification from Theorem 3 is that either the slope of the first segments or the slope of the second segments are different for the two budget sets.

COROLLARY 4: *If conditions i)-iii) of Theorem 3 are satisfied and either the first segments or the second segments of the two budget sets have different slopes then β is identified.*

Figure 5, Panel A illustrates this result. In this diagram the first segment and the kink location are the same for the two budget sets but the second segment has different slope. The figure shows the indifference curves for an individual type (value of η) choosing taxable income on the upper linear segment for both budget sets. Such a type must exist because the pdf of taxable income will be positive along the entire upper segment for both budget sets by conditions ii) and iii) of Theorem 3. That type has a different marginal tax rate at the taxable incomes chosen for the two budget sets and so β is identified from the formula in Theorem 3.

Two budget sets can also be different because only the kink is different. To state an identification result when only the kink is different let K and $\tilde{K}$ be the kinks for the two budget sets and suppose without loss of generality that $K < \tilde{K}$. Also, define $\tilde{F}_-(\tilde{K}) = \lim_{y \to \tilde{K}, y < \tilde{K}} \tilde{F}(y)$ to be the limit from the left of the CDF $\tilde{F}(y)$ for the second budget set at the kink.

THEOREM 5: *Suppose that conditions i)-iii) of Theorem 3 are satisfied. If $\tilde{F}_-(\tilde{K}) \geq F(K)$ then there is a unique y^* with $K \leq y^* < \tilde{K}$, $F(y^*) = F_-(\tilde{K})$, and $\beta = \ln(\tilde{K}/y^*)/\ln(\rho_1/\rho_2)$. Also if $\tilde{F}_-(\tilde{K}) < F(K)$ then $\beta > \underline{\beta} = \ln(\tilde{K}/K)/\ln(\rho_1/\rho_2)$ and for any $b > \underline{\beta} = \ln(\tilde{K}/K)/\ln(\rho_1/\rho_2)$ there exists a heterogeneity CDF $\Phi^b(\eta)$ that is continuous and strictly monotonic increasing such that when $\beta = b$ the CDF of taxable income is $F(y)$ for the budget set with kink K and $\tilde{F}(y)$ for the budget set with kink $\tilde{K}$.*

This result shows that $\tilde{F}_-(\tilde{K}) \geq F(K)$ is a necessary and sufficient condition for identification of β from two budget sets where only the kink is different. This condition can be interpreted in terms of the "marginal buncher" for the budget set with kink K. Letting Φ denote the CDF

of heterogeneity, note that on the left linear segment of the budget set with kink $\tilde{K}$ we have $\tilde{F}(y) = \Phi(\rho_1^{-\beta}y)$, so by continuity $\tilde{F}_-(\tilde{K}) = \Phi(\rho_1^{-\beta}\tilde{K})$. Similarly, to the right of the kink K for that budget set we will have $F(y) = \Phi(\rho_2^{-\beta}y)$, so by continuity we have $F(K) = \Phi(\rho_2^{-\beta}K)$. By strict monotonicity of Φ it follows that $\tilde{F}_-(\tilde{K}) \geq F(K)$ is equivalent to

$$\rho_1^{-\beta}\tilde{K} \geq \rho_2^{-\beta}K.$$

As discussed in Section 2, $\eta_u^K = \rho_2^{-\beta}K$ is smallest value of η such that such an individual would choose to be on the linear segment with slope ρ_2 when the kink is K. Thus η_u^K is the type of the "marginal buncher," who would locate at K even if the budget set were linear with slope ρ_2. The identification condition $\tilde{F}_-(\tilde{K}) \geq F(K)$ is thus equivalent to

$$\tilde{K} \geq \rho_1^{\beta}\eta_u^K.$$

Note that $\rho_1^{\beta}\eta_u^K$ is the taxable income that the marginal buncher (type η_u^K) would choose along an extended first segment where the slope is everywhere ρ_1. Thus, the necessary and sufficient identification condition is that the right kink $\tilde{K}$ is located beyond the point that the marginal buncher would choose along the extended first segment for the left kink K.

Figure 5, Panel B shows how the kink shift identifies β when $\tilde{K} \geq \rho_1^{\beta}\eta_u^K$. The indifference curves are those of the marginal buncher with $\eta = \eta_u^K$. This type has slope of the indifference curve equal to ρ_2 at the taxable income for the budget set with kink K and slope $\rho_1 \neq \rho_2$ at the taxable income choice for budget set with kink $\tilde{K}$. Thus identification of β follows by Theorem 3. If $\tilde{K} < \rho_1^{\beta}\eta_u^K$ then the marginal buncher would choose the kink $\tilde{K}$ where the marginal net of tax rate is the same as at K, and so will not identify β. The nonidentification result shows that none of the other types that locate at K for that budget set serve to identify β either when $\tilde{K} < \rho_1^{\beta}\eta_u^K$.

We see from Theorem 3 that not every pair of distinct budget sets will serve to identify β for isoelastic utility. We also find that a shift in the kink does provide some information in the form of a lower bound on the elasticity, where that lower bound is larger the bigger the shift in the kink and the smaller the ratio of the two tax rates.

It is also interesting to note that a shift in the kink implies strong, overidentifying restrictions on the CDF of taxable income, that it coincides for each budget set for $y < K$ and $y \geq \tilde{K}$. This restriction holds even if the budget sets have different intercepts because there is no income effect for isoelastic utility.

[16]

6 Notches

Notches have also been used to estimate the taxable income elasticity beginning with Kleven and Waseem (2013). A notch occurs at an income value where there is a discontinuity in the budget set so that the average tax rate changes. Figure 6 illustrates such a budget set with a drop in the average tax rate at the notch point K. The marginal tax rate could also change at a notch, though for notational convenience Figure 6 includes no change in the marginal tax rate. In this Section we show that, similarly to kinks, bunching at a notch provides no information about the size of the elasticity β for isoelastic utility when heterogeneity is unrestricted. We also show that a notch can identify the elasticity β for an isoelastic utility but that identification does not seem useful in practice.

Figure 6 shows how bunching at a notch can occur for isoleastic utility. As η increases from zero the choice of taxable income will move along the first segment of the budget set. The highest value η_ℓ giving a tangency solution on the first segment satisfies $K = \eta_\ell \rho_1^\beta$. As η increases beyond η_ℓ each individual will choose the kink until η equals the value $\eta_3(\beta)$ where the indifference curve passing through the notch point on the first segment is tangent to the second segment. The notch probability is

$$P_K = \int_{K\rho_1^{-\beta}}^{\eta_3(\beta)} \phi(\eta)d\eta. \tag{6.8}$$

The probability of a notch is not informative about β for similar reasons that the probability of kink is not informative. For any $P_K > 0$ and β we can choose $\phi(\eta)$ such that equation (6.8) is satisfied. Thus the probability of a notch provides no information about the size of β.

Unlike a kink, for isoelastic utility the entire distribution of taxable income does vary with β in such a way that β can be identified. To explain this result let $y_3(\beta)$ denote the tangency point on the upper segment for the indifference curve that passes through the notch point. The interval $[K, y_3(\beta)]$ is a "gap" region that no type would choose to be in. The upper limit $y_3(\beta)$ of this gap region can be shown to be a one-to-one function of β, and so β is identified. The value of $y_3(\beta)$ could be estimated as the smallest value of taxable income that exceeds K, an order statistic type of estimator, which would be consistent. We show this result for a notch where the slope of the budget frontier is ρ_1 before the notch and ρ_2 after the notch.

THEOREM 6: *Suppose that i) taxable income is chosen by maximizing isoelastic utility; ii) the CDF of η is continuous and strictly monotonic increasing on $(0, \infty)$. If the budget frontier is piecewise linear with one notch where there is an increase in the average tax rate (i.e. a jump down in the budget frontier) then β is identified.*

[17]

This identification method does depend strongly on the isoelastic specification being correct. It will fail to be useful if there is no gap region, which is typically the case in applications. The absence of a gap region could occur if the isoelastic model is not correct, there are optimization or measurement errors in taxable income, there is unobserved variation in taxable income, or for other reasons.

It is interesting to note that a notch provides for a nonparametric test of utility maximization. In the taxable income specification we are considering utility is increasing in consumption c and decreasing in y. Therefore, any point on the second budget segment with lower consumption than at the notch will not be chosen by a rational agent. Any agent would be worse off to be on this part of the budget set than they would be at the notch, so no agent would locate there.

7   Extensions

In this Section we consider how the results for isoelastic utility extend to nonparametric models with general heterogeneity and optimization errors.

7.1   Nonparametrics With General Heterogeneity

We give here an interpretation of the kink probability in terms of the compensated tax effects while allowing for general heterogeneity, the functional form of the utility to be unknown, and income effects. To do so we consider a utility function $U(c, y, \eta)$ where the functional form is unknown and where η can have any dimension. For the budget frontier $B(y)$, giving the largest amount that can be consumed for every taxable income y, the choice of taxable income will be

$$Y(B, \eta) = \arg\max_{y \geq 0} U(B(y), y, \eta).$$

In general the choice $Y(B, \eta)$ depends on the entire budget frontier, which is why B appears as an argument of $Y(B, \eta)$. Similarly, the distribution of taxable income for any particular budget frontier will depend on B. As in Section 2, the variation in η will result in variation in taxable income.

When $B(y)$ is concave there is a characterization of the taxable income distribution that does not depend on the whole budget frontier. For simplicity we give this characterization when $B(y)$ is piecewise linear and continuous with slope of the jth segment being ρ_j and $\rho_j > \rho_{j+1}$, $j = 1, ..., J - 1$. Let $\rho(y)$ denote the limit from the right of the slope of $B(y)$, which is ρ_j if y is in segment j and at a kink is equal to the slope of the segment just to the right of the kink. Also let $R(y) = B(y) - y\rho(y)$ be the intercept of the line passing through $(y, B(y))$ with slope

$\rho(y)$, i.e. the "virtual" nonlabor income. Let

$$F(y|\rho, R)$$

denote the CDF of taxable income for a linear budget constraint with slope ρ and intercept R. The following intermediate result is useful in characterizing a kink probability. For ease of exposition the Assumptions of this result and the one to follow are given in Appendix B.

THEOREM 7: *If Assumptions B1-B2 are satisfied, $B(y)$ is piecewise linear and continuous, and $\rho_j > \rho_{j+1}$, $(j = 1, ..., J-1)$ then $\Pr(Y(B, \eta) \le y) = F(y|\rho(y), R(y))$.*

Here we find that for piecewise linear, concave $B(y)$, the CDF is that for linear after tax income with slope $\rho(y)$ and nonlabor income $R(y)$. This theorem is a distributional result corresponding to the observation of Hausman (1979) that linear budget sets can be used to characterize choices when preferences are convex and $B(y)$ is continuous with increasing marginal tax rates. We note that this result allows for general heterogeneity where the dimension of η is unknown. A more general version of this result was given in Blomquist et al. (2015) and is now given in the Appendix of this paper. We will use this result to characterize kink probabilities in terms of compensated elasticities.

To characterize a kink, without loss of generality let K be the first kink so that ρ_1 is the slope from the left and ρ_2 from the right at K. Consider ρ with $\rho_2 \le \rho \le \rho_1$ and and let $R(\rho) = R_1 + K(\rho_1 - \rho)$ be the virtual income for the linear budget set with slope ρ passing through the kink. Assuming that $F(y|\rho, R(\rho))$ is differentiable in y let $f(y|\rho, R(\rho)) = \partial F(y|\rho, R(\rho))/\partial y$ denote the corresponding pdf at y. Let $Y(\rho, R, \eta)$ denote the taxable income choice for a linear budget frontier with slope ρ and intercept R and let

$$\beta(\rho, K) = E\left[\frac{\rho}{K}\left\{\frac{\partial Y(\rho, R, \eta)}{\partial \rho} - K\frac{\partial Y(\rho, R, \eta)}{\partial R}\right\}\bigg| Y(\rho, R, \eta) = K\right]_{R=R(\rho)}, \quad \tilde{\phi}(\rho) = \frac{K}{\rho}f(K|\rho, R(\rho)),$$

where the expectation is taken over the distribution of η and existence of derivatives is imposed in Assumption B3 given in Appendix B. Here $\beta(\rho, K)$ is the average compensated taxable income elasticity for those individuals facing a linear budget frontier with slope ρ that passes through the kink point, who choose to locate at the kink point. Also $\tilde{\phi}(\rho)$ is K/ρ times the pdf of taxable income at K for the taxable income distribution implied by a linear budget frontier. The following result gives a formula for the kink probability in terms of $\beta(\rho, K)$ and $\tilde{\phi}(\rho)$:

THEOREM 8: *If Assumptions B1-B3 are satisfied then*

$$P_K = \int_{\rho_2}^{\rho_1} \tilde{\phi}(\rho)\beta(\rho, K)d\rho. \tag{7.9}$$

[19]

The compensated elasticity $\beta(\rho, K)$ appears inside the integral because virtual income is being adjusted as ρ changes to stay at the kink. The virtual income adjustment needed to remain at the kink corresponds locally to the income adjustment needed to remain on the same indifference curve, as shown by Saez (2010). The pdf $f(K|\rho, R(p))$ is a counterfactual pdf, what the pdf at the kink would have been at K if the budget frontier were linear with slope ρ and passed through the kink. The formula for P_K in Theorem 8 bears some resemblance to the kink probability formulas in Saez (2010) but differs in important ways. Theorem 8 is nonparametric, takes explicit account of general heterogeneity, and allows for income effects unlike the Saez (2010) results, which are local or parametric and account for heterogeneity implicitly. Theorem 8 and the discussion in the next two paragraphs was given in Blomquist et al. (2015) and is now included here. This was the first explicit result on nonidentification of a compensated elasticity effect from a kink of which we are aware.

Theorem 8 helps clarify what can be nonparametrically learned from kinks. First, the compensated effects that enter the kink probability are only for individuals (i.e. values of η) who would choose to locate at the kink for a linear budget set with $\rho \in [\rho_2, \rho_1]$. Thus, using kinks to provide information about compensated effects is subject to the same issues of external validity as, say, regression discontinuity design (RDD). As RDD only identifies treatment effects for individuals at the jump point so kinks only provide information about compensated effects for individuals who would locate at the kink.

Second, the kink probability depends on both the average compensated elasticity $\beta(\rho, K)$ and on a pure heterogeneity effect $\tilde{\phi}(\rho)$. Intuitively, a kink probability could be large because the elasticities are large or because preferences are distributed in such a way that the pdf value $f(K|\rho, R(\rho))$ for counterfactual $\rho \in [\rho_2, \rho_1]$. Consequently it is not possible to separately identify compensated tax effects and heterogeneity effects from a kink. For example consider the weighted average elasticity

$$\bar{\beta}(K) = \frac{\int_{\rho_2}^{\rho_1} \tilde{\phi}(\rho)\beta(\rho, K)d\rho}{\int_{\rho_2}^{\rho_1} \tilde{\phi}(\rho)d\rho} = \frac{P_K}{\int_{\rho_2}^{\rho_1} \tilde{\phi}(\rho)d\rho}$$

Evidently $\bar{\beta}(K)$ depends on the denominator $\int_{\rho_2}^{\rho_1} \tilde{\phi}(\rho)d\rho$ which is needed to normalize so that $\bar{\beta}(K)$ is a weighted average of elasticities. This denominator is not identified because all $\rho \in (\rho_2, \rho_1)$ are counterfactual and so not observed in the data. Indeed, $\tilde{\phi}(\rho)$ can be any positive function over the interval (ρ_2, ρ_1), so the denominator can vary between 0 and ∞, implying that any $\bar{\beta}(K) \in (0, \infty)$ is consistent with the data. Here we see that identification from a kink for a nonparametric specification is similar to that for an isoelastic utility. Just as for the isoelastic utility, a kink is uninformative about the size of an average taxable income elasticity for the nonparametric model with general heterogeneity.

If the pdf at the kink were linear in ρ over $[\rho_2, \rho_1]$ then the denominator in $\bar{\beta}(K)$ would be

$$\int_{\rho_2}^{\rho_1} \tilde{\phi}(\rho)d\rho = K[\int_{\rho_2}^{\rho_1}\{\rho^{-1}a + b\}d\rho] = K[\ln(\rho_1/\rho_2)a + b(\rho_1 - \rho_2)],$$

$$a = \frac{\rho_1 f^+(K) - \rho_2 f^-(K)}{\rho_1 - \rho_2}, \quad b = \frac{f^-(K) - f^+(K)}{\rho_1 - \rho_2},$$

where $f^+(K) = \lim_{y \longrightarrow K, y > K} f(y)$ and $f^-(K) = \lim_{y \longrightarrow K, y < K} f(y)$. Then the average compensated elasticity $\bar{\beta}(K)$ would be identified from equation (**??**). The resulting estimator is different that Saez (2010) in assuming linearity of the pdf in ρ rather than in η and in allowing for general heterogeneity and unknown functional form for the utility function.

One can also bound the nonparametric average elasticity $\bar{\beta}(K)$ if it is known a priori that there are positive scalars $\sigma \leq 1$ and $\bar{\sigma} \geq 1$ such that for $\rho \in [\rho_2, \rho_1]$,

$$\sigma \min\{f^-(K), f^+(K)\} \leq f(K|\rho, R(\rho)) \leq \bar{\sigma} \max\{f^-(K), f^+(K)\},$$

where monotonicity is included as a special case where $\sigma = \bar{\sigma} = 1$. Note that this assumption is different than the previous assumption on the pdf of η for the isoelastic utility. It does provide one way to limit the variation in $f(K|\rho, R(\rho))$. The bounds on $f(K|\rho, R(\rho))$ imply that

$$(K/\rho)\sigma \min\{f^-(K), f^+(K)\} \leq \tilde{\phi}(\rho) \leq (K/\rho)\bar{\sigma} \max\{f^-(K), f^+(K)\}.$$

Integrating and inverting these bounds gives

$$\frac{P_K}{K \ln(\rho_1/\rho_2)\bar{\sigma} \max\{f^-(K), f^+(K)\}} \leq \bar{\beta}(K) \leq \frac{P_K}{K \ln(\rho_1/\rho_2)\sigma \min\{f^-(K), f^+(K)\}}. \quad (7.10)$$

The bounds here can be quite wide when the estimated densities from the left and right are far apart or σ is very different than $\bar{\sigma}$, similar to the bounds for the isoelastic utility function. One could tighten the bounds by restricting the pdf.

Table 2: Nonparametric Bounds

Bunch band	Linear	Lower bound	Upper bound
1000	.088	.081	.097
2000	.156	.133	.190
3000	.329	.259	.461
4000	.381	.278	.629

In Table 2, we report estimates of the nonparametric average elasticity using equation (**??**) for a taxable income pdf that is linear in ρ and non-parametric bounds using equation (7.10). To apply the formula, we need to know P_K, $f^-(K)$, and $f^+(K)$. We replace these objects with estimators like those of Saez (2010) to construct estimators. We see that the point estimates in Table 2 are close to those in Table 1. However, the bounds in Table 2 are narrower.

A main reason for the differences between the bounds here and those in Section 4 is that the nonparametric ones depend on $f^-(K)$ and $f^+(K)$ whereas the isoelastic ones depend on $f^-(y_1)$ and $f^+(y_2)$. To estimate the nonparametric bounds in applications we need to extrapolate the density towards the kink. Here we have done this.

7.2   Optimization Errors

Optimization errors are variations in taxable income away from utility maximization. There are multiple ways to model optimization errors. For labor supply Burtless and Hausman (1978) used random errors that are independent of the budget set and preferences. Chetty (2012) used optimization errors determined by the utility losses from deviating from the optimum. For concave budget frontiers and strictly convex preferences movements away from the optimum decrease utility (e.g. Lemma B2 of Appendix B), so both of these approaches have the property that optimization errors with higher utility costs are less likely. Our empirical knowledge of the form of optimization errors is meagre and there is no obvious choice for how to model them. Motivated by simplicity for agents' choices and for modeling we consider here optimization errors as additive random errors.

Such errors can result from several sources. They could be measurement errors, though many modern administrative data sets are thought to be accurate enough that measurement error is low. Optimization errors could result from the fact that only some points on the budget constraint are available and we ignore these constraints in our specification of the model. Unanticipated changes in taxable income could result in optimization errors when there are lags in adjustment. Preference variation that is not fully captured by the allowed heterogeneity might also result in optimization error. Unanticipated changes in income or preferences might be modeled as an additive, mean zero disturbance or a multiplicative disturbance with unit expectation.

Optimization errors might also result from lags in adjustment to changes in the budget set. This could produce different kinds of optimization errors depending on the nature of the change. For example, an increase in the marginal tax rate above the kink that is not adjusted to would mean that there are fewer observations at the kink than there should be. The resulting distribution of optimization errors would have a mean greater than zero and the error would only affect those at or above the kink. The opposite error, with a mean less than zero, could occur when there is a decrease in the marginal tax rate above a kink.

The presence of optimization errors makes identification more difficult because with optimization errors there is another unknown object, the distribution of the error. The kink probability will depend on the distribution of optimization errors and so there will be even

more unknown parameters to account for in bunching estimation. Cattaneo et al. (2018) give results on identification and estimation from kinks when there is an additive optimization error that is independent of the budget set. Identification from variation in budget sets would also be more difficult when there are optimization errors. One virtue of identifying tax effects from the expected value of the choice is that additive and/or multiplicative optimization errors can be allowed for, as in Blomquist and Newey (2002) and Blomquist et al. (2015).

In the bunching literature optimization errors are accounted for by estimating the kink probability from a bunching interval $[y_1, y_2]$ that contains the kink, as described earlier in Section 4. Using such a bunching interval decreases bias in the estimator of β when y_1 and y_2 are close enough to K and the true β is positive. To see this note that with continuously distributed optimization errors the probability of K will be zero, so that bunching methods estimate $\beta = 0$. For any interval with $y_1 < K < y_2$ these method will estimate $\beta > 0$. Therefore, for any bunching method that is continuous in y_1 and y_2 the estimated β will be positive and less than the true β when y_1 and y_2 are close enough to K, and so have less bias than $\beta = 0$. However, using a bunching interval generally will not fully account for optimization errors, i.e. the presence of optimization errors will lead to bias in the estimator of β.

Because bunching estimators are complicated nonlinear estimators it is difficult to derive the bias of bunching estimators resulting from optimization errors. Nevertheless it would be useful to understand the effects of optimization errors on bunching estimators. To illustrate what those effects can be we consider here examples of what is estimated by bunching estimators. In each of these examples the distribution of taxable income is that implied by an isoelastic utility, with $\beta = .2$, $K = 1000$, $\rho_1 = .7$, and $\rho_2 = .5$. This is a large kink which, according to the literature, should help identify the taxable income elasticity. We take the distribution of taxable income to be the empirical distribution of two million observations that are i.i.d. with the distributions described below. To obtain the estimated object we use the program bunchr, written by Itai Trilnick in the programming language R (that can be accessed at https:/CRAN.R-project.org/package=bunchr). That program uses the isoelastic utility specification, takes the heterogeneity density near the kink to be a polynomial similar to Chetty et al. (2011), uses a default bunching interval, and calculates the Chetty et al. (2011) estimator. Alternative ways of determining the bunching interval, such as visual inspection, could be used. As we have shown, without prior restrictions on the distribution of heterogeneity there is no information in the distribution of taxable income about the size of the elasticity. Thus, any method of determining the bunching interval cannot help with identification, though the variation across different specifications in the object estimated by bunching could be different than what is reported here.

[23]

We vary the distribution of η and of the optimization errors to see how the bunching estimators are affected. We can change the distribution in many ways; we can change the general shape, the center of the location and the variance. We can never know the shape of the preference distribution in the bunching window $[\eta_\ell, \eta_u]$. It is therefore of particular interest to consider variations in the pdf and mass in the bunching interval. We therefore construct an example so we easily can change the mass in the bunching window. We will keep the center of location constant as well as the general shape. We will see how the bunching estimate changes as we make the distribution more peaked and thereby increase the mass in the bunching window. A convenient way to accomplish this is to use a mixed normal centered at 1100, where $\phi(\eta) = \pi\varphi(\eta; 1100, 10^2) + (1 - \pi)\varphi(\eta; 1100, 140^2)$, where $\varphi(\eta; \mu, \sigma^2)$ is the pdf of a Gaussian distribution with mean μ and variance σ^2. As we vary π from .1 to .9 the distribution will become more peaked, and the mass in the bunching window will increase.

Table 3 reports the Chetty et al. (2011) elasticity formula, referred to as the bunching estimate, for a variety of heterogeneity distributions and optimization error specifications. The top row of Table 3 shows the five values of π used. The second row shows how results vary as we change π and there are no optimization errors. The simulations without optimization errors illustrate two things: First, the bunching estimate depends critically on the pdf in the bunching window and, second, there are pdfs such that the bunching estimate works well. As earlier emphasized, the amount of bunching depends on two things, the size of the bunching window $[\eta_\ell, \eta_u]$, and the pdf in the bunching window. In the simulation, we keep the kink and the taxable income elasticity constant, which means that the bunching window is constant. The only thing we vary is the pdf in the bunching window. We see that for $\pi = .1$ the bunching estimate gives a taxable income elasticity of .192. As we increase π the mass in the window increases and for $\pi = .9$ the bunching estimate is .598. Similarly we have calculated the Saez (2010) estimate from equation (5) of Saez (2010) and find .205 for $\pi = .1$ and .637 for $\pi = .9$, which are similar to the Chetty et al. (2011) estimates. This illustrates that the bunching, and the bunching estimates, depend heavily on the pdf in the window, where the pdf is not identified.

A second point that the second row of Table 3 illustrates is that there are pdfs such that the bunching estimate works well. For $\pi = .1$ the heterogeneity distribution is quite flat in the bunching window, implying that a linear approximation works well. Indeed, in this case the bunching estimate is close to the true elasticity. However, the assumed shape of the preference distribution in the bunching window is very incorrect when $\pi = .9$ and the estimate is far from the truth. Although we can easily find pdfs such that the bunching estimate works well we can never know when the true heterogeneity pdf is such that the estimates work well.

[24]

Rows 3 and 4 show results when we have added optimization errors drawn from a normal distribution with mean zero and standard deviations of 25 and 50 respectively. We see that adding this type of optimization error yields estimates of an order of magnitude smaller. In the fifth row we have only added optimization errors to taxable incomes at the kink or above, and all the optimization errors are positive. These optimization errors represent the optimization errors that would result if there had been a recent decrease in the slope of the second segment and not all individuals have been able to change their taxable income. These optimization errors mean that we observe fewer observations in the bunching window, resulting in lower estimates. This is borne out in the results. In the sixth row we illustrate what happens if there are the type of optimization errors that would arise if there had been a recent increase in the slope of the second segment and not all individuals have been able to change their taxable income. Negative optimization errors are added to taxable incomes above the kink, but there is a truncation so that no one falls below the kink because of the optimization error. By and large these last type of optimization errors do not affect the estimates much.

Table 3: Elasticities; Mixed Normals

π	.1	.3	.5	.7	.9
$\bar{\beta}$	.192	.302	.402	.500	.598
$\bar{\beta}_1$	.058	.074	.080	.081	.075
$\bar{\beta}_2$	.008	.010	.011	.013	.012
$\bar{\beta}_3$	.065	.077	.035	.013	0.0
$\bar{\beta}_4$	.238	.306	.394	.462	.530

$\bar{\beta}$ no optimization errors; $\bar{\beta}_1$ symmetric optimization errors, mean zero, std 25; $\bar{\beta}_2$ symmetric optimization errors, mean zero, std 50; $\bar{\beta}_3$ negative asymmetric optimization errors; $\bar{\beta}_4$ positive asymmetric optimization errors.

To summarize the results in the table, all data have been generated with a utility function with a taxable income elasticity of .2. The simulations illustrate that, even in the absence of optimization errors, but depending on the pdf in the bunching window, the estimates can vary between .192 and .598. Adding optimization errors in general makes the bunching estimates much smaller. Depending on the distribution of preferences and type of optimization errors the estimates vary between 0.0 and around .6.

Some might say that the heterogeneity distribution for $\pi = .9$ is quite extreme and that in reality heterogeneity distributions are not so peaked. This hypothesis does seem reasonable. Table 3 does not show that the bunching method often gives the wrong answer. The insight from Table 3 is that bunching estimates vary greatly depending on the pdf in the bunching window and the form of the optimization errors, both being constructs about which the data has nothing to say.

[25]

We have also considered examples with a variety of other parameter values, including varying the spread of optimization errors, the width of the bunching interval, the kink position relative to the mode, the difference between the slope of the first and second segments, and the elasticity. These simulation results are available upon request. In most cases, the estimated elasticity is further away from the true one when the kink is in the tails of the distribution than when it is close to the mode. Thus, the fact that we center the kink at the mode is not responsible for the large spread of elasticity estimates across specifications.

8   Conclusion

In this paper we have clarified conditions for identification of the effect of taxes on taxable income. We have shown that when the heterogeneity distribution is unrestricted the amount of bunching at a kink or a notch is not informative about the size of the taxable income elasticity for an isoelastic utility function, and neither is the entire distribution of taxable income for a concave budget frontierr. Kinks do provide information about the size of the elasticity when a priori restrictions are placed on the heterogeneity distribution. We have shown that the elasticity is identified when the heterogeneity distribution is specified across the kink. We have also provided elasticity bounds when bounds on the heterogeneity pdf are specified. We also found that distribution of taxable income for two budget sets does identify the elasticity if preferences are stable across budget sets and at least one individual type chooses income where the marginal tax rate differs across budget sets. Identification can also be obtained from a notch when there is a single budget set and isoelastic utility, although this source of identification seems not very useful in practice because it depends on gaps in the distribution of taxable income that are not observed in the data. For nonparametric utility with general heterogeneity we show that bunching at a kink is determined by compensated elasticities of individual types and a counterfactual pdf of taxable income at the kink. We also give bounds analogous to those for isoelastic utility. Identification becomes more difficult with optimization errors. We illustrate the impact of optimization errors with some simulations.

9   Appendix A: Proofs of Theorems

Proof of Theorem 1: Let $F(y)$ denote the distribution function of taxable income. Let $\Phi(\eta) = F(\rho_1^{\beta}\eta)$ for $\eta < \rho_1^{-\beta}K$ and let $\Phi(\eta) = F(\rho_2^{\beta}\eta)$ for $\eta > \rho_2^{-\beta}K$. By Theorem 1, on the lower segment where $y < K$ the distribution of taxable income will be $\Pr(\eta\rho_1^{\beta} \leq y) = \Phi(\rho_1^{-\beta}y) = F(y)$. Similarly, on the upper segment where $y > K$, the distribution of taxable income will be $\Pr(\eta\rho_2^{\beta} \leq y) = \Phi(\rho_2^{-\beta}y) = F(y)$. For $\rho_1^{-\beta}K \leq \eta \leq \rho_2^{-\beta}K$ let $\Phi(\eta)$ be any differentiable,

monotonic increasing function such that $\Phi(\rho_1^{-\beta}K) = \lim_{y\to K, y<K} F(y)$ and $\Phi(\rho_2^{-\beta}K) = F(K)$. Then by construction, we have

$$\Phi(\rho_2^{-\beta}K) - \Phi(\rho_1^{-\beta}K) = F(K) - \lim_{y\to K, y<K} F(a),$$

where the last equality holds by standard results for cumulative distribution functions.

Let $\phi(\eta) = d\Phi(\eta)/d\eta$ denote the pdf of η below $\eta_\ell = \rho_1^{-\beta}K$ and above $\eta_u = \rho_2^{-\beta}K$ respectively for the $\Phi(\eta)$ constructed above. By hypothesis there is $\varepsilon > 0$ such that $\phi(\eta)$ is bounded away from zero for

$$\eta \in N = [\eta_\ell - \varepsilon, \eta_\ell) \cup (\eta_u, \eta_u + \varepsilon].$$

Also let $\omega(\eta) = \ln\phi(\eta)$ in N and $\omega^{D-1}(\eta) = d^{D-1}\omega(\eta)/d\eta^{D-1}$ be the $D-1$ derivative. Let $\omega^- = \lim_{\eta\longrightarrow\eta_\ell, \eta<\eta_\ell} \omega^{D-1}(\eta)$ and $\omega^+ = \lim_{\eta\longrightarrow\eta_u, \eta>\eta_u} \omega^{D-1}(\eta)$. For $\eta \in [\eta_\ell, \eta_u]$ define $\omega^{D-1}(\eta)$ to be the height of the line connecting (η_ℓ, ω^-) and (η_u, ω^+). By construction $\omega^{D-1}(\eta)$ is continuous on $[\eta_\ell - \varepsilon, \eta_u + \varepsilon]$. Let $\omega(\eta)$ the $(D-1)^{th}$ integral of $\omega(\eta)$ and $\tilde\phi(\eta) = \exp(\omega(\eta))$. Let $B(\eta)$ be a D order B-spline basis function with support $[\eta_\ell, \eta_u]$ and

$$\omega_a(\eta) = [1 + aB(\eta)]\omega(\eta).$$

Then $\omega_a(\eta)$ is $D-1$ continuously differentiable, $\omega_a(\eta) = \omega(\eta)$ for $\eta \leq \eta_\ell$ or $\eta \geq \eta_u$, and $\lim_{a\longrightarrow-\infty} \omega_a(\eta) = -\infty$ and $\lim_{a\longrightarrow-\infty} \omega_a(\eta) = +\infty$ for $\eta \in (\eta_\ell, \eta_u)$. By the dominated convergence theorem $\int_{\eta_\ell}^{\eta_u} \exp(\omega_a(\eta))d\eta$ is continuous in a and

$$\lim_{a\longrightarrow-\infty} \int_{\eta_\ell}^{\eta_u} \exp(\omega_a(\eta))d\eta = 0, \quad \lim_{a\longrightarrow\infty} \int_{\eta_\ell}^{\eta_u} \exp(\omega_a(\eta))d\eta = \infty.$$

Therefore there exists a_0 such that $\int_{\eta_\ell}^{\eta_u} \exp(\omega_{a_0}(\eta))d\eta = P_K$. Then by construction a pdf satisfying the conditions of the Theorem is

$$\phi(\eta) = \exp(\omega_{a_0}(\eta)).Q.E.D.$$

Proof of Theorem 2: Let $\eta_1 = y_1\rho_1^{-\beta}$ and $\eta_2 = y_2\rho_2^{-\beta}$ be the endpoints of the bunching interval for η corresponding to y_1 and y_2 respectively. The bounds on the density are that for $\eta \in (\eta_1, \eta_2)$,

$$\underline{\sigma}\min\{\phi(\eta_1), \phi(\eta_2)\} \leq \phi(\eta) \leq \bar\sigma\max\{\phi(\eta_1), \phi(\eta_2)\}.$$

Also, $\phi(\eta_1)$ and $\phi(\eta_2)$ are given by $\phi(\eta_1) = f^-(y_1)\rho_1^\beta, \phi(\eta_2) = f^+(y_2)\rho_2^\beta$. The first conclusion of Theorem 2 then follows by

$$\begin{aligned}
P &= \Pr(y_1 \leq Y \leq y_2) = \int_{\eta_1}^{\eta_2} \phi(\eta)d\eta \leq (\eta_2 - \eta_1)\bar\sigma\max\{\phi(\eta_1), \phi(\eta_2)\} \\
&= \bar\sigma[y_2\rho_2^{-\beta} - y_1\rho_1^{-\beta}]\max\{f^-(y_1)\rho_1^\beta, f^+(y_2)\rho_2^\beta\} = \bar\sigma\max\{D^-(\beta), D^+(\beta)\}, \\
P &\geq \underline{\sigma}\min\{D^-(\beta), D^+(\beta)\}.
\end{aligned}$$

[27]

Note that both $D^-(\beta)$ and $D^+(\beta)$ are strictly monotonic increasing in β, so both $\max\{D^-(\beta), D^+(\beta)\}$ and $\min\{D^-(\beta), D^+(\beta)\}$ are as well. Also, at $\beta = 0$,

$$D^-(0) = f^-(y_1)(y_2 - y_1), \ D^+(0) = f^+(y_2)(y_2 - y_1).$$

As long as

$$P \geq \bar{\sigma} \max\{f^-(y_1), \ f^+(y_2)\}(y_2 - y_1),$$

then by strict monotonicity of $D^-(\beta)$ and $D^+(\beta)$ in β there will be unique β_ℓ and β_u satisfying

$$\bar{\sigma} \max\{D^-(\beta_\ell), D^+(\beta_\ell)\} = P, \ \sigma \min\{D^-(\beta_u), D^+(\beta_u)\} = P,$$

such that the above inequality is satisfied for all $\beta \in [\beta_\ell, \beta_u]$. If

$$\sigma \min\{f^-(y_1), f^+(y_2)\}(y_2 - y_1) < P < \bar{\sigma} \max\{f^-(y_1), f^+(y_2)\}(y_2 - y_1)$$

then we can take $\beta_\ell = 0$.

To show that sharpness, suppose

$$P < \sigma \min\left\{D^-(\beta), D^+(\beta)\right\}. \tag{9.11}$$

Then for any β satisfying the inequality (9.11) we have .

$$
\begin{aligned}
P &< \sigma[y_2\rho_2^{-\beta} - y_1\rho_1^{-\beta}]\min\{f^-(y_1)\rho_1^\beta, \ f^+(y_2)\rho_2^\beta\} \\
&= \sigma(\eta_2 - \eta_1)\min\{\phi(\eta_1), \phi(\eta_2)\} \leq \int_{\eta_1}^{\eta_2} \phi(\eta)d\eta,
\end{aligned}
$$

for any $\phi(\eta) \geq \min\{\phi(\eta_1), \phi(\eta_2)\}$ on $[\eta_1, \eta_2]$. That is, for any β satisfying (9.11) there is no pdf $\phi(\eta)$ satisfying the hypotheses of Theorem 2. It can also be shown analogously that if $P > \sigma \max\left\{D^-(\beta), D^+(\beta)\right\}$ there is also no pdf for η satisfying the hypotheses of Theorem 2. Therefore no other β than those given in the statement of Theorem 2 satisfy the restrictions of the model. $Q.E.D..$

Proof of Theorem 3: Note that $F(y) = \tilde{F}(\tilde{y})$ implies $\Phi(y\rho(y)^{-\beta}) = \Phi(\tilde{y}\tilde{\rho}(\tilde{y})^{-\beta})$, which implies $y\rho(y)^{-\beta} = \tilde{y}\tilde{\rho}(\tilde{y})^{-\beta}$ by $\Phi(\eta)$ strictly monotonic. Taking logs and solving gives the result. $Q.E..D.$

Proof of Corollary 4: Consider first the case where $\rho_2 \neq \tilde{\rho}_2$. Let $\bar{y}$ be a taxable income that is on the second segment for both budget sets. If $F(\bar{y}) < \tilde{F}(\bar{y})$ let $y > \bar{y}$ be such that $F(y) = \tilde{F}(\bar{y})$ and let $\tilde{y} = \bar{y}$. The conditions of Theorem 3 are fulfilled so β is identified. If $F(\bar{y}) \geq \tilde{F}(\bar{y})$ let $\tilde{y} \geq \bar{y}$ be such that $\tilde{F}(\tilde{y}) = F(\bar{y})$ and let $y = \bar{y}$. Then β is identified by

Theorem 3. Now suppose that $\rho_2 = \tilde{\rho}_2$ so that $\rho_1 \neq \tilde{\rho}_1$. For all $y > 0$ small enough y will be less than K and $\tilde{y} = (\tilde{\rho}_1/\rho_1)^\beta y$ will be less than $\tilde{K}$. At any such y and $\tilde{y}$ we have $\tilde{\rho}_1^{-\beta}\tilde{y} = \rho_1^{-\beta}y$ so that $\tilde{F}(\tilde{y}) = \Phi(\tilde{\rho}_1^{-\beta}\tilde{y}) = \Phi(\rho_1^{-\beta}y) = F(y)$ and since both y and $\tilde{y}$ are on the first segment, $\tilde{\rho}(\tilde{y}) = \tilde{\rho}_1 \neq \rho_1 = \rho(y)$. Then β is identified by Theorem 3. *Q.E.D.*

Proof of Theorem 5: Let ρ_1 and ρ_2 denote the slopes of the lower and upper linear segments that are equal for the two budget sets by hypothesis. Note that $F(y) = \Phi(\rho_2^{-\beta}y)$ for $y \geq K$ and $\tilde{F}(y) = \Phi(\rho_1^{-\beta}y)$ for $\tilde{y} < \tilde{K}$ so that $\tilde{F}_-(\tilde{K}) = \Phi(\rho_1^{-\beta}\tilde{K})$ by continuity of $\Phi(\eta)$. Then

$$F(\tilde{K}) = \Phi(\rho_2^{-\beta}\tilde{K}) > \Phi(\rho_1^{-\beta}\tilde{K}) = \tilde{F}_-(\tilde{K}).$$

Consider the case where $F(K) \leq \tilde{F}_-(\tilde{K})$, so that $F(K) \leq \tilde{F}_-(\tilde{K}) < F(\tilde{K})$. It follows by continuity and strict monotonicity of $\Phi(\eta)$ that there exists a unique $y^* \in [K, \tilde{K})$ with $\Phi(\rho_2^{-\beta}y^*) = F(y^*) = \tilde{F}_-(\tilde{K}) = \Phi(\rho_1^{-\beta}\tilde{K})$. By strict monotonicity of $\Phi(\eta)$ it follows that $\rho_2^{-\beta}y^* = \rho_1^{-\beta}\tilde{K}$. Taking logs and solving for β gives the first conclusion.

Next, note that $\tilde{F}_-(\tilde{K}) < F(K)$ implies $\Phi(\rho_1^{-\beta}\tilde{K}) < \Phi(\rho_1^{-\beta}\tilde{K})$, which in turn implies $\rho_1^{-\beta}\tilde{K} < \rho_2^{-\beta}K$ by strict monotonicity of Φ. Taking logs gives the second conclusion.

Now consider any $b > \underline{\beta}$, or equivalently $\rho_1^{-b}\tilde{K} < \rho_2^{-b}K$. Let $\Phi^b(\eta) = \tilde{F}(\rho_1^b\eta)$ for $\eta < \rho_1^{-b}\tilde{K}$ and $\Phi^b(\eta) = F(\rho_2^b\eta)$ for $\eta \geq \rho_2^{-b}K$. Note that

$$\lim_{\eta \longrightarrow \rho_1^{-b}\tilde{K}, \eta < \rho_1^{-b}\tilde{K}} \Phi^b(\eta) = \lim_{y \longrightarrow \tilde{K}, y < \tilde{K}} \tilde{F}(y) = \tilde{F}_-(\tilde{K}) < F(K) = \Phi^b(\rho_2^{-b}K).$$

For $\rho_1^{-b}\tilde{K} < \eta < \rho_2^{-b}K$ let $\Phi^b(\eta)$ be any strictly increasing, continuous function that is equal to $\tilde{F}_-(\tilde{K})$ at $\rho_1^{-\beta}\tilde{K}$ and $F(K)$ at $\rho_2^{-b}K$. Consider the taxable income distribution when the taxable income elasticity is b and the CDF of heterogeneity is $\Phi^b(\eta)$. Note that for the budget set with kink K, $y < K$, and taxable income Y we have $\Pr(Y \leq y) = \Phi^b(\rho_1^{-b}y) = \tilde{F}(y) = F(y)$, while for $y \geq K$, $\Pr(Y \leq y) = \Phi^b(\rho_2^{-b}y) = F(y)$. Also, for the second budget set and $y < \tilde{K}$ we have $\Pr(Y \leq y) = \Phi^b(\rho_1^{-b}y) = \tilde{F}(y)$, while for $y \geq \tilde{K} > K$ we have $\Pr(Y \leq y) = \Phi^b(\rho_2^{-b}y) = F(y) = \tilde{F}(y)$. Thus, when the taxable income elasticity is any $b > \underline{\beta}$ we have constructed a heterogeneity distribution satisfying the conditions i)-iii) of Theorem 3 where taxable income has the same distribution as the true distribution, for both budget sets. *Q.E.D.*

Proof of Theorem 6: From Figure 6 we see that as η increases individual types will choose to be at the notch until the indifference curve becomes tangent to the second budget segment. Let $\bar{\eta}$ be the η for the individual type with indifference curve passing through the notch and tangent to the second segment. Let K and $C_K = R_1 + \rho_1 K$ denote the notch and the consumption at the notch from the first budget frontier segment. Let y_3 denote the taxable

income at the tangency and $C_3 = R_2 + \rho_2 y_3$ the consumption. The type $\bar{\eta}$ being indifferent between (K, C_K) and (y_3, C_3) means that

$$C_K - \frac{\bar{\eta}}{1+\frac{1}{\beta}}\left(\frac{K}{\bar{\eta}}\right)^{1+1/\beta} = C_3 - \frac{\bar{\eta}}{1+\frac{1}{\beta}}\left(\frac{y_3}{\bar{\eta}}\right)^{1+1/\beta}.$$

The indifference curve being tangent to the second budget segment at y_3 means that

$$y_3 = \bar{\eta}\rho_2^\beta.$$

Solving for $\bar{\eta} = \rho_2^{-\beta} y_3$ and substituting in the previous equation gives

$$C_K - \frac{y_3 \rho_2^{-\beta}}{1+\frac{1}{\beta}}\left(\frac{K\rho_2^\beta}{y_3}\right)^{1+1/\beta} = C_3 - \frac{y_3 \rho_2^{-\beta}}{1+\frac{1}{\beta}}\left(\rho_2^\beta\right)^{1+1/\beta}.$$

Noting that $\rho_2^{-\beta}(\rho_2^\beta)^{1+1/\beta} = \rho_2$, it follows that

$$C_K - \frac{y_3 \rho_2}{1+\frac{1}{\beta}}\left(\frac{K}{y_3}\right)^{1+1/\beta} = C_3 - \frac{y_3 \rho_2}{1+\frac{1}{\beta}}.$$

Letting $\alpha = 1 + 1/\beta \geq 1$ it follows by collecting terms that

$$C_3 - C_K = h(\alpha), \ \ h(\alpha) = y_3 p_2 \frac{1}{\alpha}[1 - \left(\frac{K}{y_3}\right)^\alpha]. \tag{9.12}$$

Note that $h(1) = \rho_2(y_3 - K)$. Also,

$$C_3 - C_K = R_2 + \rho_2 y_3 - R_1 - \rho_1 K = R_2 + \rho_2 K - R_1 - \rho_1 K + \rho_2(y_3 - K) < h(1)$$

by $R_2 + \rho_2 K < R_1 + \rho_1 K$, i.e. by K being a notch. By $K < y_3$ it follows that $h(\alpha) \longrightarrow 0$ as $\alpha \longrightarrow \infty$. Note also that for $p = K/y_3 < 1$,

$$\frac{dh(\alpha)}{d\alpha} = -y_3 p_2 \alpha^{-2}\{1 - p^\alpha + \ln(p)p^\alpha\}.$$

For $k(p) = 1 - p^\alpha + \ln(p)p^\alpha$ note that $k(1) = 0$ and for any $0 < p \leq 1$,

$$dk(p)/dp = -\alpha p^{\alpha-1} + p^{\alpha-1} + \alpha \ln(p)p^{\alpha-1} = [(1-\alpha) + \alpha \ln(p)]p^{\alpha-1} < 0.$$

Therefore by $p < 1$ it follows that $1 - p^\alpha + \ln(p)p^\alpha > 0$ and hence that $h(\alpha)$ is monotonically decreasing. This function is also continuous so there exists a unique α that solves equation (9.12). Finally note that y_3 is identified as the smallest point in the support of the taxable income variable above K. Since there is a unique α that solves equation equation (9.12) it follows that β is identified. *Q.E.D.*

Proof of Theorem 7: Theorem 7 is a special case of Theorem B3 in Appendix B. *Q.E.D.*

Proof of Theorem 8: Let $F(y)$ be the CDF of $Y(B, \eta)$. By standard probability theory, $\Pi_{\bar{\ell}} = F(K) - \lim_{y \uparrow K} F(y)$. By Theorem B3 $F(K) = F(K | \rho_2, R_2)$ and $\lim_{y \uparrow K} F(y) = \lim_{y \uparrow K} F(y | \rho_1, R_1)$. Furthermore, by Assumption B3 $Y(\rho_1, R_1, \eta)$ is continuously distributed so that $\lim_{y \uparrow K} F(y | \rho_1, R_1) = F(K | \rho_1, R_1)$. Define $\Lambda(\rho) = F(K | \rho, R(\rho))$ for $\rho \in [\rho_2, \rho_1]$. We then have

$$B = \Lambda(\rho_2) - \Lambda(\rho_1).$$

By the chain rule, $R(\rho) = R_1 + K(\rho_1 - \rho)$, and Lemma A1 $\Lambda(\rho)$ is differentiable in ρ and

$$\frac{d\Lambda(\rho)}{d\rho} = F_\rho(K | \rho, R(\rho)) - K F_R(K | \rho, R(\rho)) = -\tilde{\phi}(\rho) \bar{\beta}(\rho, K).$$

The conclusion then follows by the fundamental theorem of calculus. *Q.E.D.*

10 Appendix B: Heterogenous, Nonparametric Taxable Income Choice for a Concave Budget Frontier

We first give conditions that impose strict quasi-concavity of utility and continuity conditions that allow us to make probability statements with general heterogeneity.

ASSUMPTION B1: *For each η, $U(c, y, \eta)$ is continuous in (c, y), increasing in c, decreasing in y, and strictly quasi-concave in (c, y). Also $Y(\rho, R, \eta) < \infty$ and $Y(\rho, R, \eta)$ is continuously differentiable in $\rho, R > 0$.*

ASSUMPTION B2: *η belongs to a complete, separable metric space and $Y(\rho, R, \eta)$, $\partial Y(\rho, R, \eta)/\partial\rho$, and $\partial Y(\rho, R, \eta)/\partial R$ are continuous in (ρ, R, η).*

ASSUMPTION B3: *$\eta = (u, \varepsilon)$ for scalar ε and Assumption A1 is satisfied for $\eta = (u, \varepsilon)$ for a complete, separable metric space that is the product of a complete separable metric space for u with Euclidean space for ε, $Y(\rho, R, \eta) = Y(\rho, R, u, \varepsilon)$ is continuously differentiable in ε, there is $C > 0$ with $\partial Y(\rho, R, u, \varepsilon)/\partial\varepsilon \geq 1/C$, $\|\partial Y(\rho, R, \eta)/\partial(\rho, R)\| \leq C$ everywhere, ε is continuously distributed conditional on u, with conditional pdf $f_\varepsilon(\varepsilon | u)$ that is bounded and continuous in ε.*

The first result of this Appendix gives useful formulae for the partial derivatives of $F(y | \rho, R)$.

LEMMA B1: *If Assumptions B1 - B3 are satisfied then $Y(\rho, R, \eta)$ is continuously distributed for each $\rho, R > 0$ and $F(y | \rho, R)$ is continuously differentiable in y, ρ, and R and for the pdf $f_{Y(\rho,R,\eta)}(y)$ of $Y(\rho, R, \eta)$ at y,*

$$\frac{\partial F(y | \rho, R)}{\partial y} = f_{Y(\rho,R,\eta)}(y),$$

$$\frac{\partial F(y | \rho, R)}{\partial(\rho, R)} = -f_{Y(\rho,R,\eta)}(y) E\left[\frac{\partial Y(\rho, R, \eta)}{\partial(\rho, R)} \Big| Y(\rho, R, \eta) = y\right].$$

Proof: This follows exactly as in the proof of Lemma A1 of Hausman and Newey (2016). Q.E.D.

The next three results are useful in the proof of Theorem 1.

LEMMA B2: *If Assumptions B1 and B2 are satisfied and $B(y)$ is concave then $Y(B, \eta)$ is unique and $U(B(y), y, \eta)$ is strictly increasing in y to the left of $Y(B, \eta)$ and strictly decreasing in y to the right of $Y(B, \eta)$.*

Proof: For notational convenience suppress the η argument, which is held fixed in this proof. Let $y^* = Y(B)$. Suppose $y^* > 0$. Consider $y < y^*$ and let $\tilde{y}$ such that $y < \tilde{y} < y^*$. Let $(\tilde{c}, \tilde{y})$ be on the line joining $(B(y), y)$ and $(B(y^*), y^*)$. By concavity of $B(\cdot)$, $\tilde{c} \leq B(\tilde{y})$, so by strict quasi-concavity and the definition of y^*,

$$U(B(y^*), y^*) \geq U(B(\tilde{y}), \tilde{y}) \geq U(\tilde{c}, \tilde{y}) > \min\{U(B(y), y), U(B(y^*), y^*)\} = U(B(y), y).$$

Thus $U(B(\tilde{y}), \tilde{y}) > U(B(y), y)$. An analogous argument gives $U(B(\tilde{y}), \tilde{y}) > U(B(y), y)$ for $y > \tilde{y} > y^*$. Q.E.D.

THEOREM B3: *If Assumptions B1 and B2 are satisfied and $B(y)$ is a concave function then for each y, $\Pr(Y(B, \eta) \leq y) = F(y|\rho(y), R(y))$.*

Proof: Consider any fixed value of y as in the statement of the Lemma and let $B(z)$ denote the value of the budget frontier for any value z of taxable income. By concavity of $B(z)$ and Rockafellar (1970, pp. 214-215), $\rho(y)$ exists and is a subgradient of $B(z)$ at y. Define $\dot{\rho} = \rho(y)$ and $\dot{B} = B(y)$. For any z let $\dot{B}(z) = \dot{B} + \dot{\rho}(z - y) = R(y) + \rho(y)z$ denote the linear budget frontier with slope $\rho(y)$ passing through $(B(y), y)$. Let $y^* = \text{argmax}_z U(B(z), z)$ where we suppress the η argument for convenience. Also let $\dot{y}^* = \arg\max_z U(\dot{B}(z), z)$. We now proceed to show that $y^* \leq y \iff \dot{y}^* \leq y$.

It follows by $\dot{\rho}$ being a subgradient at y of $B(z)$ and $B(z)$ being concave that for all z,

$$\dot{B}(z) \geq B(z).$$

Therefore $\dot{B}(y^*) \geq B(y^*)$, so that

$$U(\dot{B}(y^*), y^*) \geq U(B(y^*), y^*) \geq U(B(y), y) = U(\dot{B}(y), y).$$

Note that $\dot{B}(z)$ is linear and hence concave so that by Lemma B2, $U(\dot{B}(z), z)$ is strictly increasing to the left of $\dot{y}^*$ and decreasing to the right of $\dot{y}^*$. Suppose that $y < y^*$. Then $y < \dot{y}^*$, because otherwise $\dot{y}^* \leq y < y^*$ and the above equation contradicts that $U(\dot{B}(z), z)$ is strictly

[32]

decreasing to the right of $\dot{y}^*$. Similarly, if $y > y^*$ then $y > \dot{y}^*$, because otherwise $\dot{y}^* \geq y > y^*$ and the above equation contradicts that $U(\dot{B}(z), z)$ is strictly increasing to the left of $\dot{y}^*$.

Next suppose $y^* = y$. Let $\ddot{\rho}$ be the slope of a line that separates the set weakly preferred to $(B(y), y)$ and the budget set and let $\ddot{B}(z) = \dot{B} + \ddot{\rho}(z - y^*)$, so that $U(\dot{B}, y^*) \geq U(\ddot{B}(z), z)$ for all z. Then by Lemma B2 applied to the budget frontier $\ddot{B}(z)$, $U(\dot{B}, y^*) > U(\ddot{B}(z), z)$ for all $z \neq y^*$. Also, by Rockafellar (1970, pp. 214-215) $\ddot{\rho} \geq \dot{\rho}$. Then for any $z > y^*$ we have

$$\dot{B} + \ddot{\rho}(z - y^*) \geq \dot{B} + \dot{\rho}(z - y^*) = \dot{B}(z).$$

so that

$$U(\dot{B}(z), z) \leq U(\dot{B} + \ddot{\rho}(z - y^*), z) = U(\ddot{B}(z), z) < U(\dot{B}, y^*).$$

It follows that $\dot{y}^* \leq y^* = y$. Thus, we have shown that $y^* = y$ implies $\dot{y}^* \leq y$. Together with the implication of the previous paragraph this means that $y^* \leq y \implies \dot{y}^* \leq y$.

Summarizing, we have shown that

$$y^* \leq y \implies \dot{y}^* \leq y \text{ and } y^* > y \implies \dot{y}^* > y.$$

Therefore $y^* \leq y \iff \dot{y}^* \leq y$.

Note that y^* is the utility maximizing point on the budget frontier $B(z)$ while $\dot{y}^*$ is the utility maximizing point on the linear budget frontier $\dot{B}(z) = B(y) + \rho(y)(z - y) = R(y) + \rho(y)z$. Thus, $y^* \leq y \iff \dot{y}^* \leq y$ means that the event $Y(B, \eta) \leq y$ coincides with the event that $\arg\max_z U(R(y) + \rho(y)z, z, \eta) \leq y$, i.e. with the event the optimum on the linear budget set is less than or equal to y. The probability that the optimum on this linear budget is less than or equal to y is $F(y|\rho(y), R(y))$, giving the conclusion. *Q.E.D.*

11 References

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

Bertanha, M., A.H. McCallum, and N. Seegert (2017), "Better Bunching, Nicer Notching," working paper, Dec 2017.

Bertanha, M., A.H. McCallum, and N. Seegert (2018), "Better Bunching, Nicer Notching," working paper, March 2018.

Blomquist, S. and W.K. Newey (2002), "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica 70,* 2455–2480.

Blomquist, S. and W.K. Newey (2017), "The Bunching Estimator Cannot Identify the Taxable Income Elasticity," CEMMAP Working Paper CWP40/17, 2 October 2017; NBER Working Paper w24136, Dec 2017.

Blomquist, S. and W.K. Newey (2018), "The Kink and Notch Bunching Estimators Cannot Identify the Taxable Income Elasticity," Working paper 2018:4, Dept. of Economics, Uppsala University.

Blomquist, S. and H. Selin (2010), "Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates," *Journal of Public Economics 94*: 878–889.

Blomquist, S., A. Kumar, C.-Y. Liang and W.K. Newey (2015), "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income," CEMMAP working paper CWP 12/15.

Blundell, R.W. and A. Shephard (2012), "Employment, Hours of Work and the Optimal Taxation of Low-Income Families," *Review of Economic Studies* 79, 481–510.

Burns, S.A. and J.P. Ziliak (2017), "Identifying the Elasticity of Taxable Income," *Economic Journal* 127, 297-329.

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

Cattaneo, M., M. Jansson, X. Ma, J. Slemrod (2018), "Bunching Designs: Estimation and Inference," working paper, March 22.

Chernozhukov, V., H. Hong, E. Tamer (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica* 75, 1243-1284.

Chetty, R., Friedman, J., Olsen, T. 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, 749-804.

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

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.

Feldstein, M. (1995), "The Effects of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy 103(3)*: 551 -572.

Gelber, A., D. Jones, and D. Sacks (2017), "Estimating Earnings Adjustment Frictions: Method and Evidence from the Social Security Earnings Test," mimeo.

Gruber, J. and E. Saez (2002), "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics 84*:1–32.

Hausman, J.A. (1979), "The Econometrics of Labor Supply on Convex Budget Sets," *Economics Letters* 3, 171-174.

Hausman, J. A. (1981), "Labor Supply," in H. J. Aaron and J. A. Pechman, eds. *How Taxes Affect Economic Behavior.* The Brookings Institution.

Hausman, J.A., (1985), "The Econometrics of Nonlinear Budget Sets," *Econometrica* 53,

1255-1282.

Hausman, J.A. and W.K. Newey (2016), "Individual Heterogeneity and Average Welfare," *Econometrica* 84, 1225-1248.

Imbens, G.W. and C.F. Manski (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica* 72, 1631-1935.

Keane, M.P. (2011), "Labor Supply and Taxes: A Survey," *Journal of Economic Literature* 49, 961–1075.

Keane, M. and R. Moffitt (1998), "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review* 39, 553–589.

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, 669-723.

Kleven, H. J., M. Knudsen, C.T. Kreiner, S. Pedersen and E. Saez (2011), "Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark," *Econometrica 79*: 651–692.

Kline, P. and M. Tartari (2016), "Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach," *American Economic Review* 106, 972–1014.

Kumar, A. and C.Y. Liang (2017), "Estimating Taxable Income Responses with Elasticity Heterogeneity," Uppsala University, Department of Economics Working Paper 2017:5.

le Maire, D. and B. Schjerning (2013), "Tax Bunching, Income Shifting and Self-Employment," *Journal of Public Economics* 107, 1-18.

Manski, C. (2014), "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy," *Quantitative Economics* 5, 145-174.

Marx, B.M. (2012), "Regulatory Hurdles and Growth of Charitable Organizations: Evidence from a Dynamic Bunching Design," mimeo.

McCallum, A.H., and N. Seegert (2017), "Better Bunching, Nicer Notching," working paper, Dec 2017.

Rockafellar, R.T. (1970), *Convex Analysis*, Princeton: Princeton University Press.

Saez, E. (2003), "The Effect of Marginal Tax Rates on Income: A Panel Study of 'Bracket Creep'," *Journal of Public Economics 85*: 1231–1258.

Saez, E. (1999), "Do Taxpayers Bunch at Kink Points?" NBER Working Paper No. 7366.

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

Seim, D. (2015), "Behavioral Responses to an Annual Wealth Tax: Evidence from Sweden,"

American Economic Journal: Economic Policy, forthcoming.

Van Soest, A. (1995), "Structural Models of Family Labor Supply: a Discrete Choice Approach," *Journal of Human Resources* 30, 63-88.

Weber, C. (2014), "Toward Obtaining a Consistent Estimate of the Elasticity of Taxable Income using Difference-in-Differences," *Journal of Public Economics* 117, 90-103.
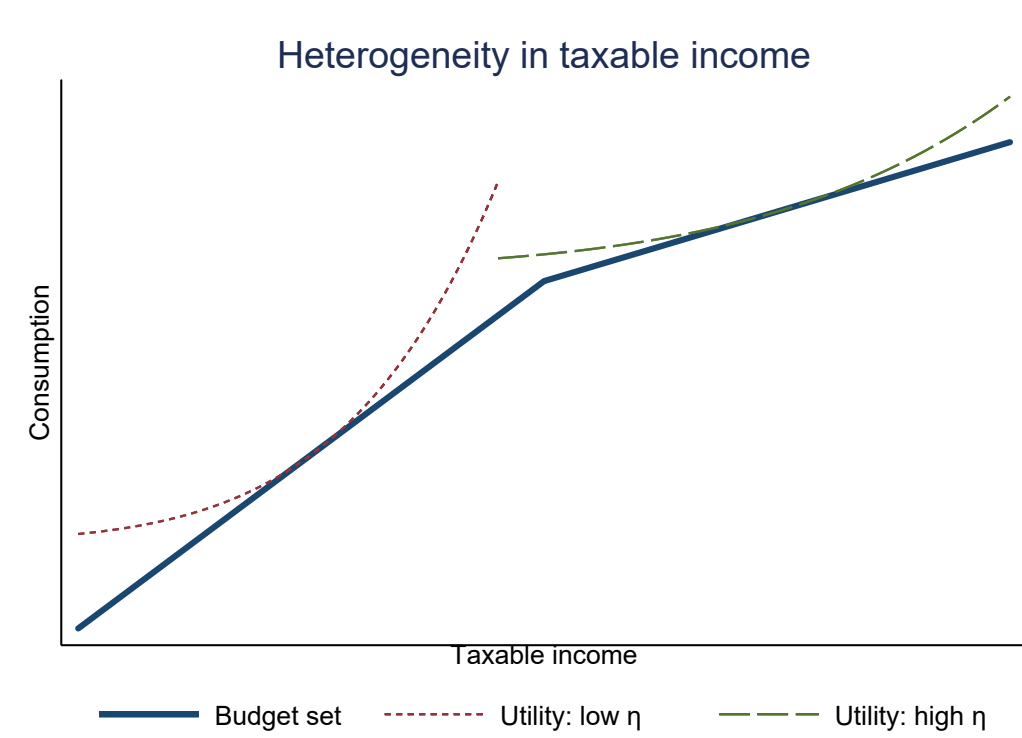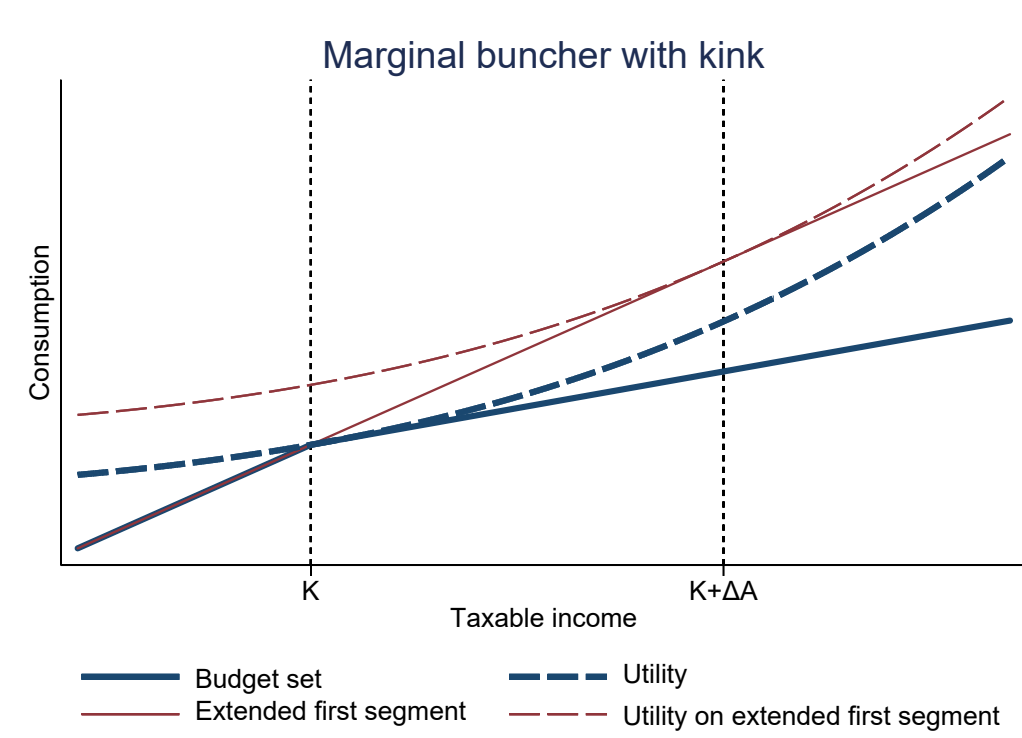
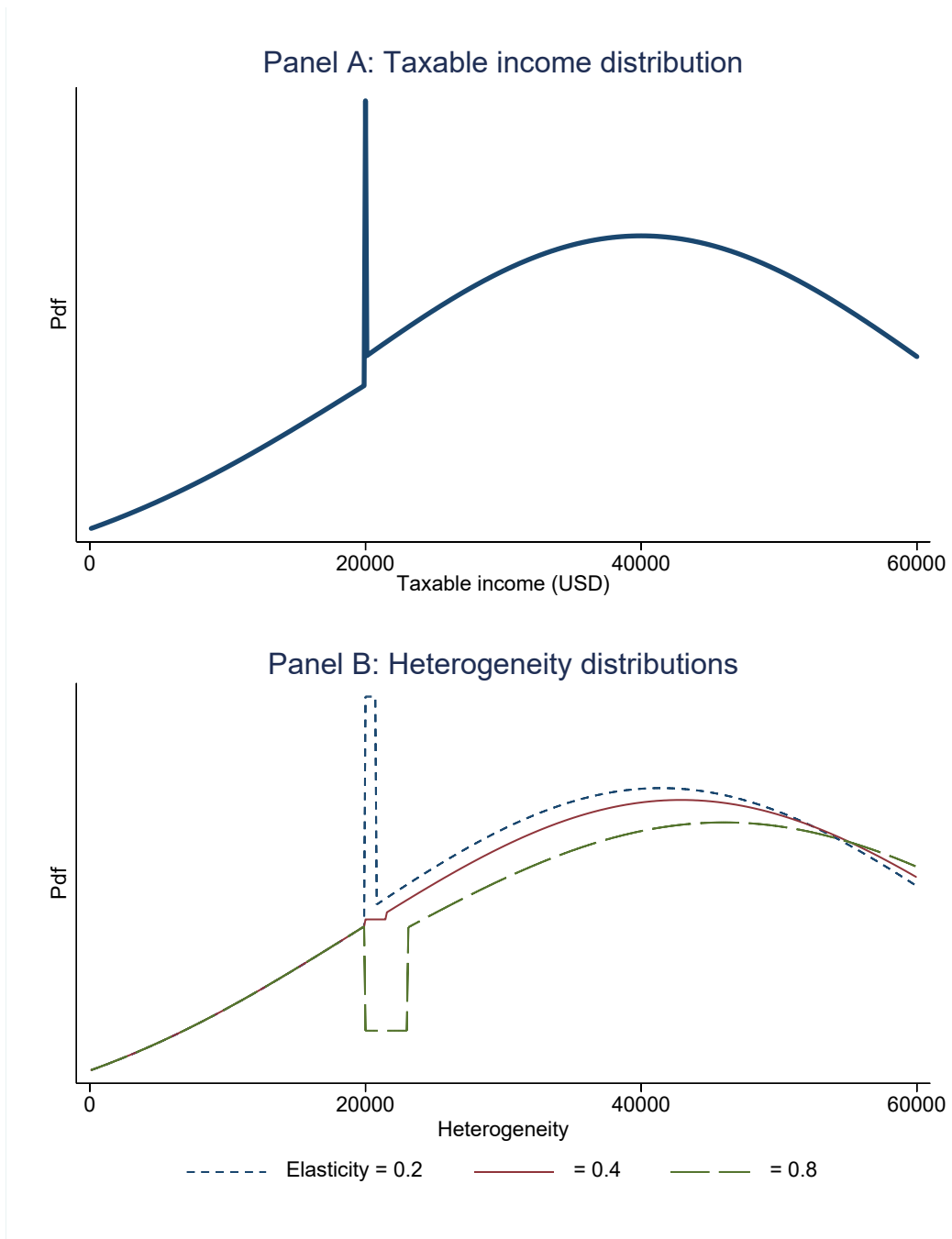Figure 1. Heterogeneity in taxable income

Figure 2. Marginal buncher with kink

Figure 3. Nonidentification: multiple combinations of elasticity and heterogeneity distribution give the same taxable income distribution
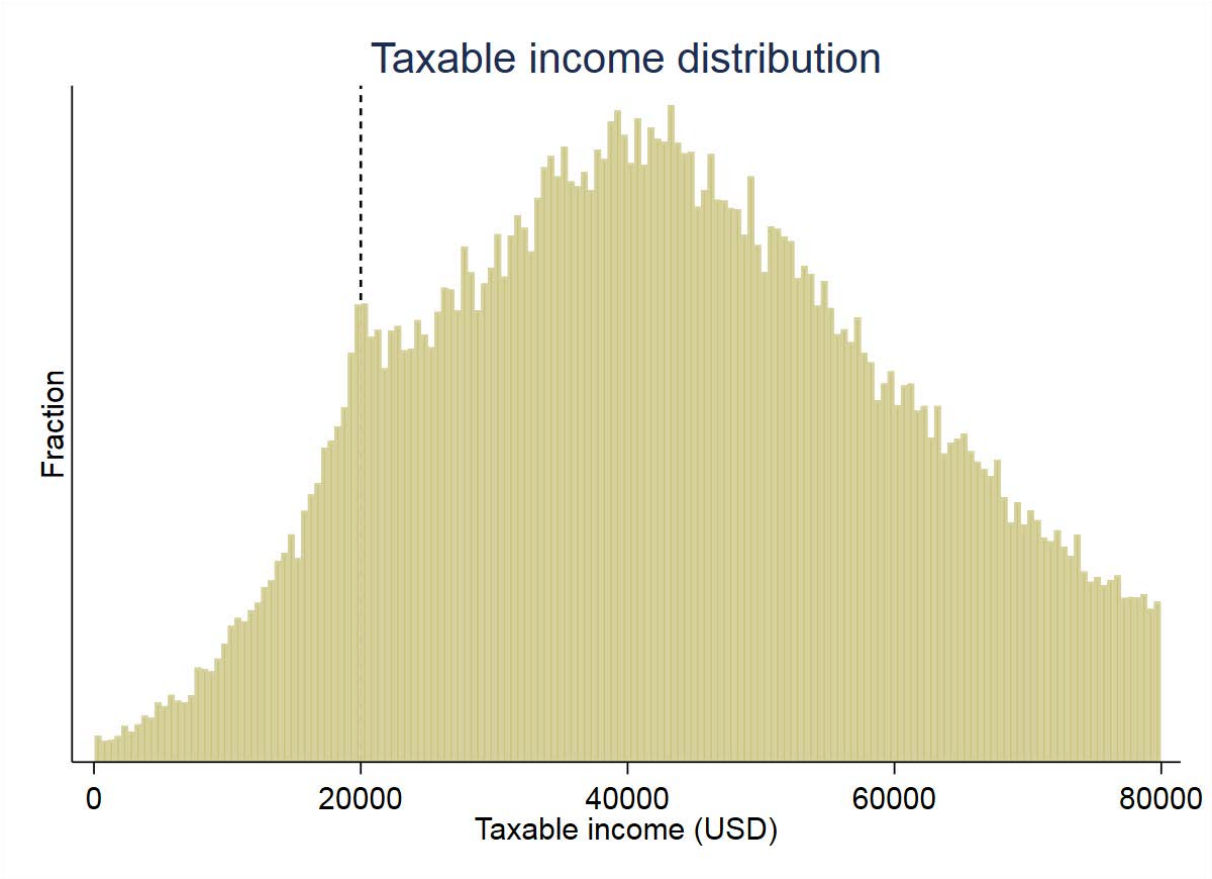
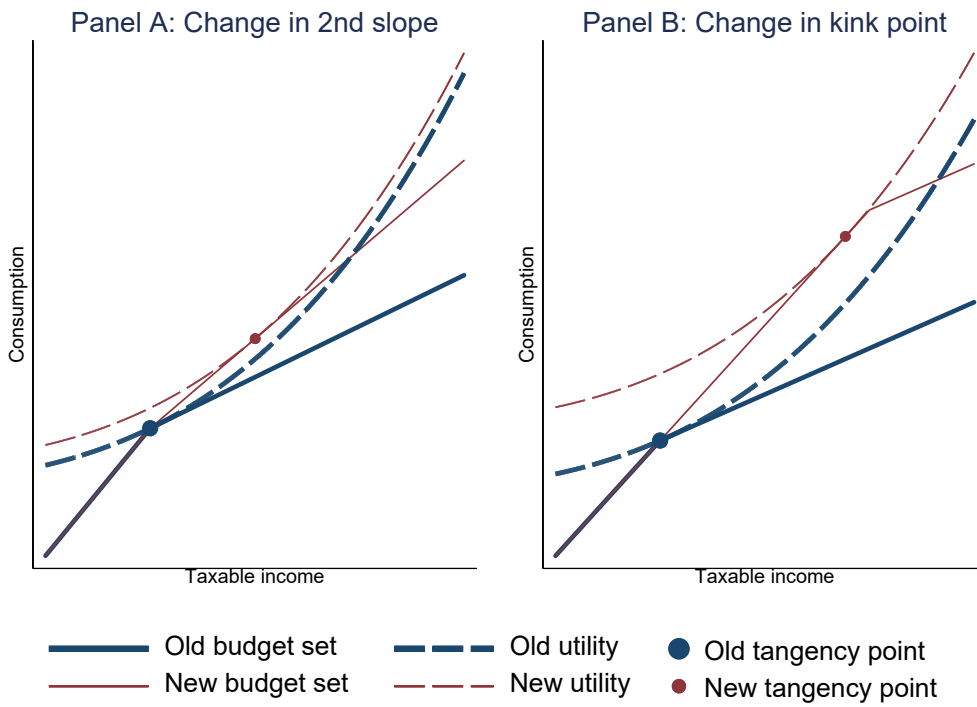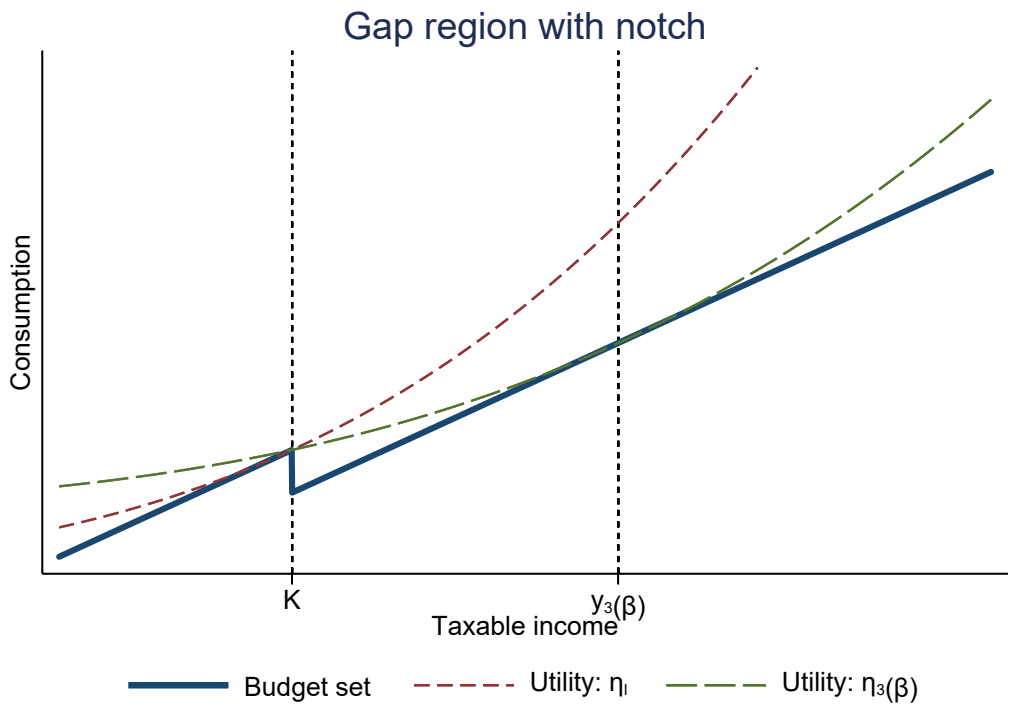Figure 4. Histogram of taxable income distribution for married tax filers 1960-1969

Figure 5. Identification from budget set variation

Figure 6. Gap region with notch