On Bunching and Identification of the Taxable Income Elasticity*

Soren Blomquist
Uppsala Center for Fiscal Studies, Department of Economics, Uppsala University

Whitney K. Newey
Department of Economics M.I.T. and NBER

Anil Kumar
Federal Reserve Bank of Dallas

Che-Yuan Liang
Institute for Housing and Urban Research, Uppsala Center for Fiscal Studies, and Department of Economics, Uppsala University

July 2020

Abstract

The elasticity of taxable income with respect to the net of tax rate 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 has been used to estimate the taxable income elasticity. We show that when the distribution of preferences 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 preference distribution. They can identify the elasticity when the preference distribution is completely and correctly specified across the kink and provide bounds under restrictions on the preference distribution. We find wide estimated bounds in an empirical example using data like Saez (2010) based on upper and lower bounds for the preference density.

JEL Classification: C14, C24, H31, H34, J22

Keywords: Bunching, kinks, notches, identification, taxable income, tax rates, heterogeneous preferences.

*The NSF provided partial financial support via grant 1757140. We are grateful for the many constructive and thoughtful comments by referees and by the editor, M. Mogstad. Helpful comments were also provided by R. Blundell, A. Finklestein, J. Hausman, H. Kleven, C. Manski, R. Matzkin, J. Poterba, E. Saez, H. Selin and participants at seminars at BC, Chicago, Georgetown, Harvard/MIT, Michigan, NYU, UC Irvine, UCL, UCLA, USC, Yale and the bunching conference at UCSD in March of 2018. This paper includes results from a previously unpublished working paper, Blomquist and Newey (2017). The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

1 Introduction

The elasticity of taxable income with respect to the net of tax rate 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 distribution of preferences 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 distinguish the taxable income elasticity from the preference distribution with a single budget set. Small kinks do not solve this problem. Whether the kink 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 preference 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 preferences 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 preference distribution. The elasticity can be identified when the preference distribution is completely and correctly specified across the kink for isoelastic utility. We also give bounds on the elasticity based on bounds on the preference density. These bounds can be viewed as measures of sensitivity of the taxable income elasticity to assumptions about the preference distribution. We find in an empirical example using data like Saez (2010) that the taxable income elasticity can be quite sensitive to the bounds on the preference distribution.

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 bounds 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 preference distribution across the kink. Imposing a known distribution of preferences seems 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 preferences. Section 4 applies these results to data like that used in Saez (2010). Proofs are given in the Appendix.

Nonparametric nonidentification of compensated tax effects from a kink was shown and bounds provided in Blomquist et al. (2015). The nonidentification results we give for parametric isoelastic utility imply those for the more general nonparametric model. McCallum and Seegert (2017) give identification results when covariates are present and preferences have a Gaussian distribution. Blomquist and Newey (2017) showed that for a parametric, isoelastic utility function the distribution of taxable income for one convex budget set provides no information about the elasticity, provided bounds, and showed identification from budget set variation holding the preference distribution fixed. These nonidentification and bounds results are incorporated in Sections 2 and 3 of this paper. Bertanha, McCallum, and Seegert (2017) consider nonidentification results and give bounds based on variability of the preference density.

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 preferences with optimization errors. Blomquist et al. (2015) showed that these results are valid with general preferences and demonstrated how to impose all the restrictions of utility maximization

[1] Bastani and Selin (2014), Gelber et al. (2020), Marx (2012), Le Maire and Schjering (2013) and Seim (2017) are a few of the recent papers that apply the bunching method.

in estimating taxable income effects. Manski (2014) and Kline and Tartari (2016) nonparametrically identified and estimated bounds on important effects. 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 by difference-in-differences. Saez et al. (2012) provided a review of this literature. Whereas early 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 (2020) provided recent estimated elasticities between 0.2 and 1.0.

2 Nonidentification from Bunching

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 $c = A(y)$,where $A(y) = y - T(y)$ for taxes $T(y)$. The utility function of an individual will be

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

where η represents unobserved heterogeneity in preferences.

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 utility is highest for that η. 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 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.

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 $(K, K+\Delta Y]$ along the first segment now choose the kink point. We refer to the individual who would have chosen $K+\Delta Y$ 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 $K+\Delta Y$. 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

$$e = \frac{\Delta Y/K}{(\rho_1 - \rho_2)/\rho_1} \tag{2.1}$$

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 ΔY. From the data we identify the proportion P_K of individuals located at the kink. Then we have

$$P_K = \int_K^{K+\Delta Y} f_1(y)dy, \tag{2.2}$$

where $f_1(y)$ is the density of taxable income along the extended first segment. If $f_1(y)$ were identified we could identify ΔY 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 ΔY 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 parameters ΔY and f_1. For every positive P_K there will be multiple values of ΔY and f_1 such that equation (2.2) is satisfied.

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

$$U(c, y, \eta) = c - \frac{\eta}{1+\frac{1}{\beta}} \left(\frac{y}{\eta}\right)^{1+1/\beta}, \eta > 0, \beta > 0, \tag{2.3}$$

where η is a scalar. Maximizing this utility function subject to a linear budget constraint $A(y) = \rho y + R$ 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.

We can see nonidentification even more clearly for the isoelastic utility function. 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 interval.* Therefore the kink probability satisfies

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

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 interval $[\eta_\ell, \eta_u]$, 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 (2.3) 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 preference distribution such that the CDF of taxable income for the model coincides with that for the data. Furthermore, we can do this with a preference 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 preferences is unrestricted. The same result can be shown for any continuous, piecewise linear budget frontier with nondecreasing marginal tax rates and each kink having positive probability.

Theorem 1 implies nonidentification of the average taxable income elasticity, averaged over any income range and over preferences, for any class of utility functions that includes isoelastic utility as a special case. This wider implication of Theorem 1 occurs because it is more difficult to identify a parameter in a more general model, so that nonidentification in the more restrictive isoelastic model implies nonidentification in the more general model. For example, Theorem 1 implies the nonidentification of the average compensated taxable income elasticity for individuals located at a kink that was shown in Blomquist et al. (2015).

Figure 3 illustrates the nonidentification 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 preference distribution below the kink that matches well the taxable income distribution there.

In Panel B of Figure 3 we graph the density of three differentiable 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 preferences densities are the same because $\rho_1 = 1$. To highlight the differences between these pdf's we choose the density to be nearly constant in the bunching interval, approximately equal to the value which makes the integral over the bunching interval equal to P_K. We see that choosing $\beta = .2$ well below the value $\beta = .4$ leads to a preference density that is large over the bunching interval, while choosing $\beta = .8$ well above $\beta = .4$ leads to a preference density that is low over the bunching interval. This pattern occurs because the length of the bunching interval $[\eta_\ell, \eta_u] = [K\rho_1^{-\beta}, K\rho_2^{-\beta}]$ is monotonic increasing in β. In order to make the integral of the preference density over the bunching interval be equal to P_K for each β we must have a large preference density across the bunching interval when β is small and a small preference density when β is large.

The taxable income elasticity is not identified no matter how close ρ_2 is to ρ_1. As long as $P_K > 0$ then for any β there will be a density of η over the bunching interval that satisfies equation (2.4). It is true that large values of β will require that the density of η be small and small values of β require that the density be large. Nevertheless, such preference densities are allowed when there is no a priori information on the distribution of preferences. In Section 3 we will consider what can be learned if there are bounds on the preference density.

The proof of Theorem 1 relies entirely upon nonidentification from a kink. For any particular β there is only one preference density 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 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 preference density over the kink, i.e. in the bunching interval $[\eta_\ell, \eta_u]$. Over the bunching interval any pdf $\phi(\eta)$ 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 preference density 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 preferences is unrestricted.

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 4 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 4 includes no change in the marginal tax rate. Figure 4 shows how bunching at a notch can occur for isoelastic 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 $\eta_g(\beta)$ where the indifference curve passing through the notch point on the first segment is tangent to the second segment. The notch probability is

$$P_K = \int_{K\rho_1^{-\beta}}^{\eta_g(\beta)} \phi(\eta) d\eta, \qquad (2.5)$$

and the bunching interval is now $[K\rho_1^{-\beta}, \eta_g(\beta)]$.

The probability of bunching at a notch is not informative about β for similar reasons that the probability of bunching at a kink is not informative. For $P_K > 0$ and any β we can choose $\phi(\eta)$ such that equation (2.5) is satisfied. Thus the probability of bunching at a a notch provides no information about the size of β. Unlike a kink, for isoelastic utility the entire distribution of taxable income does identify β from the size of the "gap" between K and $y_g(\beta)$, where no taxable income is observed, as shown in Bertanha, McCallum, and Seegert (2018) and Blomquist and Newey (2018). However this identification result for β depends crucially on the isoelastic utility specification and seems not relevant for applications where gaps are generally not present.

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, any 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 interval $[\eta_\ell, \eta_u]$ 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

$$\begin{aligned} P_K &= \int_{\eta_\ell}^{\eta_u} \phi(\eta)\, d\eta = \frac{1}{2}\left[\phi(\eta_\ell) + \phi(\eta_u)\right](\eta_u - \eta_\ell) \\ &= \frac{1}{2}\left[f^-(K)\rho_1^\beta + f^+(K)\rho_2^\beta\right]\left(K\rho_2^{-\beta} - K\rho_1^{-\beta}\right) \\ &= \frac{K}{2}\left[f^-(K) + f^+(K)(\rho_1/\rho_2)^{-\beta}\right]\left[(\rho_1/\rho_2)^\beta - 1\right]. \end{aligned} \tag{2.6}$$

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 preference density over the bunching interval $[\eta_\ell, \eta_u]$. More generally equation (2.6) imposes the trapezoid assumption

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

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 $\int_{\eta_\ell}^{\eta_u} \phi(\eta)\, d\eta$ varies with the choice of $\phi(\eta)$ in the bunching interval.

It is argued in Saez (2010, Section IB) and Kleven (2016, p. 440) that the use of a linear pdf $\phi(\eta)$ in the bunching interval should not be seen as a restrictive functional form assumption, but as an approximation that works well if the kink is sufficiently small. The idea being that, if the bunching interval is very small, any smooth density will look linear in this very small interval. However, for this to be a meaningful idea we must have some way to determine whether a given kink (bunching interval) is small or not. We can see from equation (2.4) that the bunching interval depends on the location of the kink, the slope before and after the kink, and the taxable income elasticity. We do not know the taxable income elasticity, which means that we have no operational way to tell whether a given kink (bunching interval) is small or not, so this reasoning is circular. Since, in the absence of knowledge of the taxable income elasticity, we cannot tell whether a given kink (bunching interval) is small or not, the idea that the approximation works well for small kinks is vacuous.

It is also thought, e.g. in Chetty (2011) and Kleven (2016, p. 450), that a polynomial fit to the observed distribution of taxable income, excluding data in a range around the kink point, can be used to predict a counterfactual distribution of taxable income in the kink region. However, what one is actually doing here is completely specifying the exact value of the preference density over the bunching interval to be that implied by the polynomial that fits the data away from the bunching interval. Instead of specifying a linear pdf in the bunching interval one is specifying that the pdf in the bunching interval has the polynomial form that fits the distribution of earnings outside the excluded range. If this polynomial specification is incorrect then the elasticity may be incorrect. There is no information in the data about the value of the density over the bunching interval. The polynomial approach gives just one specification of the preference density over the bunching interval but there are many specifications consistent with the data that one could choose. Theorem 1 shows that any taxable income elasticity is consistent with some specification of the preference distribution over the bunching interval.

Any choice of $\phi(\eta)$ in the bunching interval, such as a linear or polynomial density, is extrapolation, meaning an assumption about the preference distribution that is not based on data. The bunching probability is jointly determined by β and the preference density $\phi(\eta)$ in the bunching interval, so there is no information in the bunching probability about the density $\phi(\eta)$. This information must come from a source other than the observed distribution of taxable income, i.e. from extrapolation.

Identification of β from bunching should not be thought of as nonparametric. We have shown here that existing identification methods are based on completely and correctly specifying the preference density over the bunching interval. Saez (2010) implicitly specifies the preference density to be linear (or the trapezoid formula to hold). Chetty et al. (2011) explicitly specifies the preference density to be polynomial. These assumptions are more restrictive than specifying that the density over the bunching interval is parametric because they specify the exact value of the density. This is a stronger assumption than is made even in parametric models of distributions, so that identification of β from bunching cannot be considered to be nonparametric.

3 Elasticity Bounds for a Single Kink

For isoelastic utility, bounds on the preference density will imply bounds on β. Panel B of Figure 3 illustrates that in order to obtain a given probability of bunching at a given kink, the preference density will need to be large over the bunching interval when the true β is small or small when the true β is large. This pattern suggests that prior knowledge of bounds on the preference density could lead to bounds on β. We specifically consider elasticity bounds based on upper and lower bounds on the preference density. The bounds we give will also be satisfied if the preference density 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 density.

The bounds we give here depend on the correctness of the isoelastic utility specification, in particular on the taxable income elasticity being the same for all individuals. Our purpose is to use these bounds to check sensitivity of isoelastic estimates to assumptions about the density of preferences, for which purpose the isoelastic assumption seems sufficient. Bounds for the average elasticity over individuals located at a kink, with nonparametric preferences and elasticities that vary over individuals, are given in Blomquist et al. (2015).

To apply the bounds for isoelastic utility we need to account for a common feature of data that the proportion of individuals that locate exactly at a kink is very close to zero. Instead of a positive proportion of individuals at a kink data tend to display a sharp increase in the density of individuals as taxable income nears the kink. This data feature is often explained as resulting from optimization errors, meaning variations in taxable income away from utility maximization. Individuals who would locate at the kink if they were maximizing utility instead locate near the kink. The high density of individuals near the kink could also be explained by measurement error, though many modern administrative data sets are thought to be accurate enough that measurement error is low.

In the bunching literature optimization errors are accounted for by specifying an *excluded range* $[y_1, y_2]$, containing the kink, such that optimization errors can be ignored for income outside this range, at least for purposes of estimating the taxable income elasticity. We follow this approach by specifying bounds for the taxable income elasticity that only use the distribution of taxable income outside an excluded range. The idea is that the distribution of taxable income within the excluded range is potentially contaminated by optimization errors in such a way that the bounds for β would be incorrect, while the distribution outside the excluded range is not contaminated in this way. Sensitivity to the excluded range can be checked by varying $[y_1, y_2]$.

In practice the excluded range $[y_1, y_2]$ is often chosen by picking y_1 to be a value equal to or smaller than where the taxable income density begins to rise as the kink is approached from below and by picking y_2 to be a value where the density has returned back to a level that does not appear to be related to the kink. These choices are generally made by examining a graph of the taxable income density by eye, as we will illustrate in Section 4. This method implicitly assumes that for learning about the taxable income elasticity the optimization errors are accounted for by only using the distribution of taxable income outside the excluded range.

We consider bounds when $y_1 < K < y_2$ for the kink K. Let $\eta_1 = y_1\rho_1^{-\beta}$ and $\eta_2 = y_2\rho_2^{-\beta}$ denote lower and upper endpoints for η that correspond to y_1 and y_2 respectively. Also let

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

Consider the two functions

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

We have the following result:

THEOREM 2: *If $F(y)$ equals the CDF of taxable income obtained by maximizing the utility function in equation (2.3), η is continuously distributed with density $\phi(\eta)$, and there are positive scalars $\bar{\sigma} \geq 1$ and $\sigma \leq 1$ such that*

$$\sigma \min\{\phi(\eta_1), \phi(\eta_2)\} \leq \phi(\eta) \leq \bar{\sigma} \max\{\phi(\eta_1), \phi(\eta_2)\},\ \textit{for}\ \eta \in [\eta_1, \eta_2], \tag{3.7}$$

then the taxable income elasticity β satisfies

$$\sigma \min\left\{D^-(\beta), D^+(\beta)\right\} \leq \Pr(y_1 \leq Y \leq y_2) \leq \bar{\sigma}\max\{D^-(\beta), D^+(\beta)\}. \tag{3.8}$$

If $\phi(\eta)$ is monotonic then these bounds hold for $\sigma = \bar{\sigma} = 1$. If $\Pr(y_1 \leq Y \leq y_2) < \sigma(y_2 - y_1)\min\{f^-(y_1), f^+(y_2)\}$ then there is no β satisfying equation (3.8). Otherwise the set of all nonnegative β satisfying this equation is a nonempty subset of $[0, \infty)$. In addition these bounds are sharp, meaning that for any β satisfying equation (3.8) there is $\phi(\eta)$ such that $\Pr(y_1 \leq Y \leq y_2) = \int_{\eta_1}^{\eta_2}\phi(\eta)d\eta$ and equation (3.7) is satisfied.

Intuitively, upper and lower bounds on the preference density provide information about the elasticity because the length of the bunching interval increases monotonically in β. The upper bound on $\phi(\eta)$ rules out small values of β that make the bunching interval so small that the preference density would have to exceed its upper bound in order to match $\Pr(y_1 \leq Y \leq y_2)$. Similarly, the lower bound on $\phi(\eta)$ rules out large values of β that make the bunching interval so large that the preference density would have to be below its lower bound to match $\Pr(y_1 \leq Y \leq y_2)$.

The sharpness of the bounds given here depends on only using the distribution of taxable income outside the excluded range (y_1, y_2). If the distribution inside the excluded range were informative about the distribution of preferences then these bounds would not be sharp. We follow the literature in assuming that because of optimization errors the distribution of taxable income inside the excluded range is not informative so that only the data outside the excluded range should be used. In the empirical application we also consider the sensitivity of the bounds to the length of the excluded range.

The bounds given here are based on the a priori restriction in equation (3.7) for the preference density. For $\sigma = \bar{\sigma} = 1$ this inequality imposes the restriction that the density of η across the kink is bounded between the maximum and minimum of the density 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. Such an assumption might be plausible if the density of taxable income appeared to be monotonic increasing (or decreasing) before and after the excluded range. Equation (3.7) with $\sigma = \bar{\sigma} = 1$ would not be very plausible if the excluded range was thought to include a mode in the distribution of preferences. In that case it might be more plausible to set $\bar{\sigma} > 1$ equal to the ratio of an upper bound on the model to maximum density at the ends of the excluded range.

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, so that we would expect to find tighter bounds in such locations. Note though that if one wanted to include the possibility of a mode in an excluded range on the right side then one would want to choose $\bar{\sigma} > 1$, which would increase the width of the bounds.

It is difficult to economically motivate or justify the assumptions on which the bounds are based. The bounds are based on assumptions about the preference density, i.e. about tastes, over a kink, where the data provides no information about the preference density, as discussed in Section 2. Consequently bounds on the preference density must come from information other than that provided by data, i.e. are extrapolation. Extrapolation is based entirely on researchers views about how identified features extend to unidentified ones.

No assumptions or extrapolations about preferences are required for many of the bounds in the econometrics literature. For example, the Manski (1989) selection bounds impose no restrictions on the distribution of preferences or other unobservables. Also the Haile and Tamer

(2003) bounds for auctions do not impose such restrictions, but instead use economic behavior to bound the distribution of preferences, which is the distribution of auction valuation. No economic behavior is used to construct the bounds given here. Instead they are based entirely on restricting the distribution of preferences, on extrapolating from where the data provides information to where data does not.

These bounds do depend on the restrictive isoelastic utility specification. As discussed following Theorem 1, identification is more difficult in more general models. Consequently, because the bounds are sharp, we know that bounds for the average taxable income elasticity would only be wider in more general models.

One could think of these bounds as a sensitivity check on how the results are affected by allowing variation in the preference density in the bunching interval. This could make them difficult to interpret because different individuals might have different ideas about upper and lower bounds on the density of preferences. Also, these bounds are correct only if they are not affected by optimization errors outside the excluded range.

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

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

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

$$\begin{aligned}\bar{\sigma}\max\left\{\hat{D}^{-}\left(\hat{\beta}_\ell\right), \hat{D}^{+}\left(\hat{\beta}_\ell\right)\right\} &= \widehat{\Pr(y_1 \le Y \le y_2)}, \qquad (3.9)\\ \sigma\min\left\{\hat{D}^{-}\left(\hat{\beta}_u\right), \hat{D}^{+}\left(\hat{\beta}_u\right)\right\} &= \widehat{\Pr(y_1 \le Y \le y_2)}.\end{aligned}$$

4 An Application to US Data

In this Section we analyze the federal income tax application in Saez (2010, Figure 6A) for married joint tax filers for the years 1960 to 1969, for isoelastic utility. 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 5, Panels A and B, 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. We have chosen the bandwidth to be 500 as did Saez (2010). 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 filers 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 fairly low 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).

To help us select the excluded range we plot in Figure 5 the histogram and kernel density estimate with vertical lines located at $K = 20,000$ and the endpoints $K \pm 2000$ in Panel A and $K \pm 4000$ in Panel B. We look for a sharp increase and decrease in the estimated density around the kink. To us it appears that there may be some effect of the kink outside an excluded range $K \pm 2000$ and no effect of the kink outside an excluded range $K \pm 4000$. Based on these observations we report results for four choices of excluded range, $K \pm 1000$, $K \pm 2000$, $K \pm 3000$, and $K \pm 4000$.

For purposes of comparison we compute versions of the Saez (2010) and Chetty et al. (2011) bunching estimators for each of the excluded ranges. We use data outside the excluded range to estimate $f^-(K)$, $f^+(K)$, and the counterfactual density within the excluded range. The kink probability P_K is estimated as the excess mass that is the area between the observed and predicted counterfactual densities inside the excluded range. We then plug-in the estimates of $f^-(K)$, $f^+(K)$, and P_K to a formula for the elasticity obtained by specifying the distribution of preferences over the bunching interval. For the Saez (2010) method we estimate $f^-(K)$ as the average density in the interval to the left of y_1 that is the same length as $K - y_1$ and analogously on the right. In the next step we apply equation (2.6) of this paper. For the Chetty et al. (2011) method we fit a seventh-order polynomial using the entire observed distribution outside the excluded range to predict the counterfactual distribution inside the range. We then apply equation (6) in Chetty et al. (2011) which makes a different functional form assumption on the preference density inside the bunching interval compared to equation (2.6). In columns 2 and 3 of Table 1 we report the resulting Saez (2010) and Chetty et al. (2011) elasticity estimates.

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

Excluded Range	Saez	Chetty et al.	Lower bound	Upper bound
$K \pm 1000$	.088	.156	.036	.144
$K \pm 2000$	.157	.247	0	.385
$K \pm 3000$	.335	.378	0	.908
$K \pm 4000$	.390	.454	0	1.516

Notes: $K = 20,000$. The second and third columns are the Saez (2010) and Chetty et al. (2011) estimates described in this Section. The estimates of the bounds in the fourth and fifth columns are from equation (3.9).

We find an elasticity estimate of 0.157 for the $K \pm 2000$ range in column 2 of Table 1 which is quite close to Saez's (2010) preferred estimate of 0.170 using a $K \pm 1500$ range. Columns 2 and 3 of Table 1 reveal some differences between estimates from Saez (2010) and Chetty et al. (2011) methods. Also, the elasticity estimates in columns 2 and 3 increase as the excluded range increases.

In columns 4 and 5 of Table 1 we provide estimates of the bounds on β using equation (3.9). We use the same estimates of $f^-(y_1)$ and $f^+(y_2)$ as we did for the Saez (2010) estimates in column 2. We set $\sigma = \bar{\sigma} = 1$, which includes the case where the preference density is monotonic. The purpose of constructing these bounds is to provide information about the elasticity under weaker functional form assumptions than those made by previous bunching methods. The bounds are useful for assessing the sensitivity of elasticity estimates to functional form assumptions. When the excluded range is $K \pm 1000$ the Chetty estimate exceeds the upper bound, which can occur because the Chetty elasticity estimate uses different estimators of $f^-(y_1)$ and $f^+(y_2)$ than the bounds use.

The estimated bounds in columns 4 and 5 of Table 1 are quite wide and the width increases with the size of the excluded range. The estimated bounds vary as the excluded range changes for two reasons. First, as the excluded range widens, the observed mass inside the range becomes consistent with more extreme combinations of counterfactual distribution and excess mass. Each combination is in turn consistent with a range of preference densities inside the bunching interval and thus with multiple elasticities. Second, the variability of counterfactual densities inside the excluded range is constrained by the difference between the densities at the edges $f^{-}(y_1)$ and $f^{+}(y_2)$. This difference increases as the excluded range widens, since the taxable income density on either side of the excluded range is steeply increasing in this application, and this reinforces the widening of the bounds.

One could construct tighter bounds by putting more restrictions on the density of preferences. However, all such bounds are based entirely on prior information when there is only a single budget set and data inside the excluded range are not informative.

5 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 $\rho^{\beta}\eta$ being the choice of taxable income for a linear budget set with slope ρ and Theorem 2 of Blomquist et al (2015), 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 heterogeneity bunching interval for η corresponding to y_1 and y_2 respectively. The bounds on the density are that for $\eta \in (\eta_1, \eta_2)$,

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

Also, $\phi(\eta_1)$ and $\phi(\eta_2)$ are given by $\phi(\eta_1) = f^-(y_1)\rho_1^\beta$, $\phi(\eta_2) = f^+(y_2)\rho_2^\beta$. Then we have

$$\begin{aligned} P &= \Pr(y_1 \le Y \le y_2) = \int_{\eta_1}^{\eta_2} \phi(\eta)d\eta \le (\eta_2 - \eta_1)\bar{\sigma} \max\{\phi(\eta_1), \phi(\eta_2)\} \qquad (5.10)\\ &= \bar{\sigma}[y_2\rho_2^{-\beta} - y_1\rho_1^{-\beta}] \max\{f^-(y_1)\rho_1^\beta,\ f^+(y_2)\rho_2^\beta\} = \bar{\sigma} \max\{D^-(\beta), D^+(\beta)\},\\ P &\ge \sigma \min\{D^-(\beta), D^+(\beta)\}, \end{aligned}$$

where the second inequality follows similarly to the first, giving the first conclusion.

Next, 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 $\beta = 0$,

$\Pr(y_1 \le Y \le y_2) < \sigma(y_2 - y_1) \min\{f^-(y_1), f^+(y_2)\}$

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

Therefore if

$$P < \sigma(y_2 - y_1) \min\{f^-(y_1), f^+(y_2)\} = \sigma \min\{D^-(0), D^+(0)\},$$

it follows that $P < \sigma \min\{D^-(\beta), D^+(\beta)\}$ for all β, implying the second conclusion.

Next, consider the case where $P \ge \sigma \min\{D^-(0), D^+(0)\}$.

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

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 sharpness, consider any β such that equation (3.8) is satisfied. Let $\eta_2 = y_2\rho_2^{-\beta}$, $\eta_1 = y_1\rho_1^{-\beta}$, $\phi(\eta) = P/(\eta_2 - \eta_1) := \bar{\phi}$ for $\eta_1 < \eta < \eta_2$, $\phi(\eta) = f^-(\rho_1^\beta \eta)\rho_1^\beta$ for $\eta < \eta_1$, and $\phi(\eta) = f^-(\rho_2^\beta \eta)\rho_2^\beta$, $\eta > \eta_2$. As in Theorem 1 $\phi(\eta)$ is constructed so that the pdf of $Y = \eta\rho^\beta$ is $f(y)$ for $y \notin [y_1, y_2]$. Also, by construction,

$$\int_{\eta_1}^{\eta_2} \phi(\eta)d\eta = \int_{\eta_1}^{\eta_2} \frac{P}{\eta_2 - \eta_1}\phi(\eta)d\eta = P.$$

Also, by $\bar{\phi} = P/(\eta_2 - \eta_1)$ and equation (5.10)

$$\bar{\phi} = \frac{P}{\eta_2 - \eta_1} \leq \frac{(\eta_2 - \eta_1)\bar{\sigma} \max\{\phi(\eta_1), \phi(\eta_2)\}}{\eta_2 - \eta_1} \leq \bar{\sigma} \max\{\phi(\eta_1), \phi(\eta_2)\}.$$

It follows similarly that $\bar{\phi} \geq \sigma \min\{\phi(\eta_1), \phi(\eta_2)\}$, so that $\phi(\eta)$ satisfies the bounds on the density, showing sharpness. *Q.E.D.*.

6 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," draft, 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.

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.

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.

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

Haile, P. and E. Tamer (2003), "Inference With An Incomplete Model of English Auctions", *Journal of Political Economy* 111, 1-51.

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

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.

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 (2020), "Estimating Taxable Income Responses with Elasticity Heterogeneity," *Journal of Public Economis*, forthcoming.

Manksi, C. (1989), "Anatomy of the Selection Problem," *Journal of Human Resources* 24, 343-360.

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

McCallum, A.H., and N. Seegert (2017), “Better Bunching, Nicer Notching,” draft, August 2017.

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

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

Saez, E., J. Slemrod, and S. Giertz (2012), "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature* 50, 3-50.

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.

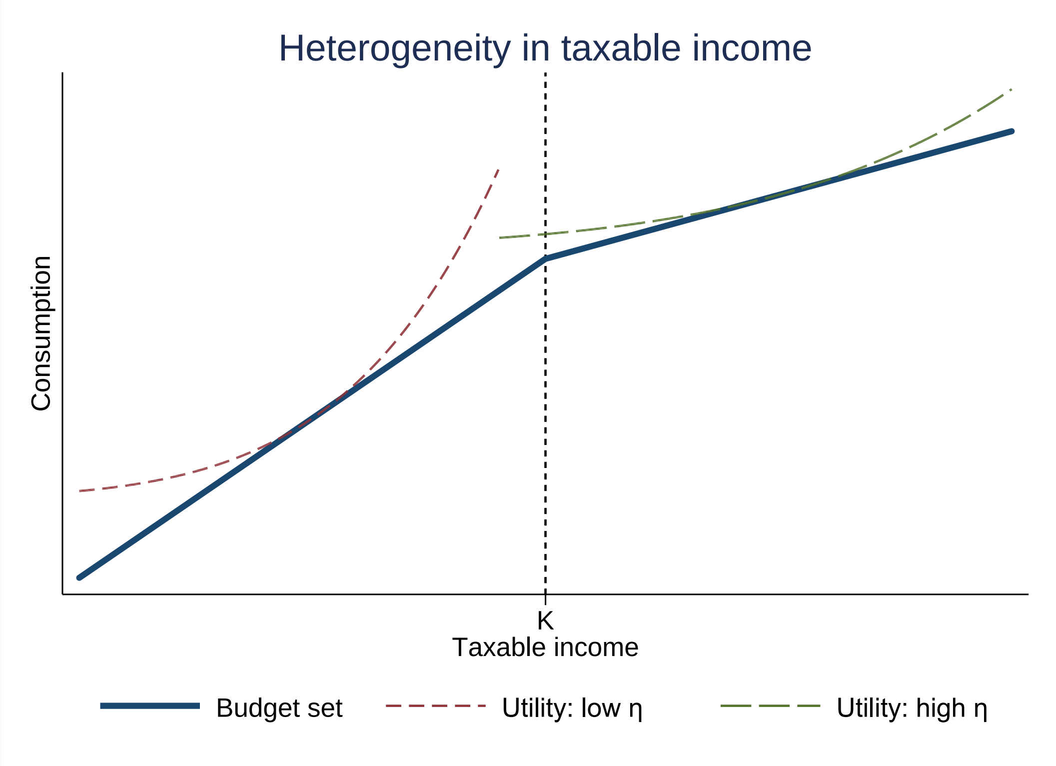

Figure 1. Heterogeneity in taxable income

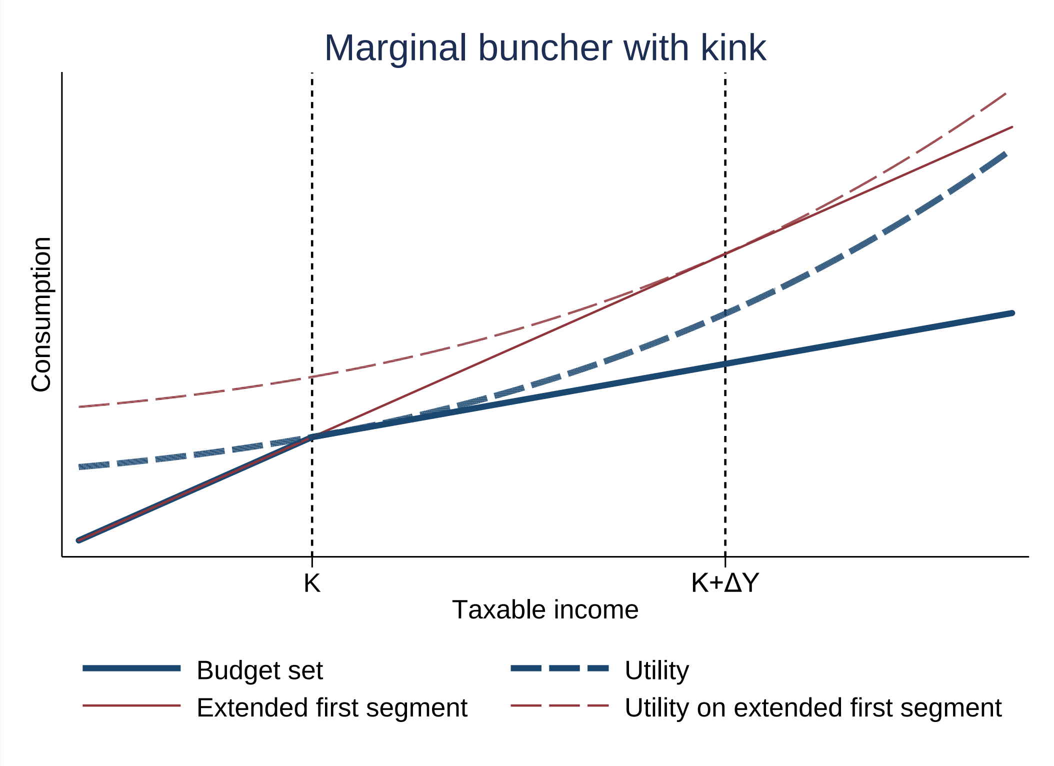

Figure 2. Marginal buncher with kink

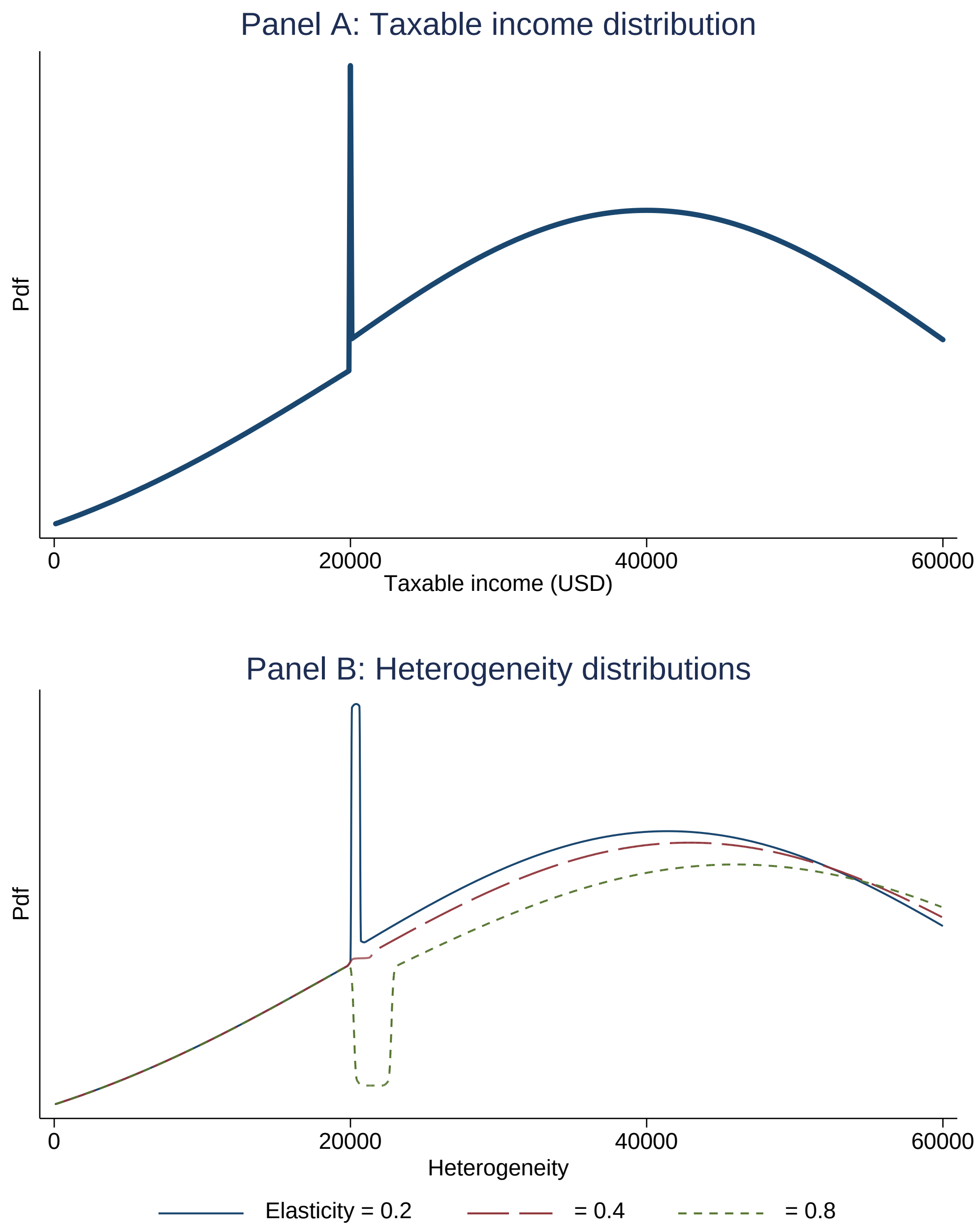

Figure 3. Nonidentification: multiple combinations of elasticity and heterogeneity distribution give the same taxable income distribution

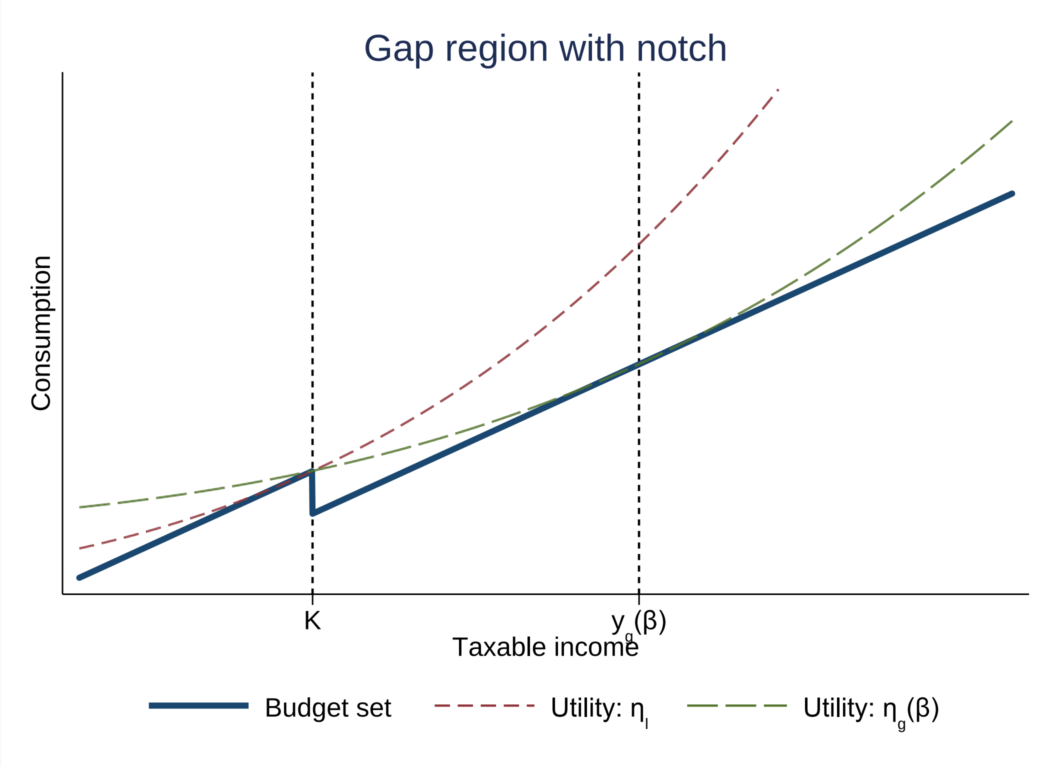

Figure 4. Gap region with notch

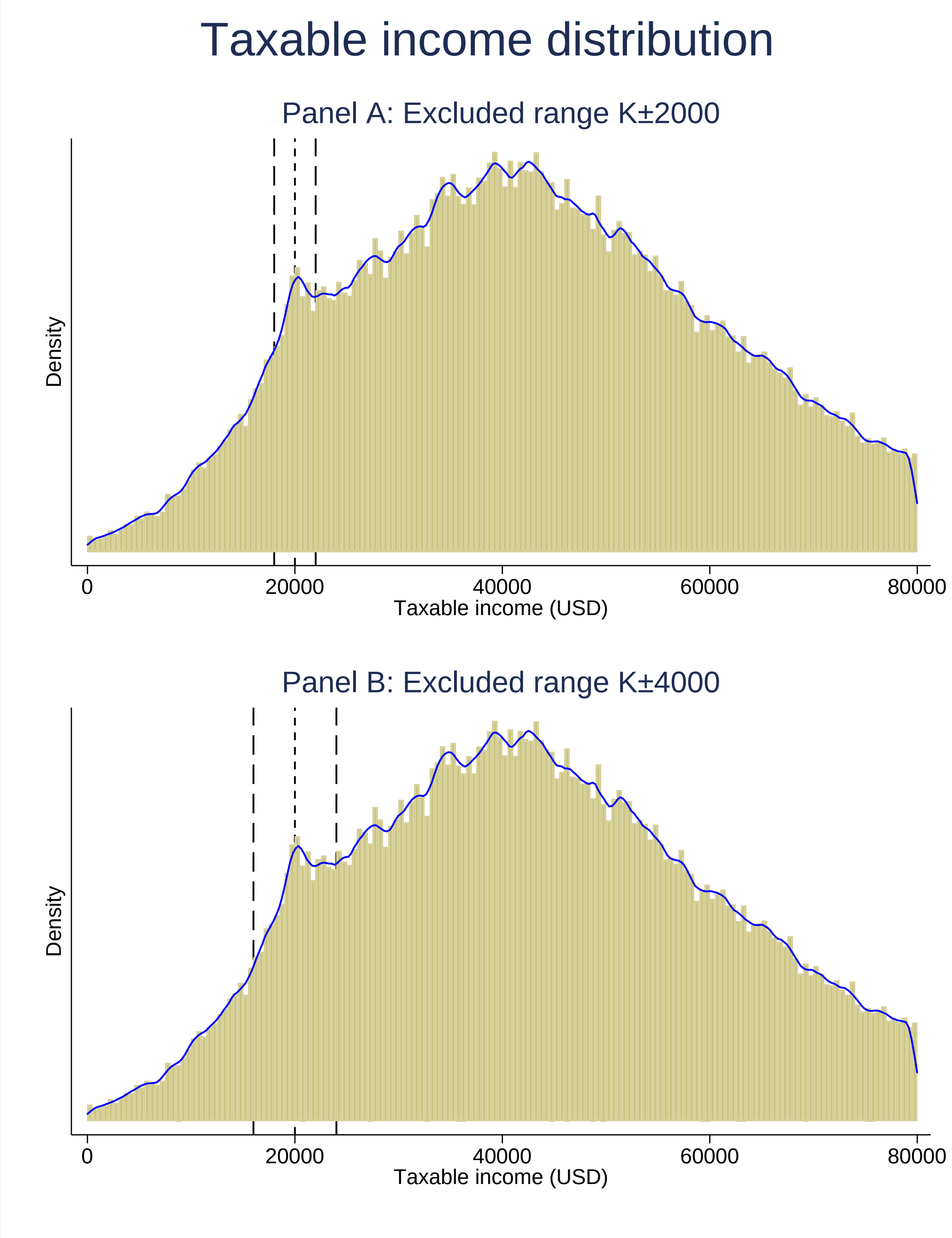

Figure 5. Taxable income distribution for joint filing married taxpayers 1960-1969

Notes: The data contain 170,973 observations. We use bins of 500 USD for the histogram and a bandwidth of 500 USD for the kernel. The kink at $K = 20{,}000$ USD is marked with a short-dashed vertical line. The long-dashed vertical lines mark the excluded ranges of $K \pm 2000$ in Panel A and $K \pm 4000$ in Panel B.