ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Blomquist, Sören; Newey, Whitney K.

Working Paper

The Bunching Estimator Cannot Identify the Taxable Income Elasticity

CESifo Working Paper, No. 6736

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Blomquist, Sören; Newey, Whitney K. (2017) : The Bunching Estimator Cannot Identify the Taxable Income Elasticity, CESifo Working Paper, No. 6736, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/173012

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CESifo WORKING PAPERS

6736
2017

November 2017

The Bunching Estimator Cannot Identify the Taxable Income Elasticity

Sören Blomquist, Whitney K. Newey

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6736*
Category 1: Public Finance

The Bunching Estimator Cannot Identify the Taxable Income Elasticity

Abstract

Saez (2010) introduced an influential estimator that has become known as the bunching estimator. Using this method one can get an estimate of the taxable income elasticity from the bunching pattern around a kink point. The bunching estimator has become popular, with a large number of papers applying the method. In this paper, we show that the bunching estimator cannot identify the taxable income elasticity when the functional form of the distribution of preference heterogeneity is unknown. We find that an observed distribution of taxable income around a kink point in a budget set can be consistent with any taxable income elasticity if the distribution of heterogeneity is unrestricted. If one is willing to assume restrictions on the heterogeneity density some information about the taxable income elasticity can be obtained. We give bounds on the taxable income elasticity based on monotonicity of the heterogeneity density and apply these bounds to the data in Saez (2010). We also consider identification from budget set variation. We find that kinks alone are still not informative even when budget sets vary. However, if the taxable income specification is restricted to be of the parametric isoelastic form assumed in Saez (2010) the taxable income elasticity can be well identified from variation among linear segments of budget sets.

Sören Blomquist
Department of Economics
Uppsala University
Sweden - 751 20 Uppsala
soren.blomquist@nek.uu.se

Whitney K. Newey
Department of Economics
M.I.T.
USA - Cambridge MA 02139
wnewey@mit.edu

2017-10-25
We want to thank Johan Lyhagen for introducing Sören Blomquist to the programming language R and get the package bunchr to run.

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. The usual way to estimate the taxable income elasticity has been to use data from several tax systems at different points in time.[1] A major challenge for this approach is to account for exogenous productivity growth, which would change the taxable income even if there were no behavioral changes.

An influential paper, Saez (2010), introduced what has become known as the bunching estimator.[2] According to this paper one can infer an interesting behavioral parameter, the taxable income elasticity, without any variation in a budget constraint. Saez derives a procedure that shows how one can get an estimate of the taxable income elasticity from the bunching pattern around a kink point. This is a quite remarkable result and differs methodologically from other empirical methods that use variation in budget constraints to identify the taxable income elasticity. If the bunching estimator worked it would be a major advance. Since data from only one point in time is needed one would not have to worry about exogenous productivity growth. The bunching estimator has become popular, and there are a large number of papers that apply Saez (2010)'s method.[3]

Unfortunately, the bunching estimator cannot identify the taxable income elasticity when the functional form of heterogeneity is unknown. The problem is that a kink probability may be high or small because of shapes of indifference curves or because more or fewer

[1] See for example Gruber and Saez (2002).
[2] Saez (1999) is a first version of the paper.
[3] 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. There are about 630 Google Scholar citations to Saez (2010), and the paper is on the curriculum in many graduate public economics courses around the world.

individuals like to have taxable income around the kink. Intuitively, for a single budget-set variation in the tax rate only occurs with variation in preferences. This conjoining of individual heterogeneity and variation in the tax rate makes it impossible to nonparametrically distinguish the taxable income elasticity from heterogeneity with a single budget set.

Nonidentification can also be explained in terms of order conditions. A kink probability is just one-reduced form parameter and so can identify just one structural parameter. The elasticity and heterogeneity parameters are not separately identified from the kink probability. We show that for the isoelastic specification the kink is completely uninformative about elasticity when the density of heterogeneity is unknown. Any elasticity is consistent with any kink probability for some choice of heterogeneity density. Furthermore, using more information about the distribution of taxable income along the budget set does not help. The order condition fails here also. The distribution of taxable income is one reduced-form "parameter," and there are two structural "parameters," the elasticity and distribution of heterogeneity. We show that for a single budget set and taxable income distribution any positive number could be the elasticity for some distribution of individual heterogeneity. The distribution of taxable income is totally uninformative about the taxable income elasticity when the distribution of heterogeneity is unknown and there is a single budget set.

A kink probability alone can identify only one structural parameter. Thus, everything about heterogeneity must come from somewhere else in order to get the elasticity from the kink probability. That is how the elasticity estimators in Saez (2010) and Chetty et. al. (2011) must work, and how they do work. Saez (2010) gets density estimates at the edge points from the budget set near the kink and then assumes the density is linear across the kink. Chetty et al. (2011) estimates a polynomial density near the kink and assumes the density is this polynomial across the kink.

A functional-form assumption for the heterogeneity density seems a very fragile assumption on which to hang identification of such an important structural parameter as taxable income elasticity. The heterogeneity is like a disturbance we might find in some other econometric model. The taxable income elasticity is an important structural parameter. It is unusual to rely so heavily on the functional form of a disturbance distribution for identification of a structural parameter. Instead we usually rely on variation in an observed variable, such as price or an instrument. Here it may seem that there is price variation as we move along the budget set, but that is incorrect. Different data points along a single budget set correspond to different individuals, so a single budget set does not allow us to distinguish the effect of changing the tax rate from heterogeneity.

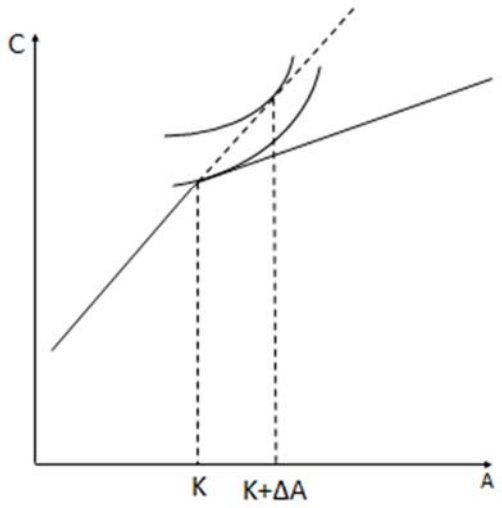
A kink may be informative about the elasticity when the heterogeneity density is restricted. We derive bounds on the elasticity when the heterogeneity density is monotonic over the kink. An assumption of monotonicity may seem reasonable when the kink occurs to one side of a unimodal distribution of taxable income. In an application like one of those in Saez (2010) we find these bounds to be very wide, so the kink is still not very informative. One could impose stronger restrictions on the heterogeneity density to shrink these bounds, like concavity. Of course all such bounds use information about the heterogeneity density to provide information about the elasticity, which is strong sensitivity of a structural parameter to disturbance distributions.

We also consider identification from budget-set variation. We find that kinks alone are still not informative when budget sets vary because the order condition is still not satisfied. In contrast, we do find that the elasticity may be identified from the distributions of taxable income from two distinct budget sets. We give a sufficient condition for identification of the elasticity for the isoelastic model, that tax rates must differ between the two budget sets over a "wide

enough" range of taxable incomes. We also discuss identification for models more general than the isoelastic specification.

Nonparametric models with general heterogeneity are considered in Blomquist et al. (2015). There it is shown that a parsimonius form for expected labor supply with scalar heterogeneity of Blomquist and Newey (2002) extends to general heterogeneity and taxable income. Also, Blomquist et al. shows how to impose all the conditions of utility maximization on expected taxable income and obtain the elasticity of expected taxable income.

Figure1



For simplicity we will focus much of the discussion on budget sets with one kink. Figure 1 illustrates such a budget set, with two linear segments with slopes $\theta_1 > \theta_2$ and a kink at K. What the researcher can observe is the income distribution along the kinked budget constraint. If there were no kink at K then there would be a smooth density function $f_1(a)$ of taxable income A along the extended first segment. However, due to the kink some individuals that otherwise would have had tangency solutions on the extended first segment are now located at the kink. A crucial step in the bunching estimation procedure is a comparison of the actual mass

of observations in an interval $(K-\delta, K+\delta)$ around the kink with the mass that would have been in the interval if there had been no kink. The actual mass in the interval can be observed. What the mass would have been in the interval, had there been no kink, must be estimated. A problem with such estimation is that all the individuals who would have been on the extended interval are now grouped at the kink.

Saez (2010) does suggest a procedure for how one can estimate $f_1(a)$ for individuals at the kink from the observed distribution of taxable income around the kink. We will see that this procedure corresponds to an assumption that the density function $f_1(a)$ is linear between the endpoints of the kink. Thus, the Saez (2010) bunching estimator depends on linearity of the density of $f_1(a)$ along the extended first segment. As mentioned, this seems to be a very strong functional-form assumption on which to hang the identification of the taxable income elasticity.

To illustrate nonidentification due to preference heterogeneity, consider the simple example in Figure 2. In this figure we show possible distributions of utility functions. In one of these distributions each individual has a large compensated taxable income elasticity, corresponding to a flat indifference curve, and the other a small taxable income elasticity corresponding to an indifference curve with larger curvature. As we have drawn the diagram, the income distributions are identical. In order not to clutter the diagram, we only show a few tangency points. We constructed the diagram such that at each tangency point we have one indifference curve corresponding to a large taxable income elasticity, the flatter indifference curves, and one corresponding to a low taxable income elasticity, the more curved indifference curves. At a point of tangency the slopes of the two indifference curves are the same, but the curvatures differ. There could be thousands, or millions, of tangency points, each constructed as the tangency points in the diagram.

Figure 2

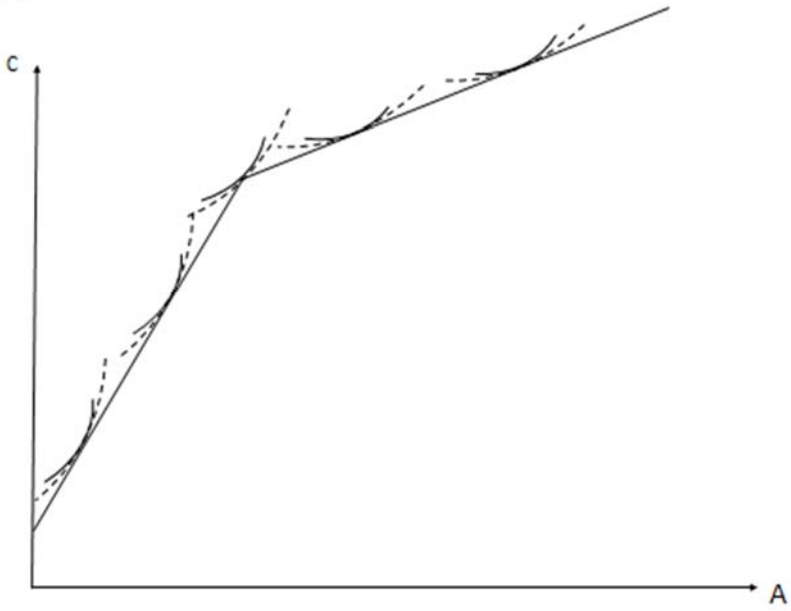


Figure 2 shows that we can have two identical income distributions where one income distribution comes from preferences with a large taxable income elasticity and the other from preferences with a low taxable income elasticity. We also assume that the indifference curves of individuals at the kink point have similar properties. The bunching estimator only uses information from the income distribution around a kink point. Hence, the bunching estimator must give the same result for the two (identical) income distributions, although they come from preferences implying different taxable income elasticities. This example shows that the bunching estimator cannot identify the taxable income elasticity.

We may also be unable to identify the taxable income elasticity because of optimization errors. That optimization errors make it problematic to estimate the structural taxable income elasticity is discussed in Saez (2010), Chetty et. al. (2011), and Kleven (2016). In what follows we discuss the impact of optimization errors.

Previous work has largely overlooked the lack of identification of the taxable income elasticity from kinks. Blomquist et al. (2015) did consider whether a kink nonparametrically identifies a weighted average of the compensated effect of taxes on taxable income for

individuals at a kink, with general preferences. That paper showed that the kink provides no information about the average tax effect, but that the effect is identified when the heterogeneity density is linear over the kink, and has identifiable bounds under monotonicity for that effect. Our results for the Saez (2010) utility function are analogous, showing kinks do not provide any information about the elasticity; that the elasticity is identified when the heterogeneity density is linear, and giving bounds under monotonicity. Our results are remarkably analogous to those of Blomquist et al. (2015), except that we integrate over scalar heterogeneity and Blomquist et al. (2015) over the slope of a budget line passing through the kink.

The rest of the paper is organized as follows. In Section 2 we describe the main ideas behind the bunching estimation procedure. In Section 3 we consider the same isoelastic utility function as Saez (2010) and show that a kink and the entire budget set provide no information about the taxable income elasticity when the heterogeneity distribution is unrestricted. We show that any positive taxable income elasticity can be obtained from a distribution of taxable income for one budget set by varying the distribution of heterogeneity. We also show that Saez (2010) implicitly assumes a linear heterogeneity density when estimating the elasticity from a kink. Section 4 illustrates how optimization errors hinder identification. We discuss various reasons for optimization errors and possible shapes for them. In Section 5 we perform a simulation exercise where we, for a given taxable income elasticity, vary the heterogeneity distribution and add various types of optimization errors. The simulations verify that the bunching estimator cannot identify the taxable income elasticity even in the absence of optimization errors. Adding optimization errors in general give estimates an order smaller in magnitude.

Section 6 gives bounds depending on the monotonicity of the heterogeneity density. Section 7 shows how observing more than one additional budget does not help with identification from kinks, but can lead to identification as a result of more comprehensive budget-set variation. Section 8 contains a brief summary and discussion.

2. The Bunching Estimator

We follow Saez (2010) when we describe the general idea behind the bunching estimator, but omit some details that are of no importance for our analysis. That paper first derives the bunching estimator for a small kink. When the analysis is extended to larger kinks a parametric, an isoelastic utility function, is used. In this section we describe the analysis for a small kink. The analysis using an isoelastic utility function follows in Section 3.

To establish how excess bunching at the kink is related to the taxable income elasticity we assume a strictly quasi-concave utility function $U(C, A, \rho)$ where C is consumption (disposable income), A taxable income, and ρ a random preference parameter following a continuous probability density function. It is assumed the taxable income function implied by the utility function is increasing in ρ. Heterogeneity of preferences is necessary in order for the bunching estimator to be of any interest. Since there is a single budget constraint, if preferences were homogenous we would have one point on a single budget constraint; no inference about preferences could be drawn from that. Since everyone faces the same budget constraint, heterogeneity is needed to create variation in taxable income.

There is a simple relationship between the taxable income elasticity and the curvature of the indifference curve. Consider an indifference curve defined by $U(C, A, \rho) = \bar{u}$ for fixed ρ and define the function $C = h(A, \rho, \bar{u})$. Let $h'(A, \rho, \bar{u}) = \partial h/\partial A$ and $h''(A, \rho, \bar{u}) = \partial^2 h/\partial A^2$. We note that h' is the slope of the indifference curve with utility level $\bar{u}$ and h'' the curvature of the indifference curve. One can show that if utility is maximized subject to a linear budget constraint with slope θ, then the compensated effect is given by $\partial A/\partial\theta = 1/f''$. The less curved an indifference curve is (small f''), the larger the $\partial A/\partial\theta$ and the taxable income elasticity are.

To derive the bunching estimator, Saez (2010) considers a counterfactual hypothetical change in a budget constraint. Suppose individuals maximize their utility subject to the

extended first segment illustrated in Figure 1. This would generate a smooth density, $f_1(a)$, of taxable income along the extended first segment. Suppose next that a kink at $A = K$ is introduced, and the slope of the budget constraint after the kink is $\theta_2 = \theta_1 + \Delta\theta, \Delta\theta < 0$. Some of the individuals who had a tangency solution above K on the extended segment will now instead choose the kink point K. This implies that there will be a mass of individuals locating at the kink, a spike in the distribution. We follow the literature and refer to this as bunching.

In Figure 1 we have drawn two indifference curves for the marginal buncher, i.e., the individual with the highest ρ that before the (hypothetical) change had a tangency on the extended first segment and after the change has a tangency at the kink. Before the (hypothetical) change in the budget constraint, the individual had a tangency on the extended segment at $K + \Delta A$, and after the change in the budget constraint a tangency on the second segment at K. The taxable income elasticity is

$$e = \frac{\Delta A/K}{\Delta\theta/\theta_1} \qquad (1)$$

However, in reality we cannot observe incomes at the individual level on the extended first segment. This means that we do not know ΔA. To overcome this lack of information Saez (2010) assumes that one can use observations along the kinked budget set to estimate the density $f_1(a)$ of taxable income along the extended first segment. This is a crucial assumption for the bunching method to work and, as we will show in the next section, it corresponds to assuming a functional form for $f_1(a)$ along the extended first segment.

We can observe the amount of bunching around the kink; we denote this bunching by B. This bunching consists of all individuals who would have had a tangency between K and $K + \Delta A$ along the extended first segment. Suppose we knew the density $f_1(a)$ of taxable income along the extended first segment. We could then use the relationship

$$B = \int_K^{K+\Delta A} f_1(a)da \qquad (2)$$

to calculate ΔA. We would then have all the pieces necessary to calculate the taxable income elasticity for the marginal buncher.

Saez (2010) notes that there might be optimization errors, which implies that some individuals might not be able to locate at the kink even if they would like to do so. This implies that instead of a pronounced spike at K, we would observe more of a hump in the distribution around the kink. Saez (2010) develops a technique for how to get an estimate of the excess bunching at the kink when there are optimization errors. Chetty et al. (2011) refines this technique. In the next section we will discuss precisely what is being assumed about the distribution of heterogeneity for these techniques.

Suppose that the distribution of taxable income A is uniform along an extended first segment $(K, K + \Delta A)$ with density $\overline{f_1}$ We can then rewrite equation (2) as $B = \overline{f_1}\Delta A$. Combining this with equation (1) we get

$$e = \frac{(B/\overline{f_1})/K}{\Delta\theta/\theta_1}. \qquad (3)$$

In the literature, the expression $B/\overline{f_1}$ is often called the excess bunching at the kink. The goal of the empirical work is to come up with an estimate of the excess bunching at the kink. Since in actual data there is rarely a spike at a kink, but more of a hump, one tries to estimate the excess bunching in an interval $(K - \delta, K + \delta)$. To achieve this, one divides the data into a number of equally-sized bins and constructs a histogram. From a visual inspection of the histogram one decides on the interval $(K-\delta, K+\delta)$. Using the distribution as measured by the number of observations in each bin one makes an estimate of the distribution along the extended first segment. In this estimation procedure one excludes the interval $(K - \delta, K + \delta)$. The excess bunching is measured as the actual number of observations in the bunching interval divided by the number predicted by the estimated counterfactual density $\bar{h}$ along the extended first segment.

The identification problem is that $\overline{f_1}$ is unknown. The density of taxable income along the extended first segment is not identified because all of those individuals who would have located there are now grouped at the kink. Furthermore, the value of $\overline{f_1}$ may be any nonnegative number, implying that the taxable income elasticity may be any non-negative number. In this sense the kink probability provides no information about the taxable income elasticity when there are no restrictions on the taxable income density.

Imposing smoothness and endpoint restrictions does not help with identification. We can fix $f_1(K)$ and $f_1(K + \Delta A)$ and their derivatives of all orders and still obtain any value of $\int_K^{K+\Delta A} f_1(a)da$ by varying $f_1(a)$ on the interior of the interval. Therefore, the taxable income elasticity may be anything depending on the value of the integral, so that the kink provides no information about the taxable income elasticity, even when the density satisfies endpoint restrictions and is continuously differentiable of all orders.

We have shown nonidentification of a discrete version of the taxable income elasticity, sometimes referred to as an arc elasticity. In the next section we will show that nonidentification also holds for the isoelastic model.

3. Nonidentification with Taxable Income Function $A = \rho\theta^{\beta}$

In this section we show that the taxable income elasticity is not identified from bunching for the isoelastic utility specification of Saez (2010). We continue to proceed under the assumption that there are no optimization errors. In the next section we will consider optimization errors. We also assume the taxable income elasticity is the same for everyone. Using these assumptions will bias the setup towards identification. Still, even with these strong restrictive assumptions we show that the bunching estimator cannot identify the taxable income elasticity. We also show that the bunching estimator of Saez (2010) is based on a linear density assumption.

The isoelastic utility function as used in Saez (2010) to derive the bunching estimator is:

$$U(C,A,\rho) = C - \frac{\rho}{1+\frac{1}{\beta}}\left(\frac{A}{\rho}\right)^{1+\frac{1}{\beta}}, \ \rho > 0 \ , \beta > 0 . \qquad (4)$$

Maximizing this utility function subject to a linear budget constraint with slope θ gives the taxable income function $A = \rho\theta^{\beta}$; the taxable income elasticity will be constant and is given by β. There are no income effects. The variable ρ represents unobserved individual heterogeneity in preferences. It is the variation in ρ that generates a distribution of income along a budget constraint. We note that A is increasing in β, ρ and θ for $\beta > 0,\ \rho > 0,\ \theta > 0$

Given the kink point $A = K$, the slope θ_1 of the segment before the kink and the slope θ_2 of the segment after the kink, we can calculate the size of the bunching window for ρ, meaning the interval of ρ for which taxable income A will be at the kink. The highest value of ρ giving a tangency solution on the first segment is given by the relation $K = \rho\theta_1^{\beta}$, and the lowest value of ρ giving a tangency solution on the second segment is given by $K = \rho\theta_2^{\beta}$. The bunching window in terms of ρ is therefore given by $[K\theta_1^{-\beta}, K\theta_2^{-\beta}]$, so the kink probability is

$$B = \Pr(A = K) = \int_{K\theta_1^{-\beta}}^{K\theta_2^{-\beta}} \phi(\rho)d\rho, \qquad (5)$$

where $\phi(\rho)$ is the density of ρ.

Here we can clearly see the identification problem. The size of the bunching window is increasing in β, which implies that for a given preference distribution, the bunching itself is increasing in β. This is the main idea behind the bunching estimator; the higher the taxable income elasticity, the more bunching there will be. 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, the larger the bunching will be. Hence, for a given value of the taxable-

income elasticity, the amount of bunching can vary a lot depending on the shape of the preference distribution.

The bunching window in terms of ρ is well defined. The bunching window in terms of taxable income depends on how we define the counterfactual budget constraint. In the bunching literature it is assumed the extended first segment is the counterfactual. In this case the bunching window will be $A \in \left(K, K\theta_1^{\beta} / \theta_2^{\beta}\right)$, to the right of the kink as shown in Figure 1. This definition of the counterfactual is, of course, quite arbitrary. One could just as well consider the extended second segment to be the counterfactual. In this case the bunching window would be to the left of the kink. Or, we could let the counterfactual be a linear budget constraint passing through the kink point with a slope intermediate between θ_1 and θ_2. In this case the bunching window would be partly to the left and partly to the right of the kink. Our analysis shows there is no need to introduce a counterfactual. However, to relate our analysis to the bunching literature we introduce a counterfactual and consider the extended first segment to be the counterfactual budget constraint.

To illustrate nonidentification we will construct two data generating processes (dgp:s) that generate identical distributions of taxable income around a kink in a budget constraint, although the underlying preferences represent different taxable income elasticities. Since the bunching estimator only uses information on the income distribution around the kink, the bunching estimator must give the same estimate for the two data generating processes, although they represent different taxable income elasticities. This shows that the bunching estimator cannot identify the taxable income elasticity.

We assume individuals maximize utility subject to a budget constraint with a kink at $A = K$ and slope θ_1 before the kink and θ_2 after the kink. Let the first dgp be defined by the cumulative distribution function $\Phi(\rho), \rho \in \left(\underline{\rho}, \overline{\rho}\right)$ for the preference parameter and an elasticity β_ρ. Let us denote by $\overline{\rho}_1$ the highest ρ that gives a tangency on the first segment and

by $\underline{\rho}_2$ the lowest ρ that gives a tangency on the second segment. Then for $\rho \in \left(\underline{\rho}, \overline{\rho}^1\right)$ there will be a tangency solution on the first segment, a kink solution at $A = K$ for $\rho \in \left(\overline{\rho}_1, \underline{\rho}_2\right)$ and a tangency on the second segment for $\rho \in \left(\underline{\rho}_2, \overline{\rho}\right)$. Since the taxable income for a linear budget set is given by $A = \rho\theta^{\beta_\rho}$, it follows by Theorem 2 of Blomquist et al. (2015) that the cumulative distribution function for taxable income on the first segment is given by $F^1(A) = \Pr\left(\rho\theta_1^{\beta_\rho} \le A\right) = \Pr\left(\rho \le A / \theta_1^{\beta_\rho}\right) = \Phi\left(A / \theta_1^{\beta_\rho}\right)$, and the pdf for A is $f^1\left(A\right) = \phi\left(A / \theta_1^{\beta_\rho}\right)\dfrac{1}{\theta_1^{\beta_\rho}}$ for $A \in \left(\underline{A}, K\right)$ where $\underline{A} = \underline{\rho}\theta_1^{\beta_\rho}$. Similarly, the cumulative distribution function for $A \in [K, \overline{A})$, where $\overline{A} = \overline{\rho}\theta_2^{\beta_\rho}$, on the second segment is $F^2\left(A\right) = \Pr(\rho\theta_2^{\beta_\rho} \le A)$ $= \Pr\left(\rho \le A / \theta_2^{\beta_\rho}\right) = \Phi\left(A / \theta_2^{\beta_\rho}\right)$, and the pdf is $f^2\left(A\right) = \phi\left(A / \theta_2^{\beta_\rho}\right)\dfrac{1}{\theta_2^{\beta_\rho}}$. The probability that taxable income is at the kink is given by

$$B = \int_{\overline{\rho}^1}^{\underline{\rho}^2} \phi\left(v\right) dv = \int_{K\theta_1^{-\beta_\rho}}^{K\theta_2^{-\beta_\rho}} \phi\left(v\right) dv = \Phi\left(K\theta_2^{-\beta_\rho}\right) - \Phi\left(K\theta_1^{-\beta_\rho}\right) = F^2\left(K\right) - F^1\left(K\right). \qquad (6)$$

This is the basic bunching equation for the constant elasticity utility function from equation (4).

The second data generating process is defined by the cumulative distribution function $\Psi\left(\eta\right)$, $\eta \in \left(\underline{\eta}, \overline{\eta}\right)$ for the preference parameter and an elasticity $\beta_\eta \ne \beta_\rho$. Following the procedure used above we can derive the cumulative distribution function $G^1(A) = \psi\left(A / \theta_1^{\beta_\eta}\right)$, and the pdf is $g^1\left(A\right) = \psi\left(A / \theta_1^{\beta_\eta}\right)\dfrac{1}{\theta_1^{\beta_\eta}}$ for taxable income on the first segment. Likewise we derive the cumulative distribution function $G^2\left(A\right) = \Psi\left(A / \theta_2^{\beta_\eta}\right)$ and the pdf $g^2\left(A\right) = \psi\left(A / \theta_2^{\beta_\eta}\right)\dfrac{1}{\theta_2^{\beta_\eta}}$ for the second segment. The probability that taxable income is at the kink is given by $\int_{\overline{\eta}^1}^{\underline{\eta}^2} \psi\left(v\right) dv = \int_{K\theta_1^{-\beta_\eta}}^{K\theta_2^{-\beta_\eta}} \psi\left(v\right) dv = \Psi\left(K\theta_2^{-\beta_\eta}\right) - \Psi\left(K\theta_1^{-\beta_\eta}\right) = G^2\left(K\right) - G^1\left(K\right)$.

We want the two data generating processes to generate identical distributions of taxable income at and around the kink? For this to be true we must have $F^1\left(A\right) \equiv G^1\left(A\right)$, $A \in \left(\underline{A}, K\right)$,

$F^2(A) \equiv G^2(A)$, $A \in (K, \overline{A})$, and there must be the same mass at $A = K$. To ensure that the two dgp:s are defined on the same intervals we must set $\underline{A} = \underline{\rho}\theta_1^{\beta_\rho} = \underline{\eta}\theta_1^{\beta_\eta}$, implying $\underline{\eta} = \underline{\rho}\theta_1^{\beta_\rho - \beta_\eta}$. We must set $K = \overline{\rho}^1\theta_1^{\beta_\rho} = \overline{\eta}^1\theta_1^{\beta_\eta}$, implying $\overline{\eta}^1 = \overline{\rho}^1\theta_1^{\beta_\rho - \beta_\eta}$, $K = \underline{\rho}^2\theta_2^{\beta_\rho} = \underline{\eta}^2\theta_2^{\beta_\eta}$ implying $\underline{\eta}^2 = \underline{\rho}^2\theta_1^{\beta_\rho - \beta_\eta}$ and finally $\overline{A} = \overline{\rho}\theta_2^{\beta_\rho} = \overline{\eta}\theta_2^{\beta_\eta}$, implying $\overline{\eta} = \overline{\rho}\theta_2^{\beta_\rho - \beta_\eta}$. The requirement $F^1(A) \equiv G^1(A)$, $A \in (\underline{A}, K)$ implies $\psi\left(A / \theta_1^{\beta_\eta}\right) = \phi\left(A / \theta_1^{\beta_\rho}\right)\theta_1^{\beta_\eta - \beta_\rho}$, $A \in (\underline{A}, K)$ and vice versa. The requirement $F^2(A) \equiv G^2(A)$, $A \in (K, \overline{A})$ implies $\psi\left(A / \theta_2^{\beta_\eta}\right) = \phi\left(A / \theta_2^{\beta_\rho}\right)\theta_2^{\beta_\eta - \beta_\rho}$, $A \in (K, \overline{A})$ and vice versa. Finally for $\Pr(A = K)$ to be the same for the two dgp:s we must have $\int_{K\theta_1^{-\beta_\rho}}^{K\theta_2^{-\beta_\rho}} \phi(v)\,dv = \int_{K\theta_1^{-\beta_\eta}}^{K\theta_2^{-\beta_\eta}} \psi(v)\,dv$. In the derivation of the bunching formula Saez (2010) assumes that the distribution of taxable income along the extended segment 1 is smooth, we therefore require the pdf:s $f(A)$ and $g(A)$ to be continuous. A necessary condition for this to hold is that the pieces that give the kink solution connect smoothly to the distributions for segments 1 and 2.

We have shown how to construct two data generating processes that generate identical taxable income distributions along a kinked budget constraint, although the two taxable income functions have different taxable income elasticities. This shows that the bunching estimator cannot identify the taxable income elasticity.

In much of the bunching literature an essential part of the estimating procedure is to get an estimate of $f(a)$ along the extended first segment using information on the distribution of taxable income around the kink. Chetty et al. (2011) suggests a procedure that has become popular. It is therefore worth noting that although the two data generating processes defined above, by construction, give rise to identical income distributions along the kinked budget

constraint, the dgp:s imply different distributions of taxable income along the extended first segment. For dgp 1 the bunching window will be $A \in \left(K, K\theta_1^{\beta_\rho} / \theta_2^{\beta_\rho}\right)$, and for this interval there is no information on the distribution of A, since the $A's$ along the extended first segment are all stacked up at the kink. The distribution could be anything. The distribution of A after $K\theta_1^{\beta_\rho} / \theta_2^{\beta_\rho}$ will be $\phi\left(A / \theta_1^{\beta_\rho}\right)\frac{1}{\theta_1^{\beta_\rho}}$. For the second dgp the bunching window will be $A \in \left(K, K\theta_1^{\beta_\eta} / \theta_2^{\beta_\eta}\right)$, and in this window the density might be anything. The distribution after $K\theta_1^{\beta_\rho} / \theta_2^{\beta_\rho}$ will be $\psi\left(A / \theta_1^{\beta_\eta}\right)\frac{1}{\theta_1^{\beta_\eta}}$. Since we have constructed the dgp:s so that the distributions of taxable income are the same along the kinked budget constraint we have the relations $\psi\left(A / \theta_1^{\beta_\eta}\right) = \phi\left(A / \theta_1^{\beta_\rho}\right)\theta_1^{\beta_\eta - \beta_\rho}$, $A \in \left(\underline{A}, K\right)$ and $\psi\left(A / \theta_2^{\beta_\eta}\right) = \phi\left(A / \theta_2^{\beta_\rho}\right)\theta_2^{\beta_\eta - \beta_\rho}$, $A \in \left(K, \overline{A}\right)$. However, these relations do not imply any relations between $\phi\left(A / \theta_1^{\beta_\rho}\right)\frac{1}{\theta_1^{\beta_\rho}}$ and $\psi\left(A / \theta_1^{\beta_\eta}\right)\frac{1}{\theta_1^{\beta_\eta}}$ along the extended first segment. This implies that from the data around the kinked budget constraint we can neither identify what the distribution of taxable income along the extended segment would be nor identify what the bunching window would be. For example, if $K = 1000$, $\theta_1 = 0.5, \theta_2 = 0.7$, $\beta_\rho = 1, \beta_\eta = 0.2$ the bunching window along the extended first segment would be $(1000,1400)$ for the first dgp and $(1000,1070)$ for the second dgp. Any attempt to estimate the density $f_1(a)$ along the extended first segment from knowledge of the distribution of taxable income around the kink is therefore doomed to fail.

Using analogous reasoning we can show why any positive number will be the taxable income elasticity for some distribution of heterogeneity. Let $b > 0$ denote a possible value of β. We now construct a distribution function $\Phi(\rho)$ of heterogeneity such that the taxable income

distribution for elasticity b and heterogeneity $\Phi(\rho)$ is the distribution function $F(a)$ from the data. Let $\Phi(\rho) = F\left(\theta_1^b \rho\right)$ for $\rho < \theta_1^{-b} K$ and let $\Phi(\rho) = F\left(\theta_2^b \rho\right)$ for $\rho > \theta_2^{-b} K$. Suppose that the taxable income for a linear budget set is $\rho\theta^b$. By Theorem 2 of Blomquist et al. (2015), on the lower segment where $a < K$ the distribution of taxable income will be $\Pr\left(\rho\theta_1^b \leq a\right) = \Phi\left(\theta_1^{-b} a\right) = F(a)$. Similarly, on the upper segment where $a > K$, the distribution of taxable income will $\Pr\left(\rho\theta_2^b \leq a\right) = \Phi\left(\theta_2^{-b} a\right) = F(a)$. For $\theta_1^{-b} K \leq \rho \leq \theta_2^{-b} K$ let $\Phi(\rho)$ be any differentiable, monotonic increasing function such that $\Phi(\theta_1^{-b} K) = \lim_{a \to K, a < K} F(a)$ and $\Phi(\theta_2^{-b} K) = F(K)$. Then by construction, have

$$\Phi\left(\theta_2^{-b} K\right) - \Phi\left(\theta_1^{-b} K\right) = F(K) - \lim_{a \to K, a < K} F(a) = \Pr(A = K),$$

where the last equality holds by standard results for cumulative distribution functions. Also, we can choose $\Phi(\rho)$ so its derivatives of any order match those of $F(\theta_1^b \rho)$ at $\rho = \theta_1^{-b} K$ and those of $F(\theta_2^b \rho)$ at $\rho = \theta_2^{-b} K$. Thus we have the following result:

THEOREM 1: *Suppose that the CDF $F(a)$ of taxable income A is continuously differentiable of order D to the right and to the left at K and $B = \Pr(A = K) > 0$. Then for any β there exists $\Phi(\rho)$ such that the CDF of taxable income obtained by maximizing the utility function in equation (4) equals $F(a)$, and $\Phi(\rho)$ 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 matches derivatives to any finite order of the CDF of heterogeneity implied by the data. Thus the failure of identification of the taxable income elasticity from one budget set is complete, in the sense

that it has no information about the elasticity, when the distribution of heterogeneity is unrestricted.

We can see from equation (5) why the density $\phi(\rho)$ must be completely specified in the bunching interval in order to estimate the taxable income elasticity from the kink probability. If $\phi(\rho)$ depended on any unknown parameters then equation (5) could result in multiple values of the elasticity.

The Saez (2010) estimator is based on assuming two assumptions: that $\phi(\rho)$ is continuous so that the density at the bunching endpoints can be estimated from the linear segments and that assuming that the density is linear in the bunching interval. By continuity its value at endpoints can be obtained from the density of taxable income. 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. Let $\underline{\rho} = K\theta_1^{-\beta}$ and $\overline{\rho} = K\theta_2^{-\beta}$ be the endpoints of the bunching interval. Accounting for the Jacobian of the transformation $a = \rho\theta_1^{\beta}$ we have $\phi(\underline{\rho}) = f^{-}(K)\theta_1^{\beta}$ and $\phi(\overline{\rho}) = f^{+}(K)\theta_2^{\beta}$. Assuming that $\phi(\rho)$ is linear on the bunching interval we then have

$$B = \int_{\underline{\rho}}^{\overline{\rho}} \phi(\rho)dt = \frac{1}{2}\left[\phi\left(\underline{\rho}\right) + \phi(\overline{\rho})\right]\left(\overline{\rho} - \underline{\rho}\right) = \frac{1}{2}\left[f^{-}(K)\theta_1^{\beta} + f^{+}(K)\theta_2^{\beta}\right]\left(K\theta_2^{-\beta} - K\theta_1^{-\beta}\right)$$

$$= \frac{K}{2}\left[f_A^{-}(K) + f_A^{+}(K)(\theta_1/\theta_2)^{-\beta}\right]\left[(\theta_1/\theta_2)^{\beta} - 1\right].$$

This is the estimating equation found in equation (5) of Saez (2010).

Here we see that the Saez (2010) formula for the taxable income elasticity corresponds to imposing linearity on the heterogeneity density over the bunching interval $\left(\underline{\rho}, \overline{\rho}\right)$. We could obtain an analogous formula for the elasticity for other functional forms. Chetty et al. (2011) uses a polynomial. The elasticity estimate will generally vary with the choice of functional form of the heterogeneity density. Every bunching elasticity estimator is based on assuming a form of the heterogeneity density over the bunching interval.

Above we assumed that the taxable income elasticity is the same for each individual. The analysis can be extended to the case with heterogeneous taxable income elasticity. Suppose we have a utility function as defined by equation (4) and a kinked budget constraint with a kink at $A = K$ and slopes $\theta_1 > \theta_2$. A pdf $\phi(\beta, \rho)$ will then imply a unique pdf $f(A)$ along the kinked budget constraint. However, the reverse is not true. One cannot deduce the distribution $\phi(\beta, \rho)$ from knowledge of $f(a)$. In fact, there is an infinity of probability density functions $\phi(\beta, \rho)$ that could have generated the given pdf $f(a)$. These different pdfs would generate different distributions $f_1(A)$ along the extended first segment. The argument is applicable to other utility functions with two or more parameters; having a more general model must make identification more difficult.

4. Optimization Errors

To illustrate how optimization errors threaten identification of the taxable income elasticity we use an example. Let us consider two data-generating processes defined by different taxable income elasticities $\beta_1 > \beta_2$, but with the *same* unknown distribution $\phi(\rho)$ of the heterogeneous preference parameter. This gives rise to two distinct distributions of taxable income around a kink. Since the distributions of ρ are the same for the two dgps the bunching around the kink would be larger for the dgp with the greater taxable income elasticity. Hence, the two dgps would not be observationally equivalent. However, assume that there is a random additive optimization error so that the realized taxable income is $A = A^d + \varepsilon$, where A^d is desired taxable income and A is realized taxable income. Suppose the pdf for the optimization error for the first data generating process is given by $\gamma_1(\varepsilon)$ and by the resulting cumulative distribution $F_1(A)$, then we can find another distribution $\gamma_2(\varepsilon)$ that gives rise to a cumulative distribution $F_2(A)$ and such that $F_1(A)$ and $F_2(A)$ are identical. Hence, the existence of optimization errors can make the taxable income elasticity unidentifiable.

Why are there optimization errors? In a sense, the term "optimization error" is a misnomer. In our models we usually assume individuals can locate at any point on the budget constraint without any adjustment costs. In reality only some points on the budget constraint might be available and often there are (short run) costs to changing behavior. If we described the budget constraint correctly there would be no optimization errors. However, in many cases it would not be feasible, or would be too costly, to describe all the details of the constraint set. The common modeling technique therefore is to use a simplified description of the choice set and denote the difference between the choice predicted by the model and the actual choice as an optimization error.[4] Another reason for what we often denote an optimization error is due to the fact that the utility function estimated by the scholar is not the utility function that the individual maximizes. In the absence of adjustment costs, it could be the case that the individual is at his optimum. However, there would still be a difference between the choice predicted by our model and the actual point chosen by the individual. This difference is really a specification error, but we usually refer to it as an optimization error.

Scholars in our profession have long been aware of adjustment costs and optimization errors. This is, for example, reflected in the vocabulary short- and long-run elasticities. The idea is that in the short run adjustment costs are high, but in the long run individuals can adapt to changes in the budget constraint. At each point in time different individuals face different adjustment costs and have different optimization errors. A common way to reflect this reality has been to model these optimization errors as an additive component in a regression function, assuming a continuous distribution of the optimization error (adjustment cost) with mean zero and zero correlation with explanatory variables.[5]

[4] Sometimes there are also be measurement errors, which often are hard to distinguish from optimization errors.
[5] See e.g., Burtless and Hausman (1978), Hausman (1979) and Hausman (1985).

Here we will discuss four different reasons optimization errors can arise. The first is because of hours constraints, implying that only a limited number of points can be chosen on the budget constraint. The second concerns short-term optimization errors due to unforeseen events. The third is due to changes in individuals' preference parameters. The fourth, which is what Chetty (2012) discusses, is the case where there has been a change in the tax schedule.[6]

The possibility of constraints on hours of work has long been studied in the labor-supply literature. One of the most popular models in this literature is a discrete-choice model of labor supply (Van Soest (1995)). In this model a set of discrete alternatives or jobs represents the budget set. These models are often estimated by the Conditional Multinomial Logit model. Translated to the taxable income framework, it would imply that only a finite number of points is available on the (kinked) budget constraint. Since in the taxable income literature model we assume that individuals can choose any point on the budget constraint, if in fact only a finite number of points can be chosen, there would be a difference between the choices indicated by our model and the actual choices; there would be optimization errors.[7]

Let us move on to the second case. An individual might at the beginning of the tax year plan for a certain taxable income, then, due to unforeseen events, that taxable income might become somewhat different. Something happens and in the short run, the remainder of the tax year, the individual cannot accommodate the random event. Unforeseen bonus paychecks, better health than expected or assigned overtime are examples of positive shocks. Unexpected sicknesses, a temporary layoff, new extended vacation plans because of a new love are

[6] Chetty (2012) develops a method to set bounds on structural elasticities, when estimates have been obtained from data generated with different budget constraints. His method to set bounds is therefore not applicable to estimates obtained from the bunching estimator, which uses data from a single budget constraint.

[7] Chetty et al. (2011) also discuss the importance of restrictions on hours of work and how this leads to optimization errors.

examples that would result in a negative shock in taxable income. We could possibly represent the distribution of this type of optimization error by a symmetric distribution with mean zero.

Now we consider the third case. The utility function that we have used above, and will use in the simulations presented in the next section, is a heroic simplification. However, we can make it slightly more flexible by letting the preference parameter ρ be a function of variables like ability (productivity), health, age, number of small children, marriage status, work status of spouse, and so on. At each point in time, we have some individuals who have had a recent change in one of the variables affecting the preference parameter and therefore want to change their taxable income. If the individual's adjustment cost is low, the individual will change his taxable income and for this person there would be no optimization error. For another person the present adjustment cost might be so large that the person does not change his/her taxable income; there would be an optimization error. However, the adjustment cost might change over time. For example, if the change in taxable income only could be achieved by moving to another living place, the adjustment cost could be in the form of children going to high school who do not want to move away from friends. There would be an optimization error. Once the children finish high school, the adjustment cost is low and a change in taxable income could take place, and there would be no optimization error. The kind of optimization error just described could possibly be represented by a random variable with a symmetric pdf with mean zero. Note that even if the distribution of preferences does not change in the population, for single individuals it will, implying that the occurrence of optimization errors will not fade over time.

In the fourth case we consider a change in the tax system. There might have been a move of a kink point or a change in a marginal tax rate, which would change the slope of a linear segment of the budget constraint. To be concrete we will consider a change in the tax rate for a segment above a kink, and we assume individuals before the tax change were at their optima. Suppose there has been an increase in the marginal tax so that the slope of the linear segment

decreases. Assuming zero income effects, this means that some individuals located on the segment would like to move to the kink, and others on the segment would like to decrease their taxable income along the segment. That is, all individuals that want to change their taxable income would like to decrease it. Some might be able to do that, but in the short- to medium-run, some would be stuck at the present level of taxable income, resulting in positive optimization errors. The resulting distribution of optimization errors would have a mean greater than zero and be downward truncated at zero. Moreover, it would only be those with their optimum at or above the kink point that would encounter this type of optimization error. This type of optimization error would lead to fewer observations at and above, close to the kink, and a lower estimate of the taxable income elasticity. In the simulations we will use a normal distribution with a downward truncation at zero to represent this type of optimization error. The opposite case, with a decrease in the marginal tax and an increase in the slope of the segment above the kink, means that some individuals would like to move out from the kink point to a tangency solution on the segment, while others would like to move up the segment. In the short to medium run some might not be able to increase their actual taxable income, implying that they would have negative optimization errors with an upward truncation at zero. This type of optimization error leads to more observations at the kink and close to the kink, above the kink, and a higher estimate of the taxable income elasticity. In the simulations we will use a normal distribution with an upward truncation at zero to represent this type of optimization error. Hence, when there are optimization errors due to a change in the marginal tax rate the distributions of optimization errors are quite different depending on whether the most recent tax change had been in the form of an increase or a decrease in marginal tax rate.

5. A Simulation Exercise

We present simulations that show how, for a given taxable income elasticity, the bunching estimates will vary as we make simple variations in the preference distribution and allow for various types of optimization errors.

To generate the data we use the quasilinear utility function given by equation (4). We use a budget constraint with a kink at 1000, a marginal tax of 0.3 before the kink and 0.5 after the kink. This is a large kink which, according to the literature, should help identify the taxable income elasticity. So as to avoid the issue of sampling variation we generate income distributions with two million observations. We tried different seeds for the random number generator. Estimates differ at most in the third decimal. To obtain the bunching estimates we used the program bunchr, written by Itai Trilnick in the programming language R.[8]

In our simulations we illustrate how the bunching estimator, for a given value of the taxable income elasticity, will vary as we change the preference distribution. We can change the preference distribution in many ways; we can change the general shape, the center of the location and the variance. Here we will keep the center of location constant as well as the general shape. We will see how the bunching estimate changes as we flatten the distribution and thereby decrease the mass in the bunching window. We set the taxable income elasticity to 0.2, which gives the bunching window in terms of taxable income $(K, K + \Delta A) = (1000, 1070)$. Expressed in terms of the preference parameter the bunching window is approximately (1074,1149). We centered the preference distribution at 1100 and represent the preference distribution with a mixed normal $\phi(\rho) = \pi \cdot n(1100, 10^2) + (1 - \pi) \cdot n(1100, 140^2), \pi \in (0,1)$. As we vary π from 0.9 down to 0.1 the distribution will flatten, and the mass in the bunching window will decrease. In the table the top row shows the five different combinations of π, $(1 - \pi)$ used. The second row shows how results vary as we change the

[8] The program can be accessed via the link https:/CRAN.R-project.org/package=bunchr .

proportions and there are no optimization errors. We see that the estimates vary from around 0.6 down to 0.19, depending on how large the part of the preference distribution that is in the bunching window is. The simulations illustrate that, even in the absence of optimization errors, the bunching estimator cannot identify the taxable income elasticity.

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 simulations. 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 optimization errors do not affect the estimates much.

TABLE: Simulations with mixed normals

π ; $(1-\pi)$	0.9 ; 0.1	0.7 ; 0.3	0.5 ; 0.5	0.3 ; 0.7	0.1 ; 0.9
$\hat{\beta}$	0.598	0.500	0.402	0.302	0.192
$\hat{\beta}_{opterror1}$	0.075	0.081	0.080	0.074	0.058
$\hat{\beta}_{opterror2}$	0.012	0.013	0.011	0.010	0.008
$\hat{\beta}_{opterror3}$	0.0	0.013	0.035	0.077	0.065
$\hat{\beta}_{opterror4}$	0.530	0.462	0.394	0.306	0.238

$\hat{\beta}$ no optimization errors; $\hat{\beta}_{opterror1}$ symmetric optimization errors, mean zero, std 25; $\hat{\beta}_{opterror2}$ symmetric optimization errors, mean zero, std 50; $\hat{\beta}_{opterror3}$ negative asymmetric optimization errors; $\hat{\beta}_{opterror4}$ positive asymmetric optimization errors.

To summarize the results of the simulations shown in the table: all data have been generated with a utility function which implies a taxable income elasticity of 0.2. The simulations illustrate that, even in the absence of optimization errors the bunching estimator cannot identify the taxable income elasticity. Adding optimization errors in general makes the bunching estimates much smaller. Depending on the distribution of preferences and optimization errors the estimates vary between 0.0 and around 0.6. The estimates are all over the place.

6. Bounds from Monotonicity

In Section 3 we showed that if the heterogeneity density is unrestricted, except for smoothness conditions, then a kink, and even the entire distribution of taxable income from a single budget set, provides no information about the taxable income elasticity. If the heterogeneity density is restricted in some way then it is possible to learn some things about the taxable income elasticity. In this Section we consider what can be learned when the heterogeneity density is monotonic over the bunching interval. Monotonicity might be a reasonable assumption in ranges where the taxable income density seems to be increasing away from the kink.

To show how monotonicity of the heterogeneity density $\phi(\rho)$ over the bunching interval can bound the elasticity, let a_ℓ and a_u denote lower and upper endpoints for a taxable income interval that includes the kink, where excess bunching may occur. Let $\rho_\ell = a_\ell \theta_1^{-\beta}$ and $\rho_u = a_u \theta_2^{-\beta}$ denote corresponding lower and upper endpoints for ρ and

$$f^-(a_\ell) = \lim_{a \to a_\ell, a < a_\ell} f(a), \quad f^+(a_u) = \lim_{a \to a_u, a > a_u} f(a).$$

Consider the two functions

$$D^-(\beta) = f^-(a_\ell)\left[a_u \left(\frac{\theta_1}{\theta_2}\right)^\beta - a_\ell\right], D^+(\beta) = f^+(a_u)\left[a_u - a_\ell \left(\frac{\theta_2}{\theta_1}\right)^\beta\right].$$

We have the following result:

THEOREM 2: *If $\phi(\rho)$ is monotonic on $[\rho_\ell, \rho_u]$ then the taxable income elasticity satisfies*

$$\min\{D^-(\beta), D^+(\beta)\} \leq \Pr(\, a_u \leq A \leq a_\ell) \leq max\{D^-(\beta), D^+(\beta)\}. \qquad (7)$$

If $\Pr(\, a_u \leq A \leq a_\ell) < \min\{D^-(0), D^+(0)\}$ *then there is no β satisfying equation (7). Otherwise the set of all nonnegative β satisfying equation (7) is a subset of* $[0, \infty)$.

For estimation we just plug in nonparametric estimators $\hat{f}^{-}(a_\ell)$ and $\hat{f}^{+}(a_u)$ to obtain

$$\widehat{D}^{-}(\beta) = \hat{f}^{-}(a_\ell)\left[a_u(\theta_1/\theta_2)^{\beta} - a_\ell\right],$$

$$\widehat{D}^{+}(\beta) = \hat{f}^{+}(a_u)\left[a_u - a_\ell(\theta_2/\theta_1)^{\beta}\right].$$

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

$$\max\{\widehat{D}^{-}(\hat{\beta}_\ell), \widehat{D}^{+}(\hat{\beta}_\ell)\} = \hat{P}\,, \min\{\widehat{D}^{-}(\hat{\beta}_u), \widehat{D}^{+}(\hat{\beta}_u)\} = \hat{P}.$$

Standard errors for these bounds are given in the Appendix.

As an example we apply these bounds to the kink at zero taxable income for married tax filers shown in Panel A of Figure 7 of Saez (2010). We take the lower endpoint of the excess bunching interval to be $a_\ell = -5000$ and the upper endpoint to be $a_u = 5000$. We approximate the graph by a function that is linear between each of the following pairs of points:

$$(-5000, .21),\ (-2500, .35),\ (0, .44),\ (2000, .35),\ (5{,}000,\ .35).$$

We take the taxable income density over (-5000,5000) to be the piecewise-linear function connecting these points, up to scale. We also take $\hat{f}^{-}(a_\ell) = .21$ and $\hat{f}^{+}(a_u) = .35$. The estimated taxable income elasticity bounds under monotonicity as described above are

$$\hat{\beta}_\ell = 0,\ \hat{\beta}_u = 3.85.$$

These bounds are very wide. Thus, for the married filers in Panel A of Figure 7 of Saez (2010), bounds based on monotonicity of the heterogeneity density do not provide much information.

Bounds based on monotonicity of the heterogeneity pdf can be quite wide when $f^{-}(a_\ell)$ and $f^{+}(a_u)$ are far apart, as in the example from the previous paragraph. One could construct tighter bounds by putting more restrictions on the heterogeneity pdf, such as concavity. However, all such bounds are based entirely on supposition. As we have discussed the data

provides no information on the heterogeneity density for individuals at the kink. Thus, information about the density at the kink must come from a source other than the data. One could extrapolate properties of the kink density from the observed distribution of taxable income, but such extrapolation necessarily is based on information other than the data. As we continue to emphasize, the heterogeneity density for individuals at the kink is not identified.

7. Identification

Given the identification difficulties for bunching it seems important to consider what will identify the taxable elasticity. We know from Section 3 that some variation in the budget set is required, even in the case of scalar separable heterogeneity and a parametric utility function. In this section we consider how much budget set variation suffices for identification.

The elasticity cannot be identified only by variation in the kinks, even from multiple budget sets. Intuitively, the order condition is still not satisfied if only information about kinks is used. Note that each kink probability is just one number. Each kink probability will depend on the pdf of heterogeneity over an interval. Except in rare cases, each interval will have some part that is not shared by all other kinks. Varying the pdf over that interval will allow the kink probability to be anything for any elasticity. Thus, kinks from multiple budget sets are generally no more informative than a single kink.

To identify the elasticity β in equation (4) it can suffice to have just two budget sets. An order condition again provides insight. If there are two budget sets the data identifies two functions, the CDF of taxable income along each of the two budget sets. In the utility specification of equation (4) there is one unknown function, the CDF of ρ, and one unknown parameter, the taxable income elasticity β. Two functions can be more than enough to identify

one function and one parameter. In fact the taxable income elasticity can be overidentified and strong restrictions imposed on the distribution of taxable income across the two budget sets.

We give an identification result for the isoelastic specification when there are two convex budget sets. We again utilize the results of Blomquist et al. (2015) that characterize choice with convex budget sets in terms of choice for linear budget sets. Let $\theta(a)$ and $\tilde{\theta}(a)$ denote the slope from the right of the budget frontier for each of the budget sets, i.e. the net-of-tax rate for a small increase in taxable income (which exists by concavity of the budget frontier; see Rockafellar, 1970, pp. 214-215). Let $F(a)$ and $\tilde{F}(a)$ be the corresponding distributions of taxable income for the two budget sets. Since the choice for a linear budget set is $\rho\theta^{\beta}$ it follows by Theorem 2 of Blomquist et al. (2015) that

$$F(a) = \Pr(A \leq a) = \Pr\left(\rho\theta(a)^{\beta} \leq a\right) = \Phi\left(a\theta(a)^{-\beta}\right), \qquad \tilde{F}(a) = \Phi\left(a\tilde{\theta}(a)^{-\beta}\right).$$

Here we see that the two distributions are the same except for a scalar multiple of the taxable income a. Changing the tax rate simply scales up or down the taxable income for a linear budget set with the amount of the scale adjustment determined by β. We can use this feature to obtain β from the size of the scale adjustment when the tax rate changes.

THEOREM 3: *If $\Phi(\rho)$ is continuous and strictly monotonically increasing and there exists a and $\tilde{a}$ such that $F(a) = \tilde{F}(\tilde{a})$ and $\theta(a) \neq \tilde{\theta}(\tilde{a})$ then*

$$\beta = \frac{\ln\left(\frac{\tilde{a}}{a}\right)}{\ln\left(\frac{\tilde{\theta}(\tilde{a})}{\theta(a)}\right)}.$$

Here we see that β is identified from any pair of taxable incomes a and $\tilde{a}$ with the same value of the distribution for the first and second budget sets but a different marginal tax rate.

One example is provided by two piecewise budget sets with one kink that is the same, common θ_1 up to the kink K, and different slopes $\theta_2 > \tilde{\theta}_2$ respectively, beyond the kink. Note that $F(K) < \tilde{F}(K) < 1$. Let $\tilde{a} \geq K$. By $\Phi(\rho)$ strictly monotonic and continuous there is $a > \tilde{a}$ such that $F(a) = \tilde{F}(\tilde{a})$, so by Theorem 3 we have $\beta = \ln\left(\frac{\tilde{a}}{a}\right) / \ln\left(\frac{\tilde{\theta}_2}{\theta_2}\right)$. Indeed, this equation is satisfied for any $\tilde{a} \geq K$ so that there is a continuum of identifying equations for β. In this example the elasticity is highly overidentified.

In another example the two budget sets have the same tax rates but the kink is different, say $K < \widetilde{K}$. Note here that the slopes of the budget set differ only in the interval $[K, \widetilde{K})$, so that there must exist points a and $\tilde{a}$ in this interval to apply Theorem 3. Specifically, to apply Theorem 3 there must exist $\tilde{a} < \widetilde{K}$ such that $F(K) = \tilde{F}(\tilde{a})$. In this case $\beta = \ln\left(\frac{\tilde{a}}{a}\right) / \ln\left(\frac{\theta_1}{\theta_2}\right)$. Intuitively, if we think of $\widetilde{K}$ as a right shift of K, then Theorem 3 gives identification of the elasticity when the shift is beyond the end of the extended first segment. Equivalently, identification holds when some individual who was on the linear segment beyond the original kink K experiences a tax change. If $\tilde{F}(\tilde{a})$ remains bounded above by something slightly smaller than $F(K)$ as $\tilde{a}$ approaches $\widetilde{K}$ from the left then choices for some individuals located at the first kink will not be observed away from a kink.

It would be useful to have identification conditions for specifications more general than the isoelastic case. One more general specification has taxable income for a linear budget set given by $A = \rho h(\theta)$ where $h(\theta)$ is an unknown, strictly monotonic increasing function. Here variation in budget sets can identify $h(\theta)$ up to scale at values of θ corresponding to different budget sets.

THEOREM 4: *If* $\Phi(\rho)$ *is continuous and strictly monotonic increasing and there exists* a *and* $\tilde{a}$ *such that* $F(a) = \tilde{F}(\tilde{a})$ *and* $\theta(a) \neq \tilde{\theta}(\tilde{a})$ *then* $\frac{h(\tilde{\theta}(\tilde{a}))}{h(\theta(a))} = \frac{a}{\tilde{a}}$.

This result allows us to identify taxable income effects that scale the taxable income according to the net of tax rate θ in the model $A = \rho h(\theta)$. These effects would be for discrete tax changes corresponding to rates along the two budget sets. Obtaining continuous tax effects for small tax changes would require continous variation in piecewise linear budget sets.

Although the specification $A = \rho h(\theta)$ is nonparametric it still has the strong "scaling" property, where changes in the tax rate shift the scale of taxable income. To obtain models that are not restricted in this way one needs to allow heterogeneity to enter in a more general way than multiplicatively. For instance, one could let both ρ and β vary over individuals, giving a linear random coefficients specification $\ln(A) = \ln(\rho) + \beta \ln(\theta)$, where both $\ln(\rho)$ and β are random. If the budget sets were linear then a least squares regression of $\ln(A)$ on a constant and $\ln(\theta)$ would identify the expected elasticity, as is well known. If the budget sets are nonlinear then identification of the expected elasticity would be more difficult. The most general case, with heterogeneity that could affect the taxable income in any way, is considered in Blomquist et al. (2015).

7. Summary

In this paper we first described the bunching estimation procedure of Saez (2010). We then showed nonidentification of the taxable income elasticity when the distribution of heterogeneity is unrestricted. For this purpose we used both a diagrammatic, non-parametric example and theoretical analysis showing a kink is not informative about the taxable income elasticity, even if one is willing to assume a parametric utility function of the type used in Saez (2010). The failure of identification of the taxable income elasticity from one budget set is complete in the

sense that it has no information about the elasticity when the distribution of heterogeneity is unrestricted. The fundamental reason for this lack of identification is that movements along one budget set correspond to variations across individuals so that one cannot separate heterogeneity effects from price effects. We also showed that optimization errors hinder identification of the taxable income elasticity.

If one is willing to put restrictions on the heterogeneity distribution then bunching can be informative about the taxable income elasticity. We showed that the Saez (2010) estimator corresponds to assuming that the heterogeneity density is linear over the kink. Linearity of the heterogeneity density seems a strong restriction on which to hang identification of the taxable income elasticity. Bounds on the taxable income elasticity can be obtained if the heterogeneity density is restricted or known not to be too variable. For example, we show how to derive bounds if one is willing to assume that the heterogeneity distribution is monotonic. However, applying these results to the data in Saez (2010) gives very wide bounds.

We performed a small simulation exercise where, for a given taxable income elasticity, we varied the heterogeneity distribution. We also studied the effect of adding various types of optimization errors. The simulations verify that the bunching estimator cannot identify the taxable income elasticity even in the absence of optimization errors. Adding optimization errors in general gives estimates an order smaller in magnitude.

The negative results on the possibility of identifying the taxable income from bunching around a single kink raise the question, how can the taxable income elasticity be identified? We show that using bunching from several kinks does not help. However, variation in budget sets can be used to identify the taxable income elasticity. We show that, given that one is willing to put functional form restrictions on the utility function, such as Saez (2010)'s quasilinear form, the distribution of taxable income along linear segments provides strong identifying information without optimization errors. With optimization errors and preference heterogeneity

there are estimation methods using variation in budget constraints that can be used to identify slope effects. It is true that most studies using such methods have used parametric assumptions for the errors as well as for the utility function. However, Blomquist and Newey (2002) developed a non-parametric model that fully allows for measurement errors, optimization errors and preference heterogeneity. This model was shown to work with general heterogeneity and was extended to the taxable income setting in Blomquist et al. (2015). Variation in budget sets in practice often means that one uses data from several points in time, implying that one must have some way to account for exogenous productivity growth.[9]

Appendix: Proofs of Theorems

Proof of Theorem 1: Given preceding the statement of Theorem 1. *Q.E.D.*

Proof of Theorem 2: Monotonicity implies that for $\rho \in (\rho_\ell, \rho_u)$,

$$\min\{\phi(\rho_\ell), \phi(\rho_u)\} \leq \phi(\rho) \leq \max\{\phi(\rho_\ell), \phi\,(\rho_u)\}.$$

Also, $\phi(\rho_\ell)$ and $\phi(\rho_u)$ are given by $\phi(\rho_\ell) = f^-(a_\ell)\theta_1^\beta,\ \phi(\rho_u) = f^+(a_u)\theta_2^\beta$. The first conclusion of Theorem 2 then follows by

$$P = \Pr(a_\ell \leq A \leq a_u) = \int_{\rho_\ell}^{\rho_u} \phi(\rho)d\rho \ \leq (\rho_u - \rho_\ell)\max\{\phi(\rho_\ell), \phi(\rho_u)\}$$

$$= \left[a_u\theta_2^{-\beta} - a_\ell\theta_1^{-\beta}\right]\max\left\{f^-(a_\ell)\theta_1^\beta, f^+(a_u)\theta_2^\beta\right\} = \max\{D^-(\beta), D^+(\beta)\},$$

$$P \geq \min\{D^-(\beta), D^+\,(\beta)\}.$$

[9] The issue of exogenous productivity growth is a problem for studies of taxable income, not studies of hours of work.

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 β = 0,

$$D^{-}(0) = f^{-}(a_\ell)(a_u - a_\ell), D^{+}(0) = f^{+}(a_u)(a_u - a_\ell).$$

As long as

$$P \geq \max\{f^{-}(a_\ell), f^{+}(a_u)\}(a_u - a_\ell),$$

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

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

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

$$\min\{f^{-}(a_\ell), f^{+}(a_u)\}\,(a_u - a_\ell) < P < \max\{f^{-}(a_\ell), f^{+}(a_u)\}(a_u - a_\ell)$$

then we can take $\beta_\ell = 0$. *Q.E.D.*

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

Proof of Theorem 4: Note that $F(a) = \tilde{F}(\tilde{a})$ implies $\Phi(ah(\theta(a))) = \Phi\left(\tilde{a}h(\tilde{\theta}(a))\right)$, which implies $ah(\theta(a)) = \tilde{a}h(\tilde{\theta}(a))$ by $\Phi(\rho)$ strictly monotonic. Solving gives the result. Q.*E.D.*

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.

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

Blomquist, S., A. Kumar, C-Y, Liang and W.K. Newey (2015), "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income," CESifo Working Paper No. 5320.

Burtless, G. and J. Hausman (1978), "The Effect of Taxes on Labor Supply," *Journal of Political Economy* 86, 1103-l130.

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

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.

Eklöf, M. and H. Sacklen, (2000), "The Hausman-MaCurdy Controversy: Why Do the Results Differ across Studies?" *Journal of Human Resources* 35, 204-220.

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," *Econonomic Letters* 3, 171-174.

Hausman, J.A., (1985), "The Econometrics of Nonlinear Budget Sets," *Econometrica* 53, 1255-1282.

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

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

MaCurdy, T., D. Green and H. Paarsch, (1990), "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply," *Journal of Human Resources* 25, 414-490.

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

Mirrlees, J. A. **(**1971), "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies* 38, 175-208.

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

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.