2017-09-15

The Bunching Estimator Cannot Identify the Taxable Income Elasticity

Sören Blomquist* and Whitney K. Newey**

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.

***** Uppsala Center for Fiscal Studies, Department of Economics, Uppsala University.

** Department of Economics, M.I.T.

1. Introduction

The taxable income elasticity is a key parameter when predicting the effect of tax reform or designing an income tax. A large literature has developed over several decades which attempts to estimate this elasticity. However, due to a large variation in results between different empirical studies there is still some controversy over the size of the elasticity. 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 600 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.

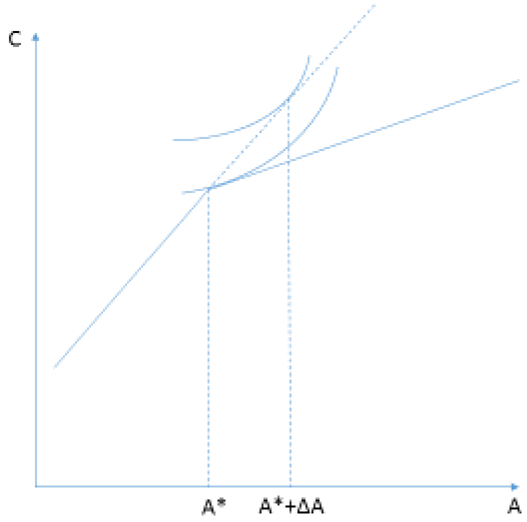
A kink may be informative about the elasticity when the heterogeneity density is restricted. We derive bounds on the elasticity when the heterogeneity density is monotonic over the kink. An assumption of monotonicity may seem reasonable when the kink occurs to one side of a unimodal distribution of taxable 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.

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.

Figure 1



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

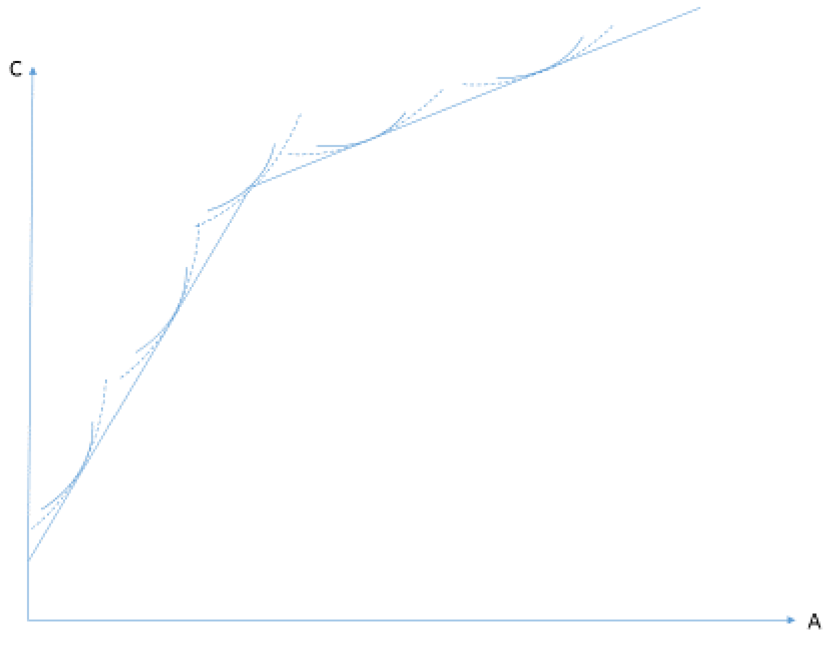


Figure 2 shows that we can have two identical income distributions where one income distribution comes from preferences with a large taxable income elasticity and the other from preferences with a low taxable income elasticity. We also assume that the indifference curves of individuals at the kink point 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 also discuss various reasons for optimization errors and possible shapes for them. In Section 5we perform a simulation exercise where, for a given taxable income elasticitywe 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 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

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

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 also show that the bunching estimator of Saez (2010) is based on a linear density assumption. We continue to proceed under the assumption that there are no optimization errors. In the next section we will consider optimization errors.

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

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

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 ρ and θ for $\rho > 0,\ \theta > 0$ and decreasing in β for $0 < \theta < 1$.

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

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

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 $\Phi(\rho), \rho \in (\underline{\rho}, \overline{\rho})$ 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 $\rho \in (\underline{\rho}, \overline{\rho}^1)$ there will be a tangency solution on the first segment, a kink solution at $A = K$ for $\rho \in (\overline{\rho}_1, \underline{\rho}_2)$ and a tangency on the second segment for $\rho \in (\underline{\rho}_2, \overline{\rho})$. 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 $f^1(A) = \phi\left(A/\theta_1^{\beta_\rho}\right)\dfrac{1}{\theta_1^{\beta_\rho}}$ for $A \in (\underline{A}, K)$ 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 $F^2(A) = \Pr(\rho\theta_2^{\beta_\rho} \le A)$ $= \Pr\left(\rho \le A/\theta_2^{\beta_\rho}\right) = \Phi\left(A/\theta_2^{\beta_\rho}\right)$, and the pdf is $f^2(A) = \phi\left(A/\theta_2^{\beta_\rho}\right)\dfrac{1}{\theta_2^{\beta_\rho}}$. The probability that taxable income is at the kink is given by

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

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

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 $F^1(A) \equiv G^1(A)$, $A \in (\underline{A}, K)$, $F^2(A) \equiv G^2(A)$, $A \in (K, \overline{A})$, 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 $\overline{A} = \overline{\rho}\theta_2^{e_\rho} = \overline{\eta}\theta_2^{e_\eta}$, implying $\overline{\eta} = \overline{\rho}\theta_2^{e_\rho - e_\eta}$. The requirement $F^1(A) \equiv G^1(A)$, $A \in (\underline{A}, K)$ implies $\psi\left(A/\theta_1^{e_\eta}\right) = \phi\left(A/\theta_1^{e_\rho}\right)\theta_1^{e_\eta - e_\rho}$, $A \in (\underline{A}, K)$ 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

[4] Sometimes there are also be measurement errors, which often are hard to distinguish from optimization errors.

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]

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

[5] See e.g., Burtless and Hausman (1978), Hausman (1979) and Hausman (1985).

[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.

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 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

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

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

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

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 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*

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

If $\Pr(\ a_u \le A \le a_\ell) < \min\{D^-(0), D^+(0)\}$ *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 marginal 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

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

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 $\rho \in (\rho_\ell, \rho_u)$,

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

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

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

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

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

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,

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

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

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.

Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income*

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

Anil Kumar
Federal Reserve Bank of Dallas

Che-Yuan Liang
Uppsala Center for Fiscal studies, Department of Economics, Uppsala University

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

First Draft: January 2010
This Draft: May 2015

Abstract

Many studies have estimated the effect of taxes on taxable income. To account for nonlinear taxes these studies either use instrumental variables approaches that are not fully consistent or impose strong functional form assumptions. None allow for general heterogeneity in preferences. In this paper we derive the expected value and distribution of taxable income conditional on a nonlinear budget set, allowing general heterogeneity and optimization error in taxable income. We find an important dimension reduction and use that to develop nonparametric estimation methods. We show how to nonparametrically estimate the expected value of taxable income imposing all the restrictions of utility maximization and allowing for measurement errors. We characterize what can be learned nonparametrically from kinks about compensated tax effects. We apply our results to Swedish data and estimate for prime age males a significant net of tax elasticity of 0.21 and a significant nonlabor income effect of about -1. The income effect is substantially larger in magnitude than it is found to be in other taxable income studies.

JEL Classification: C14, C24, H31, H34, J22
Keywords: Nonlinear budget sets, nonparametric estimation, heterogeneous preferences, taxable income, revealed stochastic preference.

*The NSF provided partial financial support. We are grateful for comments by R. Blundell, G. Chamberlain, P. Diamond, J. Hausman, C. Manski, R. Matzkin, J. Poterba, H. Selin and participants at seminars at UCL (Jan 2010), BC, Harvard/MIT, and NYU.

1 Introduction

Behavioral responses to tax changes are of great policy interest. In the past much of this interest was focused on hours of work, and the central question was how labor supply responds to tax reform. In a set of influential papers, Feldstein (1995, 1999) emphasized that traditional measures of deadweight loss based just on labor supply are biased downward as they ignore many other important behavioral responses like work effort, job location, tax avoidance and evasion. Inspired by Feldstein's work, which showed that the taxable income elasticity is sufficient for estimating the marginal deadweight loss from taxes, a large number of studies have produced a wide range of estimates.[1]

Although the conventional estimates of the taxable income elasticity provide information on how taxable income reacts to a marginal change in a *linear* budget constraint, they are less useful for estimating the effect of tax reforms on taxable income. In a real world of nonlinear tax systems with kinks in individuals' budget constraints, tax reforms often result in changes in kink points as well as in marginal tax rates for various brackets. There has been extensive research on estimating the effect of such complicated changes in the tax systems on labor supply using parametric structural models with piecewise linear budget sets, often estimated by maximum likelihood methods. More recent labor supply studies have estimated a utility function which can be used to predict the effect of taxes in the presence of piecewise linear budget sets. These studies focusing on labor supply, however, not only ignore other margins of behavioral responses to taxation but also rely on strong distributional and functional form assumptions when they use parametric models.

We nonparametrically identify and estimate the expected value of taxable income conditional on nonlinear budget frontiers while allowing for general heterogeneity. The heterogenous preferences are assumed to be strictly convex and statistically independent of the budget frontier, but are otherwise unrestricted. We also allow for optimization errors in estimation of the expected value. Identification is straightforward. The object of interest is the conditional expectation of an observable variable (taxable labor income) conditional on another observable variable (the budget frontier). As usual the conditional expectation is identified over all values of the budget frontier that are observed in the data.

Estimation is challenging because the conditioning variable (the budget frontier) is infinite dimensional. We approach this problem by using the restrictions of utility maximization to simplify the conditional expectation. We find that for a piecewise linear, progressive tax

[1] The estimates range from -1.3 (Goolsbee, 1999) to 3 (Feldstein, 1995) with more recent studies closer to 0.5 (Saez, 2003; Gruber and Saez, 2002; Kopczuk, 2005, Giertz, 2007). Blomquist and Selin, 2010 find an elasticity of about .2; See Saez, Slemrod and Giertz (2012) for a comprehensive review of the literature.

schedule the expected value depends on low dimensional objects. One form of these objects is exactly analogous to the expected value of labor supply from Blomquist and Newey (2002, BN henceforth), which was derived when preference heterogeneity can be represented by a scalar. Consequently, it turns out that the labor supply results of BN are valid under general preference heterogeneity. However the BN form did not impose all the restrictions imposed by utility maximization and we show how to do this. Also we show how to allow for nonconvexities in the budget set.

We derive the distribution of taxable income at points where the budget frontier is concave over an open interval. We analyze kinks, showing how kinks of any size depend nonparametrically on the density of taxable income as well as on average compensated effects for individuals at the kink. We also find that the conditional distribution of taxable income given the budget set is the same as for a linear budget set with the net of tax rate equal to the slope from the right of the budget set and income equal to virtual income. This finding dramatically reduces the dimension of the nonparametric estimation problem, because it shows that the conditional distribution and expected value depend only on a two or three dimensional object rather than depending on the whole budget set. Also, we find that varying convex budget sets provides the same information about preferences as varying linear budget sets.

The model here is like the revealed stochastic preference model of McFadden (2005) for continuous outcomes in having general heterogenous preferences and budget sets that are statistically independent of preferences. We differ in considering only the two good case with strictly convex preferences. We find that for two goods, smooth strictly convex preferences, and convex budget sets, necessary and sufficient conditions for utility maximization are that the CDF given a linear budget set satisfies a Slutzky property.

In independent work Manski (2014) considered the identification of general preferences when labor supply is restricted to a finite choice set, with the goal of evaluating tax policy. He concluded that nonidentification of preferences makes tax policy evaluation difficult. This paper and BN reach a different conclusion than Manski (2014) because we consider different policies than he did. We find that tax policy evaluation under general preferences is feasible and useful when it is based on comparing expected values across observed budget sets. Imposing the restrictions implied by utility maximization makes such policy evaluation feasible. Importantly, we and BN also differ from Manski (2014) by allowing for a source of variation in the outcome other than preference variation (e.g. measurement error), which has long been thought to be important for labor supply and taxable income; see Burtless and Hausman (1978). The way we allow for such variation is not based on utility maximization unlike Chetty (2012), but it does allow for low probabilities for kinks, as is often found in data.

Keane and Moffitt (1998), Blundell and Shephard (2012), and Manski (2014) have considered labor supply when hours are restricted to a finite set. Our taxable income setup could accommodate such constraints, though we do not do this for simplicity. As long as the grid of possible labor supply values was rich enough the expected value of taxable income that we derive would be approximately correct. It appears to be harder to incorporate the bilateral contracting framework of Blundell and Shephard (2012).

To evaluate the effect of taxes on taxable income we focus on elasticities that apply to changes in nonlinear tax systems. Real-world tax systems are non-linear, and it is variations in non-linear tax systems that we observe. Therefore, it is easiest to nonparametrically identify elasticities relevant for changes in nonlinear tax systems. BN did show that with labor supply it may be possible to identify labor supply elasticities for changes in linear budget sets, but for taxable income we find that the conditions for identifying average elasticities for a linear budget set are very stringent and not likely to be satisfied in applications. Here we propose effects defined by an upward shift of the non-linear budget constraint, in either slope or intercept. These effects are relevant for changes in non-linear budget constraints. We find that these can be estimated with a high degree of accuracy in our application.

In the taxable income setting it is important to allow for productivity growth. To nonparametrically separate out the effect of exogenous productivity growth from changes in taxable income that are due to changes in individual behavior is one of the hardest problems in the taxable income literature. We provide a way to do this and show how it matters for the results.

Our application is to Swedish data from 1993-2008 with third party reported taxable *labor* income. This means that the variation in the taxable income that we observe for Sweden is mainly driven by variations in effort broadly defined and by variations in hours of work and not by variations in tax evasion.[2] We estimate a statistically significant tax elasticity of 0.21 and a significant income effect of -1. This income effect is significantly larger than in many taxable income studies, many of which find a small effect or assume no effect.

The rest of our paper is organized as follows. Section 2 reviews the taxable income literature. Section 3 lays out a model of individual behavior where there are more decision margins than hours of work. Section 4 derives the distribution of taxable income for nonlinear budget sets. Section 5 analyzes kinks. Section 6 derives the expected value of taxable income and shows how to approximate it in a way that imposes utility maximization. Section 7 describes the policy effects we consider and how we allow for productivity growth. Section 8 explains how these results can be empirically implemented. In Section 9 we describe the Swedish data we use and

[2] Kleven et al. (2011) find that the tax evasion rate is close to zero for income subject to third-party reporting.

present our estimates. Section 10 concludes.

2 Previous literature

Lindsey (1987) used 1981 ERTA as a natural experiment to estimate a taxable income elasticity of about 1.6 using repeated cross sections from 1980-1984. In his influential paper that brought the taxable income elasticity to the center stage of research on behavioral effects of taxation, Feldstein (1995) used a panel of NBER tax returns and variation from TRA 1986 to estimate elasticity greater than 1 and even higher for high-income individuals for a sample of married individuals with income over $30,000. Navratil (1995) also used the 1980–1983 waves of NBER tax panel and using variation from 1981 ERTA on a sample of married people with income more than $25,000 he estimated an elasticity of 0.8. Feldstein and Feenberg (1995) used OBRA 1993 as a source of identifying variation and used IRS data from 1992 and 1993 and estimated an elasticity of 1.

Other papers have found much lower taxable income elasticities. Auten and Caroll (1999) used treasury tax panel from 1985 and 1989, i.e., before and after TRA 1986 to find an elasticity of 0.5. They restricted their sample to individuals earning more than $15,000. Sammartino and Weiner (1997) also used treasury tax panel from 1991 and 1994 and variation from OBRA 1993 to estimate zero taxable income elasticity. Goolsbee (1999) used a panel of high-income corporate executives with earnings higher than $150000 before and after OBRA 1993. His estimate of the elasticity was close to 0.3 in the long run but close to 1 in the short run. Carroll (1998) also used the treasury tax panel from 1985 to 1989 and found an elasticity of 0.5. Goolsbee (1999) used a long data set from 1922-1989 and used multiple tax reforms as a source of identification to find a taxable elasticity ranging from -1.3 to 2 depending on the tax reform.

Moffitt and Wilhelm (2000) used the SCF waves of 1983 and 1989 and exploited TRA 1986 to estimate a much larger elasticity of 2. Gruber and Saez (2002) used alternative definitions of taxable income and used variation from ERTA 1981 and TRA 1986 using the Continuous Work History Files from 1979-1990. Their elasticity estimates were in the range of 0.12-0.4. However, for high-income individuals the elasticity was 0.57 compared with 0.18 for the lower-income individuals. Sillamaa and Veall (2000) used Canadian data from 1986-1989 and identified the taxable income elasticity using the Tax Reform Act of 1988. They found taxable income elasticity ranging from 0.14-1.30.

Saez (2003) used the University of Michigan tax panel from 1979-1981 and made use of the "bracket-creep" due to high inflation to compare income changes of those at the top of the bracket who experienced a change in their marginal tax rate as they crept into an upper bracket to those at the bottom of the tax bracket whose marginal tax rates remained relatively unchanged. Since the two groups are very close in their incomes, these estimates are robust to biases due to increasing income inequality. He estimated an elasticity of 0.4 using taxable income as the definition of income. However, the estimated elasticity was zero once the definition was changed to wage income.

More recent studies have also estimated low taxable income elasticities. Kopczuk (2005) used the University of Michigan tax panel to yield an estimate of -0.2-0.57. More recently Eissa and Giertz (2006) used the Treasury tax panel from 1992-2003 and data from executive compensation. They used variation from multiple tax reforms during this period –TRA 1986, OBRA and EGTRRA on a sample of executives and the top 1 percent of the tax panel. Their elasticity estimates were small for the long run (0.19), but 0.82 for the short run. Using data from SIPP and the NBER tax panel, Looney and Singhal (2006) also estimate a somewhat larger elasticity of 0.75. More recently Giertz (2007) used Continuous Work History Survey data from 1979 to 2001 and using methods similar to those of Gruber and Saez (2002) estimated taxable income elasticity of 0.40 for the 1980s and 0.26 for the 1990s. Using a broader definition of income, the elasticities were 0.21 for the 80s and 0.13 for the 90s. Blomquist and Selin (2010) used the Swedish Level of Living Survey combined with register data to estimate an elasticity for taxable income of 0.19-0.21 for men and 0.96-1.44 for females. Using the University of Michigan Tax Panel from 1979-1990 and instrumental variable methods, Weber (2014) found a taxable income elasticity between 0.86 and 1.36 in different specifications.

3 The Model

Feldstein (1995) argued that individuals have more margins than hours of work to respond to changes in the tax. For example, individuals could exert more effort on the present job, switch to a better paid job that requires more effort, or could move geographically to a better-paid job. The choice of compensation mix (cash versus fringe benefits) and tax avoidance/evasion are still other margins. Our data is such that we do not need to worry about tax evasion but allowing for an effort margin seems useful and is important for accounting correctly for productivity growth over time.

To describe the model let c denote consumption, e effort, and h hours of work. Also let R denote nonlabor income and for a linear tax let τ denote the tax rate and $\rho = 1 - \tau$ the net of tax rate for income. We let the wage be $w(e)$ for effort level e. Let $u(c, e, h)$ denote an individual's utility function, assumed to be strictly quasi-concave, increasing in c and decreasing in e and

h. The individual choice problem is

$$\operatorname*{Max}_{c,e,h} u(c,e,h) \quad s.t. \quad c = w(e)h\rho + R, c \geq 0, e \geq 0, h \geq 0. \tag{3.1}$$

This problem can be reformulated as a choice of consumption and taxable income $y = w(e)h$. Since $w(e) = y/h$, if the wage function $w(e)$ is one-to-one then inverting gives $e = w^{-1}(y/h)$. Noting that only y enters the constraint we can concentrate h out of the choice problem by choosing h to maximize $u(c, w^{-1}(y/h), h)$ and then maximizing over c and y. Letting $U(c,y) = \max_h u(c, w^{-1}(y/h), h)$ be a concentrated utility function, the choice of c and y is obtained by solving

$$\operatorname*{Max}_{c,y} U(c,y) \quad s.t. \quad c = y\rho + R, c \geq 0, y \geq 0. \tag{3.2}$$

The solution gives taxable income $y(\rho, R)$ as a function of the net of tax rate ρ and nonlabor income R.

In the taxable income literature one usually starts with individual choice of consumption c and taxable income y as given by equation (3.2). We will also adopt this approach for much of the paper. We do return to the original effort specification when we incorporate productivity changes. We do this because productivity affects wages as a function of effort. Also, through much of the paper we will assume that $U(c,y)$ is strictly quasi-concave. This condition is not equivalent to $u(c,e,h)$ being strictly quasi-concave but instead corresponds to an additional restriction on $u(c,e,h)$. Nevertheless we will assume strict quasi-concavity of $U(c,y)$ throughout consistent with our focus on taxable income.

We allow for general heterogeneity that affects both preferences and wages. Let η denote a vector valued random variable of any dimension that represents an individual. We specify the utility function of an individual as $u(c,e,h,\eta)$ and the wage rate as $w = g(e,\eta)$. We impose no restriction on how η enters the utility or wage function, thus allowing for distinct heterogeneity in both preferences and the wage function (e.g. ability), with different components of η entering u and $w(e)$. The individual's optimization problem for a linear budget set is now to maximize $u(c,e,h,\eta)$ subject to $c = w(e,\eta)h\rho + R$. As before we concentrate out hours using $U(c,y,\eta) = \max_h u(c, w^{-1}(y/h,\eta), h, \eta)$. The choice of taxable labor income $y(\rho,R,\eta)$ for an individual η for a linear budget set is then given by

$$y(\rho, R, \eta) = \arg\max_{c,y} U(c,y,\eta) \quad s.t. \quad c = y\rho + R, c \geq 0, y \geq 0.$$

This is the same choice problem as before except that the concentrated utility function U now depends on η, and hence so does the taxable income function $y(\rho, R, \eta)$.

This specification allows for preferences to vary across individuals in essentially any way at all. For example income and level effects can vary separately, as in Burtless and Hausman

(1978). We do need to restrict η and $U(c,y,\eta)$ so that probability statements can be made but these are technical side conditions that do not affect our interpretation of η as representing general heterogeneity and are reserved for the Appendix. Here we make the following Assumption about $U(c,y,\eta)$ and $y(\rho,R,\eta)$:

ASSUMPTION 1: *For each η, $U(c,y,\eta)$ is continuous in (c,y), increasing in c, decreasing in y, and strictly quasi-concave in (c,y). Also $y(\rho,R,\eta)<\infty$ and $y(\rho,R,\eta)$ is continuously differentiable in $\rho,R>0$.*

The strict quasi concavity of $U(c,y,\eta)$ is essentially equivalent to uniqueness of $y(\rho,R,\eta)$. The continuity and monotonicity conditions are standard in the taxable income literature. Also, we will use continuous differentiability for some of the results to follow.

The analysis to follow will focus on the CDF of taxable income for a fixed budget set as η varies. For a linear budget set this CDF is that of $y(\rho,R,\eta)$ for fixed ρ and R. Let G denote the distribution of η. The taxable income CDF $F(y|\rho,R)$ for a linear budget set is given by

$$F(y|\rho,R)=\int 1(y(\rho,R,\eta)\leq y)G(d\eta) \tag{3.3}$$

This CDF plays a pivotal role in the analysis to follow.

The model we are analyzing is a random utility model (RUM) of the kind considered by McFadden (2005) for continuous choice (see also McFadden and Richter, 1991, for discrete choice). The model here specializes the RUM to $U(c,y,\eta)$ that are strictly quasi-concave and $y(\rho,R,\eta)$ that is smooth in ρ and R. Single valued, smooth demand specifications are often used in applications. In particular, smoothness has often proven useful in applications of nonparametric models and it will here.

McFadden (2005) derived restrictions on $F(y|\rho,R)$ that are necessary and sufficient for a RUM. With choice over two dimensions (c and y) there is a simple, alternative characterization of the RUM. The characterization is that the CDF satisfy a Slutzky like condition, referred to henceforth as the Slutzky condition. The following result holds under technical conditions that are given in Assumption A2 of the Appendix. Let $F_\rho(y|\rho,R)=\partial F(y|\rho,R)/\partial\rho$ and $F_R(y|\rho,R)=\partial F(y|\rho,R)/\partial R$ when these partial derivatives exist.

THEOREM 1: *If Assumptions 1, A1, and A2 are satisfied then $F(y|\rho,R)$ is continuously differentiable in ρ and R and*

$$F_\rho(y|\rho,R)-yF_R(y|\rho,R)\leq 0. \tag{3.4}$$

Also, if for all $\rho,R>0$, $F(y|\rho,R)$ is continuously differentiable in y, ρ, R, the support of $F(y|\rho,R)$ is $[y_\ell,y_u]$, $\partial F(y|\rho,R)/\partial y>0$ on (y_ℓ,y_u), and equation (3.4) is satisfied then there is a RUM satisfying Assumption 1.

In this sense, for two goods and single valued smooth demands, the revealed stochastic preference conditions are that the CDF satisfies the Slutzky condition. This result will be used in the analysis to follow and is of interest in its own right. Dette, Hoderlein, and Neumeyer (2011) showed that each quantile of $y(\rho, R, \eta)$ satisfies the Slutzky condition for demand functions under conditions similar to those of Assumption A2. Hausman and Newey (2014) observed that when a quantile function satisfies the Slutzky condition there is always a demand model with that quantile function. Theorem 1 is essentially those results combined with the inverse function theorem, that implies that the CDF satisfies the Slutzky condition if and only if the quantile satisfies the Slutzky condition.

4 Nonlinear Budget Sets

In practice tax rates vary with income, so the budget frontier is nonlinear. To describe choice in this setting let $B(y)$ denote the maximum obtainable consumption for income y allowed by a tax schedule that we will refer to as the budget frontier. The set of points $\{(y, B(y)) : y \geq 0\}$ will be the frontier of the budget set $\mathcal{B} = \{(y, c) : 0 \leq y, 0 \leq c \leq B(y)\}$. Under the monotonicity condition of Assumption 1 that utility is strictly increasing in consumption, the choice $y(B, \eta)$ of taxable income by individual η will lie on the budget frontier. This choice is given by

$$y(B, \eta) = \text{argmax}_y \, U(B(y), y, \eta).$$

When the budget frontier B is concave the choice $y(B, \eta)$ will be unique by strict quasi-concavity of preferences. In general, when B is not concave the choice $y(B, \eta)$ could be a set. Here we will assume the set valued choices occur with probability zero in the distribution of η and so ignore them.

In practice most tax systems have a finite number of rates that change at certain income values. In such cases the budget frontier is piecewise linear. A continuous, piecewise-linear budget set with J segments, indexed by j, can be described by a vector $(\rho_1, \ldots, \rho_J, R_1, \ldots, R_J)$ of net-of-tax rates ρ_j (slopes) and virtual incomes R_j (intercepts). It will have kink points $\ell_0 = 0$, $\ell_J = \infty$, and $\ell_j = (R_{j+1} - R_j)/(\rho_j - \rho_{j+1})$, $(1 \leq j \leq J-1)$. The budget frontier will be

$$B(y) = \sum_{j=1}^{J} 1(\ell_{j-1} \leq y < \ell_j)(R_j + \rho_j y).$$

In what follows we will also present some results for the case where budget sets need not be piecewise linear.

In general the CDF of taxable income $y(B, \eta)$ will depend on the entire frontier function B. An important simplification occurs around points where $B(y)$ is concave, i.e. where the marginal tax rate is increasing. Let $\bar{\mathcal{B}}$ denote the convex hull of the budget set and $\bar{B}(y) = \max_{(c,y)\in\bar{\mathcal{B}}} c$ denote the corresponding budget frontier. Note that by standard convex analysis results $\bar{B}(y)$ will be a concave function. Let

$$\rho(y) = \lim_{z\downarrow y} \left[\bar{B}(z) - \bar{B}(y)\right]/(z-y),\ R(y) = \bar{B}(y) - \rho(y)y,$$

denote the slope from the right $\rho(y)$ of $\bar{B}(y)$ and $R(y)$ the corresponding virtual income, where the limit $\rho(y)$ exists by Rockafellar (1970, pp. 214-215). Also let $F(y|B) = \int 1(y(B,\eta) \leq y)G(d\eta)$ denote the CDF of taxable income for a budget frontier B.

Theorem 2: *If Assumptions 1 and A1 are satisfied then for all y such that there is $\Delta > 0$ with $\bar{B}(z) = B(z)$ for $z \in [y, y+\Delta]$ we have $F(y|B) = F(y|\rho(y), R(y))$.*

Here we find that the CDF is that of a linear budget set at the right slope $\rho(y)$ and corresponding virtual income $R(y)$ at any value y where the frontier coincides with the frontier of the convex hull on a neighborhood to the right of y. The slope from the right $\rho(y)$ and the neighborhood to the right of y appear here because of the weak inequality in the definition of the CDF. This theorem is a distributional result corresponding to observations of Hall (1973) and Hausman (1979) that linear budget sets can be used to characterize choices when preferences are convex and the budget frontier is concave.

This result is an important dimension reduction in the way the CDF depends on the budget set. In principle $F(y|B)$ can depend on the entire frontier B, an infinite dimensional object. When the frontier is locally concave (to the right of y) the CDF depends only on the slope $\rho(y)$ and virtual income $R(y)$ instead of on the entire budget set. Furthermore, the CDF is that for a linear budget set. This result has a number of useful implications that are discussed in the rest of the paper. For example, this dimension reduction makes it possible to nonparametrically estimate how the expected value of taxable income varies with convex budget sets, as we do in the application below.

In many applications nonconvexities occur only at small values of income. Theorem 2 could be used to nonparametrically quantify how the CDF depends on the budget set at higher values of y where the conditions of Theorem 2 are satisfied. For example, one could nonparametrically estimate the revenue effect of changing taxes on higher income earners. Such an object would be of interest because most of the revenue often comes from those paying higher taxes. We leave this use of Theorem 2 to future work.

Theorem 2 implies a revealed stochastic preference result for convex budget sets. As shown by Theorem 1, for linear budget sets and preference satisfying the conditions of Assumptions 1 and A1, a necessary and sufficient condition for a RUM is that the CDF satisfy the Slutzky condition. An implication of Theorem 2 is that this result is also true for convex budget sets. The CDF of taxable income for convex budget sets is consistent with a RUM if and only if the CDF satisfies the Slutzky condition for linear budget sets.

Theorem 2 can also be used to derive identification results for the CDF and conditional expectation of taxable income for a linear budget set. Let $\mathcal{S}$ denote a set of budget frontiers and $X(y) = \{(\rho(y), R(y)) : B \in \mathcal{S}\}$. Then $F(y|\rho, R)$ is identified for $(\rho, R) \in X(y)$. Also, the conditional mean for a linear budget set $\int yF(dy|\rho, R)$ is identified for $(\rho, R) \in \cap_y X(y)$.

In many applications the budget set may be nonconvex. It would be useful to know how the CDF depends on the budget set in these cases. We can show that the CDF only depends on $B(y)$ over the the values of y where $B(y)$ is not concave. For simplicity we show this result for the case where $B(y)$ has only one nonconcave segment. Let $[\underline{y}(B), \bar{y}(B)]$ denote the interval where $B(y)$ may not be concave and let $\breve{B} = \{\underline{y}(B), \bar{y}(B), B(z)|_{z\in[\underline{y}(B),\bar{y}(B)]}\}$ denote the interval endpoints and the budget frontier over the interval.

Theorem 3: *If Assumption 1 and A1 are satisfied then for all B such that $\bar{B}(z) = B(z)$ except possibly for $z \in (\underline{y}(B), \bar{y}(B))$ and for $y \in [\underline{y}(B), \bar{y}(B))$ we have $\Pr(y(B, \eta) \leq y)$ depends only on $\breve{B}$.*

This result shows that the CDF of taxable income depends on the budget frontier over the entire nonconvex interval, for any point in the interval. Thus, for a piecewise linear budget constraint the CDF would depend on the slope and virtual income of all the segments that affect that nonconvex interval, when y is in that interval.

5 Kinks and Nonparametric Compensated Tax Effects

Kinks have been used by Saez (2010) and others to provide information about compensated tax effects for small kinks or parametric models. In this section we derive the nonparametric form of a kink probability with general heterogeneity and show how it is related to compensated effects. We also consider in our nonparametric setting how the Slutzky condition is related to a positive kink probability and the density of taxable income being positive.

Consider a kink $\bar{\ell}$ for a piecewise linear budget frontier where the frontier coincides with that of the convex hull in a neighborhood of the kink and let $\Pi_{\bar{\ell}}$ denote the kink probability. Let ρ_- and ρ_+ be the slope of the budget frontier at $\bar{\ell}$ from the left and right respectively. Consider ρ between ρ_- and ρ_+ and let $R(\rho) = R_- + \bar{\ell}(\rho_- - \rho)$ be the virtual income for the linear budget set with slope ρ passing through the kink. Assuming that $y(\rho, R(\rho), \eta)$ is continuously distributed, let

$$\phi(\rho)=\frac{\partial F(\bar{\ell}|\rho,R(\rho))}{\partial y},\delta(\rho)=E\left[\frac{\partial y(\rho,R,\eta)}{\partial\rho}-\bar{\ell}\frac{\partial y(\rho,R,\eta)}{\partial R}\middle|y(\rho,R,\eta)=\bar{\ell}\right]\Big|_{R=R(\rho)}$$

where the expectation is taken over the distribution of η.

Theorem 4: *If Assumptions 1, A1, and A2 are satisfied then*

$$\Pi_{\bar{\ell}}=\int_{\rho_+}^{\rho_-}\phi(\rho)\delta(\rho)d\rho \text{ and } \phi(\rho)\delta(\rho)=-F_\rho(\bar{\ell}|\rho,R(\rho))+\bar{\ell}\cdot F_R(\bar{\ell}|\rho,R(\rho)). \quad (5.5)$$

The $\delta(\rho)$ in Theorem 4 is an average compensated effect of changing ρ for a linear budget set. This compensated effect appears here because virtual income is being adjusted as ρ changes to stay at the kink. The virtual income adjustment needed to remain at the kink corresponds locally to the income adjustment needed to remain on the same indifference curve, as shown by Saez (2010). The formula for $\Pi_{\bar{\ell}}$ bears some resemblance to the kink probability formulas in Saez (2010) but differs in important ways. Theorem 4 is global, nonparametric, and takes explicit account of general heterogeneity, unlike the Saez (2010) results, which are local or parametric and account for heterogeneity implicitly.

Theorem 4 helps clarify what can be nonparametrically learned from kinks. First, the compensated effects that enter the kink probability are only for individuals who would choose to locate at the kink for a linear budget set with $\rho\in[\rho_+,\rho_-]$. Thus, using kinks to provide information about compensated effects is subject to the same issues of external validity as, say, regression discontinuity design (RDD). As RDD only identifies treatment effects for individuals at the jump point so kinks only provide information about compensated effects for individuals who would locate at the kink.

Second, the kink probability depends on both a compensated tax effect $\delta(\rho)$ and on a pure heterogeneity effect $\phi(\rho)$. Intuitively, a kink probability could be large because the compensated tax effect is large or because preferences are distributed in such a way that many like to be at the kink. Information about compensated effects from kinks depends on knowing something about pure heterogeneity effects.

Third, the pure heterogeneity effect, and hence compensated effects, is not identified when ρ_- and ρ_+ do not vary in the data. One cannot identify the pdf $\phi(\rho)$ for $\rho\in(\rho_+,\rho_-)$ because observations will not be available for such ρ values. Because of this it may be impossible to say anything about compensated effects from kinks. An example can be used to illustrate. Suppose that the parameter θ of interest is a weighted average (over ρ) of compensated effects $\theta=\int_{\rho_+}^{\rho_-}\phi(\rho)\delta(\rho)d\rho/\int_{\rho_+}^{\rho_-}\phi(\rho)d\rho$. Theorem 4 gives

$$\theta=\frac{\Pi_{\bar{\ell}}}{\int_{\rho_+}^{\rho_-}\phi(\rho)d\rho}.$$

Evidently θ depends on the denominator $\int_{\rho_+}^{\rho_-} \phi(\rho)d\rho$. If ρ_- and ρ_+ are fixed then $\phi(\rho)$ is not identified for $\rho \in (\rho_+, \rho_-)$ so that $\phi(\rho)$ can be anything at all over that interval and the denominator can vary between 0 and ∞. In this setting the kink probability provides no information about θ.

If $\phi(\rho)$ is assumed to satisfy certain conditions then a kink probability can provide information about θ when ρ_- and ρ_+ are fixed. We can continue to illustrate using the parameter θ. As in Saez (2010), $\phi(\rho_-)$ and $\phi(\rho_+)$ may be identified from the pdf of taxable income to the left and right of the kink respectively. If $\phi(\rho)$ is assumed to be monotonic for $\rho \in (\rho_+, \rho_-)$ then we have bounds on θ of the form

$$\frac{\Pi_{\bar{\ell}}}{(\rho_- - \rho_+)\max\{\phi(\rho_-), \phi(\rho_+)\}} \leq \theta \leq \frac{\Pi_{\bar{\ell}}}{(\rho_- - \rho_+)\min\{\phi(\rho_-), \phi(\rho_+)\}}.$$

If $\phi(\rho)$ is assumed to be linear on $\rho \in (\rho_+, \rho_-)$ then

$$\theta = \frac{\Pi_{\bar{\ell}}}{(\rho_- - \rho_+)[\phi(\rho_-) + \phi(\rho_+)]/2}.$$

Thus we see that assumptions about $\phi(\rho)$ can used to obtain information about θ from the kink.

In some data the kink may remain fixed while ρ_- and/or ρ_+ varies. This could occur in a cross section due to variation in local tax rates. In such cases it may be possible to obtain information about $\phi(\rho)$ from the data as ρ varies. This information may then be combined with kink probabilities to obtain information about compensated effects. For brevity we will not consider this kind of information here.

This example of a weighted average compensated effect θ is meant to highlight the importance of the pure heterogeneity term $\phi(\rho)$ in recovering compensated effects from kink probabilities. Similar issues would arise for measures of compensated effects other than θ. Nonparametrically recovering information about compensated effects from kink probabilities generally requires assuming or knowing something about the pure heterogeneity term.

Theorem 4 can also be used to relate positivity of the kink probability $\Pi_{\bar{\ell}}$ to the Slutzky condition. One could specify a CDF $F(y|\rho, R)$ for taxable income for a linear budget set and derive the probability of a kink from equation (5.5). Then the Slutzky condition is sufficient but not necessary for positivity of $\Pi_{\bar{\ell}}$, because an integral can be positive without the function being integrated being positive. In this sense the kink probability can be positive without all the conditions for utility maximization being satisfied.

A similar thing happens for the pdf of taxable income for a smooth, concave budget frontier. By Theorem 2 the CDF of taxable income for a smooth budget set is $F(y|\rho(y), R(y))$. By the chain rule the pdf of taxable income implied by the model will be

$$\frac{\partial F(y|\rho(y), R(y))}{\partial y} = F_y(y|\rho(y), R(y)) + \rho_y(y)[F_\rho(y|\rho(y), R(y)) - yF_R(y|\rho(y), R(y))], \qquad (5.6)$$

where $\rho_y(y) = \partial\rho(y)/\partial y$. One could specify a CDF $F(y|\rho, R)$ for taxable income for a linear budget set and derive the pdf from equation (5.6). The first term F_y is nonnegative because it is a pdf. The $\rho_y(y)$ is nonpositive because it is the derivative of the slope of a concave function. Then the Slutzky condition is sufficient for a positive pdf because it means that the second term will be nonnegative and hence so will the sum. However, the Slutzky condition is not necessary for positivity of the pdf of taxable income because the positivity of the pdf F_y can result in positivity of the sum of the two terms even when the second term is negative. In this sense the pdf of taxable income for a smooth concave budget frontier can be positive without all the conditions for utility maximization being satisfied.

This analysis shows that a coherent nonparametric model, one with a positive pdf and kink probabilities, can be constructed without imposing all the conditions of utility maximization. In particular, the distribution of taxable income implied by a particular $F(y|\rho, R)$ can be coherent without the Slutzky condition for the CDF being satisfied. This analysis is consistent with most of the comments of Keane (2011) about a previous literature concerning the relationship between positive likelihoods and utility maximization. We do differ in finding that positive kink probabilities are possible without a Slutzky condition, which could be attributed to our nonparametric framework.

6 The Expected Value of Taxable Income

The expected value $\mu(B) = \int y(B, \eta)G(d\eta)$ of taxable income for a given budget set is useful for identifying important policy effects. It can be used to predict the effect of tax changes on average taxable income. Furthermore, the presence of an additive mean zero disturbance in taxable income can be allowed for. The presence of such an additive disturbance is one way to account for the common occurrence that individuals do not choose to be at kinks, as noted by Burtless and Hausman (1978).

In this Section we derive $\mu(B)$ for piecewise linear budget sets and show how it can be approximated for estimation purposes. To describe the expected value, recall that $F(y|\rho, R)$ is the CDF of taxable income for a linear budget set. Define

$$\begin{aligned}
\bar{y}(\rho, R) &= \int yF(dy|\rho, R), \\
\nu(\rho, R, \ell) &= \int 1(y < \ell)(y - \ell)F(dy|\rho, R), \lambda(\rho, R, \ell) = \int 1(y > \ell)(y - \ell)F(dy|\rho, R).
\end{aligned}$$

These objects are integrals over the CDF $F(y|\rho, R)$ for a linear budget set. The expected value of taxable income given a piecewise linear, convex budget set depends on them in the way shown in the following result:

THEOREM 5: *If Assumption 1 and A1 are satisfied,* $\int |y(\rho, R, \eta)|G(d\eta) < \infty$ *for all* $\rho, R > 0$*, and* $B(y)$ *is piecewise linear and concave then*

$$\begin{aligned} \mu(B) &= \bar{y}(\rho_J, R_J) + \sum_{j=1}^{J-1}[\nu(\rho_j, R_j, \ell_j) - \nu(\rho_{j+1}, R_{j+1}, \ell_j)] \\ &= \bar{y}(\rho_1, R_1) + \sum_{j=1}^{J-1}\left[\lambda(\rho_{j+1}, R_{j+1}, \ell_j) - \lambda(\rho_j, R_j, \ell_j)\right]. \end{aligned} \tag{6.7}$$

The first equality in the conclusion is exactly analogous to the conclusion of Theorem 2.1 of BN. As discussed there, this additive decomposition of the conditional mean makes it feasible to nonparametrically estimate the conditional expectation as a function of the budget set. The fact that the conditional expectation only depends on one two-dimensional function $\bar{y}(\rho, R)$ and one three-dimensional function $\nu(\rho, R, \ell)$ (or $\lambda(\rho, R, \ell)$) means the curse of dimensionality can be avoided by using a nonparametric estimator that imposes the structure in the formula for $\mu(B)$.

Theorem 5 generalizes Theorem 2.1 of BN by allowing general heterogeneity and zero hours of work, whereas BN assumed scalar η. Consequently, the empirical conclusions drawn by BN about the average labor supply effect of a large Swedish tax reform are valid under general heterogeneity. To the best of our knowledge that makes the tax policy estimates of BN the first that are valid with general preference heterogeneity.

We can use Theorem 2, which implies that the expectation depends only on the CDF $F(y|\rho, R)$ for a linear budget set, to construct a more parsimonious approximation to $\mu(B)$ than BN. The definitions of $\bar{y}(\rho, R)$ and $\nu(\rho, R, \ell)$ (or $\lambda(\rho, R, \ell)$) and the conclusion of Theorem 5 give the precise form of the dependence on $F(y|\rho, R)$. Replacing $F(y|\rho, R)$ by a series approximation in those definitions and plugging the result into the formula in Theorem 5 gives a more parsimonious approximation than BN.

The series approximation we use is a linear in parameters approximation to the conditional CDF of taxable income for a linear budget set. For a positive integer A let $F_1(y), ..., F_A(y)$ be CDF's and $x = (\rho, R)$. Let $r_1(x), ..., r_B(x)$ denote approximating functions, such as splines or polynomials. Let β_{ab}, $(a = 2, ..., A; b = 1, ..., B)$ be coefficients of a series approximation to be specified below and $w_a(x, \beta) = \sum_{b=1}^{B} \beta_{ab} r_b(x)$. We consider an approximation to the conditional CDF of the form

$$
\begin{aligned}
F(y|x) &\approx F_1(y) + \sum_{a=2}^{A} w_a(x,\beta)[F_a(y) - F_1(y)] \\
&= \sum_{a=1}^{A} w_a(x,\beta)F_a(y), w_1(x,\beta) = 1 - \sum_{a=2}^{A} w_a(x,\beta).
\end{aligned}
$$

This could be thought of as a mixture approximation to the conditional CDF with weights $w_a(x,\beta), a = 1, ..., A$. We have normalized the weights to sum to one by choosing $w_1(x,\beta)$ as above. Because of this normalization the conditional CDF approximation will go to 1 as y grows for all x. We do not impose that the weights $w_a(x,\beta)$ be nonnegative. We are primarily interested in approximating the expected value and so are not concerned that the underlying approximation to the conditional CDF be everywhere increasing.

We obtain an approximation to the conditional mean by plugging the CDF approximation into the respective formulas for $\bar{y}(\rho,R)$ and $\nu(\rho,R,\ell)$ and then into the formula for the mean in Theorem 5. Let

$$
\bar{y}_a = \int yF_a(dy), \nu_a(\ell) = \int 1(y<\ell)(y-\ell)F_a(dy), (a = 1, ..., A).
$$

Substituting the CDF approximation in the expression for the conditional mean from Theorem 5 gives

$$
\begin{aligned}
\mu(B) &\approx \bar{y}_1 + \sum_{a=2}^{A} w_a(x_J,\beta)(\bar{y}_a - \bar{y}_1) + \sum_{a=2}^{A}\sum_{j=1}^{J-1}[w_a(x_j,\beta) - w_a(x_{j+1},\beta)][\nu_a(\ell_j) - \nu_1(\ell_j)] \\
&= \bar{y}_1 + \sum_{a=2}^{A}\sum_{b=1}^{B}\beta_{ab}\{r_b(x_J)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J-1}[r_b(x_j) - r_b(x_{j+1})][\nu_a(\ell_j) - \nu_1(\ell_j)]\}.
\end{aligned}
$$

This is a series approximation, where the regressor corresponding to β_{ab} is a linear combination of the approximating function evaluated on the last segment and differences of approximating functions between segments. A series estimator can be obtained by running least squares of y_i on these regressors.

A series estimator based on this approximation imposes the restrictions that the same CDF for a linear budget set appears in both $\bar{y}(\rho_J, R_J)$ and in $\nu(\rho,R,\ell)$. This approximation is more parsimonious than BN (based on Theorem 5) because it does not use a separate approximation to $\bar{y}(\rho,R)$. It makes use of Theorem 2, being based on an approximation to the CDF of taxable income for a linear budget set, which is the underlying nonparametric object determining the distribution of taxable income.

By Theorem 1 the one additional restriction imposed by utility maximization is that the CDF for a linear budget set satisfies the Slutzky condition. It is straightforward to impose this restriction on a grid of values for x and y, say $x_1, ..., x_C$, and $y_1, ..., y_D$. The CDF of taxable income for a linear budget set with slope and intercept x that corresponds to this approximation is $F_1(y) + \sum_{a=2}^{A} w_a(x, \beta)[F_a(y) - F_1(y)]$. The Slutzky condition for the CDF approximation at the values of x and y is then

$$\sum_{a=2}^{A} \left[\frac{\partial w_a(x_c, \beta)}{\partial \rho} - y_d \frac{\partial w_a(x_c, \beta)}{\partial R} \right] [F_a(y_d) - F_1(y_d)] \leq 0, (c = 1, ..., C; d = 1, ..., D). \tag{6.8}$$

These are a set of linear in parameters, inequality restrictions on the coefficients β of the weights $w_a(x, \beta)$. A series approximation as above with coefficients satisfying these Slutzky inequalities is an approximation to the expected value that approximately satisfies all the restrictions of utility maximization. Because the only restriction imposed by utility maximization is that the Slutzky condition is satisfied for $F(y|x)$ we know that approximately imposing all those conditions approximately imposes all the conditions of utility maximization.

To show that this approximation works we give a rate result for specific types of CDF's $F_a(y)$ and functions $r_b(x)$. We view this result as a theoretical justification of the approach though it may not be the best for applications. In our application we use different CDF's that are more closely linked to the data we have. The rate result is based on choosing $F_a(y)$ to be integrals of b-splines that are positive and normalized to integrate to one and on $r_b(x)$ also being splines. We also require that y and x be contained in bounded sets $\mathcal{Y}$ and $\mathcal{X}$ and that the conditional pdf of taxable income for a linear budget set $f(y|x)$ be smooth.

THEOREM 6: *If $\mathcal{Y}$ and $\mathcal{X}$ are compact, $f(y|x)$ is zero outside $\mathcal{Y} \times \mathcal{X}$ and is continuously differentiable to order s on $\mathcal{Y} \times \mathcal{X}$, and $dF_a(y)/dy, (a = 1, ..., A)$ and $r_b(x), (b = 1, ..., B)$ consist of tensor product b-splines of order s on $\mathcal{Y} \times \mathcal{X}$ then there exist a constant C and β_{ab} such that for all piecewise linear, concave budget frontiers with $x_j \in \mathcal{X}$, $(j = 1, ..., J)$,*

$$\begin{aligned} & \left| \mu(B) - \bar{y}_1 - \sum_{a=2}^{A} \sum_{b=1}^{B} \beta_{ab} \{ r_b(x_J)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J-1} [r_b(x_j) - r_b(x_{j+1})][\nu_a(\ell_j) - \nu_1(\ell_j)] \} \right| \\ \leq \; & C A^{1-s} B^{(1-s)/2}. \end{aligned}$$

This result shows that the series approximation we have proposed does indeed approximate the expected value of taxable income for concave, piecewise linear budget frontiers. The approximation rate is uniform in the number J of budget segments. The rate of approximation corresponds to a multivariate b-spline approximation to a function and its derivative, where approximating the derivative is useful for making the rate uniform in the number of budget segments.

We can also make allowance for nonconcave budget frontiers. For example, suppose that the budget frontier is nonconcave over only two segments. From Theorem 3 we see that the distribution over those segments will depend only on the slope and intercept of those two segments. Because the expected value is a sum of integrals over different segments it would take the form

$$\mu(B) = \bar{y}(\rho_J, R_J) + \sum_{j=1}^{J-1}[\nu(\rho_j, R_j, \ell_j) - \nu(\rho_{j+1}, R_{j+1}, \ell_j)] + \varsigma(R_{\tilde{j}-1}, \rho_{\tilde{j}-1}, R_{\tilde{j}}, \rho_{\tilde{j}}),$$

where $\tilde{j}$ and $\tilde{j}-1$ index the segments where the nonconcavities occur. The ς term represents the deviation of the mean from what it would be if the budget frontier were concave. It can be accounted for in the approximation by separately including series terms that depend just on $R_{\tilde{j}-1}$, $\rho_{\tilde{j}-1}$, $R_{\tilde{j}}$, and $\rho_{\tilde{j}}$. If the nonconcavities are small or few people have taxable income where they occur then ς would be small and including terms to account for ς will lead to little improvement. The integration across individuals to obtain the expected value reduces the importance of nonconcavities.

7 Policy Effects and Productivity Growth

It is common practice to measure behavioral effects in terms of elasticities. We are used to linear budget constraints and elasticities with respect to the net of tax rate and non-labor income of a linear budget constraint. One problem with nonlinear budget constraints is that this elasticity may not be identified. The elasticity for a linear budget constraint would often be thought of as corresponding to $\bar{y}(\rho, R)$. From the discussion following Theorem 2 we see that this function is only identified at any ρ that is equal to $\rho(y)$ of a budget frontier in the data for every value of y. The set of such net of tax rates could well be very small, even empty. Therefore we must look for other kinds of elasticities to hope for identification. Furthermore, since everyone generally faces a nonlinear budget set, and policy changes are not likely to eliminate this nonlinearity, it makes sense to focus on effects of changes in a nonlinear budget set.

For motivation we first consider effects for the average taxable income $\bar{y}(\rho, R)$ for a linear budget set. As usual, the average net-of-tax effect will be $d\bar{y}(\rho, R)/d\rho$ and the average effect of nonlabor income will be $d\bar{y}(\rho, R)/dR$. Next consider the case where the expected taxable income is a function of a piecewise-linear budget constraint, say $\mu(B) = g(\rho_1, \ldots, \rho_J, R_1, \ldots, R_J)$ for a function g. Assume that the budget constraint is continuous so that the kink points will be well defined by the net-of-tax rates and virtual incomes and are given by $\ell_j = (R_j - R_{j+1})/(\rho_{j+1} - \rho_j)$. Let $G(a, b) = g(\rho_1 + a, \ldots, \rho_J + a, R_1 + b, \ldots, R_J + b)$. The parameter a tilts the budget constraint, and the parameter b shifts the budget constraint vertically, both while holding fixed the kink points. For policy purposes a is like a change in a local proportional tax rate, and b is like a change in unearned income. Identification of effects for changes in a and b only requires variation in the overall slopes and intercepts of the budget constraint across individuals and time periods. This is a common source of variation in nonlinear budget sets due to variations in local tax rates and in nonlabor income, so effects of such a change should be identified.

Consider the derivative of $G(a,b)$ with respect to a evaluated at $a = b = 0$, given by $\partial G/\partial a = \sum_{j=1}^{J} \partial g/\partial \rho_j$. This is the effect on the expectation of tilting the budget constraint. To obtain an elasticity we multiply this derivative by a constant $\tilde{\rho}$ that represents the vector of net-of-tax rates by a single number and then divide by $\mu(B)$. The construction of $\tilde{\rho}$ can be done in many different ways. We use the sample averages of the net-of-tax rates and virtual incomes for the segments where individuals are actually located. Our elasticity $(\partial G/\partial a)(\tilde{\rho}/\mu(B))$ is an aggregate elasticity which is the policy relevant measure as argued in Saez et. al. (2012). We also consider the effect of unearned income R given by $\partial G/\partial b = \sum_{j=1}^{J} \partial g/\partial R_j$. Due to ambiguity on how to normalize this we do not report an elasticity. Instead we will simply report estimates of the unearned income effect $\partial G/\partial b$.

In the long run, exogenous wage growth is a major determinant of individuals' real incomes. Such growth may be caused by factors such as technological development, physical capital, and human capital. It is important to account for such growth when identifying the effects of taxes on taxable income using variation over time as we do in the application below. We do so by assuming that productivity growth is the same in percentage terms for all individuals. We assume the wage rate in period t is given by $w = g(e,\eta)\phi(t)$ with $\phi(0) = 1$. The function $\phi(t)$ is a function that captures exogenous productivity growth, i.e., percentage changes in an individual's wage rate that do not depend on the individual's behavior.

With productivity growth and heterogeneity the individual's optimization problem is:[3]

$$\operatorname{Max} u(c,e,h,\eta) \quad \text{s.t.} \quad c = g(e,\eta)\phi(t)h\rho + R \tag{7.9}$$

This problem can be solved similarly to previous ones, by letting $y = wh$, inverting the wage function, and choosing hours of work to maximize $u(c, g^{-1}(y/(h\phi(t)),\eta), h, \eta)$ over h. Concentrating out hours of work gives the concentrated utility function $U(c, y/\phi(t), \eta)$. In a second step the individual solves $\operatorname{Max} U(c, y/\phi(t), \eta)$ s.t. $c = y\rho + R$.

A feature of this problem is that the concentrated utility shifts over time. Our approach to repeated cross section data depends on using a preference specification invariant to individuals

[3]Note that we are still considering an atemporal model of individual behavior. An individual considers a sequence of one-period optimization problems. The purpose of the extension here is to show how to account for exogenous productivity growth.

and time. A simple way to do that is to focus on taxable income net of productivity growth, given by $\tilde{y} = y/\phi(t)$. Then the reduced-form maximization problem becomes $\text{Max}\, U(c, \tilde{y}, \eta)$ s.t. $c = \tilde{y}\tilde{\rho} + R$ for $\tilde{\rho} = \phi(t)\theta$. Here the productivity growth appears in the budget set, multiplying the net of tax rate. From the tax authorities' point of view the taxable income is $y = \phi(t)\tilde{y}$. However, to keep things stationary over time we study the behavior of $\tilde{y}$.

Although the function $U(c, \tilde{y}, \eta)$ does not shift over time, it depends on a base year and a normalization of $\phi(0)$ to one. If we use another base year we would have another concentrated utility function $U(c, y, \eta)$.

This way to account for productivity growth is similar to that used in log-linear models. Suppose that $y = [\phi(t)\rho]^{\beta}\eta$, where β is the net-of-tax elasticity of interest and that there are no income effects. Taking logarithms gives $\ln y = \beta \ln \phi(t) + \beta \ln \rho + \ln \eta$. Here $\phi(t)$ enters as a time effect and β can be identified in a regression involving the logarithm of the uncorrected variables y and ρ. This is, more or less, how productivity growth has been accounted for in previous models. Including time effects in log-linear models corresponds to the productivity growth specification we adopt here.

To implement the corrections on the net-of-tax rates and the dependent variable we need to know the wage/productivity growth. Unfortunately there are few good measures of the exogenous wage/productivity growth. The productivity measures available in the literature have in general not separated out the change in wages that is due to behavioral effects of tax changes. We will therefore use our data to estimate exogenous wage growth. To not use up too much identifying information when doing this we constrain the annual productivity growth to be the same every year, where $\phi(t) = e^{gt}$ for some constant g. This may well be misspecified. However, to do a more refined correction of the budget constraints would use up much of the information in the data. In particular we would lose much of the identifying power of changes in the overall tax rate across years. We do not think there are wide swings in the productivity growth rate from year to year so that the misspecification would not be very large for any individual year's budget constraints.

In the long run changes in tax rates can be swamped by productivity growth. For example, over say a twenty-year period, if the annual productivity growth is 0.02, $\phi(20)/\phi(0)$ will be 1.5, corresponding to an increase in the net-of-tax rate of a factor of 1.5. In the short run, changes in tax rates can swamp short-run changes in $\phi(t)$. For example, a change in the tax rate from, say, 0.6 to 0.4 raises ρ by a factor of 1.5.

In a linear budget set, productivity growth and tax-rate changes have the same kind of effect on net-of-tax rates. It can therefore be difficult to nonparametrically separate the two kinds of effects. In a nonlinear budget set the situation is different. Consider an example with two budget segments. The budget constraint can then be written as $c = \tilde{y}\phi(t)\rho_1 + R_1$ for $\tilde{y} < \phi(t)^{-1}\ell_1$ and $c = \tilde{y}\phi(t)\rho_2 + R_2$ for $\tilde{y} > \phi(t)^{-1}\ell_1$. In this specification productivity changes shift both slopes and kinks, a different effect than just a change in slopes. These effects are also present for budget sets with many segments. Thus, productivity changes have different effects on the budget sets than just changing slopes, so it may be possible to separate out the effect of productivity growth and tax rate changes in our estimates.

It would be interesting to allow for the productivity growth rate to vary with individuals, i.e. to allow for heterogeneity in productivity growth. We do this in the empirical application by allowing productivity growth to vary between more and less educated individuals. We find that allowing this observed heterogeneity does not impact our tax and income effect estimates. We could also allow for unobserved heterogeneity by specifying that $c = g(e,\eta)\phi(t,\eta)h\rho + R$. Unfortunately this does not lead to a stable utility function over time and so would not allow us to use time variation in taxes to identify tax effects. This variation appears to be important for the ability to estimate tax effects in our data so we do not pursue models with unobserved heterogeneity in productivity growth.

8 Empirical Application

The previous results are based on the expected value and distribution of taxable income for a given budget frontier. These objects are identified when the budget sets in the data are independent of preferences. We can also allow for an additive disturbance with mean zero. That is, suppose that the data consist of observations on taxable income and budget frontiers for individuals (Y_i, B_i), $(i = 1, ..., n)$ with $Y_i = y(B_i, \eta_i) + \epsilon_i$, B_i and η_i are statistically independent, and $E[\epsilon_i|B_i] = 0$. In that case

$$E[Y_i|B_i = B] = \mu(B).$$

Here the expected value of taxable income for a given frontier is the conditional expectation of taxable income in the data. The expected value can then be estimated by the nonparametric series estimator described earlier. Specifically, for $\bar{y}_a$, $\nu_a(\ell)$, and $r_b(x)$ as previously defined, $Y_i - \bar{y}_1$ could be regressed on

$$r_b(x_{i,J_i})(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J_i-1}[r_b(x_{ij}) - r_b(x_{i,j+1})][\nu_a(\ell_{ij}) - \nu_1(\ell_{ij})], a = 2, ..., A; b = 1, ..., B, \quad (8.10)$$

where J_i is the number of budget segments for the i^{th} observation, $x_{ij} = (\rho_{ij}, R_{ij})$ is the slope and intercept of the j^{th} segment, and ℓ_{ij} is the location of the j^{th} knot. For the coefficients $\hat{\beta}_{ab}$

obtained from this regression, the estimator of the expected value of taxable income for any concave piecewise linear budget frontier is then

$$\hat{\mu}(B) = \bar{y}_1 + \sum_{a=2}^{A}\sum_{b=1}^{B}\hat{\beta}_{ab}\{r_b(x_J)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J-1}[r_b(x_j) - r_b(x_{j+1})][\nu_a(\ell_j) - \nu_1(\ell_j)]\}.$$

Independence of η_i and B_i can be relaxed to allow for covariates and/or control functions. For covariates w we consider an index specification where there is a vector of functions $v(w,\delta)$ such that η_i and $(B_i, v(w_i,\delta_0))$ are independent for some parameter value δ_0. These covariates might include demographic variables that represent observed components of the utility. For example, one could use a single, linear index $v(w,\delta) = w_1 + w_2^T\delta$, with the usual scale and location normalization imposed. Covariates can be allowed by letting the functions $r_b(x,v)$ depend on the index v as well as the slope and intercept x. This specification corresponds to preferences that are allowed to depend on $v(w,\delta_0)$. The parameters β_{ab} and δ could then be estimated by nonlinear least squares, as the minimizers of

$$\begin{aligned}&\sum_{i=1}^{n}\{Y_i - \bar{y}_1 - \sum_{a=2}^{A}\sum_{b=1}^{B}\beta_{ab}\{r_b(x_{i,J_i}, v(w_i,\delta))(\bar{y}_a - \bar{y}_1)\\ &+\sum_{j=1}^{J_i-1}[r_b(x_{ij}, v(w_i,\delta)) - r_b(x_{i,j+1}, v(w_i,\delta))][\nu_a(\ell_j) - \nu_1(\ell_j)]\}\}^2.\end{aligned}$$

An estimable control variable can be used to account for endogeneity. Such a control variable would be ξ_i such that B_i and η_i are independent conditional on ξ_i and the conditional support of ξ_i given B_i equals the marginal support of ξ_i. In that case it follows as in Blundell and Powell (2006) that

$$\int E[Y_i|B_i = B, \xi_i = \xi]F_\xi(d\xi) = \mu(B),$$

where $F_\xi(\xi)$ is the CDF of ξ_i. This integral can be estimated by letting the functions $r_b(x,\xi)$ depend on ξ and then including $\hat{\xi}_i$ to form the regressors

$$r_b(x_{i,J_i}, \hat{\xi}_i)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J_i-1}[r_b(x_{ij}, \hat{\xi}_i) - r_b(x_{i,j+1}, \hat{\xi}_i)][\nu_a(\ell_{ij}) - \nu_1(\ell_{ij})], a = 2, ..., A; b = 1, ..., B.$$

The expected value of taxable income could then be estimated by averaging over the observations $\hat{\xi}_i$, with

$$\begin{aligned}\hat{\mu}(B) &= \bar{y}_1 + \sum_{a=2}^{A}\sum_{b=1}^{B}\hat{\beta}_{ab}\{\bar{r}_b(x_J)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J-1}[\bar{r}_b(x_j) - \bar{r}_b(x_{j+1})][\nu_a(\ell_j) - \nu_1(\ell_j)]\},\\ \bar{r}_b(x) &= \sum_{i=1}^{n} r_b(x, \hat{\xi}_i)/n.\end{aligned}$$

Although conditions for existence of a control variable are quite strong (see Blundell and Matzkin, 2014), this approach does provide a way to allow for some forms of endogeneity.

In the application we use $F_a(y)$ that are constructed from the data in a way described in the next Section. We use power series terms for $r_b(x, v, \xi)$, with

$$r_b(x, v, \xi) = \rho^{m_1(b)} R^{m_2(b)} v^{m_3(b)} \xi^{m_4(b)},$$

where $m_k(b)$ are nonnegative integers.

An important problem for practice is the selection of which approximating functions to include in estimation. For this purpose we use a hybrid of standard Lasso and cross-validation approaches to model selection. To describe the method let i index the observations and $p_i^K = (p_{i1}^K, ..., p_{iK}^K)'$ be the vector of approximating functions, say from equation (8.10) when there are no covariates or control functions. Here K denotes the number of approximating functions, which will be $(A-1)B$ in the case of equation (8.10). We adopt an approach similar to Belloni and Chernozhukov (2013) in using Lasso as a model selection method. The goal is to select a subvector of p_i^K to use in the least squares estimation described above. Note that this process begins with a choice of K, which is generally reasonably large. The regressors need to be normalized so that they all have sample second moment equal to one. This can be done by dividing each observation p_{ik}^K on the k^{th} regressor by $\{\sum_{i=1}^n (p_{ik}^K)^2/n\}^{1/2}$. Here is a multi-step description of the model selection method we use:

a) Get a first estimate of $\hat{\mu}(B)$ by using all K terms, where K is chosen by cross-validation or some other method.

b) Calculate the sample residuals $\hat{\varepsilon}_i = Y_i - \hat{\mu}(B_i)$.

c) Draw a sample $(\hat{\varepsilon}_1^b, ..., \hat{\varepsilon}_n^b)$ of size n from the empirical distribution of the residuals.

d) Calculate $\hat{S}_b = 2\max_{k\leq K} \left|\sum_{i=1}^n p_{ik}^K \hat{\varepsilon}_i^b/n\right|$.

e) Repeat c) and d) B times to obtain $\hat{S}_1, ..., \hat{S}_B$. We could set B to be a few hundred, say 500, to help estimate the .95 quantile of the distribution of $\hat{S}$.

f) Choose $\hat{S}$ to be the $B * .95$ order statistic of $\hat{S}_1, ..., \hat{S}_B$. This is an estimate of the .95 quantile of the distribution of $\hat{S}$.

g) Let $\hat{\lambda} = cn\hat{S}$ where $c > 1$. We use $c = 2$ and try different values of c, mostly smaller.

h) Minimize

$$\frac{\sum_{i=1}^n (Y_i - p_{ik}^{K\prime}\beta)^2}{n} + \frac{\hat{\lambda}}{n}\sum_{j=1}^J |\beta_j|.$$

using the Matlab program for lasso.

i) Do OLS using only those elements of p_{ik}^K which had nonzero coefficients in the previous step.

j) Try other values of λ as suggested in g) above corresponding to c closer to 1 and slightly bigger than 2.

k) For each choice of λ compute the cross-validation criteria using the OLS estimates from i) for each model, to provide some goodness of fit comparisons.

Once we have estimated coefficients $\hat{\beta}_{ab}$ in hand by following the above procedure, we can construct taxable income elasticities and income effects in the way described previously. It is straightforward to construct the nonlinear budget set effects described earlier from the estimated parameters for a given budget set, control function, and residual function. We average over the sample distribution of budget sets, single-index and control functions to obtain an average effect. We also use an analogous procedure to estimate the effect of some tax reforms on the expected value of taxable income. We estimate standard errors using the delta method.

9 Application to Sweden

9.1 Data

We use data from HEK (Hushållens Ekonomi) provided by Statistics Sweden, which is a combined register and survey data set. The data set contains repeated cross sections of approximately 17,000 randomly-sampled individuals from the population and members of their households each year. The response rate is approximately seventy percent. The register component contains income, tax, and demographic data used by the authorities for taxation purposes. The survey component primarily contains housing variables required to construct several housing-related budget set variables such as the housing allowance, which are important components of the budget sets.

We use data covering a period of sixteen years, from 1993 to 2008. In the estimation, we limit the sample to married or cohabiting men between 21 and sixty years of age. This is the economically most significant group with respect to labor income. We exclude those receiving medical-leave benefits, parental benefits, income from self-employment, or student financial aid above half of the average monthly gross labor income, which was 17,607 SEK in 2008. We use this limit instead of zero, as that would result in a large loss of observations. Out of 102,630 married or cohabiting men between 21 and sixty years of age, 81,718 observations remain after this sample restriction.

Our labor income definition primarily includes third-party reported earned income and income from self-employment. It excludes, however, medical-leave benefits and parental benefits, unlike previous studies using Swedish data. Note that those individuals with large amounts of income from these sources are excluded from the sample.

To construct the individual budget sets, we use a micro simulation model, FASIT, developed by Statistics Sweden, which in principle captures all of the features of the Swedish tax and transfer system relevant for individuals. FASIT is used by, e.g., the Swedish Ministry of Finance, to simulate the mechanical effects of various tax policies including potential future policies. Single cross sections of this model have been previously employed by Flood et al. (2007), Aaberge and Flood (2008), and Ericson et al. (2009). We construct the budget sets by iteratively letting FASIT calculate net family incomes by varying individuals' gross labor incomes. When doing this, we set medical-leave compensation to zero as this is a component that is difficult to predict for the individuals in the beginning of the year when planning how much to work during the year.

We set non-labor income as the net income the family would receive if the husband had no labor income. This component includes the spouse's net labor income, family's net capital income, and various welfare benefits the family would receive if the husband had no labor income. For capital incomes, we set capital gains and losses to zero for the same reason as we set sick leave benefits to zero. Additionally, we include the implicit income from residence-owned housing in nonlabor income.

Nonlabor income may be endogenous. We instrument nonlabor income using transfers received at zero labor income. This includes, e.g., housing and child allowances and social assistance. Like the tax system, the transfer system is beyond the control of individuals. The transfers that would be received at zero labor income vary between individuals depending on demographics. Because we control for such factors, most of the variation arises due to changes in the transfer system between years and how these changes affect different individuals differently. The control variable we use is the residual from a linear regression of nonlabor income on the instrument and the demographic variables described below.

We also adjust the budget sets for indirect taxation. Payroll taxes are generated by FASIT, while we make a simple rudimentary correction for consumption taxes using the quotient of aggregate value-added-tax revenues divided by aggregate private consumption for each year separately. These additional corrections are similar to those in Blomquist and Newey (2002).

The data set contains many demographic background variables. We control for age (eight groups), educational level (seven groups), socioeconomic occupational groups (eight groups), spousal income (twenty groups), county of residence (22 groups), whether the individual has children below age six, and whether he was born abroad. These variables are all included as covariates in the index function described above.

In Table B1 in Appendix B, we report sample statistics. We report the mean values of gross labor income, some variables characterizing the budget sets, some demographic variables, and the instrument for nonlabor income. We report statistics for the entire sample, as well as for the years 1993, 1998, 2003, and 2008 separately, to illustrate the development over time.

9.2 Results

We do nonparametric estimation using a power series approximation with selection among series terms up to a fourth order using the procedures described in Section 8. For the approximation we use three CDF's $F_a(y), (a = 1, 2, 3)$, corresponding to the marginal empirical distribution of taxable income for the whole sample, for the smallest 1/5 of the sample, and for the largest 1/5 of the sample. For constructing $\mu_a(\ell)$ we use the sample CDF for a subsample of 5,000 observations evenly spaced by gross labor income rank.

We report the net-of-tax elasticity described in Section 7. The marginal effect is evaluated at each of the sample individual budget sets and we report the sample average marginal effects scaled by the sample mean income and marginal net-of-tax rate. We report the marginal effect for the income effect, i.e., the change in gross labor income as net nonlabor income increases by one SEK. We do this because the income elasticity is highly dependent on the scaling parameters. The natural scaling factor, the average marginal virtual income is extremely high in our case compared to other studies that report income elasticities because the marginal tax rate is low and nonlabor income is high. Nonlabor income is high because we include the spouse's net income and implicit income from residence-owned housing. The marginal income effect is therefore more informative.

We implicitly estimate the productivity growth rate by estimating specifications assuming different productivity growth rates and by selecting the specification that maximizes the cross-validation value. We vary the productivity growth rate between 0% and 1.4% in steps of 0.1% when doing this. In Table 1, we report the estimated cross-validation values at different growth rates and specifications with different Lasso lambdas. We normalize so that larger cross-validation values correspond to better fits.

Table 1: Cross-validation Values

Growth/Lambda	2	1.5	1	0.5	0.2	0.1
0%	0.1003	0.0891	0.0945	0.1026	-0.6284	-0.2095
.1%	0.1001	-1.3758	-1.3758	0.0915	-0.1517	-3.95
.2%	0.0987	0.1039	0.1039	0.0106	-2.4467	-10.82
.3%	0.0968	0.1067	0.1067	0.1112	-0.1563	-0.0227
.4%	0.0945	0.0997	0.0997	-0.583	-10.471	-1.8175
.5%	0.0919	0.0933	0.0933	0.1036	-1.3474	-9.5351
.6%	0.0889	0.0942	-1.9278	-0.0397	0.0728	-1.0969
.7%	0.0858	0.0678	0.0824	0.1004	0.0773	-0.4132
.8%	0.0824	0.0878	0.0878	0.0702	0.1095	0.0028
.9%	0.0756	-0.0432	-0.0432	0.0916	-0.7532	-0.4593
1.0%	0.0755	0.0807	0.0807	0.0699	0.107	-0.9698
1.1%	0.072	0.0773	0.0771	0.0809	-2.6009	-0.0756
1.2%	0.0401	0.083	0.083	0.073	-0.8163	-0.4652
1.3%	0.0649	0.07	0.0683	0.0792	-0.0764	-22.41
1.4%	0.0518	0.0671	0.0409	0.067	-0.9842	-0.4108

We observe that the cross-validation criterion is maximized at the productivity growth rate .3% (at Lasso lambda .5). Gross labor income increased by a factor of 1.23 during the sample period implying an average geometric annual growth rate of 1.4%. The way we have estimated productivity growth here means that we would attribute most of the increase in gross labor income to average responses to changes in the tax structure.

In Table 2, we report the net-of-tax elasticity point estimates in the different specifications. The elasticities are evaluated at the sample mean income of 420,329 SEK and marginal net-of-tax rate of .32. Standard errors constructed using the delta method are either .08 or .09 in all specifications. We observe that the estimated elasticity decreases as we increase the productivity growth rate. This reflects that, overall, taxes have decreased over the sample period at the same time as gross labor incomes have increased. Using a specification that assumes a too low (high) productivity growth rate (such as assuming no growth) would therefore result in a positive (negative) bias in the estimated elasticities. Accounting for productivity growth appropriately is therefore important. At the specification with largest cross validation value the net-of-tax elasticity is .21 and statistically significant, with a .3% growth rate. This is roughly the same elasticity found by Blomquist and Selin (2010). The highest cross validation values give estimated net-of-tax elasticities of 0.38 for a growth rate of 0.2% and 0.07 for a slightly higher growth rate of .4%.

Table 2: Estimated Net of Tax Elasticities.

Growth/Lambda	2	1.5	1	0.5	0.2	0.1
0%	0.76	0.71	0.71	0.68	0.72	0.79
.1%	0.62	0.55	0.55	0.55	0.56	0.62
.2%	0.38	0.38	0.38	0.37	0.36	0.46
.3%	0.22	0.52	0.52	0.21	0.27	0.25
.4%	0.07	0.07	0.07	0.05	0.68	0.05
.5%	-0.05	-0.05	-0.05	-0.07	-0.07	0.6
.6%	0.05	-0.15	-0.17	-0.07	-0.14	-0.16
.7%	-0.21	-0.22	-0.2	-0.21	-0.21	-0.24
.8%	-0.14	-0.25	-0.25	-0.24	-0.23	-0.23
.9%	-0.24	-0.24	-0.24	-0.25	-0.22	-0.25
1.0%	-0.29	-0.29	-0.29	-0.25	-0.25	-0.2
1.1%	-0.31	-0.3	-0.25	-0.26	0	-0.32
1.2%	-0.28	-0.28	-0.28	-0.34	-0.34	-0.29
1.3%	-0.38	-0.35	-0.01	-0.38	-0.34	-0.34
1.4%	-0.44	-0.4	-0.4	-0.41	-0.41	-0.41

In Table 3, we report the marginal nonlabor income effect estimates in the different specifications. We see that the estimated marginal effect is around -1 and fairly stable across specifications. In the specification with largest cross-validation value, the estimated marginal income effect is -1.02. The standard errors are around 0.13 and 0.14 throughout Table 3.

Table 3: Estimated Income Effects

Growth/Lambda	2	1.5	1	0.5	0.2	0.1
0%	-0.7	-0.69	-0.69	-0.73	-0.74	0.01
.1%	-0.85	-0.79	-0.79	-0.93	-0.87	-0.09
.2%	-0.9	-0.9	-0.9	-1	-0.94	-0.66
.3%	-0.99	-0.9	-0.9	-1.02	-1.09	-1.12
.4%	-1.07	-1.07	-1.07	-1.11	-0.86	-1.07
.5%	-1.14	-1.14	-1.14	-1.17	-1.18	-0.96
.6%	-1.16	-1.19	-1.23	-1.1	-1.17	-1.2
.7%	-1.22	-1.23	-1.15	-1.18	-1.19	-1.26
.8%	-1.26	-1.24	-1.24	-1.18	-1.2	-1.18
.9%	-1.15	-1.15	-1.15	-1.17	-1.23	-1.17
1.0%	-1.22	-1.23	-1.23	-1.15	-1.16	-1.15
1.1%	-1.2	-1.21	-1.11	-1.13	-1.49	-1.23
1.2%	-1.09	-1.09	-1.09	-1.2	-1.2	-1.11
1.3%	-1.1	-1.11	-1.37	-1.14	-1.08	-1.08
1.4%	-1.01	-1.04	-1.04	-1.06	-1.06	-1.06

We also checked to see if the Slutzky condition of equation (6.8) was satisfied. We did this by plugging in estimated parameters in that equation and evaluating at x_c and y_d values given by marginal net-of-tax rate, marginal virtual income, and gross labor income quartiles, resulting in 27 inequalities for each specification. We found that each of these inequalities is satisfied until the growth rate is 0.5%. The Slutzky conditions therefore hold in the specification with the highest cross validation values and nearby specifications.

An income effect of -1 is larger in magnitude than found or assumed in most taxable income studies. Using similar data but a linear model, Blomquist and Selin (2010) estimate a statistically significant negative income effect that is much closer to zero. Some studies just assume there is no income effect. It is important to account correctly for income effects in evaluating tax policy, because they have an impact on deadweight loss and other policy relevant variables. For example, by equation (2.10) of Auerbach (1985), the marginal deadweight loss in compensating variation terms is $(1-\theta)(y/\theta)e^*$ where e^* is the compensated elasticity of taxable income. The compensated elasticity satisfies $e^* = e - \theta * \partial y/\partial R$ where e is the uncompensated elasticity. Suppose that taxes were linear, the net of tax elasticity was .21, and the income effect -1, corresponding to our estimates. In our data the net of tax rate is .32 on average. Thus, a compensated elasticity corresponding to our estimates is $.53 = .21 - (.32) * (-1)$. This estimate is much larger than the uncompensated elasticity estimate of .21. It is also substantially larger than in Blomquist and Selin (2010), where a similar uncompensated elasticity but smaller income effect was estimated. Thus, allowing for general heterogeneity and accounting correctly for nonlinear taxes leads to a larger deadweight loss estimate for linear taxes through a larger income effect.

In Table 4 and 5, we investigate the influence of the productivity growth rate, covariates, instruments, and the regressor selection procedure on the estimates. We report net-of-tax elasticity (Table 4) and marginal income effect (Table 5) estimates from the highest cross-validation values corresponding to productivity growth rates .2, .3, and .4 respectively. We also report specifications using a first- and a second-order power series approximation of the budget set. We start with specifications without demographic control variables and without instrumenting nonlabor income using the control function approach. We then add demographic control variables and finally also instrument nonlabor income by including the control function.

Table 4. Net-of-tax elasticity estimates by method

	Lasso .2%	Lasso .3%	Lasso .4%	First-order	Second-order
No controls	3.23	2.83	2.63	2.19	1.73
	(0.09)	(0.09)	(0.09)	(0.07)	(0.08)
Demo	0.49	0.32	0.18	0.39	0.37
	(0.09)	(0.09)	(0.09)	(0.07)	(0.08)
Demo+IV	0.38	0.21	0.07	0.32	0.27
	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)

Table 5. Marginal income effects by method

	Lasso .2%	Lasso .3%	Lasso .4%	First-order	Second-order
No controls	0.04	0.12	0.04	0.04	0.19
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Demo	0.03	0.09	0.03	0.03	0.13
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Demo+IV	-0.9	-1.02	-1.07	-0.67	-0.92
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)

We observe that accounting for demographics has a drastic effect on the estimated net-of-tax elasticity. This could occur because of high correlation of the budget set with background variables. Furthermore, instrumentation is crucial for obtaining negative marginal income effects. This is consistent with a positive correlation between different likely endogenous components of nonlabor income, such as capital income and gross labor income. The specification using a first- or a second-order power series approximation produces elasticity and marginal effect estimates of the same magnitude as in the Lasso specifications.

We have also performed subsample estimation by splitting the sample by education level into a high- and a low-education subsample. In these estimations, we allow for different productivity growth in the two subsamples. Although estimated productivity growth differ slightly (higher for the high-education sample), the elasticities obtained are similar and not statistically significantly different from each other. We have experimented with alternative similar sample restrictions, definitions of labor income, and ways to handle the impact of the transfer system on the budget sets. The results are not sensitive to these issues.

9.3 Comparison with Linear Estimates

To assess the importance of appropriately accounting for a nonlinear budget set and productivity growth, we have estimated some different specifications based on linearizing the budget set at observed gross income levels. In these specifications, we regress gross labor income against the marginal net-of-tax rate and virtual income (the nonlabor income for the linearized budget set). To account for the endogeneity of the marginal net-of-tax rate to the budget set, we instrument it using either the net-of-tax rate of the first segment or the last segment. To account for endogeneity of virtual income we instrument it using the same instrument as in the control variable estimates given above. We also control for county fixed effects, include year fixed effects, and the same covariates as used in nonparametric regression estimates above. We estimate specifications that are either linear or logarithmic in the gross labor income, in the marginal net-of-tax rate, and in virtual income.

This type of specification is fairly common for taxable income estimation in repeated cross sections. The typical approach is, however, to use panel data like in Gruber and Saez (2002).

This would allow for linear individual effects by differencing and enable us to use instruments based on lagged income. Note though that the typical approach does not allow for other individual heterogeneity, such as in the net of tax elasticity, while our estimates above allow for general individual heterogeneity. The linear specifications we consider here are meant only to provide a comparison with our nonlinear estimates but are not meant to be best possible estimates based on the typical linearization method.

The estimates based on instrumental variables linearization specifications are reported in Table 6. Standard errors constructed using the delta method are reported in parenthesis. The elasticities are evaluated at the sample mean income of 420,329 SEK. We report the same type of net-of-tax elasticity as before, but now for the linearized budget sets. We observe that the estimated elasticity is sensitive to the instrument used and functional form.

Table 6. Estimated elasticities using instrumental variables linearization specifications

Instrument	Functional form	Net-of-tax elasticity	Income effect
First segment	Linear	-.69	-.92
		(1.1)	(.48)
First segment	Logarithmic	4.3	.44
		(.18)	(.061)
Last segment	Linear	1.3	.55
		(0.30)	(.28)
Last segment	Logarithmic	1.0	.075
		(0.10)	(.041)

In comparison with the nonparametric regression estimates using our method, we note that the estimated net-of-tax elasticity is much higher here while there are smaller income effects of opposite sign than previously. We find a statistically insignificant income effect when we use the log-log specification and instrument using the last segment tax rate. For our data, using our more sophisticated method therefore matters a lot. The overall picture is that correctly accounting for nonlinear budget sets and general heterogeneity are important in this data.

9.4 Reform Estimates

The expected value of taxable income is a useful tool for evaluating the effect of a tax reform on average taxable income. To illustrate this we use our estimated taxable income function to evaluate a tax program that was introduced in 2007 to 2008, which are the last two years of the sample period. In 2006, a center-right wing coalition government came into power in Sweden, and it launched a broader reform package to encourage labor supply among primarily low-income groups. The cornerstone of the package was an earned income tax credit (EITC) program that drastically lowered taxes on labor income. The credit was introduced in 2007 and reinforced in the subsequent three years. We investigate the effect of the 2008 version of the EITC on the labor supply of our sample individuals of married or cohabiting men using estimates from our preferred specifications.

The Swedish EITC program in 2008 is outlined in Table 7 for a representative individual. We observe that marginal tax rates were decreased quite a lot, by 16.40%, in a small low income region between 23,200 and 54,300 SEK, and slightly, by 2.19%, in a large medium income region between 148,900 and 431,600 SEK, where most married or cohabiting men are located. This construction creates positive substitution effects in these income ranges. However, it also creates income effects at most income levels. In particular, at higher income levels, there is only an income effect. For a sample of mostly full-time working individuals like ours, these income effects may be the most apparent effect on their labor supply. The Swedish EITC is different from the American and British equivalents by being universal, i.e., the same credit formula applies to everybody, and by not having a phase-out region where the credits received at lower income levels are phased out at higher income levels, creating stronger income effects.

Table 7. Swedish EITC program in 2008

Gross labor income	Net credits	Marginal tax rate change
0-23,200	0	0
23,200-54,300	(Income-23,200)*0.1640	-0.164
54,300-148,900	5,100	0
148,900-431,600	5,100+(Income-148,900)*0.0219	-0.0219
>431,600	11,300	0

The estimated relative effects on gross labor income and government revenues are reported in Table 8. For gross labor income, we report the mean difference in the individual predicted income levels given their budget sets with (post-reform) and without (pre-reform) the EITC program respectively relative to the predicted income levels without the program (pre-reform). For tax revenue effects, we make the tax revenue predictions at the predicted income levels before calculating the mean relative effect in the same way as for gross labor income. We also report the average relative mechanical tax revenue effect in the absence of behavioral effects at predicted income levels without the program (pre-reform). Standard errors constructed using the delta method are reported for the reform effect on income in parenthesis.

Table 8. Estimated EITC effect on labor income and tax revenues.

	Lasso 0.2%	Lasso 0.3%	Lasso 0.4%	First order	Second order
Gross income	3.8	2.86	1.96	0.69	1.29
	(0.45)	(0.44)	(0.43)	(0.39)	(0.4)
Tax revenues	0.28	-0.83	-1.89	-3.39	-2.66
Mechanical	-4.04	-4.08	-4.1	-4.14	-4.12

We observe that the reform effect estimates vary with productivity growth rate as for the elasticity estimates in the Lasso selected specifications. The estimated reform effect on gross income decreases when productivity growth rate increases, which is a similar pattern as for the net-of-tax elasticity. At the specification with largest cross-validation values, where the productivity growth rate is 0.3%, we obtain a statistically significant and positive reform effect on gross labor income of 2.86%. However, the reform decreases tax revenues by 0.83%. On the other hand, if there were no behavioral effects, tax revenues would have decreased by 4.08%. The behavioral effects therefore reduce the cost of the reform by 80%. The reform effect estimates on gross labor income are much smaller when using a first- or a second-order approximation of the budget set. This contrasts with the elasticity estimates where the Lasso specification produced similar results as linear and quadratic.

Two remarks are appropriate here. First, we include indirect taxation in our income and tax revenue measures. Revenues from these sources are larger than from direct labor taxes. Second, in nonlinear budget sets, tax revenues depend on the distribution of income under different tax regimes. In progressive tax systems, average revenues at predicted expected income levels are a downward biased estimate of expected revenues over the entire distribution. Because this bias may not be the same for the pre- and post-reform tax revenues, the reform effect tax revenue estimates may be biased.

10 Conclusion

In this paper we develop a method to nonparametrically estimate the expected value of taxable income as a function of a nonlinear budget set while allowing for general heterogeneity and optimization errors. We apply this approach to Swedish data and find a significant taxable income elasticity of .21 and an income effect of -1. This income effect is more substantial than found in most taxable income studies.

This method could be extended to estimate the expected value of taxes. As with taxable income, for concave budget frontiers the expected value of taxes will depend only on the CDF of taxable income for a linear budget set. A straightforward calculation can be carried out to derive the expected value of taxes, which could then be estimated by a procedure analogous that the described above for taxable income.

11 Appendix A: Proofs

The following two technical conditions are referred to in the text and used in the proofs.

Assumption A1: *η belongs to a complete, separable metric space and $y(\rho, R, \eta)$, $\partial y(\rho, R, \eta)/\partial \rho$, and $\partial y(\rho, R, \eta)/\partial R$ are continuous in (ρ, R, η).*

Assumption A2: *$\eta = (u, \varepsilon)$ for scalar ε and Assumption A1 is satisfied for $\eta = (u, \varepsilon)$ for a complete, separable metric space that is the product of a complete separable metric space for u with Euclidean space for ε, $y(\rho, R, \eta) = y(\rho, R, u, \varepsilon)$ is continuously differentiable in ε, there is $C > 0$ with $\partial y(\rho, R, u, \varepsilon)/\partial\varepsilon \geq 1/C$, $\|\partial y(\rho, R, \eta)/\partial(\rho, R)\| \leq C$ everywhere, ε is continuously distributed conditional on u, with conditional pdf $f_\varepsilon(\varepsilon|u)$ that is bounded and continuous in ε.*

Before proving Theorem 1 we state a result on the derivatives of $F(y|\rho, R)$ with respect to ρ and R.

Lemma A1: *If Assumptions 1 and A2 are satisfied then $y(\rho, R, \eta)$ is continuously distributed for each $\rho, R > 0$ and $F(y|\rho, R)$ is continuously differentiable in y, ρ, and R and for the pdf $f_{y(\rho,R,\eta)}(y)$ of $y(\rho, R, \eta)$ at y,*

$$\begin{aligned}
\frac{\partial F(y|\rho, R)}{\partial y} &= f_{y(\rho,R,\eta)}(y), \\
\frac{\partial F(y|\rho, R)}{\partial(\rho, R)} &= -f_{y(\rho,R,\eta)}(y)E[\frac{\partial y(\rho, R, \eta)}{\partial(\rho, R)}|y(\rho, R, \eta) = y].
\end{aligned}$$

Proof: This follows exactly as in the proof of Lemma A1 of Hausman and Newey (2014). *Q.E.D.*

Proof of Theorem 1: Note that $y(\rho, R, \eta)$ is differentiable by Assumption 1. Also, ρ behaves like the negative of a price (increasing ρ increases utility), so that the Slutzky condition for taxable income is

$$-\frac{\partial y(\rho, R, \eta)}{\partial\rho} + y(\rho, R, \eta)\frac{\partial y(\rho, R, \eta)}{\partial R} \leq 0.$$

By Assumption 1 and standard utility theory this inequality must be satisfied for all η and all $\rho, R > 0$. Then by Lemma A1 $F(y|\rho, R)$ is differentiable in ρ and R and

$$\begin{aligned}
&\frac{\partial F(y|\rho, R)}{\partial\rho} - y\frac{\partial F(y|\rho, R)}{\partial R} \\
&= -f_{y(\rho,R,\eta)}(y)\left\{E[\frac{\partial y(\rho, R, \eta)}{\partial\rho}|y(\rho, R, \eta) = y] - y \cdot E[\frac{\partial y(\rho, R, \eta)}{\partial R}|y(\rho, R, \eta) = y]\right\} \\
&= f_{y(\rho,R,\eta)}(y)E[-\frac{\partial y(\rho, R, \eta)}{\partial\rho} + y(\rho, R, \eta)\frac{\partial y(\rho, R, \eta)}{\partial R}|y(\rho, R, \eta) = y] \leq 0,
\end{aligned}$$

where the inequality follows by $f_{y(\rho,R,\eta)}(y) \geq 0$. This argument shows the first conclusion.

To show the second conclusion, for $0 < \tau < 1$ let $Q(\tau|\rho, R) = F^{-1}(\tau|\rho, R)$, which inverse function exists by $F(y|\rho, R)$ strictly increasing in y on (y_ℓ, y_u) and $[y_\ell, y_u]$ being the support of $F(y|\rho, R)$. By the inverse function theorem, for all $\rho, R > 0$,

$$\begin{aligned}
&-\frac{\partial Q(\tau|\rho, R)}{\partial\rho} + Q(\tau|\rho, R)\frac{\partial Q(\tau|\rho, R)}{\partial R} \\
&= f_{y(\rho,R,\eta)}(Q(\tau|\rho, R))^{-1}\left\{\frac{\partial F(Q(\tau|\rho, R)|\rho, R)}{\partial\rho} - Q(\tau|\rho, R)\frac{\partial F(Q(\tau|\rho, R)|\rho, R)}{\partial R}\right\} \leq 0.
\end{aligned}$$

Therefore it follows by Hurwicz and Uzawa (1971) that for each τ with $0 < \tau < 1$ there is a utility function $U(c, y, \tau)$ with for all $\rho, R > 0$,

$$Q(\tau|\rho, R) = \arg\max_{c,y} U(c, y, \tau) \quad s.t. \quad c = y\rho + R, c \geq 0, y \geq 0.$$

Let η be distributed uniformly on $(0, 1)$ and define

$$y(\rho, R, \eta) = Q(\eta|\rho, R).$$

Then

$$\Pr(Q(\eta|\rho, R) \leq y) = \Pr(\eta \leq F(y|\rho, R)) = F(y|\rho, R).$$

Thus, the RUM $U(c, y, \eta)$ has $F(y|\rho, R)$ as its CDF. *Q.E.D.*

LEMMA A2: *If Assumptions 1 and A1 are satisfied and $B(y)$ is concave then $y(B, \eta)$ is unique and $U(B(y), y, \eta)$ is strictly increasing to the left of $y(B, \eta)$ and strictly decreasing to the right of $y(B, \eta)$.*

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

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

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

LEMMA A3: *If Assumptions 1 and A1 are satisfied and the budget frontier B is concave then for each y, $\Pr(y(B, \eta) \leq y) = F(y|\rho(y), R(y))$.*

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

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

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

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

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

Note that $\dot{B}(z)$ is linear and hence concave so that by Lemma A2, $U(\dot{B}(z), z)$ is strictly increasing to the left of $\dot{y}^*$ and decreasing to the right of $\dot{y}^*$. Suppose that $y < y^*$. Then $y < \dot{y}^*$, because otherwise $\dot{y}^* \leq y < y^*$ and the above equation contradicts that $U(\dot{B}(z), z)$ is strictly decreasing to the right of $\dot{y}^*$. Similarly, if $y > y^*$ then $y > \dot{y}^*$, because otherwise $\dot{y}^* \geq y > y^*$ and the above equation contradicts that $U(\dot{B}(z), z)$ is strictly increasing to the left of $\dot{y}^*$.

Next suppose $y^* = y$. Let $\ddot{\rho}$ be the slope of a line that separates the set weakly preferred to $(B(y), y)$ and the budget set and let $\ddot{B}(z) = \dot{B} + \ddot{\rho}(z - y^*)$, so that $U(\dot{B}, y^*) \geq U(\ddot{B}(z), z)$ for all z. Then by Lemma A2 applied to the budget frontier $\ddot{B}(z)$, $U(\dot{B}, y^*) > U(\ddot{B}(z), z)$ for all $z \neq y^*$. Also, by Rockafellar (1970, pp. 214-215) $\ddot{\rho} \geq \dot{\rho}$. Then for any $z > y^*$ we have

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

so that

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

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

Summarizing, we have shown that

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

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

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

LEMMA A4: *If Assumptions 1 and A1 are satisfied then for all y such that there is $\Delta > 0$ with $\bar{B}(z) = B(z)$ for $z \in [y, y + \Delta]$ we have $\Pr(y(B, \eta) \leq y) = \Pr(y(\bar{B}, \eta) \leq y)$.*

Proof: Note that $\bar{B}(z) \geq B(z)$ for all z. For notational simplicity suppress the η argument and let $U(c, y) = U(c, y, \eta)$. Let

$$y^* \overset{def}{=} \text{argmax}_z U(B(z), z), \bar{y}^* \overset{def}{=} \text{argmax}_z U\left(\bar{B}(z), z\right).$$

Suppose first that $y^* = y$. Then for any $z \in (y, y+\Delta]$,

$$U(\bar{B}(y), y) = U(B(y), y) \geq U(B(z), z) = U(\bar{B}(z), z).$$

By Lemma A2 we cannot have $\bar{y}^* > y$ because then the above inequality is not consistent with $U(\bar{B}(z), z)$ being strictly monotonically increasing to the left of $\bar{y}^*$. Therefore $\bar{y}^* \leq y$. Suppose next that $y^* < y$. Then

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

Then by similar reasoning as before $\bar{y}^* \leq y$. Thus, we have shown that

$$y^* \leq y \Longrightarrow \bar{y}^* \leq y.$$

Next, suppose that $y^* > y$. Then there is $z \in (y, y+\Delta]$ with $y < z < y^*$

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

Again by Lemma A2 we cannot have $\bar{y}^* \leq y$ because then $\bar{y}^* < z < y^*$ and the above inequality is not consistent with $U(\bar{B}(z), z)$ being strictly monotonic decreasing to the right of $\bar{y}^*$. Therefore, $\bar{y}^* > y$, and we have shown that

$$y^* > y \Longrightarrow \bar{y}^* > y.$$

Therefore we have $y^* \leq y \Longleftrightarrow \bar{y}^* \leq y$, so the conclusion follows similarly to the conclusion of Lemma A2. *Q.E.D.*

Proof of Theorem 2: Combining the conclusions of Lemmas A3 and A4 gives

$$\Pr(y(B, \eta) \leq y) = \Pr(y(\bar{B}, \eta) \leq y) = F(y|\rho(y), R(y)). \ Q.E.D.$$

Proof of Theorem 3: For notational simplicity we suppress the η as before. Let $\hat{y} = \underline{y}(B), \check{y} = \bar{y}(B)$ denote the lower and upper bound respectively for the set where $B(z)$ is not concave. Consider

$$\tilde{y}^* = \arg\max_z U(B(z), z) \text{ s.t. } \hat{y} \leq z \leq \check{y}.$$

By construction the CDF of $\tilde{y}$ depends only on $\hat{y}$, $\check{y}$, and $B(z)$ for $z \in [\hat{y}, \check{y}]$. We will show that for $y \in [\hat{y}, \check{y})$, the CDF of $y^* = \arg\max U(B(z), z)$ coincides with that of $\tilde{y}^*$, which will then prove the result. For the CDF's to coincide it is sufficient to show that for $y \in [\hat{y}, \check{y})$, $y^* \leq y$ if and only if $\tilde{y}^* \leq y$. Consider $y \in [\hat{y}, \check{y})$.

Suppose first that $y^* \leq y$. If $y^* \geq \hat{y}$ then y^* is within the constraint set $[\hat{y}, \check{y}]$ so that $y^* = \tilde{y}^*$. If $y^* < \hat{y}$, then by Lemma A2, for all $z \in (\hat{y}, \check{y}]$

$$U(B(y^*), y^*) > U(B(\hat{y}), \hat{y}) > U(\bar{B}(z), z) \geq U(B(z), z),$$

so that $\tilde{y}^* = \hat{y} \leq y$. Therefore, $y^* \leq y \Longrightarrow \tilde{y}^* \leq y$.

Next, suppose that $y^* > y$. If $y^* \leq \check{y}$ then $\tilde{y}^* = y^*$ so that $\tilde{y}^* > y$. Suppose $y^* > \check{y}$. Then by Lemma A1 and $y < \check{y}$, for all $z \in [\hat{y}, \check{y})$ we have

$$U(B(y^*), y^*) > U(B(\check{y}), \check{y}) > U(\bar{B}(z), z) \geq U(B(z), z),$$

so that $\tilde{y}^* = \check{y} > y$. Therefore $y^* > y \Longrightarrow \tilde{y}^* > y$. Summarizing, for $y \in [\hat{y}, \check{y})$ we have $y^* \leq y \Longleftrightarrow \tilde{y}^* \leq y$, which proves the result. *Q.E.D.*

Proof of Theorem 4: Let $F(y)$ be the CDF of $y(B, \eta)$. By standard probability theory, $\Pi_{\bar{\ell}} = F(\bar{\ell}) - \lim_{y \uparrow \bar{\ell}} F(y)$. By Theorem 2 $F(\bar{\ell}) = F(\bar{\ell}|\rho_+, R_+)$ and $\lim_{y \uparrow \bar{\ell}} F(y) = \lim_{y \uparrow \bar{\ell}} F(y|\rho_-, R_-)$. Furthermore, by Assumption A2 $y(\rho_-, R_-, \eta)$ is continuously distributed so that $\lim_{y \uparrow \bar{\ell}} F(y|\rho_-, R_-) = F(\bar{\ell}|\rho_-, R_-)$. Define $\Lambda(\rho) = F(\bar{\ell}|\rho, R(\rho))$ for $\rho \in [\rho_+, \rho_-]$. We then have

$$\Pi_{\bar{\ell}} = \Lambda(\rho_+) - \Lambda(\rho_-).$$

By the chain rule, $R(\rho) = R_- + \bar{\ell}(\rho_- - \rho)$, and Lemma A1 of Hausman and Newey (2014), $\Lambda(\rho)$ is differentiable and

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

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

Proof of Theorem 5: Let $F_j(y) = F(y|\rho_j, R_j)$. By Theorem 2, the CDF of $y(B, \eta)$ on (ℓ_{j-1}, ℓ_j) is $F_j(y)$. Therefore,

$$\mu(B) = \sum_{j=1}^{J-1} \left[\int 1(\ell_{j-1} < y < \ell_j) y F_j(dy) + \ell_j \Pr(Y(B, \eta) = \ell_j) \right] + \int 1(\ell_{J-1} < y) y F_J(dy).$$

Note that

$$\int 1(\ell_{J-1} < y) y F_J(dy) = \bar{y}(\rho_J, R_J) - \int 1(y \leq \ell_{J-1}) y F_J(dy).$$

In addition, by $\ell_0 = 0$ we have $\int 1(y \leq \ell_0) y F_1(dy) = 0$, so that

$$\begin{aligned}
\sum_{j=1}^{J-1} \int 1(\ell_{j-1} &< y < \ell_j) y F_j(dy) + \int 1(\ell_{J-1} < y) y F_J(dy) \\
&= \sum_{j=1}^{J-1} \int [1(y < \ell_j) - 1(y \leq \ell_{j-1})] y F_j(dy) + \int 1(\ell_{J-1} < y) y F_J(dy) \\
&= \bar{y}(\rho_J, R_J) + \sum_{j=1}^{J-1} \left[\int 1(y < \ell_j) y F_j(dy) - \int 1(y \leq \ell_j) y F_{j+1}(dy) \right],
\end{aligned}$$

Also, it follows from Theorem 2, similarly to the proof of Theorem 4, that

$$\begin{aligned}\Pr(y(B,\eta) &= \ell_j) = F_{j+1}(\ell_j) - \lim_{y\uparrow\ell_j} F_j(y) = \int 1(y \le \ell_j)F_{j+1}(dy) - \int 1(y < \ell_j)F_j(dy) \\ &= \int 1(y \ge \ell_j)F_j(dy) - \int 1(y > \ell_j)F_{j+1}(dy).\end{aligned} \tag{11.11}$$

Combining these results we have

$$\mu(B) = \bar{y}(\rho_J, R_J) + \sum_{j=1}^{J-1}[\int 1(y < \ell_j)(y-\ell_j)F_j(dy) - \int 1(y \le \ell_j)(y-\ell_j)F_{j+1}(dy)].$$

Noting that $\int 1(y \le \ell_j)(y-\ell_j)F_{j+1}(dy) = \int 1(y < \ell_j)(y-\ell_j)F_{j+1}(dy)$ then gives the first conclusion.

To show the second conclusion, note that by $\ell_0 = 0$ we have $\bar{y}(\rho_1, R_1) = \int 1(y > \ell_0)yF_1(dy)$. Then it follows that

$$\begin{aligned}\sum_{j=1}^{J-1}\int 1(\ell_{j-1} &< y < \ell_j)yF_j(dy) + \int 1(\ell_{J-1} < y)yF_J(dy) \\ &= \sum_{j=1}^{J-1}\int \{1(y > \ell_{j-1}) - 1(y \ge \ell_j)\}yF_j(dy) + \int 1(y > \ell_{J-1})yF_J(dy) \\ &= \bar{y}(\rho_1, R_1) + \sum_{j=1}^{J-1}\left[\int 1(y > \ell_j)yF_{j+1}(dy) - \int 1(y \ge \ell_j)yF_j(dy)\right].\end{aligned}$$

Combining this with the second equality in eq. (11.11) then gives

$$\mu(B) = \bar{y}(\rho_1, R_1) + \sum_{j=1}^{J-1}\left[\int 1(y > \ell_j)(y-\ell_j)F_{j+1}(dy) - \int 1(y \ge \ell_j)(y-\ell_j)F_j(dy)\right].$$

Noting that$\int 1(y \ge \ell_j)(y-\ell_j)F_j(dy) = \int 1(y > \ell_j)(y-\ell_j)F_j(dy)$ then gives the second conclusion. *Q.E.D.*

Proof of Theorem 6: By Edmunds and Evans (1989) there exists C such that for each A and B there is (β_{ab}) such that for $\varepsilon = CA^{1-s}B^{(1-s)/2}$ and $p^{AB}(y,x) = \sum_{a=1}^{A}\sum_{b=1}^{B}\beta_{ab}f_a(y)r_b(x)$,

$$\sup_{\mathcal{Z}}\left|f(y|x) - p^{AB}(y,x)\right| + \sup_{\mathcal{Z}}\left|f_\rho(y|x) - p_\rho^{AB}(y,x)\right| + \sup_{\mathcal{Z}}\left|f_R(y|x) - p_R^{AB}(y,x)\right| \le \varepsilon,$$

where the subscripts denote partial derivatives and $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$. Let

$$\tilde{p}^{AB}(y,x) = f_1(y) + \sum_{a=2}^{A} w_a(x,\beta)[f_a(y) - f_1(y)] = p^{AB}(y,x) + [1 - \sum_{a=1}^{A} w_a(x,\beta)]f_1(y).$$

Note that by $\int f_a(y)dy = 1$ for each a and by $\mathcal{Y}$ bounded,

$$\sup_{\mathcal{X}}\left|1-\sum_{a=1}^{A}w_a(x,\beta)\right| = \sup_{\mathcal{X}}\left|\int[f(y|x)-p^{AB}(y,x)]dy\right| \le \sup_{\mathcal{X}}\int\left|f(y|x)-p^{AB}(y,x)\right|dy \le C\varepsilon.$$

Also,

$$\sup_{\mathcal{X}}\left|\frac{\partial}{\partial\rho}\sum_{a=1}^{A}w_a(x,\beta)\right| = \sup_{\mathcal{X}}\left|\frac{\partial}{\partial\rho}\int[f(y|x)-p^{AB}(y,x)]dy\right| \le \sup_{\mathcal{X}}\int\left|f_\rho(y|x)-p_\rho^{AB}(y,x)\right|dy \le C\varepsilon,$$

and $\sup_{\mathcal{X}}\left|\frac{\partial}{\partial R}\sum_{a=1}^{A}w_a(x,\beta)\right| \le C\varepsilon$ similarly. Recall that $\bar{y}(x) = \int yf(y|x)dy$ and $\nu(x,\ell) = \int 1(y<\ell)(y-\ell)f(y|x)dy$. Define

$$\bar{y}^{AB}(x) = \int y\cdot\tilde{p}^{AB}(y,x)dy, \nu^{AB}(x,\ell) = \int 1(y<\ell)(y-\ell)\tilde{p}^{AB}(y,x)dy.$$

Note that for any pdf $f_1(y)$, $\int yf_1(y)dy \le \sup_{\mathcal{Y}} y = C$, so for all $x\in\mathcal{X}$,

$$\begin{aligned}
\left|\bar{y}(x)-\bar{y}^{AB}(x)\right| &= \left|\int y[f(y|x)-p^{AB}(y,x)]dy + [1-\sum_{a=1}^{A}w_a(x,\beta)]\int yf_1(y)dy\right| \\
&\le \int y\left|f(y|x)-p^{AB}(y,x)\right|dy + \left|1-\sum_{a=1}^{A}w_a(x,\beta)\right|\int yf_1(y)dy \le C\varepsilon.
\end{aligned}$$

Let $\Delta(x,\ell) = \nu(x,\ell)-\nu^{AB}(x,\ell)$. Note that by $\mathcal{Y}$ bounded there is C such that for all $\ell\in\mathcal{Y}$ and any pdf $f_1(y)$ with support contained in $\mathcal{Y}$, $\left|\int 1(y<\ell)(y-\ell)f_1(y)dy\right| \le C$. Therefore,

$$\begin{aligned}
&\left|\frac{\partial}{\partial\rho}\Delta(x,\ell)\right| \\
=&\left|\frac{\partial}{\partial\rho}\int 1(y<\ell)(y-\ell)[f(y|x)-p^{AB}(y,x)]dy + \frac{\partial}{\partial\rho}\left[1-\sum_{a=1}^{A}w_a(x,\beta)\right]\int 1(y<\ell)(y-\ell)f_1(y)dy\right| \\
\le& C\varepsilon
\end{aligned}$$

Therefore we have

$$\begin{aligned}
&\left|\sum_{j=1}^{J-1}[\nu(x_j,\ell_j)-\nu(x_{j+1},\ell_j)] - \sum_{j=1}^{J-1}[\nu^{AB}(x_j,\ell_j)-\nu^{AB}(x_{j+1},\ell_j)]\right| \\
=& \left|\sum_{j=1}^{J-1}[\Delta(x_j,\ell_j)-\Delta(x_{j+1},\ell_j)]\right| = \left|\sum_{j=1}^{J-1}[\frac{\partial\Delta(\bar{x}_j,\ell_j)}{\partial x}]^T(x_j-x_{j+1})]\right| \\
\le& C\left(\sup_{\mathcal{X}}\left|\frac{\partial\Delta(\bar{x}_j,\ell_j)}{\partial\rho}\right| + \sup_{\mathcal{X}}\left|\frac{\partial\Delta(\bar{x}_j,\ell_j)}{\partial R}\right|\right)\sum_{j=1}^{J-1}[\rho_j-\rho_{j+1}+R_{j+1}-R_j] \\
\le& C\varepsilon[\rho_1-\rho_J+R_J-R_1] \le C\varepsilon.
\end{aligned}$$

To conclude the proof, note that

$$\begin{aligned}\sum_{j=1}^{J-1}[\nu^{AB}(x_j,\ell_j)-\nu^{AB}(x_{j+1},\ell_j)] &= \sum_{j=1}^{J-1}[\int 1(y<\ell_j)(y-\ell_j)[p^{AB}(y,x_j)-p^{AB}(y,x_{j+1})]dy \\ &= \sum_{j=1}^{J-1}[\int 1(y<\ell_j)(y-\ell_j)\sum_{a=2}^{A}[w_a(x_j,\beta)-w_a(x_{j+1},\beta)][f_a(y)-f_1(y)]dy \\ &= \sum_{a=2}^{A}\sum_{j=1}^{J-1}[w_a(x_j,\beta)-w_a(x_{j+1},\beta)][\nu_a(\ell_j)-\nu_1(\ell_j)] \\ &= \sum_{a=2}^{A}\sum_{b=1}^{B}\beta_{ab}\sum_{j=1}^{J}[r_b(x_j)-r_b(x_{j+1})][\nu_a(\ell_j)-\nu_1(\ell_j)].\end{aligned}$$

From these two equations we see that the expression in the statement of the theorem is

$$\mu(B)-\bar{y}^{AB}(x_J)-\sum_{j=1}^{J-1}[\nu^{AB}(x_j,\ell_j)-\nu^{AB}(x_{j+1},\ell_j)].$$

The conclusion then follows by the triangle inequality, Theorem 5, and the above bonds on $\left|\bar{y}(x)-\bar{y}^{AB}(x)\right|$ and $\left|\sum_{j=1}^{J-1}[\nu(x_j,\ell_j)-\nu(x_{j+1},\ell_j)]-\sum_{j=1}^{J-1}[\nu^{AB}(x_j,\ell_j)-\nu^{AB}(x_{j+1},\ell_j)]\right|$. *Q.E.D.*

12 Appendix B: Sample Statistics Sweden

Table B1. Sample statistics

Variable	All years	1993	1998	2003	2008
Number of observations	78,268	5,003	4,244	5,403	4,915
Gross labor income	4.18	3.79	4.11	4.23	4.62
1-st net-of-tax rate	0.52	0.54	0.50	0.55	0.53
1-st virtual income	1.86	1.76	1.77	1.73	1.69
Marginal net-of tax-rate	0.32	0.32	0.29	0.33	0.33
Marginal virtual income	2.17	2.07	2.08	2.03	2.09
Last net-of-tax rate	0.25	0.29	0.25	0.25	0.24
Last virtual income	2.42	2.15	2.20	2.38	2.52
Age	43.80	42.26	43.77	44.23	44.23
Dummy children $<$ 6 years	0.35	0.44	0.36	0.30	0.33
Dummy foreign born	0.13	0.10	0.12	0.14	0.17
Wife's net labor income	1.45	1.25	1.28	1.54	1.86

Notes: Gross labor income, wife's net labor income, and virtual incomes are expressed in 100,000 SEK.

13 References

Aaberge, R. and L. Flood (2008), "Evaluation of an In-work Tax Credit Reform in Sweden: Effects on Labor Supply and Welfare Participation of Single Mothers," IZA Discussion Paper

3736.

Auerbach, A. (1985), "The Theory of Excess Burden and Optimal Taxation," in A. Auerbach and M. Feldstein, eds., *Handbook of Public Economics Vol. 1*, 61–128.

Auten, G. and R. Carroll (1999), "The Effect of Taxes on Household Income," *Review of Economics and Statistics 81*: 681–693.

Belloni, A. and V. Chernozhukov (2013), "Least Squares After Model Selection in High-Dimensional Sparse Models," *Bernoulli* , 521-547.

Blomquist, S. and W. Newey (2002), "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica 70(6)*: 2455–2480.

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.

Blundell, R.W. and R. Matzkin (2014), "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics* 5, 271–295.

Blundell, R.W. and J.L. Powell (2006), "Endogeneity in Nonparametric and Semiparametric Regression Models," in Dewatripont, M., L.P. Hansen, and S.J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II,* Cambridge University Press.

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.

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

Carroll, R. (1998), "Do Taxpayers Really Respond to Changes in Tax Rates? Evidence from the 1993 Tax Act," Office of Tax Analysis Working Paper No. 78., Washington, D.C.: U.S. Department of Treasury.

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

Dette, H., S. Hoderlein, and N. Nuemeyer (2011), "Testing Multivariate Economic Restrictions Using Quantiles: The Example of Slutsky Negative Semidefiniteness," working paper, Boston College.

Edmunds, D. E., and W. Evans (1989), "Entropy Numbers and Approximation Numbers in Function Spaces," *Proceedings of the London Mathematical Society* 58, 137–152.

Eissa, N. and S. Giertz (2006), "Trends in High Incomes and Behavioural Responses to Taxation: Evidence from Executive Compensation and Statistics of Income Data," Congressional Budget Office Working Paper No. 2006-14.

Ericson, P., R. Wahlberg, and L. Flood (2009), "SWEtaxben: A Swedish Tax/Benefit Micro

Simulation Model and an Evaluation of a Swedish Tax Reform," IZA Discussion Paper 4106.

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.

Feldstein, M. (1999), "Tax Avoidance and the Deadweight Loss of the Income Tax," *Review of Economics and Statistics 81(4)*: 674–680.

Feldstein, M. and D. Feenberg, D. (1995), "The Effects of Increased Tax Rates on Taxable Income and Economic Efficiency: A Preliminary Analysis of the 1993 Tax Rate Increases," NBER Working Paper No. W5370.

Flood, L., R. Wahlberg, and E. Pylkkänen, (2007), "From Welfare to Work: Evaluating a Tax and Benefit Reform Targeted at Single Mothers in Sweden," *Labour 21*: 443–471.

Giertz, S. (2007), "The Elasticity of Taxable Income Over the 1980s and 1990s," *National Tax Journal 60(4)*: 743-768.

Goolsbee, A. (1999), "Evidence on the High-Income Laffer Curve from Six Decades of Tax Reform," *Brookings Papers on Economic Activity, No. 2* (1999), 1–64.

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

Hall, R.E. (1973), "Wages, Income, and Hours of Work in the U.S. Labor Force," In G. Cain and H. Watts, eds., *Income Maintenance and Labor Supply* 102–62, Chicago: Markham.

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

Hausman, J.A. and W.K. Newey (2014), "Individual Heterogeneity and Average Welfare", working paper.

Hurwicz, L. and H. Uzawa (1971), "On the Integrability of Demand Functions," in Chipman, L. Hurwicz, M. K. Richter, and H. F. Sonnenschein, eds. 1971. *Preferences, Utility, and Demand: A Minnesota Symposium.* New York: Harcourt, Brace, Jovanovich, 114–148.

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

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

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

Kopczuk, W. (2005), "Tax Bases, Tax Rates and the Elasticity of Taxable Income," *Journal of Public Economics 89(11-12)*: 2093–2119.

Lindsey, L. (1987), "Individual Taxpayer Response to Tax Cuts: 1982–1984, with Implications for the Revenue Maximizing Tax Rate," *Journal of Public Economics 33*: 173–206.

Looney, A. and M. Singhal (2006), "The Effect of Anticipated Tax Changes on Intertemporal Labor Supply and the Realization of Taxable Income," NBER Working Paper 12417, National Bureau of Economic Research.

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

McFadden, D. (2005), "Revealed Stochastic Preference: A Synthesis," *Economic Theory* 26, 245-264.

McFadden, D. and K. Richter (1991), "Stochastic Rationality and Revealed Stochastic Preference," in J. Chipman, D. McFadden, K. Richter (eds) *Preferences, Uncertainty, and Rationality*, Westview Press, 161-186.

Moffitt, R. and M. Wilhelm (2000), "Taxation and the Labor Supply Decisions of the Affluent" in *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, J. Slemrod (ed.), New York: Cambridge University Press.

Navratil, J. (1995), "The Economic Recovery Tax Act of 1981: Evidence on Individual Taxpayer Behavior from Panel Tax Return Data," PhD thesis, Department of Economics, Harvard University.

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

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

Saez, E. (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.

Sammartino, F. and D. Weiner (1997), "Recent Evidence on Taxpayers' Response to the Rate Increases in the 1990's," *National Tax Journal 50(3)*: 683-705.

Sillamaa, M. and M. Veall (2000), "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1988 Tax Flattening in Canada," QSEP Research Report, No. 354.

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.