Identifying the Effect of Taxes on Taxable Income*

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

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

Anil Kumar
Federal Reserve Bank of Dallas

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

March 2018

Abstract

Kinks and notches can identify whether or not taxes affect behavior. When preferences are continuously distributed then bunching at kinks and notches provides evidence of a positive taxable income elasticity. Unfortunately, the size of a kink is not informative about the size of the elasticity when the distribution of heterogeneity is unrestricted. Kinks do provide information about the size of the elasticity when a priori restrictions are placed on the heterogeneity distribution. They can identify the elasticity when the functional form of the heterogeneity distribution is specified across the kink and provide bounds under nonparametric restrictions on the heterogeneity. We also show that variation in budget sets can identify the taxable income elasticity when the distribution of preferences is unrestricted. With optimization errors identification becomes more difficult. We use variation in budget sets to nonparametrically identify tax effects in the presence of optimization errors via the conditional mean of taxable income. We apply this approach to estimation of tax effects using a Swedish repeated cross-section data set.

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

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

---

*The NSF provided partial financial support. We are grateful for comments by R. Blundell, A. Finklestein, J. Hausman, H. Kleven, C. Manski, R. Matzkin, J. Poterba, E. Saez, H. Selin and participants at seminars at UCL in 2010, BC, Chicago, Georgetown, Harvard/MIT, Michigan, NYU, UC Irvine, UCLA, USC, Yale and the bunching conference at UCSD in March of 2018. This paper combines two previously unpublished working papers, Blomquist and Newey (2017) and Blomquist et al. (2015).
1 Introduction

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

Kinks and notches can identify whether taxes affect behavior. When preferences are continuously distributed bunching at kinks and notches provides evidence of a positive taxable income elasticity. We show that the size of a kink is not informative about the size of the elasticity when the distribution of heterogeneity is unrestricted. Kinks do provide information about the size of the elasticity when a priori restrictions are placed on the heterogeneity distribution. Kinks can identify the elasticity when the functional form of the heterogeneity distribution is specified across the kink and provide bounds under nonparametric restrictions on the heterogeneity.

We show that variation in budget sets can identify the taxable income elasticity when the distribution of preferences is unrestricted. With optimization errors identification becomes more difficult. We use variation in budget sets to identify taxable income effects via the conditional mean of taxable income while accounting for all the restrictions of utility maximization. This method identifies tax effects with general preferences while allowing for optimization errors.

Bunching estimators of the taxable income elasticity were developed and extended in influential work by Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013). Unfortunately, bunching estimators do not identify the size of the taxable income elasticity when the heterogeneity distribution is unrestricted. The problem is that a kink or notch probability may be large or small because of the size of the elasticity or because more or fewer individuals like to have taxable income around the kink or notch. Intuitively, for a single budget set, variation in the tax rate only occurs with variation in preferences. The conjoining of variation in the tax rate and preferences makes it impossible to nonparametrically distinguish the taxable income elasticity from heterogeneity with a single budget set.

A kink or notch probability is just one reduced form parameter and so can identify just one structural parameter. Thus, everything about heterogeneity must come from somewhere else in order to get the elasticity from bunching. Saez (2010) combines density values at the edges of the bunching interval with assuming that the density is linear across the kink to identify the elasticity.

\footnote{Bastani and Selin (2014), Gelber et al. (2017), Marx (2012), Le Maire and Schijering (2013) and Seim (2015) are a few of the recent papers that apply the bunching method.}
elasticity. Chetty et al. (2011) assume that the density is a polynomial near and across the kink. These results impose a known heterogeneity distribution across the kink which seems an unusually strong functional form assumption for identifying an important structural parameter.

Bunching may provide interval information about the size of the elasticity under nonparametric, a priori restrictions on the heterogeneity distribution. We give bounds on the elasticity under prior restrictions on the heterogeneity density, including monotonicity or known bounds on the density. Of course all such elasticity bounds are sensitive to the a priori assumptions one makes about the heterogeneity density.

Budget set variation can identify taxable income effects with unrestricted heterogeneity. For an isoelastic model we find that the elasticity is identified from two convex budget sets if preferences have the same distribution for the two budget sets (i.e. budget set variation is independent of preferences) and the marginal tax rate differs at the chosen taxable income for at least one individual. This is an intuitive condition for identification, that the marginal tax rate varies for some individual. We also find that kinks alone may not be informative when budget sets vary.

In data it is often observed that there is little or no bunching at kinks. This feature of the data has been accounted for by allowing departures from utility maximization, referred to as optimization errors. Hausman’s (1981) specification included an additive disturbance to account for the lack of bunching. Saez (2010) considered a bunching window that included the kink in its interior. Optimization errors make identification more difficult because there are more things to identify from the same data. We give examples showing that optimization errors can have large effects on bunching estimators. Cattaneo et al. (2018) give results on identification of tax effects from bunching when there are optimization errors and the heterogeneity density has a known functional form across the kink.

The expected value of taxable income can be used to nonparametrically identify taxable income effects when there are additive or multiplicative optimization errors. We show how to identify the expected value of taxable income conditional on nonlinear budget sets with nonparametrically heterogenous preferences that are strictly convex and statistically independent of the budget sets. We use utility maximization to impose restrictions that make nonparametric estimation feasible. We demonstrate how to check all of the restrictions of utility maximization on the conditional mean. Our specification is based on convex budget sets but allows for nonconvexities.

We show how to estimate the expected value of taxable income from a repeated cross section. We account for possible productivity growth, estimating that from the data. We allow for endogeneity of unearned income through a control function and allow for covariates. To
evaluate the effect of taxes on taxable income we focus on changes in nonlinear tax systems. We consider effects defined by an upward shift of the nonlinear budget constraint, in either slope or intercept. We find that these effects can be estimated with a high degree of accuracy in our application.

We give an application to Swedish data from 1993-2008 with third party reported taxable labor income. This means that the variation in the taxable income in our data is mainly driven by variation in effort broadly defined and by hours of work, and not by variation in tax evasion.\footnote{Kleven et al. (2011) find that the tax evasion rate is close to zero for income subject to third-party reporting.} We estimate a statistically significant uncompensated tax elasticity of 0.21 and a significant income effect of -1, implying a compensated elasticity of 0.53. The income effect is larger than in many taxable income studies but has similar size to several estimates in the labor supply literature.

The rest of this paper is organized as follows. In the remainder of this Section we give a brief literature review. Section 2 lays out the model of individual behavior we consider, where we focus primarily on two polar specifications; isoelastic utility with scalar heterogeneity or nonparametric utility with general heterogeneity. Section 3 discusses nonidentification from a single budget set. Section 4 gives partial identification results, i.e. bounds, for a single budget set. Section 5 shows how variation in budget sets helps identify the taxable income elasticity. Section 6 considers the role of optimization errors and gives numerical results showing their impact on bunching estimators. Section 7 gives the expected value of taxable income and shows how to check all the restrictions of utility maximization. Section 8 gives the policy effects we consider and explains how they can be estimated when there is productivity growth. Section 9 outlines how to estimate the expected value of taxable income in practice. Section 10 gives the empirical results and Section 11 offers some conclusions.

Previous work has mostly not focused on the lack of identification of the taxable income elasticity from kinks. An exception is Blomquist et al. (2015), where nonparametric identification of an average compensated tax effect from a kink was considered. That paper showed that the kink provides no information about that average tax effect, but that the effect is identified when the taxable income density is linear across the kink as a function of the tax rate, and gave bounds under monotonicity in the tax rate of the taxable income density. Those results are now incorporated in this paper. Our partial identification results for the isoelastic utility function are analogous, showing kinks do not provide any information about the size of the elasticity, that the elasticity is identified when the heterogeneity density is linear, and giving bounds under monotonicity. Einav et al. (2017) provided recent empirical evidence on the sensitivity of
policy effects to kink modeling assumptions for the elderly in Medicare Part D, where there is substantial bunching around the famous “donut hole.” In work that first appeared following Blomquist and Newey (2017, 2018), Bertanha, McCallum, and Seegert (2018) also give some nonidentification and bounds results.

Nonparametric models are considered in Blomquist and Newey (2002), Blomquist et al. (2015), Manski (2014), and Kline and Tartari (2016). Blomquist and Newey (2002) nonparametrically identify and estimate important policy effects under scalar heterogeneity and optimization errors. Optimization errors are not allowed for in Manski (2014) or Kline and Tartari (2016). Blomquist et al. (2015) show that the results of Blomquist and Newey (2002) are valid with general heterogeneity and demonstrate how to check or use all the restrictions on expected taxable income implied by utility maximization. Manski (2014) and Kline and Tartari (2016) give bounds on effects.

Van Soest (1995), Keane and Moffitt (1998), Blundell and Shephard (2012), and Manski (2014) have considered labor supply when hours are restricted to a finite set. The expected value of taxable income could accommodate such constraints, though we do not do this for simplicity. It appears to be harder to incorporate the bilateral contracting framework of Blundell and Shephard (2012).

Many estimates of compensated taxable income elasticities exist in the literature. For purposes of comparison with our compensated elasticity estimate of .53 we briefly survey the existing results. 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. 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.

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. 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.

The most recent work has also found a range of elasticities. 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. For the same data set, Kumar and Liang (2017) estimate a weighted-average of heterogeneous elasticities of approximately 0.7. Burns and Ziliak (2017) use matched panels from the Current Population Survey along with grouping instrumental variables to estimate an elasticity in the range of 0.4-0.55.
2 Modeling Taxable Income

We consider individuals with preferences defined over after-tax income $c$ (value of consumption) and before tax income $y$ (cost of effort). After and before tax income are related by

$$c = B(y),$$

where $B(y) = y - T(y)$ is after tax income for taxes $T(y)$. The utility function of an individual will be

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

where $\eta$ is a possibly multi-dimensional vector representing individual preferences. We will assume throughout that for each $\eta$ the utility function $U(c, y, \eta)$ is increasing in $c$, decreasing in $y$, and strictly quasi-concave. Strict quasi-concavity is equivalent to the taxable income choice being unique for any nonlabor income and linear tax rate, as is generally assumed in the literature. For an individual with preferences $\eta$ we denote the choice of $y$ that maximizes utility $U(B(y), y, \eta)$ by

$$y(B, \eta),$$

where we assume that the maximizing value generally exists and is unique. Here we allow the choice $y(B, \eta)$ to depend on the whole after tax function $B$ as it may.

Figure 1 illustrates a budget set that has two linear segments with slopes (net of tax rates) $\rho_1 > \rho_2$ and a kink at $K$. An individual with preferences $\eta$ will choose the point on the budget set where their utility is highest. Different individuals may have different utility functions and so choose different taxable incomes. The distribution of taxable income along the budget set comes from variation in preferences. Heterogeneity of preferences is necessary in order to have a distribution of taxable income along a single budget set. If preferences were homogenous we would have one point on a single budget constraint; no inference about preferences could be drawn from that.

There is a simple relationship between the taxable income elasticity and the curvature of the indifference curve. Consider an indifference curve defined by $U(c, y, \eta) = u$ for fixed $\eta$ and define the function $c = h(y, \eta, u)$. Let $h'(y, \eta, u) = \partial h/\partial y$ and $h''(y, \eta, u) = \partial^2 h/\partial y^2$. We note that $h'$ is the slope of the indifference curve with utility level $u$ and $h''$ the curvature of the indifference curve. One can show that if utility is maximized subject to a linear budget constraint with slope $\rho$, then the compensated effect is given by $\partial y/\partial \rho = 1/h''$. The less curved an indifference curve is (small $h''$), the larger the $\partial y/\partial \rho$ and the taxable income elasticity are.
An important special case is the isoelastic utility function considered by Saez (2010),

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

where \( \eta \) is a scalar. Maximizing this utility function subject to a linear budget constraint \( B(y) = \rho y + R \) with slope (net of tax rate) \( \rho \) and intercept (nonlabor income) \( R \) gives the taxable income function \( y(\rho, \eta) = \rho^\beta \eta \). The taxable income elasticity \( \partial \ln y(\rho, \eta)/\partial \ln \rho = \beta \) is constant for this specification and there is no income effect of changing \( R \). The variable \( \eta \) represents unobserved individual heterogeneity in preferences. We note that \( y(\rho, \eta) \) is decreasing in \( \rho \) (by \( \rho < 1 \)) and increasing in \( \eta \) and \( \rho \). We will also consider cases where the elasticity \( \beta \) may vary over individuals.

The taxable income for a linear budget set will be

\[ y(\rho, R, \eta) = \arg \max_y U(\rho y + R, y, \eta), \quad y \geq 0. \]

This taxable income for linear taxes has an important role in the identification and estimation results to follow. For convex budget sets (i.e. non decreasing marginal tax rates) the distribution of taxable income over \( \eta \) will be determined by the distribution of \( y(\rho, R, \eta) \) over \( \eta \) for certain values of \( \rho \) and \( R \).

The general specification \( U(c, y, \eta) \) of the utility function allows preferences to vary across individuals in essentially any way at all. The isoelastic utility function above is included as special case, with \( \eta \) a scalar. If the elasticity \( \beta \) in that specification is also allowed to vary over individuals then \( \eta \) would be two dimensional with \( \beta \) being one of the components of \( \eta \). The Burtless and Hausman (1978) specification is also included as a special case where income and level effects can vary separately and \( \eta \) is two dimensional. In the fully nonparametric specification we allow \( \eta \) to be of unknown dimension. We do need to restrict \( \eta \) 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 \( \eta \) as representing general heterogeneity and are reserved for the Appendix.

In practice most tax systems have a finite number of rates that change at certain income values. In such cases the after tax function \( B(y) \) is piecewise linear. A piecewise-linear \( B(y) \) with \( J \) segments, indexed by \( j \), can be described by a vector \((\rho_1, \ldots, \rho_J, R_1, \ldots, R_J, \ell_1, \ldots, \ell_{J-1})\) of net-of-tax rates \( \rho_j \) (slopes), virtual incomes \( R_j \) (intercepts), and kinks or notches \( \ell_j \), with \( \ell_0 = 0 \) and \( \ell_J = \infty \). The after tax function \( B(y) \) will be continuous at each \( \ell_j \) where \( \ell_j = (R_{j+1} - R_j)/(\rho_j - \rho_{j+1}), \quad (1 \leq j \leq J - 1) \). We can represent \( B(y) \) as

\[ B(y) = \sum_{j=1}^{J} \mathbf{1}(\ell_{j-1} \leq y < \ell_j)(R_j + \rho_j y), \]

[7]
where we assume that tax rates can change at \( y = \ell_j, \) \( (j = 0, \ldots, J) \). We will also consider \( B(y) \) that need not be piecewise linear.

In general the CDF of taxable income \( y(B, \eta) \) will depend on the entire after tax function \( B \). An important simplification occurs when \( B(y) \) is continuous and marginal tax rates are increasing with \( \rho_j > \rho_{j+1}, \) so that the budget set is convex and \( B(y) \) a concave function. Let

\[
\rho(y) = \sum_{j=1}^{J} 1(\ell_{j-1} \leq y < \ell_j)\rho_j, \quad \rho(y) = B(y) - \rho(y)y,
\]

where \( \rho(y) \) is the slope from the right of \( B(y) \) and \( R(y) \) is the corresponding virtual income. Also let \( F(y|B) = \int 1(y(B, \eta) \leq y)G(d\eta) \) denote the CDF of taxable income for an after tax function \( B \) and \( F(y|\rho, R) \) the CDF of taxable income for \( B(y) = \rho y + R \).

**Theorem 1:** If Assumption A1 is satisfied, \( B(y) \) is piecewise linear and continuous, and \( \rho_j > \rho_{j+1}, \) \( (j = 1, \ldots, J - 1) \) then \( F(y|B) = F(y|\rho(y), R(y)). \)

Here we find that for concave \( B(y) \), i.e. for a convex budget set, the CDF is that for a linear after tax income with slope \( \rho(y) \) and nonlabor income \( R(y) \). At a kink where \( y = \ell_j \) for some \( j \) the slope from the right is used because of the weak inequality in the definition of the CDF. This theorem is a distributional result corresponding to the observation of Hausman (1979) that linear budget sets can be used to characterize choices when preferences are convex and \( B(y) \) is continuous with increasing marginal tax rates. We note that this result allows for general heterogeneity when the dimension of \( \eta \) is unknown. We will use this result to characterize kink probabilities as well as distributions along linear segments. A more general version of this result was given in Blomquist et al. (2015) and is presented in the Appendix of this paper.

### 3 Bunching Does Not Identify the Size of the Taxable Income Elasticity with Unrestricted Heterogeneity

Bunching estimators estimate the taxable income elasticity from the proportion \( P_K \) of individuals at or near a kink. Our discussion of bunching estimators will focus on budget sets with one kink for simplicity, as illustrated in Figure 1. What the researcher can observe is the income distribution along the kinked budget constraint. If there were no kink at \( K \), then the density function \( f_1(y) = \partial F(y|\rho_1, R_1)/\partial y \) of taxable income \( y \) along the extended first segment with \( y > K \) would be identified. 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 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 individuals who would have been on the extended first segment are now grouped at the kink.

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 goes here.

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 taxable income elasticity cannot be identified from a kink probability when the distribution of heterogeneity is unrestricted.

We follow Saez (2010) when we describe the general idea behind identifying the taxable income elasticity from a kink probability, but omit some details that are of no importance for our analysis. Saez (2010) considers a counterfactual, hypothetical change in a budget constraint. We return to Figure 1 and consider individuals maximizing their utility for a linear budget set with slope $\rho_1$ and intercept $R_1$ giving a density $f_1(y)$ of taxable income along the extended first segment. Suppose next that a kink at $y = K$ is introduced, and the slope of the budget constraint after the kink is $\rho_2 = \rho_1 + \Delta \rho$, $\Delta \rho < 0$. Suppose that individuals who would have been in the interval $(K, K + \Delta A]$ along the first segment now choose the kink point. We refer
to the individual who would have chosen $K + \Delta A$ as the marginal buncher. In Figure 1 we have drawn two indifference curves for the marginal buncher. Before the (hypothetical) change in the budget constraint, the individual had a tangency on the extended segment at $K + \Delta A$, and after the change in the budget constraint a tangency on the second segment at $K$. The discrete (e.g. arc) taxable income elasticity of the marginal buncher is 

$$e = \frac{\Delta A / K}{\Delta \rho / \rho_1}$$  

(3.2)

However, in reality we cannot observe incomes at the individual level on the extended first segment so that we do not know $\Delta A$. We do observe $P_K$, the proportion of individuals located at the kink. Because the individuals at the kink are those that would have located in $(K, K+\Delta A]$ along the first segment we have 

$$P_K = \int_{K}^{K+\Delta A} f_1(y)dy.$$  

(3.3)

If $f_1(y)$ were identified we could identify $\Delta A$ from this equation. The problem is $f_1(y)$ is not identified because it is a density for those grouped at the kink. This means that there are two structural parameters, the $\Delta A$ and the density $f_1(y)$, but only one equation involving the reduced form parameter. It is impossible to identify two structural parameters from one reduced form parameter. The order condition of having as many reduced form parameters as structural parameters is not satisfied.

We can see this nonidentification even more clearly for the isoelastic utility function where $y(\rho, \eta) = \rho^\beta \eta$. Using Theorem 1 we can calculate the bunching window for $\eta$, meaning the interval of $\eta$ for which taxable income $y$ will be at the kink. The highest value of $\eta$ giving a tangency solution on the first segment is given by the relation $K = \eta \rho_1^{-\beta}$, and the lowest value of $\eta$ giving a tangency solution on the second segment is given by $K = \eta \rho_2^{-\beta}$. The bunching window in terms of $\eta$ is therefore given by $[K \rho_1^{-\beta}, K \rho_2^{-\beta}]$, so the kink probability is 

$$P_K = \Pr(y = K) = \int_{K \rho_1^{-\beta}}^{K \rho_2^{-\beta}} \phi(\eta) d\eta,$$  

(3.4)

where $\phi(\eta)$ is the density of $\eta$. Here we can clearly see the identification problem. The size of the bunching window is increasing in $\beta$, which implies that for a given preference distribution, the bunching itself is increasing in $\beta$. 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.
Using more information about the distribution of taxable income, such as its density to the right and left of the kink, does not help with identification of $\beta$ in the isoelastic model when $\phi(\eta)$ is unrestricted. In fact we can show that given any distribution of taxable income with positive kink probability and any $b > 0$ there is a distribution of heterogeneity such that $b$ is the taxable income elasticity.

**Theorem 2:** Suppose that the CDF $F(y)$ of taxable income $y$ is continuously differentiable of order $D$ to the right and to the left at $K$ and $P_K = \Pr(y = K) > 0$. Then for any $\beta$ there exists a CDF $\Phi(\eta)$ of $\eta$ such that the CDF of taxable income obtained by maximizing the utility function in equation (2.1) equals $F(y)$, and $\Phi(\eta)$ is continuously differentiable of order $D$.

Theorem 2 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 we find that the distribution of taxable income for one budget set with one kink has no information about the taxable income elasticity when the distribution of heterogeneity is unrestricted. The same result can be shown for any continuous, piecewise linear $B(y)$ with nondecreasing marginal tax rates and each kink having positive probability. This lack of identification can also be thought of as failure of an order condition, with the distribution of taxable income being the one reduced form parameter but there being two structural parameters, the elasticity and the distribution of heterogeneity.

The relationship between the kink probability and compensated elasticities is also interesting in the nonparametric case. Taxable income elasticities are also not identified in this case, as one would expect by the nonparametric specification being less restrictive than the isoelastic formula, but there is a useful formula for the kink probability. Consider a budget set with kink $K$ and slopes $\rho_1$ and $\rho_2$ of $B(y)$ at $K$ from the left and right respectively. Consider $\rho$ with $\rho_2 \leq \rho \leq \rho_1$ and and let $R(\rho) = R_1 + K(\rho_1 - \rho)$ be the virtual income for the linear budget set with slope $\rho$ passing through the kink. Assuming that $y(\rho, R(\rho), \eta)$ is continuously distributed, let $f(y|\rho, R(\rho)) = \partial F(y|\rho, R(\rho))/\partial y$ denote its pdf and

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

where the expectation is taken over the distribution of $\eta$ and existence of derivatives is imposed in Assumption A2 given in the Appendix. This $\beta(\rho, K)$ is the average taxable income elasticity for those individuals facing a linear $B(y)$ with slope $\rho$ that passes through the kink point, who choose to locate at the kink point. The following result gives a formula for the kink probability in terms of $\beta(\rho, K)$ and $\bar{\phi}(\rho)$:
**Theorem 3:** If Assumptions A1 and A2 are satisfied then

\[
P_K = \int_{\rho_2}^{\rho_1} \tilde{\phi}(\rho) \beta(\rho, K) d\rho \quad \text{and} \quad \tilde{\phi}(\rho) \beta(\rho, K) = -F_\rho(K|\rho, R(\rho)) + K \cdot F_R(K|\rho, R(\rho)). \quad (3.5)
\]

The compensated elasticity appears here because virtual income is being adjusted as \( \rho \) 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 \( P_K \) in Theorem 3 bears some resemblance to the kink probability formulas in Saez (2010) but differs in important ways. Theorem 3 is global, nonparametric, takes explicit account of general heterogeneity, and allows for income effects unlike the Saez (2010) results, which are local or parametric and account for heterogeneity implicitly. Theorem 3 and the discussion in the next two paragraphs was given in Blomquist and Newey (2015).

Theorem 3 helps clarify what can be nonparametrically learned from kinks. First, the compensated effects that enter the kink probability are only for individuals (i.e. values of \( \eta \)) who would choose to locate at the kink for a linear budget set with \( \rho \in [\rho_2, \rho_1] \). Thus, using kinks to provide information about compensated effects is subject to the same issues of external validity as, say, regression discontinuity design (RDD). As RDD only identifies treatment effects for individuals at the jump point so kinks only provide information about compensated effects for individuals who would locate at the kink.

Second, the kink probability depends on both the average compensated elasticity \( \beta(\rho, K) \) and on a pure heterogeneity effect \( \tilde{\phi}(\rho) \). Intuitively, a kink probability could be large because the elasticities are large or because preferences are distributed in such a way that many like to be at the kink, i.e. so that \( \tilde{\phi}(\rho) \) is large. Consequently it is not possible to separately identify compensated tax effects and heterogeneity effects from a kink. For example consider the weighted average elasticity

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

Evidently \( \bar{\beta}(K) \) depends on the denominator \( \int_{\rho_2}^{\rho_1} \tilde{\phi}(\rho) d\rho \) which is needed to normalize so that \( \bar{\beta}(K) \) is a weighted average of elasticities. This denominator is not identified because no \( \rho \in (\rho_2, \rho_1) \) is observed in the data. Indeed, \( \tilde{\phi}(\rho) \) can be any positive function over the interval so the denominator can vary between 0 and \( \infty \), meaning that any \( \bar{\beta}(K) \in (0, \infty) \) is consistent with the data. Here we see that identification from a kink for a nonparametric specification is similar to that for an isoelastic utility. Just as for the isoelastic utility, a kink is uninformative.
about the size of an average taxable income elasticity for the nonparametric model with general heterogeneity.

Saez (2010) and Chetty et al. (2011) do give estimators of the taxable income elasticity from a kink. By the order condition for identification we know that to obtain the elasticity from the one reduced form parameter (the kink probability), everything about the density must come from outside the kink. Saez (2010) obtains an estimator by implicitly assuming that the density \( \phi(\eta) \) is linear over the bunching interval and is continuous from the left and from the right at the respective lower and upper endpoints \( K\rho_1^{-\beta} \) and \( K\rho_2^{-\beta} \). To demonstrate that these assumptions give the Saez (2010) estimator, 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 \( \bar{\eta} = K\rho_1^{-\beta} \) and \( \bar{\eta} = K\rho_2^{-\beta} \) be the endpoints of the bunching interval. Accounting for the Jacobian of the transformation \( y = \eta \rho_1^{\beta} \) we have \( \phi(\eta) = f^{-}(K)\rho_1^{\beta} \) and \( \phi(\bar{\eta}) = f^{+}(K)\rho_2^{\beta} \). Assuming that \( \phi(\eta) \) is linear on the bunching interval we then have

\[
P_K = \int_{\bar{\eta}}^{\eta} \phi(t) \, dt = \frac{1}{2} \left[ \phi(\eta) + \phi(\bar{\eta}) \right] (\bar{\eta} - \eta) = \frac{1}{2} \left[ f^{-}(K)\rho_1^{\beta} + f^{+}(K)\rho_2^{\beta} \right] (K\rho_2^{-\beta} - K\rho_1^{-\beta})
\]

\[
= \frac{K}{2} \left[ f^{-}(K) + f^{+}(K) (\rho_1/\rho_2)^{-\beta} \right] \left[ (\rho_1/\rho_2)^{\beta} - 1 \right].
\]

This is the equation for \( \beta \) 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 \((\eta, \bar{\eta})\). 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.

One could also construct an estimator of the average nonparametric elasticity \( \bar{\beta}(K) \) analogous to the Saez (2010) estimator for isoelastic utility. Assume that \( f^{-}(K) \) and \( f^{+}(K) \) are identified as before. If \( f(K|\rho, R(\rho)) \) is assumed to be linear in \( \rho \) for \( \rho \in [\rho_2, \rho_1] \) then on that interval

\[
\bar{\phi}(\rho) = (K/\rho)f(K|\rho, R(\rho)) = K(a + b), a = \frac{\rho_1f^{+}(K) - \rho_2f^{-}(K)}{\rho_1 - \rho_2}, b = \frac{f^{-}(K) - f^{+}(K)}{\rho_1 - \rho_2}.
\]

Integrating then gives

\[
\bar{\beta}(K) = \frac{P_K}{K[a \ln(\rho_1/\rho_2) + f^{-}(K) - f^{+}(K)]}.
\]

This gives an nonparametric average taxable income elasticity with general heterogeneity for individuals located at the kink.
There appears to be no way to use additional information to construct an estimator of \( \tilde{\beta}(K) \) analogous to Chetty et al. (2011) from identifying \( \hat{\phi}(\rho) \) at points other than \( \rho_1 \) and \( \rho_2 \). The problem is that there is no data in a single budget set about the density of taxable income at the kink point except for \( \rho_1 \) and \( \rho_2 \). All that is identified is \( f^{-}(K) \) and \( f^{+}(K) \). Since we only identify \( \hat{\phi}(\rho) \) at the two points \( \rho_1 \) and \( \rho_2 \) the only unique polynomial that can be fit is a line. In this way the nonparametric case with general heterogeneity appears to be different than isoelastic utility, in that only linear interpolation of the weighting function \( \hat{\phi}(\rho) \) is possible across the kink point.

Notches have also been used to estimate the taxable income elasticity beginning with Kleven and Waseem (2013). A notch occurs at an income value where the average tax rate changes so there is a discontinuity in the budget set. The marginal tax rate may also change at a notch. Similarly to kinks, for isoelastic utility the bunching at a notch provides no information about the size of the elasticity when heterogeneity is unrestricted.

A notch occurs at \( K \) when there is a drop discontinuity in \( B(y) \) at \( K \) with \( B(y) \) being lower to the right of \( K \) than to the left of \( K \). Consider an indifference curve that passes through the notch point and is tangent to the segment beyond the notch, with no part of the budget set being greater than or equal to the indifference curve at any other taxable income. Let \( y^1(\beta) \) denote the tangency point, which depends just on \( \beta \). This tangency point is determined by the tangency of the indifference curve at \( y^1(\beta) \) and the utility at the notch point being equal to the utility at \( y^1(\beta) \) on the second segment. For brevity we omit the detailed formula for \( y_1(\beta) \). We also let \( \eta_2(\beta) \) denote the value of \( \eta \) corresponding to this indifference curve.

To see that bunching at the notch provides no information about the elasticity note that the bunching window is \([K\rho_1^{-\beta}, \eta_2(\beta)]\), so that

\[
P_K = \int_{K\rho_1^{-\beta}}^{\eta_2(\beta)} \phi(\eta) d\eta.
\]

As for bunching at a kink, the mass at the notch will be increasing in the taxable income \( \beta \). Also as for the kink, for any \( \beta \) we can find a pdf such that this equation is satisfied. Consequently, as with a kink, bunching at a notch provides no information about the size of the taxable income elasticity.

Unlike a kink, for isoelastic utility the entire distribution of taxable income does vary with \( \beta \) in such a way that \( \beta \) can be identified. With a notch there is a "gap" region \((K, y^1(\beta))\) where no one would choose to be. The upper limit \( y^1(\beta) \) of this gap region can be shown to be a one-to-one function of \( \beta \), and so \( \beta \) is identified. The value of \( y^1(\beta) \) could be estimated at the smallest value of taxable income that exceeds \( K \), an order statistic type of estimator, which would be consistent. Details of this identification argument can be found on pp. 36-37
of Blomquist and Newey (2018).

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

4 Partial Identification for a Single Budget Set

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 size of the taxable income elasticity. We also showed that if certain parametric forms for distribution across the kink are specified then the elasticities are identified. In this Section we show that nonparametric a prior restrictions on the density across the kink can provide some information about the elasticity. Specifically we consider what can be learned if the density is known to be bounded above by a scalar multiple of the maximum of the densities at the boundaries of the bunching interval and below by another scalar multiple of the minimum of the densities at the boundary. These bounds include as a special case monotonicity where the upper and lower scalars are each one.

For isoelastic utility we consider bounds when the heterogeneity density $\phi(\eta)$ is known to be bounded above and below over the bunching interval. We also follow the literature and consider a range around the kink rather than just the kink itself. Let $y_\ell$ and $y_u$ denote lower and upper endpoints for a taxable income interval that includes the kink, where excess bunching may occur. Let $\eta_\ell = y_\ell \rho_1^{-\beta}$ and $\eta_u = y_u \rho_2^{-\beta}$ denote corresponding lower and upper endpoints for

$$f^-(y_\ell) = \lim_{y \rightarrow y_\ell, y < y_\ell} f(y), \quad f^+(y_u) = \lim_{y \rightarrow y_u, y > y_u} f(y).$$

Consider the two functions

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

We have the following result:

**Theorem 4:** If there are positive scalars $\bar{\sigma} \geq 1$ and $\sigma \leq 1$ such that for $\eta \in [\eta_\ell, \eta_u]$,

$$\sigma \max\{\phi(\eta), \phi(\eta_u)\} \leq \phi(\eta) \leq \bar{\sigma} \max\{\phi(\eta_\ell), \phi(\eta_u)\}$$

then the taxable income elasticity $\beta$ satisfies

$$\sigma \min\{D^-(\beta), D^+(\beta)\} \leq \Pr(y_\ell \leq y \leq y_u) \leq \bar{\sigma} \max\{D^-(\beta), D^+(\beta)\}. \quad (4.6)$$

[15]
If \( \phi(\eta) \) is monotonic then these bounds hold for \( \sigma = \bar{\sigma} = 1 \). If \( \Pr(y_\ell \leq y \leq y_u) < \min\{D^-(0), D^+(0)\} \) then there is no \( \beta \) satisfying this equation. Otherwise the set of all nonnegative \( \beta \) satisfying this equation is a subset of \([0, \infty)\).

For estimation we just plug in nonparametric estimators \( \hat{f}^- (y_\ell) \) and \( \hat{f}^+ (y_u) \) to obtain
\[
\hat{D}^- (\beta) = \hat{f}^- (y_\ell) \left[ y_u (\rho_1 / \rho_2)^\beta - y_\ell \right], \\
\hat{D}^+ (\beta) = \hat{f}^+ (y_u) \left[ y_u - y_\ell (\rho_2 / \rho_1)^\beta \right].
\]
Estimated bounds for \( \beta \) are \( \hat{\beta}_\ell \) and \( \hat{\beta}_u \) that solve
\[
\bar{\sigma} \max \left\{ \hat{D}^- \left( \hat{\beta}_\ell \right), \hat{D}^+ \left( \hat{\beta}_\ell \right) \right\} = P_K, \quad \sigma \min \left\{ \hat{D}^- \left( \hat{\beta}_u \right), \hat{D}^+ \left( \hat{\beta}_u \right) \right\} = P_K.
\]

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 approximate the graph by a function that is linear between each of the following pairs of points:
\[
(-5000, .21), (-2500, .35), (0, .44), (2000, .35), (5000, .35).
\]
We take the taxable income density over \((-5000, 5000)\) to be the piecewise-linear function connecting these points, up to scale. The bounds of Theorem 4 are for the isoelastic utility which requires positive taxable income. From Panel A of Figure 7 it appears that density is about zero at \(-22,000\), so we take that to be the lowest possible value of taxable income. To apply those bounds we normalized so that \( y \) in the isoelastic utility function is replaced by \( y + 22,000 \), i.e. we replace each taxable income by its value plus 22,000. We then calculated the bounds of Theorem 4 under monotonicity, where \( \sigma = \bar{\sigma} = 1 \).

We considered two bunching intervals. For \( y_\ell = 17,000 \), \( f^- (y_\ell) = .21 \), \( y_u = 27,000 \), and \( f^+ (y_u) = .35 \) we obtain bounds
\[
\hat{\beta}_\ell = .021, \quad \hat{\beta}_u = 1.01.
\]
These bounds are quite wide. For \( y_\ell = 19,500 \), \( f^- (y_\ell) = .35 \), \( y_u = 24,000 \), and \( f^+ (y_u) = .35 \) we obtain bounds
\[
\hat{\beta}_\ell = .106, \quad \hat{\beta}_u = .135.
\]
These bounds are narrower but perhaps less plausible, because the narrower bunching window means optimization errors could have a bigger impact as we discuss in Section 6. Overall the bounds suggest that the large elasticities found in some studies may not be plausible.

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

[16]
where monotonicity is included as a special case where \( \sigma = \bar{\sigma} = 1 \). Note that this assumption is different than the previous assumption on the pdf of \( \eta \) for the isoelastic utility. It does provide one way to limit the variation in \( f(K|\rho, R(\rho)) \). The bounds on \( f(K|\rho, R(\rho)) \) imply that

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

Integrating and inverting these bounds gives

\[
\frac{P_K}{K \ln(\rho_1/\rho_2)\sigma \max\{f^-(K), f^+(K)\}} \leq \hat{\beta}(K) \leq \frac{P_K}{K \ln(\rho_1/\rho_2)\sigma \min\{f^-(K), f^+(K)\}}.
\]

Each of the bounds we have given can be quite wide when the estimated densities from the left and right are far apart or \( \sigma \) is very different than \( \bar{\sigma} \). The difference in density estimates explains the width of the bounds in the application for the wider bunching interval. One could construct tighter bounds by putting more restrictions on the heterogeneity pdf, such as concavity. However, all such bounds are based entirely on prior information when there is only a single budget set. As we have discussed, the data provides no information on the heterogeneity density for individuals at the kink. Thus, information about the density at the kink must come from a source other than the data, when there is only one budget set.

5 Identification from Budget Set Variation

Given the identification difficulties for bunching it seems important to consider what will identify the taxable income elasticity. We know from Section 3 that 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 generally 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. In most 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 \( \beta \) for the isoelastic utility function 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. For isoelastic utility there is one unknown function, the CDF of \( \eta \), and one unknown parameter, the taxable income elasticity \( \beta \). Two functions can be more than enough to identify one function.
and one parameter. In fact, the taxable income elasticity can be overidentified with strong restrictions being imposed on the distribution of taxable income across the two budget sets.

We give an identification result for the isoelastic specification when the data include two net of tax functions \( B(y) \) and \( \tilde{B}(y) \) that are both continuous with decreasing marginal tax rates and the distribution of \( \eta \) does not depend on which tax schedule is in place. Here we are assuming that the distribution of heterogeneity is independent of the budget set. Let \( \rho(y) \) and \( \tilde{\rho}(y) \) denote the slope from the right of \( B(y) \) and \( \tilde{B}(y) \) respectively and let \( F(y) \) and \( \tilde{F}(y) \) be the corresponding distributions of taxable income for the two budget sets. Since the choice for a linear budget set is \( \rho \downarrow \) it follows by Theorem 1 that

\[
F(y) = \Pr(\eta \rho(y)\beta \leq y) = \Phi(y \rho(y)^{-\beta}), \quad \tilde{F}(y) = \Phi(y \tilde{\rho}(y)^{-\beta}),
\]

where \( \Phi(\eta) \) is the CDF of \( \eta \). Here we see that the two distributions are the same except for a scalar multiple of the taxable income \( y \). 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 \( \beta \). We can use this feature to obtain \( \beta \) from the size of the scale adjustment when the tax rate changes.

Theorem 5: If taxable income is chosen by maximizing isoelastic utility, \( \Phi(\eta) \) is continuous and strictly monotonic increasing on \((0, \infty)\) and there exist \( y \) and \( \tilde{y} \) such that \( F(y) = \tilde{F}(\tilde{y}) \) and \( \rho(y) \neq \tilde{\rho}(y) \) then

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

Here we see that \( \beta \) is identified from any pair of taxable incomes \( y \) and \( \tilde{y} \) with the same value of the distribution for the first and second budget sets but a different marginal tax rate. Note that \( F(y) = \tilde{F}(\tilde{y}) \) is the same as \( \Phi(y \rho(y)^{-\beta}) = \Phi(y \tilde{\rho}(\tilde{y})^{-\beta}) \), which means that \( y \) and \( \tilde{y} \) correspond to the same point in the distribution of \( \eta \). Also, since \( \Phi(\eta) \) is strictly monotonic, the same point in the distribution of \( \eta \) could be thought of as the same value of \( \eta \), i.e. as the same type of individual. Therefore, the identification assumption of this result is that there is an individual type \( \eta \) that faces different marginal tax rates at the values of taxable income chosen in the respective budget sets. This is an intuitive condition for identification of the taxable income elasticity, that the budget sets vary in such a way that some individual faces different marginal tax rates under their choice for the two budget sets.

One example is provided by two piecewise linear budget sets with one kink that is the same, with common \( \rho_1 \) up to the kink \( K \), and different slopes \( \tilde{\rho}_2 < \rho_2 \) respectively, beyond the kink. Here identification follows easily. Consider any \( \tilde{y} \geq K \). By \( F(K) = \Phi(K \rho_2^{-\beta}) < \Phi(K \tilde{\rho}_2^{-\beta}) = \Phi(K \rho_2^{-\beta}) \)
\( \hat{F}(K) < 1 \) and \( \Phi(\eta) \) strictly monotonic and continuous there is \( y > \tilde{y} \) such that \( F(y) = \hat{F}(\tilde{y}) \), so by Theorem 3 we have \( \beta = \ln(\tilde{y}/y) / \ln(\hat{\rho}_2/\rho_2) \). Indeed, such an \( y \) exists for any \( \tilde{y} \geq K \) so that there is a continuum of identifying equations for \( \beta \). In this example the elasticity is highly overidentified.

A more complicated example is provided by two piecewise linear budget sets each with one kink, where the two tax rates are the same but the second kink \( \hat{K} \) is different than the first kink \( K \), say \( K < \hat{K} \). The slope from the right of these two budget sets differ only in the interval \( [K, \hat{K}] \). Thus, to apply Theorem 3 there must exist \( y < \tilde{y} \), both in \( [K, \hat{K}] \), with \( F(y) = \hat{F}(\tilde{y}) \). We can generalize this condition slightly for the purposes of this example. Let \( \hat{F}_-(\hat{K}) = \lim_{y \to \hat{K}, y < \hat{K}} \hat{F}(y) \). It turns out that the necessary and sufficient condition for identification of \( \beta \) is \( \hat{F}_-(\hat{K}) \geq F(K) \). Intuitively, identification holds when some individual who was on the linear segment beyond the original kink \( K \), or just on the border of that, experiences a tax change. This would occur when the new kink \( \hat{K} \) is at or beyond the end \( K + \Delta A \) of the extended first segment shown in Figure 1. If \( \hat{F}_-(\hat{K}) < F(K) \) then the new kink will be left of the end of the extended first segment, so there will be no individual for whom there must be a change in the tax rate across budget sets. In that case there will be a bound on \( \beta \) but there will be no information about \( \beta \) other than that bound.

**Theorem 6:** Suppose that taxable income is chosen by maximizing isoelastic utility and \( \Phi(\eta) \) is continuous and strictly monotonic increasing on \( (0, \infty) \). If \( \hat{F}_-(\hat{K}) \geq F(K) \) then there is a unique \( y^* > K \) with \( F(y^*) = F_-(\hat{K}) \) and \( \beta = \ln(\hat{K}/y^*) / \ln(\hat{\rho}_1/\rho_2) \). If \( \hat{F}_-(\hat{K}) < F(K) \) then \( \beta > \beta = \ln(\hat{K}/K) / \ln(\hat{\rho}_1/\rho_2) \) and for any \( b > \beta \) there exists \( \Phi^b(\eta) \) such that when \( \beta = b \) and \( \Phi^b(\eta) \) is the CDF of \( \eta \), the CDF of \( y \) is \( F(y) \) for the budget set with kink \( K \) and is \( \hat{F}(y) \) for the budget set with kink \( \hat{K} \).

This result shows that not every pair of budget sets will serve to identify the taxable income elasticity, even for isoelastic utility. It is interesting to note that a shift in the kink does provide at least some information in the form of a lower bound on the elasticity, where that lower bound is larger the bigger the shift in the kink and smaller the bigger the ratio of the two tax rates. Of course as the shift gets larger one would also move towards a situation where the elasticity is point identified. Also, a shift in the kink implies strong, testable restrictions on the CDF of taxable income, that it coincides for each budget set for \( y < K \) and \( y \geq \hat{K} \). This restriction holds even when \( \beta \) is not identified and is a consequence of scalar heterogeneity and no income effect.

One could let both \( \eta \) and \( \beta \) vary over individuals, giving a linear random coefficients specification \( \ln(y) = \ln(\eta) + \beta \ln(\rho) \), where both \( \ln(\eta) \) and \( \beta \) are random. If the budget sets were
linear then a least squares regression of $\ln(y)$ on a constant and $\ln(\rho)$ 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. In what follows we will focus on identifying the effect of taxes on the expected value of taxable income.

6 Optimization Errors

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

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

In the bunching literature optimization errors are accounted for by choosing a bunching window that is larger than just the kink, as we did in the formulation of Theorem 4. When the taxable income elasticity is positive changing the bunching window from being just the kink to something larger reduces bias in that it generally gives a positive elasticity rather than one that is zero. However, using a bunching window need not account fully for optimization errors.

The presence of optimization errors makes identification more difficult because with optimization errors there is another unknown object, the distribution of the error. The kink probability will depend on the distribution of optimization errors and so there will be even more unknown parameters to account for in bunching estimation. Cattaneo et al. (2018) give results on identification and estimation from kinks when there is an additive optimization error that is independent of the budget set. Identification from variation in budget sets would also be more difficult when there are optimization errors. One virtue of identifying tax effects from the expected value of taxable income is that additive and/or multiplicative optimization errors are
allowed for. Indeed one motivation for a focus on the expected value of taxable income is that taxable income effects will be identified in the presence of optimization errors. It would also be interesting to allow for optimization errors in estimation using the distribution of taxable income, but that is beyond the scope of this paper.

We present calculations that show how, for a given taxable income elasticity, the object estimated by bunching estimators will vary as we make simple variations in the preference distribution and allow for various types of optimization errors. We simulate many observations from the isoelastic utility function with optimization errors and apply bunching estimation to determine the object of estimation. 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 objects estimated by bunching methods we used the program bunchr, written by Itai Trilnick in the programming language R.\(^3\) That program uses the isoelastic utility specification, takes the heterogeneity density near and across the kink to be a polynomial similar to Chetty et al. (2011), and uses a default bunching window. Alternative ways of determining the bunching window, such visual inspection, could be used. As we have shown, without prior restrictions on the distribution of heterogeneity there is no information in the distribution of taxable income about the size of the elasticity. Thus, any method of determining the bunching window cannot help with identification, though the variation across different specifications in the object estimated by bunching could be different than what is reported here.

In our simulations we illustrate how the bunching object, for a given value of the taxable income elasticity, will vary as we change the preference distribution. We can change the preference distribution in many ways; we can change the general shape, the center of the location and the variance. Here we will keep the center of location constant as well as the general shape. We will see how the bunching object 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 \((K, K + \Delta y) = (1000, 1070)\). Expressed in terms of \(\eta\) the bunching window is approximately \((1074,1149)\). We centered the preference distribution at 1100 and represent the preference distribution with a mixed normal \(\phi(\eta) = \pi \cdot n(1100, 10^2) + (1 - \pi) \cdot n(1100, 140^2), \pi \in (0,1)\). As we vary \(\pi\) 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 \(\pi, (1 - \pi)\) used. The second row shows

\(^3\)The program can be accessed via the link https://CRAN.R-project.org/package=bunchr.
how results vary as we change the proportions and there are no optimization errors. We see that the object varies 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 object estimated by the bunching can be far from 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 objects 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 results very much.

TABLE: Simulations with Mixed Normals

<table>
<thead>
<tr>
<th>$\pi; (1 - \pi)$</th>
<th>0.9; 0.1</th>
<th>0.7; 0.3</th>
<th>0.5; 0.5</th>
<th>0.3; 0.7</th>
<th>0.1; 0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}$</td>
<td>0.598</td>
<td>0.500</td>
<td>0.402</td>
<td>0.302</td>
<td>0.192</td>
</tr>
<tr>
<td>$\hat{\beta}_{\text{opt error}1}$</td>
<td>0.075</td>
<td>0.081</td>
<td>0.080</td>
<td>0.074</td>
<td>0.058</td>
</tr>
<tr>
<td>$\hat{\beta}_{\text{opt error}2}$</td>
<td>0.012</td>
<td>0.013</td>
<td>0.011</td>
<td>0.010</td>
<td>0.008</td>
</tr>
<tr>
<td>$\hat{\beta}_{\text{opt error}3}$</td>
<td>0.0</td>
<td>0.013</td>
<td>0.035</td>
<td>0.077</td>
<td>0.065</td>
</tr>
<tr>
<td>$\hat{\beta}_{\text{opt error}4}$</td>
<td>0.530</td>
<td>0.462</td>
<td>0.394</td>
<td>0.306</td>
<td>0.238</td>
</tr>
</tbody>
</table>

$\hat{\beta}$ no optimization errors; $\hat{\beta}_{\text{opt error}1}$ symmetric optimization errors, mean zero, std 25; $\hat{\beta}_{\text{opt error}2}$ symmetric optimization errors, mean zero, std 50; $\hat{\beta}_{\text{opt error}3}$ negative asymmetric optimization errors; $\hat{\beta}_{\text{opt error}4}$ 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. 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.
7 The Taxable Income Regression

The expected value \( \mu(B) = \int y(B, \eta)G(d\eta) \) of taxable income for a given net of tax function \( B(y) \) is useful for identifying important policy effects, where \( G(\eta) \) is the CDF of \( \eta \). The function \( \mu(B) \) can be used to predict the effect of tax changes (i.e. changes of \( B \)) on average taxable income. Furthermore, the conditional mean will have the form implied by utility maximization in the presence of certain kinds of optimization errors, such as an additive error with conditional mean zero or a multiplicative error with conditional mean one (or both). The allowance for optimization errors is one good reason to focus on the conditional mean of taxable income rather than attempt to model the conditional distribution of taxable income, which is much more complicated with optimization errors.

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

\[
\bar{y}(\rho, R) = \int yF(dy|\rho, R),
\]

\[
\nu(\rho, R, \ell) = \int 1(y < \ell)(y - \ell)F(dy|\rho, R), \lambda(\rho, R, \ell) = \int 1(y > \ell)(y - \ell)F(dy|\rho, R).
\]

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

**Theorem 7:** If Assumption A1 is satisfied, \( \int |y(\rho, R, \eta)||G(d\eta) < \infty \) for all \( \rho, R > 0 \), and \( B(y) \) is continuous with increasing marginal tax rates then

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

(7.7)

\[
= \bar{y}(\rho_1, R_1) + \sum_{j=1}^{J-1} [\lambda(\rho_{j+1}, R_{j+1}, \ell_j) - \lambda(\rho_j, R_j, \ell_j)].
\]

The first equality in the conclusion is exactly analogous to the conclusion of Theorem 2.1 of Blomquist and Newey (2002, BN henceforth). 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 7 generalizes Theorem 2.1 of BN by allowing general heterogeneity and zero hours of work, whereas BN assumed scalar $\eta$. 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 1, 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 $\mu(B)$. The definitions of $\bar{y}(\rho, R)$ and $\nu(\rho, R, \ell)$ (or $\lambda(\rho, R, \ell)$) and the conclusion of Theorem 7 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 7 gives a more parsimonious approximation than simply using the additive form of the conditional mean.

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 $\beta_{ab}, (a = 2,...,A; b = 1,...,B)$ 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

$$F(y|x) \approx F_1(y) + \sum_{a=2}^{A} w_a(x, \beta) [F_a(y) - F_1(y)]$$

$$= \sum_{a=1}^{A} w_a(x, \beta) F_a(y), w_1(x, \beta) = 1 - \sum_{a=2}^{A} w_a(x, \beta).$$

One can think of this 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 $w_a(x, \beta)$ are positive, so the implicit CDF approximation may not be monotonic increasing in $y$. We are primarily interested in approximating the expected value and so are not concerned that associated 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 7. Let

$$\bar{y}_a = \int y F_a(dy), \nu_a(\ell) = \int 1(y < \ell)(y - \ell) F_a(dy), (a = 1,...,A).$$
Substituting the CDF approximation in the expression for the conditional mean gives

\[ \mu(B) \approx \bar{y}_1 + \sum_{a=2}^{A} w_a(x, \beta)(\bar{y}_a - \bar{y}_1) + \sum_{a=2}^{A} \sum_{j=1}^{J-1} \left[ w_a(x_j, \beta) - w_a(x_{j+1}, \beta) \right] [\nu_a(\ell_j) - \nu_1(\ell_j)] \]

\[ = \bar{y}_1 + \sum_{a=2}^{A} \sum_{b=1}^{B} \beta_{ab} \left[ r_b(x_j)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J-1} [r_b(x_j) - r_b(x_{j+1})] [\nu_a(\ell_j) - \nu_1(\ell_j)] \right]. \]

This is a series approximation, where the regressor corresponding to \( \beta_{ab} \) 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 \( \bar{y}_i \) 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 \( \bar{y}^\rho(\rho, R) \) and in \( \nu(\rho, R, \ell) \). We show that this approximation is valid in Theorem A6 of Appendix A. There we give an approximation error bound for \( \mu(B) \) based on spline approximation. We find that the bound does not depend on the number of segments.

Utility maximization imposes the restriction on \( F(y|\rho, R) \) that this CDF satisfy a Slutzky like condition

\[ \frac{\partial F(y|\rho, R)}{\partial \rho} - y \frac{\partial F(y|\rho, R)}{\partial R} \leq 0. \]

In Theorem B1 of Appendix B we show that this condition is necessary and sufficient for \( F(y|\rho, R) \) to be the CDF of taxable income for individuals maximizing utility with general heterogeneity. We can test or impose this condition on a grid \( x_1, \ldots, x_C, \) and \( y_1, \ldots, y_D \) in connection with the conditional mean approximation. 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

\[ \sum_{a=2}^{A} \left[ \frac{\partial w_a(x_c, \beta)}{\partial \rho} - y_d \frac{\partial w_a(x_c, \beta)}{\partial R} \right] [F_a(y_d) - F_1(y_d)] \leq 0, \quad (c = 1, \ldots, C; d = 1, \ldots, D). \quad (7.8) \]

These are a set of linear in parameters, inequality restrictions on the coefficients \( \beta \) of the weights \( w_a(x, \beta) \). These restrictions can be checked in a particular application, as we will do. Also, 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 imposing all those conditions approximately imposes all the conditions of utility maximization.

We can also make allowance for nonincreasing marginal tax rates. We show in the Appendix that when the marginal tax rate only decreases at one place and the smallest concave function
greater than or equal to $B(y)$ only departs from $B(y)$ over the segments immediately above and below the increase point then

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

where the marginal tax rate decreases at $\ell_j$. The $\zeta$ term represents the deviation of the mean from what it would be if marginal tax rates were always increasing. This extra term can be accounted for in the approximation by separately including series terms that depend just on $(R_{j-1}, \rho_{j-1}, R_j, \rho_j)$. If relatively few individuals have taxable income on these segments or marginal tax rates only increase for low income values, as is often true in applications, then adding such terms would have little impact on the approximation of $\mu(B)$. The integration across individuals to obtain the expected value reduces the importance of decreasing marginal tax rates.

8 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. Identification of $\bar{y}(\rho, R)$ only occurs if the support of the lowest tax rates overlaps with the support of the highest tax rates as discussed in BN, which is not likely to happen in practice. Therefore we must look for useful elasticities for nonlinear budget sets. 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.

Consider the case where the budget set has $J$ segments and is continuous so that $\mu(B) = g(\rho_1, \ldots, \rho_J, R_1, \ldots, R_J)$ for a function $g$. 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 changes should be identified. In the Swedish data set we use there is tax rate variation from local tax rates and tax changes over time and wide variation in unearned income, so we expect these effects are well identified in our application.
Consider the derivative of $G(a, b)$ with respect to $a$ evaluated at $a = b = 0$, given by 
$$\frac{\partial G}{\partial a} = \sum_{j=1}^{J} \frac{\partial g}{\partial \rho_j}.$$ 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 for the segments where individuals are actually located. Our elasticity $(\frac{\partial G}{\partial a})(\tilde{\rho}/\mu(B))$ 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
$$\frac{\partial G}{\partial b} = \sum_{j=1}^{J} \frac{\partial g}{\partial R_j}.$$ 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 $\frac{\partial G}{\partial b}$.

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 leading to a utility specification of the form $U(c, y/\phi(t), \eta)$, where $\phi(t)$ is a productivity index for period $t$ with a normalization that $\phi(t_0) = 1$ for some time period $t_0$ and $\phi(t) > 1$ for $t > t_0$. Intuitively, due to productivity growth the same taxable income requires less effort in later periods with $\phi(t)$ being a productivity factor that is the same for all individuals. In the Appendix we show that this specification can follows from a model including effort where $\phi(t)$ represents wage growth.

A feature of this problem is that the utility function shifts over time. Our approach to repeated cross section data depends on using a preference specification invariant to individuals 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 individual utility problem becomes $\max \tilde{y} U(\tilde{B}(\tilde{y}), \tilde{y}, \eta)$ where $\tilde{B}(\tilde{y})$ is the net of tax income function for $\tilde{\rho} = \phi(t)\rho$. 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 $\phi(t_0)$ to one. Changing the base year would change the utility function.

This way to account for productivity growth is similar to that used in log-linear models. For isoelastic utility with linear $B(y)$ we would have $y = [\phi(t)\rho]^\beta\eta$, where $\beta$ is the net-of-tax elasticity of interest and 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 $\beta$ can be identified in a regression involving
the logarithm of the uncorrected variables \( y \) and \( \rho \). 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.

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 \( \rho \) 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.
9 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 conditional mean one. That is, suppose that the data consist of observations on taxable income and budget frontiers for individuals \((Y_i, B_i, \text{\eta}_i)\), \((i = 1, \ldots, n)\) with \(Y_i = \epsilon_{i2} y(B_i, \eta_i) + \epsilon_{i1}\), \(B_i\) and \(\eta_i\) are statistically independent, \(E[\epsilon_{i1}|B_i] = 0\), and \(E[\epsilon_{i2}|B_i, \eta_i] = 1\). 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. In this way estimating the conditional mean allows for optimization errors in the form of \(\epsilon_{i1}\) and \(\epsilon_{i2}\) described here.

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

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

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 \(\ell_{ij}\) is the location of the \(j^{th}\) kink. 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

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

Independence of \(\eta_i\) 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 \(\eta_i\) and \((B_i, v(w_i, \delta_0))\) are independent for some parameter value \(\delta_0\). 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 \(\beta_{ab}\) and \(\delta\) could then be
estimated by nonlinear least squares, as the minimizers of
\[
\sum_{i=1}^{n} \{Y_i - \bar{y}_1 - \sum_{a=2}^{A} \sum_{b=1}^{B} \beta_{ab} \{r_b(x_{i,a}, v(w_i, \delta))(\bar{y}_a - \bar{y}_1) \nolinebreak + \sum_{j=1}^{J_i-1} \rho_b(x_{ij}, v(w_i, \delta)) - r_b(x_{ij+1}, v(w_i, \delta)) \nu_a(\ell_j) - \nu_1(\ell_j) \}^2.
\]

An estimable control variable can be used to account for endogeneity. Such a control variable would be \( \xi_i \) such that \( \beta \) and \( \eta \) are independent conditional on \( \xi_i \) and the conditional support of \( \xi_i \) given \( B_i \) equals the marginal support of \( \xi_i \). 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 \( \xi_i \). This integral can be estimated by letting the functions \( r_b(x, \xi) \) depend on \( \xi \) and then including \( \hat{\xi}_i \) to form the regressors
\[
\tilde{r}_b(x) = n \sum_{i=1}^{n} \sum_{a=2}^{A} \sum_{b=1}^{B} \beta_{ab} \{r_b(x_{i,a}, \hat{\xi}_i)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J_i-1} \rho_b(x_{ij}, \hat{\xi}_i) - r_b(x_{ij+1}, \hat{\xi}_i) \nu_a(\ell_j) - \nu_1(\ell_j) \}, a = 2, ..., A; b = 1, ..., B.
\]
The expected value of taxable income could then be estimated by averaging over the observations \( \hat{\xi}_i \), with
\[
\hat{\mu}(B) = \bar{y}_1 + \sum_{a=2}^{A} \sum_{b=1}^{B} \beta_{ab} \{r_b(x, \hat{\xi}_i)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J_i-1} \rho_b(x_j) - r_b(x_{j+1}) \nu_a(\ell_j) - \nu_1(\ell_j) \}, a = 2, ..., A; b = 1, ..., B.
\]

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
\[
r_b(x, v, \xi) = \rho^{m_1(b)} R^{m_2(b)} \psi^{m_3(b)} v^{m_4(b)} \xi^{m_4(b)} ,
\]
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 \( p^K_i = (p^K_{i1}, ..., p^K_{ik})' \) be the vector of approximating functions, say from equation (9.9) 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 (9.9). 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 $p_{ik}^K$ 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 $\left\{\sum_{i=1}^n (p_{ik}^K)^2/n\right\}^{1/2}$. 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 $(\hat{\varepsilon}_1^h, ..., \hat{\varepsilon}_n^h)$ of size $n$ from the empirical distribution of the residuals.

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

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 $B \times .95$ 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

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

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 $\lambda$ as suggested in g) above corresponding to $c$ closer to 1 and slightly bigger than 2.

k) For each choice of $\lambda$ 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 $\beta_{ab}$ 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.
We note that this approach accounts well for intensive margin effects but does not separately allow for extensive margin effects. For our application to married or cohabitating men we expect small extensive margin effects, so that our approach seems appropriate.

10 Application to Sweden

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 we set medical 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.

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

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 the 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.

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

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
We also checked to see if the Slutsky condition of equation (7.8) was satisfied. We did this by plugging in estimated parameters in that equation and evaluating at $x_c$ and $y_{ld}$ 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 Slutsky conditions therefore hold in the specification with the highest cross validation values and nearby specifications.

The income effect we find is much larger than what is usually reported in studies focusing on taxable income. However, in the labor supply literature one can find studies with similar effects. To get at a concept that is similar to our income effect we have to look at $w = \partial h/\partial R$ for wage rate $w$ and hours of work $h$, which usually is called the marginal propensity to earn out of unearned income MPE. According to Pencavel (1987) the estimates reported in Wales and Woodland (1979) imply a MPE of -.7. The estimates reported in Hausman (1981) imply a MPE of -0.95 to -1.0.

It is a common misperception that the marginal propensity to consume out of unearned income (MPC) and the MPE should add up to one; this relationship only holds if earned income is not taxed. The average marginal tax evaluated at the points on the budget constraints where individuals are located is in our sample around 0.7. Hence, our finding of an MPE of around -1 implies a MPC of around 0.7. Because the slope of the budget constraint is low, it is not very costly in terms of consumption to reduce the taxable income. Given the slope of the budget constraint, the numbers 0.7 and -1 for the MPC and MPE respectively are not unreasonable.

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

<table>
<thead>
<tr>
<th>Growth/Lambda</th>
<th>2</th>
<th>1.5</th>
<th>1</th>
<th>0.5</th>
<th>0.2</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>-0.7</td>
<td>-0.69</td>
<td>-0.69</td>
<td>-0.73</td>
<td>-0.74</td>
<td>0.01</td>
</tr>
<tr>
<td>.1%</td>
<td>-0.85</td>
<td>-0.79</td>
<td>-0.79</td>
<td>-0.93</td>
<td>-0.87</td>
<td>-0.09</td>
</tr>
<tr>
<td>.2%</td>
<td>-0.9</td>
<td>-0.9</td>
<td>-0.9</td>
<td>-1</td>
<td>-0.94</td>
<td>-0.66</td>
</tr>
<tr>
<td>.3%</td>
<td>-0.99</td>
<td>-0.9</td>
<td>-0.9</td>
<td>-1.02</td>
<td>-1.09</td>
<td>-1.12</td>
</tr>
<tr>
<td>.4%</td>
<td>-1.07</td>
<td>-1.07</td>
<td>-1.07</td>
<td>-1.11</td>
<td>-0.86</td>
<td>-1.07</td>
</tr>
<tr>
<td>.5%</td>
<td>-1.14</td>
<td>-1.14</td>
<td>-1.14</td>
<td>-1.17</td>
<td>-1.18</td>
<td>-0.96</td>
</tr>
<tr>
<td>.6%</td>
<td>-1.16</td>
<td>-1.19</td>
<td>-1.23</td>
<td>-1.1</td>
<td>-1.17</td>
<td>-1.2</td>
</tr>
<tr>
<td>.7%</td>
<td>-1.22</td>
<td>-1.23</td>
<td>-1.15</td>
<td>-1.18</td>
<td>-1.19</td>
<td>-1.26</td>
</tr>
<tr>
<td>.8%</td>
<td>-1.26</td>
<td>-1.24</td>
<td>-1.24</td>
<td>-1.18</td>
<td>-1.2</td>
<td>-1.18</td>
</tr>
<tr>
<td>.9%</td>
<td>-1.15</td>
<td>-1.15</td>
<td>-1.15</td>
<td>-1.17</td>
<td>-1.23</td>
<td>-1.17</td>
</tr>
<tr>
<td>1.0%</td>
<td>-1.22</td>
<td>-1.23</td>
<td>-1.23</td>
<td>-1.15</td>
<td>-1.16</td>
<td>-1.15</td>
</tr>
<tr>
<td>1.1%</td>
<td>-1.2</td>
<td>-1.21</td>
<td>-1.11</td>
<td>-1.13</td>
<td>-1.49</td>
<td>-1.23</td>
</tr>
<tr>
<td>1.2%</td>
<td>-1.09</td>
<td>-1.09</td>
<td>-1.09</td>
<td>-1.2</td>
<td>-1.2</td>
<td>-1.11</td>
</tr>
<tr>
<td>1.3%</td>
<td>-1.1</td>
<td>-1.11</td>
<td>-1.37</td>
<td>-1.14</td>
<td>-1.08</td>
<td>-1.08</td>
</tr>
<tr>
<td>1.4%</td>
<td>-1.01</td>
<td>-1.04</td>
<td>-1.04</td>
<td>-1.06</td>
<td>-1.06</td>
<td>-1.06</td>
</tr>
</tbody>
</table>
\[(1 - \rho)y/\rho|e^*\] where \(e^*\) is the compensated elasticity of taxable income. The compensated elasticity satisfies 
\[e^* = e - \rho \ast \partial y/\partial R\] 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) \ast (-1). This estimate is much larger 
than the uncompensated elasticity estimate of .21. It is well within the range of elasticities 
estimated by others. It is substantially larger than in Blomquist and Selin (2010), where 
a similar uncompensated elasticity but smaller income effect was estimated. In comparison, 
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, 
instrumentations, 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>
<thead>
<tr>
<th>Table 4. Net-of-tax elasticity estimates by method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>No controls</td>
</tr>
<tr>
<td>Demo</td>
</tr>
<tr>
<td>Demo+IV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5. Marginal income effects by method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>No controls</td>
</tr>
<tr>
<td>Demo</td>
</tr>
<tr>
<td>Demo+IV</td>
</tr>
</tbody>
</table>

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 vari-
ables. 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. 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.

10.1 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
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.

<table>
<thead>
<tr>
<th></th>
<th>Lasso 0.2%</th>
<th>Lasso 0.3%</th>
<th>Lasso 0.4%</th>
<th>First order</th>
<th>Second order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross income</td>
<td>3.8</td>
<td>2.86</td>
<td>1.96</td>
<td>0.69</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(0.44)</td>
<td>(0.43)</td>
<td>(0.39)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>Tax revenues</td>
<td>0.28</td>
<td>-0.83</td>
<td>-1.89</td>
<td>-3.39</td>
<td>-2.66</td>
</tr>
<tr>
<td></td>
<td>(0.4)</td>
<td>(0.4)</td>
<td>(0.4)</td>
<td>(0.4)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>Mechanical</td>
<td>-4.04</td>
<td>-4.08</td>
<td>-4.1</td>
<td>-4.14</td>
<td>-4.12</td>
</tr>
</tbody>
</table>

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.

11 Conclusion

In this paper we clarify conditions for identification of the effect of taxes on taxable income. We build on the well established result that the presence of bunching provides evidence of tax effects. We find that for a parametric, isoelastic utility function and for a nonparametric utility specification, kinks provide no information about the size of the taxable income elasticity when the distribution of heterogeneity is unrestricted. In contrast, the elasticity can be point identified if the density of heterogeneity across the kink is known except for parameters that come from outside the kink. We also found that a priori nonparametric information can give partial identification of the elasticity. Optimization errors make identification more difficult.

Budget set variation is an important source of identification. We give conditions for identification of the elasticity in an isoelastic specification. We give 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 and imposing or checking all of the restrictions of utility maximization. We apply this approach to Swedish data and find a compensated taxable income elasticity of .53

12 Appendix A: Proofs and Additional Results

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

**Assumption A0**: For each $\eta$, $U(c,y,\eta)$ is continuous in $(c,y)$, increasing in $c$, decreasing in $y$, and strictly quasi-concave in $(c,y)$. Also $y(\rho,R,\eta) < \infty$ and $y(\rho,R,\eta)$ is continuously differentiable in $\rho, R > 0$.

**Assumption A1**: $\eta$ 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 $(\rho,R,\eta)$.

**Assumption A2**: $\eta = (u, \varepsilon)$ for scalar $\varepsilon$ 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 $\varepsilon$, $y(\rho,R,\eta) = y(\rho,R,u,\varepsilon)$ is continuously differentiable in $\varepsilon$, 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, $\varepsilon$ is continuously distributed conditional on $u$, with conditional pdf $f_\varepsilon(\varepsilon|u)$ that is bounded and continuous in $\varepsilon$.

The first result of the Appendix gives useful formulae for the partial derivatives of $F(y|\rho,R)$.
Lemma A1: If Assumptions A0 - 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 \( y, \rho, \) and \( R \) and for the pdf \( f_y(\rho, R, \eta)(y) \) of \( y(\rho, R, \eta) \) at \( y, \)
\[
\frac{\partial F(y|\rho, R)}{\partial y} = f_y(\rho, R, \eta)(y),
\]
\[
\frac{\partial F(y|\rho, R)}{\partial (\rho, R)} = -f_y(\rho, R, \eta)(y)E[\frac{\partial y(\rho, R, \eta)}{\partial (\rho, R)}|y(\rho, R, \eta) = y].
\]

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

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

Lemma A2: If Assumptions A0 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 \( \eta \) argument, which is held fixed in this proof. Let \( y^* = y(B) \). Suppose \( y^* > 0 \). Consider \( y < y^* \) and let \( \tilde{y} \) such that \( y < \tilde{y} < y^* \). Let \((\tilde{c}, \tilde{y})\) be on the line joining \((B(y), y)\) and \((B(y^*), y^*)\). By concavity of \( B(\cdot) \), \( \tilde{c} \leq B(\tilde{y}) \), so by strict quasi-concavity and the definition of \( y^* \),
\[
U(B(y^*), y^*) \geq U(B(\tilde{y}), \tilde{y}) \geq U(\tilde{c}, \tilde{y}) > \min\{U(B(y), y), U(B(y^*), y^*)\} = U(B(y), y).
\]
Thus \( U(B(\tilde{y}), \tilde{y}) > U(B(y), y) \). An analogous argument gives \( U(B(\tilde{y}), \tilde{y}) > U(B(y), y) \) for \( y > \tilde{y} > y^* \). Q.E.D.

Theorem A3: If Assumptions A0 and A1 are satisfied and \( B(y) \) is a concave function then for each \( y \), \( \Pr(y(B, \eta) \leq y) = F(y|\rho(y), R(y)) \).

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

It follows by \( \hat{\rho} \) being a subgradient at \( y \) of \( B(z) \) and \( B(z) \) being concave that for all \( z, \)
\[
\hat{B}(z) \geq B(z).
\]
Therefore $\hat{B}(y^*) \geq B(y^*)$, so that

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

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

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

$$\hat{B} + \hat{p}(z - y^*) \geq \hat{B} + \hat{p}(z - y^*) = \hat{B}(z).$$

so that

$$U(\hat{B}(z), z) \leq U(\hat{B} + \hat{p}(z - y^*), z) = U(\hat{B}(z), z) < U(\hat{B}, y^*).$$

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

Summarizing, we have shown that

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

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

Note that $y^*$ is the utility maximizing point on the budget frontier $B(z)$ while $\hat{y}^*$ is the utility maximizing point on the linear budget frontier $\hat{B}(z) = B(y) + \rho(y)(z - y) = R(y) + \rho(y)z$. Thus, $y^* \leq y \iff \hat{y}^* \leq y$ means that the event $y(B, \eta) \leq y$ coincides with the event that $\arg \max \ni U(R(y) + \rho(y)z, z, \eta) \leq y$, 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.

Theorem A3 can be used to derive identification results for the CDF and conditional expectation of taxable income for a linear budget set. Let $S$ denote a set of budget frontiers and $X(y) = \{(\rho(y), R(y)) : B \in S\}$. 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 $(\rho, R) \in \cap_y X(y)$.

Proof of Theorem 1: Theorem 1 specializes Theorem A3 to piecewise linear budget sets. Q.E.D.
Theorem A3 can be generalized to $B(y)$ that are not everywhere concave. Let $\tilde{B}$ denote the convex hull of the budget set $B = \{(c, y) : 0 \leq c \leq B(y), 0 \leq y\}$ and $\bar{B}(y) = \max_{(c, y) \in B} c$ denote the corresponding budget frontier. By standard convex analysis results $\bar{B}(y)$ will be the smallest concave function such that $\tilde{B}(y) \geq B(y)$ for all $y$. Let

$$\rho(y) = \lim_{z \to y} [\bar{B}(z) - \tilde{B}(y)] / (z - y),$$

$$R(y) = \bar{B}(y) - \rho(y)y,$$

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 A4:** If Assumptions A0 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 $F(y|B) = F(y|\rho(y), R(y))$.

This result generalizes Theorem 1 to $B(y)$ that may be nonlinear and/or nonconcave over some ranges of $y$. In many applications nonconvexities occur only at small values of income. Theorem A4 could be used to nonparametrically quantify how the CDF depends on the budget set at higher values of $y$ where the conditions of Theorem A4 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 A4 to future work.

**Proof of Theorem A4:** We will proceed to prove that $\Pr(y|B, \eta) \leq y = \Pr(y|\bar{B}, \eta) \leq y)$, so the conclusion follows by Theorem A3 and $\bar{B}(y)$ concave. Note that $\bar{B}(z) \geq B(z)$ for all $z$. For notational simplicity suppress the $\eta$ argument and let $U(c, y) = U(c, y, \eta)$. Let

$$y^* \overset{def}{=} \arg\max_z U(B(z), z), \tilde{y}^* \overset{def}{=} \arg\max_z U(\tilde{B}(z), z).$$

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

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

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

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

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

$$y^* \leq y \implies \tilde{y}^* \leq y.$$
Next, suppose that \( y^* > y \). Then there is \( z \in (y, y + \Delta] \) with \( y < z < y^* \)
\[
U(\tilde{B}(y^*), y^*) \geq U(B(y^*), y^*) \geq U(B(z), z) = U(\tilde{B}(z), z).
\]
Again by Lemma A2 we cannot have \( \tilde{y}^* \leq y \) because then \( \tilde{y}^* < z < y^* \) and the above inequality is not consistent with \( U(\tilde{B}(z), z) \) being strictly monotonic decreasing to the right of \( \tilde{y}^* \). Therefore, \( \tilde{y}^* > y \), and we have shown that
\[
y^* > y \implies \tilde{y}^* > y.
\]
Therefore we have \( y^* \leq y \iff \tilde{y}^* \leq y \), so the conclusion follows similarly to the conclusion of Lemma A2. Q.E.D.

The next result makes explicit allowance for nonconcavity of \( B(y) \). It shows that the CDF only depends on \( B(y) \) over wvalues 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 \([y(B), \tilde{y}(B)]\) denote the interval where \( B(y) \) may not be concave and let \( \tilde{B} = \{y(B), \tilde{y}(B), B(z)\}_{z \in [y(B), \tilde{y}(B)]} \) denote the interval endpoints and the budget frontier over the interval.

**Theorem A5:** If Assumption 1 and A1 are satisfied then for all \( B \) such that \( \tilde{B}(z) = B(z) \) except possibly for \( z \in (y(B), \tilde{y}(B)) \) and for \( y \in [y(B), \tilde{y}(B)) \) we have \( \Pr(y(B, \eta) \leq y) \) depends only on \( \tilde{B} \).

**Proof of Theorem A5:** For notational simplicity we suppress the \( \eta \) as before. Let \( \hat{y} = y(B), \tilde{y} = \tilde{y}(B) \) denote the lower and upper bound respectively for the set where \( B(z) \) is not concave. Consider
\[
\tilde{y}^* = \arg \max_{z} U(B(z), z) \text{ s.t. } \hat{y} \leq z \leq \tilde{y}.
\]
By construction the CDF of \( \tilde{y} \) depends only on \( \hat{y}, \tilde{y} \), and \( B(z) \) for \( z \in [\hat{y}, \tilde{y}] \). We will show that for \( y \in [\hat{y}, \tilde{y}] \), 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 \( y \in [\hat{y}, \tilde{y}] \), \( y^* \leq y \) if and only if \( \tilde{y}^* \leq y \). Consider \( y \in [\hat{y}, \tilde{y}] \).

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

Next, suppose that \( y^* > y \). If \( y^* \leq \hat{y} \) then \( \tilde{y}^* = y^* \) so that \( \tilde{y}^*>y \). Suppose \( y^* > \tilde{y} \). Then by Lemma A1 and \( y < \tilde{y} \), for all \( z \in [\hat{y}, \tilde{y}) \) we have
\[
U(B(y^*), y^*) > U(B(\tilde{y}), \tilde{y}) > U(\tilde{B}(z), z) \geq U(B(z), z),
\]
so that \( \tilde{y}^* = \tilde{y} > y \). Therefore \( y^* > y \implies \tilde{y}^* > y \). Summarizing, for \( y \in [\tilde{y}, \tilde{y}) \) we have \( y^* \leq y \iff \tilde{y}^* \leq y \), which proves the result. \( Q.E.D. \)

**Proof of Theorem 2:** Let \( F(y) \) denote the distribution function of taxable income. Let \( \Phi(\eta) = F(\rho_1^1 \eta) \) for \( \eta < \rho_1^1 K \) and let \( \Phi(\eta) = F(\rho_2^2 \eta) \) for \( \eta > \rho_2^2 K \). By Theorem 1, on the lower segment where \( y < K \) the distribution of taxable income will be \( \Pr(\eta \rho_1^1 \leq y) = \Phi(\rho_1^1 y) = F(y) \). Similarly, on the upper segment where \( y > K \), the distribution of taxable income will be \( \Pr(\eta \rho_2^2 \leq y) = \Phi(\rho_2^2 y) = F(y) \). For \( \rho_1^1 K \leq \eta \leq \rho_2^2 K \) let \( \Phi(\eta) \) be any differentiable, monotonic increasing function such that \( \Phi(\rho_1^1 K) = \lim_{y \to K, y < K} F(y) \) and \( \Phi(\rho_2^2 K) = F(K) \). Then by construction, we have

\[
\Phi(\rho_2^2 K) - \Phi(\rho_1^1 K) = F(K) - \lim_{y \to K, y < K} F(a),
\]

where the last equality holds by standard results for cumulative distribution functions. Also, we can choose \( \Phi(\eta) \) so its derivatives of any order match those of \( F(\rho_1^1 \eta) \) at \( \eta = \rho_1^1 K \) and those of \( F(\rho_2^2 \eta) \) at \( \eta = \rho_2^2 K \).

**Proof of Theorem 3:** Let \( F(y) \) be the CDF of \( y(B, \eta) \). By standard probability theory, \( \Pi_\ell = F(K) - \lim_{y \uparrow K} F(y) \). By Theorem A5 \( F(K) = F(K|\rho_2, R_2) \) and \( \lim_{y \uparrow K} F(y) = \lim_{y \uparrow K} F(y|\rho_1, R_1) \). Furthermore, by Assumption A2 \( y(\rho_1, R_1, \eta) \) is continuously distributed so that \( \lim_{y \uparrow K} F(y|\rho_1, R_1) = F(K|\rho_1, R_1) \). Define \( \Lambda(\rho) = F(K|\rho, R(\rho)) \) for \( \rho \in [\rho_2, \rho_1] \). We then have

\[
B = \Lambda(\rho_2) - \Lambda(\rho_1).
\]

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

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

The conclusion then follows by the fundamental theorem of calculus. \( Q.E.D. \)

**Proof of Theorem 4:** Let \( \eta_\ell = y_\ell \rho_1^{\beta} \) and \( \eta_u = y_u \rho_2^{\beta} \) be the endpoints of the bunching interval for \( \eta \) corresponding to \( y_\ell \) and \( y_u \) respectively. The bounds on the density are that for \( \eta \in (\eta_\ell, \eta_u) \),

\[
\sigma \max\{\phi(\eta_\ell), \phi(\eta_u)\} \leq \phi(\eta) \leq \sigma \max\{\phi(\eta_\ell), \phi(\eta_u)\}.
\]

Also, \( \phi(\eta_\ell) \) and \( \phi(\eta_u) \) are given by \( \phi(\eta_\ell) = f^{-}(y_\ell) \rho_1^{\beta}, \phi(\eta_u) = f^{+}(y_u) \rho_2^{\beta} \). The first conclusion of Theorem 2 then follows by

\[
P = \Pr(y_\ell \leq y \leq y_u) = \int_{\eta_\ell}^{\eta_u} \phi(\eta) d\eta \leq (\eta_u - \eta_\ell) \sigma \max\{\phi(\eta_\ell), \phi(\eta_u)\}
\]

\[
= \sigma (y_u \rho_2^{\beta} - y_\ell \rho_1^{\beta}) \max\{f^{-}(y_\ell) \rho_1^{\beta}, f^{+}(y_u) \rho_2^{\beta}\} = \sigma \max\{D^{-}(\beta), D^{+}(\beta)\},
\]

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

$$D^{-}(0) = f^{-}(y)e(y_u - y), \quad D^{+}(0) = f^{+}(y_u)(y_u - y).$$

As long as

$$P \geq \sigma \max\{f^{-}(y), f^{+}(y_u)\}(y_u - y),$$

then by strict monotonicity of $D^{-}(\beta)$ and $D^{+}(\beta)$ in $\beta$ there will be unique $\beta_{\ell}$ and $\beta_u$ satisfying

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

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

$$\sigma \min\{f^{-}(y), f^{+}(y_u)\}(y_u - y) < P < \sigma \max\{f^{-}(y), f^{+}(y_u)\}(y_u - y)$$

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

**Proof of Theorem 5:** Note that $F(y) = \tilde{F}(\tilde{y})$ implies $\Phi(y\rho(y)^{-\beta}) = \Phi(\tilde{y}\tilde{\rho}(\tilde{y})^{-\beta})$, which implies $y\theta(y)^{-\beta} = \tilde{y}\tilde{\rho}(\tilde{y})^{-\beta}$ by $\Phi(\eta)$ strictly monotonic. Taking logs and solving gives the result. Q.E.D.

**Proof of Theorem 6:** Note first that the right slope of the two budget sets are equal, and hence $F(y) = \tilde{F}(\tilde{y})$, at any $y \in [0, K] \cup [K, \infty)$. Also, $F(K) = \Phi(\rho_2^{-b}K)$. Also, $\tilde{F}(\tilde{y}) = \Phi(\rho_2^{-b}\tilde{y})$ for $\tilde{y} < \tilde{K}$ so that $\tilde{F}_{-}(\tilde{K}) = \Phi(\rho_1^{-b}\tilde{K})$. Next, consider the case with $\tilde{F}_{-}(\tilde{K}) \geq F(K)$. Note that for $y \geq K$ we have $F(y) = \Phi(\rho_2^{-b}y)$ so that $F(K) = \Phi(\rho_2^{-b}\tilde{K}) > \Phi(\rho_1^{-b}\tilde{K}) = \tilde{F}_{-}(\tilde{K})$. It follows by continuity and strict monotonicity of $\Phi(\eta)$ that there exists a unique $y^{*} \in [K, \tilde{K})$ with $\Phi(\rho_2^{-b}y^{*}) = F(y^{*}) = \tilde{F}_{-}(\tilde{K}) = \Phi(\rho_1^{-b}\tilde{K})$, the first conclusion. This equality implies $\rho_2^{-b}y^{*} = \rho_1^{-b}\tilde{K}$. Taking logs and solving for $\beta$ gives the second conclusion.

Next, consider the case with $\tilde{F}_{-}(\tilde{K}) < F(K)$. This inequality implies that $\rho_1^{-b}\tilde{K} < \rho_2^{-b}K$. Taking logs gives the third conclusion. Now consider any $b > b$, or equivalently $\rho_1^{-b}\tilde{K} < \rho_2^{-b}K$. Let $\Lambda(\eta) = \tilde{F}(\rho_1^{-b}\eta)$ for $\eta < \rho_1^{-b}\tilde{K}$, $\Lambda(\eta) = F(\rho_2^{-b}\eta)$ for $\eta \geq \rho_2^{-b}K$, and for $\rho_1^{-b}\tilde{K} < \eta < \rho_2^{-b}K$ let $\Lambda(\eta)$ be any increasing, continuous function that is continuous at $\rho_1^{-b}\tilde{K}$ and $\rho_2^{-b}K$. Consider the taxable income distribution when the taxable income elasticity is $b$ and the CDF of heterogeneity is $\Lambda(\eta)$. Note that for the first budget set and $y < K$ we have $Pr(Y \leq y) = \Lambda(\rho_1^{-b}y) = \tilde{F}(y) = F(y)$, while for $y \geq K$, $Pr(Y \leq y) = \Lambda(\rho_2^{-b}y) = F(y)$. Also, for the second budget set and $y < \tilde{K}$ we have $Pr(Y \leq y) = \Lambda(\rho_1^{-b}y) = \tilde{F}(y)$, while for $y \geq \tilde{K} > K$ we have $Pr(Y \leq y) = \Lambda(\rho_2^{-b}y) = F(y) = \tilde{F}(y)$. Thus, when the taxable income elasticity is any $b > b$ we have constructed a heterogeneity distribution where taxable income has the same distribution as the true distribution, for both budget sets. Q.E.D.
Proof of Theorem 7: Let $F_j(y) = F(y|\rho_j, R_j)$. By Theorem 1, the CDF of $y(B, \eta)$ on $(\ell_{j-1}, \ell_j)$ 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

$$\sum_{j=1}^{J-1} \int 1(\ell_{j-1} < y < \ell_j) y F_j(dy) + \int 1(\ell_{J-1} < y) y F_J(dy)$$

$$= \sum_{j=1}^{J-1} \int [1(y < \ell_j) - 1(y \leq \ell_{j-1})] y F_j(dy) + \int 1(\ell_{J-1} < y) y F_J(dy)$$

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

Also, it follows from Theorem 1 that

$$\Pr(y(B, \eta) = \ell_j) = F_{j+1}(\ell_j) - \lim_{y\to\ell_j} F_j(y) = \int 1(y \leq \ell_j) F_{j+1}(dy) - \int 1(y < \ell_j) F_j(dy) \quad \text{(4.10)}$$

Combining these results we have

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

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

To show the second conclusion, note that by $\ell_0 = 0$ we have $\bar{y}(\rho_1, R_1) = \int 1(y > \ell_0) y F_1(dy)$. Then it follows that

$$\sum_{j=1}^{J-1} \int 1(\ell_{j-1} < y < \ell_j) y F_j(dy) + \int 1(\ell_{J-1} < y) y F_J(dy)$$

$$= \sum_{j=1}^{J-1} \int [1(y > \ell_{j-1}) - 1(y \geq \ell_j)] y F_j(dy) + \int 1(y > \ell_{J-1}) y F_J(dy)$$

$$= \bar{y}(\rho_1, R_1) + \sum_{j=1}^{J-1} \left[ \int 1(y > \ell_j) y F_{j+1}(dy) - \int 1(y \geq \ell_j) y F_j(dy) \right].$$
Combining this with the second equality in eq. (12.10) then gives

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

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

Here we give a rate result for the approximation for the conditional mean for specific types of CDF’s \( F_\alpha(y) \) 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 \( F_\alpha(y) \) 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 A6:** 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 \( dF_\alpha(y)/dy \), \( (a = 1, ..., A) \) 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 \( \beta_{ab} \) such that for all piecewise linear, concave budget frontiers with \( x_j \in \mathcal{X} \), \( (j = 1, ..., J) \),

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

This result shows that the series approximation we have proposed does indeed approximate the expected value of taxable income for concave, piecewise \( B(y) \). 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.

**Proof of Theorem A6:** By Edmunds and Evans (1989) there exists \( C \) such that for each \( A \) and \( B \) there is \( (\beta_{ab}) \) such that for \( \varepsilon = CA^{1-s}B^{(1-s)/2} \) and \( p^{AB}(y, x) = \sum_{a=1}^{A} \sum_{b=1}^{B} \beta_{ab} f_a(y) r_b(x) \),

\[
\sup_Z |f(y|x) - p^{AB}(y, x)| + \sup_Z |f_p(y|x) - p^{AB}(y, x)| + \sup_Z |f_R(y|x) - p^{AB}(y, x)| \leq \varepsilon,
\]

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

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

[48]
Note that by $\int f_a(y)dy = 1$ for each $a$ and by $\mathcal{Y}$ bounded,

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

Also,

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

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

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

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

$$|\bar{y}(x) - \bar{y}^{AB}(x)| = \left| \int y[f(y|x) - p_{AB}(y, x)]dy + [1 - \sum_{a=1}^{A} w_a(x, \beta)] \int y f_1(y)dy \right| \leq \int y |f(y|x) - p_{AB}(y, x)| dy + [1 - \sum_{a=1}^{A} w_a(x, \beta)] \int y f_1(y)dy \leq C\varepsilon.$$ 

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 $\mathcal{Y}$, $|\int 1(y < \ell)(y-\ell)f_1(y)dy| \leq C$. Therefore,

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

Therefore we have

\[ \sum_{j=1}^{J-1} [\nu(x_j, \ell_j) - \nu(x_{j+1}, \ell_j)] - \sum_{j=1}^{J-1} [\nu^{AB}(x_j, \ell_j) - \nu^{AB}(x_{j+1}, \ell_j)] \]

\[ = \sum_{j=1}^{J-1} [\Delta(x_j, \ell_j) - \Delta(x_{j+1}, \ell_j)] = \sum_{j=1}^{J-1} \left| \frac{\partial \Delta(x_j, \ell_j)}{\partial x} T(x_j - x_{j+1}) \right| \]

\[ \leq C \left( \sup_{\mathcal{X}} \left| \frac{\partial \Delta(x_j, \ell_j)}{\partial \rho} \right| + \sup_{\mathcal{X}} \left| \frac{\partial \Delta(x_j, \ell_j)}{\partial R} \right| \right) \sum_{j=1}^{J-1} [\rho_j - \rho_{j+1} + R_{j+1} - R_j] \]

\[ \leq C\varepsilon[\rho_1 - \rho_J + R_J - R_1] \leq C\varepsilon. \]
To conclude the proof, note that

\[
\sum_{j=1}^{J-1} [\nu^AB(x_j, \ell_j) - \nu^AB(x_{j+1}, \ell_j)] = \sum_{j=1}^{J-1} \left[ \int (y < \ell_j)(y - \ell_j) [p^AB(y, x_j) - p^AB(y, x_{j+1})] dy \right]
\]

\[
= \sum_{j=1}^{J-1} \int (y < \ell_j)(y - \ell_j) \sum_{a=2}^{A} [w_a(x_j, \beta) - w_a(x_{j+1}, \beta)][f_a(y) - f_1(y)] dy
\]

\[
= \sum_{a=2}^{A} \sum_{j=1}^{J-1} [w_a(x_j, \beta) - w_a(x_{j+1}, \beta)][\nu_a(\ell_j) - \nu_1(\ell_j)]
\]

\[
= \sum_{a=2}^{A} \sum_{b=1}^{B} \sum_{j=1}^{J} [r_b(x_j) - r_b(x_{j+1})][\nu_a(\ell_j) - \nu_1(\ell_j)].
\]

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 \(|\bar{y}(x) - \bar{y}^AB(x)|\) and \(\sum_{j=1}^{J-1} [\nu(x_j, \ell_j) - \nu(x_{j+1}, \ell_j)] - \sum_{j=1}^{J-1} [\nu^AB(x_j, \ell_j) - \nu^AB(x_{j+1}, \ell_j)]\). Q.E.D.

13 Appendix B: Stochastic Revealed Preference with Two Goods and Nonlinear Budget Sets.

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 \(\rho\) 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 \((c \text{ and } y)\) 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 B1:** If Assumptions A0-A2 are satisfied then \(F(y|\rho, R)\) is continuously differentiable in \(\rho\) and \(R\) and

\[
F_\rho(y|\rho, R) - yF_R(y|\rho, R) \leq 0.
\]  

(13.11)
Also, if for all $\rho, R > 0$, $F(y|\rho, R)$ is continuously differentiable in $y$, $\rho$, $R$, the support of $F(y|\rho, R)$ is $[y_t, y_u]$, $\partial F(y|\rho, R)/\partial y > 0$ on $(y_t, y_u)$, and equation (13.11) 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 B1 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.

Theorem 1 implies a revealed stochastic preference result for convex budget sets. As shown by Theorem B1, for linear budget sets and preference satisfying the conditions of Assumptions A0 and A1, a necessary and sufficient condition for a RUM is that the CDF satisfy the Slutzky condition. An implication of Theorem 1 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.

**Proof of Theorem B1:** Note that $y(\rho, R, \eta)$ is differentiable by Assumption 1. Also, $\rho$ behaves like the negative of a price (increasing $\rho$ 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 $\eta$ and all $\rho, R > 0$. Then by Lemma A1 $F(y|\rho, R)$ is differentiable in $\rho$ and $R$ and

$$\frac{\partial F(y|\rho, R)}{\partial \rho} - y \frac{\partial F(y|\rho, R)}{\partial R} = -f_y(\rho, R, \eta)(y) \left\{ E[\frac{\partial y(\rho, R, \eta)}{\partial \rho}|y(\rho, R, \eta) = y] - y \cdot E[\frac{\partial y(\rho, R, \eta)}{\partial R}|y(\rho, R, \eta) = y] \right\}$$

$$= f_y(\rho, R, \eta)(y) E[\frac{\partial y(\rho, R, \eta)}{\partial \rho} + y(\rho, R, \eta) \frac{\partial y(\rho, R, \eta)}{\partial R}|y(\rho, R, \eta) = y] \leq 0,$$

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 $(y_t, y_u)$ and $[y_t, y_u]$ being the support of
By the inverse function theorem, for all \( \rho, R > 0 \),
\[
- \frac{\partial Q(\tau | \rho, R)}{\partial \rho} + Q(\tau | \rho, R) \frac{\partial Q(\tau | \rho, R)}{\partial R} = f_{y(\rho, R, \eta)}(Q(\tau | \rho, R))^{-1} \left\{ \frac{\partial F(Q(\tau | \rho, R) | \rho, R)}{\partial \rho} - Q(\tau | \rho, R) \frac{\partial F(Q(\tau | \rho, R) | \rho, R)}{\partial R} \right\} \leq 0.
\]
Therefore it follows by Hurwicz and Uzawa (1971) that for each \( \tau \) with \( 0 < \tau < 1 \) there is a utility function \( U(c, y, \tau) \) with for all \( \rho, R > 0 \),
\[
Q(\tau | \rho, R) = \arg \max_{c,y} U(c, y, \tau) \text{ s.t. } c = y\rho + R, c \geq 0, y \geq 0.
\]
Let \( \eta \) 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.

Theorem B1 can also be used to relate positivity of the kink probability \( \Pi_\ell \) 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 (3.5). Then the Slutzky condition is sufficient but not necessary for positivity of \( \Pi_\ell \), 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 A 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
\[
\frac{\partial F(y | \rho(y), R(y))}{\partial y} = F_y(y | \rho(y), R(y)) + \rho_y(y) [F_\rho(y | \rho(y), R(y)) - yF_R(y | \rho(y), R(y))],
\] (13.12)
where \( \rho_y(y) = \frac{\partial \rho(y)}{\partial y} \). One could specify a CDF \( F(y | \rho, R) \) for taxable income for a linear budget set and derive the pdf from equation (13.12). 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 Slutsky 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 Slutsky condition, which could be attributed to our nonparametric framework.

14 Appendix C: Effort and Accounting for Productivity Growth

To motivate the way we allow for productivity growth we embed the utility model in one where effort enters the utility. 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 $\tau$ 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

$$\max_{c, e, h} u(c, e, h) \quad s.t. \quad c = w(e)h\rho + R, c \geq 0, e \geq 0, h \geq 0. \quad (14.13)$$

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

$$\max_{c, y} U(c, y) \quad s.t. \quad c = y\rho + R, c \geq 0, y \geq 0. \quad (14.14)$$

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

[53]
In the taxable income literature one usually starts with individual choice of consumption $c$ and taxable income $y$ as given by equation (14.14). 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 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:

$$\begin{align*}
\text{Max } u(c, e, h, \eta) & \quad \text{s.t. } c = g(e, \eta)\phi(t)h\rho + R \\
\end{align*}$$

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 $\text{Max } U(c, y/\phi(t), \eta)$ s.t. $c = y\rho + R$.

15 Appendix D: Sample Statistics Sweden

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

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

16 References


metrica 79: 651–692.


Layard, eds., Handbook of Labor Economics Volume 1, Amsterdam: North-Holland, 3-102.


