Individual heterogeneity, nonlinear budget sets, and taxable income

Soren Blomquist
Anil Kumar
Che-Yuan Liang
Whitney Newey

Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income[*]

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

Anil Kumar
Dallas Federal Reserve

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

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

Abstract

Given the key role of the taxable income elasticity in designing an optimal tax system there are many studies attempting to estimate this elasticity. To account for nonlinear taxes these studies either use instrumental variables approaches that are not fully consistent, or impose strong functional form assumptions. None allow for general heterogeneity in preferences. In this paper we derive the mean and distribution of taxable income, conditional on a nonlinear budget set, allowing general heterogeneity and optimization errors for the mean. We find an important dimension reduction and use that to develop nonparametric estimation methods. We show how to nonparametrically estimate the conditional mean of taxable income imposing all the restrictions of utility maximization and allowing for measurement errors. We apply this method to Swedish data and estimate for prime age males a significant net of tax elasticity of 0.6 and a significant income elasticity of -0.08.

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

Keywords: Nonlinear budget sets, nonparametric estimation, heterogeneous preferences, taxable income, revealed stochastic preference.

1   Introduction

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

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

We nonparametrically identify and estimate taxable income effects with nonlinear budget sets while allowing for general heterogeneity. The heterogenous preferences are assumed to be strictly convex and statistically independent of the budget set, but are otherwise unrestricted. We also allow for optimization errors in estimation of the conditional mean. We find that the conditional mean of taxable income can be estimated by a low dimensional nonparametric regression. One form of this regression is exactly analogous to the conditional mean of labor supply from Blomquist and Newey (2002, BN henceforth), which was derived when preference heterogeneity can be represented by a scalar. Consequently, it turns out that the labor supply results of BN are valid under general preference heterogeneity. However the BN form did not impose all the restrictions imposed by utility maximization and we show how to do this.

We also derive the distribution of taxable income at points where the budget frontier is

---

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

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

The model here is like the revealed stochastic preference model of McFadden (2005) in having general heterogenous preferences and budget sets that are statistically independent of one another. We differ in considering only the two good case with strictly convex preferences. We find that for two goods, smooth strictly convex preferences, and convex budget sets, necessary and sufficient conditions for utility maximization are that the CDF given the budget set is a CDF for a linear budget set, satisfying a Slutzky property, and evaluated at (right) slope and virtual income. This result advances a program suggested by McFadden (2005), of studying stochastic revealed preference under restrictions like strictly convex preferences.

In independent work Manski (2013) has recently considered labor supply in the revealed stochastic preference setting where budget sets are independent of a finite number of preference types. We do not restrict heterogeneity to a finite number of types but do impose strictly convex preferences. We find that, because of the dimension reduction, nonparametric regression can be used to predict the effect of a change in tax structure on average taxable income within the sample range of budget sets. In fact, because the method of BN is valid with general heterogeneity, such nonparametric prediction was already carried out in the application of BN to labor supply. Thus, the nonparametric method we consider has and can be used to estimate interesting policy effects in a revealed stochastic preference setting. Where it applies, this approach leads to a simpler and more optimistic view of nonparametric policy analysis than in Manski (2013). We find that nonparametric policy analysis is practical for some policies under stochastic revealed preference.

To evaluate the effect of taxes on taxable income we focus on elasticities that apply to changes in nonlinear tax systems. Real-world tax systems are non-linear, and it is variations in non-linear tax systems that we observe. Therefore, it is easiest to nonparametrically identify elasticities relevant for changes in nonlinear tax systems. BN did show that with labor supply it may be possible to identify labor supply elasticities for changes in linear budget sets, but for taxable income we find that the conditions for identifying average elasticities for a linear budget set are very stringent and not likely to be satisfied in applications. Here we propose

elasticities defined by an upward shift of the non-linear budget constraint, in either slope or intercept. These elasticities are relevant for changes in non-linear budget constraints. We find that these can be estimated with high accuracy in our application.

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

Our application is to Swedish data from 1993-2008 with third party reported taxable *labor* income. This means that the variation in the taxable income that we observe for Sweden is mainly driven by variations in effort broadly defined and variations in hours of work and not by variations in tax evasion.[2] We estimate a statistically significant tax elasticity of 0.60 and a significant income elasticity of -0.08.

The rest of our paper is organized as follows. Section 2 reviews the taxable income literature. Section 3 lays out a model of individual behavior where there are more decision margins than hours of work. Section 4 analyzes kinks and Section 5 derives the distribution of taxable income conditional on the budget set. Section 6 describes the policy effects we consider. In Section 7 we show how to estimate the conditional mean imposing all the restrictions implied by utility maximization. In section 8 we describe the Swedish data we use and present our estimates. Section 9 concludes.

2 Previous literature

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

---

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

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 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 Tax Reform Act of 1988. They found taxable income elasticity ranging from 0.14-1.30.

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

More recent studies have also estimated low taxable income elasticities. Kopczuk (2005) used the University of Michigan tax panel to yield an estimate of -0.2-0.57. More recently Eissa and Giertz (2006) used the Treasury tax panel from 1992-2003 and data from executive compensation. They used variation from multiple tax reforms during this period –TRA 1986, OBRA and EGTRRA on a sample of executives and 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 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 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.

3   The Model

Feldstein (1995) argued that individuals have more margins than hours of work to respond to changes in the tax. For example, they 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. It would distract from the main focus of the paper to deal with all these margins. Our Swedish data is such that we do not need to worry about tax evasion. However, we believe it yields important insights to consider a model where the individual chooses both hours of work and effort, where the effort level affects the hourly wage rate. Effort can be quite broadly interpreted.

In the taxable income literature it is common to start with a utility function $U(c, y)$, where c is consumption and y is taxable income, assumed to be quasi-concave. We believe it is of value to study the underlying problem that generates a reduced-form utility function $U(c, y)$, as this gives us insight about properties of the function $U(c, y)$. We start by considering an individual's utility optimization problem given a linear tax system. Let e and h denote effort and hours of work. We write the utility maximization problem as

$$\text{Max}\, u(c, e, h) \quad s.t. \quad c = g(e)h\rho + R \tag{3.1}$$

Here $\rho = 1 - \tau$ is the net-of-tax rate (for tax rate τ) and R is nonlabor income. For now we assume that $w = g(e)$. In this formulation, u is strictly decreasing in e and h. We assume that u is strictly quasi-concave in its arguments.

The problem can be solved in the following way. Let $y = wh$, implying that $w = y/h$. Inverting the wage function, we get $e = g^{-1}(w) = g^{-1}(y/h)$. For given c and y we chose h to maximize $u(c, g^{-1}(y/h), h)$. This gives h as a function of c and y. Sticking this function back into the direct utility function we get the reduced-form utility function $U(c, y)$. In a second step the individual solves

$$\text{Max}\, U(c, y) \quad s.t. \quad c = y\rho + R \tag{3.2}$$

This gives the taxable income function $y(\rho, R)$.

In the empirical taxable income literature one usually starts out with the utility optimization problem as given by equation (3.2), taking for granted that $U(c, y)$ is quasi-concave. Generally, quasi-concavity of $u(c, e, h)$ does not imply quasi-concavity of $U(c, y)$, because the budget

constraint is nonlinear in e and h. Nevertheless, we will follow the literature in assuming that $U(c, y)$ is quasi-concave.

We allow for general heterogeneity that affects both preferences and wages. To describe this set up let η denote a vector valued random variable of any dimension that represents individual type. We specify the utility function of an individual as $u(c, e, h, \eta)$ and the wage rate as $w = g(e, \eta)$. We impose no restriction on how η enters the utility or wage function, thus allowing for distinct heterogeneity in both preferences and ability, with different components of η entering u and g. The individual's optimization problem for a linear budget set is now

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

The problem can be solved similar to before. Holding fixed c and $y = wh$ fixed, solving for e from the budget constraint, and then maximizing over h we obtain a reduced-form utility function $U(c, y, \eta)$. The η in this function includes heterogeneity in the original utility function as well as ability terms from the wage function. The optimal consumption and income are obtained by maximizing $U(c, y, \eta)$ s.t. $c = y\rho + R$. This is the same choice problem as before except that the reduced form utility function U now depends on η.

This specification allows for preferences to vary across individuals in essentially any way at all. For example income and level effects can vary separately (as in Burtless and Hausman, 1978), income and tax elasticities are not restricted to be on a one-dimensional curve (unlike BN), and the number of types is not restricted to be finite (unlike Manski, 2013).We do need to specify η so that probability statements can be made but these are technical side conditions that do not affect our interpretation of η as representing general heterogeneity.

In applications the tax rates vary with income, corresponding to nonlinear budget sets. A continuous, piecewise-linear budget set with J segments, indexed by j, can be described by a vector $(\rho_1, \ldots, \rho_J, R_1, \ldots, R_J)$ of net-of-tax rates ρ_j (slopes) and virtual incomes R_j (intercepts). It will have kink points $\ell_0 = 0$, $\ell_J = \infty$, $\ell_j = (R_{j+1} - R_j)/(\rho_j - \rho_{j+1})$, $(1 \leq j \leq J - 1)$. In what follows we will also give some results for the case where budget sets need not be piecewise linear. Throughout we will denote the net income function by $B(y)$, it being the amount that can be consumed net of taxes, including unearned income. The budget set of the individual is $B = \{(c, y) : 0 \leq c \leq B(y), y \geq 0\}$. Under the conditions we impose, the choice of y for individual η will be

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

We impose the following condition on preferences and how they vary in the data we observe. We denote an outcome for the budget set, heterogeneity, and taxable income by B_i, η_i, and y_i,

where i indexes the observations. Our notation for random variables is an i subscript, consistent with the convention that outcomes are random variables.

ASSUMPTION 1: *i) η_i is contained in a separable metric space and $U(c, y, \eta)$ is continuous in (c, y, η); ii) $U(c, y, \eta)$ is increasing in c, decreasing in y, and strictly quasi-concave in (c, y); iii) the data $(y_i, B_i), (i = 1, \ldots, n)$ are identically distributed and B_i and η_i are statistically independent;*

The conditions in i) are used to ensure that probability statements concerning η and $y(B, \eta)$ are well defined. Part ii) imposes strict monotonicity and convexity of preferences. Independence of budget sets and heterogeneity in part iii) is important for obtaining information about price and income effects from cross section data. For piecewise linear budget sets this condition requires that the number of budget segments, the tax rates, and the intercepts for the budget sets be distributed independently of tastes and ability. Independence of budget and preferences encompasses a statistical version of a hypothesis of utility theory that preferences do not vary with the budget set. It is the working hypothesis of the revealed stochastic preference literature, see McFadden (2005). We can replace full independence with independence conditional on control functions and a rank condition, as discussed below.

4   Kinks and Nonparametric Compensated Tax Effects

Kinks have been used by Saez (2010) to identify compensated tax effects for small kinks or parametric models. In this section we derive the nonparametric form of a kink probability with general heterogeneity and show how that is related to compensated effects. The form of the kink will also be used later to derive the conditional mean of taxable income.

The taxable income when the budget set is linear is an important building block for the analysis to follow. Let $y(\rho, R, \eta_i)$ denote the taxable income when the budet frontier is linear with slope ρ and intercept R, i.e. when $B(y) = R + \rho \cdot y$. Also let $F(y|\rho, R) = \Pr(y(\rho, R, \eta_i) \leq y)$ denote the CDF of taxable income for a linear budget set and $f_{\rho,R}(y)$ the corresponding pdf.

We will derive the kink probability for a convex budget set with piecewise linear frontier. For such a budget set the budget frontier $B(y)$ will be piecewise linear and concave. Denote the slope and intercept of the jth segment of $B(y)$ by ρ_j and R_j respectively. By concavity of $B(y)$ we have $\rho_j > \rho_{j+1}$, corresponding to increasing marginal tax rates.

Hausman (1979) gives an important characterization of taxable income for such budget sets that is useful for our purposes. Let $y_j(\eta_i) = y(\rho_j, R_j, \eta_i), (j = 1, \ldots, J)$ be the taxable income when the budget set is linear with slope ρ_j and intercept R_j. Then by Hausman (1979) we know

that taxable income $y(B, \eta_i)$ satisfies $y(B, \eta_i) = y_j(\eta_i)$ for the j such that $\ell_{j-1} < y_j(\eta_i) < \ell_j$ and $y(B, \eta_i) = \ell_j$ for $y_j(\eta_i) \geq \ell_j, y_{j+1}(\eta_i) \leq \ell_j$.

This analysis implies that the probability of the kink point ℓ_j is $\Pr(y_j(\eta_i) \geq \ell_j, y_{j+1}(\eta_i) \leq \ell_j | B_i)$. This appears to be a two-dimensional integral. However, by revealed preference and strict monotonicity of preferences, $y_j(\eta_i) < \ell_j$ and $y_{j+1}(\eta_i) > \ell_j$ cannot occur for the same individual. As a result the probability of a kink point is a difference of two one dimensional integrals.

LEMMA 1: *If Assumption 1 is satisfied and B_i is piecewise-linear and concave then*

$$
\begin{aligned}
\Pr(y(B_i, \eta_i) \;\; &= \;\; \ell_j | B_i) = \Pr(y_{j+1}(\eta_i) \leq \ell_j | B_i) - \Pr(y_j(\eta_i) < \ell_j | B_i) \\
&= \;\; F(\ell_j | \rho_{j+1}, R_{j+1}) - \lim_{y \uparrow \ell_j} F(y | \rho_j, R_j).
\end{aligned}
$$

We can use this result to relate the kink probability to compensated effects. To do so, assume that $F(y | \rho, R)$ is one-to one on the support of $y(\rho, R, \eta_i)$ and that all other needed regularity conditions for the following discussion are satisfied (including existence of various derivatives). Then for $0 < \tau < 1$, the τth conditional quantile of of $y(\rho, R, \eta_i)$ given ρ and R is $F^{-1}(\tau | \rho, R)$. Note that $y(\rho, R, \eta_i)$ is a nonseparable function of a multi-dimensional disturbance η_i that is independent of ρ and R, so by Hoderlein and Mammen (2007),

$$
\frac{\partial F^{-1}(\tau | \rho, R)}{\partial \rho} = E[\frac{\partial y(\rho, R, \eta_i)}{\partial \rho} | y(\rho, R, \eta_i) = F^{-1}(\tau | \rho, R)].
$$

A corresponding equation also holds for the derivative with respect to R. Then by the inverse function theorem,

$$
\begin{aligned}
\frac{\partial F(y | \rho, R)}{\partial \rho} &= -f_{\rho, R}(y) E[\frac{\partial y(\rho, R, \eta_i)}{\partial \rho} | y(\rho, R, \eta_i) = y], &\qquad (4.3) \\
\frac{\partial F(y | \rho, R)}{\partial R} &= -f_{\rho, R}(y) E[\frac{\partial y(\rho, R, \eta_i)}{\partial R} | y(\rho, R, \eta_i) = y].
\end{aligned}
$$

One implication of these formulae is that $F(y | \rho, R)$ has a Slutzky property. Utility maximization implies that the compensated net of tax derivative, given by $D(\rho, R, \eta_i) = \partial y(\rho, R, \eta_i)/\partial \rho - y(\rho, R, \eta_i)\partial y(\rho, R, \eta_i)/\partial R$, is nonnegative, which is the Slutzky condition for taxable income. From the previous equation it then follows that

$$
\frac{\partial F(y | \rho, R)}{\partial \rho} - y\frac{\partial F(y | \rho, R)}{\partial R} = -f_{\rho, R}(y) E[D(\rho, R, \eta_i) | y(\rho, R, \eta_i) = y] \leq 0. \qquad (4.4)
$$

This is a Slutzky property of the CDF, that the familiar linear combination of derivatives with respect to ρ and y is nonpositive. Dette, Hoderlein, and Nuemeyer (2011) previously showed

[8]

that the quantile function satisfies the Slutzky condition. The above derivation of the Slutzky property of the CDF is really a combination of their result and the inverse function theorem.

As discussed in Hausman and Newey (2013), the Slutzky condition for the quantile $F^{-1}(\tau|\rho, R)$ is all the restrictions implied by utility maximization with linear budget sets and two goods. The above reasoning shows that the Slutzky property of the CDF is equivalent to the Slutzky condition for quantiles over all ρ, R, y with $f_{\rho,R}(y) > 0$. As we show below, the conditional CDF with convex budget sets depends only on $F(y|\rho, R)$, so the Slutzky property of $F(y|\rho, R)$ also gives all the restriction implied by utility maximization with convex budget sets. We will use this result to develop an estimator for the conditional mean of taxable income with convex piecewise linear budget sets that imposes all the restrictions of utility maximization.

The CDF derivatives imply an interesting formula for the kink. Differentiating the kink probability with respect to ρ_{j+1} while setting $R_{j+1} = R_j + (\rho_j - \rho_{j+1})\ell_j$ to hold the kink fixed, gives

$$\left.\frac{\partial \Pr(y(B_i, \eta_i) = \ell_j | B_i = B)}{\partial \rho_{j+1}}\right|_{R_{j+1}=R_j+(\rho_j-\rho_{j+1})\ell_j.} = \frac{\partial F(\ell_j | \rho_{j+1}, R_{j+1})}{\partial \rho_{j+1}} - \ell_j \frac{\partial F(\ell_j | \rho_{j+1}, R_{j+1})}{\partial R_{j+1}}$$

$$= -f_{\rho_{j+1}, R_{j+1}}(\ell_j) E[D(\rho_{j+1}, R_{j+1}, \eta_i)| y(\rho_{j+1}, R_{j+1}, \eta_i) = \ell_j]$$

by Lemma 1 and equation (4.3). For $\rho_j \geq \rho \geq \rho_{j+1}$ let

$$\bar{D}(\rho) = E[D(\rho, R_j + (\rho_j - \rho)\ell_j, \eta_i)| y(\rho, R_j + (\rho_j - \rho)\ell_j, \eta_i) = \ell_j], w(\rho) = f_{\rho, R_j+(\rho_j-\rho)\ell_j}(\ell_j)$$

denote the average compensated net of tax effect and the pdf for individuals at the kink. Then by the fundamental theorem of calculus (letting ρ_{j+1} vary) and by the above expression for the derivative of the kink probability we have

$$\Pr(y(B_i, \eta_i) = \ell_j | B_i = B) = \int_{\rho_{j+1}}^{\rho_j} w(\rho) \bar{D}(\rho) d\rho. \tag{4.5}$$

Thus we see that the kink probability is an integral over the net of tax rate of a pdf times an average compensated effect. Hence the kink probability is determined by two features of the distribution of preferences across individuals. These two features are compensated tax effects for individuals at the kink (quantified by $\bar{D}(\rho)$) and the density of individuals at the kink point (quantified by $w(\rho)$).

A relation between the kink probability and compensated effects had previously been derived by Saez (2010) for small kinks or for a particular functional form. Equation (4.5) is a nonparametric formula for kinks of any size with general heterogeneity. We do see that the kink probability depends on compensated tax effects even with large kinks and income effects. We also find that the compensated effect is the average only for those individuals who locate

[9]

at the kink. Importantly, we can also see that the relationship between the kink probability and compensated effects depends on the size of $w(\rho)$. This $w(\rho)$ can be thought of as a pure heterogeneity effect. It is possible that some individuals may be located at a kink because they like it there, not because of the size of the average compensated effect.

It is clear from this discussion that recovering compensated effects from kinks requires knowing $w(\rho)$. Unfortunately, even when $f_{\rho,R}(y)$ is continuous in y, $w(\rho)$ is only identified at ρ_j and ρ_{j+1}, as $w(\rho_j) = \lim_{y\uparrow\ell_j} f_{\rho_j,R_j}(y)$ and $w(\rho_{j+1}) = \lim_{y\downarrow\ell_j} f_{\rho_{j+1},R_{j+1}}(y)$. In between the endpoints $w(\rho)$ is not identified for a given budget set B, because it is the density of taxable income at net of tax rates that are not observed for budget set B. Therefore, nonparametrically recovering compensated income effects from kinks necessarily requires untestable assumptions about $w(\rho)$.

One can make assumptions about $w(\rho)$ that allow recovery of interesting compensated effects from kink probabilities. For example, one might assume that $w(\rho)$ is linear on $[\rho_{j+1}, \rho_j]$, in which case $\int_{\rho_{j+1}}^{\rho_j} w(\rho)d\rho = (\rho_j - \rho_{j+1})[f_{\rho_j,R_j}(\ell_j) + f_{\rho_{j+1},R_{j+1}}(\ell_j)]/2$. Then

$$\frac{\int_{\rho_{j+1}}^{\rho_j} w(\rho)\bar{D}(\rho)d\rho}{\int_{\rho_{j+1}}^{\rho_j} w(\rho)d\rho} = \frac{\Pr(y(B_i,\eta_i) = \ell_j | B_i = B)}{(\rho_j - \rho_{j+1})[f_{\rho_j,R_j}(\ell_j) + f_{\rho_{j+1},R_{j+1}}(\ell_j)]/2}.$$

In this example the kink probability divided by a normalization factor gives a weighted average of compensated effects. However, this interpretation depends critically on the assumed linearity of $w(\rho)$. Other assumptions would lead to other normalizations that would be needed to obtain a weighted average of compensated effects from the kink probability.

Identifying compensated effects from kinks has some similarities and differences to identifying treatment effects from regression discontinuity design. A kink only identifies average compensated effects for individuals at the kink, just as regression discontinuity only identifies average treatment effects for individuals at the discontinuity. Also, identification of $w(\rho)$ at the endpoints rests on continuity of $f_{\rho,R}(y)$, like identification of average treatment effects rests on continuity of the conditional expectation of the outcome with and without treatment. Unlike regression discontinuity, nonparametrically recovering compensated effects from a kink requires assumptions about how $w(\rho)$ behaves on (ρ_{j+1}, ρ_j), a range of net of tax rates that may not be observed in the data.

This analysis does not account for the possibility of optimization errors. Saez (2010) shows how to do that by replacing the kink probability with the probability of "excess bunching." We will also allow for optimization errors in the conditional mean specification we consider, in a different way.

[10]

5   The Conditional Mean and Distribution of Taxable Income

The conditional mean and distribution of taxable income are useful for identifying important policy effects from variation in budget sets, as we discuss in the next Section. In this section we derive the conditional mean and distribution. Recall that $F(y|\rho, R)$ is the CDF of taxable income for a linear budget set. Define

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

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

THEOREM 2: *If Assumption 1 is satisfied, $B(y)$ is piecewise linear and concave, $y_i = y(B_i, \eta_i) + \varepsilon_i, E[\varepsilon_i|B_i] = 0$, and $E[|y(\rho, R, \eta_i)|] < \infty$ for all (ρ, R) then*

$$
\begin{aligned}
E[y_i|B_i] &= \bar{y}(\rho_{J_i i}, R_{J_i i}) + \sum_{j=1}^{J_i - 1} [\mu(\rho_{ji}, R_{ji}, \ell_{ji}) - \mu(\rho_{j+1,i}, R_{j+1,i}, \ell_{ji})] \qquad (5.6) \\
&= \bar{y}(\rho_{1i}, R_{1i}) + \sum_{j=1}^{J_i - 1} \left[\lambda(\rho_{j+1,i}, R_{j+1,i}, \ell_{ji}) - \lambda(\rho_{ji}, R_{ji}, \ell_{ji}) \right].
\end{aligned}
$$

The first equality in the conclusion is exactly analogous to the conclusion of Theorem 2.1 of BN. As discussed there, the additive decomposition of the conditional mean makes nonparametric estimation feasible. The fact that the conditional expectation only depends on one two-dimensional function $\bar{y}(\rho, R)$ and one three-dimensional function $\mu(\rho, R, \ell)$ (or $\lambda(\rho, R, \ell)$) means the curse of dimensionality can be avoided. In fact, the conditional mean depends only on $F(y|\rho, R)$, because each of $\bar{y}$, λ, and μ do. This fact makes the conditional mean even more parsimonious. We will describe below how dependence of the conditional mean on just $F(y|\rho, R)$ can be used in estimation.

Theorem 2 generalizes Theorem 2.1 of BN in several ways. First, and most important, Theorem 2 is valid with general heterogeneity, whereas BN only derived the additive decomposition for scalar η. Theorem 2 also generalizes BN by allowing for zero taxable income (corresponding to zero hours of work), though it does not allow for other kinds of sample selection. Thus, Theorem 2 shows that the analysis of BN is correct with general heterogeneity and zero hours of work, and not just the scalar heterogeneity they considered. Consequently, the BN empirical results are valid with general heterogeneity, including the policy evaluation of the effect of Swedish tax reform on average labor supply they carried out.

[11]

An important feature of Theorem 2 is that it allows for an additive disturbance ε_i. This additive disturbance represents optimization errors, e.g. as in Hausman (1981). The presence of ε_i helps the model fit the data better, because there are few people located at the kinks.

Next we derive the conditional distribution of taxable income given the budget set. We first do this at a point y where $B(y)$ is on the frontier of the convex hull of the budget set in a neighborhood of y. Let $\bar{B}_i$ denote the convex hull of the budget set B_i and $\bar{B}_i(y) = \max_{(c,y)\in\bar{B}_i} c$ denote its frontier. Also, for $B_i(z)$ concave in a neighborhood on the right of y, let $\rho_i(y) = \lim_{z\downarrow y}[B_i(z) - B_i(y)]/(z - y)$ and $R_i(y) = B_i(y) - \rho_i(y)y$, where $\rho_i(y)$ exists by Rockafellar (1970, pp. 214-215).

THEOREM 3: *If Assumption 1 is satisfied then for all B_i such that $\bar{B}_i(z) = B_i(z)$ for $z \in [y, y + \Delta]$ with $\Delta > 0$ we have $\Pr(y(B_i, \eta_i) \leq y|B_i) = F(y|\rho_i(y), R_i(y))$.*

Here we find that at any y and budget set satisfying the conditions of this result, the CDF of taxable income given the budget is the CDF for a linear budget set evaluated at the right slope $\rho(y)$ and corresponding intercept $R(y)$. The right slope of $B(y)$ appears here because of the weak inequality in $\Pr(y(B_i, \eta_i) \leq y|B_i)$.

The form of the CDF involves an important dimension reduction. Note that the CDF conditional on the budget set depends on only the CDF of taxable income for a linear budget set, rather than rather than the infinite dimensional object that is the entire budget set. This makes nonparametric estimation of the conditional CDF feasible over the range of budget sets and y values satisfying the conditions. For example, in many applications nonconvexities in the budget only occur at smaller incomes. Consequently, the conditions of Theorem 3 are likely to be satisfied for higher values of y. This result then shows that nonparametric estimation of the conditional distribution is feasible at higher values of y, where nonconvexities do not occur. This could be used for estimating the revenue effect of tax changes, because much of the revenue often comes from those paying higher taxes.

It is interesting to note that the CDF of taxable income depends on y in each of y, ρ, and R arguments of $F(y|\rho, R)$. Monotonicity of this CDF in y thus depends on what happens as ρ and R change. Utility maximization turns out to be sufficient but not necessary for monontonicity in y of $F(y|\rho(y), R(y))$. This property can easily be seen when the budget frontier $B(y)$ is twice differentiable and concave. In this case $\rho(y) = B_y(y)$, where the subscript denotes a partial derivative, and $R_y(y) = -\rho_y(y)y$. Assuming continuous differentiability of the function $F(y|\rho, R)$, the chain rule then gives

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

[12]

The F_y term is positive because it is the pdf of $Y(\rho, R, \eta)$ at y. As discussed above, utility maximization implies $F_\rho - yF_R \leq 0$, the Slutzky property of the CDF. Also $\rho_y(y) \leq 0$ by the concavity of $B(y)$. Thus utility maximization implies that the second term is nonnegative, and hence that $\partial F(y|\rho(y), R(y))/\partial y \geq 0$. Note though that utility maximization is not necessary for monotonicity of the function $F(y|\rho(y), R(y))$. If F_y is large enough then $\partial F(y|\rho(y), R(y))/\partial y$ could be nonnegative even when $F(y|\rho, R)$ does not satisfy the Slutzky property. This feature of the CDF is consistent with Blomquist's (1995) observation that the likelihood derived from utility maximization can be positive even without the conditions for utility maximization being satisfied.

Theorem 3 makes it easy to understand the conditions under which the CDF and conditional mean for a linear budget set are identified. Note that $\Pr(y(B_i, \eta) \leq y|B_i)$ is from the data. Then from the equality in Theorem 3 we see that for each y, $F(y|\rho, R)$ is identified at each (ρ, R) such that $(\rho, R) = (\rho_i(y), R_i(y))$ for a budget B_i in the data satisfying the hypotheses of Theorem 3. That is, $F(y|\rho, R)$ is identified when there is a budget set B_i in the data that is locally convex at y (in the sense of Theorem 2) where the right slope $\rho(y)$ at y is ρ with corresponding intercept R. Also, since the conditional mean for a linear budget involves all values of y, it will be identified only for those ρ and R such that for every y there is a budget set in the data with (right) slope ρ and intercept R at y. For simplicity we give this result for convex budget sets.

COROLLARY 4: *If Assumption 1 is satisfied and B_i is convex with probability one then for the support Y of y_i and for every $y \in Y$ the function $F(y|\rho, R)$ is identified on the support $X(y)$ of $(\rho_i(y), R_i(y))$. Also, $\int yF(dy|\rho, R)$ is identified on $\cap_{y \in Y} X(y)$.*

One might also try to identify the preference distribution by varying budget sets. Another implication of Theorem 3 is that with strictly convex preferences the set of convex budget sets are no more informative than the set of linear budget sets. This result follows from the fact that the conditional CDF of taxable income only depends on the CDF for linear budget sets, so varying the budget sets over all convex budget sets does no more than trace out the CDF for linear budget sets. Hausman and Newey (2013) show that the distribution of preferences is not identified from linear budget sets, so it will not be identified from convex ones.

Theorem 3 and the Slutzky condition for the CDF $F(y|\rho, R)$ can be combined to give a revealed stochastic preference result for the smooth, strictly convex preferences. As noted in Hausman and Newey (2013), with two goods, smooth stochastic preferences that satisfy the Slutzky condition and produce a quantile function satisfying certain regularity conditions are consistent with the data distribution for linear budget sets if and only if the quantile function

also satisfies the Slutzky condition. Theorem 3 extends this result to convex budget sets. Smooth, strictly convex stochastic preferences that satisfy the Slutzky condition and satisfy other regularity conditions are consistent with the data if and only if the conclusion of Theorem 3 is satisfied for all convex budget sets for an $F(y|\rho, R)$ satisfying the Slutzky like condition $F_\rho - yF_R \leq 0$.

It may also be useful to know the distribution of taxable income at points where the budget set does not have the local convexity property of Theorem 3. We can show that the CDF only depends on $B(y)$ over the the values of y where $B(y)$ is not concave. For simplicity we show this result the case where $B(y)$ has only one nonconcave segment. Let $[\underline{y}(B), \bar{y}(B)]$ denote the interval where $B(y)$ may not be concave and let $\breve{B} = \{\underline{y}(B), \bar{y}(B), B(z)|_{z \in [\underline{y}(B), \bar{y}(B)]}\}$ denote a conditioning set consisting of the interval endpoints and the budget frontier over the interval.

THEOREM 5: *If Assumption 1 is satisfied then for all B_i such that $\bar{B}_i(z) = B_i(z)$ except possibly for $z \in (\underline{y}(B), \bar{y}(B))$ and for $y \in [\underline{y}(B), \bar{y}(B))$ we have* $\Pr(y(B_i, \eta_i) \leq y|B_i) = \Pr(y(B_i, \eta_i) \leq y|\breve{B}_i)$.

This result shows that the CDF of taxable income depends on the budget frontier over the entire nonconvex interval, for any point in the interval. Consequently the conditional mean will depend on the budget frontier over the nonconvex interval. Because the conditional mean can be viewed as a sum of integrals over different intervals, this interval effect will enter in an additive way. For example suppose the budget set is piecewise linear with $\rho_{\tilde{j}-1} < \rho_{\tilde{j}}$ for just one $\tilde{j}$. Then we would have

$$E[y_i|B_i] = \bar{y}(\rho_{J_i i}, R_{J_i i}) + \sum_{j=1}^{J_i-1}[\mu(\rho_{ji}, R_{ji}, \ell_{ji}) - \mu(\rho_{j+1,i}, R_{j+1,i}, \ell_{ji})] + \gamma(R_{\tilde{j}-1}, \rho_{\tilde{j}-1}, R_{\tilde{j}}, \rho_{\tilde{j}}).$$

The γ term represents the deviation of the conditional mean from what it would be if the budget set were convex. If the nonconvexities are small, and ignoring it would be fine. Here the integration over individual heterogeneity reduces the importance of nonconvexities.

6 Policy Effects

It is common practice to measure behavioral effects in terms of elasticities. We are used to linear budget constraints and elasticities with respect to the net of tax rate and non-labor income of a linear budget constraint. One problem with nonlinear budget constraints is that this elasticity may not be identified. The elasticity for a linear budget constraint would often be thought of as corresponding to $\bar{y}(\rho, R)$. From Theorem 4 we see that this function is only identified where for every y the value ρ is the (right) slope of a budget for some individual. The set of such net

of tax rates could well be very small. Therefore for identified object we must look for other kinds of elasticities. 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.

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

Consider the derivative of $G(a,b)$ with respect to a evaluated at $a = b = 0$, given by $\partial G/\partial a = \sum_{j=1}^{J} \partial g/\partial \rho_j$. This is the effect on the conditional mean 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 $E[y_i|B_i = B]$. The construction of $\tilde{\rho}$ can be done in many different ways. We use the sample averages of the net-of-tax rates and virtual incomes for the segments where individuals are actually located. Our elasticity $(\partial G/\partial a)(\tilde{\rho}/E[y|B])$ is an aggregate elasticity which is the policy relevant measure as argued in Saez et. al. (2012). The relationship between this elasticity and that for a linear budget set is spelled out and compared in Appendix A.

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. Here we do so by assuming that productivity growth is the same in percentage terms for all individuals. We assume the wage rate in period t is given by $w = g(e, \eta)\phi(t)$ with $\phi(0) = 1$. The function $\phi(t)$ is a function that captures exogenous productivity growth, i.e., percentage changes in an individual's wage rate that do not depend on the individual's behavior.

[15]

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

$$\text{Max}\, u(c, e, h, \eta) \quad \text{s.t.} \quad c = g(e, \eta)\phi(t)h\rho + R \tag{6.7}$$

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. Inserting the hours of work function back into the direct utility function gives the reduced-form 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. $c = y\rho + R$.

A feature of this problem is that reduced-form utility shifts over time. Our approach to repeated cross section and panel 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 reduced-form maximization problem becomes $\text{Max}\, U(c, \tilde{y}, \eta)$ s.t. $c = \tilde{y}\tilde{\rho} + R$ for $\tilde{\rho} = \phi(t)\theta$. Here the productivity growth appears in the budget set, multiplying the tax rate. From the tax authorities' point of view the taxable income is $y = \phi(t)\tilde{y}$. However, to keep things stationary over time we study the behavior of $\tilde{y}$.

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

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

To implement the corrections on the net-of-tax rates and the dependent variable we need to know the exogenous wage/productivity growth. Unfortunately there are no 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. When doing this, we constrain the annual productivity growth to be the same every year. This is clearly misspecified. However, we do not have information that allows us to do a more refined correction

---

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

[16]

of the budget constraints. Luckily, because data is yearly, the error we impose is not very large for any individual year's budget constraints.

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

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

7 Estimation

The results we have given on the distribution of taxable income given the budget set are important for estimation. They give a dimension reduction, where the conditional mean and distribution only depend on the CDF for taxable income with linear budget constraints rather than the whole budget set. In this section we show how these restrictions can be used in feasible nonparametric estimation. We focus on estimation of the conditional mean with piecewise linear budget sets because our results for the mean allow for an additional additive disturbance, that accounts for optimization errors, and because piecewise linear budget sets are common in applications.

We use series estimates because it is straightforward to impose the kind of dimension reductions we have derived. We will first follow BN and show how to impose the form derived in the conclusion of Theorem 2. We also show how to impose the full implications of utility maximization on the conditional mean, including the dependence only on the CDF for linear budget sets.

To impose the additivity type restrictions of equation (5.6) we include two types of terms in the series approximation, one type to approximate the first term $\bar{y}(\rho_J, R_J)$ and the second type to approximate the other term. We approximate the first term $\bar{y}(\rho_J, R_J)$ by including

functions of just p_J and R_J. For power series these terms take the form

$$p_{kK}(B) = \rho_J^{m(k)} R_J^{q(k)},$$

where $m(k)$ and $q(k)$ are nonnegative integers. Linear combinations of these will approximate the first term. To approximate the second term $\sum_{j=1}^{J-1}[\mu(\rho_j, R_j, \ell_j) - \mu(\rho_{j+1}, R_{j+1}, \ell_j)]$ we use powers of $\rho, R,$ and ℓ and apply the same linear transformation to these terms as is applied to as $\mu(\rho, R, \ell)$. Linear combinations of powers will approximate $\mu(\rho, R, \ell)$ and so, as shown in BN, linear combinations of the transformation will approximate the second term uniformly in the number of budget segments. Consider $\mu_{kK}(\rho, R, \ell) = \rho^{m(k)} R^{q(k)} \ell^{r(k)}$. Applying the differencing transformation to this function gives

$$p_{kK}(B) = \sum_{j=1}^{J-1}[\mu_{kK}(\rho_j, R_j, \ell_j) - \mu_{kK}(\rho_{j+1}, R_{j+1}, \ell_j)].$$

A series estimator can then be formed from these approximating terms in the usual way. For a vector $p^K(B) = (p_{1K}(B), ..., p_{KK}(B))'$ consisting of distinct approximating functions for both terms, let $P = [p^K(B_1), ..., p^K(B_n)]'$ be the matrix of observations on the approximation functions for the budget sets $B_1, ..., B_n$ of all individuals in the sample and $Y = (y_1, ...y_n)'$. A series estimator of $E[y_i | B_i = B]$ is then given by

$$\widehat{E[y_i | B_i = B]} = p^K(B)'\hat{\beta}, \hat{\beta} = (P'P)^- P'Y,$$

where $(P'P)^-$ is any symmetric generalized inverse.

An important and interesting feature of this approach is that the number of budget segments may vary across individuals in a completely flexible way without affecting K. This feature allows the number of budget segments to be as large as needed. The reason that we are able to do this is that this estimator imposes the additivity and equality restrictions. Consequently, it will have convergence properties that are uniform in the number of budget segments, as discussed in BN.

This estimator only imposes some of the restrictions implied by utility maximization. It is also possible to impose all of the restrictions of utility maximization using a series approximation to the conditional pdf of taxable income for linear budget sets. Let $f_1(y), ..., f_A(y)$ be pdf's for a positive integer A and $x = (\rho, R)$. Consider an approximation of the form

$$f(y|x) \approx f_1(y) + \sum_{a=2}^{A} w_a(x, \beta)[f_a(y) - f_1(y)], w_a(x, \beta) = \sum_{b=1}^{B} \beta_{ab} r_b(x),$$

to the conditional pdf of taxable income given a linear budget set, where $r_b(x)$ are approximating functions and β denotes a vector of coefficients in the approxmaton. This is a conditional pdf

[18]

approximation that is a weighted combination of pdf's

$$f(y|x) \approx \sum_{a=1}^{A} w_a(x,\beta) f_a(y), w_1(x,\beta) = 1 - \sum_{a=2}^{A} w_a(x,\beta),$$

where the weights $w_a(x,\beta)$ sum to one over a. If the weights $w_a(x,\beta)$ are positive then this is a mixture approximation to the conditional pdf. We can substitute this approximation in the expression for the conditional mean to obtain an approximation that uses the fact that the integrals in the conditional mean are all taken over $f(y|x)$. Let

$$\bar{y}_a = \int y f_a(y) dy, \mu_a(\ell) = \int 1(y < \ell)(y - \ell) f_a(y) dy.$$

Substituting this approximation to the conditional pdf in the expression for the conditional mean from Theorem 2 gives

$$
\begin{aligned}
E[y_i|B_i \;=\; B] &\approx \bar{y}_1 + \sum_{a=2}^{A} w_a(x_J,\beta)(\bar{y}_a - \bar{y}_1) + \sum_{a=2}^{A} \sum_{j=1}^{J-1} [w_a(x_j,\beta) - w_a(x_{j+1},\beta)][\mu_a(\ell_j) - \mu_1(\ell_j)] \\
&= \bar{y}_1 + \sum_{a=2}^{A} \sum_{b=1}^{B} \beta_{ab}\{r_b(x_J)(\bar{y}_a - \bar{y}_1) + \sum_{j=1}^{J-1} [r_b(x_j) - r_b(x_{j+1})][\mu_a(\ell_j) - \mu_1(\ell_j)]\}.
\end{aligned}
$$

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

$$p_{kK}(B) = \rho_J^{m(k)} R_J^{q(k)}(\bar{y}_{a(k)} - \bar{y}_1) + \sum_{j=1}^{J-1} [\rho_j^{m(k)} R_j^{q(k)} - \rho_{j+1}^{m(k)} R_{j+1}^{q(k)}][\mu_{a(k)}(\ell_j) - \mu_1(\ell_j)]. \qquad (7.8)$$

A series estimator based on these approximating functions imposes the restrictions that the same conditional pdf appears in $\bar{y}(\rho_J, R_J)$ and in $\mu(\rho, R, \ell)$. This restriction is one of those that arises from utility maximization. Another restriction is that the conditional CDF satisfies the Slutzky condition. This can be imposed by requiring that it holds for the approximation at specific values of x, say $x_1, ..., x_C$, and for y, say $y_1, ..., y_D$. Let $F_a(y) = \int_{-\infty}^{y} f_a(y) dy$. The CDF of taxable income conditional on 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 at the values of x and y is then

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

[19]

These are a set of linear in parameters, inequality restrictions. There are also inequality restrictions that correspond to the weights being nonnegative. They take the form

$$w_a(x_c, \beta) \geq 0, (a = 1, ..., A; c = 1, ..., C).$$

Doing least squares with the approximating functions as given in equation (7.8) and imposing all the inequality imposes the full set of restrictions on the conditional mean that come from utility maximization.

A vector of covariates z could be allowed for by letting the conditional pdf for a linear budget set depend on z as well as $x = (p, R)$. The form of the conditional mean and CDF would be as in Theorem 2 and 3, with the covariates z included as additional argument in $\bar{y}$ and F. With this specification the addition of each covariate would raise the dimension of the nonparametric function, making it impractical to allow for more than a very few covariates. An alternative approach would be to specify that z enters through only an index $z'\pi_0$, where π_0 is an unknown vector of parameters, analogously to Ichimura (1993). We follow this index approach to allowing for covariates in the application given below. For the conditional mean the coefficients π_0 could be estimated using a nonlinear version of the least squares estimator described above, with the location and scale of π_0 normalized in some way.

We can allow for endogeneity of the variables determining the budget set using control variates as in Blundell and Powell (2006). Here $E[y_i|B_i = B]$ is precisely the average structural function they consider, and can be identified in the way they describe by including a control variable and then averaging over it holding the other regressors fixed. This approach is now quite familiar so for brevity we omit a description.

An important issue in series estimation is the choice of approximating terms. Here we have the usual trade-off between bias and variance in nonparametric estimation, with a richer set of approximating functions leading to less bias but more variance. 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 $p^K(B_i)$ 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 the all have sample second moment equal to one. This can be done by diving each observation $p_{kK}(B_i)$ on the kth regressor by $\{\sum_{i=1}^{n} p_{kK}(B_i)^2/n\}^{1/2}$.

Here is a multi-step description of the model selection method we use:

a) Get a first estimate of $E[y_i|B_i = B]$ by using all K terms by cross-validation or some other method.

b) Calculate the residuals $\hat{\varepsilon}_i = y_i - \widehat{E[y_i|B_i]}$.

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

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

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

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

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

h) Minimize

$$\frac{\sum_{i=1}^{n}(y_i - p^K(B_i)'\beta)^2}{n} + \frac{\hat{\lambda}}{n} \sum_{j=1}^{J} |\beta_j|.$$

using the Matlab program for lasso.

i) Do OLS using only those elements of $p^K(B_i)$ which had nonzero coefficients in the previous step.

j) Try other values of λ as suggested in g) above corresponding to c closer to 1 and slightly bigger than 2. Compute the cross-validation criteria using the OLS estimates from i) for each model, to provide some goodness of fit comparisons.

We use this procedure to pick the approximating functions for the unrestricted least squares regression. It is not clear whether it would apply when imposing the inequality restrictions implied by utility maximization. We will investigate this in future work.

The estimated taxable income function can be used to predict the conditional mean taxable income given budget set variables and covariates, after integrating over the empirical distribution of the covariates. However, more often one is interested in a functional of the taxable income function, such as elasticities, or the effect of tax reform. It is straightforward to construct the nonlinear budget set elasticities as described on p. 12 from the estimated parameters for a given budget set, control function, and residual function. We integrate over the population distribution of budget sets, single-index and control functions to obtain an aggregate elasticity, and we use a similar procedure for tax reforms. We compute standard errors by bootstrapping.

8 Empirical Application: Sweden

8.1 Data

We use data from HEK (Hushållens Ekonomi) provided by Statistics Sweden, which is a combined register and survey data set. The data set contains repeated cross sections of approximately 17,000 randomly-sampled individuals from the population and members of their households each year. The response rate is approximately seventy percent. The register component contains income, tax, and demographic data used by the authorities for taxation purposes. The

[21]

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 earned income, which was approximately 18,000 SEK in 2008. We use this limit instead of zero, as that would result in a large loss of observations. We also exclude individuals with nonlabor income, as defined below, above 1 million SEK, and below 0 SEK. Out of 102,647 married or cohabiting men between 21 and sixty years of age, 78,268 observations remain after the sample restrictions.

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

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

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

As discussed before, nonlabor income may be endogenous. We instrument nonlabor income (using the control function procedure) using transfers received at zero labor income. This includes, e.g., housing allowances, 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.

For the whole budget set, we also correct 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. These additional corrections are similar to those in Blomquist and Newey (2002).

The data set contains many demographic background variables and in the single-index control function we include dummies for age groups (eight groups), educational level groups (seven groups), socioeconomic occupational groups (eight groups), spousal income groups (twenty groups), whether the individual has children below age six, and whether she is born abroad.

We have experimented with alternative similar sample restrictions, definitions of labor income, ways to handle the impact of the transfer system on the budget sets, and the exact set of covariates in the single-index function. The results are not sensitive to these issues, except for the upper limit on non-labor income. When not using instrumental variables, the income elasticity generally decreases when the upper limit on non-labor income decreases.

We do nonparametric estimation using a series approximation of the BN conditional mean specification as described at the beginning of the last section. We use power series approximations of up to the fourth-order, that imposes just the additivity structure in the conditional mean but not the fact that it comes from a conditional distribution satisfying the Slutzky conditions. We first do this without the single-index and control functions. We then include the covariates linearly to the optimal specification. After this, we form terms including interaction terms between the budget set terms and the single-index and control functions. We then select the optimal specification using Lasso and cross-validation allowing for this expanded set of approximating functions. This procedure is performed separately for different assumptions on the productivity growth rate.

In Table C1 in Appendix C, we report sample statistics. We report the mean values of gross labor income, some variables characterizing the budget sets, and some demographic variables for the whole sample, as well as for the years 1993, 1998, 2003, and 2008.

The average increase in gross labor income over this period is by a factor of 1.22. This implies that the average geometric annual productivity growth is 1.33% (1.0133^{16}=1.22). As discussed in section 3.2, this growth may be due to productivity growth or due to increased work hours and effort. We implicitly estimate the productivity growth in our estimations by estimating models with different productivity growth rates and then selecting the one that

maximizes the cross-validation value. We vary the productivity growth rate between 0% and 1.4%, which is above the growth rate of gross labor income when doing this.

8.2   Results

The cross-validation values for specifications with different productivity growth rates and Lasso lambda values are reported in Table 1. The productivity growth rate increased vertically, and the Lasso lambda decreased horizontally. All reported specifications include the single-index control function and the nonparametric instrumental-variable residual control function estimated in a first stage and their interactions with the budget set regressors. We observe that the cross-validation value increases as the Lasso penalty coefficient λ decreases, up to 0.01 or 0.003. The pattern for different productivity growth rates is not as clear when lambda is large. For the smaller lambdas, cross-validation is, however, clearly maximized around 0.004. The global maximum occurs when the productivity growth rate is 0.4% and the Lasso lambda is 0.003, which is the preferred specification. The cross-validation values for the basic setup without the single-index and control functions, not reported here, display a similar pattern, with a global optimum at the productivity growth rate 0.4%. The optimal Lasso lambda is, however, larger, in that specification.

Table 1. Cross-validation values

Growth/Lambda	1	0.3	0.1	0.03	0.01	0.003	0.001
0.000	0.3992	0.4018	0.4108	0.4170	0.4177	0.3637	0.3637
0.002	0.4003	0.4018	0.4119	0.4162	0.4180	0.4190	0.3243
0.004	0.3967	0.3990	0.4103	0.4164	0.4188	0.4193	0.4181
0.006	0.3939	0.4027	0.4109	0.4159	0.4183	0.4174	0.4173
0.008	0.3916	0.4037	0.4103	0.4140	0.4174	0.4174	0.0163
0.010	0.3929	0.4034	0.4101	0.4134	0.4169	0.4178	0.4168
0.012	0.3923	0.4041	0.4100	0.4134	0.4163	0.4172	0.4165
0.014	0.3938	0.4030	0.4101	0.4129	0.4165	0.4168	0.3141

Notes: Growth refers to imposed productivity growth rate. Lambda refers to the lambda used in Lasso.

The net-of-tax and income elasticities for the preferred specification and at other productivity growth rates at the optimal 0.003 Lasso lambda are reported in Table 2. The marginal effects are evaluated at each of the individual budget sets, and at the sample mean income, marginal net-of-tax rate, and marginal virtual incomes. We observe that the net-of-tax and income elasticities decrease as we increase the productivity growth rate, up to a growth rate of 0.4%, after which it is fairly stable. Correcting for productivity growth appropriately therefore matters. At the

[24]

preferred specification with the productivity growth rate of 0.4%, the net-of-tax elasticity is 0.60, and the income elasticity is -0.08. Both are statistically significant.

Table 2. Estimated elasticities by productivity growth rate

Productivity growth	Lasso lambda	Net-of-tax elasticity	Income elasticity
0.000	0.003	1.16	0.14
		(0.09)	(0.03)
0.002	0.003	0.66	-0.06
		(0.14)	(0.03)
0.004	0.003	0.60	-0.08
		(0.13)	(0.03)
0.006	0.003	0.57	-0.08
		(0.12)	(0.03)
0.008	0.003	0.50	-0.07
		(0.11)	(0.03)
0.010	0.003	0.55	-0.06
		(0.12)	(0.03)
0.012	0.003	0.55	-0.04
		(0.10)	(0.03)
0.014	0.003	0.56	-0.04
		(0.09)	(0.03)

Notes: Bootstrapped standard errors (50 replications) are reported in parenthesis. The elasticities are evaluated at the sample mean income 462,174 SEK, marginal net-of-tax rate 0.3174, and virtual income 208,513 SEK.

In Table 3, we instead vary the Lasso lambda to examine the effect of the number of approximating terms. We observe that the net-of-tax elasticity first decreases and then increases as the lambda decreases. The income elasticity decreases as lambda decreases. Decreasing lambda below 0.003 has, however, small effects on the elasticities. We have experimented with decreasing lambda further. The net-of-tax elasticities remain between 0.60 and 0.66 and the income elasticity around -0.08 when doing this (these estimates are not reported here).

Table 3. Estimated elasticities by Lasso lambda

Productivity growth	Lasso lambda	Net-of-tax elasticity	Income elasticity
0.004	1	0.97	-0.21
		(0.06)	(0.04)
0.004	0.3	1.00	-0.25
		(0.06)	(0.05)
0.004	0.1	0.38	-0.13
		(0.08)	(0.02)
0.004	0.03	0.50	-0.10
		(0.07)	(0.03)
0.004	0.01	0.38	-0.11
		(0.10)	(0.03)
0.004	0.003	0.60	-0.08
		(0.13)	(0.03)
0.004	0.001	0.66	-0.08
		(0.13)	(0.03)

Notes: Bootstrapped standard errors (50 replications) are reported in parenthesis. The elasticities are evaluated at the sample mean income 462,174 SEK, marginal net-of-tax rate 0.3174, and virtual income 208,513 SEK.

The elasticities are weighted averages of the marginal effects multiplied by a scaling factor. In the current framework, the marginal effects are, however, allowed to vary with budget sets and with demographics.

In Table 4, we investigate the effects of the different methods used. In Basic, we report the estimates when only using budget set regressors. The next set of estimates additionally includes the single-index function, and the final set of estimates also includes the residual from the first stage.

Table 4. Estimated elasticities and EITC effect by method

Method	Net-of-tax elasticity	Income elasticity
Basic	0.50	0.40
	(0.08)	(0.01)
+Single index	0.61	0.16
	(0.15)	(0.01)
+Control function	0.60	-0.08
	(0.13)	(0.03)

Notes: Bootstrapped standard errors (50 replications) are reported in parenthesis. The elasticities are evaluated at the sample mean income 462,174 SEK, marginal net-of-tax rate 0.3174, and virtual income 208,513 SEK. Productivity growth rate is set to 0.004. Lasso lambda is set to 0.003.

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 significant from each other.

We observe that the net-of-tax elasticity is fairly stable between 0.50 and 0.61. We have performed similar exercises for larger Lasso lambdas. For these specifications, not reported here, the basic net-of-tax elasticity is much higher, up to 1.0, and the single-index control function brings the elasticity down. This indicates that either the single-index control function or additional terms alone do a good job of adjusting the net-of-tax elasticity towards the ex-post preferred number. The income elasticity decreases, however, when including the single-index function from 0.40 to 0.16 and becomes negative when instrumenting non-labor income. Because the instrumentation directly affects virtual incomes only, it mainly affects the income elasticity.

The conditional mean of taxable income is an ideal tool to evaluate the effect of a tax reform on taxable income. To illustrate this we use our estimated conditional mean to evaluate a temporary tax reform that took place in 1995-1998. Prior to 1995, there was a central government top income tax of 20% for (post-payroll tax) earned income above approximately SEK 230,000. 63% of the individuals in the sample had income above this level. Total top marginal tax rates were 71% including indirect taxes. In 1995 the central tax rate was increased by 5% to 25%. In 1999, it was decreased back to 20% for income below approximately SEK 400,000, but kept at 25% for income above this level. We investigate the effects of the additional top income tax of 5% in 1995-1998, by using our estimated taxable income function. We can do this by predicting expected taxable incomes for each individual in the 1995-1998 sample imposing the pre-reform top income tax of 20% and then the post-reform top income tax of 25%. The effects on the sample average taxable income, government revenues assuming no behavioral effects, and government revenues including behavioral effects (under some distributional assumptions) are reported in Table 5.

Table 5. Estimates pre- and post-reform 1995-1998

	Taxable income	Revenues mechanic	Revenues behavior
Pre-reform	379,900	242,000	244,400
Post-reform	376,800	246,400	246,400
% change	-0.810	+1.807	+0.822

A difficulty in calculating informative sample averages is that the expected prediction error for 1995-1998 is not zero since more years are used in the estimation. We do, however, observe

taxable income and can calculate the prediction error for each individual in the sample as the difference between observed and predicted taxable income for the post-reform tax system in place. We report the sample average observed taxable income for the post-reform tax system $(y_{post} = \hat{y}_{post} + \hat{\varepsilon}_{post})$ which was SEK 376,800. For the counterfactual pre-reform tax system, we report the sample average predicted taxable income plus the same prediction errors as in the post-reform period $(\hat{y}_{pre} + \hat{\varepsilon}_{post})$ and obtain SEK 379,900. Because the prediction errors are kept constant, $\hat{y}_{post} - \hat{y}_{pre} = (\hat{y}_{post} + \hat{\varepsilon}_{post}) - (\hat{y}_{pre} + \hat{\varepsilon}_{post})$. Note that the individual predictions plus prediction errors may be problematic because they may yield individual taxable incomes below zero. The difference between the pre-reform and post-reform sample averages is, however, an unproblematic measure of the reform effect. We see that the reform decreased taxable income by 0.8%.

Another difficulty arises if we want to calculate effects that depend on the distribution of taxable incomes under different tax regimes since we do not estimate parameters related to the error terms. For instance, the distribution of expected taxable incomes in the pre-reform counterfactual tax regime cannot be obtained, which is problematic when calculating the implications of the taxable income effects on government revenues. For the post-reform tax regime, we know the sample distribution and can calculate average revenues at observed taxable incomes $(T_{post}(y_{post}))$, which is SEK 246,400. We calculate the average pre-reform revenues assuming that there are no behavioral effects, keeping taxable incomes at the observed post-reform level $(T_{pre}(y_{post}))$ and obtain average revenues of SEK 242,000. The reform hence implies a 1.8% mechanical increase in tax revenues. It is possible to try to calculate tax revenues based on the pre-reform predicted taxable incomes. Because of the nonlinearity of the tax system, for a given budget set, revenues at predicted taxable income would be a biased estimate of expected revenues. It would be downward biased because of the progressivity of the tax system. If the post-reform prediction error is added to the pre-reform predicted taxable income, an additional issue is what to do with taxable incomes below zero. The biases created by those issues would not be the same in the pre- and post-reform regimes, and consequently, the estimate of the change in revenues would be biased.

The same types of difficulties arise when we want to infer the effect on average revenues including behavioral effects from the unproblematic estimate of average reform effect on taxable income. For the post-reform tax regime, we can again use the observed sample distribution of taxable incomes to infer average revenues $(T_{post}(y_{post}))$ of 246,400 SEK. For the pre-reform tax regime, we proceed by making the distributional assumption that the change in average reform effect on taxable income is entirely due to behavioral effects of individuals in the top tax bracket for which the reform changes their marginal tax rate. We can then calculate pre-reform tax

revenues as $T_{post}(y_{post}) + (\hat{y}_{post} - \hat{y}_{pre}) * $"top marginal tax rate" which gives SEK 244,400. The reform therefore increases tax revenues by approximately 0.8%. Thus, the behavioral response reduces the percentage increase in tax revenue from 1.8% to 0.8%. It is possible to evaluate pre-reform tax revenues under alternative distributional assumptions. If the response is uniform in the population, it can be calculated as $T_{post}(y_{post}) + (\hat{y}_{post} - \hat{y}_{pre}) * $"average marginal tax rate" . This would produce a difference of less than SEK 100.

9   Summary and conclusion

Building on Blomquist and Newey (2002) we develop a method to nonparametrically estimate expected taxable income as a function of a nonlinear budget set while allowing for general heterogeneity and optimization errors. The heterogenous preferences are assumed to be strictly convex and statistically independent of the budget set, but are otherwise unrestricted. Using restrictions implied by utility maximization we find that the conditional mean of taxable income can be estimated by a low dimensional nonparametric function.

We also derive the distribution of taxable income at points where the budget frontier is concave. We find that this distribution is the same as for a linear budget set at these points, dramatically reducing the dimension of the nonparametric estimation problem. Also, we find that varying concave budget sets provides the same information about preferences that linear budgets do. We also analyze kinks, showing that kink derivatives are like regression discontinuities, identifying effects for those at the kink under continuity conditions on unobserved outcomes. In addition we give simple identification results for the conditional distribution and mean for linear budget sets.

To evaluate the effect of taxes on taxable income we focus on elasticities that apply to changes in nonlinear tax systems. Real-world tax systems are non-linear, and it is variations in non-linear tax systems that we observe. Therefore, it is easiest to nonparametrically identify elasticities relevant for changes in nonlinear tax systems. BN did show that with labor supply it may be possible to identify labor supply elasticities for changes in linear budget sets, but we show that the conditions are stringent, and may not be satisfied for taxable income. Here we propose elasticities defined by an upward shift of the non-linear budget constraint, in either slope or intercept. These elasticities are relevant for changes in non-linear budget constraints.

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

We apply our method to Swedish data from 1993-2008 with third party reported taxable

labor income. For prime age Swedish males we estimate a statistically significant net of tax elasticity of 0.60 and a significant income elasticity of -0.08.

Often one also wants to know how expected tax revenue changes as there are changes in the tax system. If the tax function is linear it is sufficient to know expected taxable income, but if the tax is nonlinear one needs to know the distribution of taxable income. Therefore, in future work we will attempt to derive a way to non-parametrically estimate the CDF (cumulative distribution function) for taxable income as a function of the budget set. (We believe this is simpler than to non-parametrically estimate quantile functions.) An alternative approach that we also will try is to derive an estimable form for expected tax revenue as a function of the nonlinear budget set.

10 Appendix A: Comparing nonlinear and linear budget set elasticities

Let the taxable income supply function for a linear budget constraint be given by $\pi(R, \theta, v)$ with $\pi(\cdot)$ increasing in v. For simplicity we consider the case with only one heterogeneity term. Further we assume the form of $\pi(R, \theta, v)$ and the distribution of $v \in [\underline{v}, \overline{v}]$ are such that corner solutions are never an issue. The expected taxable income for a linear budget constraint would be given by $E(Y) = \overline{\pi}(R, \theta) = \int_{\underline{v}}^{\overline{v}} \pi(R, \theta, v) f(v) dv$. Suppose the tax system consists of two parts, the federal income tax and state income tax, so that $\theta = 1 - \tau - t$, where t is the federal income tax and τ the state income tax. Let $a = 1 - \tau$ and $\theta = a - t$. We are interested in the elasticity of the expected taxable income w.r.t. a. This elasticity would be given by $\frac{dE(Y)}{da} \frac{a}{E(Y)} = \frac{a}{E(Y)} \int_{\underline{v}}^{\overline{v}} \frac{\partial \pi(R, \theta, v)}{\partial \theta} f(v) dv$.

Let us next consider a kinked budget constraint, with net-of-tax rates θ_1, θ_2, and θ_3 with virtual incomes R_1, R_2, and R_3. Define the subsets $S1 = (\underline{v}, \overline{v}_1)$, $S2 = (\underline{v}_2, \overline{v}_2)$ and $S3v \in (\underline{v}_3, \overline{v})$ such that Y will be on segment i for $v \in Si$. Likewise, define the sets $K1 = (\overline{v}_1, \underline{v}_2)$ and $K2 = (\overline{v}_2, \underline{v}_3)$ such that Y will be at the kink point $Y1$ for $v \in K1$, and Y will be at kink point $Y2$ for $v \in K2$. The expected taxable income is given by $E(Y) = \sum_{i=1}^{3} \int_{Si} \pi(R_i, \theta_i, v) f(v) dv + \sum_{i=1}^{2} Ki \int_{Ki} f(v) dv$. Doing the same kind of parametrization as for the linear budget constraint we rewrite this as $E(Y) = \sum_{i=1}^{3} \int_{Si} \pi(R_i, a - t_i, v) f(v) dv + \sum_{i=1}^{2} Ki \int_{Ki} f(v) dv$. Forgetting about changes in integration limits, which we will consider in the future, we would have the elasticity given by

$$\frac{dE(Y)}{da} \frac{a}{E(Y)} = \frac{a}{E(Y)} \sum_{1=1}^{3} \int_{Si} \frac{\partial \pi(R_i, \theta_i, v)}{\partial \theta_i} f(v) dv.$$

.

Suppose that we instead linearize the second segment and calculate the elasticity as a variation of this segment, then we would obtain

$$\eta_a = \frac{dE(Y)}{da}\frac{a}{E(Y)} = \frac{a}{E(Y; R_2, \theta_2)}\int_{\underline{v}}^{\bar{v}}\frac{\partial \pi\left(R_2, \theta_2, v\right)}{\partial \theta_2}f(v)dv$$

To get an idea about the numerical difference in the two elasticity concepts we consider a functional form for the taxable income function that has been estimated in the taxable income literature often: $y = \theta^\beta e^v$. This implies that we have assumed away income effects. Taking logs it has the form $\ln y = \beta \ln \theta + v$. For simplicity we assume that v is uniformly distributed on $[0, 1]$. For a linear budget constraint with net-of-tax rate θ we get the expected taxable income

$$E(y) = \int_0^1 \theta^\beta e^v dv = \theta^\beta \left[e^1 - e^0\right] = \theta^\beta \left[e - 1\right]$$

and the taxable income elasticity for a linear budget constraint becomes:

$$\eta_\ell = \frac{dE(y)}{d\theta}\frac{\theta}{E(y)} = \frac{\beta\theta^{\beta-1}\left[e-1\right]\theta}{\theta^\beta\left[e-1\right]} = \beta.$$

Let us next consider what the elasticity is for a piecewise-linear budget constraint with three segments. We denote the break points $Y1$ and $Y2$. For simplicity we consider a budget constraint defining a convex budget set.

With this form for the taxable income function we will have $S1 = \left(0, \ln\left(Y1/\theta_1^\beta\right)\right)$, $S2 = \left(\ln\left(Y1/\theta_2^\beta\right), \ln\left(Y2/\theta_2^\beta\right)\right)$ and $S3 = \left(\ln\left(Y2/\theta_3^\beta\right), 1\right)$. Likewise define the sets $K1 = \left(\ln\left(Y1/\theta_1^\beta\right), \ln\left(Y1/\theta_2^\beta\right)\right)$ and $K2 = \left(\ln\left(Y2/\theta_2^\beta\right), \ln\left(Y2/\theta_3^\beta\right)\right)$.

The integral over $S1$ will be $\theta_1^\beta \int_0^{\ln\left(\frac{Y1}{\theta_1^\beta}\right)} e^v dv = \theta_1^\beta\left[e^{\ln\left(\frac{Y1}{\theta_1^\beta}\right)} - e^0\right] = \theta_1^\beta\left[\frac{Y1}{\theta_1^\beta} - 1\right] = \left[Y1 - \theta_1^\beta\right].$

In a similar way the integral over $S2$ will be $\theta_2^\beta\left[\frac{Y2}{\theta_2^\beta} - \frac{Y1}{\theta_2^\beta}\right] = [Y2 - Y1]$ and over $S3$ $\left[\theta_3^\beta e - Y2\right]$. For $K1$ we get $Y1\left[\ln\theta_1^\beta - \ln\theta_2^\beta\right]$ and for $K2$, $Y2\left[\ln\theta_2^\beta - \ln\theta_3^\beta\right]$. Summing over the segments we get sum segments $= \theta_3^\beta e - \theta_1^\beta$. Summing over the two kink points we obtain sum kink points $= \beta\left(Y2 - Y1\right)\ln\theta_2 + \beta Y1\ln\theta_1 - \beta Y2\ln\theta_3$. (Note that the sum over the kink points is zero if $\theta_1 = \theta_2 = \theta_3$.) Hence, the expression for $E(Y)$ becomes:

$$E(Y) = \theta_3^\beta e - \theta_1^\beta + \beta\left(Y2 - Y1\right)\ln\theta_2 + \beta Y1 n\theta_1 - \beta Y2\ln\theta_3.$$

We make the parametrization $\theta_i = a - t_i$ and note that $d\theta_i/da = 1$. We use the superscript pl to denote the elasticity for a piecewise-linear budget constraint and the subscript a to denote that it is the elasticity with respect to a.

[31]

The elasticity becomes

$$\eta_a^{pl} = \frac{dE(Y)^{pl}}{da}\frac{a}{E(Y)_i^{pl}} = \frac{a\left(\beta\theta_3^{\beta-1}e - \beta\theta_1^{\beta-1} + \beta\left(Y2 - Y1\right)/\theta_2 + \beta Y1/\theta_1 - \beta Y2/\theta_3\right)}{\theta_3^{\beta}e - \theta_1^{\beta} + \beta\left(Y2 - Y1\right)\ln\theta_2 + \beta Y1\ln\theta_1 - \beta Y2\ln\theta_3}.$$

Note that this elasticity takes into account that the number of individuals at a kink point will change as we change a.

If we linearize around, say, the second segment we come up with a linear budget constraint with elasticity equal to β.

Let us exemplify with some numbers: Let $\beta = 0.4$, $Y1 = 1.2$, $Y2 = 1.4$, $\theta_1 = 0.7$, $\theta_2 = 0.5$ and $\theta_3 = 0.3$. The expected value for Y will be 1.256. We set $a = \theta_2 = 0.5$. The elasticity w.r.t. a will then be 0.28. This is significantly less than the elasticity of 0.4 that we would get from a linearized budget constraint. There is no reason to believe that the marginal deadweight loss is directly related to η^{ℓ}. Is it directly related to η^{pl}? Our conjecture is that it is.

The intuition for the lower elasticity for the curved budget constraint is the following. The elasticity is a weighted average of the elasticities for all the different v. On a linear budget constraint the elasticity is β for any v. For individuals with a v such that they are at a kink point, their elasticity is zero.

11 Appendix B: Proofs

Proof of Lemma 1: We suppress the i subscript and conditioning on B for notational convenience. Consider the disjoint sets $A = \{\eta : Y_j(\eta) \geq \ell_j, Y_{j+1}(\eta) \leq \ell_j\}$, $B = \{\eta : Y_j(\eta) \geq \ell_j, Y_{j+1}(\eta) > \ell_j\}$, $C = \{\eta : Y_j(\eta) < \ell_j, Y_{j+1}(\eta) \leq \ell_j\}$, and $D = \{\eta : Y_j(\eta) < \ell_j, Y_{j+1}(\eta) > \ell_j\}$.

First suppose that $Y_j(\eta) < \ell_j$. Consider any $\bar{y} > \ell_j$. Note that by $\rho_j > \rho_{j+1}$

$$\begin{aligned}
R_j + \rho_j\bar{y} &= R_j + \rho_j\ell_j + \rho_j(\bar{y} - \ell_j) = R_j + \rho_{j+1}\ell_j + \rho_j(\bar{y} - \ell_j) \\
&> R_j + \rho_{j+1}\ell_j + \rho_{j+1}(\bar{y} - \ell_j) = R_{j+1} + \rho_{j+1}\bar{y}.
\end{aligned}$$

Therefore, by strict monotonicity of preferences we have

$$U(R_j + \rho_j Y_j(\eta), Y_j(\eta), \eta) \geq U(R_j + \rho_j\bar{y}, \bar{y}, \eta) > U(R_{j+1} + \rho_{j+1}\bar{y}, \bar{y}, \eta).$$

It follow similarly that for any $\tilde{y} < \ell_j$,

$$U(R_{j+1} + \rho_{j+1}Y_{j+1}(\eta), Y_{j+1}(\eta), \eta) > U(R_j + \rho_j\tilde{y}, \tilde{y}, \eta).$$

If $Y_j(\eta) < \ell_j < Y_{j+1}(\eta)$ then choosing $\bar{y} = Y_{j+1}(\eta)$ gives

$$U(R_j + \rho_j Y_j(\eta), Y_j(\eta), \eta) > U(R_{j+1} + \rho_{j+1}Y_{j+1}(\eta), Y_{j+1}(\eta), \eta),$$

[32]

while choosing $\tilde{y} = Y_j(\eta)$ gives

$$U(R_{j+1} + \rho_{j+1}Y_{j+1}(\eta), Y_{j+1}(\eta), \eta) > U(R_j + \rho_j Y_j(\eta), Y_j(\eta), \eta),$$

a contradiction. Therefore $Y_j(\eta) < \ell_j < Y_{j+1}(\eta)$ cannot occur, i.e. $\Pr(D) = 0$. Therefore,

$$
\begin{aligned}
\Pr(Y(B, \eta) &= \ell_j) = \Pr(A) = \Pr(A \cup B) - \Pr(B) = \Pr(A \cup B) - \Pr(B \cup D) \\
&= \Pr(Y_j(\eta) \geq \ell_j | B) - \Pr(Y_{j+1}(\eta) > \ell_j | B),
\end{aligned}
$$

where the second equality follows by A and B disjoint and the third by $\Pr(D) = 0$. The conclusion then follows by $\Pr(Y_j(\eta) \geq \ell_j | B) = 1 - \Pr(Y_j(\eta) < \ell_j | B)$ and $\Pr(Y_{j+1}(\eta) > \ell_j | B) = 1 - \Pr(Y_{j+1}(\eta) \leq \ell_j | B)$. Q.E.D.

Proof of Theorem 2: For convenience we suppress the i subscript and the B_i argument. Let $F_j(y) = F(y | \rho_j, R_j)$. Note that by Hausman (1979),

$$
\begin{aligned}
E[y] &= \sum_{j=1}^{J-1} \left[\int 1(\ell_{j-1} < Y_j(\eta) < \ell_j) Y_j(\eta) F(d\eta) + \ell_j \Pr(Y(B, \eta) = \ell_j) \right] \\
&+ \int 1(\ell_{J-1} < Y_J(\eta)) Y_J(\eta) F(d\eta) \\
&= \sum_{j=1}^{J-1} \left[\int 1(\ell_{j-1} < y < \ell_j) y F_j(dy) + \ell_j \Pr(Y(B, \eta) = \ell_j) \right] \\
&+ \int 1(\ell_{J-1} < y) y F_J(dy).
\end{aligned}
$$

By Lemma 1,

$$\Pr(Y(B, \eta) = \ell_j) = \int 1(y \leq \ell_j) F_{j+1}(dy) - \int 1(y < \ell_j) F_j(dy)$$

Also, note that by $\ell_0 = 0$ we have $\int 1(y \leq \ell_0) y F_1(dy) = 0$. Therefore, by a change of summation index.

$$
\begin{aligned}
\sum_{j=1}^{J-1} \int 1(\ell_{j-1} < y < \ell_j) y F_j(dy) &= \sum_{j=1}^{J-1} \int \{1(y < \ell_j) - 1(y \leq \ell_{j-1})\} y F_j(dy) \\
&= \sum_{j=1}^{J-1} \int 1(y < \ell_j) y F_j(dy) - \sum_{j=1}^{J-2} \int 1(y \leq \ell_j) y F_{j+1}(dy)
\end{aligned}
$$

Furthermore we have $\int 1(\ell_{J-1} < y) F_J(dy) = \bar{y}(\rho_J, R_J) - \int 1(y \leq \ell_{J-1}) y F_J(dy)$. Collecting terms then gives

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

[33]

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

To show the second conclusion note that also by Lemma 1,

$$\Pr(Y(B, \eta) = \ell_j) = \int 1(y \geq \ell_j)F_j(dy) - \int 1(y > \ell_j)F_{j+1}(dy).$$

Note that $\bar{y}(\rho_1, R_1) = \int 1(y > \ell_0)yF_1(dy)$. Then we have

$$\begin{aligned}
\sum_{j=1}^{J-1} \int 1(\ell_{j-1} < y < \ell_j)yF_j(dy) &= \sum_{j=1}^{J-1} \int \{1(y > \ell_{j-1}) - 1(y \geq \ell_j)\}yF_j(dy) \\
&= \sum_{j=1}^{J-1} \int 1(y > \ell_{j-1})yF_j(dy) - \sum_{j=1}^{J-1} \int 1(y \geq \ell_j)yF_j(dy) \\
&= \bar{y}(\rho_1, R_1) + \sum_{j=1}^{J-2} \int 1(y > \ell_j)yF_{j+1}(dy) - \sum_{j=1}^{J-1} \int 1(y \geq \ell_j)yF_j(dy).
\end{aligned}$$

Collecting terms then gives

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

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.*

We use the following three lemmas to prove Theorem 3.

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

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

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

An analogous argument gives $U(B(y^*), y^*) > U(B(y), y)$ for $y > y^*$. *Q.E.D.*

LEMMA A2: *If Assumption 1 is satisfied and B_i is convex then $\Pr(y(B_i, \eta) \leq y|B_i) = F(y|\rho_i(y), R_i(y))$.*

[34]

Proof: Let $y^* = \text{argmax}_{\tilde{y}} U(B(\tilde{y}), \tilde{y})$ where we suppress the η argument for convenience. Let $\tilde{B} = B(y)$, $\bar{\rho} = \rho(y)$, and let $\bar{y}^*$ denote the utility maximizing value on the line $\tilde{B}(\tilde{y}) = \tilde{B} + \bar{\rho}(\tilde{y} - y)$, i.e.

$$\bar{y}^* = \text{argmax}_{\tilde{y}} U(\tilde{B}(\tilde{y}), \tilde{y}).$$

By Rockafellar (1970, pp. 214-215) $\rho(y)$ exists and is a subgradient of $B(\tilde{y})$ at y. Therefore, by concavity of $B(\tilde{y})$, $\tilde{B}(\tilde{y}) \geq B(\tilde{y})$ for all $\tilde{y} \geq 0$, and hence $\tilde{B}(y^*) \geq B(y^*)$. It follows by monotonicity of preferences that

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

If $y < y^*$ then $y < \bar{y}^*$, as otherwise $\bar{y}^* \leq y < y^*$, so that Lemma A1 implies $U(\tilde{B}(y), y) > U(\tilde{B}(y^*), y^*)$ contradicting the above equation. Similarly, $y^* < y$ implies $\bar{y}^* < y$.

Next suppose $y = y^*$. Let ρ^* be the slope of a line that separates the set weakly preferred to $(B(y), y)$ and the budget set. By Rockafellar (1970, pp. 214-215) $\rho^* \geq \bar{\rho}$. Then for any $\tilde{y} > y^*$ we have

$$\tilde{B} + \rho^*(\tilde{y} - y^*) \geq \tilde{B} + \bar{\rho}(\tilde{y} - y) = \tilde{B}(\tilde{y}),$$

so that

$$U(\tilde{B}(\tilde{y}), \tilde{y}) \leq U(\tilde{B} + \rho^*(\tilde{y} - y^*), \tilde{y}) < U(\tilde{B}, y^*),$$

where the strict inequality follows by Lemma A1 applied to the linear and hence convex budget frontier $B(y) = \tilde{B} + \rho^*(\tilde{y} - y)$, for which y^* is optimal. Therefore $\bar{y}^* \leq y$.

Summarizing, we have shown that

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

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

Note that y^* is the utility maximizing point on the budget frontier $B(y)$ while $\bar{y}^*$ is the utility maximizing point on the linear budget set passing through the point $(B(y), y)$ with slope $\rho(y)$. Thus, $y^* \leq y \Longleftrightarrow \bar{y}^* \leq y$ means that the event $y(B, \eta) \leq y$ coincides with the event that the optimum on the linear budget set is less than or equal to y. The probability that the optimum on this linear budget is less than or equal to y is $F(y|\rho(y), R(y))$, giving the conclusion. $Q.E.D.$

LEMMA A3: *If the hypotheses of Theorem 3 are satisfied then* $\Pr(y(B_i, \eta) \leq y|B_i) = \Pr(y(\bar{B}_i, \eta) \leq y|B_i)$.

Proof: Let $\bar{B}$ be convex hull of B and $\bar{B}(y) = \{\max\{c = (c, y) \in \bar{B}\}\}$ be the upper boundary. For notational simplicity suppress the η argument and let $U(c, y) = U(c, y, \eta)$. The budget B

is a subset of its convex hull so that $\bar{B}(z) \geq B(z)$ for all z. Define

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

where y^* may be a set.

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

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

By Lemma A1 we cannot have $\bar{y}^* > y$ because then the above inequality is not consistent with $U(\bar{B}(z), z)$ being strictly monotonically decreasing to the left of $\bar{y}^*$. Therefore $\bar{y}^* \leq y$. Suppose next that $y^* < y$. Then

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

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

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

Next, suppose that $y^* > y$. Then for all $z \in [y, y + \Delta]$

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

Again by Lemma A1 we cannot have $\bar{y}^* \leq y$ because this inequality is not consistent with $U(\bar{B}(z), z)$ being strictly monotonic decreasing to the right of $\bar{y}^*$. Therefore, $\bar{y}^* > y$, and we have shown that

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

Therefore we have $y^* \leq y \Longleftrightarrow \bar{y}^* \leq y.Q.E.D.$

Proof of Theorem 3: Combining the conclusions of Lemmas A2 and A3 gives

$$\Pr(y(B_i, \eta) \leq y | B_i) = \Pr(y(\bar{B}_i, \eta) \leq y | B_i) = F(y | \rho_i(y), R_i(y)).Q.E.D.$$

Proof of Theorem 4: Given in the paper.

Proof of Theorem 5: For notational simplicity we suppress the η as before. Let $\hat{y} = y(B), \check{y} = \bar{y}(B)$ denote the lower and upper bound respectively. Let $\tilde{B}(B) = \{(c, y) : c \leq B(y), \hat{y} \leq y \leq \check{y}\}$, $y^* = y(B, \eta)$, and consider

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

[36]

By construction, for all y

$$F_{\tilde{y}(B,\eta)}(y|B) = F_{\tilde{y}(B,\eta)}(y|\tilde{B}).$$

Now consider any $y \in [\hat{y}, \check{y})$. We will prove that $y^* \leq y$ if and only if $\tilde{y} \leq y$.

Suppose first that $y^* \leq y$. If $y^* \geq \hat{y}$ then $y^* = \tilde{y}^*$, so that $\tilde{y}^* \leq y$. If $y^* < \hat{y}$, then by Lemma A1, for all $z \in (\hat{y}, \check{y}]$ and the upper boundary $\bar{B}(z)$ of the convex hull $\bar{B}$ of B,

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

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

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

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

so that $\tilde{y}^* = \check{y} > y$. Therefore $y^* > y \implies \tilde{y}^* > y$.

Summarizing, for $y \in [\hat{y}, \check{y})$ we have $y^* \leq y \iff \tilde{y}^* \leq y$. Therefore, adding back the η notation, we have

$$F_{y(B,\eta)}(y|B) = F_{\tilde{y}(B,\eta)}(y|B) = F_{\tilde{y}(B,\eta)}(y|\tilde{B}).Q.E.D.$$

12 Appendix C: Sample Statistics Sweden

Table C1. Sample statistics

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

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

13   References

Aaberge, R. and L. Flood (2008), "Evaluation of an In-work Tax Credit Reform in Sweden: Effects on Labor Supply and Welfare Participation of Single Mothers," IZA Discussion Paper 3736.

Andrews, D.W.K. (1991), "Asymptotic Optimality of Generalized CL, Cross-validation, and Generalized Cross-validation in Regressions with Heteroskedastic Errors," *Journal of Econometrics 47*: 359–377.

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

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

Blomquist, S. (1995), "Restrictions in Labor Supply Estimation: Is the MaCurdy Critique Correct?" *Economics Letters 47,* 229-235.

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

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

Blundell, R. and T. MaCurdy (1999), "Labor Supply: A Review of Alternative Approaches," in *Handbook of Labor Economics*, Volume 3, O. Ashenfelter and D. Card (Eds.), Elsevier Science B.V.

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

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

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

Chetty, R., (2008), "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," NBER Working Paper 14399, National Bureau of Economic Research.

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

College.

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

Ericson, P., R. Wahlberg, and L. Flood (2009), "SWEtaxben: A Swedish Tax/Benefit Micro Simulation Model and an Evaluation of a Swedish Tax Reform," IZA Discussion Paper 4106.

Feenberg, D.R. and E. Coutts (1993), "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management 12(1)*: 189–194.

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

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

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

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

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

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

Goolsbee, A. (2000), "What Happens when you Tax the Rich? Evidence from Executive Compensation," *Journal of Political Economy, 108(2)*: 352–378.

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

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

Hausman, J. (1981), "Labor Supply: How Taxes Affect Economic Behavior," in: H. Aaron and J. Pechman (eds.), *Tax and the Economy.* Washington, DC: Brookings Institution.

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

Hausman, J. and W. Newey (2013), "Individual Heterogeneity and Average Welfare", working paper.

Heim, B.T. and B.D. Meyer (2003) "Structural Labor Supply Models when Budget Constraints are Nonlinear," Working Paper, Harris School of Public Policy, University of Chicago.

Hoderlein, S. and E. Mammen (2007): "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," *Econometrica* 75, 1513–1518.

Ichimura, H. (1993) "Estimation of Single Index Models," *Journal of Econometrics* 58, 71-120.

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

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

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

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

Manski, C. (2013) "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy," *Quantitative Economics*, forthcoming.

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

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

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

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

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

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

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

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

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

[40]

Slemrod, J. (1998), "Methodological Issues in Measuring and Interpreting Taxable Income Elasticities," *National Tax Journal 51(4)*: 773-788.