Nonlinear Budget Set Regressions for the Random Utility Model

Soren Blomquist, Anil Kumar, Che-Yuan Liang and Whitney K. Newey

Working Paper 2219 **September 2022**

Research Department

https://doi.org/10.24149/wp2219

Nonlinear Budget Set Regressions for the Random Utility Model*

Soren Blomquist†, Anil Kumar‡, Che-Yuan Liang§ and Whitney K. Newey±

September 2022

Abstract

This paper is about the nonparametric regression of a choice variable on a nonlinear budget set when there is general heterogeneity, i.e., in the random utility model (RUM). We show that utility maximization makes this a three-dimensional regression with piecewise linear, convex budget sets with a more parsimonious specification than previously derived. We show that the regression allows for measurement and/or optimization errors in the outcome variable. We characterize all of the restrictions of utility maximization on the budget set regression and show how to check these restrictions. We formulate nonlinear budget set effects that can be identified by this regression and give automatic debiased machine learners of these effects. We find that in practice nonconvexities in the budget set have little effect on these estimates. We use control variables to allow for endogeneity of budget sets and adjust for productivity growth in taxable income. We apply the results to estimate .52 as the elasticity of an overall tax rate change in Sweden. We also find that the restrictions of utility maximization are satisfied at the choices made by nearly all individuals in the data.

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

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

* The NSF provided support via Grant 1757140. This paper is a revision of "Individual Heterogeneity, Nonlinear Budget Sets, and Taxable Income," and supersedes that paper. We are grateful for comments by R. Blundell, G. Chamberlain, P. Diamond, J. Hausman, C. Manski, R. Matzkin, J. Poterba, H. Selin and participants at seminars at Chicago, UCL (Jan 2010), BC, Harvard/MIT, NYU, and UC Irvine. The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

† Soren Blomquist, Uppsala Center for Fiscal Studies, Department of Economics, Uppsala University, soren.blomquist@nek.uu.se.

‡ Anil Kumar, Federal Reserve Bank of Dallas, anil.kumar@dal.frb.org.

§ Che-Yuan Liang, Uppsala Center for Fiscal Studies, Department of Economics, Uppsala University, Che-Yuan.Liang@nek.uu.se.

± Whitney K. Newey, MIT Department of Economics and NBER, wnewey@mit.edu.

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 a central question was how labor supply responds to tax reform. In a set of influential papers, Feldstein (1995) emphasized that studies focusing on just hours of work neglect many important margins that are distorted by taxes. By estimating how taxable income reacts to changes in the marginal tax rate one captures a larger set of relevant margins, like work effort, job location, tax avoidance and evasion. Inspired by Feldstein's work a large number of studies have produced a wide range of estimates.

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

The use of nonparametric models and methods allows one to avoid errors that can arise from functional form misspecification. Nonparametric models of choice with nonlinear budget sets is potentially important because these models are quite nonlinear with parametric results potentially depending heavily on functional form specifications. In this paper we consider the implications of utility maximization with nonparametric utilities and general heterogeneity across individuals, i.e. in a nonparametric random utility model (RUM) like those of McFadden and Richter (1991), McFadden (2005), Blomquist, Kumar, Liang, and Newey (2015), Blundell, Kristensen, and Matzkin (2014), Manski (2014), Hausman and Newey (2016), Kline and Tartari (2016), and Kitamura and Stoye (2018). A particular focus of this paper is the budget set regression that is the conditional expectation of a choice variable given the budget set, under RUM. This regression is the average over preferences and measurement and/or optimization errors of the choice given the budget set. This regression can be used to identify policy effects that are changes in averages resulting from changes within the range of budget sets observed in the data.

This regression is subject to a potentially severe curse of dimensionality because the conditioning "variable," i.e. the budget set, is a high dimensional object. Blomquist and Newey (2002, BN) showed that with piecewise linear, convex budget sets and scalar, monotonic heterogeneity this regression is only a three dimensional function. This paper shows that the result of BN holds with general heterogeneity and nonparametric utility, i.e. for the RUM. This paper also innovates by giving a more parsimonious specification for the budget set regression than in BN that leads to more precise estimates. We provide corresponding regressors that are easy to calculate and show that the approximation rate for these regressors does not depend on the number of budget set segments or how that varies over individuals. We also show how to check all of the restrictions that the RUM imposes on the budget set regression. The budget set regression also allows for optimization and or measurement errors for the choice variable. Furthermore we show that in practice nonconvexities should have little effect because they are located at low taxable income values

This nonparametric specification avoids errors due to misspecified functional forms and/or heterogeneity distributions. The importance of nonparametric specifications for labor supply was shown by BN who found that the estimated effect of a Swedish tax reform was significantly different for a nonparametric specification rather than a parametric specification. Allowance for general heterogeneity is motivated by Kumar and Liang (2020) who used instrumental variables to estimate average taxable income elasticities that vary widely across different groups of individuals corresponding to different instruments. The allowance for measurement and optimization errors was shown to be important by Burtless and Hausman (1978) and Blomquist (1983) for labor supply. The way we allow for measurement and/or optimization errors is not based on utility maximization as in Chetty (2012) but does allow for low probabilities for kinks, as is often found in data.

For taxable income it is important to allow for productivity growth because over time 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. To nonparametrically separate out the effect of exogenous productivity growth from changes in individual behavior is one of the hardest problems in the taxable income literature. We show how this can be done using the budget set regression.

To evaluate the effect of taxes on taxable income we focus on 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 effects for changes in nonlinear tax systems. We give estimators of the effect of increasing all slopes or intercepts of budget segments on expected taxable income. We use Lasso to estimate the budget set regression and automatic debiased machine learning as in Chernozhukov, Newey, and Singh (2022) to estimate the effects of changing all slopes or intercepts.

We give an application to Swedish data from 1993-2008 for third party reported taxable *labor* income. This means that the variation in the taxable income that we observe for Sweden is mainly driven by variations in effort broadly defined and by variations in hours of work and not by variations in tax evasion.[1]

Budget sets may depend on choices of individuals making it important to allow for endogenous budget sets. We show how endogeneity can be allowed for via control variables while keeping a small dimension for the budget set regression. In the application we allow for endogeneity that arises from spouses income by including a control variable in the budget set regression. We also allow for endogeneity from the choice of locality by including dummy variables for each locality in the budget set regression. For Sweden we do not need to account for endogeneity of the budget set due to choice of whether to own a house or not, unlike US data. This choice only affects the lowest segment and we show that such endogeneity will not affect the results.

As is common in the taxable income literature we do not model the extensive choice; see e.g. Gruber and Saez (2002). We view extending the model to allow for the extensive choice as important but beyond the scope of this paper. Also, we do not expect that extensive choices that primarily affect those with small income will have much impact on the effects we consider that are for averages over the entire distribution of taxable income.

In independent work Manski (2014) considered the identification of general preferences when labor supply is restricted to a finite choice set, with the goal of evaluating tax policy. He finds that tax policy evaluation is difficult when it requires estimation of preferences. BN and this paper bypass estimation of preferences by focusing on policy effects that can be evaluated using the conditional mean, including the effect of tax schedule changes on average taxable income. Imposing the dimension reduction implied by utility maximization makes such policy evaluation feasible when the policy is included within the range of budget sets in the data. Importantly, we and BN also differ from Manski (2014) in allowing for measurement and/or optimization errors in the outcome, which has long been thought to be important for labor supply and taxable income; see Burtless and Hausman (1978).

Aaberge et al. (1995), van Soest (1995), Keane and Moffitt (1998), Blundell and Shephard (2012), and Manski (2014) have considered labor supply when hours are restricted to a finite set. Our taxable income setup could accommodate such constraints, though we do not do this

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

for simplicity. As long as the grid of possible labor supply values was rich enough the expected value of taxable income that we derive would be approximately correct. It appears to be harder to incorporate the bilateral contracting framework of Blundell and Shephard (2012).

The previous literature on taxable income has produced a wide range of estimates. Earlier studies used major tax reforms in the U.S. such as Economic Tax Recovery Act of 1981 and Tax Recovery Act of 1986 as natural experiments and, using lower income groups as control groups for the treated group of higher income taxpayers, estimated elasticities close to 1 or higher (Lindsey, 1987; Feldstein, 1995; Eissa, 1995; Navratil, 1995; Feldstein and Feenberg, 1995).

To address the identification challenges of the natural experiment approach, e.g., highlighted in Goolsbee (1999, 2000), subsequent research used panel data spanning multiple tax reforms and employed synthetic tax changes as instruments to address the endogeneity of marginal tax rate changes, finding more modest elasticities mostly ranging from 0.2-0.5 (Auten and Carroll, 1999; Gruber and Saez, 2002; Saez, 2003; Kopczuk, 2005; Giertz, 2007).

However, such synthetic tax changes were still dependent on prior year income, which could be problematic if the error term was serially correlated or if elasticities were heterogeneous across income groups. To circumvent these concerns, recent papers proposed more refined instruments (Blomquist and Selin, 2010;Weber, 2014; Burns and Ziliak, 2017; Kumar and Liang, 2020). Except for the Blomquist and Selin study, these latter studies find higher elasticties ranging from 0.6 to 1.4

Our work is also related to a large literature that, following Burtless and Hausman (1978), estimated the effect of taxes on labor supply using maximum likelihood techniques to model labor supply choice over piecewise-linear budget constraints (Hausman 1980, 1981a; Hausman, 1985; Blomquist and Hansson-Brusewitz, 1990; Bourgiugnon and Magnac, 1990; Triest,1990; Van Soest et al., 1990; Colombino and Del Boca, 1990, Heim, 2009).

The rest of our paper is organized as follows. Section 2 gives the model of individual behavior we consider. Section 3 derives the budget set regression and gives regressors that can be used to estimate it. Section 4 describes the effects we consider and automatic debiased machine learners of them. Section 5 gives the application to Swedish data. Section 6 gives the stochastic revealed preference results. Section 7 gives additional RUM theory and shows how that lessens the importance of nonconvexities in the budget set. Section 8 concludes.

2 RUM Choice for Linear Budget Sets

Feldstein (1995) argued that individuals have more margins than hours of work to respond to changes in the tax. For example, individuals could exert more effort on the present job, switch to a better paid job that requires more effort, or could move geographically to a better-paid job. The choice of compensation mix (cash versus fringe benefits) and tax avoidance/evasion are still other margins. The application we consider is for taxable earned income where tax evasion is not a concern, but allowing for an effort margin seems useful especially for allowing for productivity growth over time.

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

$$\underset{c,e,h}{\text{Max}}\, u(c, e, h) \quad s.t. \quad c = w(e)h\rho + R, c \geq 0, e \geq 0, h \geq 0. \tag{2.1}$$

This problem can be reformulated as a choice of consumption and taxable income $y = w(e)h$ when the wage function is strictly monotonic increaseing in e. Since $w(e) = y/h$ inverting gives $e = w^{-1}(y/h)$. Noting that only y enters the constraint we can concentrate h out of the choice problem by choosing h to maximize $u(c, w^{-1}(y/h), h)$ and then maximizing over c and y. Let $U(c, y) = \max_{h \geq 0} u(c, w^{-1}(y/h), h)$ be the resulting concentrated utility function. The choice of c and y will maximize this concentrated utility function subject to the linear budget constraint by solving

$$\underset{c,y}{\text{Max}}\, U(c, y) \quad s.t. \quad c = y\rho + R, c \geq 0, y \geq 0. \tag{2.2}$$

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

In the taxable income literature one usually starts with individual choice of consumption c and taxable income y as given by equation (2.2) and specifies that $U(c, y)$ is strictly quasi-concave. We first specified utility as a function of effort to illustrate how the taxable income specification allows for another margin of choice. Strict quasi-concavity of $u(c, e, h)$ does not necessarily imply strict quasiconcavity of $U(c, y)$, but we simply follow the taxable income literature by specifying that $U(c, y)$ is strictly quasi-concave for each individual.

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

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

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

This specification allows for preferences to vary across individuals in essentially any way at all. For example the effect of unearned income on y may vary with individuals analogously to the way the effect of unearned income on labor supply varies with individuals in Burtless and Hausman (1978). We do need to restrict η and $U(c, y, \eta)$ so that probability statements can be made but these are technical side conditions that do not affect our interpretation of η as representing general heterogeneity and are reserved for the Appendix. Here we make the following Assumption about $U(c, y, \eta)$ and $y(\rho, R, \eta)$:

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

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

This model is a RUM as discussed in the Introduction. The model here specializes the RUM to $U(c, y, \eta)$ that are strictly quasi-concave and $y(\rho, R, \eta)$ that is smooth in ρ and R, while leaving the effect of heterogeneity η on $y(\rho, R, \eta)$ entirely unrestricted. Single valued, smooth demand specifications are often used in applications and we follow that practice here. In particular, smoothness has often proven useful in applications of nonparametric models and it will here.

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

$$F(y|\rho, R) = \int 1(y(\rho, R, \eta) \leq y) G(d\eta) \tag{2.3}$$

This CDF plays a pivotal role in RUM choice theory for nonlinear budget sets in Sections 3 and 7 to follow.

It is important to account for productivity growth when identifying the effects of taxes on taxable income using variation over time as we do in the application below. We account for productivity growth by assuming the existence of a function $\phi(t)$ such that the wage in period t for individual η with effort level e is $w = g(e,\eta)\phi(t)$. This specification allows for heterogeneity in the effect of productivity growth on individuals through the η in $g(e,\eta)$. For example, if $g(e,\eta) = g_1(e,\eta_1)\eta_2$ then the wage in period t would be $g_1(e,\eta_1)\eta_2\phi(t)$ with $\eta_2\phi(t)$ being an effect of productivity growth on wages that varies with individuals. This specification is restrictive in specifying all individuals with the same effort level in two periods will have the same wage growth rate over those two periods, i.e. that the productivity growth rate does not depend on the individual. The function $\phi(t)$ captures exogenous productivity growth that does not depend on individual behavior. We impose the condition that $\phi(\bar{t}) = 1$ for some time period $\bar{t}$.

With productivity growth and heterogeneity the individual's optimization problem is:[2]

$$\text{Max}\, u(c,e,h,\eta) \quad \text{s.t.} \quad c = g(e,\eta)\phi(t)h\rho + R \tag{2.4}$$

Here $g(e,\eta)\phi(t)$ is the wage in period t for individual η that is determined both by effort, the time period, and by heterogeneity η. This utility maximization problem can be solved similarly to previous one, by letting $y = wh$, inverting the wage function, and choosing hours of work to maximize $u(c, g^{-1}(y/(h\phi(t)),\eta), h, \eta)$ over h. Concentrating out hours of work gives the concentrated utility function $U(c, y/\phi(t), \eta)$. In a second step the individual solves $\text{Max}\, U(c, y/\phi(t), \eta)$ s.t. $c = y\rho + R$.

A feature of this problem is that the concentrated utility shifts over time. Our approach to repeated cross section data depends on using a preference specification invariant to individuals and time. A simple way to do that is to focus on taxable income net of productivity growth, given by $\tilde{y} = y/\phi(t)$. Here the reduced-form maximization problem becomes $\text{Max}\, U(c,\tilde{y},\eta)$ s.t. $c = \tilde{y}\tilde{\rho} + R$ for $\tilde{\rho} = \phi(t)\rho$. Thus, the productivity growth appears in the budget set, multiplying the net of tax rate. From the tax authorities' point of view the taxable income is $y = \phi(t)\tilde{y}$. However, to keep things stationary over time we study the behavior of $\tilde{y}$. Although the function $U(c,\tilde{y},\eta)$ does not shift over time, it depends on a base year $\bar{t}$ and a normalization of $\phi(\bar{t})$ to one. If we use another base year we would have another concentrated utility function.

This way to account for productivity growth is similar to that used in log-linear models. Suppose that $\tilde{y}(\rho,R,\eta) = [\phi(t)\rho]^{\beta}\eta$, where β is the net-of-tax elasticity that is the same for all individuals. Taking logarithms gives $\ln y = (\beta+1)\ln\phi(t) + \beta\ln\rho + \ln\eta$. 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

[2] 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.

previous models of taxable income. Including time effects in log-linear models corresponds to the productivity growth specification we adopt here.

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 that it may be possible to separate out the effect of productivity growth and tax rate changes in our estimates.

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

It would be interesting to know whether the productivity growth specification given here is sufficiently rich in the way it allows productivity growth to differ across individuals. We checked for this in the application by allowing productivity growth to vary between more and less educated individuals. We found that allowing this observed heterogeneity does not impact our tax and income effect estimates. More generally the use of a single budget set regression to explain a time series of cross section relies on a stable utility function over time making it difficult to allow more general heterogeneity than considered here.

3 The Budget Set Regression for Piecewise Linear Budget Sets

The budget set regression is the expected value of taxable income conditional on the budget frontier. This regression can be used for estimating policy effects when the regression is stable across different policy regimes. It can be used to predict the effect of tax schedule changes on average taxable income when there is sufficient variation in budget frontiers to identify such effects. In this Section we describe this regression when the budget frontier is piecewise linear and continuous.

The budget set regression has many regressors including the intercepts and slopes of all the segments. It can be difficult to estimate high dimensional regressions. Here we use restrictions from utility maximization to reduce the dimension of the budget set regression. This dimension reduction makes budget set regression feasible in practice.

We focus first on increasing marginal tax rates and give generalizations that allow for marginal tax rates to decrease. We can describe a continuous, piecewise linear budget frontier as one with J segments with slope ρ_j and intercept R_j for the j^{th} segment, $(j = 1, ..., J)$. By continuity of the budget frontier the j^{th} segment ends at the kink $\ell_j = (R_j - R_{j+1})/(\rho_{j+1} - \rho_j)$, $(j = 1, ..., J-1)$. The budget frontier can be thought of as a $2J+1$ vector of the form $B = (J, \rho_1, R_1, ..., \rho_J, R_J,)'$. To describe the budget set regression, recall that $F(y|\rho, R)$ is the CDF of taxable income for a linear budget set. Define

$$\begin{aligned} \bar{y}(\rho, R, F) &= \int y F(dy|\rho, R), \qquad (3.5) \\ \nu(\rho, R, \ell, F) &= \int 1(y < \ell)(y - \ell) F(dy|\rho, R), \\ \lambda(\rho, R, \ell, F) &= \int 1(y > \ell)(y - \ell) F(dy|\rho, R). \end{aligned}$$

The first object is the conditional mean of taxable income for a linear budget set with slope ρ and intercept R. The other objects are integrals over values of Y below or above a possible kink point ℓ. Note that all of these objects are known functions of the conditional CDF $F(y|\rho, R)$ of taxable income for a linear budget set. Thus, these objects depend only on $F(y|\rho, R)$.

The budget set regression is given in the following result. Let $\mu(B) = E[Y|B]$.

THEOREM 1: *If Assumption 1 and A1 are satisfied,* $\int |y(\rho, R, \eta)| G(d\eta) < \infty$ *for all* $\rho, R > 0$, $\rho_1 > ... > \rho_J$, *and* B *is independent of* η *then*

$$\begin{aligned} \mu(B) &= \bar{y}(\rho_J, R_J, F) + \sum_{j=1}^{J-1} [\nu(\rho_j, R_j, \ell_j, F) - \nu(\rho_{j+1}, R_{j+1}, \ell_j, F)] \qquad (3.6) \\ &= \bar{y}(\rho_1, R_1, F) + \sum_{j=1}^{J-1} \left[\lambda(\rho_{j+1}, R_{j+1}, \ell_j, F) - \lambda(\rho_j, R_j, \ell_j, F)\right]. \end{aligned}$$

This result shows the budget set regression depends only on a single three dimensional function, the CDF of taxable income $F(y|\rho, R)$ for a linear budget set. Based on this feature we will give a three dimensional series approximation in this Section, which is much smaller dimensional that the $2J+1$ dimensions of B. In this way utility maximization reduce the dimension of the budget set regression.

Theorem 1 generalizes Theorem 2.1 of BN by allowing general heterogeneity, whereas BN assumed scalar η. Theorem 1 shows that the BN decomposition of the budget set regression into the sum of a two dimensional and a three dimensional function holds without the restriction of scalar heterogeneity. Consequently, the empirical conclusions drawn by BN about the average labor supply effect of the 1980 to 1991 Swedish tax reform are valid under nonparametric utility and general heterogeneity as stated in Section 2.

Theorem 1 innovates relative to BN in the budget regression depending only on a single three dimensional function. In contrast BN showed that the budget set regression was the sum of a two-dimensional function (of ρ_J, R_J) and a three dimensional function ($v(\rho, y, \ell)$), i.e. that the expression following the first equality holds. Theorem 1 shows that there is a cross function restriction where the two functions $\bar{y}(\cdot)$ and $v(\cdot)$ each depend on $F(y|\rho, R)$ in a known way. This more parsimonious specification makes the budget set regression easier to estimate. Theorem 1 also innovates by allowing for zero hours of work, rather than the scalar heterogeneity considered in BN.

The budget set regression of Theorem 1 allows for measurement and/or optimization errors of multiplicative and/or additive forms. Allowing for such disturbances is consistent with the fact that in most data sets there is no bunching at kink points of budget sets, as in Burtless and Hausman (1978). To explain, let Y^* denote an observed choice of an individual in the absence of disturbances and assume that equation (3.6) is satisfied with Y^* replacing Y. Suppose that the observed Y satisfies $Y = \varepsilon_1 + \varepsilon_2 Y^*$ where ε_1 and ε_2 are disturbances with $E[\varepsilon_1|B] = 0$ and $E[\varepsilon_2|B, Y^*] = 1$. Then by iterated expectations

$$E[Y|B] = E[\varepsilon_1|B] + E[E[\varepsilon_2|B, Y^*]Y^*|B] = 0 + E[Y^*|B] = \mu(B).$$

It is true that disturbances ε_1 and ε_2 having conditional means 0 and 1 would not generally arise from optimization errors as modeled by Chetty (2012) but optimization errors need not have that form. The type of errors allowed here are plausible in allowing random departures of taxable income from utility maximization. A virtue of modeling the conditional mean rather than the entire conditional distribution of taxable income is that such errors are allowed for.

We can use Theorem 1 to construct a linear in parameters approximation to the budget set regression by specifying a linear in parameters approximation to the conditional CDF. To do so let M be a positive integer and $F_1(y), ..., F_M(y)$ be CDF's chosen by the researcher. For example, these CDF's could be chosen to give more weight to different intervals of taxable income values. They are used here to approximate the counterfactual distribution of taxable income given a linear budget set but they could be based on the observed taxable income. For example $F_1(y)$ could be the empirical CDF of taxable income in the data and $F_m(y)$ could be empirical CDF's of lower and upper parts of the data. The use of the data in this way will not bias the budget set regression because all that is required to estimate the regression is a sufficiently rich approximation.

Also let $x := (\rho, R)$ and let $r_1(x), ..., r_K(x)$ denote approximating functions such as splines or power series. For example a linear approximation could specify

$$(r_1(x), ..., r_K(x))' = (1, \rho, R).$$

Let β_{mk}, $(m = 2, ..., M; k = 1, ..., K)$ be coefficients of a series approximation to be specified below and $w_m(x, \beta) = \sum_{k=1}^{K} \beta_{mk} r_k(x)$. An approximation to the conditional CDF can be constructed as

$$\begin{aligned} F(y|x, \beta, K, M) &= F_1(y) + \sum_{m=2}^{M} w_m(x, \beta)[F_m(y) - F_1(y)] \qquad (3.7) \\ &= \sum_{m=1}^{M} w_m(x, \beta) F_m(y), \; w_1(x, \beta) = 1 - \sum_{m=2}^{M} w_m(x, \beta). \end{aligned}$$

This could be thought of as a mixture approximation to the conditional CDF with weights $w_m(x, \beta)$, $(m = 1, ..., M)$, that are linear in parameters. The normalization that the weights sum to 1 is imposed by the formula for $w_1(x, \beta)$. This leads to a CDF approximation having the important property that for all β,

$$\lim_{y \longrightarrow \infty} F(y|x, \beta, K, M) = 1.$$

The conditional pdf could also be restricted to be nonnegative by requiring that $w_m(x, \beta) \geq 0$ on a grid of values for x. For computational simplicity we do not require that this nonnegativity restriction hold. We are primarily interested in approximating the budget set regression and the CDF approximation in equation (3.7) gives a linear in parameters approximation to the budget set that can be estimated by linear regressions. Imposing nonnegative weights would complicate estimation of the budget set regression.

We obtain an approximation to the budget set regression by plugging the CDF approximation into the respective formulas for $\bar{y}(\rho, R)$ and $\nu(\rho, R, \ell)$ and these into the formula for conditional mean in Theorem 1. Let

$$\bar{y}_m = \int y F_m(dy), \; \nu_m(\ell) = \int 1(y < \ell)(y - \ell) F_m(dy), \; (m = 1, ..., M).$$

Substituting the CDF approximation in the expression for the conditional mean $\mu(B)$ in Theorem 1 give

$$\begin{aligned} \mu(B, \beta, K, M) &= \bar{y}_1 + \sum_{m=2}^{M} w_m(x_J, \beta)(\bar{y}_m - \bar{y}_1) \\ &\quad + \sum_{m=2}^{M} \sum_{j=1}^{J-1} [w_m(x_j, \beta) - w_m(x_{j+1}, \beta)][\nu_m(\ell_j) - \nu_1(\ell_j)] \\ &= \bar{y}_1 + \sum_{m=2}^{M} \sum_{k=1}^{K} \beta_{mk} p_{mk}(B), \\ p_{mk}(B) &= r_k(x_J)(\bar{y}_m - \bar{y}_1) + \sum_{j=1}^{J-1} [r_k(x_j) - r_k(x_{j+1})][\nu_m(\ell_j) - \nu_1(\ell_j)]. \end{aligned}$$

This has a linear regression form where the regressor with coefficient β_{mk} is $p_{mk}(B)$. Each such regressor is a linear combination of the k^{th} approximating function evaluated at the last segment $r_k(x_J)$ and the differences $r_k(x_j) - r_k(x_{j+1})$, $(j = 1, ..., J-1)$ for adjacent segments. The dimension reduction is achieved by the regressors being linear combinations of differences across adjacent segments.

For example with three segments and a linear approximation where $(r_1(x), ..., r_K(x)) = (1, \rho, R)$ there are three regressors for each m given by

$$\bar{y}_m - \bar{y}_1,\ \rho_3(\bar{y}_m - \bar{y}_1) + (\rho_1 - \rho_2)[\nu_m(\ell_1) - \nu_1(\ell_1)] + (\rho_2 - \rho_3)[\nu_m(\ell_2) - \nu_1(\ell_2)],$$
$$R_3(\bar{y}_m - \bar{y}_1) + (R_1 - R_2)[\nu_m(\ell_1) - \nu_1(\ell_1)] + (R_2 - R_3)[\nu_m(\ell_2) - \nu_1(\ell_2)].$$

Also, the regressors for each m will include a constant and it is only necessary to include one constant term in the regression, so the total number of regressors in this example is $2(M-1) + 1 = 2M - 1$.

These regressors will vary across individuals through variation in the budget sets across individuals. The regressors described here are specific to individual budget sets and so vary with those budget set. These regressors may vary across individuals by the number of segments J and by slopes and intercepts ρ_j, R_j, and hence also by ℓ_j and $v_m(\ell_j)$, $m = (2, ..., M)$. For each individual there will be $(M-1)(K-1) + 1$ regressors with the form of those regressors depending on the individual's budget set.

This regression specification must be modified to account for productivity growth. As discussed in Section 2 we allow for productivity growth by assuming that in period t the individuals are maximizing utility over $\tilde{y} = y/\phi(t)$ for a utility function of the form $U(c, \tilde{y}, \eta)$, where η is drawn from the same distribution in each time period and for every individual and $\phi(t)$ is productivity growth. Consequently each individual in time period t will behave as if each of their ρ_j is $\tilde{\rho}_j = \phi(t)\rho_j$ with corresponding kink points $\tilde{\ell}_j(t) = (R_j - R_{j+1})/(\tilde{\rho}_{j+1} - \tilde{\rho}_j) = \phi(t)^{-1}\ell_j$. Also the budget set regression is that for $\tilde{Y}$ which must be multiplied by $\phi(t)$ to obtain the regression for Y. Consequently with productivity growth the budget set regression depends on

t and for $\tilde{x}(t) = (\phi(t)\rho, R)$ is given by

$$\begin{aligned}
\mu(B,\beta,K,M,t) &= \phi(t)\{\bar{y}_1 + \sum_{m=2}^{M} w_m(\tilde{x}_J(t),\beta)(\bar{y}_m - \bar{y}_1) \\
&\quad + \sum_{m=2}^{M}\sum_{j=1}^{J-1}[w_m(\tilde{x}_j(t),\beta) - w_m(\tilde{x}_{j+1}(t),\beta)][\nu_m(\tilde{\ell}_j(t)) - \nu_1(\tilde{\ell}_j(t))]\} \\
&= \phi(t)\bar{y}_1 + \sum_{m=2}^{M}\sum_{k=1}^{K}\beta_{mk}p_{mk}(B,t), \\
p_{mk}(B,t) &= \phi(t)\{r_k(\tilde{x}_J(t))(\bar{y}_m - \bar{y}_1) + \sum_{j=1}^{J-1}[r_k(\tilde{x}_j(t)) - r_k(\tilde{x}_{j+1}(t))][\nu_m(\tilde{\ell}_j(t)) - \nu_1(\tilde{\ell}_j(t))]\}.
\end{aligned}$$

Here the regressor $p_{mk}(B,t)$ depends on both the budget set variables B and the time period t. It is important to note that none of these regressors is constant over time even when $r_k(x) = 1$ for some k. Because the conditional mean of $\tilde{Y} = Y/\phi(t)$ has been multiplied by $\phi(t)$ any regressors for $\tilde{Y}$ that are constant have been multiplied by $\phi(t)$.

The budget set regression we have described can be theoretically justified by showing that the specification provides an approximation rate to the true regression that is uniform in individual budgets sets. For simplicity we do this for specific choices of distributions $F_m(y)$ and approximating functions $r_k(x)$. We expect that this approximation result would also apply to other distributions and approximating functions. We consider $F_m(y)$ to be integrals of b-splines that are positive and normalized to integrate to one and $r_b(x)$ that are also splines. Splines are piecwise polynomial hthat are continuously differentiable to order one less the order of the polynomial and b-splines have this form while being nonnegative and nonzero on a bounded set. We also require that y and x be contained in bounded sets $\mathcal{Y}$ and $\mathcal{X}$ and that the conditional pdf of taxable income for a linear budget set $f(y|x)$ be smooth. For simplicity we state the result assuming there is no productivity growth with the understanding that the conclusion also applies with productivity growth.

Theorem 2: *If $\mathcal{Y}$ and $\mathcal{X}$ are compact, $f(y|x)$ is zero outside $\mathcal{Y} \times \mathcal{X}$ and is continuously differentiable to order s on $\mathcal{Y} \times \mathcal{X}$, and $dF_m(y)/dy, (m = 1, ..., M)$ and $r_k(x), (k = 1, ..., K)$ consist of tensor product b-splines of order s on $\mathcal{Y} \times \mathcal{X}$ then there exist a constant C and β_{km} such that for all piecewise linear, concave budget frontiers with $x_j \in \mathcal{X}$, $(j = 1, ..., J)$,*

$$\left|\mu(B) - \bar{y}_1 - \sum_{m=2}^{M}\sum_{k=1}^{K}\beta_{mk}p_{mk}(B)\right| \leq C(MK)^{(1-s)/3}.$$

This result shows that the series approximation we have proposed does approximate the expected value of taxable income for piecewise linear budget sets with increasing marginal tax rates. We emphasize that approximation rate is uniform in the number J of budget segments and other aspects of the budget set under the conditions given here. Thus budget sets may vary over individuals in quite general ways while maintaining low bias from using the budget set regressions given here. The rate of approximation corresponds to a multivariate b-spline approximation to a function and its derivative, where approximating the derivative is useful for making the rate uniform in the number of budget segments.

We can also make allowance for nonconcave budget frontiers. For example, suppose that the budget frontier is nonconcave over only two segments. As shown in Section 7, the distribution over those segments will depend only on the slope and intercept of those two segments. Because the budget set regression is a sum of integrals over different segments it would take the form

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

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

Endogeneity of budget sets is a potentially important problem to overcome. In the data we consider spouse's income is included in nonlabor income which creates endogeneity due to correlation in couples' preferences. Also, tax rates vary across locations creating endogeneity due to correlation between preferences for location and the choice of taxable income. Control variables can be used to account for endogeneity. A control variable is ξ such that B and η are independent conditional on ξ and the conditional support of ξ given B equals the marginal support of ξ. In that case it follows as in Blundell and Powell (2006) and Hausman and Newey (2016) that

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

where $F_\xi(\xi)$ is the marginal CDF of ξ. This integral can be estimated by letting the functions $r_b(x,\xi)$ depend on ξ and letting the regressor corresponding to coefficient β_{mk} be

$$p_{mk}(B,\xi,t) = \phi(t)\{r_k(\tilde{x}_J(t),\xi)(\bar{y}_m - \bar{y}_1) + \sum_{j=1}^{J-1}[r_k(\tilde{x}_j(t),\xi) - r_k(\tilde{x}_{j+1}(t),\xi)][\nu_m(\tilde{\ell}_j(t)) - \nu_1(\tilde{\ell}_j(t))]\}.$$

A budget set regression with endogeneity controls is then obtained from regressing taxable income on these budget set regressors to obtain coefficients β_{mk}, $(m = 1, ..., M; k = 1, ..., K)$.

Integrating over the control variable then gives an approximation to the average taxable income function obtained by integrating over the marginal distribution of preferences, i.e.

$$\mu(B) \approx \phi(t)\bar{y}_1 + \sum_{m=2}^{M}\sum_{k=1}^{K}\beta_{mk}\bar{p}_{mk}(B,t),\ \bar{p}_{mk}(B,t) = \int p_{mk}(B,\xi,t)F_{\xi}(d\xi).$$

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

4 Estimation of Budget Set Effects

It is common practice to measure behavioral effects in terms of elasticities. Often the object of interest has been the elasticity with respect to the net of tax rate for a linear budget constraint. A 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 $\bar{y}(\rho, R)$. As discussed in Section 7, stringent conditions are required to identify this function. There it is shown that the set of net of tax rates where this function is identified can be very small, even empty. Therefore we must look for other kinds of elasticities to hope for identification. Furthermore, since everyone generally faces a nonlinear budget set, and policy changes are not likely to eliminate this nonlinearity, effects from changes in nonlinear budgets are important object of interest.

For nonlinear budget sets the expected taxable income will be $\mu(B) = \mu(J, \rho_1, \ldots, \rho_J, R_1, \ldots, R_J)$. We consider the effect on expected taxable income of an increase in all slopes of the budget set segments. When $\mu(B)$ is a differentiable function of the slopes this object is

$$\theta_{\rho 0} = E\left[\frac{\partial\mu(J, \rho_1 + a, \ldots, \rho_J + a, R_1, \ldots, R_J)}{\partial a}\right] = E\left[\sum_{j=1}^{J}\frac{\partial\mu(J, \rho_1, \ldots, \rho_J, R_1, \ldots, R_J)}{\partial\rho_j}\right].$$

by the chain rule, where the expectation is over B. It is plausible that this object could be identified from variation in the overall slopes in the data. Such variation could arise from changes in the size of taxes. Similarly one could consider a nonlabor income effect that is the average effect of increasing each R_j,

$$\theta_{R0} = E\left[\sum_{j=1}^{J}\frac{\partial\mu(J, \rho_1, \ldots, \rho_J, R_1, \ldots, R_J)}{\partial R_j}\right].$$

The $\theta_{\rho 0}$ is the average effect of rotating the budget frontier and θ_{R0} the average effect of shifting the budget frontier vertically, both while holding fixed the kink points. For policy purposes

$\theta_{\rho 0}$ is like a change in a proportional tax rate, and θ_{R0} is like a change in unearned income. Identification of $\theta_{\rho 0}$ and θ_{R0} 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 overall degree of taxation and in nonlabor income, so it is plausible that effects of such changes should be identified.

One can also specify an elasticity object corresponding to $\theta_{\rho 0}$ by multiplying θ_ρ by a constant $\tilde{\rho}$ that summarizes the vector of net-of-tax rates in a single number and then dividing by $\mu(B)$ to obtain

$$e_\rho = \tilde{\rho}\theta_{\rho 0}/\mu(B).$$

The construction of $\tilde{\rho}$ can be done in many different ways. We use the averages of the net-of-tax rates and virtual incomes for the segments where individuals are actually located. The elasticity e_ρ is an aggregate elasticity which is the policy relevant measure as argued in Saez et. al. (2012). For nonlabor income we stick with θ_{R0} as an income effect because there is more ambiguity in how to construct an income elasticity.

The parameters of interest $\theta_{\rho 0}$ and θ_{R0} are average derivatives of the budget set regression. For computational purposes it is useful to work with corresponding difference quotients. Also, to describe the estimators of these objects it is useful to work with slightly different notation for the budget set regression than we have used previously. For this purpose let Z denote the vector consisting of all the budget set variables B and controls ξ adjusted for productivity growth according to the time period t corresponding to Z. Also let Z_ρ^δ denote Z with δ added to each ρ_j in B and Z_R^δ denote Z with δ added to each R_j. Difference quotient versions of $\theta_{\rho 0}$ and θ_{R0} are

$$\theta_{\rho 0} = E\left[\frac{\mu(Z_\rho^\delta) - \mu(Z_\rho^{-\delta})}{2\delta}\right],\ \theta_{R0} = E\left[\frac{\mu(Z_R^\delta) - \mu(Z_R^{-\delta})}{2\delta}\right].$$

The budget set regressors described in Section 3 depends on slopes and intercepts of budget sets in various ways and are quite high dimensional once control variables of various kinds are included. For this reason we use Lasso regression which is known to have good prediction properties for high dimensionsal regression while allowing for high multicollinearity among regressors. Lasso regression estimators are biased by model selection and shrinkage, so to reduce that bias for estimating the parameters of interest $\theta_{\rho 0}$ and θ_{R0} we use automatic debiased machine learning as in Chernozhukov, Newey, and Singh (2022, CNS henceforth). This method corrects the shrinkage and model selection bias of Lasso in a way that is specific to the parameter of interest $\theta_{\rho 0}$ or θ_{R0}.

To describe the estimator let i denote the observation, so that the adjustment for productivity growth for observation i is captured in the vector Z_i, and let n denote the total number of observations. Let $b(Z)$ denote a $q \times 1$ dictionary of functions of Z that includes the productivity adjusted regressors we have described and their interactions with controls, where q is determined by the number M of CDF's and the number K of approximating functions in the budget set regressors as well as the number of controls. As usual for Lasso we normalize each regressor $b_j(Z)$ to have sample mean 0 and sample variance 1, $(j = 1, ..., q)$. As in CNS we use cross-fitting where averaging takes place over observations different than used in the budget set regression. To describe the cross fitting let I_ℓ, $(\ell = 1, ..., L)$, be a partition of the observation index set $\{1, ..., n\}$ into L distinct subsets of about equal size. In the application we use $L = 5$ (5-fold) partitioning as is also used in CNS and Chernozhukov et al. (2018). The Lasso regression estimator $\hat{\mu}_\ell$ constructed from the observations that are *not* in I_ℓ is given by

$$\hat{\mu}_\ell(z) = \bar{Y}_\ell + b_\ell(z)'\hat{\beta}_\ell,\ \hat{\beta}_\ell = \arg\min_\beta \left\{ \frac{\sum_{i \notin I_\ell}[Y_i - b_\ell(Z_i)'\beta]^2}{n - n_\ell} + 2\lambda_\ell \sum_{j=1}^{p} |\beta_j| \right\},$$

where n_ℓ is the total number of observations in I_ℓ, $\bar{Y}_\ell$ is the sample mean of Y_i over $i \notin I_\ell$, $b_\ell(z)$ each have sample mean 0 and variance 1 over $i \notin I_\ell$, and λ_ℓ is a penalty degree. In the application λ_ℓ is chosen by cross-validation over $i \notin I_\ell$.

The automatic debiased machine learner of CNS makes use of a bias correction term constructed by a Lasso minimum distance procedure. To describe this bias correction term let $\mu(Z)$ denote any function of Z and define $m(Z, \mu)$ to be either $[\mu(Z_\rho^\delta) - \mu(Z_\rho^{-\delta})]/2\delta$ or $[\mu(Z_R^\delta) - \mu(Z_R^{-\delta})]/2\delta$ corresponding to θ_ρ or θ_R respectively. Thus, $m(Z, b_j)$ is the difference quotient applied to the j^{th} dictionary function $b_j(z)$. Let $m(Z_i, b) = (m(Z_i, b_1), ..., m(Z_i, b_p))'$. The bias correction term of CNS is constructed from observations not in I_ℓ as

$$\hat{\alpha}_\ell(z) = b_\ell(z)'\hat{\rho}_\ell,\ \hat{\rho}_\ell = \arg\min_\rho \{-2\hat{M}_\ell' + \rho'\hat{G}_\ell\rho + 2r\sum_{j=1}^{J} |\rho_j|\}, \tag{4.8}$$
$$\hat{M}_\ell = \frac{1}{n - n_\ell}\sum_{i \notin I_\ell} m(Z_i, b),\ \hat{G}_\ell = \frac{1}{n - n_\ell}\sum_{i \notin I_\ell} b(Z_i)b(Z_i)',$$

where $r > 0$ is a positive scalar that specifies a degree of penalization. The r may be chosen by a cross-validation method or a plug-in method as described in CNS.

An automatic debiased machine learner of an object of interest is straightforward to construct from the Lassso regression $\hat{\mu}_\ell(z)$ and bias correction $\hat{\alpha}_\ell(z)$ for $\ell = 1, ..., L$. Let $m(z, \hat{\mu}_\ell(z))$ be the difference quotient applied to the Lasso regression for each ℓ. An estimator $\hat{\theta}$ of an object of interest and an asymptotic variance estimator $\hat{V}$ are given by

$$\hat{\theta} = \frac{1}{n}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\{m(Z_i,\hat{\mu}_\ell) + \hat{\alpha}_\ell(Z_i)[Y_i - \hat{\mu}_\ell(Z_i)]\}, \tag{4.9}$$

$$\hat{V} = \frac{1}{n}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\hat{\psi}_{i\ell}^2,\ \hat{\psi}_{i\ell} = m(Z_i,\hat{\mu}_\ell) - \hat{\theta} + \hat{\alpha}_\ell(Z_i)[Y_i - \hat{\gamma}_\ell(Z_i)].$$

Under regularity conditions given in CNS, for the true value θ_0 of the parameter of interest it will be the case that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V),\ \hat{V} \xrightarrow{p} V,$$

so that large sample confidence intervals and tests concerning θ_0 can be constructed from $\hat{\theta}$ and $\hat{V}$ in the usual way.

5 Application to Sweden

5.1 Data

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

We use data covering a period of sixteen years, from 1993 to 2008. In the estimation, we limit the sample to married or cohabiting men between 21 and sixty years of age. This is the economically most significant group with respect to labor income. We exclude those receiving medical-leave benefits, parental benefits, income from self-employment, or student financial aid above half of the average monthly gross labor income, which was 17,607 SEK in 2008. We use this limit instead of zero, as that would result in a large loss of observations. Out of 102,630 observations, 81,717 observations remain after these 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 individuals with large amounts of income from these sources are excluded from the sample.

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

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

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

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

Summary statistics of the sample used for estimation is reported in Table 1 of Appendix C. We report the mean values along with standard deviations of gross labor income, some variables characterizing the budget sets, key demographic variables, and transfers at zero labor income, which is the instrument for nonlabor income. We report statistics for the entire sample, as well as for the years 1993, 1998, 2003, and 2008 separately, to illustrate the development over time. Figure 1 presents the variation in net-of-tax-rates in the estimation sample. After declining in the early 1990's the first segment and marginal net-of-tax-rates rose sharply in the late 1990's and early 2000's, though their growth was notably slower in the late 2000's. On the other hand, the last segment net-of-tax rate saw a more significant decline in the early 1990's, followed by modest fluctuations.

In a given time period the budget constraint varies across individuals for several reasons, differences in local tax rates, differences in itemized deductions and different demographic characteristics. As mentioned above we attempt to correct for endogeneity issues caused by the differences in local tax rates by using county dummy variables as covariates

5.2 Regression Results

We use Lasso regression with a power series approximation for terms up to a fourth order using the procedures described in Section 3. For the approximation we use three CDF's $F_m(y),(m = 1,2,3)$, corresponding to the marginal empirical distribution of taxable income for the whole sample, for the smallest third of the sample, and for the largest third of the sample. For constructing $\mu_a(\ell)$ we use the sample CDF for a subsample of 5,000 observations evenly by gross labor income rank. We use a power series approximation with $r_k(x) = \rho^{m_1(k)}R^{m_2(k)}$, $m_1(k) + m_2(k) \leq 4$ for $m_1(k)$ and $m_2(k)$ equal to nonnegative integers.

To account correctly for productivity growth we need to estimate it. Unfortunately there are few good measures of the exogenous wage/productivity growth, as the productivity measures available in the literature generally do not disentangle the change in wages due to behavioral effects of tax changes. We will therefore use our data to estimate exogenous productivity growth. To guard against using too much identifying information when doing this, we constrain the annual productivity growth to be the same every year, i.e. $\phi(t) = e^{gt}$ for some constant g. While a constant productivity growth assumption may lead to some misspecification, a more refined correction of the budget constraints would use up much of the information in the data. In particular we would lose much of the identifying power of changes in the overall tax rate across years. We do not think there are wide swings in the productivity growth rate from year to year so that the misspecification would not be very large for any individual year's budget constraints.

Marginal effects are evaluated at each of the sample individual budget sets and the net-of-tax-rate elasticity is based on sample average marginal effects scaled by the sample mean income and marginal net-of-tax rate. The elasticities are evaluated at the sample mean income of 420,000 SEK and marginal net-of tax rate of .42. We also report the marginal effect of nonlabor income, i.e., the change in gross labor income as net nonlabor income increases by one SEK. For nonlabor income, we report the marginal effect rather than the elasticity because the income elasticity is highly dependent on the scaling parameters. The natural scaling factor for nonlabor income—the average marginal nonlabor incxome—is extremely high in our case compared to other studies that report income elasticities because the marginal tax rate is low and nonlabor income is high. Nonlabor income is high because we include the spouse's net income and implicit income from residence-owned housing. The marginal income effect for nonlabor income is therefore more informative.

Given the challenges associated with estimating any one productivity growth over the entire sample period, we estimate specifications over a fine grid of productivity growth rates ranging from 0.5% and 2% in steps of 0.1%. We can then select a preferred specification as one that minimizes the root mean squared error (RMSE). Alternatively, we could select a specification corresponding to plausible productivity growth estimates available from official sources. Previous estimates of productivity growth during the sample period are around 1%-2%, therefore we focus on estimates corresponding to this range.

Table 1 reports the double/debiased machine learning estimates of the net-of-tax rate elasticity and the marginal effect of virtual income for different productivity growth rates. Estimated net-of-tax-rate elasticities decline monotonically with productivity growth rates and range from a large and statistically significant estimate of 1.1 when productivity growth is 0.5% to a small and imprecise estimate of -0.14 when the productivity growth is set to 2%.

The decline in estimated elasticities with productivity growth reflects that, overall, taxes have decreased over the sample period at the same time as gross labor incomes have increased. Using a specification that assumes a too low (high) productivity growth rate (such as assuming no growth) would therefore result in a positive (negative) bias in the estimated elasticities. Accounting for productivity growth appropriately is therefore important.

Table 1: Debiased Estimation of Taxable Income Effects

Growth Rate	Slope Effect	SE	Income Effect	SE
.5	1.151	.296	-4.98	4.80
.6	1.096	.291	-2.16	2.01
.7	1.034	.285	-2.10	1.93
.8	.978	.278	-4.95	4.77
.9	.928	.287	-5.41	5.21
1.0	.890	.288	-2.32	2.17
1.1	.830	.289	-2.48	2.32
1.2	.789	.300	-2.62	2.46
1.3	.705	.274	-2.75	2.59
1.4	.521	.176	-1.59	1.53
1.5	.363	.135	-1.70	1.64
1.6	.276	.154	-3.16	3.06
1.7	.181	.163	-2.14	2.05
1.8	.0580	.220	-6.72	6.61
1.9	-.000278	.235	-1.75	1.65
2.0	-.138	.311	-1.65	1.54

Figure 2 plots the estimated net-of-tax-rate elasticities along with their 95% confidence intervals and shows that for the more plausible range of 1%-2% productivity growth the net of tax rate elasticities decline from a statistically significant estimate of 0.89 to an imprecise estimate of -0.13. Marginal effects of nonlabor income for different productivity growth rates are plotted in Figure 3, which shows that the income effect estimates are negative and imprecisely estimated. In the more plausible range of productivity growth rates of 1%-2%, income effects are between -1.7 to -2.3, though the confidence interval include zero. In this respect our findings are quite similar to much of the previous literature that also estimates small and/or imprecise income effects.

RMSE for the estimated specifications plotted in Figure 4 shows that after rising initially across low productivity growth rates, it achieves a local minimum at a productivity growth rate of 1.4%. It is worth noting that there is some empirical support from our data for a 1.4% productivity growth. Gross labor income increased by a factor of 1.23 during the sample period, implying an average geometric annual growth rate of 1.4%. Most of the overall increase in taxable income over this period can thus be attributed to exogenous productivity growth and not to changes in the tax rate. Between years the changes in the taxes help us identify the effect of the overall tax rate while the overall increase is attributed to productivity growth.

Our preferred estimate of net-of-tax-rate elasticity corresponding to 1.4% productivity growth is 0.52 and is statistically significant; the income effect for this specification is -1.5—an estimate that is not statistically significant.

An income effect of -1.5 is larger in magnitude than found or assumed in most taxable income studies. Using similar data but a linear model, Blomquist and Selin (2010) estimate a statistically significant negative income effect that is much closer to zero. Some studies just assume there is no income effect. Our negative income effect is imprecise and not statistically different from zero.

While we have so far focused on specifications based on power series terms in $r_k(x)$ up to fourth order, we also estimated specifications with power series terms up to fifth order. The results presented in Table 2 of Appendix D suggest that estimates of net-of-tax-rate elasticities are qualitatively similar to those in Table 1 for specifications with productivity growth rates smaller than 1.1%. Unfortunately, estimates of net-of-tax-rate elasticities for productivity growth rates higher than 1.1% are highly imprecise and, therefore, not comparable to analogous estimates in the more parsimonious fourth order specifications. Just like the fourth order specifications, all income effects for the fifth order power series specifications are highly imprecise. All in all, because the RMSE for fifth order power series specifications remained largely unchanged compared with the fourth order power series, we stick with the estimates from the more parsimonious fourth order power series specifications as our preferred estimates.

5.3 Instrumental Variable Estimates for Linearized Budget Sets

To assess the importance of appropriately accounting for a nonlinear budget set and productivity growth, we have estimated some other specifications based on linearizing the budget set at observed gross income levels. In these specifications, we regress taxable labor income against the marginal net-of-tax rate and virtual income (the nonlabor income for the linearized budget set). To account for the endogeneity of the marginal net-of-tax rate to the budget set, we instrument it using three alternative instruments—net-of-tax rates of the first segment, the last segment, and the mean net-of-tax-rates across all segments. To account for endogeneity of virtual income we instrument it using the same instrument as in the control variable estimates discussed above. We also control for county fixed effects, include year fixed effects, and the same covariates as used in nonparametric regression estimates above. We estimate specifications of gross labor income that are linear in the marginal net-of-tax rate, and in virtual income. Estimates from specifications with logarithm of gross labor income are not reported because they were qualitatively similar.

Such specifications are fairly common for taxable income estimation in repeated cross sections. The typical approach is, however, to use panel data like in Gruber and Saez (2002). This would allow for linear individual effects by differencing and enable us to use instruments based on lagged income. Note though that this standard approach does not allow for other individual heterogeneity, such as in the net of tax elasticity, while our estimates above allow for general individual heterogeneity. The linear specifications we consider here are meant only to provide a comparison with our regression estimates but are not meant to be best possible estimates based on the typical linearization method.

Table 2. Linear Instrumental Variables Estimates

	(1)	(2)	(3)
Slope Elasticity	0.0431	1.338	0.0612
	(0.700)	(3.210)	(0.718)
Income Effect	-1.281	2.931	-1.222
	(0.758)	(8.296)	(0.893)

Standard errors are in parentheses

The estimates based on instrumental variables linearization specifications are reported in Table 2. Standard errors constructed using the delta method are reported in parenthesis. The elasticities are evaluated at the sample mean income of 420,000 SEK. We report the same type of net-of-tax elasticity as before, but now for the linearized budget sets. We observe that the estimated elasticity is sensitive to the instrument used.

While the estimated net-of-tax elasticity is very small for specifications in columns (1) and (3), it is quite substantial for the specification with the last segment tax rate as instrument in column (2), though none of the estimates are significant. Estimates of income effects also vary substantially across the three models and are highly imprecise. Evidently, for our data, using a more sophisticated method therefore matters a lot. The overall picture is that correctly accounting for nonlinear budget sets and general heterogeneity are important in this data.

6 Slutzky Conditions for the Budget Set Regression

The estimator of the budget set regression is based on an approximation to the conditional CDF of the taxable income for a linear budget set in connection with Theorem 1. This approximation can be used to construct an estimator of that CDF. Under the RUM that CDF satisfies a Slutzky condition described in this Section. This Slutzky condition can be checked from data and we do so in this Section for our application.

McFadden (2005) derived restrictions on conditional CDF $F(y|\rho, R)$ for a linear budget set that are necessary and sufficient for a RUM. With choice over two dimensions (c and y) there is a simple, alternative characterization of the RUM. The characterization is that the CDF satisfy a Slutzky condition given in the following result. Let $F_\rho(y|\rho, R) = \partial F(y|\rho, R)/\partial\rho$ and $F_R(y|\rho, R) = \partial F(y|\rho, R)/\partial R$ when these partial derivatives exist.

Theorem 3: *If Assumptions 1, A1, and A2 are satisfied then $F(y|\rho, R)$ is continuously differentiable in ρ and R and*

$$F_\rho(y|\rho, R) - yF_R(y|\rho, R) \leq 0. \tag{6.10}$$

Also, if for all $\rho, R > 0$, $F(y|\rho, R)$ is continuously differentiable in y, ρ, R, the support of $F(y|\rho, R)$ is $[y_\ell, y_u]$ and is bounded, $\partial F(y|\rho, R)/\partial y > 0$ on (y_ℓ, y_u), and equation (6.10) is satisfied then there is a RUM satisfying Assumption 1.

In this sense, for two goods and single valued smooth demands, the revealed stochastic preference conditions are that the CDF satisfies the Slutzky condition. This result will be used in the analysis to follow and is of interest in its own right. Dette, Hoderlein, and Neumeyer (2011) showed that each quantile of $y(\rho, R, \eta)$ satisfies the Slutzky condition for demand functions under conditions similar to those of Assumption A2. Hausman and Newey (2016) observed that when a quantile function satisfies the Slutzky condition there is always a demand model with that quantile function. Theorem 1 is essentially those results combined with the inverse function theorem, that implies that the CDF satisfies the Slutzky condition if and only if the quantile satisfies the Slutzky condition.

The Slutzky condition can be checked by using the linear in parameters approximation to the conditional CDF of Y on which the budget set regression is based. It is straightforward to do this for a grid of values for x and y, say $x_1, ..., x_C$, and $y_1, ..., y_D$. The CDF of taxable income for a linear budget set that corresponds to the budget set regression is $F_1(y) + \sum_{m=2}^{M} w_m(x, \beta)[F_m(y) - F_1(y)]$. The Slutzky condition for the CDF approximation at the values of x and y is then

$$\sum_{m=2}^{M} \left[\frac{\partial w_m(x_c, \beta)}{\partial \rho} - y_d \frac{\partial w_m(x_c, \beta)}{\partial R} \right] [F_m(y_d) - F_1(y_d)] \leq 0, (c = 1, ..., C; d = 1, ..., D). \quad (6.11)$$

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

To check to see if the Slutzky restrictions are satisfied we plugged in estimated post-Lasso parameters and evaluated at individual-specific values of x_c and y_d, given by the observed marginal net-of-tax rate, marginal virtual income, and gross labor income for each individual observation in our sample. Table 3 presents the sample mean and key percentiles of the distribution of values obtained from evaluating equation (6.11) across all individuals in the data for specifications with different productivity growth rates. In Table 3 the Slutzky condition holds for a vast majority of individuals in the sample, with the value being negative not only at the mean and median but also at the 95th percentile.

Table 3: Quantiles of estimated Slutzky conditions

Growth Rate	Mean	Median	75th percentile	95th percentile
0	-1.869	-1.713	-0.852	-0.128
.5	-1.638	-1.574	-0.812	-0.124
1.0	-1.291	-1.291	-0.670	-0.107
1.4	-0.660	-0.660	-0.343	-0.055
2.0	-0.299	-0.299	-0.155	-0.025

7 RUM Choice with General Nonlinear Budget Sets

So far we have focused on piecewise linear budget sets which is the most important kind of budget set in practice. In this Section we give some properties of more general budget sets that also may occur. These properties are also helpful for understanding choice when part of the budget frontier is nonconcave.

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

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

When the budget frontier B is concave the choice $y(B, \eta)$ will be unique by strict quasi-concavity of preferences. In general, when B is not concave the choice $y(B, \eta)$ could be a set. Here we will assume the set valued choices occur with probability zero in the distribution of η. Piecwise linear budget sets are a special case where $B(y)$ is piecewise linear.

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

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

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

Theorem 4: *If Assumptions 1 and A1 are satisfied then for all y such that there is $\Delta > 0$ with $\bar{B}(z) = B(z)$ for $z \in [y, y + \Delta]$ we have $F(y|B) = F(y|\rho(y), R(y))$.*

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

This result is an important dimension reduction in the way the CDF depends on the budget set. In principle $F(y|B)$ can depend on the entire frontier B, an infinite dimensional object. When the frontier is locally concave to the right of y, as in Theorem 4, the CDF depends only on the slope $\rho(y)$ and virtual income $R(y)$ instead of on the entire budget set. Furthermore, the CDF is that for a linear budget set. This result has a number of useful implications. It shows that the dimension reduction for the regression in Theorem 1 also holds for the CDF of taxable income. Also it limits the impact of a nonconcave budget frontier on the conditional distribution of taxable income.

In many applications nonconcavities occur only at small values of income so that $B(y)$ and $\bar{B}(y)$ coincide for y above some realtively small income level. In that case the distribution of taxable income will be determined by the concave budget frontier $\bar{B}(y)$ for most income levels so nonconcavities have a small impact. Also, Theorem 4 could also be used to nonparametrically quantify how the CDF depends on the budget set at higher values of y. For example, one could nonparametrically estimate the revenue effect of changing taxes on higher income earners. Such an object would be of interest because most of the revenue often comes from those paying higher taxes. We leave this use of Theorem 4 to future work.

Theorem 4 implies a revealed stochastic preference result for convex budget sets. As shown by Theorem 3, for linear budget sets and preference satisfying the conditions of Assumptions 1 and A1, a necessary and sufficient condition for a RUM is that the CDF satisfy the Slutzky condition. An implication of Theorem 4 is that this result is also true for convex budget sets. The CDF of taxable income for convex budget sets is consistent with a RUM if and only if the CDF satisfies the Slutzky condition for linear budget sets.

Theorem 4 can also be used to derive identification results for the CDF and conditional expectation of taxable income for a linear budget set. Let $\mathcal{S}$ denote a set of budget frontiers and $X(y) = \{(\rho(y), R(y)) : B \in \mathcal{S}\}$. Then $F(y|\rho, R)$ is identified for $(\rho, R) \in X(y)$. Also, the conditional mean for a linear budget set $\int yF(dy|\rho, R)$ is identified for $(\rho, R) \in \cap_y X(y)$. For taxable income this set may easily be empty due to productivity growth, so the conditional mean for a linear budget set is not identified for any (ρ, R).

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

THEOREM 5: *If Assumption 1 and A1 are satisfied then for all B such that $\bar{B}(z) = B(z)$*

except possibly for $z \in (y(B), \bar{y}(B))$ *and for* $y \in [y(B), \bar{y}(B))$ *we have* $\Pr(y(B,\eta) \leq y)$ *depends only on* $\breve{B}$.

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

8 Conclusion

In this paper we develop a method to nonparametrically estimate the expected value of taxable income as a function of a nonlinear budget set while allowing for general heterogeneity and optimization/measurement errors. We apply this approach to Swedish data and find a significant elasticity .52 of an overall change in the tax rate.

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

9 Appendix A: Proofs of Theorems 1-5

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

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

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

Before proving Theorems 1-5 we give a several useful Lemmas. First we state a result on the derivatives of $F(y|\rho, R)$ with respect to ρ and R.

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

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

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

Next we give some other useful Lemmas.

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

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

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

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

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

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

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

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

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

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

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

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

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

so that

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

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

Summarizing, we have shown that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof of Theorem 1: Let $F_j(y) = F(y|\rho_j, R_j)$. By Lemmas A3 and A4 it follows (similarly to the proof of Theorem 4 to follow) that the CDF of $y(B, \eta)$ on (ℓ_{j-1}, ℓ_j) is $F_j(y)$. Therefore,

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

Note that

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

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

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

Also, it follows that

$$\begin{aligned}\Pr(y(B,\eta) &= \ell_j) = F_{j+1}(\ell_j) - \lim_{y\uparrow\ell_j} F_j(y) = \int 1(y \le \ell_j)F_{j+1}(dy) - \int 1(y < \ell_j)F_j(dy) \quad (9.12)\\ &= \int 1(y \ge \ell_j)F_j(dy) - \int 1(y > \ell_j)F_{j+1}(dy).\end{aligned}$$

Combining these results we have

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

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

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

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

Combining this with the second equality in eq. (9.12) then gives

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

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

Proof of Theorem 2: By Edmunds and Evans (1989) there exists C such that for each A and B there is (β_{ab}) such that for $\varepsilon = C(MK)^{(1-s)/3}$ and $p^{MK}(y,x) = \sum_{a=1}^{M}\sum_{b=1}^{K}\beta_{ab}f_a(y)r_b(x)$,

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

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

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

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

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

Also,

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

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

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

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

$$\begin{aligned} \left|\bar{y}(x) - \bar{y}^{MK}(x)\right| &= \left|\int y[f(y|x) - p^{MK}(y,x)]dy + [1 - \sum_{a=1}^{M} w_a(x,\beta)] \int y f_1(y)dy\right| \\ &\le \int y \left|f(y|x) - p^{MK}(y,x)\right| dy + \left|1 - \sum_{a=1}^{M} w_a(x,\beta)\right| \int y f_1(y)dy \le C\varepsilon. \end{aligned}$$

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

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

Therefore we have

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

To conclude the proof, note that

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

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

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

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

Proof of Theorem 3: Note that $y(\rho,R,\eta)$ is differentiable by Assumption 1. Also, ρ behaves like the negative of a price (increasing ρ increases utility), so that the Slutzky condition for taxable income is

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

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

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

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

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

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

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

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

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

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

Then

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

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

Proof of Theorem 4: Combining the conclusions of Lemmas A3 and A4 gives

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

Proof of Theorem 5: For notational simplicity we suppress the η as before. Let $\hat{y} = \underline{y}(B), \check{y} = \bar{y}(B)$ denote the lower and upper bound respectively for the set where $B(z)$ is not concave. Consider

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

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

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

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

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

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

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

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

10 Appendix B: Kinks and Nonparametric Compensated Tax Effects

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

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

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

where the expectation is taken over the distribution of η.

Theorem B1: *If Assumptions 1, A1, and A2 are satisfied then*

$$\Pi_{\bar{\ell}} = \int_{\rho_+}^{\rho_-} \phi(\rho)\delta(\rho)d\rho \text{ and } \phi(\rho)\delta(\rho) = -F_\rho(\bar{\ell}|\rho, R(\rho)) + \bar{\ell} \cdot F_R(\bar{\ell}|\rho, R(\rho)). \tag{10.13}$$

The $\delta(\rho)$ in Theorem B1 is an average compensated effect of changing ρ for a linear budget set. This compensated effect appears here because virtual income is being adjusted as ρ changes to stay at the kink. The virtual income adjustment needed to remain at the kink corresponds locally to the income adjustment needed to remain on the same indifference curve, as shown by Saez (2010). The formula for $\Pi_{\bar{\ell}}$ bears some resemblance to the kink probability formulas in Saez (2010) but differs in important ways. Theorem B1 is global, nonparametric, and takes explicit account of general heterogeneity, unlike the Saez (2010) results, which are local or parametric and account for heterogeneity implicitly.

Theorem B1 helps clarify what can be nonparametrically learned from kinks. First, the compensated effects that enter the kink probability are only for individuals who would choose to locate at the kink for a linear budget set with $\rho \in [\rho_+, \rho_-]$. Thus, using kinks to provide information about compensated effects is subject to the same issues of external validity as, say, regression discontinuity design (RDD). As RDD only identifies treatment effects for individuals at the jump point so kinks only provide information about compensated effects for individuals who would locate at the kink.

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

Third, the pure heterogeneity effect, and hence compensated effects, is not identified when ρ_- and ρ_+ do not vary in the data. One cannot identify the pdf $\phi(\rho)$ for $\rho \in (\rho_+, \rho_-)$ because observations will not be available for such ρ values. Because of this it may be impossible to say anything about compensated effects from kinks. An example can be used to illustrate. Suppose that the parameter θ of interest is a weighted average (over ρ) of compensated effects $\theta = \int_{\rho_+}^{\rho_-} \phi(\rho)\delta(\rho)d\rho / \int_{\rho_+}^{\rho_-} \phi(\rho)d\rho$. Theorem B1 gives

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

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

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

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

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

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

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

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

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

Theorem B1 can also be used to relate positivity of the kink probability $\Pi_{\bar{\ell}}$ to the Slutzky condition. One could specify a CDF $F(y|\rho, R)$ for taxable income for a linear budget set and derive the probability of a kink from equation (10.13). Then the Slutzky condition is sufficient but not necessary for positivity of $\Pi_{\bar{\ell}}$, because an integral can be positive without the function being integrated being positive. In this sense the kink probability can be positive without all the conditions for utility maximization being satisfied.

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

$$\frac{\partial F(y|\rho(y), R(y))}{\partial y} = F_y(y|\rho(y), R(y)) + \rho_y(y)[F_\rho(y|\rho(y), R(y)) - yF_R(y|\rho(y), R(y))], \quad (10.14)$$

where $\rho_y(y) = \partial\rho(y)/\partial y$. One could specify a CDF $F(y|\rho, R)$ for taxable income for a linear budget set and derive the pdf from equation (10.14). The first term F_y is nonnegative because it is a pdf. The $\rho_y(y)$ is nonpositive because it is the derivative of the slope of a concave function. Then the Slutzky condition is sufficient for a positive pdf because it means that the second term will be nonnegative and hence so will the sum. However, the Slutzky condition is not necessary for positivity of the pdf of taxable income because the positivity of the pdf F_y can result in positivity of the sum of the two terms even when the second term is negative. In this sense the pdf of taxable income for a smooth concave budget frontier can be positive without all the conditions for utility maximization being satisfied.

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

Proof of Theorem B1: Let $F(y)$ be the CDF of $y(B, \eta)$. By standard probability theory, $\Pi_{\bar{\ell}} = F(\bar{\ell}) - \lim_{y\uparrow\bar{\ell}} F(y)$. By Theorem 4 $F(\bar{\ell}) = F(\bar{\ell}|\rho_+, R_+)$ and $\lim_{y\uparrow\bar{\ell}} F(y) = \lim_{y\uparrow\bar{\ell}} F(y|\rho_-, R_-)$.

Furthermore, by Assumption A2 $y(\rho_-, R_-, \eta)$ is continuously distributed so that $\lim_{y \uparrow \bar{\ell}} F(y|\rho_-, R_-) = F(\bar{\ell}|\rho_-, R_-)$. Define $\Lambda(\rho) = F(\bar{\ell}|\rho, R(\rho))$ for $\rho \in [\rho_+, \rho_-]$. We then have

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

By the chain rule, $R(\rho) = R_- + \bar{\ell}(\rho_- - \rho)$, and Lemma A1 of Hausman and Newey (2016), $\Lambda(\rho)$ is differentiable and

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

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

11 Appendix C: The Swedish Tax System and Summary Statistics

In Sweden, it is the individual, and not the household, that is the basic tax unit, and the income tax system is a so-called dual tax system, meaning that earned income and capital income are taxed separately. Capital income is taxed at a flat rate of 30% and earned income is taxed by a progressive tax schedule.

The shape of individuals' budget constraints is determined by the income tax, the transfer system, and the employee paid social security contributions. The earned income tax consists of two parts, a proportional local tax rate that applies to all income and a federal income tax that kicks in for income above a certain level.The local tax rate varies across communes, but with a thick clustering around 31-32%. The average local tax has been almost constant during the time-period our sample covers, increasing from 31.04% in 1993 to 31.44% in 2008. There is some variation in the local tax rate across communes and there is a concern that this variation causes endogeneity; to account for this we use county dummy variables as covariates.

In 1993 the federal income tax was 20% and kicked in at a level corresponding to roughly the 75th percentile of the income distribution. In 1995 this federal tax was raised to 25%. In 1999 it was reduced back to 20% for incomes lower than incomes at around the 95th percentile, but the 25% tax rate was kept for income above this level; thus, from 1999 there were two levels for the federal income tax. After 1999 the kink points of the tax schedule have been changed several times, implying further time variation in tax rates.

In 2007 an EITC was introduced in Sweden and further expanded in 2008 and later years. The Swedish EITC differs in design in several respects from that in other countries; it is universal in the sense that all citizens below 65 face the same tax credit schedule, also, in 2007 and 2008 there was no phase out of the credit. The EITC was gradually phased in up to an income corresponding to the median wage of a white-collar worker in the private sector. Although those with higher incomes did not get a lower marginal tax, they did get the full credit lump sum, hence, for high income people there was no substitution effect, just an income effect.

There is also a fairly complex set of rules governing what deductions an individual can make. There are some standard deductions that are applied automatically by the tax authorities, but also itemized deductions that you can claim. These rules have changed over time, which also contributes to the time variation in net of tax rates.

Two important components of the transfer system are welfare payments and housing allowances. These transfers are phased out as income increases thereby decreasing the slope of the budget constraint at low levels of income. For families with children the child allowances are important; these allowances are not income dependent and do not affect slopes.

Employee social security contributions to finance health care, pensions and unemployment benefits was introduced in 1993 at a rate of 0.95% of income and over the years gradually increased to 7% in 2000. These "contributions" are paid on income up to a ceiling, and zero above the ceiling.

References

Aaberge, R., J. Dagsvik and S. Ström (1995): "Labor Supply Responses and Welfare Effects of Tax Reforms," *Scandinavian Journal of Economics* 97, 635-659, 1995. https://doi.org/10.2307/3440547

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.

Auten, G. and R. Carroll (1999): "The Effect of Taxes on Household Income," *Review of Economics and Statistics 81* : 681—693. https://doi.org/10.1162/003465399558409

Belloni, A. and V. Chernozhukov (2013): "Least Squares After Model Selection in High-Dimensional Sparse Models," *Bernoulli*, 521-547. https://doi.org/10.3150/11-BEJ410

Blomquist, N. S. (1983): "The effect of income taxation on the labor supply of married men in Sweden," Journal of Public Economics 22, 169-197. https://doi.org/10.1016/0047-2727(83)90064-6

Blomquist, N. S., and U. Hansson-Brusewitz, U. (1990): "The effect of taxes on male and female labor supply in Sweden," *Journal of Human Resources*, 317-357. https://doi.org/10.2307/145987

Blomquist, S. and W. Newey (2002): "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica 70* : 2455—2480. https://doi.org/10.1111/j.1468-0262.2002.00445.x

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. https://doi.org/10.1016/j.jpubeco.2010.08.001

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

Blundell, R.W. and R. Matzkin (2014): "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics* 5, 271—295. https://doi.org/10.3982/QE281

Blundell, R., D. Kristensen, D., and R. Matzkin (2014): "Bounding quantile demand functions using revealed preference inequalities," *Journal of Econometrics* 179, 112-127. https://doi.org/10.1016/j.jeconom.2014.01.005

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

Blundell, R.W. and A. Shephard (2012): "Employment, hours of work and the optimal taxation of low-income families," *Review of Economic Studies* 79, 481—510. https://doi.org/10.1093/restud/rdr034

Bourguignon, F., and T. Magnac (1990): "Labor supply and taxation in France," *Journal of Human Resources*, 358-389. https://doi.org/10.2307/145988

Burns, S. K., and J.P. Ziliak (2017): "Identifying the elasticity of taxable income," *The Economic Journal* 127, 297-329. https://doi.org/10.1111/ecoj.12299

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. https://doi.org/10.1086/260730

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.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W.K. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal* 21. https://doi.org/10.1111/ectj.12097

Chernozhukov, V., W.K. Newey, and R. Singh (2022): "Automatic debiased machine learning of causal and structural effects," *Econometrica* 90, 967-1027. https://doi.org/10.3982/ECTA18515

Chetty, R. (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica* 80, 969-1018. https://doi.org/10.3982/ECTA9043

Colombino, U., and D. Del Boca (1990): "The effect of taxes on labor supply in Italy," *Journal of Human Resources*, 390-414. https://doi.org/10.2307/145989

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

Edmunds, D. E., and W. Evans (1989): "Entropy Numbers and Approximation Numbers in Function Spaces," *Proceedings of the London Mathematical Society* 58, 137—152. https://doi.org/10.1112/plms/s3-58.1.137

Eissa, N. (1995): "Taxation and labor supply of married women: the Tax Reform Act of 1986 as a natural experiment,"

Eissa, N. and S. Giertz (2006): "Trends in High Incomes and Behavioural Responses to Taxation: Evidence from Executive Compensation and Statistics of Income Data," Congressional

Budget Office Working Paper No. 2006-14.

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

Feldstein, M. (1995): "The Effects of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy 103(3)*: 551 -572. https://doi.org/10.1086/261994

Feldstein, M. (1999): "Tax Avoidance and the Deadweight Loss of the Income Tax," *Review of Economics and Statistics 81(4)*: 674—680. https://doi.org/10.1162/003465399558391

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. https://doi.org/10.3386/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. https://doi.org/10.1111/j.1467-9914.2007.00386.x

Giertz, S. (2007): "The Elasticity of Taxable Income Over the 1980s and 1990s," *National Tax Journal 60(4)*: 743-768. https://doi.org/10.17310/ntj.2007.4.04

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. https://doi.org/10.2307/2534678

Goolsbee, A. (2000): "What happens when you tax the rich? Evidence from executive compensation" Journal of Political Economy 108, 352-378. https://doi.org/10.1086/262122

Gruber, J. and E. Saez (2002): "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics 84*:1—32. https://doi.org/10.1016/S0047-2727(01)00085-8

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

Hausman, J.A (1979): "The Econometrics of Labor Supply on Convex Budget Sets," *Economics Letters* 3, 171-174. https://doi.org/10.1016/0165-1765(79)90112-5

Hausman J.A. (1980): "The effect of wages, taxes, and fixed costs on women's labor force participation," *Journal of Public Economics* 14, 161-194.

https://doi.org/10.1016/0047-2727(80)90039-0

Hausman, J.A. (1981): "The Effect of Taxes on Labor Supply," in *How Taxes Affect Eco-*

nomic Behavior, ed. H. Aaron and J. Pechman, Brookings Papers on Economic Activity.

Hausman, J. A. (1985): "Taxes and labor supply," in *Handbook of Public Economics* 1, 213-263. Elsevier. https://doi.org/10.1016/S1573-4420(85)80007-0

Hausman, J.A. and W.K. Newey (2016): "Individual Heterogeneity and Average Welfare" *Econometrica* 84, 1225-1248. https://doi.org/10.3982/ECTA11899

Heim, B. T. (2009): "Structural estimation of family labor supply with taxes estimating a continuous hours model using a direct utility specification," *Journal of Human Resources* 44, 350-385. https://doi.org/10.3368/jhr.44.2.350

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

Keane, M.P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature* 49, 961—1075. https://doi.org/10.1257/jel.49.4.961

Keane, M. and R. Moffitt (1998): "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review* 39, 553—589. https://doi.org/10.2307/2527390

Kitamura, Y., and J. Stoye (2018): "Nonparametric analysis of random utility models," *Econometrica* 86, 1883-1909. https://doi.org/10.3982/ECTA14478

Kline, P., and M. Tartari, (2016): "Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach," *American Economic Review* 106, 972-1014. https://doi.org/10.1257/aer.20130824

Kumar, A. and C.-Y. Liang (2020): "Estimating Taxable Income Responses with Elasticity Heterogeneity," *Journal of Public Economics* 188. https://doi.org/10.1016/j.jpubeco.2020.104209

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. https://doi.org/10.3982/ECTA9113

Kopczuk, W. (2005): "Tax Bases, Tax Rates and the Elasticity of Taxable Income," *Journal of Public Economics* 89 (11-12): 2093—2119. https://doi.org/10.1016/j.jpubeco.2004.12.005

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. https://doi.org/10.1016/0047-2727(87)90073-9

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. https://doi.org/10.3386/w12417

Manski, C. (2014): "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy," *Quantitative Economics* 5, 145 - 174. https://doi.org/10.3982/QE262

McFadden, D. (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory* 26, 245-264. https://doi.org/10.1007/s00199-004-0495-3

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

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

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

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

Saez, E. (2003): "The Effect of Marginal Tax Rates on Income: A Panel Study of 'Bracket Creep'," *Journal of Public Economics* 85: 1231—1258. https://doi.org/10.1016/S0047-2727(01)00178-5

Saez, E. (2010): "Do Taxpayers Bunch at Kink Points," *American Economic Journal: Economic Policy* 2, 180-212. https://doi.org/10.1257/pol.2.3.180

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. https://doi.org/10.1257/jel.50.1.3

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. https://doi.org/10.1086/NTJ41789288

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.

Triest, R. K. (1990): "The effect of income taxation on labor supply in the United States," *Journal of Human Resources,* 491-516. https://doi.org/10.2307/145991

Van Soest, A. (1995), "Structural Models of Family Labor Supply: a Discrete Choice Approach," *Journal of Human Resources* 30, 63-88. https://doi.org/10.2307/146191

Van Soest, A., I. Woittiez, and A. Kapteyn (1990): "Labor supply, income taxes, and hours restrictions in the Netherlands," *Journal of Human Resources*, 517-558. https://doi.org/10.2307/145992

Weber, C. (2014): "Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences," *Journal of Public Economics* 117, 90-103. https://doi.org/10.1016/j.jpubeco.2014.05.004

Figure 1a: Variation in First Segment and Marginal Net of Tax Rates

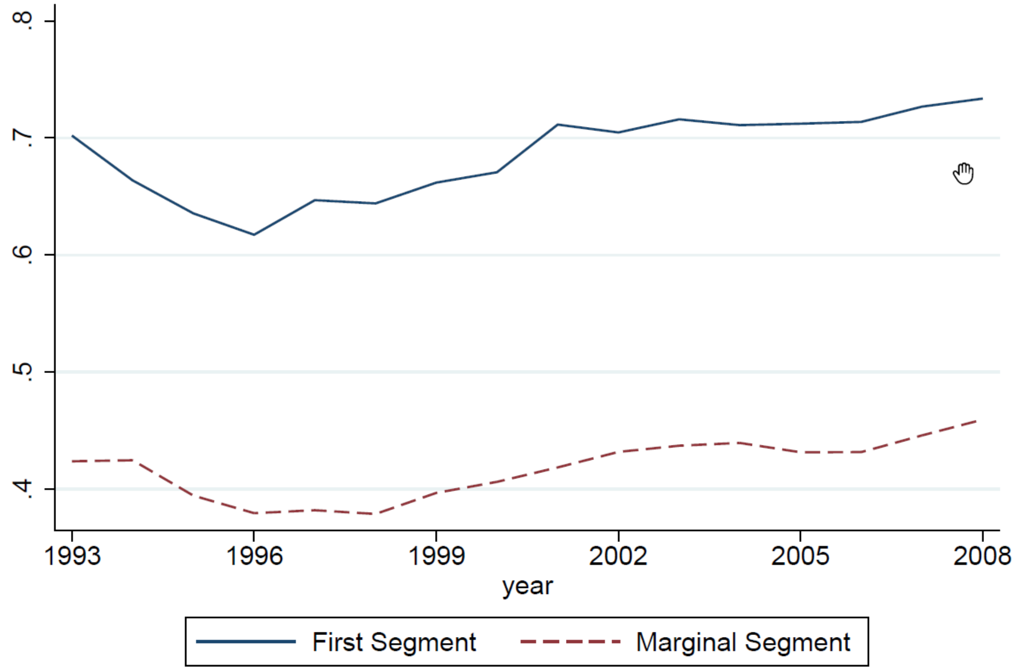


Figure 1b: Variation in Last Segment Net of Tax Rates

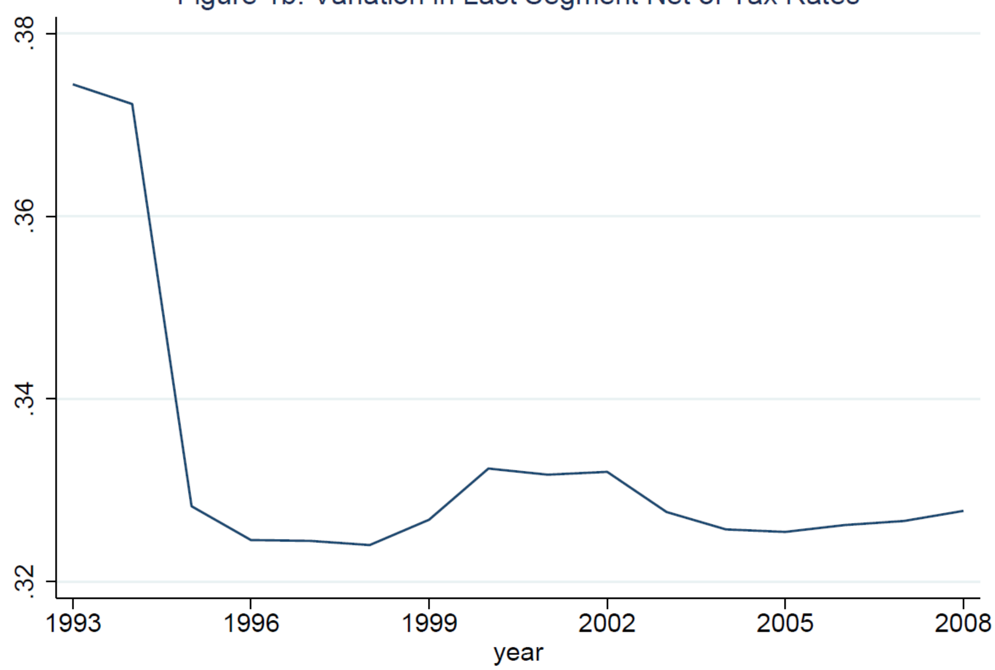
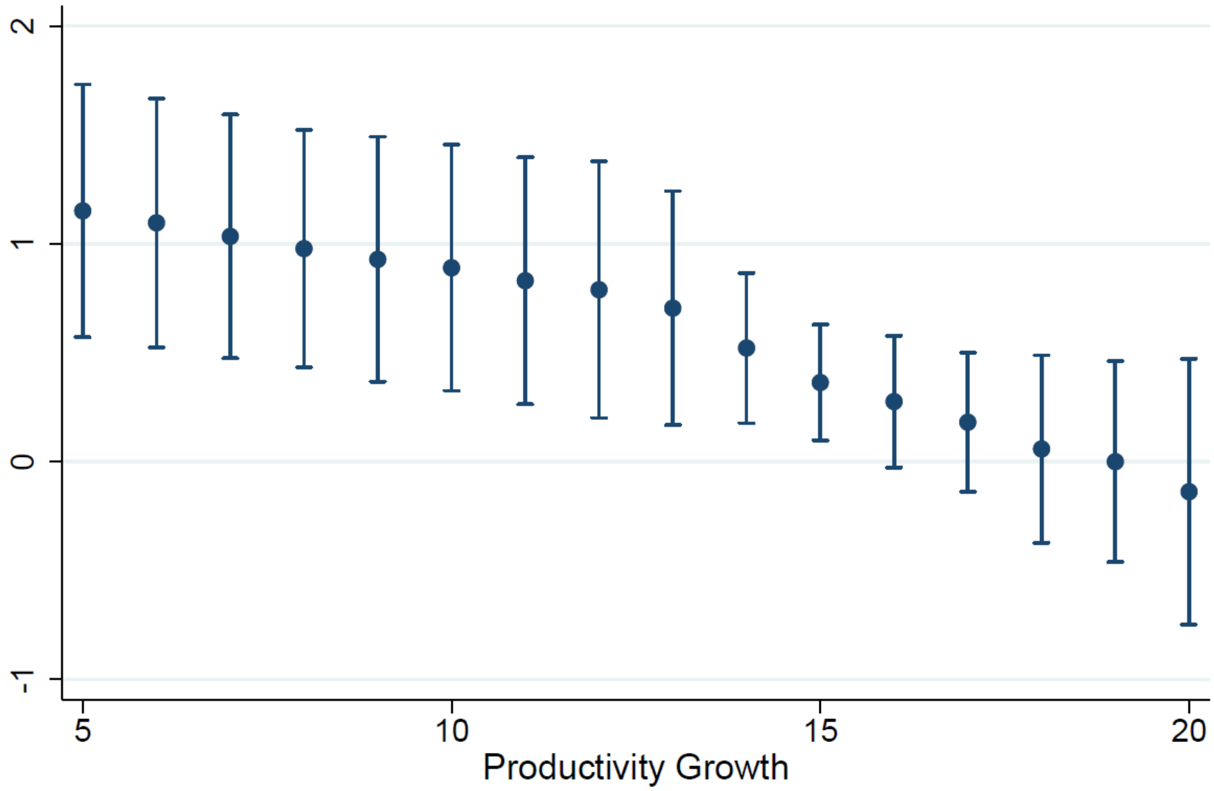
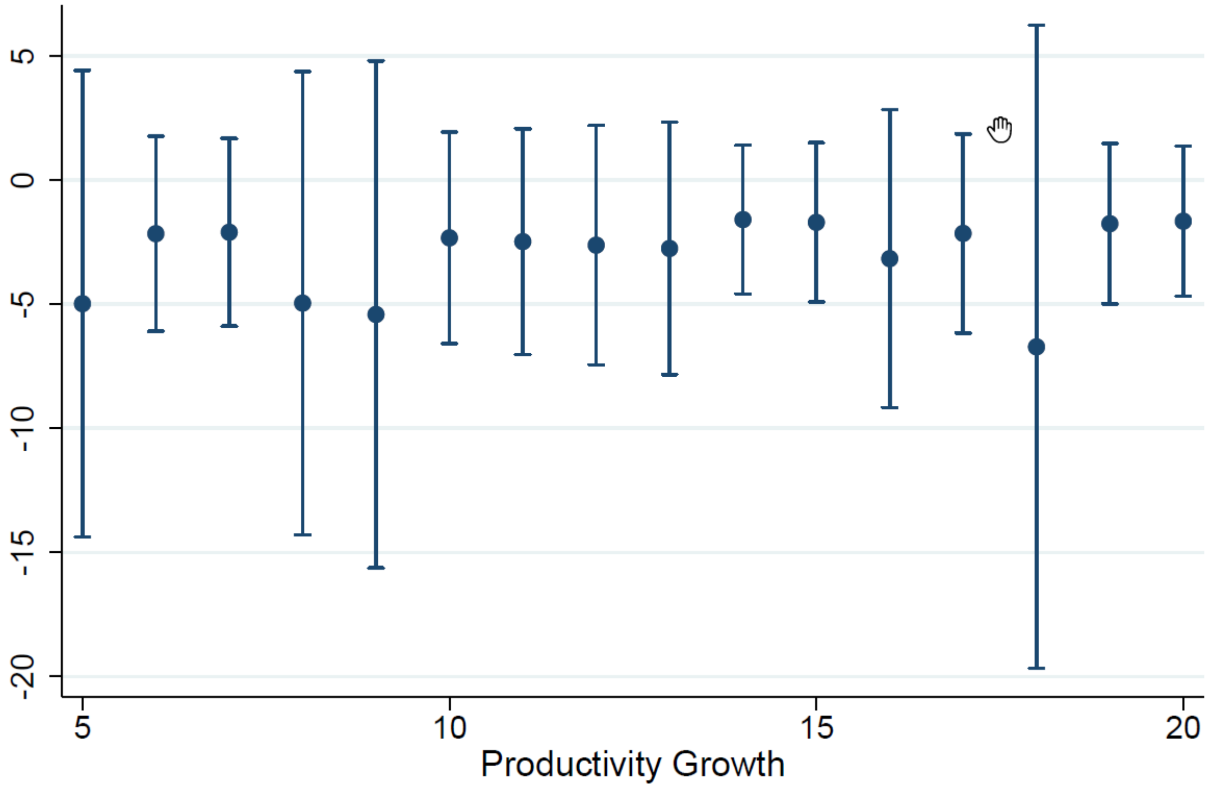


Figure 2: Estimates of Slope Elasticity



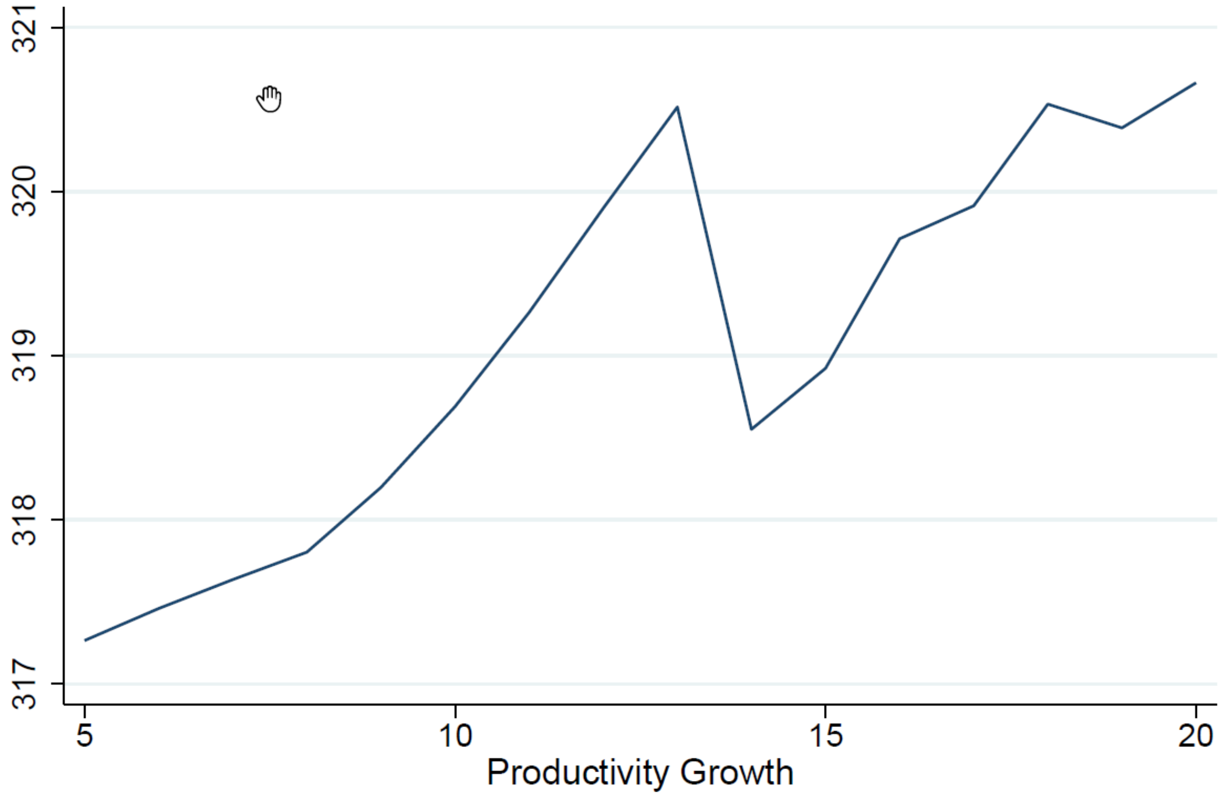
Note: Budget Set Polynomial Degree 4

Figure 3: Estimates of Income Effect



Note: Budget Set Polynomial Degree 4

Figure 4: Root Mean Squared Error



Note: Budget Set Polynomial Degree 4

Appendix C Table 1: Summary Statistics

	All Years	1993	1998	2003	2008
Gross labor income	420.329	377.856	406.396	428.010	464.490
1-st net-of-tax rate	0.687	0.702	0.644	0.716	0.734
1-st virtual income	254.250	219.443	222.637	261.824	247.039
Marginal net-of tax-rate	0.418	0.424	0.379	0.437	0.459
Marginal Virtual Income	295.007	259.160	262.941	301.510	301.535
Last net-of-tax rate	0.332	0.374	0.324	0.328	0.328
Last virtual income	327.578	269.726	278.517	347.411	360.649
Age	43.873	42.431	43.774	44.325	44.275
Dummy children<6 Yrs	0.357	0.440	0.367	0.299	0.333
Dummy foreign born	0.141	0.108	0.131	0.148	0.170
Wife's net labor income	141.718	124.647	124.004	149.965	184.022
Transfers at zero labor inc	21.754	23.222	19.767	18.785	20.454
Observations	81717	5192	4473	5592	5086

Notes: The table presents means of key variables. Gross labor income, virtual income, wife's net labor income, and transfers at labor income are expressed in 1,000 SEK.

Appendix D Table 2: Debiased Estimates of Taxable Income Elasticity
(Budget Set Polynomial Degree 5)

	(1)		(2)	
Productivity Growth	Slope Elasticity	Std. Err.	Income Effect	Std. Err.
0.5	0.866	0.117	7.950	7.738
0.6	1.952	1.113	-1.052	1.004
0.7	0.822	0.131	-0.565	0.506
0.8	0.709	0.122	8.737	8.523
0.9	0.657	0.124	9.046	8.833
1.0	0.611	0.134	1.664	1.662
1.1	0.548	0.128	1.992	1.994
1.2	1.358	0.828	2.335	2.342
1.3	-1.971	2.417	2.708	2.717
1.4	0.0220	0.411	1.885	1.918
1.5	-0.0766	0.456	-2.623	2.547
1.6	0.333	0.140	-1.649	1.561
1.7	-0.243	0.547	-1.751	1.663
1.8	-0.349	0.585	-1.730	1.636
1.9	-0.431	0.623	-1.578	1.483
2.0	2.277	2.140	-1.518	1.419

Lasso used for debiasing using procedure in Chernozhukov et. al (2022).