working paper
department
of economics

NONPARAMETRIC ESTIMATION WITH NONLINEAR BUDGET SETS

Soren Blomquist
Whitney K. Newey

No. 99-03 July, 1999

massachusetts
institute of
technology
50 memorial drive
cambridge, mass. 02139

WORKING PAPER
DEPARTMENT
OF ECONOMICS

NONPARAMETRIC ESTIMATION WITH NONLINEAR BUDGET SETS

Soren Blomquist
Whitney K. Newey

No. 99-03 July, 1999

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

50 MEMORIAL DRIVE
CAMBRIDGE, MASS. 02142

NONPARAMETRIC ESTIMATION WITH NONLINEAR BUDGET SETS *

Sören Blomquist Uppsala University Sweden
Whitney Newey MIT, E52-262D Cambridge, MA 02139

September, 1998
Revised, February 1999

Abstract

Choice models with nonlinear budget sets are important in econometrics. In this paper we propose a nonparametric approach to estimation of choice models with nonlinear budget sets. The basic idea is to think of the choice, in our case hours of labor supply, as being a function of the entire budget set. Then we can account nonparametrically for a nonlinear budget set by estimating a nonparametric regression where the variable in the regression is the budget set. We reduce the dimensionality of this problem by exploiting additive structure implied by utility maximization with convex budget sets. This structure leads to a polynomial convergence rate for the estimator. We give asymptotic normality results also. The usefulness of the estimator is demonstrated in Monte Carlo and empirical work, where we find it can have a large impact on estimated effects of tax changes.

JEL Classification: C14, C24

Keywords: Nonlinear budget sets, nonparametric estimation, additive models.

*Financial support from the Bank of Sweden Tercentenary Foundation is gratefully acknowledged. We are grateful to Matias Eklof for competent research assistance. We thank participants at the Harvard-MIT econometrics workshop and the NBER for helpful comments.

1. Introduction

Choice models with nonlinear budget sets are important in econometrics. They provide a precise way of accounting for the ubiquitous nonlinear tax structures when estimating demand. This is important for testing economic theory and formulating policy conclusions when budget sets are nonlinear. Estimation of such models presents formidable challenges, because of the inherent nonlinearity. The most common approach has been maximum likelihood under specific distributional assumptions, as exposited by Hausman (1985). This approach provides precise estimates when the assumptions of it are correct, but is subject to specification error when the distribution or other aspects of the model are wrong. Also, the likelihood is quite complicated, so that the MLE presents computational challenges as well.

In this paper we propose a nonparametric approach to estimation of choice models with nonlinear budget sets. This approach should be less sensitive to specification of disturbance distributions. Also, it is computationally straightforward, being based on nonparametric modeling of the conditional expectation of the choice variable. The basic idea is to think of the choice, in our case hours of labor supply, as being a function of the entire budget set. Then we can account nonparametrically for a nonlinear budget set by estimating a nonparametric regression where the variable in the regression is the budget set. Assuming that the budget set is piecewise linear, the budget sets will be characterized by two or more numbers. For instance, a linear budget constraint is characterized by the intercept and slope. More generally, a piecewise linear budget constraint will be characterized by the intercept and slope of each segment. Nonparametric regression on these slopes and intercepts should yield an estimate of how choice depends on the budget set.

A well-known problem of nonparametric estimation is the "curse of dimensionality," referring to the difficulty of nonparametric estimation of high dimensional functions. Budget sets with many segments have a high dimensional characterization, so for nonparametric estimation to be successful it will be important to find a more parsimonious approach. One feature that is helpful is that under utility maximization with convex preferences, the conditional expectation of the choice variable will be additive, with each additive component depending only on a few variables. This feature helps reduce the curse of dimensionality, leading to estimators that have faster convergence rates. We also consider approximating budget constraints with many segments by budget constraints with only a few segments (like three or four). Often in applications there will be only a few sources of variation in the data, which could be captured by budget constraints with few segments.

An advantage of nonparametric estimation is that it should allow utility consistent functions that are more flexible than some parametric specifications, where utility maximization can impose severe restrictions. For instance, it is well known that utility maximization with convex preferences implies that the linear labor supply function $h = a + bw + cy + e$ must satisfy the restrictions $b > 0$ and $c < b/H$, where w is the wage, y nonlabor income and H is the maximum number of hours. Relaxing the parametric form for the labor supply function should substantially increase its flexibility while allowing for utility consistent functional forms. In the paper we do not impose utility maximization, but we can test for utility consistency using our approach.

The rest of the paper is organized as follows. In section two we present a particular data generating process and derive an expression for expected hours of work. The estimation procedure we propose is described in section 3. Asymptotic properties of the estimator are discussed in section 4 and small sample properties, based on Monte Carlo simulations, in section 5. In section 6 we apply the method to Swedish data. We use estimated labor supply functions to calculate the effect of income tax reform in section 7. Section 8 concludes.

2. Data generating process and expected hours of work

Our estimation method is to nonparametrically estimate the conditional mean of hours given the budget set. That is, if h_i is the hours of the ith individual and B_i represents their budget set, our goal is to estimate

$$E[h_i \mid B_i] = \bar{h}(B_i).$$

This should allow us to predict the average effect on hours of changes in the budget set that are brought about by some policy, such as a change in the tax structure. Also depending on the form of the unobserved heterogeneity in h_i , one can use $\bar{h}(B_i)$ to test utility maximization and make utility consistent predictions, such as for consumer surplus.

In comparison with the maximum likelihood approach, ours imposes fewer restrictions but only uses first (conditional) moment information. This comparison leads to the usual trade-off between robustness and efficiency. In particular, most models in the literature have a labor supply function of the form

$$h_i = h(B_i, v_i) + \varepsilon_i,$$

where v_i represents individual heterogeneity, and ε_i is measurement error. The typical maximum likelihood specification relies on an assumption that v_i and ε_i are normal and homoskedastic, while all that we would require is that v_i is independent of B_i and $E[\varepsilon_i \mid B_i] = 0$, in which case $\bar{h}(B_i) = \int h(B_i, v)G(dv)$. This should allow us to recover some features of $h(B, v)$ under much weaker conditions than normality of the disturbance. Of course, these more general assumptions come at the expense of efficiency of the estimates. In particular maximum likelihood would also use other moment information, so that we would expect to have to use more data to get the same precision as maximum likelihood estimation would give.

Our approach to estimation will be valid for quite general data generating processes. In particular, it is neither necessary that data are generated by utility maximization nor that the data generating budget constraints are convex. However, without imposing a simplifying structure on the expected hours of work function it will in general be infeasible to estimate the function due to a severe dimensionality problem. We will therefore derive expressions for expected hours of work given the assumption that data are generated by utility maximization subject to piece wise linear convex budget constraints. This will help in constructing parsimonious specifications for $\bar{h}(B)$ and in understanding utility implications of the model. These restrictions can then be tested, as we do in the empirical work.

Assume data are generated by utility maximization with globally convex preferences subject to a piecewise linear budget constraint. To simplify the exposition, let us consider a budget constraint with three segments defining a convex budget set. We show such a budget constraint in Figure 1. The budget constraint is defined by the slopes and intercepts of the three segments. These segments also define two kink points. The kink points are related to the slopes and intercepts as: $\ell_1 = (y_2 - y_1)/(w_2 - w_1)$ and $\ell_2 = (y_3 - y_2)/(w_3 - w_2)$.

We will derive an expression for expected hours of work given this data generating process. Let desired hours of work for a linear budget constraint be given by $h_j^* = \pi(y_j, w_j) + v$, where v is a random preference variable. Let $g(t)$ be the density of v, $G(v)$ the c.d.f of v, $H(v) = \int_{-\infty}^{v} tg(t)dt$ and $J(v) = H(v) - vG(v)$. We assume that $H(\infty) = 0$, i.e., $E(v) = 0$. We further assume labor supply is generated by utility maximization with globally convex preferences. Then desired hours will equal zero if $\pi_1 + v \leq 0$. Desired hours will fall on the first segment if $0 \leq \pi_1 + v \leq \ell_1$ and be located at kinkpoint ℓ_1 if $\pi(y_1, w_1) + v \geq \ell_1$, and $\pi(y_2, w_2) + v \leq \ell_1$ i.e. if $\ell_1 - \pi(y_1, w_1) \leq v \leq \ell_1 - \pi(y_2, w_2)$. Desired hours will be on the second segment if $\ell_1 < \pi(y_2, w_2) + v < \ell_2$, etc. This implies that we can write expected hours of work as:

$$\begin{aligned}
E(h^*) &= 0 \cdot G(-\pi_1) \\
&+ \underbrace{[G(\ell_1 - \pi_1) - G(-\pi_1)]}_{\text{probability that } h^* \text{ is on first segment}} \times \{\pi_1 + E(v) \mid -\pi_1 \leq v \leq \ell_1 - \pi_1\} \\
&+ \ell_1 \cdot \underbrace{[G(\ell_1 - \pi_2) - G(\ell_1 - \pi_1)]}_{\text{probability that desired hours are at kinkpoint } \ell_1} \\
&+ \underbrace{[G(\ell_2 - \pi_2) - G(\ell_1 - \pi_2)]}_{\text{probability that } h^* \text{ is on the second segment}} \times \{\pi_1 + E(v) \mid \ell_1 - \pi_2 \leq v \leq \ell_2 - \pi_2\} \\
&+ \ell_2 [G(\ell_2 - \pi_3) - G(\ell_2 - \pi_2)] \\
&+ \underbrace{[1 - G(\ell_2 - \pi_3)]}_{\text{probability that desired hours are on third segment}} \times \{\pi_3 + E(v) \mid v > \ell_2 - \pi_3\}
\end{aligned} \tag{1'}$$

We see from this expression that $E(h^*)$ is a continuous, differentiable function in ℓ_1, π_1, ℓ_2 π_2, ℓ_3, π_3.[1] Since π_i is differentiable in y_i, w_i it follows that $E(h^*)$ is continuous and differentiable in ℓ_1, w_1, y_1, ℓ_2, w_2, ℓ_3, w_3, y_3.

Using the $J(v)$ notation and setting $\ell_0 = 0$ we can rewrite (1') as:

$$E(h^*) = -J(-\pi_1) + \sum_{k=1}^{2} [J(\ell_k - \pi_k) - J(\ell_k - \pi_{k+1})] + \pi_3 \tag{2.1}$$

This expression generalizes straightforwardly for the case with more segments. The particular form of expression (1) follows from the assumption that hours of work are generated by utility maximization with globally convex preferences. For particular c.d.fs of v we can derive properties of the $J(v)$ function. For example, if v is uniformly distributed $J(v)$ will be quadratic. Independent of the form of the c.d.f. for v, $J(v)$ will always be decreasing and concave and lie below its

[1] Expression (1') is derived under the assumption that there is no upper limit $\bar{H}$ for hours of work. If we introduce an upper limit $\bar{H}$ for hours of work, we would get one more term, and the last term would be slightly different. If $\bar{H}$ is set at a high value, say, 6000 hours a year, it would not matter for empirical applications whether we use expression (1) or an expression with an upper limit $\bar{H}$ included.

asymptotes which is 0 if v goes to minus infinity and a line through the origin with slope -1 for v going to plus infinity.

There are two important aspects of expression (1) that we want to emphasize. One is that the strong functional form restrictions implied by utility maximization and a convex budget set, as shown in equation (1), can be used to test the assumption of utility maximization. For example, we can test the utility maximization hypothesis by testing the separability properties of the function shown in equation (1).

The second aspect is that equation (1) suggests a way to recover the underlying preferences when utility maximization holds. If the budget constraint is linear we can regard this as a piecewise linear budget constraint where the slopes and virtual incomes of the budget constraint are all equal. This implies that all the π_k are equal, and equation (1) simplifies to $\pi - J(-\pi)$. Also, if the probability of no work is zero then the hours equation becomes π. This can occur if the support of v is bounded. Furthermore, if the probability of zero hours of work is very small, then setting all of the virtual incomes and wages to be equal will approximately give π.

This aspect does not depend on the convexity of the budget sets, since identical virtual incomes and wages will give the expected hours for a linear budget set. What it does depend on is that there is at least some data where the budget constraint is approximately linear. Consistency of a nonparametric estimator at any particular point, such as a linear budget constraint, depends on there being data in a neighborhood of that point. In practice, the estimator will smooth over data points near to the one of interest, which provides information that can be used to estimate expected hours at a linear budget constraint. Thus, data with approximately linear budget constraints will be useful for identification. Standard errors could be used to help to determine whether there is sufficient data to be reliable, because the standard errors will be large when there is little data.

It can be computationally complicated to do a nonparametric regression imposing all the constraints implied by expression (1). A simpler approach is to only take into account the separability properties implied by utility maximization. Going back to (1') we note that there is additive separability so we can write expected hours of work as

$$E(h^*) = f_1(\ell_1, w_1, y_1) + f_2(\ell_1, w_2, y_2) + f_3(\ell_2, w_2, y_2) + f_4(\ell_2, w_3, y_3). \qquad (2.2)$$

That is, there are four additive terms, with ℓ_1 appearing in two terms and ℓ_2

appearing in two terms.

Alternatively we can write expected hours of work as:

$$E(h^*) = \gamma_1(\ell_1, w_1, y_1) + \gamma_2(\ell_1, \ell_2, w_2, y_2) + \gamma_3(\ell_2, w_3, y_3) \tag{2.3}$$

Noting that $\ell_i = \frac{y_{i+1} - y_i}{w_i - w_{i+1}}$ we can also write $E(h^*)$ as

$$E(h^*) = \phi_1(y_1, w_1, y_2, w_2) + \phi_2(y_2, w_2, y_3, w_3) \tag{2.4}$$

That is, by giving up some of the separability properties we can reduce the dimensionality of the problem from 8 to 6. It is worth noting that if we use (2) or (3) there is an exact (nonlinear) relationship between some of the independent variables.

Equation (1) gives an expression for expected desired hours. However, we would normally expect that there also are measurement and/or optimization errors. If these errors are additive it is simple to take these errors into account. Let observed hours be given by: $\hat{h} = h^* + \varepsilon$, where $E(\varepsilon \mid x, v) = 0$. It follows that the expectation of observed hours will be the same as the expectation of desired hours.

The expressions above were derived under the assumption of a convex budget set. If the budget set is nonconvex we can do a similar, but somewhat more complicated derivation. The separability properties will weaken, but it is still true that expected hours of work is a function of the net wage rates, virtual incomes and kink points. We have also assumed that v is distributed independently of the budget sets and utility maximization holds. This condition will generally require that v have a bounded support.

3. Estimation method

If data were generated by a linear budget constraint defined by the slope w and intercept y, the expected hours of work would be given by $E(h \mid w, y) = g(w, y)$. If we do not know the functional form of $g()$, we can estimate it by, for example, kernel estimation. A crucial question is: how can we do nonparametric estimation when we have a nonlinear budget constraint. From the previous section we know that if the data-generating process is utility maximization with globally convex preferences, then the expected value of hours of work can be written as equation (1). If we do not know the functional form of (1) we can *in principle* estimate (1) by kernel estimation. However, because of the curse of dimensionality, this will usually be impossible in practice. In the study by Blomquist and Hansson-Brusewitz (1990) Swedish data with budget constraints consisting of up to 27 segments were used. To describe such a budget constraint we need 54 variables! Nonparametric estimation using actual budget constraints consisting of 27 segments would require a huge amount of data. To obtain a practical estimation procedure we therefore have to reduce the dimensionality of the problem.

Another reason to look for a more parsimonious specification is that when there are many budget segments relative to the sample size there may not be sufficient variation in the budget sets to allow us to estimate separate effects for each segment. That is, there may be little independent movement in the virtual incomes and wages for different segments. Therefore it is imperative that we distill the budget set variation, so that we capture the essential features of the data.

The estimation technique we suggest is a two-step procedure. In the first step each actual budget constraint is approximated by a budget constraint that can be represented by only a few numbers. In the second step nonparametric estimation via series approximation is applied, using the approximate budget constraints as data.

We consider two approaches to the first step of the estimator, the approximation of the true budget set by a smaller dimensional one.

i. *The least squares method.* Take a set of points $h_j, j = 1, ..., K$. Let $C(h_j)$ denote consumption on the true budget constraint and $\hat{C}(h_j)$ consumption on the approximating budget constraint. The criterion to choose the approximating budget constraint is $\text{Min} \sum_j [\hat{C}(h_j) - C(h_j)]^2$.

ii. *Interpolation method.* Take three values for hours of work: h_1, h_2 and h_3. Let $w(h_j)$, be the slope of the true budget constraint at h_j. Define linear budget constraints passing through h_j and with slope $w(h_j)$. The approximating budget constraint is given as the intersection of the three budget sets, defined by the linear budget constraints. The approximation depends on how the h_i are chosen and on how the slopes $w(h_j)$ are calculated.[2]

With the budget set approximation in hand we can proceed to the second step, which is nonparametric estimation of the labor supply function carried out as if the budget set approximation were true. The nonparametric estimator we

[2] One can, of course, use many other methods to approximate the budget constraints. One procedure would be to take the intercept of the budget constraint and 3 other points on the budget constraint and connect these points with linear segments.

consider is a series estimator, obtained by regressing the hours of work on several functions of the virtual income and wages. We use a series estimator rather than another type of nonparametric estimator, because it is relatively easy to impose additivity on that estimator.

To describe a series estimator let $x = (y_1, w_1, ..., y_J, w_J)'$ be the vector of virtual incomes and wage rates, and let $p^K(x) = (p_{1K}(x), ..., p_{KK}(x))'$ be a vector of approximating functions, each of which satisfies the additivity restrictions implied in equations (2), (3), or (4). For data (x_i, h_i), $(i = 1, ..., n)$, let $P = (p^K(x_1), ..., p^K(x_n))'$ and $H = (h_1, ...h_n)'$. A series estimator of $g(x) = E(h \mid x)$ is given by

$$\begin{aligned} \hat{g}(x) &= p^K(x)'\hat{\beta} \\ \hat{\beta} &= (P'P)^{-}P'H, \end{aligned} \tag{3.1}$$

where B^- denotes any symmetric generalized inverse.

Two types of approximating functions that can be used in constructing series estimators are power series and regression splines. In this paper we will focus on power series in the theory and application. For power series the components of $p^K(x)$ will consist of products of powers of adjacent pairs of the kinkpoint, virtual income, and wages. We also follow the common, sensible practice of using lower powers first.

Even with the structure implied by utility maximization there are very many terms in the approximation even for low orders. To help further with keeping the equation parsimonious it is useful to take the first few terms from a functional form implied by a particular distribution. Suppose for the moment that the budget approximation contains three segments, as it does in the application. Suppose also that the disturbance v was uniformly distributed on $[-u/2, u/2]$. Then, as shown in Appendix A,

$$\bar{h}(B) = [\ell_1(\pi_1 - \pi_2) + \ell_2(\pi_2 - \pi_3)] + (\pi_3 + u)^2/(2u).$$

Also suppose that $\pi(y, w) = \gamma_1 + \gamma_2 y + \gamma_3 w$. Then for $dy = \ell_1(y_1 - y_2) + \ell_2(y_2 - y_3)$ and $dw = \ell_1(w_1 - w_2) + \ell_2(w_2 - w_3)$,

$$\bar{h}(B) = \beta_1 + \beta_2 dy + \beta_3 dw + \beta_4 y_3 + \beta_5 w_3 + \beta_6 y_3^2 + \beta_7 w_3^2 + \beta_8 y_3 w_3, \tag{3.2}$$

where the coefficients of this equation satisfy, for $c = \gamma_1 + u$,

$$\begin{aligned} \beta_1 &= c^2/2u, & \beta_5 &= c\gamma_3/u, \\ \beta_2 &= \gamma_2/u, & \beta_6 &= (\gamma_2)^2/2u, \\ \beta_3 &= \gamma_3/u, & \beta_7 &= (\gamma_3)^2/2u, \\ \beta_4 &= c\gamma_2/u & \beta_8 &= \gamma_2\gamma_3/u. \end{aligned}$$

This function satisfies the additivity properties discussed earlier. We use this function by specifying the first eight terms in the series estimator to be one of the eight functions on the right-hand side of equation (6). Further flexibility is then obtained by adding other functions of virtual income and wages to the set of approximating functions. The estimator attains nonparametric flexibility by allowing for higher-order terms to be included, so that for large enough sample size the approximation might be as flexible as desired.

To make use of the nonparametric flexibility of series estimators it is important to choose the number of terms based on the data. In that way the nonparametric feature of the estimator becomes active, because a data-based choice of approximation allows adaptation to conditions in the data. Here we will use cross-validation to choose both the number of terms and to compare different specifications. The cross-validation criteria is

$$\begin{aligned} \hat{CV}(K) &= 1 - SSE(K)/[\textstyle\sum_{i=1}^{n}(h_i - \bar{h})^2], \\ SSE(K) &= \textstyle\sum_{i=1}^{n}[h_i - \hat{g}(x_i)]^2/[1 - p^K(x_i)'(P'P)^- p^K(x_i)]^2. \end{aligned}$$

The term $SSE(K)$ is the sum of squares of one-step ahead forecast errors, where all the observations other than the i^{th} are used to form coefficients for predicting the i^{th}. It has been divided by the sample sum of squares for h to make the criteria invariant to the scale of h. Cross-validation is known to have optimality properties for choosing the number of terms in a series estimator (e.g. see Andrews, 1991). We will choose the order of the series approximation by maximizing $CV(K)$, and also compare different models using this criterion.

4. Econometric theory

The estimator we have proposed is based on series estimation with virtual incomes and wages from a budget set approximation. This estimator uses two approximations. One is piecewise linear approximation of the true budget. The other is approximation of labor supply by a series regression. Here we derive convergence rates that account for both approximations. We also develop asymptotic normality results for the case where the budget set is exact.

For the budget set approximation we will focus on the case where the true budget sets are smooth and convex. Piecewise linear approximation of smooth budget sets seems a useful way to model the case in our empirical work where there are many linear segments that are being approximated by only a few segments. Also, the leading non-smooth budget set case is the piecewise linear one, where the budget set approximation error simply disappears when the number of segments is large enough. We restrict attention to the convex budget set case because the nonconvex case is inherently more difficult. Labor supply will no longer have the additive structure described earlier, so that the series approximation may require many more terms. However, if the non-convexities are not too pronounced, the convex approximation should be satisfactory. For example, in our empirical work the results were not affected much by convexifying the budget constraints. Also, the asymptotic normality results assume piecewise linear true budget sets, and do not rely on convexity of the budget sets.

The labor supply specification we consider is that of equation (1). We also focus on the nonparametric model described in Section 2, where the labor supply for a linear budget set is $\pi(y,w)+v$, where $\pi(y,w)$ is an unknown function and v is distributed independently of the budget set. This is a quite general model, subsuming many from the literature, and has enough structure to allow us to derive precise results.

4.1. Mean square convergence and the budget set approximation

We first derive convergence rates for the estimator while accounting for the budget set approximation. A fundamental property of $\bar{h}(B)$ that is important in controlling the budget set approximation error is that it is Lipschitz in B. To state that result we need some extra notation. Here we limit attention to convex budget sets where the budget frontier, $B(\ell), \ell \in \mathcal{L} = [0, \bar{\ell}]$ is concave and continuous. A concave function always has a right derivative $B_\ell^+(\ell)$ and a left derivative $B_\ell^-(\ell)$ at each ℓ, with $B_\ell^+(\ell) \le B_\ell^-(\ell)$. Define a norm of the budget frontier to be

$$\|B\| = \sup_{\ell \in \mathcal{L}} \left(|B(\ell)| + \left|B_\ell^+(\ell)\right| + \left|B_\ell^-(\ell)\right| \right).$$

With this notation the labor supply function is given by the solution $\ell(B,v)$ to

$$\pi(B(\ell) - \ell B_\ell^-(\ell), B_\ell^-(\ell)) + v \geq \ell \geq \pi(B(\ell) - \ell B_\ell^+(\ell), B_\ell^+(\ell)) + v,$$

where $B_\ell^-(0)$ is anything greater than $B_\ell^+(0)$ and $B_\ell^+(\bar{\ell})$ anything less than $B_\ell^-(\bar{\ell})$. This condition reduces to the equality $\ell = \pi(B(\ell) - B_\ell(\ell)\ell, B_\ell(\ell)) + v$ when $B(\ell)$ is differentiable at ℓ. There $B(\ell) - B_\ell(\ell)\ell$ and $B_\ell(\ell)$ are the virtual income and wage. A solution with $B_\ell^+(\ell) < B_\ell^-(\ell)$corresponds to a kink point. A solution will generally exist under weak conditions, e.g. if $\partial\pi(y,w)/\partial y < 0$. Here we will just assume that the solution exists.

To derive the results it is useful to impose some regularity conditions on the budget sets and the labor supply function $\pi(y, w) + v$.

Assumption 1: $\pi(y, w)$ is continuously differentiable with bounded derivatives. Also, there is a set $\mathcal{B}$ of concave budget frontiers $B : [0, \bar{\ell}] \to \Re$, and sets $\mathcal{Y}$, $\mathcal{W}$, and $\mathcal{V}$ such that $\mathcal{V}$ contains the support of v, $\mathcal{Y}$x$\mathcal{W}$ contains $(B(\ell)-\ell B_\ell^+(\ell), B_\ell^+(\ell))$ and $(B(\ell)-\ell B_\ell^-(\ell), B_\ell^-(\ell))$ for all $B \in \mathcal{B}$ and $\ell \in [0, \bar{\ell}]$, and $\pi(y,w)+v$ satisfies the Slutzky condition $\pi_w(y, w) - [\pi(y, w)+ v]\pi_y(y, w) > 0$, for all $(y, w, v) \in \mathcal{Y}x\mathcal{W}x\mathcal{V}$.

The Slutzky condition is helpful for bounding the effect of the budget set on labor supply. Here this economic restriction helps determine the continuity properties of labor supply.

Lemma 4.1. *If Assumption 1 is satisfied and $B(\ell)$ is twice continuously differentiable with $B \in \mathcal{B}$, then there is a constant C such that for any $\tilde{B} \in \mathcal{B}$ and $v \in \mathcal{V}$, $\left|\ell(\tilde{B}, v) - \ell(B, v)\right| \leq C\|\tilde{B} - B\|$.*

This result says that the labor supply is Lipschitz in the budget set, in terms of the norm $\|B\|$. It follows immediately from this result that $\left|\bar{h}(\tilde{B}) - \bar{h}(B)\right| \leq C\|\tilde{B}-B\|$. Thus, average labor supply at a general, smooth and convex budget set will be approximated by average labor supply at a close piecewise linear set, with an approximation error that is the same order as the budget set approximation error.

The budget set approximation can be combined with a series approximation of labor supply to obtain a total approximation error. Consider the formulation in equation (4), where labor supply is a sum of four dimensional functions of the triples

$$(w_j, y_j, w_{j+1}, y_{j+1}),\ (j = 1, ..., L - 1).$$

Let $x^L = (w_1 y_1, ..., w_L, y_L)$ and let $p^K(x^L)$ denote a $K \times 1$ vector of approximating functions, each of which depends only on one of the $(L-1)$ quadruples above. Here we assume that $p^K(x^L)$ is a four-dimensional power series, although it could be a tensor product spline. Assuming that the polynomials have comparable order for each j the order of the entire polynomial will be $(K/L)^{1/4}$. By Lorentz (1986, Theorem 8) it follows that the approximation error of an s-times differentiable function will be of the order $(K/L)^{-s/4}$. Combining this result with the budget set approximation rate leads to a rate of approximation of the true labor supply. Suppose that the following condition holds.

Assumption 2: $J(v)$ and $\pi(y, w)$ are s times continuously differentiable and for the subset $\mathcal{B}_2$ of $\mathcal{B}$ consisting of twice differentiable functions the derivative $B_{\ell\ell}(\ell)$ is uniformly bounded.

We now obtain the approximation rate result:

Lemma 4.2. *If Assumptions 1 and 2 are satisfied, then there is a constant C and for each K a vector β_K such that for every $B \in \mathcal{B}_2$ there is a piecewise linear budget set with associated x^L_B such that* $\sup_{B \in \mathcal{B}_2} \left|\bar{h}(B) - p^K(x^L_B)'\beta_L\right| \leq C\left(\frac{1}{L} + L\left(\frac{K}{L}\right)^{-s/4}\right)$.

This approximation rate result leads to a mean-square error (MSE) convergence rate for the nonparametric estimator. The following condition is useful for deriving that rate:

Assumption 3: $(h_1, x_1), ..., (h_n, x_n)$ are i.i.d. and $\mathrm{Var}(h|B)$ is bounded.

The bounded conditional variance is standard in the series estimation literature, and relaxing this condition would be difficult. Let $\hat{h}_i = p(x^L_{B_i})'\hat{\beta}_L$ and $\bar{h}_i = h(B_i)$.

Theorem 4.3. *If Assumptions 1 - 3 are satisfied then for each i there is $x^L_{B_i}$ such that* $\sum_{i=1}^n (\hat{h}_i - \bar{h}_i)^2/n = O_p(\frac{K}{n} + \frac{1}{L^2} + L^2\left(\frac{K}{L}\right)^{-2s/4})$.

The K/n term in the statement of the theorem is a variance term. The other two terms are bias terms that correspond to Lemma 2. These terms depend on both K and L. The best attainable convergence rate is obtained by choosing them so that each term converges to zero at the same rate. When this is done we obtain

$$\sum_{i=1}^n (\hat{h}_i - \bar{h}_i)^2/n = O_p(n^{-2s/(8+3s)}).$$

Here we find that the convergence rate is a power of n, in spite of the infinite dimensional nature of the budget set. As the number of derivatives of the supply function (i.e. its smoothness) increases, the convergence rate increases, approaching $n^{-1/3}$ as s grows. This bound on the rate is smaller than the usual one of $n^{-1/2}$, being limited by the use of a piecewise linear approximation to the budget set and its derivative. In particular $n^{-1/3}$ is the best rate that could be attained by a linear spline approximation of a function and its derivative, as in Stone (1985).

Applying this result in practice would requires choosing a piecewise linear budget set approximation that satisfies the conditions of Lemma 2. This could be done by choosing the approximate budget set B_i^L so that $\|B_i^L - B_i\|$ was within $1/L$ of its infimum. The least squares approximation used in the empirical work is a way of implementing such an approximation, because mean-square error and supremum norms are equivalent for functions with uniformly bounded derivatives, and when convex functions are close in a supremum norm their derivatives are also close.

4.2. Asymptotic Normality

In deriving asymptotic normality results it is difficult to account for the budget set approximation. The difficulty is a technical one, due to the relatively slow approximation of the true budget set by a piecewise linear one. The best available series asymptotic normality results, in Newey (1997), have upper bounds for K that do not allow the bias to shrink fast enough. This difficulty could be overcome by using other kinds of budget set approximations, leading to different empirical methods. We leave these extensions to future work.

The following conditions are useful for the asymptotic normality results:

Assumption 4: The support of x is a Cartesian product of compact connected intervals on which x has a probability density function that is bounded away from zero.

This assumption can be relaxed by specifying that it only holds for a component of the distribution of x (which would allow points of positive probability in the support of x), but it appears difficult to be more general. It is somewhat restrictive, requiring that there be some independent variation in each of the individual virtual incomes and wages. Also, it requires that the upper bound and lower bounds for the virtual incomes not overlap with each other.

These conditions allow us to derive population MSE and uniform convergence rates that complement the rates given above. These rates are for different criteria than above, but do not allow for the budget set approximation. Let X denote the support of x, and $F_0(x)$ the distribution function of x_i.

Theorem 4.4. *If Assumptions 2 - 4 are satisfied and $K^3/n \to 0$ then*

$$\begin{aligned} \int [\hat{g}(x) - g_0(x)]^2 dF_0(x) &= O_p(\frac{K}{n} + K^{-2s/4}) \\ \sup_{x \in X} |\hat{g}(x) - g_0(x)| &= O_p(K[\sqrt{\frac{K}{n}} + K^{-s/4}]) \end{aligned}$$

This result gives mean square and uniform convergence rates for the estimated expected labor supply function. The different terms in the convergence rates correspond to bias and variance. If the number of terms is set so that the mean square convergence rate is as fast as possible, with K proportional to $n^{2/(s+2)}$, the mean square convergence rate is $n^{-s/(s+2)}$. This rate attains Stone's (1982) bound for the four-dimensional case, that is, the rate is as fast as possible for a four-dimensional function. Thus, the additivity of the expected-hours equation leads to a convergence rate which corresponds to a four-dimensional function, rather than the potentially very slow $2J$ dimensional rate.

To show asymptotic normality we need to be precise about the object of estimation. Also, an important use of these results is in asymptotic inference, where a consistent estimator of the asymptotic variance is needed. Suppose that a quantity of interest can be represented as $\theta_0 = a(g_0)$ where $a(g)$ depends on the function g and is linear in g. For example, $a(g)$ might be the derivative of the function at a particular point, or an average derivative. The corresponding estimator is

$$\hat{\theta} = a(\hat{g}). \tag{4.1}$$

A standard error for this estimator can be constructed in the usual way for least squares. Let $A = (a(p_{1K}), ..., a(p_{KK}))'$ and

$$\begin{aligned} \hat{V} &= A'\hat{Q}^{-}\hat{\Sigma}\hat{Q}^{-}A, \\ \hat{Q} &= P'P/n, \\ \hat{\Sigma} &= \sum_{i=I}^{n} p^K(x_i)p^K(x_i)'[h_i - \hat{g}(x_i)]^2/n. \end{aligned} \tag{4.2}$$

This estimator is just the usual one for a function of least squares coefficients, with $\hat{Q}^{-}\hat{\Sigma}\hat{Q}^{-}$ being the White (1980) estimator of the least-squares asymptotic variance for a possibly misspecified model. This estimator will lead to correct asymptotic inferences because it accounts properly for variance, and because bias will be small relative to variance under the regularity conditions discussed below.

Some additional conditions are important for the asymptotic normality result.

Assumption 5: $E[\{h - g_0(x)\}^4|x]$ is bounded, and $\text{Var}(h|x)$ is bounded away from zero.

This assumption requires that the fourth conditional moment of the error is bounded, strengthening Assumption 1.

Assumption 6: $a(g)$ is a scalar, there exists C such that $|a(g)| < C \ \sup_{x \in X} |g(x)|$, and there exists $g_K(x) = p^K(x)'\tilde{\beta}$ such that $E[g_K(x)^2] \to 0$ and $a(g_K)$ is bounded away from zero.

This assumption says that $a(g)$ is continuous in the supremum sense, but *not* in the mean-square norm $(E[g(x)^2])^{1/2}$. The lack of mean-square continuity is a useful regularity condition and will also imply that the estimator $\hat{\theta}$ is not $\sqrt{n}$-consistent. Another restriction imposed is that $a(g)$ is a scalar, which is general enough to cover many cases of interest.

To state the asymptotic normality result it is useful to work with an asymptotic variance formula. Let $\sigma^2(x) = \text{Var}(h \mid x)$. Let

$$\begin{aligned} V_K &= A'Q^{-1}\Sigma Q^{-1}A, \qquad (4.3) \\ Q &= E[p^K(x)p^K(x)'], \\ \Sigma &= E[p^K(x)p^K(x)'\sigma(x)^2]. \end{aligned}$$

Theorem 4.5. *If Assumptions 3-6 are satisfied, $K^3/n \to 0$, and $\sqrt{n}K^{-s/4} \to 0$ then $\hat{\theta} = \theta_0 + O_p(K^{3/2}/\sqrt{n})$ and*

$$\begin{aligned} \sqrt{n}V_K^{-1/2}(\hat{\theta} - \theta_0) &\xrightarrow{d} N(0,1), \\ \sqrt{n}\hat{V}_K^{-1/2}(\hat{\theta} - \theta_0) &\xrightarrow{d} N(0,1). \end{aligned}$$

This result can be used to construct an asymptotic confidence interval of the form $(\hat{\theta} - z_{\alpha/2}\sqrt{\hat{V}}, \hat{\theta} + z_{\alpha/2}\sqrt{\hat{V}})$, where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. The two rate conditions are those of Newey (1997). The first ensures convergence in probability of the second moment matrix of the approximating functions, after a normalization. The second ensures that the bias is small relative to $\sqrt{n}$. The existence of K satisfying both conditions requires $s > 6$, a smoothness condition that is somewhat stronger than for asymptotic normality of other nonparametric estimators. The convergence rate for $\hat{\theta}$ is only a bound, so it may be possible to derive more precise results. In particular, one obtains $\sqrt{n}$ consistency under slightly different conditions.

The following condition is crucial for $\sqrt{n}$-consistency.

Assumption 7: There is $v(x)$ with $E[v(x)v(x)']$ finite and nonsingular such that $a(g_0) = E[v(x)g_0(x)]$, $a(p_{kK}) = E[v(x)p_{kK}(x)]$, for all k and K, and there is $\tilde{\beta}_K$ with $E[\|v(x) - p^K(x)'\tilde{\beta}_K\|^2] \to 0$.

This condition allows for $a(g)$ to be a vector. It requires a representation of $a(g)$ as an expected outer product, when g is equal to the truth or any of the approximating functions, and for the functional $v(x)$ in the outer product representation to be approximated in mean-square by some linear combination of the functions. This condition and Assumption 6 are mutually exclusive, and together cover most cases of interest (i.e. they seem to be exhaustive). A sufficient condition for Assumption 7 is that the functional $a(g)$ be mean-square continuous in g over some linear domain that includes the truth and the approximating functions, and that the approximation functions form a basis for this domain. The outer product representation in Assumption 7 will then follow from the Riesz representation theorem. The asymptotic variance of the estimator will be determined by the function $v(x)$ from Assumption 7. It will be equal to

$$V = E[v(x)v(x)' \operatorname{Var}(h \mid x)]. \tag{4.4}$$

Theorem 4.6. *If Assumptions 2 - 5 and 7 are satisfied, $K^3/n \to 0$, and $\sqrt{n}K^{-s/4} \to 0$ then*

$$\begin{aligned} &\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V), \\ &\hat{V} \xrightarrow{p} V. \end{aligned} \tag{4.5}$$

5. Sampling Experiments

There are three questions we want to study. First, suppose we do not have to approximate budget constraints, how well would then an estimation method that regresses hours of work on the slopes and intercepts of the budget constraint work? Second, how much "noise" is introduced in the estimation procedure if we instead of actual budget constraints use approximated budget constraints. The answer to the second question depends on how the approximation is done. Hence, we would like to study the performance of the estimation procedure for various methods tc approximate budget constraints. Third, we would like to know how well a nonparametric labor supply function can predict the effect of tax reform. We have studied these three questions using both actual and simulated data. To judge the performance of our suggested estimation procedure we use the cross-validation measure previously presented.

Evaluation of budget approximation methods using actual data

We have performed extensive estimations on actual data from 1973, 1980 and 1990 to compare the relative performance of the least squares and the interpolation methods where performance is measured by the cross-validation criteria. For the least squares method we must specify the set of points h_i, $i = 1,..,K$. We have subdivided this into the choice of the number of points to use, the type of distribution from which the h_i are chosen and the length of the interval defined by the highest and lowest values for the h_i. We tried three types of distributions: a uniform distribution, a triangular distribution and the square root of the observed distribution. For the interpolation method we must specify three points h_1, h_2, h_3 and how to calculate the slope of the actual budget constraint at the chosen points. We have used a function linear in virtual incomes and net wage rates to evaluate the various approximation methods.

Using data from 1981 one particular specification of the interpolation method works best of all methods attempted. Unfortunately, this specification works quite badly for data from 1990. Hence, the interpolation method is not robust in performance across data generated by different types of tax systems. Since we want to use our estimated function to predict the effect of tax reform this is a clear disadvantage of the interpolation method. The least squares method is more robust across data from different years. We have not found a specification of the least squares method that is uniformly best across data from different years. However, the least squares method using a uniform distribution over the interval 0-5000 hours and represented by 21 points has a relatively good cross-validation performance for data from all years. This is the approximation method we use in the rest of the study.

Monte Carlo Simulations

We perform two sets of Monte Carlo simulations. In the first set of simulations we use data from only one point in time, namely data from LNU 1981. For 864 males in ages 20 to 60 we use the information on their gross wage rates and non-labor income to construct budget constraints and generate hours of work using the preferences estimated and reported in Blomquist and Hansson-Brusewitz (1990). It should be noted that for a majority of individuals the budget sets are nonconvex.

The basic supply function is given by: $h^* = 1.857 + v + 0.0179w - 3.981 * 10^{-4}y + 4.297 * 10^{-3}AGE + 2.477 * 10^{-3}NC$, where $v \sim N(0, 0.0673)$, hours of work are measured in thousands of hours, the wage rate is given in 1980 SEK and the virtual income in thousands of 1980 SEK. AGE is an age dummy , NC a dummy for number of children living at home and SEK is a shorthand for Swedish kronor. Observed hours of work is given by $h = h^* + \varepsilon$, where $\varepsilon \sim N(0, 0.0132)$.

We use the following four types of data generating processes (DGP):

i. Fixed preferences; no measurement error. (That is we assume all individuals have identical preferences.)

ii. Fixed preferences and measurement errors;

iii. Random preferences; no measurement error.

iv. Random preferences and measurement errors.

The simulations presented in Table 1 show how well the procedure works if we use *actual* budget constraints in the estimation. Hence, when generating the data we use budget constraints consisting of three linear segments. These budget constraints were obtained as approximations of individuals' 1981 budget constraints. The constructed data are then used to estimate labor supply functions. The same budget constraints that were used to generate the data are used to estimate the nonparametric regression. The following 5 functional forms were estimated:[3]

1. linear in w_i, y_i, $i = 1, 2, 3$.
2. linear in w_i, y_i, $i = 1, 2, 3$ and ℓ_1 and ℓ_2.
3. quadratic form in w_i, y_i, $i = 1, 2, 3$.
4. quadratic form in w_i, y_i, $i = 1, 2, 3$ and linear in ℓ_1 and ℓ_2

[3] We also tried some other functions. Adding more terms, like squares of the kink points and more interaction terms increase the coefficient of determination but yields a lower cross validation measure.

5. linear form in *const.*, dy, dw, w_3, y_3, w_3^2, y_3^2.

In the first row we present results from simulations with a DGP with no random terms. The variation in hours of work across individuals only depends on the variation in budget constraints. The reason why the coefficient of determination is less than one is that we use an incorrect specification of the function relating hours of work as a function of the net wage rates, virtual incomes and kink points. As we add more random terms to the DGP the values for the coefficient of determination and the cross validation measure decrease. Looking across columns, we see that in terms of the coefficient of determination the functions containing many quadratic and interaction terms do well. However, looking at the cross validation measure the simpler functional forms containing only linear terms perform best. For the DGP with both random preferences and measurement error function 2 performs slightly better than function 1.

Table1. Evaluation of Estimation Method using constructed "actual" budget constraints. Coefficient of determination and Cross validation used as performance meaure. Average over 500 replications.

DGP		**function 1**	**function 2**	**function 3**	**function 4**	**function 5**
No random	Average R^2	0.601	0.604	0.644	0.658	0.450
terms	Average CV	0.581	0.576	0.556	0.536	0.392
Measurement	Average R^2	0.215	0.218	0.245	0.252	0.163
error	Average CV	0.194	0.190	0.136	0.123	0.128
Random	Average R^2	0.125	0.137	0.167	0.184	0.083
preferences	Average CV	0.103	0.106	0.010	0.013	0.052
Random pref	Average R^2	0.098	0.107	0.135	0.149	0.066
+meas. error	Average CV	0.075	0.078	-0.016	-0.015	0.037

Suppose data are generated by budget constraints consisting of z number of segments. How well does our method do if we use approximated budget constraints in the estimation procedure? The simulations presented in Table 2 show how well the procedure works if we generate data with budget constraints consisting of up to 27 linear segments, but in the estimation use approximated budget constraints consisting of only three segments. We use the OLS procedure described above to approximate the actual data generating budget constraints. The weight system is a uniform distribution over the interval 0-5000 hours. We use 21 points to represent the distribution. We use the same functional forms as in Table 1.

Comparing the results presented in table 2 with those in Table 1 we find, somewhat surprisingly, that the R^2s and CVs in Table 2 in general are higher than those in Table 1. This is especially so for the case when there are random preferences but no measurement error. The fact that we in the estimation use approximated budget constraints does not impede the applicability of the estimation procedure.

Table 2. Evaluation of Estimation Method using approximated budget constraints in the est Coefficient of determination and Cross validation used as performance meaure. Averages ove 500 replications.

DGP		function 1	function 2	function 3	function 4	function 5
No random	Average R^2	0.746	0.757	0.781	0.785	0.668
terms	Average CV	0.738	0.748	0.715	0.671	0.633
Measurement	Average R^2	0.183	0.187	0.209	0.212	0.165
error	Average CV	0.165	0.165	0.100	0.084	0.139
Random	Average R^2	0.420	0.428	0.480	0.481	0.372
preferences	Average CV	0.398	0.400	0.325	0.314	0.320
Random pref	Average R^2	0.157	0.161	0.195	0.196	0.141
+meas. error	Average CV	0.136	0.135	0.059	0.049	0.107

Why are the R^2s and CVs higher in Table 2 than in Table 1, especially when there are random preferences? We provide the following explanation. If the budget constraint is linear, the effect of random preferences is the same as the measurement error. If there is one sharp kink in the budget constraint, desired hours will be located at this kink for a large interval of v. That is, the kink will reduce the dispersion in hours of work as compared with a linear budget constraint. In the DGP used for the simulations presented in Table 2 we use budget constraints with up to 27 linear segments. The presence of so many kinks greatly reduces the effect of the random preferences on the dispersion of hours of work. It is true that for the three-segment budget constraints used for the simulations presented in Table 1 the kinks are more pronounced. On balance it turns out that the DGP used in Table 2 is affected less by the random preferences than what is the DGP used for the simulations presented in Table 1.

Looking across rows in Table 2 we see that adding more of random terms to the DGP decreases both the R^2s and CVs. However, while in Table 1 the inclusion of random preferences reduced the R^2s and CVs most, in Table 2 it is the inclusion of measurement error that decreases the R^2s and CVs most. Looking across columns and approximating functions we find that the coefficient of determination increases as we include more squares and interactions, while the cross validation decreases. In terms of the cross validation measure a linear form in virtual incomes, net wage rates and the kink points shows the best performance. This is the same result as in Table 1.

Much of the interest in labor supply functions stems from a wish to be able to predict the effect of changes in the tax system on labor supply. We have therefore performed a second set of simulations to study how well a function estimated with the estimation procedure suggested can predict the effect of tax reform on hours of work. For these simulations we use data from three points in time:

i. We use individuals' actual budget constraints from 1973, 1980 and 1990 in combination with the labor supply model estimated and presented in Blomquist and Hansson-Brusewitz (1990). (See the labor supply function shown on pp. 19 above.) This model contains both random preferences and measurement errors. Thus, the data-generating process is utility maximization subject to nonconvex budget constraints.

ii. The generated data are used to estimate both parametric and nonparametric labor supply functions. We estimate eight different functional forms for the nonparametric function.

iii. We perform a tax reform. We take the 1991 tax system as described in Section 7 and appendix D to construct post-tax budget constraints for the 1980 sample. Using the labor supply model from Blomquist and Hansson-Brusewitz (1990) we calculate "actual" post tax hours for all individuals in the 1980 sample.

iv. Approximating the post-tax reform budget constraints we then apply our estimated function to predict after tax reform hours.

Let

H_{BTR} = actual average hours of work before the tax reform.
H_{ATR} = actual average hours of work after the tax reform.
$\hat{H}_{BTR}$ = predicted before tax reform average hours of work.
$\hat{H}_{ATR}$ = predicted after tax reform average hours of work.

The actual percentage change in average hours of work is given by

$$M = (H_{ATR} - H_{BTR})/H_{BTR}.$$

We can calculate the predicted percentage change in hours of work in two ways

$$\begin{aligned} M1 &= (\hat{H}_{ATR} - \hat{H}_{BTR})/\hat{H}_{BTR}, \\ M2 &= (\hat{H}_{ATR} - H_{BTR})/H_{BTR}. \end{aligned}$$

The average value of M is 0.0664. In table 3 we show the average values of $M1$, $M2$ and the CV over 100 iterations.

When researchers predict the effect of tax reform the before tax reform hours are usually known. In actual practice a measure like $M2$ is often calculated. There are proponents for a measure where the before tax reform hours also are predicted. In this simulation, as is common in actual practice, the predicted before tax reform hours is a within-sample prediction, whereas the after-tax-reform prediction is an out-of-sample prediction. It is not shown in the table, but the predicted before-tax-reform hours are predicted quite well. The error in the after tax reform hours is larger.

Table 3. Average values of M1, M2 and CV over 100 iterations

	Model	M1	M2	CV
function 1	$const., dy, dw$	-0.0171	0.0044	0.0121
function 2	above and w_3, y_3	0.0554	0.0538	0.1147
function 3	above and y_3^2	0.0546	0.0532	0.1147
function 4	above and w_3^2	0.0506	0.0521	0.1189
function 5	above and w_3, y_3	0.0506	0.0521	0.1183
function 6	above and ℓ_1, ℓ_2	0.0517	0.0530	0.1157
function 7	above and y_2, w_1, w_2	0.0511	0.0517	0.1328
function 8	above and ℓ_1^2, ℓ_2^2	0.0625	0.0621	0.1416
Maximum likelihood estimate		0.0784	0.0704	

According to Table 3, function 8 performs on average best. In fact, in 99 of the iterations function 8 achieved the highest CV. In one iteration function 7 had a slightly higher CV than function 8. We see that the nonparametric estimation method can predict the effect of the tax reform quite well. The actual change in hours of work is 6.64% while the predicted change on average is 6.25%. The maximum likelihood based prediction slightly over predicts the effect.

In Table 4 we use the same *DGP* as in table 3, except for the measurement error. The measurement error used to generate data for Table 4 is a simple transformation of the random terms in the previous *DGP*. The measurement error χ is given by $\chi = \varepsilon^2/5$. The likelihood function used is the same as for Table 3. This means that the likelihood function is misspecified. We see that the nonparametric estimates in Tables 3 and 4 are very close. However, the maximum likelihood estimate over predicts the effect of tax reform when the likelihood function is incorrectly specified. In Table 4 the ML estimate predicts an increase in hours of work of 11.40% as measured by M1 and 9.72% as measured by M2 although the true increase is 6.64%.

Table 4.

Model	**M1**	**M2**	**Average CV**
const, *dy*, *dw*	-0.0172	0.0433	0.0204
above and w_3, y_3	0.0554	0.0538	0.1852
above and y_3^2	0.0547	0.0532	0.1853
above and w_3^2	0.0507	0.0521	0.1924
above and w_3, y_3	0.0507	0.0521	0.1916
above and ℓ_1, ℓ_2	0.0515	0.0527	0.1879
above and y_2, w_1, w_2	0.0511	0.0517	0.2171
above and ℓ_1^2, ℓ_2^2	0.0627	0.0622	0.2324
Maximum likelihood estimate	0.1140	0.0972	

6. Estimation on Swedish data

6.1. Data source

We use data from three waves of the Swedish "Level of Living" survey. The data pertain to the years 1973, 1980 and 1990. The surveys were performed in 1974, 1981 and 1991. The 1974 and 1981 data sources are briefly described in Blomquist (1983) and Blomquist and Hansson-Brusewitz (1990) respectively. The 1990 data is based on a survey performed in the spring of 1991. The sample consists of 6,710 randomly chosen individuals aged 18-75. The response rate was 79.1%. Certain information, like taxation and social security data, were acquired from fiscal authorities and the National Social Insurance Board.[4]Sample statistics are provided in appendix B.

In the estimation we only use data for married or cohabiting men in ages 20-60. Farmers, pensioners, students, those with more than 5 weeks of sickleave, those who were liable for military service and the self employed are excluded. This leaves us with 777 observations for 1973, 864 for 1980, and 680 for 1990.

The tax systems for 1973 and 1980 are described in Blomquist (1983) and Blomquist and Hansson-Brusewitz (1990). The tax system for 1990 is described in Appendix C. Housing allowances have over time become increasingly important. For 1980 and 1990 we have therefore included the effect of housing allowances on the budget constraints. The housing allowances increase the marginal tax rates in certain intervals and also create nonconvexities.

The fact that we pool data from three points in time has the obvious advantage that the number of observations increases. Another important advantage is that we obtain a variation in budget sets that is not possible with data from just one point in time. The tax systems were quite different in the three time periods which generates a large variation in the shapes of budget sets.

6.2. Parametric estimates

We pool the data for the three years and estimate our parametric random preference model described in, for example, Blomquist and Hansson-Brusewitz (1990). The data from 1973 and 1990 were converted into the 1980 price level. We have also convexified the budget constraints for data from 1980 and 1990. We show the results in equation (14). The elasticities E_w and E_y are calculated at the mean values of hours of work, net wages and virtual incomes. The means are taken over all years. t-values are given in parenthesis beneath each coefficient.[5]'

$$\begin{array}{llllll} h= & 1.914 & +0.0157w & -8.65*10^{-4}y & -9.96*10^{-3}AGE & -3.46*10^{-3}NC \\ & (62.09) & (8.96) & (-5.95) & (-0.53) & (-0.44) \end{array} \tag{6.1}$$

[4]Detailed information on the 1990 data source can be found in Fritzell and Lundberg (1994).

[5]The variance-covariance matrix for the estimated parameter vector is calculated as the inverse of the Hessian of the log-likelihood function evaluated at the estimated parameter vector. We have had to resort to numerically calculated derivatives. It is our experience that the variance-covariance matrix obtained by numerical derivatives give less reliable results than when analytic derivatives are used.

6.3. Nonparametric estimates

Below we report results when we have pooled data for the three years.[6] We use a series estimator. As our criterion to choose the estimating function we use the cross validation measure presented earlier. We have used two different procedures to approximate individuals' budget constraints. In the first procedure we apply the least squares approximation to individuals' original budget constraints. In the second procedure we first convexify the budget constraints by taking the convex hull and then apply the least squares approximation. The budget constraints from 1980 and 1990 are nonconvex, so the two procedures differ. To approximate the budget constraints we have used the least squares method with the span from 0 to 5000 hours and with 21 equally-spaced points. It turns out that the results are very similar whether we approximate the original or the convexified constraints. As shown in Table 5 the cross validation measure is a little bit higher for the best performing approximating functions when we approximate the original budget constraints without first convexifying. In the following we therefore only report the results for the functions estimated on approximated budget constraints from original budget constraints. We only report results for functions estimated on approximated budget constraints consisting of three piece-wise linear segments. We have also tried approximations with four segments, but these approximations yielded lower cross validation measures.

In Table 5 we present a partial listing of how the cross validation measure varies w.r.t. the specification of the estimating function. In Table 6 we report the estimated coefficients for the two specifications with the highest cross-validation measure.[7] We have also used the data to test restrictions implied by utility maximization with convex budget sets. This test was performed by estimating a function allowing for interactions between the regressors that violates the separability properties. (See the discussion on p. 6.) These interaction terms were not significant.

[6] We have also estimated nonparametric functions for individual years. However, the standard errors are considerably larger for the individual years as compared to when we pool the data.

[7] We note that the functional form with the highest CV differs between Table 5 and, say, Tables 3 and 4. This is not surprising since the DGP for the actual data presumably is different from the one used in the simulations presented in Tables 3 and 4. We also see that the functional form with the highest CV differ between Tables 1 and 2 versus Tables 3 and 4. However, Tables 1 and 2 are based on only the 1980 data, while Tables 3 and 4 use data from all three years, and one would expect that the form with highest CV might have more terms in the larger data sets.

Table 5. Nonparametric estimation on all years. Cross-validation values

Variables included	**Original budget constraints nonconvex**	**Original budget constraints convexified**
const., dy, dw	0.0073	0.0057
above and w_3 , y_3	0.0323	0.0291
above and y_3^2	0.0373	0.0350
above and w_3^2	0.0366	0.0341
above and $w_3 y_3$	0.0360	0.0340
above and ℓ_1, ℓ_2	0.0358	0.0336
above and y_2, w_1, w_2	0.0278	0.0310
above and ℓ_1^2, ℓ_2^2	0.0268	0.0288

It would be of interest to have a summary measure of how these functions predict hours of work to change as budget constraints change. For data generated by linear budget constraints one often reports wage and income elasticities. These are summary measures of how hours of work react to a change in the slope and intercept of a linear budget constraint. Can we calculate similar summary measures for the functions reported in Table 6? The functions reported in Table 6 are estimated on nonlinear budget constraints, and are useful for predicting changes in hours of work as such constraints change. However, we could regard a linear budget constraint as a limiting case of a nonlinear one. If the wage rates and virtual incomes for the three segments approach a common value the budget constraint approaches a linear one. It turns out that if the wage rates and virtual incomes are the same for all three segments the terms dw and dy drop out of the functions. We are left with the w_3 and y_3 terms. The coefficients for these terms can be used to calculate wage and income elasticities.. The elasticities reported are calculated at the mean of hours of work, the wage rate and virtual income. The means are taken for the segments where individuals are observed and calculated over all three years. Hence, all elasticities are evaluated at the same values for the wage rate, virtual income and hours of work. The fact that the first three functions include a term with the wage rate squared implies that the wage elasticity measure is very sensitive to the point at which the elasticity is evaluated.

In comparison with the parametric estimates, the nonparametric ones show less sensitivity of the hours supplied to the wage rate, and more sensitivity to nonlabor income. Both the elasticity and coefficient estimates show this pattern.

The nonparametric elasticity estimate is smaller than the parametric one for the wage rate and larger for non-labor income. Also, for the nonparametric estimates in the first column of Table 6, the coefficient of w_3 is smaller than is the wage coefficient for the parametric estimate in equation (14). As previously noted, the coefficient of w_3 gives the wage effect for a linear budget set, because dw is identically zero in that case.

The wage and income elasticities are evaluated at the mean of the net wage rates and virtual incomes from the segments where individuals observed hours of work are located.[8] Of course, the wage and income elasticities are summary measures of how the estimated functions predict how changes in a linear budget constraint affect hours of work. None of the budget constraints used for the estimation are linear, and we actually never observe linear budget constraints. It is therefore of larger interest to see how the predictions differ between the parametric and nonparametric labor supply functions for discrete changes in nonlinear budget constraints. In section 7 we use the estimated functions to predict the effect on hours of work of Swedish tax reform.

[8] Ackum Agell and Meghir (1995), using another data source and an instrumental variables estimation technique, present wage elasticities that are quite similar to those presented here.

Table 6. Nonparametric estimates using pooled data

Variables	Best function	Next-best function
Const.	2.064	2.097
	(49.85)	(39.69)
dy	-0.00210	-0.00204
	(-4.37)	(-4.28)
dw	-0.00145	-0.00131
	(-1.16)	(-1.06)
y_3	-0.0036	-0.0037
	(-3.95)	(-4.01)
w_3	0.00964	0.00560
	(6.61)	(1.40)
y_3^2	1.98×10^{-5}	2.00×10^{-5}
	(3.40)	(3.42)
w_3^2		1.16×10^{-4}
		(1.01)
wage elasticity	0.075	0.074
	(6.61)	(6.60)
income elasticity	-0.038	-0.040
	(-4.31)	(-4.37)
Cross validation	0.0373	0.0366
R^2	0.0435	0.0440

t-values in parentheses. The delta method was used to calculate the t-values for the elasticities.

In Table 7 we report estimates of the basic supply function $\pi(y, w)$ when we impose the functional form for the conditional mean implied by utility maximization and specific distributions of individual heterogeneity. The estimates are obtained by estimating equation (1) given an assumption on the distribution of v. We recover $\pi(\cdot)$ from the relation $E(h^*) = \pi - J(\pi)$, which shows expected hours of work if data are generated by a linear budget constraint. Surprisingly, the coefficient estimates for both the wage and nonlabor income are substantially lower for the parametric regression specification in Table 7 than for either the maximum likelihood or the nonparametric estimation procedure. This provides some evidence against the distributional assumptions that are imposed on the estimates in Table 7. The standard errors for the Gaussian conditional mean estimates are not reported because they were implausibly large. For the uniform estimates, assuming homoskedasticity leads to a simple Hausman test of the distributional assumption. Comparing the coefficient of w_3 in the first column of Table 6 with the coefficient of w in the first column of Table 7 gives a Hausman statistic 6.53 that should be a realization of a standard normal distribution. This is an implausibly large value, providing evidence against the uniform distributional model.

7. Tax reform

In this section we use the estimated functions to predict the effect of recent changes in the Swedish income tax.[9] The purpose is not to give a detailed evaluation of Swedish tax reform, but rather to see the difference in predictions across estimated functions.[10] Around 1980 the Swedish tax system reached a peak in terms of high marginal tax rates. Then, gradually during the '80's the marginal tax rates were lowered with a quite large change in the tax system between 1990 and 1991. We will use the actual distribution of gross wage rates and non-labor income from the 1980 data set to calculate the effect of the changes in the tax system between 1980 and 1991. The 1980 income tax system is described in Blomquist and Hansson-Brusewitz (1990). We present the most important aspects of the 1991 income tax system in Appendix D.

The income tax consists of two parts. There is a proportional local income tax which has been largely unchanged since 1980. The average local income tax rate has increased from 29.1% to 31%. The federal income tax consists of two important parts. First, the marginal tax rates have fallen significantly. Secondly, in 1980 interest payments were fully deductible against labor income, while in 1991 30% of interest payments were deductible from other taxes. We will study the effect of the change in the income tax schedule, but we will not take account of the change in deduction rules. There have also been changes in the VAT and the payroll tax. These changes are of course also important for the shape of individuals' budget constraints. We could model the effect of the change in VAT and the payroll tax as a change in the real wage rate. However, we have chosen to represent it as a change in the proportional income tax rate. In Appendix D

[9] There exist alternative approaches to evaluate the effect of tax reform on labor supply. Blundell et. al. (1998) and Eissa (1995) use difference in differences estimators to estimate the effect of tax reform on female hours of work.

[10] Agell *et al.* (1995) contain a broad evaluation of the Swedish tax reform. Aronsson and Palme (1995) also contain a description of tax reform in Sweden. They present labor supply functions derived from a household model and estimated by a maximum likelihood technique.

we describe how this is done. Taking account of the change in VAT and payroll taxes the income tax reform implies a decrease in the highest federal tax rate from 58% to 25%.

Predictions based on parametric and nonparametric labor supply functions

We use the labor supply function estimated on pooled data from 1973, 1980 and 1990 by the maximum likelihood method and shown as equation (14). The estimation method used assumes the budget sets are convex, so the function is estimated on convexified budget sets. However, since we estimate a well-defined direct utility function we can when we calculate the effect of tax reform either use the original nonconvex budget sets or convexified ones. It turns out that the difference in predictions is negligible. Using the original nonconvex budget sets the prediction is that average hours of work increase by 6.1%, from 2073 to 2200.[11]

Table 8 gives the predictions for various nonparametric specifications along with standard errors. We find that the prediction is not very sensitive to functional form specification. The functions shown in Table 8 are estimated on approximated budget constraints where some of the original budget constraints are nonconvex. We have also estimated supply functions on approximated budget constraints where we first have convexified the original budget constraints. The results are very close. For example, for the specification in Table 8 that predicts an increase in hours of work of 2.98% the prediction obtained using convexified original budget constraints is 2.43%. The standard error for both predictions is around 0.009. Hence, the difference in the predictions is slightly more than half a standard deviation. It does not seem to be important whether we use the original nonconvex or convexified budget constraints in our estimation procedure. The prediction obtained from the nonparametric labor supply function is considerably lower than that obtained from the parametric labor supply function.

The nonparametric estimates of the policy shift are less than half the size of the parametric estimates. We can construct a Hausman test statistic to check for statistical significance of this difference. Under a null hypothesis that the parametric model is correct the parametric estimator of the policy shift will be the MLE of the policy shift, by invariance of MLE, and is therefore an efficient estimator. Under the alternative of misspecification the nonparametric estimator will be consistent and is also asymptotically normal because it is an average like that considered in Theorem 3. Therefore, we can construct a test statistic from the difference of the parametric and nonparametric estimators divided by the square

[11] The averages are taken over 20 simulations with different drawings of the random preference terms in each simulation. The standard error of the simulation error is estimated to 0.0065.

root of the difference of their variance estimates.

We constructed a standard error for the parametric prediction by using the delta method, with numerical derivatives of the prediction with respect to the likelihood parameters. The standard error was sensitive to the finite differences chosen for the numerical derivative, ranging from .0073 to .013. The larger values are bigger than the nonparametric standard error, making it impossible to construct the Hausman statistic in those cases (and providing further evidence of misspecification). Nevertheless, it is easy to bound the possible values of the Hausman statistic. It can be shown that under general conditions it is possible to construct a standard deviation for an efficient estimator that is less than the standard deviation for the inefficient estimator. The Hausman statistic constructed in this way will be no smaller in absolute value than the difference of the estimators divided by the standard error of the inefficient estimator. In our case this bound is -3.4, which is a large value for a standard normal. Alternatively, the value of the Hausman test at the smaller standard error of .0073 is -5.7, which rejects the null hypothesis of correct specification even more soundly. Thus, this Hausman test of parametric versus nonparametric models provides evidence against the parametric specification.

The difference in the nonparametric and parametric estimates seems too large to be explained away by the downward bias of the nonparametric estimates and upward bias of the parametric estimates that was found in the Monte Carlo results. The size of the bias found in Table 3 is much smaller than that. On the other hand, the differences between parametric and nonparametric estimates are comparable with the biases found in Table 4, where the maximum likelihood specification is incorrect. In Table 4, the maximum likelihood estimator of the shift is slightly over twice the size of the nonparametric estimator, as in the Swedish data. A feature of Table 4 that is not shared by the Swedish data results is the size of the nonparametric estimates. The empirical estimates of the policy shift are much smaller than those of the Monte Carlo. Of course, that is consistent with misspecification of the likelihood in the empirical application.

Table 8.

	M1	STD	T	CV
const., dy, dw	-0.0214	0.0062	-3.45	0.0073
above and w_3, y_3	0.0247	0.0091	2.73	0.0323
above + y_3^2	0.0298	0.0091	3.27	0.0373
above + w_3^2	0.0278	0.0090	3.10	0.0366
above + $w_3 y_3$	0.0278	0.0093	3.00	0.0360
above and ℓ_1, ℓ_2	0.0251	0.0099	2.52	0.0358
above and y_2, w_1, w_2	0.0247	0.0105	2.36	0.0278
above and ℓ_1^2, ℓ_2^2	0.0262	0.0145	1.80	0.0268

8. Conclusion

In this paper we have proposed a nonparametric model and estimator for labor supply with a nonlinear budget set. The estimator is formed in two steps: 1) approximating each budget set by a piece-wise linear set with a few segments; 2) running a nonparametric regression of hours on the parameters of the piecewise linear set. We exploit the additive structure implied by utility maximization by imposing the additivity on the nonparametric regression. This estimator is not based on a likelihood specification, and so is relatively simple to compute and robust to distributional misspecification.

We apply our nonparametric method on Swedish data and use the estimated nonparametric function to predict the effect of recent Swedish tax reform. We compare our method with a parametric maximum likelihood method. The differences between the maximum likelihood and nonparametric estimates provide an example where the flexibility of nonparametric estimation has a substantial impact on the conclusions of empirical work. Here we find that the nonparametric policy prediction is less than half the parametric one, and the difference is statistically significant. The designed flexibility of our nonparametric approach to allowing for nonlinear budget sets lends credence to the idea that the maximum likelihood estimates overstate the size of the effect of Swedish tax reform. More generally, the simplicity of our approach, together with its flexibility, should make it quite useful for sensitivity analysis for maximum likelihood estimation with nonlinear budget sets. A simple, powerful adjunct to, or even replacement of, maximum likelihood estimation would be nonparametric estimation using the approximation to the budget sets that is described here.

References

[1] Ackum Agell, S. and C. Meghir (1995), "Male Labour Supply in Sweden: Are Incentives Important?" *Swedish Economic Policy Review 2*, 391-418.

[2] Agell J., P. Englund and J. Södersten (1995), *Svensk skattepolitik i teori och praktik: 1991 års skattereform.* SOU 1995:104, *bilaga 1.*

[3] Andrews, D.W.K. (1991), "Asymptotic Optimality of Generalized CL, Cross-Validation and Generalized Cross-Validation in Regression with Heteroskedastic Errors," *Journal of Econometrics 47*, 359-377.

[4] Aronsson, T. and M. Palme (1998), "A Decade of Tax and Benefit Reforms in Sweden: Effects on Labour Supply, Welfare and Inequality," *Economica 65*, 39-67.

[5] Blomquist, S. (1983), "The Effect of Income Taxation on the Labor Supply of Married Women in Sweden," *Journal of Public Economics 22*, 169-97.

[6] Blomquist, S. and U. Hansson-Brusewitz (1990), "The Effect of Taxes on Male and Female Labor Supply in Sweden," *Journal of Human Resources 25*, 317-357.

[7] Blundell, R., A. Duncan and C. Meghir (1998), "Estimating Labour Supply Responses Using Tax Reforms," *Econometrica 66,* 827-861.

[8] Eissa, N. (1995), "Taxation and Labor Supply of Married Women: The Tax Reform Act of 1986 as a Natural Experiment," NBER Working Paper #5032.

[9] Fritzell J. and O. Lundberg (1994), *Vardagens villkor: Levnadsförhållanden i Sverige under tre decennier*, Brombergs förlag.

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

[11] Lorentz, G.G. (1986), *Approximation of Functions,* New York: Chelsea Publishing Company.

[12] Hausman, J. and W. Newey (1995) "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica 63*, 1445-1476.

[13] Newey, W.K. (1997), "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics 79*, 147-168.

[14] Stone, C.J. (1982), "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics 10*, 1040-1053.

[15] Stone, C.J. (1985), "Additive Regression and Other Nonparametric Models," *Annals of Statistics 13*, 689-705.

[16] White, H. (1980), "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review 21*, 149-170.

Appendix A. Expected hours of work for a special case

Suppose data are generated by utility maximization subject to a convex budget constraint consisting of three piece-wise linear segments. Suppose further that the basic supply function is linear and that there is an additive random preference term that is uniformly distributed, i.e. the pdf for the random preference term is given by:

$$g(t) = \frac{1}{u} 1 \left(-\frac{u}{2} \leq t \leq \frac{u}{2} \right).$$

The expression for expected hours of work will then take the form:

$$E(h^{*}) = \frac{\ell_1(\pi_1 - \pi_2)}{u} + \frac{\ell_2(\pi_2 - \pi_3)}{u} + \frac{1}{2u}(\pi_3 + u)^2.$$

If we know expected hours of work has this form but we do not know the parameters of the basic supply function, the estimating function would take the form:

$$h = const. + b_1 dy + b_2 dw + b_3 y_3 + b_4 w_3 + b_5 y_3^2 + b_6 w_3^2 + b_7 w_3 y_3,$$

where $dy = \ell_1(y_1 - y_2) + \ell_2(y_2 - y_3)$ and $dw = \ell_1(w_1 - w_2) + \ell_2(w_2 - w_3)$.

Appendix B. Sample statistics.

Hours of work are measured in thousands of hours, virtual income in thousands of SEK and the wage rate in SEK. The marginal wage rates and virtual incomes are calculated at observed hours of work for each individual. The economic variables are expressed in the 1980 price level.

Variable	Mean	Variance
1973		
# of observations: 777		
Hours of work	2.133	0.0656
Marginal wage rate	16.27	19.67
Virtual income	36.34	331.06
1980		
# of observations: 864		
Hours of work	2.098	0.0605
Marginal wage rate	14.90	31.02
Virtual income	69.19	840.48
1990		
# of observations: 680		
Hours of work	2.120	0.1067
Marginal wge rate	19.77	30.27
Virtual income	55.51	399.43
All years combined		
# of observations: 2321		
Hours of work	2.116	0.0760
Marginal wage rate	16.55	27.93
Virtual income	54.18	731.79

Appendix C. The Swedish 1990 income tax and transfer system

Income tax.

Note that the figures below are expressed in the 1990 price level. In our calculations we have deflated all figures to the 1980 price level. We use the following income definitions. *Gross income* refers to the individuals income before tax and deductions. *Assessed income* is defined as the gross income, minus deductions of income related costs. *Taxable income* is defined as assessed income, minus personal allowances. Finally, *Capital income* is the income from rents, dividends, interest etc., minus capital losses and interest payments.

The income related deductions are the registered deficits in income sources, plus a standard deduction of 10% of earned income (at maximum 3000 SEK). The personal allowances equal 10000 SEK and there is a standard capital allowance equal to 1600 SEK.

The standard federal and local taxes are levied on taxable income. The local taxes vary across municipalities, but are in general close to 30%. The standard federal marginal tax rate equals 3% on taxable income between zero and 75000 SEK and 10% on taxable income above 75000 SEK. In addition to the standard federal tax there is an additional federal tax levied on taxable income omitting the deductions relating to deficits in capital income. The additional federal marginal tax rate is 14% on the modified taxable income between 140000 and 190000 SEK and 25% on the income above 190000 SEK.

Housing allowance

Housing allowances are only granted households with children and households where the head is no more than 28 years old. The housing allowance is calculated in two steps. First, the maximum allowance is calculated and second, the allowance is reduced depending on the economic status of the applicant. The maximum allowance is based on the monthly housing costs and the family composition. The monthly housing cost is defined as the monthly rent payments or a calculated standard monthly cost for owned-occupied homes. The monthly housing costs for owner-occupied homes (and tenant-owned flats) are based on the operating costs, the implicit income to the owner, the leasehold right etc. The calculations also account for the tax effects of deficits in capital incomes due to mortgages. Table C.1. presents the lower, middle and upper bounds of the monthly housing costs that serve as base for the calculation of the allowance. The monthly allowance equals 80% of the monthly costs between the lower and middle bound, and 60% of the costs between the middle and upper bound.

Table C.1. Interval bounds of monthly housing costs.

	80%		60%
No. of children	Lower bound	Middle bound	Upper bound
0	700	2400	2400
1	1000	2400	2600
2	1000	2400	2900
3	1000	2600	3200
4	1000	2600	3500
>4	1000	3125	3800

Furthermore, a household with children receives an additional housing allowance according to Table C.2.

Table C.2. Additional annual housing allowance to households with children

No. of children	Additional allowance
1	3180
2	6360
3	9540
4	7920
>4	3180

It should be noted that the decrease in housing allowance for a household with many children is compensated by an increase in child allowance.

The reduction of the allowance is based on the assessed income of the household in 1987. If, however, the applicant's economical status in 1990 differs substantially from the status 1987, then the calculations are based on a modified income definition. In particular, for households with children, if the household earned income in 1990 increased by more than 75000 SEK or decreased by more than 15000 SEK, then the allowance is based on the household estimated assessed income 1990 (minus a deduction of 30000 SEK if there was an increase in earnings). For households without children the increase (or decrease) refers to the difference in assessed income 1987 and estimated assessed income 1990 of the household. Furthermore, an amount of 20% of the household wealth exceeding 180000 SEK is added to the assessed income. The allowance is reduced by 33.3% of the household income above 38000 SEK for households without children and by 20% of the household income above 63000 SEK for households with children. It should be noted that the construction of the housing allowance creates non-convexities as well as non-continuities.

Appendix D. 1991 Income tax system

The local income tax was roughly as in 1980. In the federal income tax schedule there was a basic standard deduction of SEK 10,000. For taxable income up to

SEK 180,000 the federal tax was zero. For taxable income above 180,000 the federal tax rate was 20%. Denoting labor income by x, taking account of the standard deduction and deflating to the 1980 price level gives the tax schedule.

x	**Marginal tax**
-77661	0
77661-	0.20

Between 1980 and 1991 there was also a base broadening for the VAT and an increase of the VAT rate from 21.34% to 25%.[12] In credue terms assuming the increase in the VAT tax is completely rolled over onto consumers, the combined effect of the base broadening and increase in the VAT tax rate is equivalent to an increase in proportional income tax with four percentage points. There was also a change in payroll taxes from a rate of 35.25% in 1980 to 37.4% in 1991. The rates are in terms of income net of the payroll tax. Expressed as a percentage of gross labor income the percentages are 26.06% and 27.26% respectively. In Sweden there is a discussion of whether the payroll taxes should be fully regarded as taxes or if some part should be treated as a fee for insurance. Here we treat the payroll taxes as taxes. In crude terms the change in payroll taxes between 1980 and 1991 is equivalent to an increase in a proportional income tax with 1.2 percentage points. The combined effects of the change in VAT and payroll taxes is hence equivalent to an increase of a proporational income tax with 5 percentage points. We treat the changes in the VAT and the payroll tax in a simplified way and represent the changes as an increase by five percentage points in a proporational income tax. We then obtain the following tax schedule.

Tax schedule including the effect of increased VAT and payroll taxes.

x	**Marginal tax**
-77661	0.05
77661-	0.25

[12]There was a change of the VAT rate in 1980. 21.34% is a weighted average for thge year.

Appendix E: Proofs of Theorems

Proof of Lemma 1: Let

$$\begin{aligned} f(\ell) &= \ell - \pi(B(\ell) - \ell B_\ell(\ell), B_\ell(\ell)), \\ \tilde{f}^+(\ell) &= \ell - \pi(\tilde{B}(\ell) - \ell \tilde{B}_\ell^+(\ell), \tilde{B}_\ell^+(\ell)), \\ \tilde{f}^-(\ell) &= \ell - \pi(\tilde{B}(\ell) - \ell \tilde{B}_\ell^-(\ell), \tilde{B}_\ell^-(\ell)). \end{aligned}$$

By the chain rule, $f(\ell)$ is differentiable and

$$f_\ell(\ell) = 1 - [\pi_y(B(\ell) - \ell B_\ell(\ell), B_\ell(\ell))\,(-\ell) + \pi_w(B(\ell) - B_\ell(\ell), B_\ell(\ell))]\, B_{\ell\ell}(\ell).$$

By the Slutzky equation, the term in square brackets is equal to the derivative of the Hicksian (utility-constant) labor supply, and hence is positive. Since $B(\ell)$ is concave, $B_{\ell\ell}(\ell) \leq 0$ so that $f_\ell(\ell) \geq 1$. For $\tilde{\ell} = \ell(\tilde{B}, v)$, the mean-value theorem gives $f(\ell) = f(\tilde{\ell}) + f_\ell(\bar{\ell})(\ell - \tilde{\ell})$, so equation (7) gives

$$\tilde{f}^+(\tilde{\ell}) \leq f(\tilde{\ell}) + f_\ell(\bar{\ell})(\ell - \tilde{\ell}) \leq \tilde{f}^-(\tilde{\ell}).$$

It then follows by subtracting $f(\tilde{\ell})$ from both sides, by $|f_\ell(\ell)| \geq 1$, by taking absolute values, and by $\pi(y, w)$ Lipschitz, that

$$\left|\ell - \tilde{\ell}\right| \leq \frac{1}{\left|f_\ell(\bar{\ell})\right|} \max\left\{\left|\tilde{f}^+(\tilde{\ell}) - f(\tilde{\ell})\right|, \left|\tilde{f}^-(\tilde{\ell}) - f(\tilde{\ell})\right|\right\} \leq C\|\tilde{B} - B\|. \qquad Q.E.D.$$

Proof of Lemma 2: By equation (1) and s-times continuous differentiability of $J(v)$ and $\pi(y, w)$, the derivatives of the additive components $h(B^L(B))$ with respect to x^L are bounded *uniformly in L*. It follows by Theorem 8 of Lorentz (1986) that each component has an approximation error $C\left(\frac{K}{L}\right)^{-s/3}$. Summing these gives an approximation error of $CL\left(\frac{K}{L}\right)^{-s/3}$. Furthermore, $\left|h(B^L(B)) - h(B)\right| \leq C\|B^L(B) - B\|$. Select $B^L(B)$ so $\|B^L(B) - B\| \leq \frac{1}{L}$, using spline approximation results. The triangle inequality then gives the result. *Q.E.D.*

Proof of Theorem 3: Choose x_L^B satisfying the conclusions of Lemma 2 and let x_i be x_L^B for the ith individual. Let $p_i = p^K(x_i), P = [p_1, ..., p_n], \hat{h} = (\hat{h}_1, ..., \hat{h}_n)'$, and $\bar{h} = (\bar{h}_1, ..., \bar{h}_n)'$. Then $\hat{h} = Qy$ for $Q = P(P'P)^- P'$. Note that $\hat{h} - \bar{h} = Q(y - \bar{h}) - (I - Q)\bar{h} = Q\varepsilon - (I - Q)\bar{h}$ for $\varepsilon = y - \bar{h}$, so that $\sum_{i=1}^n (\hat{h}_i - \bar{h}_i)^2/n = (\hat{h} - \bar{h})'(\hat{h} - \bar{h})/n = \varepsilon' Q\varepsilon/n + \bar{h}'(I - Q)\bar{h}/n$. By the conditional variance of ε bounded,

$E[\varepsilon' Q\varepsilon | B_1, ..., B_n] = tr(QE[\varepsilon\varepsilon' | B_1, ..., B_n]) \leq CK$, so that $\varepsilon' Q\varepsilon/n = O_p(K/n)$. Also, by Lemma 2 and $I - Q$ idempotent,

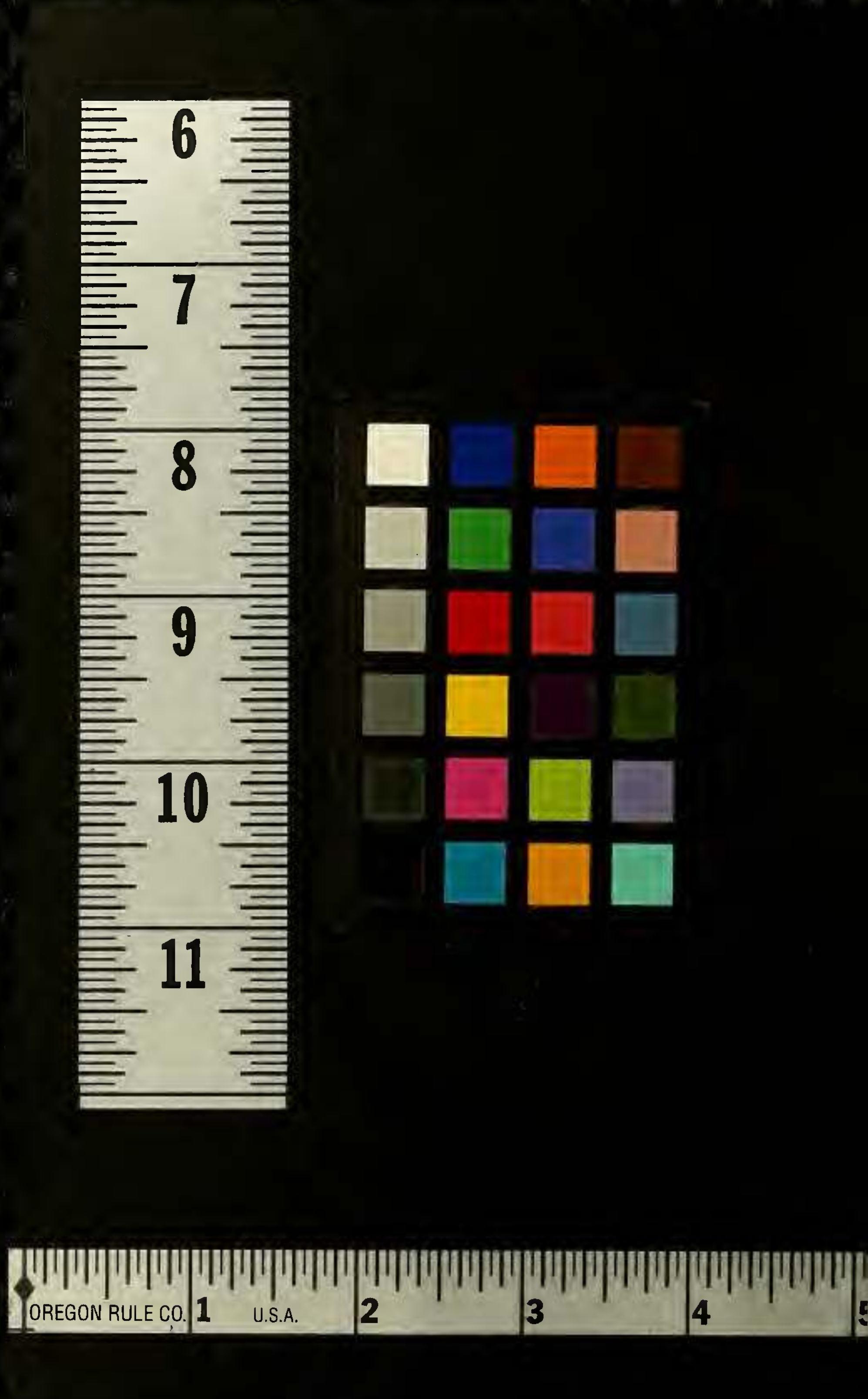
$$\begin{aligned} \bar{h}'(I-Q)\bar{h}/n &= (\bar{h} - P\beta_L)'(I-Q)(\bar{h} - P\beta_L)/n \leq (\bar{h} - P\beta_L)'(\bar{h} - P\beta_L)/n \\ &= O(\frac{1}{L^2} + L^2(\frac{K}{L})^{-2s/4}). \end{aligned}$$

The conclusion then follows by the triangle inequality. *Q.E.D.*

Proof of Theorem 4: Assumptions 2, 3, and 4 correspond to Assumptions 9, 1, and 8 respectively of Newey (1997). The conclusion of Theorem 4 of Newey (1997) for r = 4 then gives the conclusion.

Proof of Theorem 5: Assumptions 2, 3, 4, 5, and 6 correspond to Assumptions 9, 1, 8, 4, and 6 of Newey (1997) The conclusion of Theorem 5 of Newey (1997) for r = 4 then gives the conclusion.

Proof of Theorem 6: Assumptions 2, 3, 4, 5, and 7 correspond to Assumptions 9, 1, 8, 4, and 7 of Newey (1997) The conclusion of Theorem 6 of Newey (1997) for r = 4 then gives the conclusion.


6
7
8
9
10
11
OREGON RULE CO. 1 U.S.A. 2 3 4 5