1997-09-04

NONPARAMETRIC ESTIMATION OF LABOR SUPPLY FUNCTIONS GENERATED BY PIECE WISE LINEAR BUDGET CONSTRAINTS

by

Sören Blomquist* and Whitney Newey**

ABSTRACT

The basic idea in this paper is that labor supply can be viewed as a function of the entire budget set, so that one way to account non-parametrically for a nonlinear budget set is to estimate a nonparametric regression where the variable in the regression is the budget set. In the special case of a linear budget constraint, this estimator would be the same as nonparametric regression on wage and nonlabor income. Nonlinear budget sets will in general be charac-terized by many variables. An important part of the estimation method is a procedure to reduce the dimensionality of the regression problem. It is of interest to see if nonparametrically estimated labor supply functions support the result of earlier studies using parametric methods. We therefore apply parametric and nonparametric labor supply functions to calculate the effect of recent Swedish tax reform. Qualitatively the nonparametric and parametric labor supply functions give the same results. Recent tax reform in Sweden has increased labor supply by a small but economically important amount.

Keywords: Nonparametric estimation, Labor supply, Nonlinear budget constraints, Tax Reform.
JEL Classification: C14, J22, H24.

Financial support from the Bank of Sweden Tercentenary Foundation is gratefully acknowl-edged. We are grateful to Matias Eklöf for competent assistance.
* Department of Economics, Uppsala University, Box 513, SE-751 20 Uppsala, phone +46-18-471 11 02, fax +46 18 471 14 78, e-mail soren.blomquist@nek.uu.se.
** Massachusetts Institute of Technology.

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 one way to account nonparametrically for a nonlinear budget set is to estimate 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. Thus, nonparametric regression on these characterizing variables 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. Thus, this more parsimonious approach should help us capture the features of the choice variable that are identified from the data.

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 the first part of section 4 and small sample properties, based on Monte Carlo simulations, in the latter part. In section 5 we apply the method to Swedish data. We use estimated labor supply functions to calculate the effect of income tax reform in section 6. Section 7 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

$$E\left[h_i \middle| B_i\right] = \bar{h}(B_i).$$

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 tradeoff between robustness and efficiency. In particular, most models in the literature have a labor supply function of the form

$$h_i = h(B_i,v_i) + e_i ,$$

where v_i represents individual heterogeneity and e_i is measurement error. The typical maximum likelihood specification relies on an assumption that v_i and e_i are normal and homoskedastic, while all that we would require is that v_i is independent of B_i and $E\left[e_i \middle| B_i\right] = 0$, in which case $\bar{h}(B_i) = \int h(B_i, v)F(dv)$. 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, as a starting point we will 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.

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 w_i and intercepts y_i of the three segments. These segments also define two kink points. The kink points are related to the slopes and intercepts as: $\ell_1 = \left(y_2 - y_1\right)/\left(w_2 - w_1\right)$ and $\ell_2 = \left(y_3 - y_2\right)/\left(w_3 - w_2\right)$.

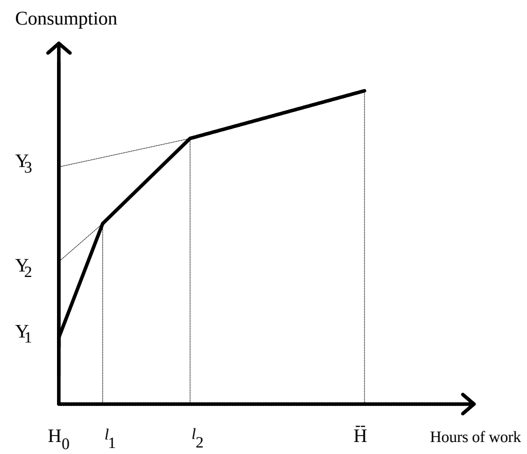

Figure 1.

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 $h_j^* = p(y_j, w_j) + v$, where n 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 $p() + v$ is generated by utility maximization with globally convex preferences. Then desired hours will equal zero if

$p_1+v\le 0$. Desired hours will fall on the first segment if $0\le p_1+v\le \ell_1$ and be located at kinkpoint ℓ_1 if $p(y_1,w_1)+v\ge \ell_1$, and $p(y_2,w_2)+v\le \ell_1$ i.e. if $\ell_1-p(y_1,w_1)\le v\le \ell_1-p(y_2,w_2)$. Desired hours will be on the second segment if $\ell_1<p(y_2,w_2)+v<\ell_2$, etc. This implies that we can write expected hours of work as:

$$
\begin{aligned}
E(h^*) &= 0\cdot G(-p_1)\\
&+ \underbrace{\left[G(\ell_1-p_1)-G(-p_1)\right]}_{\text{probability that h* is on first segment}} \times \left\{p_1+E(v)\,|\,-p_1\le v\le \ell_1-p_1\right\}\\
&+\ell_1\cdot \underbrace{\left[G(\ell_1-p_2)-G(\ell_1-p_1)\right]}_{\text{probability that desired hours are at kinkpoint } \ell_1.} +\\
&+ \underbrace{\left[G(\ell_2-p_2)-G(\ell_1-p_2)\right]}_{\text{probability that h* is on the second segment}}\cdot \left\{p_2+E(v)\,|\,\ell_1-p_2\le v\le \ell_2-p_2\right\}\\
&+ \ell_2\left[G(\ell_2-p_3)-G(\ell_2-p_2)\right]\\
&+ \underbrace{\left[1-G(\ell_2-p_3)\right]}_{\text{probability that desired hours are on third segment}} \times\left\{p_3+E(v)\,|\,v>\ell_2-p_3\right\}
\end{aligned}
\qquad (1')
$$

Wee see from this expression that $E(h^*)$ is a continuous, differentiable function in ℓ_1, p_1, ℓ_2, p_2, ℓ_3,p_3.[1] Since p_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:

$$E(h^*)=-J(-p_1)+\sum_{k=1}^{2}\left[J(\ell_k-p_k)-J(\ell_k-p_{k+1})\right]+p_3 \qquad (1)$$

This expression generalizes straightforwardly for the case with more segments. The particular form of (1) follows from the assumption that hours of work are generated by utility maximization with globally convex preferences. For particular c.d.f:s of v we can derive properties of the $J(v)$ function. For example, if v is uniformly distributed

[1] Expression (1') is derived under the assumption that there is no upper limit $\overline{H}$ for hours of work. If we introduce an upper limit $\overline{H}$ for hours of work, we would get one more term, and the last term would be slightly different. If $\overline{H}$ 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 $\overline{H}$ included.

$J(v)$ will be quadratic. Independent of the form of the c.d.f. for v, $J(v)$ will always be concave and lie below it's 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 p_k are equal and equation (1) simplifies to $p - J(-p)$. Also, if the probability of no work is zero then the hours equation becomes p. 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 p.

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

$$E(h^*) = g_1(\ell_1,w_1,y_1) + g_2(\ell_1,w_2,y_2) + g_3(\ell_2,w_2,y_2) + g_4(\ell_2,w_3,y_3) \tag{2}$$

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:

$$E(h^*) = g_1(\ell_1,w_1,y_1) + g_2(\ell_1,\ell_2,w_2,y_2) + g_3(\ell_2, w_3, y_3) \tag{3}$$

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

$$E(h^*) = f_1(y_1,w_1,y_2,w_2) + f_2(y_2,w_2, y_3,w_3) \tag{4}$$

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: $h = h^* + e$, where $E(e|x,h) = 0$. 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.

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 $E(h|w,y) = g(w,y)$. 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 eq. (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 con-straints consisting of up to 27 segments were used. To describe such a budget con-straint we need 54 variables! Nonparametric estimation using actual budget con-straints 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, say, only 5-6 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 approxi-mation 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 $C(h_j)$ consumption on the approximating budget constraint. The criterion to choose the approximating budget constraint is $Min \sum_j \left[C(h_j) - C(h_j)\right]^2$.

ii. Interpolation method

Take three values for hours of work: h_1, h_2 and h_3. Let $w(h^1{}_j)$, be the slope of the true budget constraint at $h^1{}_j$. 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 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 $p^K(x) = (p_{1K}(x),...,p_{KK})'$ 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

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.

$$g(x) = p^K(x)'b, \quad b = (P'P)^- P'H, \tag{5}$$

where B^- denotes any symmetric generalized inverse. Under conditions given below, $P'P$ will be nonsingular with probability approaching one, and hence $(P'P)^-$ will be the standard 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 parsimonius 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 the appendix,

$$\bar{h}(B) = \left[[\ell_1(p_1 - p_2) + \ell_2(p_2 - p_3)]/u\right] + (p_3 + u)^2/(2u).$$

Also suppose that $p(y,w) = g + gy + gw$. Then for $dy = \ell_1(y_1 - y_2) + \ell_2(y_2 - y_3)$ and dw $= \ell_1(w_1 - w_2) + \ell_2(w_2 - w_3)$,

$$\bar{h}(B) = b_1 + b_2 dy + b_3 dw + b_4 y_3 + b_5 w_3 + b_6 y_3^2 + b_7 w_3^2 + b_8 y_3 w_3, \tag{6}$$

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

$$b_1 = c^2/2u, \quad b_2 = g_2/u, \; b_3 = g_3/u, \; b_4 = c\,g_2/u, \; b_5 / c\,g_3/u$$

$$b_6 = (\gamma_2)^2/2u\,,\;\; b_7 = (\gamma_3)^2/2u\,,\;\; b_8 = \gamma_2\,\gamma_3/u\;.$$

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

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

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**

4.1 Asymptotic theory

As previously noted, utility maximization with convex, piecewise linear budget constraints leads to expected hours being additive in virtual wages and income. In this section we present asymptotic theory for a series estimator of one of these additive specifications, that of equation (4). We are mindful that piecewise linear budget constraints may only be an approximation. Here we do not take explicit account of this approximation error, because of the depth of this topic. We leave this task to future work.

Generalizing equation (4) to allow for J budget segments leads to

$$E(h^*) = \sum_{j=1}^{J-1} f_j(y_j, w_j, y_{j+1}, w_{j+1}). \tag{7}$$

Newey (1995) has developed theory for series estimators of additive models that can be applied here to obtain convergence rates and asymptotic normality results. The following assumptions list the regularity conditions that lead to this result:

Assumption 1: $(h_1,x_1),..., (h_n,x_n)$ are i.i.d. and $Var(h|x)$ is bounded.

The bounded conditional variance assumption is difficult to relax without affecting the convergence rates.

Assumption 2: 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.

Assumption 3: $g_0(x) = E[h|x]$ is continuously differentiable of order s on the support of x.

This condition specifies that the expected hours function is smooth.

These conditions and a limit on the growth rate of the number of terms K leads to the following convergence rates. Let χ be the support of x, and $F_0(x)$ its distribution function.

Theorem 1: *If Assumptions 1, 2, and 3 are satisfied and* K^3/n ® 0 *then*

$$\int [g_0(x) - g(x)]^2 dF_0(x) = O_p\left(K/n + K^{-s/2}\right), \tag{8}$$
$$\sup_{x \in c} |g(x) - g_0(x)| = O_p\left(K\left[\sqrt{K}/\sqrt{n} + K^{-s/4}\right]\right).$$

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, with the variance being increasing in K and the bias decreasing. 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.

The asymptotic theory also leads to approximate inference methods. 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

$$q = a(g). \tag{9}$$

This estimator can be combined with a consistent standard error for inference. Let $A = \left(a(p_{1K}),\ldots,a(p_{KK})\right)'$ and

$$V = A'Q^{-}\Sigma Q^{-}A, \quad Q = P'P/n, \quad \Sigma = \sum_{i=I}^{n} p^K(x_i)p^K(x_i)'\left[h_i - g(x_i)\right]^2/n. \qquad (10)$$

This estimator is just the usual one for a function of least squares coefficients, with $Q^{-}\Sigma Q^{-}$ 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 4: $E[\{h-g_0(x)\}^4|x]$ is bounded, and $Var(h|x)$ is bounded away from zero.

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

Assumption 5: $a(g)$ is a scalar, there exists C such that $|a(g)| < C\sup_{\chi}|g(x)|$, and there exists $g_K(x) = p^K(x)'\tilde{b}_K$ 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 will imply that the estimator q is not $\sqrt{n}$-consistent, and is also a useful regularity condition. 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 $\sigma^2(x) = Var(h|x)$. The asymptotic variance formula is

$$V_K = A'Q^{-1}\Sigma Q^{-1}A, \;\; Q = E[p^K(x)p^K(x)'], \;\; \Sigma = E[p^K(x)p^K(x)'\sigma(x)^2]. \quad (11)$$

Theorem 2: *If Assumptions 1-5 are satisfied,* $K^3/n \to 0$, *and* $\sqrt{n}K^{-s/4} \to 0$ *then* $q = q_0 + O_p(K^{3/2}/\sqrt{n})$ *and*

$$\sqrt{n}V_K^{-1/2}(q - q_0) \xrightarrow{d} N(0,1), \quad \sqrt{n}V^{-1/2}(q - q_0) \xrightarrow{d} N(0,1),$$

There are also cases where q is $\sqrt{n}$-consistent, that are useful to consider separately. Under the following condition this will occur.

Assumption 6: There is $\nu(x)$ with $E[\nu(x)\nu(x)']$ finite and nonsingular such that $a(g_0) = E[\nu(x)g_0(x)]$, $a(p_{kK}) = E\left[n(x)p_{kK}(x)\right]$ for all k and K, and there is $\tilde{b}_K$ with $E\left[\left\|n(x) - \tilde{b}_K p^K(x)\right\|^2\right] \to 0.$

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 $\nu(x)$ in the outer product representation to be approximated in mean-square by some linear combination of the functions. This condition and Assumption 5 are mutually exclusive, and together cover most cases of interest (i.e. they seem to be exhaustive).

A sufficient condition for Assumption 6 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 6 will then follow from the Riesz representation theorem. The asymptotic variance of the estimator will be determined by the function $\nu(x)$ from Assumption 6. It will be equal to

$$V = E[\nu(x)\nu(x)'Var(h|x)]. \quad (12)$$

Theorem 3: *If Assumptions 1-4 and 6 are satisfied,* $K^3/n \rightarrow 0$, *and* $\sqrt{n}K^{-s/4} \rightarrow 0$ *then*

$$\sqrt{n}(q - q_0) \xrightarrow{d} N(0,V), \quad V \xrightarrow{p} V . \tag{13}$$

4.2 Small sample properties

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 to 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 R^2 and 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 OLS and the interpolation methods where performance is measured by the cross-validation criteria. For the OLS 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 OLS method is more robust across data from different years. We have not found a specification of the OLS method that is uniformly best across data from different years. However, the OLS 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 nonlabor 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: $h^* = 1.857 + n + 0.0179w - 3.981 * 10^{-4} y$
$+4.297 * 10^{-3} AGE + 2.477 * 10^{-3} NC$, where $n \sim$ N(0, 0.0673), hours of work is 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 $h = h^* + e$ where e ~ N(0, 0.0132).

We use the following four types of 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.
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 depend 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

[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.

both random preferences and measurement error function 2 performs slightly better than function 1.

Table 1. Evaluation of Estimation Method using constructed "actual" budget constraints. Coefficient of determination and Cross validation used as performance measure. Averages over 500 replications.

DGP		function 1	function 2	function 3	function 4	function 5
No random terms	Average R^2	0.601	0.604	0.644	0.658	0.450
	Average CV	0.581	0.576	0.556	0.536	0.392
Measurement error	Average R^2	0.215	0.218	0.245	0.252	0.163
	Average CV	0.194	0.190	0.136	0.123	0.128
Random preferences	Average R^2	0.125	0.137	0.167	0.184	0.083
	Average CV	0.103	0.106	0.010	0.013	0.052
Random pref +meas. error	Average R^2	0.098	0.107	0.135	0.149	0.066
	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 pro-cedure 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^2:s and CV:s in table 2 in general are higher than those in table 1. This is especially so for the case when there is 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 estimation. Coefficient of determination and Cross validation used as performance measure. Averages over 500 replications.

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

Why are the R^2:s and CV:s higher in table 2 than in table 1, especially when there is 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 n. 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^2 :s and CV:s. However, while in table 1 the inclusion of random preferences reduced the R^2 :s and CV:s most, in table 2 it is the inclusion of measurement error that decreases the R^2 :s and CV:s most. Looking across columns and approximating functions we find that the coefficient of determination increase as we include more squares and interactions while the cross validation decrease. 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 p. 18 above.) This model contains both random preferences and measurement errors. Thus, the datagenerating 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 1990 tax system as described in section 6 and appendix B 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.

$H_{BTR} =$ predicted before tax reform average hours of work.

$H_{ATR} =$ 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

$$M1 = (H_{ATR} - H_{BTR}) / H_{BTR} .$$
$$M2 = (H_{ATR} - H_{BTR}) / H_{BTR} .$$

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 c is given by $c = e^2 / 5$. 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	

5. Estimation on Swedish data

5.1 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]

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 self employed are excluded. This leaves us with 777 observations for 1973, 864 for 1980 and 680 for 1990.

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

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 A. 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 increase. Another important advantage is that we obain 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.

5.2 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 eq. (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] [6]

$$\underset{(62.09)}{h = 1.914} + \underset{(8.96)}{0.0157w} - \underset{(-5.95)}{8.65*10^{-4}\,y} - \underset{(-0.53)}{9.96*10^{-3}\,AGE} - \underset{(-0.44)}{3.46*10^{-3}\,NC} \qquad (14)$$

---

[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] Net wage rates and virtual income are expressed in the 1980 price level for all years. The wage and income elasticities are evaluated at the average net wage rate and virtual income. The net wage rate and virtual income being calculated for the segment where observed hours are located.

$$\ln L = -225.43 \qquad s_h = 0.270 \qquad s_e = 0.105 \qquad E_w = 0.123 \qquad E_y = -0.022$$
$$\qquad\qquad\qquad\qquad (42.12) \qquad\quad (11.81) \qquad\quad (8.96) \qquad\quad (-5.95)$$

5.3 Nonparametric estimates

Below we report results when we have pooled data for the three years.[7] We use a series estimator. As our criterion to choose the estimating function we use the cross validation measure presented on p. 11. 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 1973 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.[8] We have also used the data to test the utility maximization hypothesis. This

---

[7] 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.

[8] 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

test was performed by estimating a function allowing for interactions between the regressors that violates the separability properties implied by utility maximization. (See the discussion on p. 6.) These interaction terms were not significant. Hence, the data are consistent with utility maximization.

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 $w_3, \quad y_3$	0.0323	0.0291
above + y_3^2	0.0373	0.0350
above + w_3^2	0.0366	0.0341
above+ $w_3 y_3$	0.0360	0.0340
above and ℓ_1, ℓ_2	0.0358	0.0336
above and $y_2, \quad w_1, \quad w_2$	0.0278	0.0310
above and $\ell_1^2, \quad \ell_2^2$	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

between tables 1and 2 vs. tables 3 and 4. However, the DGP:s used for tables 1 and 2 vs. tables 3 and 4 are different.

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 dy and dw 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 elasticiteis. 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 indiviuals 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 nonlabor 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.[9] 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

[9] 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.

labor supply functions for discrete changes in nonlinear budget constraints. In section 6 we use the estimated functions to predict the effect on hours of work of Swedish tax reform.

Table 6. Nonparametric estimates using pooled data.

Variables	**Best function**	**Next best function**
Const.	2.064 (49.85)	2.097 (39.69)
dy	-0.00210 (-4.37)	-0.00204 (-4.28)
dw	-0.00145 (-1.16)	-0.00131 (-1.06)
y_3	-0.0036 (-3.95)	-0.0037 (-4.01)
w_3	0.00964 (6.61)	0.00560 (1.40)
y_3^2	$1.98\text{x}10^{-5}$ (3.40)	$2.00\text{x}10^{-5}$ (3.42)
w_3^2		$1.16\text{x}10^{-4}$ (1.01)
wage elasticity	0.075 (6.61)	0.074 (6.60)
income elasticity	-0.038 (-4.31)	-0.040 (-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 $p(y, w)$ when we impose the functional form for the conditional mean implied by utility maximization.

The estimates are obtained by estimating equation (1) given an assumption on the distribution of v. We recover $p(\cdot)$ from the relation $E(h^*) = p - J(p)$, which shows expected hours of work if data are generated by a linear budget constraint.

Table 7. All years. Estimates obtained when constraints implied by utility maximization are imposed.

Variables	n **uniformly distributed** p **linear**	n **uniformly distributed** p **linear**	n **normally distributed** p **linear**
Constant	0.999	0.986	0.942
	(36.28)	(49.29)	
y	-0.000529	-0.000517	-0.00053
	(3.34)	(3.30)	
w	0.0030	0.0028	0.0009
	(2.87)	(2.71)	
U/2	0.5660	0.5219	
	(2.56)	(2.50)	
# of children		0.0021	
		(0.69)	
Age		0.00036	
		(1.13)	
CV	0.0286	0.0273	0.0009

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.

6. Tax reform

In this section we use the estimated functions to predict the effect of recent changes in the Swedish income tax. 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 nonlabor 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 B.

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 is progressive and has undergone substantial change. The change in 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%

[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.

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 has also been changes in the VAT and the pay roll 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 pay roll tax as a change in the real wage rate. However, we have chosen to represent it as a change in a proportional income tax rate. In appendix B 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 eq. (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 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. This 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

7. Conclusion

[11] The averages are taken over ten simulations with different drawings of the random preference terms in each simulation.

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 piecewise 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 specifi-cation, and so is relatively simple to compute and robust to distributional misspecifi-cation.

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. 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

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

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

Andrews, D.W.K. (1991), "Asymptotic Optimality of Generalized C_L, Cross-Validation and Generalized Cross-Validation in Regression with Hetero-skedastic Errors," *Journal of Econometrics* 47, pp. 359-377.

Aronsson, T. and M. Palme (1995), "A Decade of Tax and Benefit Reforms in Sweden -- Effects on Labour Supply, Welfare and Inequality", Tax Reform Evaluation Report No. 18, Konjunktur Institutet.

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

Blomquist, S. (1991), "Flexible Functional Forms and Coherency for Labor Supply Models". Working Paper 1991:13. Department of Economics, Uppsala University.

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.

Blundell, R. , A. Duncan and C. Meghir (1995), "Estimating Labour Supply Responses Using Tax Reforms", Institute for Fiscal Studies. Mimeo.

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

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

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

Newey, W.K. (1995), "Convergence rates and asymptotic normality for series estimators". Working Paper, Department of Economics, MIT.

Stone, C.J. (1982), "Optimal global rates of convergence for nonparametric regression". *Annals of Statistics* 10, 1040-1053.

White, H. (1980), "Using least squares to approximate unknown regression functions". *International Economic Review* 21, 149-170.

Appendix A. 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 observation: 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 wage 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 B 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 10000. For taxable income up to SEK 180000 the federal tax was zero. For taxable income above 180000 the federal tax rate was 20%. Denoting labor income by x, taking account of the standard deduction and deflating to the 1980 price level this 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 crude 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 a proportional income tax with four percentage points. There was also a change in pay roll taxes from a rate of 35.25% in 1980 to 37.47% in 1991. The rates are in terms of income net of the pay roll 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 pay roll taxes should be fully regarded as taxes or if some part should be treated as a fee for insurance. Here we treat the pay roll taxes as taxes. In crude terms the change in pay roll 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 proportional income tax with 5 percentage points. We treat the changes in the VAT and the pay roll tax in a simplified way and represent the changes as an increase by five percentage points in a proportional 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 the year.

Appendix C. 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:

$g(t) = \frac{1}{u} 1(-\frac{u}{2} \le t \le \frac{u}{2})$. The expression for expected hours of work will then take the form:

$E(h^*) = \frac{\ell_1(p_1 - p_2)}{u} + \frac{\ell_2(p_2 - p_3)}{u} + \frac{1}{2u}(p_3 + u)^2$. 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 $dw = \ell_1(w_1 - w_2) + \ell_2(w_2 - w_3)$.