Estimation of a censored regression panel data model using conditional moment restrictions efficiently

Charlier, G.W.P.; Melenberg, B.; van Soest, A.H.O.

Publication date:
1995

Estimation of a Censored Regression Panel Data Model Using Conditional Moment Restrictions Efficiently

by

Erwin Charlier, Bertrand Melenberg and Arthur van Soest,[1]

Tilburg University
Department of Econometrics and CentER
P.O. Box 90153
5000 LE, Tilburg
The Netherlands

September 1995

Abstract

A new semiparametric estimator for the censored regression panel data model with fixed effects is introduced. It is based upon an estimator proposed by Honoré for the case of two time periods combined with ideas of Newey to improve the efficiency. The estimator is more efficient than Honoré's and is generalized to the case of a balanced or unbalanced panel of more than two waves. Estimation is performed in two steps. Using Honoré's estimator in a first step, efficient GMM using conditional moment restrictions is applied. The performance of this estimator is compared to that of Honoré's and other existing estimators in an empirical example concerning labour income of married females, using panel data from the Dutch Socio-Economic Panel, 1984-1988. Attention is paid to specification testing and the sensitivity of the results for the choice of smoothness parameters.

Keywords : panel data, censored regression, semiparametric efficiency, unbalanced panel

[1] We thank Bo Honoré, Marno Verbeek, Myoung-Jae Lee and Gerard Pfann for helpful comments. All remaining errors are ours. Research of the third author is possible due to a fellowship of the Netherlands Royal Academy of Arts and Sciences (KNAW). We are grateful to Statistics Netherlands (CBS) for providing the data.

1. Introduction

For a censored regression model with individual fixed effects and two time periods, Honoré (1992) derived two semiparametric estimators based on two conditional moment restrictions (CMRs). From the CMRs, unconditional moment restrictions (UMRs) are constructed such that the estimators can be obtained from minimization of a strictly convex objective function. The UMRs used by Honoré (1992) do not lead to a semiparametrically efficient estimator: as shown by Honoré (1993), the estimator does not attain the semiparametric efficiency bound for the model at hand. A more efficient estimator can be obtained by using an optimal set of UMRs based upon the given CMR, along the lines of Newey (1993). This is a two step estimator, using Honoré's estimator as the first step.

This type of improvement is analyzed in this paper. Starting point is the smooth CMR underlying one of the Honoré (1992) estimators. First, we consider the efficient Generalized Method of Moments (GMM) estimator based on this CMR for the case of two panel waves. Second, we generalize this estimator to the case of more than two panel waves, allowing for balanced as well as unbalanced panels. Applications of the Honoré (1992) estimator are still scarce. Some examples are Udry (1995a,b) and Alderman et al. (1995). In an empirical example, we compare the performance of our GMM estimator with that of Honoré's estimator and with some parametric estimators.

The paper is structured as follows. In section 2, we introduce the model and discuss the merits and drawbacks of a number of existing parametric and semiparametric estimators. Section 3 explains in detail how the efficiency of the Honoré (1992) estimator can be improved upon, and describes the limit distribution of the resulting two step GMM estimator for the case of two time periods. In section 4, we compare the performance of Honoré's estimator and our GMM estimator in a small Monte Carlo experiment. Section 5 deals with the empirical application for two time periods. We explain weekly earnings of Dutch married females, using data drawn from the Dutch Socio-Economic Panel (SEP). Stoker (1992) uses earnings of married females as the prototype example of a censored regression model in a cross-section framework. Since earnings partly reflect female labour supply, it is natural in a life cycle context to add fixed effects (cf. Heckman and MaCurdy, 1980). Parametric estimates are compared to the (semiparametric) estimates proposed by Honoré (1992) and our (asymptotically) more efficient estimates. In section 6, the GMM-estimator is extended to panel data with more than two waves. We propose estimators for both balanced and unbalanced panel data. We analyze the performance of the estimator for the same empirical example, using five waves of SEP data, from 1984 to 1988. In Section 7, we interpret the economic results. In section 8, some conclusions are drawn.

2. Model and Existing Estimators

The censored regression model for panel data with individual effects is defined as follows.

$$\begin{cases} y_{it}^* = \alpha_i + \beta' x_{it} + u_{it} & i = 1,..,N;\ t = 1,..,T \\ y_{it} = \max\{0, y_{it}^*\} \end{cases} \qquad \textbf{(1)}$$

Here i denotes the individual and t denotes the time period. y_{it} is the endogenous variable to be explained, x_{it} is a vector of covariates, α_i is the individual effect, u_{it} is an error term, y_{it}^* is an underlying latent variable, and β is the unknown parameter vector of interest. We only observe (y_{it},x_{it}). We are interested in asymptotic results for fixed T, while N tends to infinity. In all cases we assume independence across individuals, but not necessarily over time. We discuss a number of models with varying sets of assumptions and corresponding estimators.

In models with random effects, α_i is assumed to be independent of $x_i=(x_{i1}',...,x_{iT}')'$. Various parametric models with random effects and corresponding estimators for β have been proposed. $\alpha_i \sim N(0,\sigma_\alpha^2)$ and $u_i=(u_{i1},...,u_{iT})' \sim N(0,\Sigma)$, with $\Sigma=\sigma_u^2 I_T$ and I_T the T-dimensional identity matrix, yields the specification of equi-correlation, see Heckman and Willis (1976). Here Maximum Likelihood (ML) can be applied (a one dimensional integral has to be computed). If milder restrictions on Σ are imposed, estimators can be based on numerical integration over $T-1$ dimensions or on simulation (e.g. simulated ML or simulated moments; see Gourieroux and Monfort, 1993).

If the underlying distributional assumptions are satisfied, these random effects approaches lead to consistent asymptotically normal estimators for β. Drawbacks, however, are the assumptions of normality and independence between (u_i,α_i) and x_i. Due to the nonlinear nature of the models (censoring), violation of either of these may lead to inconsistency of ML (see Arabmazar and Schmidt, 1981, 1982).

Parametric models with fixed effects can be divided into two categories. In the first, no restrictions on the distribution of α_i conditional on x_i are imposed. The α_i are then usually considered as nuisance parameters which can be estimated. In the second category, some restrictions on the distribution of α_i are imposed allowing for dependence between α_i and x_i.[2]

In models of the first category, it is usually assumed that u_{it}, i=1,..,N, t=1,..,T, are i.i.d. and independent of x_i. Since the α_i are parameters to be estimated, models in this category suffer from the incidental parameter problem, see Neyman and Scott (1948). In the censored regression model, unlike the binary choice model, the incidental parameter problem cannot be solved by conditional

[2] Some studies only refer to the first category as fixed effects models, and refer to the second category as random effects models (see Manski, 1987, and Chamberlain, 1984).

ML, since no sufficient statistic is known, irrespective of the distribution of u_{it}.[3]

An example of a model in the second category is given by Chamberlain (1984). He assumes that α_i depends in a linear way on x_i: $\alpha_i = a'x_i + w_i$, with $w_i \sim N(0,\sigma^2_w)$, $u_i \sim N(0,\Sigma)$ without restrictions on Σ, and w_i, u_i and x_i independent. A two stage estimation procedure is used to estimate β. First the model is estimated for each t separately (ignoring cross-equation restrictions). The second step is a minimum distance step, taking account of the cross-equation restrictions. This estimator allows for a specific form of correlation between α_i and x_i but still assumes normality of w_i and u_i. If a=0 this model simplifies to the random effects model. Thus for the random effects model, Chamberlain's estimation procedure is an alternative to simulated ML.

To avoid the normality and independence assumptions, semiparametric estimation methods can be used. For T=2, Honoré (1992) derives two CMRs under the following basic assumption (with i subscript suppressed from now on):

Conditional symmetry assumption

The distribution of $(u_1+\alpha, u_2+\alpha)$, conditional on (x_1,x_2) is absolutely continuous and symmetric (i.e., has conditional density f, with $f(v_1,v_2|x_1,x_2)=f(v_2,v_1|x_1,x_2)$ for all (v_1,v_2) and (x_1,x_2)).[4]

This is the assumption used in Honoré (1993). It replaces the somewhat weaker identifying assumption (E.3) in Honoré (1992). The conditional symmetry assumption allows for nonnormality and dependence between errors and/or the fixed effect and x_i. In this sense it is more general than the assumptions needed by Chamberlain (1984).[5]

The estimators developed by Honoré are obtained by constructing UMRs from CMRs, where the UMRs are chosen in such a way that their empirical counterparts are the first-order conditions of a strictly convex objective function. Each CMR yields its own estimator for β; the two CMRs are not combined. One of the two objective functions is not differentiable in a finite number of points. To obtain the asymptotic characteristics of the corresponding estimator requires then an approach as proposed by Pakes and Pollard (1989).

---

[3] In principle, β could be estimated by transforming the endogenous variables into binary choice dummies, and by applying conditional ML to the binary choice model with logistically distributed error terms. However, in this way identification of the variance of the error term is sacrificed and the slope parameters are only estimated up to scale.

[4] A primitive condition on the density of (u_1,u_2,α) conditional on (x_1,x_2) that is sufficient for conditional symmetry of $(u_1+\alpha,u_2+\alpha)$ is that $f(u_1,u_2,\alpha \mid x_1,x_2)=f(u_2,u_1,\alpha \mid x_1,x_2)$ for all (u_1,u_2,α) and (x_1,x_2).

[5] On the other hand, contrary to Honoré (1992), Chamberlain (1984) does not impose that the conditional variances of u_1 and u_2 are equal.

Because the objective function for the estimator based upon the other CMR is twice differentiable in all but a finite number of points, deriving the limit distribution of this estimator is straightforward, as is estimation of its covariance matrix. The related CMR will be referred to as the ‘smooth’ CMR. The two estimators share the property that (β,x_1,x_2) appears only as $\beta'(x_1-x_2)$, so identification hinges on variation in x_1-x_2. For example, coefficients related to time constant regressors are not identified. Since the estimates can be obtained by minimizing a strictly convex objective function, a local search algorithm can be used in locating the minimum.

The Honoré estimators are easy to obtain. But they lack (semiparametric) efficiency. In this paper, we construct more efficient estimators. One way to do this might be to construct the efficient scores, see Honoré (1993). However, estimation of the efficient scores appears to be hard in general: Honoré (1993) needs a specific distributional assumption concerning $(u_1+\alpha,u_2+\alpha)$, conditional on (x_1,x_2), to obtain a semiparametrically efficient estimator. This approach therefore does not seem to be generally applicable.

Instead of using the efficient scores to obtain an estimator that asymptotically attains the semiparametric efficiency bound, the approach suggested by Newey (1991) could be followed. Similar to Chamberlain (1987), he starts with the notion that conditional symmetry leads to infinitely many CMRs and also to infinitely many UMRs. The idea is to let the number of CMRs used in estimation grow to infinity at an appropriate rate as N tends to infinity. Newey shows that using this approach, the semiparametric efficiency bound can be attained asymptotically. However, in finite samples this approach requires making many choices: which of the infinitely many CMRs to use, and which functions of the conditioning variables to use to form UMRs.

An easier approach can be based on Newey (1993). This approach starts from a model identified by a given finite set of CMRs. These are used to construct UMRs using optimal instruments. The resulting UMRs can be used in a GMM estimation procedure. We will only use one CMR: the smooth CMR of Honoré (1992). As a consequence, we might loose efficiency compared to the semiparametric efficiency bound. Stated more precisely: our estimator attains the semiparametric efficiency bound in a class of models leading to this single CMR; this class may be larger than the one satisfying the conditional symmetry assumption.

3. Identification, consistency, efficiency and GMM estimation

Let T=2. Define $\Delta x=x_1-x_2$. In the remainder, we assume that Honoré’s conditions (those discussed previously and additional regularity conditions given in Honoré, 1992) are satisfied. Under these assumptions, Honoré (1992) derives the following smooth CMR

$$E\left\{\rho(y_1,y_2,\beta'\Delta x)\,|\,x_1,x_2\right\} \;=\; 0 \tag{2}$$

where

$$\begin{aligned}\rho(y_1,y_2,\beta'\Delta x) \;=\; & 1((y_1,y_2)\in\mathbf{A}_1)[y_1 - y_2 - \beta'\Delta x] \;-\; 1((y_1,y_2)\in\mathbf{A}_2)[y_2 - \max(0,-\beta'\Delta x)] \\ & +\; 1((y_1,y_2)\in\mathbf{B}_1)[y_1 - y_2 - \beta'\Delta x] \;+\; 1((y_1,y_2)\in\mathbf{B}_2)[y_1 - \max(0,\beta'\Delta x)]\end{aligned} \tag{3}$$

with

$A_1 = \{(y_1,y_2)\,|\; y_1 > \max\{0, \beta'\Delta x\},\ y_2 > y_1 - \beta'\Delta x\};$

$A_2 = \{(y_1,y_2)\,|\; 0 \le y_1 \le \beta'\Delta x,\ y_2 > 0 \text{ if } \beta'\Delta x \ge 0;\ y_1 = 0,\ y_2 > -\beta'\Delta x \text{ if } \beta'\Delta x < 0\ \}$

$B_1 = \{(y_1,y_2)\,|\; y_1 > \max\{0, \beta'\Delta x\},\ \max\{0, -\beta'\Delta x\} < y_2 < y_1 - \beta'\Delta x\};$

$B_2 = \{(y_1,y_2)\,|\; y_1 > \beta'\Delta x,\ y_2 = 0 \text{ if } \beta'\Delta x \ge 0;\ y_1 > 0,\ 0 \le y_2 \le -\beta'\Delta x \text{ if } \beta'\Delta x < 0\ \}.$

Note that ρ depends on β only through $\beta'\Delta x$. Therefore, a necessary condition for identification is that $E\{\Delta x \Delta x'\}$ has full rank. This excludes time constant regressors, whose effects will be picked up by the fixed effects.

CMR (2) implies that, for any (measurable) function $A(x_1,x_2)$,

$$E\left\{A(x_1,x_2)\rho(y_1,y_2,\beta'\Delta x)\right\} \;=\; 0 \tag{4}$$

For a given choice for $A(x_1,x_2)$, UMR (4) can be used to apply GMM. A condition for consistency of the GMM estimator is that (4) has a unique solution for β. This is difficult to prove in general. Honoré (1992) avoids this problem: he chooses $A(x_1,x_2)=\Delta x$, and constructs a strictly convex objective function, whose first order derivative equals the sample analogue of the UMR. This guarantees identification and consistency of the estimator obtained by minimizing the strictly convex function. We denote the estimator for β based on $A(x_1,x_2)=\Delta x$ by $\hat{\beta}_H$. In general, an estimator based on (4) for some arbitrary choice of $A(x_1,x_2)$ is denoted by $\hat{\beta}$.

The limit distribution of $\hat{\beta}$ is given by

$$\sqrt{N}(\hat{\beta} - \beta) \;\to^{d}\; N(0, G^{-1} V G^{-1\prime}), \tag{5}$$

where

$$G = E\left\{A(x_1,x_2)\frac{\partial\rho(y_1,y_2,\beta'\Delta x)}{\partial\beta'}\right\}, \qquad V = E\left\{A(x_1,x_2)\rho(y_1,y_2,\beta'\Delta x)\rho'(y_1,y_2,\beta'\Delta x)A(x_1,x_2)'\right\} \tag{6}$$

For an arbitrary choice of $A(x_1,x_2)$, including $A(x_1,x_2)=\Delta x$, $\hat{\beta}$ is generally not efficient. The semiparametric efficiency bound using only the information provided by (2), can be attained by using an optimal choice of instruments to turn (2) into unconditional moment restrictions, see Newey (1993). He shows that the optimal choice of instruments is given by $A(x_1,x_2)=B(x_1,x_2)$, with $B(x_1,x_2)\equiv D(x_1,x_2)'\Omega(x_1,x_2)^{-1}$, where

$$D(x_1,x_2) \;=\; E\left\{\frac{\partial\rho(y_1,y_2,\beta'\Delta x)}{\partial\beta'}\;\middle|\; x_1,x_2\right\} \;=\; -\Delta x'\; E\left\{1(-y_2<\beta'\Delta x<y_1)\;|x_1,x_2\right\} \tag{7}$$

$$\Omega(x_1,x_2) \;=\; E\left\{\rho(y_1,y_2,\beta'\Delta x)\rho(y_1,y_2,\beta'\Delta x)' \;\middle|\; x_1,x_2\right\} \tag{8}$$

As shown by Chamberlain (1987), the efficient estimator based on efficient GMM is not only efficient in the class of GMM estimators, but is also asymptotically efficient in the wider class of all consistent and asymptotically normal (regular) estimators. Therefore, the components of $B(x_1,x_2)\rho(y_1,y_2,\beta'\Delta x)$ can be interpreted as the efficient scores.

The optimal instruments are generally unobserved. Newey shows that, when applying GMM or some asymptotically equivalent method, the optimal instruments $B(x_1,x_2)$ may be replaced by a consistent (nonparametric) estimates $\hat{B}(x_1,x_2)$, without affecting the asymptotic characteristics of the resulting estimator. The resulting GMM estimator is:

$$\dot{\beta} = \underset{\beta\in\mathbf{B}}{\operatorname{argmin}}\; \left\{\sum \rho(y_1,y_2,\beta'\Delta x)'\hat{B}(x_1,x_2)'\left[\sum \hat{B}(x_1,x_2)\hat{B}(x_1,x_2)'\right]^{-1} \sum \hat{B}(x_1,x_2)\rho(y_1,y_2,\beta'\Delta x)\right\} \tag{9}$$

Its limit distribution is given by

$$\sqrt{N}(\dot{\beta} - \beta) \rightarrow^{d} N(0,\Lambda)\;,\; \text{where} \qquad \Lambda = \left(E\left\{D(x_1,x_2)'\Omega(x_1,x_2)^{-1}D(x_1,x_2)\right\}\right)^{-1} \tag{10}$$

Alternatively, one Newton-Raphson step can be performed using $\hat{\beta}_H$ as the starting solution. This yields an estimator that is asymptotically equivalent to efficient GMM. Denote this estimator by $\tilde{\beta}$. It is given by:

$$\tilde{\beta} = \hat{\beta}_H - \left[\sum \hat{\mathbf{B}}(x_1,x_2)\frac{\partial\rho(y_1,y_2,\hat{\beta}_H'\Delta x)}{\partial\beta}\right]^{-1} \sum \hat{\mathbf{B}}(x_1,x_2)\rho(y_1,y_2,\hat{\beta}_H'\Delta x) \ . \tag{11}$$

Particularly the latter approach is computationally convenient: $\hat{\beta}_H$ is easy to obtain due to strict convexity of the objective function it minimizes, and no numerical optimization is required in the second step. Therefore, this is the estimator we will focus on.

To apply (11), we need $\hat{B}(x_1,x_2)$. Newey proposes to use nearest neighbour or series approximation. In the nearest neighbour case, Newey proposes to estimate $B(x_1,x_2)$ by estimating $\Omega(x_1,x_2)$ and $D(x_1,x_2)$ separately. Let $\hat{\Omega}(x_1,x_2)$ be the nearest neighbour regression estimate based upon (8), using $\hat{\beta}_H$ instead of β ((4.2) in Newey, 1993). Newey proposes to split $\hat{D}(x_1,x_2)$ in a part depending on a finite dimensional nuisance parameter η and an additive remaining part ((4.4) of Newey, 1993). But (7) yields no natural way to parameterize some part of $D(x_1,x_2)$. Therefore, estimating $D(x_1,x_2)$ boils down to estimating $E\{1(-y_2<\beta'\Delta x<y_1 \mid x_1,x_2\}$ nonparametrically (after replacing β by $\hat{\beta}_H$). $B(x_1,x_2)$ is then estimated by $\hat{D}(x_1,x_2)'\hat{\Omega}(x_1,x_2)^{-1}$.

In the series approximation case, Newey proposes to estimate $B(x_1,x_2)$ as a whole, again using a part possibly depending on a finite dimensional parameter η and using a series approximation to approximate the multiplicative remaining term, see (5.1) of Newey (1993). In our case we can write $B(x_1,x_2) = -\Delta x F(x_1,x_2)$, and

$$F(x_1,x_2) = E\left\{1(y_2<\beta'\Delta x<y_1)\,|\,x_1,x_2\right\}\left[E\left\{[\rho(y_1,y_2,\beta'\Delta x)]^2\,|\,x_1,x_2\right\}\right]^{-1} \tag{12}$$

The real valued function $F(x_1,x_2)$ will be approximated by a series along the lines of Newey (1993) (after replacing β by $\hat{\beta}_H$). Thus, in neither of the two cases, we see a natural way to parameterize some part of $D(x_1,x_2)$ or $B(x_1,x_2)$. We therefore do not make use of the nuisance parameter η.

The following two theorems now follow from Newey (1993).

Theorem 1 (nearest neighbours):

If the conditions stated in assumptions 4.1, 4.3, 4.4 and theorem 1 of Newey (1993) are satisfied, then

$$\sqrt{N}(\tilde{\beta} - \beta) \rightarrow^d N(0,\Lambda), \ where\ \Lambda = \left(E\left\{D(x_1,x_2)'\Omega(x_1,x_2)^{-1}D(x_1,x_2)\right\}\right)^{-1} \tag{13}$$

A consistent estimator for Λ is given by

$$\hat{\Lambda} = \left[\frac{1}{N}\sum \hat{\boldsymbol{D}}(x_1,x_2)'\hat{\Omega}(x_1,x_2)^{-1}\hat{\boldsymbol{D}}(x_1,x_2)\right]^{-1} \tag{14}$$

The assumptions required for theorem 1 can be divided into assumptions that can easily be checked for the specific model of interest, and regularity conditions that are hard to check in practice. Those that can be checked are special cases of Newey's assumptions for the general case. We discuss them in the appendix.

Theorem 2 (series approximation):

Assume that the conditions stated in assumptions 4.1, 4.3, 5.1, theorem 2 and either assumptions 5.2 and 5.4 or 5.3 and 5.5 of Newey (1993) are satisfied. Then

$$\sqrt{N}(\tilde{\beta} - \beta) \rightarrow^{d} N(0,\Lambda),\ \textit{where}\ \Lambda = \left(E\left\{\boldsymbol{D}(x_1,x_2)'\Omega(x_1,x_2)^{-1}\boldsymbol{D}(x_1,x_2)\right\}\right)^{-1} \tag{15}$$

A consistent estimator for Λ *is given by*

$$\hat{\Lambda} = \left[\frac{1}{N}\sum \hat{\boldsymbol{B}}(x_1,x_2)\rho(y_1,y_2,\hat{\beta}'\Delta x)\rho(y_1,y_2,\hat{\beta}'\Delta x)'\hat{\boldsymbol{B}}(x_1,x_2)'\right]^{-1} \tag{16}$$

The assumptions, drawn from Newey (1993), are discussed in the appendix.

The main problem that occurs in practice is how to construct the estimators $\hat{B}(x_1,x_2)$ of the optimal instruments $B(x_1,x_2)$. Newey (1993) sketches a general procedure for nearest neighbours as well as series approximations. Among other things, this involves the choice of smoothness parameters. Newey (1993) provides some criteria to select the number of neighbours or the number of terms in the series approximation to use in estimation.

The nearest neighbours procedure is intuitively easier to understand than the series approximations, since it relies on two nonparametric regressions. We shall focus on nearest neighbours. In the empirical application, however, we also apply the series approximation procedure. Computational details of the latter can be found in the appendix.

4. The Nearest Neighbours GMM estimator and some Monte Carlo Results

The GMM estimator in Theorem 1 uses nearest neighbours to construct estimates for $D(x_1,x_2)$ and $\Omega(x_1,x_2)$. This requires the choice of the numbers of nearest neighbours. Newey (1993) proposes to choose the same number in $D(x_1,x_2)$ and $\Omega(x_1,x_2)$, using some type of cross-validation, with a criterion based on the linearized difference between "true" and estimated scores.[6]

Apart from the number of nearest neighbours, the norm (determining the distance function) and the weights have to be chosen. We used two norms: $\|x\|_1 = (x'S_{xx}^{-1}x)^{½}$ (norm 1), where S_{xx} is the sample covariance matrix of $x=(x_1,x_2)$. This norm is invariant to (nonsingular) linear transformations of x. And $\|x\|_2 = (x'\Delta^{-1}x)^{½}$ (norm 2), where Δ is a diagonal matrix with the sample variances of the components of x on the diagonal. This norm (proposed by Newey) is only invariant to the scale of x. We used the three choices for the weights given in Robinson (1987) (uniform, triangular and quartic).[7]

For both norms and all three weights, Newey's suggestion for determining the optimal number of nearest neighbours failed to work in the empirical application (see below). The value of the cross-validation objective function was decreasing in the number of nearest neighbours. Using 900 nearest neighbours in a sample of size 938, led to parameter estimates that are very different from the Honoré estimates and to huge standard errors. When we applied Newey's criterion using different numbers of nearest neighbours for D and Ω, the same problem occurred. An alternative selection method for the number of nearest neighbours is to perform cross-validation for D and Ω separately. This is what we use in the remainder.

To indicate how well this estimation procedure can perform in practice, a small Monte Carlo experiment is conducted. We use two time periods and a combination of specifications 3 and 5 of the Monte Carlo sudy performed by Honoré (1992). Because the dimension of the nonparametric regression is twice the number of explanatory variables, only two explanatory variables are included. Assuming that the explanatory variables are independently normally distributed might result in too optimistic Monte Carlo results (Chesher, 1995). Instead, we assume them to be independently chi-square distributed (specification 3 of Honoré, 1992). We allow for correlation between error term and fixed effect as in specification 5 of Honoré (1992). To be precise, write $x_t=(x_{1t},x_{2t})$, t=1,2, then $x_{1t}=\alpha+\eta_t$ and the random variables $\alpha,\eta_1,\eta_2,x_{21}$ and x_{22} are independent and all distributed χ^2_3, standardized to have mean zero and variance 1. Conditional on α, u_1 and u_2 are distributed $N(0,½+½\alpha^2)$.

The results with 1000 replications and different sample sizes are presented in table 1. The table reports the true parameter values, the estimated bias and root mean squared error (RMSE), and the root mean squared error implied by the asymptotic theory (ARMSE). The quartiles and the median

[6] the expression presented in Newey (1993) contains a small error on the top of page 433: ...+B(x)[...].. should read ...−B(x)[...]..

[7] Let m be the number of nearest neighbours. Uniform weights give all neighbours equal weight 1/m, triangular weights give weight $(m-j+1)/[½m(m+1)]$ to the j^{th} nearest neighbour, j=1,..,m, and quartic weights give weight $[m^2-(j-1)^2]/[m(m^2-(m-1)(2m-1)/6)]$ to the j^{th} nearest neighbour, j=1,..,m.

absolute error of the estimator (MAE), and the median absolute error predicted by the asymptotic distribution (AMAE) are also reported. In the nonparametric estimation of $D(x_1,x_2)$ and $\Omega(x_1,x_2)$ we used nearest neighbours with uniform weights and norm 1 (invariant to linear transformations of x).

We find that for a small sample (N=200) ARMSE is smaller for the efficient GMM estimator, but because RMSE is much larger than ARMSE, the asymptotic approximation is not accurate. For N=500 this improves: RMSE decreases substantially for the efficient GMM estimator, although the asymptotic distribution still does not appear to be an accurate approximation. The bias of the efficient GMM estimator is smaller in absolute value than the bias of the Honoré estimator. For N=5000, the asymptotic approximation to the root mean square error is quite accurate for both estimators. The GMM estimator now clearly outperforms its inefficient counterpart. The bias is much lower for the efficient GMM estimates.

In terms of mean absolute error (MAE) instead of mean square error, GMM already performs better for sample size 500. For N=200, it already performs as good as the Honoré estimator. The reason for the difference is that, for the smaller sample sizes, GMM in some replications leads to estimates which are far from the true parameter values. These get a larger weight in the RMSE criterion than in the MAE criterion.

Comparing the quartiles (LQ and UQ) we conclude that the distribution of the Honoré estimator is skewed to the right in small samples, which also occurred in the Monte Carlo study in Honoré (1992). The efficient GMM estimator is skewed to the left. The bias in the Honoré estimates is positive whereas it is negative for the efficient GMM estimates.

In spite of the limitations of the Monte Carlo set up, we are tempted to conclude that the estimator based on efficient GMM performs quite well provided that the sample size is large relative to the number of parameters. The latter condition is due to the nonparametric estimation of the optimal instruments.

5. Empirical Application

We want to explain earnings of married Dutch females in age between 18 and 65. Earnings are positive if the female works and zero if she does not, so the dependent variable is censored at zero. A static microeconomic model leading to a linear latent variable model for earnings is presented in Stoker (1992) for the cross-section case. His explanatory variables include family and individual characteristics affecting preferences and human capital variables correlated with the (potential) wage. An intertemporal choice model leading to (1) (including the individual effects) can easily be obtained as in MaCurdy (1981). In the remainder we will use the natural logarithm of after tax earnings+1 as the dependent variable. The log transformation is used to capture the usual lognormal model as a special case, and to prevent a large impact of outliers. The +1 is added to account for the zeros.[8]

As explanatory variables we include the logarithm of weekly other family income (including the husband's earnings; again, +1 is added to account for zeros), the number of hours per week that the male is working, and a dummy indicating whether the male is working. These variables characterize the effect of the husband's behaviour on the wife's labour supply. A dummy indicating whether the family contains children younger than six captures the effect of household composition.[9] Also included are calender time, age, and the female's education level. Data on experience are not available. Note that in the fixed effects model, the human capital variables related to the (potential) wage cannot be included due to lack of time variation.

We use data of the Dutch Socio-Economic Panel (SEP) of the years 1984 to 1988. Table 2 contains the definitions of all variables considered and sample statistics for the whole five year period.[10] In this section, we only use the data of the 1987 and 1988 waves. Our results are based upon 2278 individuals who are in both waves.

We started with the Chamberlain (1984) model, with the modification that only part of x could be used in explaining the 'fixed effect' α: $\alpha = a'z + w$, where $a=(a_1',a_2')'$, $z=(z_1',z_2')'$, z_t is a subvector of x_t (the same for all i=1,..,N, t=1,2), and w is a random error term. Variables that are time-invariant or change linearly over time cannot be included in z_t. However, this is not essential as long as we are not interested in the β parameters for these variables (as they are not identified). A constant term and the variables TIME, LOI, HM, DCH6, IEM, AGE, and EDF are included in x_t, whereas only the variables LOI, HM, DCH6, and IEM are included in z_t.

In the first stage of the estimation procedure all cross-equation restrictions are ignored and the model $y^*_{it}=\beta' x_{it}+a'z_i+w_i+u_{it}$, $y_{it}=\max\{0,y^*_{it}\}$, with w_i and u_{it} independently normally distributed, is estimated separately for each t. In the second stage, the cross-equation restrictions are taken into

[8] For the chosen scaling of earnings (Dfl per week), the +1 has a negligible effect on the positive earnings values.

[9] Preliminary results indicated that other variables related to children had no significant effect.

[10] Over time and on average, the variables log other family income (LOI), hours worked by the male (HM) and employment of the husband (IEM) hardly change. The dummy for children under six years (DCH6) tends to decrease, education level of the female (EDF) and her age (AGE) tend to increase. In each year approximately 35% of the married females has a job. The following numbers give some insight in participation mobility. Comparing consecutive years, we find that approximately 62.5% does not work in both years, 31% works in both years, 3% switches from not working to working and 3.5% switches the other way around. Comparing waves which are two years apart, these percentages are 60%, 28%, 6%, and 6%, respectively. If the time difference is three years, the percentages become 59%, 25.5%, 7.5%, and 8%. For a time difference of four years, they are 56%, 25%, 9%, and 10%.

account and a minimum distance step is performed. The estimation results with a=0 imposed (random effects) are presented in the second column of table 3. The estimates for the full model are presented in the third column. Here $\hat{\mu}_t^2$ is an estimate for Var$\{w_i+u_{it}\}$, t=1,2. Because some of the coefficients in the fixed effect are significant, the random effects specification (col. 2) is rejected against the fixed effect specification (col. 3). Comparing the results suggests that the random effects specification leads to an overestimation of the effect of children (DCH6) and employment of the husband (IEM). Other variables such as the education level of the female (EDF), for example, are not included in z_{it}, and thus their impact may also be interpreted as a fixed effect.

The assumptions of normality and homoskedasticity were tested after the first round, using the tests of Chesher and Irish (1987). The form of heteroskedasticity that was tested for was Var$\{w_i+u_{it}\}=\exp(\lambda' x_{it})$. Both assumptions were strongly rejected (both with a=0 and a≠0). This implies that the estimates of β may be inconsistent, which justifies the use of estimators based on weaker assumptions. Moreover, a test on overidentifying restrictions can be performed in the second step, by comparing the objective function value with the critical value of a χ^2_7 distribution. The hypothesis that the overidentifying restrictions are valid, is rejected at the 5 percent level. However, it should be realized that this test is only valid when the first round estimators are consistent, which may not be the case here as indicated by the tests after the first round.

The estimates proposed by Honoré (1992), based on (4) with $A(x_1,x_2)=x_1-x_2$, are presented in the fourth column of table 3. Because β affects ρ only through $\beta'\Delta x$, the coefficients related to TIME and AGE are not simultaneously identified. Hence only TIME will be included in the estimation. For the same reason, the effect of education level cannot be estimated, since EDF hardly changes over time. The remaining parameters are related to preferences and not to (potential) wages, and we are basically estimating labour supply responses.[11]

Only two effects are significant: the presence of a young child and the husband's hours of work have similar impact as in the fixed effects Chamberlain (1984) model. As before, the effect of family income is negative, but its significance level has dropped. The time trend can be compared to the sum of the time trend and the age effect in the Chamberlain model. It appears to have a large standard error, much larger than AGE and TIME in the Chamberlain model. The low significance levels might be caused by the fact that the estimator is not efficient. To show

[11] This raises the question why we do not estimate a model for hours worked instead of earnings. In some waves of SEP, for those who do not change jobs, hours worked are not measured in each wave but taken from the previous wave. This makes hours worked infeasible for a panel data analysis. Earnings are measured independently in each wave for all respondents. Moreover, hours worked are measured per week and show large spikes at 20 and 40 hours (see Van Soest, 1995, for example). A censored regression model is not appropriate to deal with this.

that it is important to take the censoring into account we also present the OLS estimates on the first differences in the fifth column of table 3. It is clear that the estimates differ substantially; the OLS estimate of the parameter related to DCH6 is significantly positive!

Before turning to efficient GMM, note that the observations for which $y_{i1}=y_{i2}=0$ contribute zero to $\rho(y_1,y_2,\beta'\Delta x)$, whatever the value of β. These observations are discarded in the nearest neighbours estimation, which reduces computer time substantially. Discarding these observations reduces the dataset to 938 observations. AGE and EDF could in principle be included in x_t, t=1,2, although the related coefficients in β are not identified. Including them may affect the weights in nearest neighbour estimation. It would also increase the dimension of the nonparametric regressions, however. In the remainder we therefore do not include them.

GMM with Nearest Neighbours

For both norms 1 and 2 and all three weights and for D as well as Ω (see Section 4), our cross-validation criterion function appeared to be U-shaped. The optimal numbers of neighbours varied from 5 to 7 for D, and from 46 to 68 for Ω (see table 4). Using these numbers of neighbours and performing one Newton-Raphson step starting from $\hat{\beta}_H$, led to the results reported in table 4. Reported standard errors are based on (14). For norm 1, the parameters related to DCH6 and HM are significant at the 5% level. For norm 2, only the parameter related to DCH6 is significant. The choice of weights does not affect the sign of the parameters and has a modest effect on the significance levels. For norm 1, the estimates and their standard errors are very similar for different choices of the weights. For norm 2, there is some variation with the weights, in particular for the estimates related to T and IEM. The significantly negative estimates for DCH6 and HM are quite robust. The estimates related to T, HM and IEM are rather different for the two norms, but the differences are not significant.

We investigated the sensitivity of the results w.r.t. the number of nearest neighbours used in estimation. Because the results with norm 1 seem to lead to rather low standard errors and the choice of weights does not matter very much, we only look at the model with norm 1 and uniform weights (to be referred to as the benchmark model). The sensitivity of the results w.r.t. the smoothness parameters is presented in table 5. Keeping the number of nearest neighbours optimal for Ω and varying this number for D, results in most cases in estimates and standard errors that are similar to the results with the optimal choice. An exception is the coefficient of IEM. Changing the number of neighbours for Ω influences the parameter estimates related to T and AGE2, whereas standard errors are hardly affected. The parameters related to LOI, HM and DCH6 remain significant and negative in all cases but one. Decreasing the number of neighbours in estimating D leads to an increase of the significance level of the parameter related to LOI. We conclude that the results are not very sensitive to the numbers of nearest neighbours.

Comparing the benchmark results in table 4 to the Honoré estimates in table 3, we find that all standard errors have decreased. All t-ratios have increased, and with uniform weights and norm 1 the estimate related to LOI is now significant. To test for model misspecification, a Hausman-type specification test is performed.[12] Comparing the Honoré estimates in table 3 with the results in table 4, the null hypothesis of correct specification of the model could not be rejected at the 5% level. This result was obtained for all six specifications in table 4. Except for the parameter related to TIME, the parameter estimates have not changed much. For all parameters, the 95% confidence intervals based on the estimates in tables 3 and 4 overlap.

GMM with Series Approximation.

The alternative approach to estimate $B(x_1,x_2)$ is to use series approximations. Newey (1993) shows how to approximate $B(x_1,x_2)$ directly instead of estimating $D(x_1,x_2)$ and $\Omega(x_1,x_2)$ separately. We apply his procedure. See the appendix for details on the computation and the choice of 'smoothness parameters', i.e. the choice which polynomials to include in the series, the so-called polynomial base.

Results are presented in table 6. Columns 3 to 6 contain the results for different choices of the polynomial base. Standard errors are based on (16). The results with a constant, IEM87, IEM88 and IEM87*IEM88 in the base, led to the over-all lowest value for Newey's cross-validation type of criterion function. These results will be considered as 'best' in the remainder. The estimates related to TIME, LOI, DCH6 and IEM are sensitive to the choice of base, although the t-ratios hardly change. Significant estimates of about −0.029 for HM and of about −1.9 for DCH6 are robust across base choice.

Comparing the results of series approximation with the results of nearest neighbours, we find that the same parameters are significant except for LOI. Estimated standard errors are lower for some parameters, but larger for others. Comparing the standard errors to the Honoré estimates in table 3, we again see that they slightly fall. In terms of t-ratios, we 'gained' a little bit more when using nearest neighbours than when using series approximation.

We performed a Hausman-type specification test based on the series approximation estimates in table 6. As in the nearest neighbours case, the null hypothesis of correct specification of the model

[12] The Hausman test requires a positive definite estimate for the covariance matrix of $\tilde{\beta}-\hat{\beta}$. We follow the standard approach to use (11) which implies that $\sqrt{N}(\tilde{\beta}-\hat{\beta})=\hat{C}v_N$, with $\hat{C}\to^p C$, $v_N\to^d N(0,\Sigma)$ as $N\to\infty$. Obtaining a positive definite estimator $\hat{\Sigma}$ for Σ is straightforward. It follows that $\hat{C}\hat{\Sigma}\hat{C}'$ is a positive (semi-) definite estimator for the covariance matrix of $\sqrt{N}(\tilde{\beta}-\hat{\beta})$.

could not be rejected at the 5% level.

An Alternative Specification Test

An alternative specification test for the fixed effects model (1) with the conditional symmetry assumption, is based upon combining two smooth CMRs of Honoré (1992). The first is the CMR for the censored regression model, equations (2) and (3), the basis for the Honoré estimator. The second is a similar smooth CMR for the case that only information on observations with $y_1>0$ and $y_2>0$ is used (the smooth CMR for the truncated model, equation (2.3) in Honoré (1992)). These two CMRs can be used to construct the following UMRs:

$$E\{\rho(y_1,y_2,\beta'\Delta x)\}=0$$

and

$$E\{1((y_1,y_2)\in A_1)[y_1-y_2-\beta'\Delta x]+E\{1((y_1,y_2)\in B_1)[y_1-y_2-\beta'\Delta x]\}=0.$$

The sample equivalents of these moments are evaluated at the Honoré (1992) estimator for β. Following Newey (1985), it is straightforward to derive a test statistic based upon the overidentifying restrictions, which, under the null of no misspecification, is chi squared distributed. The null was rejected at the 5% level, contrary to the result on the basis of the Hausman tests. An interpretation of this result is that the data support CMR (2)-(3), but do not support the more specific model assumptions (1) and conditional symmetry. Because of this, we only consider estimators based upon (2) and (3), and do not try to improve efficiency by using more CMRs.

6. Extension to a panel with more than two waves

6.1 Balanced panel

In this section we extend our analysis to more than two waves. We first look at the balanced panel. We assume that it is random (no attrition on the basis of the endogenous variable). The basic idea is to combine the conditional moment restrictions in (2) for each pair of panel waves. A sufficient assumption for this, together with regularity conditions similar to those for the two waves case, is the following generalization of Honoré's symmetry condition:

For all $s,t \in \{1,...,T\}$, $s\neq t$, the distribution of $(v_s,v_t) = (u_s+\alpha,u_t+\alpha)$, conditional on $x=(x_1',...,x_T')'$, is absolutely continuous and symmetric.[13]

This assumption is rather general and allows for all kinds of correlation structures between the random errors u_t. For example, it is less restrictive than the assumption of complete interchangeability, that, conditional on x, $v=(v_1,..,v_T)$ has the same distribution as $(v_{\pi(1)},..,v_{\pi(T)})$ for any permutation π. The latter allows for equicorrelated errors, but, for example, not for errors with first order autocorrelation.

Let $\Delta x_{st}=x_s-x_t$ and

$$\rho_{st}(\beta) \;=\; \rho(y_s,y_t,\beta'\Delta x_{st}) \tag{17}$$

where ρ is defined in (3). Then, for all $1\leq s<t\leq T$,

$$E\{\rho_{st}(\beta)|x_1,..,x_T\} \;=\; 0. \tag{18}$$

These CMRs can be stacked into one vector defining

$$\rho(\beta) \;=\; \left[\rho_{12}(\beta) \;\; \rho_{13}(\beta) \;\; . \;\; . \;\; \rho_{1T}(\beta) \;\; \rho_{21}(\beta) \;\; . \;\; . \;\; \rho_{T-1,T}(\beta)\right]' \tag{19}$$

Then, for any A(x), we have the UMR

$$E\{\mathbf{A}(x)\,\rho(\beta)\} \;=\; 0 \tag{20}$$

The optimal choice for A(x) is $B(x)\equiv D(x)'\Omega(x)^{-1}$, where

$$\mathbf{D}(x) = E\left\{\frac{\partial\rho(\beta)}{\partial\beta'}\bigg|x\right\} \;,\;\; \Omega(x) = E\{\rho\,\rho'|x\} \tag{21}$$

Estimation of the optimal instruments requires a preliminary estimator for β. Honoré (1992) suggests to construct such an estimator on the basis of

$$\mathbf{A}(x)=\begin{bmatrix} \Delta x_{12} & 0 & 0 & . & . & 0 \\ 0 & \Delta x_{13} & 0 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & . & . & . & . & \Delta x_{T-1,T} \end{bmatrix} \tag{22}$$

Combining (20) and (22) more moments than parameters are used in estimation, so, for example,

[13] For some of our estimators, it is sufficient to impose the slightly weaker condition of symmetry conditional upon x_s and x_t instead of upon $x_1,...,x_T$.

GMM with the optimal weighting matrix can be used. This requires estimating the optimal weighting matrix. For this, a consistent preliminary estimator for β can be constructed giving equal weights to the moments $\Delta x_{st}\rho_{st}$. This latter choice is convenient because the estimator can be obtained by minimizing a strictly convex objective function. Given this preliminary estimator, we can estimate the optimal weighting matrix and perform one Newton-Raphson step towards the solution of the optimal GMM estimator based on (20) and (22). We refer to this estimator, which is asymptotically equivalent to the GMM estimator with the optimal weighting matrix, as the Honoré estimator. The many moments used in estimation can be used to test for overidentifying restrictions.

The Honoré estimator for β can then be used to perform efficient GMM with the optimal choice for A(x), i.e. B(x). Of course, our estimator consists of calculating one Newton-Raphson step towards the solution of the efficient GMM estimator. We refer to this final estimator as efficient GMM. Its drawback is the large dimension of the nonparametric estimation of B(x) if the dimension of x_t or the number of time periods is large, as in our case.

Alternatively, we can use that $E\{\rho_{st}(\beta) \mid x_s,x_t\}=0$ for each $1\leq s<t\leq T$ and apply the estimation procedure for time periods for each combination (s,t), $1\leq s<t\leq T$ separately. To reduce the computational burden we determine the smoothness parameters for one particular pair (s,t) and use the outcome for all pairs. The final step in estimation is then Asymptotic Least Squares (ALS), to restrict the estimates for β to be the same for each combination (s,t). This strategy, referred to as the ALS estimator, might asymptotically be less efficient than efficient GMM, but is easier from a practical point of view. Moreover, it can also be applied to unbalanced panels.

To compare the estimation procedures, we applied them to the balanced subpanel for the years 1986-1988 (T=3). The dataset (with at least one non-zero observation on the dependent variable) consists of 823 observations. We use norm 1 and nearest neighbours with uniform weights. Cross-validation was used to determine the optimal numbers of nearest neighbours for D and Ω (see table 7, row 8; for ALS these numbers are based on 1987 and 1988 only).

The Honoré estimate for panel data is presented in the second column of table 7. Only the parameter related to the dummy for the presence of young children (DCH6) is significant, with value −2.357. The results for the other two estimation procedures using all elements but time (T) in calculating distances are presented in columns 3 and 4 of table 7. Efficient GMM leads to significant parameter estimates except for employment of the husband (IEM). ALS leads to significance of the presence of young children (DCH6) only, with estimate −2.297. Compared to the results based on 1987 and 1988 only, the impact of the husband’s hours worked (HM) disappears. Compared to efficient GMM, we see large differences in standard errors. The parameter estimates also differ but they are all insignificant.

We have only 823 observations, but the efficient GMM estimates in column 2 use a nonparametric estimator for a conditional mean, conditional on a twelve dimensional vector. To avoid this problem of dimensionality, we present results based on conditioning on all periods' values for HM and DCH6 only (cols. 5 and 6 of table 7). In section 5, these two appeared to be the main explanatory variables. This reduces the dimension of the nonparametric regression from 12 to 6. Excluding LOI and IEM from the set of conditioning variables in efficient GMM estimation, leads to significant changes in estimates for TIME, HM and IEM. For two-stage ALS the parameter estimates change substantially but not significantly for most parameters. The significant negative impact of young children is still robust, although its magnitude has changed. The ALS results show no significant changes compared to the Honoré estimates.

The objective function value (bottom row of table 7) can be used to perform a test on overidentifying restrictions in the Honoré estimates and the two-stage ALS estimator. In both cases, the hypothesis of no misspecification is rejected at the 5 percent level, but not at the 1 percent level. Comparing the efficient GMM estimates with the consistent Honoré estimates, a Hausman test can be performed. For both efficient GMM estimates, the null hypothesis of correct specification was rejected.

We also considered the balanced panel for the years 1984-1988 (T=5). Selecting the observations for which all information for all 5 waves was available and that had at least one nonzero observation on y_t, left us with a sample of only 243 observations. We found this too small for a sensible analysis involving higher dimensional nonparametrics.

Unbalanced panel

Let $c_{st}=1$ if (y_s,y_t,x_s,x_t) is fully observed and zero otherwise. We assume that the distributions of c_{st} and y_s,y_t are conditionally independent for given $x=(x_1',..,x_T')'$ (no selection or attrition bias). We then have

$$E\{c_{st}\rho_{st}(\beta)\,|\,x_1,..,x_T\} = 0, \text{ for all } s,\ t, \text{ with } 1 \le s < t \le T \qquad \textbf{(23)}$$

Because we do not observe $x_1,..,x_T$ for all individuals, we use the weaker CMR

$$E\{c_{st}\rho_{st}(\beta)\,|\,x_s,x_t\} = 0, \text{ for all } s,\ t, \text{ with } 1 \le s < t \le T \qquad \textbf{(24)}$$

We apply the two waves estimation procedure for each (s,t) separately and use ALS to estimate β. To reduce the computational burden we determine the smoothness parameters for one particular pair (s,t) and use the outcome for all pairs of waves.

The unbalanced panel (for 1984-1988) consists of those individuals who are observed in at least two waves, with positive earnings at least once. This leads to a sample of 1351 individuals. We use uniform weights and norm 1 and all elements in (x_s,x_t) but TIME are included in calculating distances. The optimal numbers of nearest neighbours used in estimation are the same as in section 5, i.e. 5 for $D(x_s,x_t)$ and 50 for $\Omega(x_s,x_t)$, $1\leq s<t\leq T$.

Estimation results for the Honoré estimator for panel data are presented in the seventh column of table 7. Again, DCH6 is significant only at the 5% level with a parameter estimate of −2.68. The same holds for the two-stage ALS estimates. Compared to the balanced sub-panel 1986-1988 we see that the standard errors have decreased and that no significant differences in parameter estimates can be found. Again, the similarity between the Honoré estimate and the two stage ALS estimate is striking. Even more striking is that standard errors have increased while we are trying to improve efficiency. Although we did not perform a Monte Carlo study for the two-stage ALS this might be caused by the relatively small amount of data per combination of years (on average about 900 individuals per combination of years with positive earnings in at least one of the two years). At the 5% level, the test on overidentifying restrictions results in rejecting the hypothesis of a correct specification for the Honoré estimate ($69.11>\chi^2_{45;0.05}=60.61$) whereas it is not rejected with two stage ALS ($58.67<\chi^2_{45;0.05}=60.61$).

7. Economic Interpretation

Our model explains earnings of married females, which are determined by hours worked and hourly wages. The Chamberlain (1984) estimates in table 3 already suggest that fixed effects are substantial, the random effects model being clearly rejected against the fixed effects alternative. Fixed effects in the labour supply decision have a clear interpretation in a life cycle context. The hourly wage is mainly determined by human capital variables that hardly vary independently over time, so that fixed effects and human capital effects on hourly wages cannot be distinguished.

In the fixed effect models, only the variables of the time varying regressors can be identified. These mainly refer to the labour supply decision. From the results we conclude that, *ceteris paribus*, the presence of a child less than 6 years old has a strong negative effect on the female's labour supply. The magnitude of the effect, though, is much smaller than in the random effects model. This is the most robust finding in the paper. It confirms with the common finding in the female labour supply literature. Obviously, the assumption of exogeneity could be criticized here (see Mroz, 1987). Our data and the semiparametric nature of our models do not allow to test this.

According to most of the estimates, other family income (mainly husband's earnings) has a negative effect, which is often significant. According to the results in table 7 column 8, the elasticity (for observations with positive earnings and other income) would be about −0.15. Note that in a standard life cycle model without uncertainty, the elasticity should be zero, because family consumption can be smoothed for changes in family income. Our results suggest that changes in other family income could at least partly be unanticipated, and lead to adjustments of permanent income.

We find some evidence suggesting that, *ceteris paribus*, the number of hours the male is working has a negative effect on the wife's labour supply. The average elasticity would be about −0.4, but the estimate is never very accurate. This result suggests that male and female leisure are substitutes. To disentangle the impact of the husband's hours worked and the husband's participation, we also included a dummy for the husband's employment. Its impact never appeared to be significant.

We find that the joint impact of time and age is insignificant. Our fixed effects model does not allow to distinguish between the (probably positive) time trend and the (probably negative) age effect. We also estimated the model with additional explanatory variables, such as the number of children in the family younger than 18, and age squared. In none of the estimation results these variables were significant. Including them had little effect on the other estimates.

Finally, it should be noted that for T=2 most specification tests led to the conclusion that the censored regression fixed effects model cannot be rejected. This is somewhat surprising, since in cross-section settings, the censored regression model is often found to be inferior to a less restrictive sample selection model (see, for example, Melenberg and Van Soest, 1995). Apparently, fixed effects may make a large difference here. On the other hand, tests including more time periods observations often led to rejection of the censored regression fixed effects model.

8. Conclusions

We have considered various estimators for the censored regression model, and applied them to panel data on earnings of married females. In the case of two panel waves, we have focused on the semiparametric estimator for models with fixed effects designed by Honoré (1992), and efficient GMM estimators based upon Newey (1993). Monte Carlo results suggest that these techniques work quite well in practice, although many observations are needed before we can gain some efficiency compared to the Honoré estimator. For the case of more than two time periods, we have considered an estimator proposed by Honoré (1992), an efficient GMM estimator for a balanced panel, and have looked at Asymptotic Least Squares estimators for both the balanced and the unbalanced case.

Our empirical results show that taking account of fixed effects substantially changes the conclusions on the sensitivity of female labour supply for the presence of children, other family income, the husband's hours of work, etc.

Contrary to the Honoré (1992) estimator, the efficient GMM estimator requires nonparametric estimation, involving the choice of smoothness parameters. Our sensitivity analysis for a panel with two waves shows that the results are not very sensitive to the choice of these parameters. Our results are somewhat mixed. Where the efficient GMM estimators should, at least asymptotically, be more efficient than Honoré's estimator, it does not lead to unambiguously smaller (estimates of) standard errors, although t-values do tend to increase.

The efficiency gains are obtained by an optimal construction of unconditional moment restrictions, given the choice of a conditional moment restriction. An alternative would be to consider more conditional moment restrictions. Honoré (1992) notes that there is an infinite number of conditional moments one could consider. However, from the Hausman specification test it follows that the conditional moment restriction used by Honoré is valid but that the assumption of conditional symmetry might not hold. Therefore we do not include more conditional moment restrictions based on this assumption in estimation. Another direction of future work would be to relax the model assumptions and consider selection models. Kyriazidou (1994) introduces a consistent estimator allowing for a general structure of fixed effects. More efficient estimators using this estimator as a starting point, could be obtained along the same lines as described in this paper.

References

Alderman, H., J. Hoddinot, L. Haddad and C. Udry, "Gender Differentials in Farm Productivity: Implications for Household Efficiency and Agricultural Policy," *Food Policy*, forthcoming.

Arabmazar, A. and P. Schmidt (1981), "Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity," *Journal of Econometrics*, 17, 253-258.

Arabmazar, A. and P. Schmidt (1982), "An Investigation of the Robustness of the Tobit Estimator to Non-Normality," *Econometrica*, 50, 1055-1063.

Chamberlain, G. (1984), "Panel Data." In *Handbook of econometrics*, ed. Z. Griliches and M. Intrilligator, vol.2, 1248-1318, Amsterdam: North-Holland Publishing Co.

Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-334.

Chesher, A. (1995), "A Mirror Image Invariance for M-Estimators," *Econometrica*, 63, 207-213.

Chesher, A., and M. Irish (1987), "Residual Analysis in the Grouped Data and Censored Normal Linear Model," *Journal of Econometrics*, 34, 33-61.

Gourieroux, C. and A. Monfort (1993), "Simulation-Based Inference, A Survey with Special Reference to Panel Data Models," *Journal of Econometrics*, 59, 5-33.

Heckman, J., and R. Willis (1976), "Estimation of a Stochastic Model of Reproduction: An Econometric Approach." In *Household Production and Consumption*, ed. N. Terleckyj, New York: National Bureau of Economic Research.

Honoré, B. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533-565.

Honoré, B. (1993), "Efficiency Considerations in Tobit Models with Fixed Effects." In *Panel Data and Labour Market Dynamics*, eds. H. Bunzel, P. Jensen and N. Westergård-Nielsen, Amsterdam, Elsevier Science Publishers.

Kyriazidou, E. (1994), "Estimation of a Panel Data Sample Selection Model", paper presented at Review of Economic Studies Tour, 1995.

Manski, C. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357-362.

MaCurdy, T. (1981), "An empirical Model of Labour Supply in a Life-Cycle Setting," *Journal of Political Economy*, 89, 1059-1085.

Melenberg, B. and A. van Soest (1995), "A semi-parametric model of vacation expenditures," *Journal of Applied Econometrics*, forthcoming.

Mroz, Th. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765-799.

Newey, W. (1985), "Maximum Likelihood Specification Testing and Conditional Moment tests", *Econometrica*, 53, 1047-1070.

Newey, W. (1991), "Estimation of Tobit Models under Conditional Symmetry." In *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, eds. W. Barnett, J. Powell and G. Tauchen, Cambridge, Cambridge University Press.

Newey, W. (1993), "Efficient Estimation of Models With Conditional Moment Restrictions." In *Handbook of Statistics, volume 11*, eds. G. Maddala, C. Rao and H. Vinod, Amsterdam, Elsevier Science Publishers.

Neyman, J. and E.L. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1-32.

Pakes, A., and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1057.

Robinson, P. (1987), "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," *Econometrica*, 55, 875-893.

Van Soest, A. (1995), "Discrete choice models of family labour supply," *Journal of Human Resources*, 30, 63-88.

Stoker, T. (1992), *Lectures on Semiparametric Econometrics*. CORE Lecture Series, Core Foundation, Louvain-la-Neuve, Belgium.

Udry, C. (1995a), "Gender, Agricultural Production and the Theory of the Household," working paper, Northwestern University.

Udry, C. (1995b), "Risk and Saving in Northern Nigeria," *American Economic Review*, forthcoming.

Table 1: Monte Carlo simulation. Honoré (1992) estimates and efficient GMM estimates with nearest neighbours, uniform weights, norm 1.

			True	Bias	RMSE	ARMSE	LQ	Median	UQ	MAE	AMAE
N=200	Honoré	b1	1.000	0.075	0.399	0.302	0.863	1.026	1.231	0.176	0.204
		b2	1.000	0.046	0.373	0.293	0.842	0.991	1.180	0.165	0.198
	Efficient GMM (nearest neighbours)	b1	1.000	−0.075	0.916	0.205	0.795	0.952	1.123	0.175	0.138
		b2	1.000	−0.062	0.794	0.173	0.813	0.952	1.097	0.155	0.117
N=500	Honoré	b1	1.000	0.033	0.204	0.186	0.902	1.020	1.130	0.114	0.125
		b2	1.000	0.026	0.211	0.186	0.892	1.000	1.120	0.113	0.125
	Efficient GMM (nearest neighbours)	b1	1.000	−0.018	0.232	0.111	0.908	0.988	1.063	0.081	0.075
		b2	1.000	−0.020	0.236	0.094	0.902	0.981	1.057	0.080	0.063
N=5000	Honoré	b1	1.000	0.007	0.060	0.058	0.962	1.004	1.046	0.041	0.039
		b2	1.000	0.005	0.060	0.058	0.963	1.001	1.041	0.040	0.039
	Efficient GMM (nearest neighbours)	b1	1.000	0.000	0.029	0.033	0.981	0.999	1.020	0.019	0.023
		b2	1.000	−0.001	0.028	0.029	0.981	0.998	1.018	0.019	0.019

Number of nearest neighbours used in nonparametric estimation of $D(x_1,x_2)$ and $\Omega(x_1,x_2)$.
(For each sample size and each number of neighbours, the fraction of the 1000 replications is given for which this number of neighbours was optimal.)

$D(x_1,x_2)$	2	4	6	8	10	12	14	16	18	20
N=200	0.14	0.37	0.29	0.14	0.05	0.01				
N=500	0.04	0.27	0.34	0.22	0.10	0.03	0.01			
N=5000		0.02	0.09	0.25	0.28	0.22	0.08	0.05	0.01	

$\Omega(x_1,x_2)$	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170
N=200	0.31	0.52	0.11	0.04	0.02												
N=500	0.01	0.48	0.41	0.09	0.01												
N=5000							0.02	0.07	0.19	0.24	0.22	0.13	0.07	0.03	0.02	0.01	

Table 2 : Variable definitions and sample statistics, 10976 observations

variable	description	Mean	Standard deviation	Minimum	Maximum
LINC*	log after tax earnings of the female (Dfl per week)	5.37	0.75	2.37	7.17
TIME	time (in years after 1900)	86.22	1.35	84	88
LOI	log of after tax other family income, excluding female's earnings and unemployment benefits and earnings of children, including husband's earnings and benefits (Dfl per week)	6.24	0.88	0	9.26
HM	male's number of hours worked per week	34.93	17.39	0	97
DCH6	dummy, indicating whether the family contains one or more children with an age less than 6 years (DCH6=1) or not (DCH6=0)	0.30	0.46	0	1
IEM	dummy, IEM=1 if the husband works, IEM=0 otherwise	0.84	0.37	0	1
AGE	age of the female	38.43	10.93	18	64
EDF	education level of the female (from 1: primary school only, to 5: university level)	2.31	0.97	1	5

* based on 3837 positive observations only

Table 3: Estimation Results Chamberlain Model and Honoré estimates (T=2) (standard errors in parentheses)

PARAMETER	Random effects estimates	Fixed effects estimates[14]	Honoré estimates	Linear model
TIME	0.060** (0.011)	0.059** (0.012)	−0.052 (0.078)	−0.676** (0.013)
LOI	−0.228 (0.130)	−0.338* (0.160)	−0.212 (0.149)	−0.699** (0.013)
HM	−0.052** (0.014)	−0.030* (0.013)	−0.031* (0.014)	−0.005* (0.002)
DCH6	−4.864** (0.284)	−1.970** (0.466)	−1.813** (0.356)	0.153** (0.038)
IEM	3.500** (0.721)	0.452 (0.659)	0.526 (0.728)	0.429 (0.076)
AGE	−0.205** (0.012)	−0.206** (0.013)		
EDF	1.336** (0.130)	1.343** (0.130)		
$\hat{\mu}_1^2$	4.976[a] (1.014)	4.918[a] (1.014)		
μ_2^2	5.102[a] (0.997)	4.893[a] (1.003)		
Obj. function	9.08	10.53		

* significant at the 5% level

** significant at the 1% level

[a] no test on significance carried out

[14] For the fixed effect, the coefficient in a_1 related to DCH6 and in a_2 related DCH6 were significantly negative at the 1% level and the coefficient in a_2 related to IEM was significantly positive at the 5% level, indicating that it is important to allow for correlation between the individual effect and the regressors.

Table 4 : Nearest Neighbours Estimates
(standard errors in parentheses)

Norm \ Weights		uniform (benchmark model)		triangular		quartic	
invariant to linear transformations		(5,50)[a]		(7,68)[a]		(7,66)[a]	
	TIME	0.063	(0.057)	0.058	(0.058)	0.048	(0.058)
	LOI	−0.237*	(0.104)	−0.214	(0.108)	−0.203	(0.107)
	HM	−0.031*	(0.013)	−0.031*	(0.013)	−0.030*	(0.014)
	DCH6	−2.007**	(0.342)	−1.957**	(0.349)	−1.983**	(0.344)
	IEM	0.505	(0.654)	0.580	(0.672)	0.503	(0.684)
invariant to multiplication (cf. Newey, 1993)		(5,46)[a]		(5,54)[a]		(5,50)[a]	
	TIME	−0.236**	(0.056)	−0.040	(0.056)	−0.059	(0.056)
	LOI	−0.161	(0.118)	−0.201	(0.110)	−0.189	(0.115)
	HM	−0.067**	(0.014)	−0.049**	(0.013)	−0.052**	(0.014)
	DCH6	−1.871**	(0.338)	−1.835**	(0.350)	−1.850**	(0.349)
	IEM	1.864**	(0.660)	1.114	(0.657)	1.214	(0.662)

* significant at the 5 % level.
** significant at the 1% level
[a] (b,c): b nearest neighbours used in estimation of $D(x_1,x_2)$ and c nearest neighbours used in estimation of $\Omega(x_1,x_2)$.

*Table 5 : Sensitivity of the results for the number of nearest neighbours (uniform weights)****

Norm		$(3,50)^a$		$(7,50)^a$		$(5,45)^a$		$(5,55)^a$	
invariant to linear transform.	TIME	0.055	(0.057)	0.065	(0.057)	0.103	(0.057)	0.033	(0.058)
	LOI	-0.263^{**}	(0.092)	-0.226^{*}	(0.108)	-0.244^{*}	(0.108)	−0.196	(0.103)
	HM	-0.030^{*}	(0.012)	-0.029^{*}	(0.014)	-0.031^{*}	(0.013)	-0.029^{*}	(0.013)
	DCH6	-2.007^{**}	(0.334)	-1.997^{**}	(0.337)	-1.993^{**}	(0.351)	-2.021^{**}	(0.340)
	IEM	0.785	(0.605)	0.482	(0.692)	0.502	(0.661)	0.396	(0.660)

* significant at the 5% level.
** significant at the 1% level.
a (b,c): b nearest neighbours used in estimation of $D(x_1,x_2)$ and c nearest neighbours used in estimation of $\Omega(x_1,x_2)$.
*** The cells contain parameter estimates and standard errors for the variables in the second column.

*Table 6 : Efficient GMM estimates based on optimal number of terms in the series approximation****

Base[a]	IEM87,IEM88,IEM87*IEM88		MH87,IEM87,HM88,IEM88		HM87,DCH687,HM88,DCH688,HM87*HM87,HM87*DCH687,HM87*HM88,HM87*DCH688,DCH687*HM88,DCH687*DCH688,HM88*HM88,HM88*DCH688	
TIME	−0.048	(0.079)	−0.047	(0.068)	0.060	(0.056)
LOI	−0.119	(0.099)	−0.037	(0.095)	−0.075	(0.072)
HM	-0.029^{*}	(0.012)	-0.029^{**}	(0.011)	-0.024^{*}	(0.012)
DCH6	-1.836^{**}	(0.209)	-1.919^{**}	(0.183)	-2.897^{**}	(0.171)
IEM	0.132	(0.807)	0.046	(0.789)	−0.308	(0.757)

* significant at the 5% level.
** significant at the 1% level.
a A constant term was always included in the base.
*** The cells contain the parameter estimates and standard errors based on (16).

Table 7: Estimation Results with more than two waves
(standard errors in parentheses)

Parameter	Honoré estimator	Efficient GMM, fully[a]	Two stage ALS, fully[a]	Efficient GMM, partly[b]	Two stage ALS, partly[b]	Honoré estimator	Two stage ALS, fully[a]
Panel	Balanced, '86-'88	Balanced, '86-'88	Balanced, '86-'88	Balanced, '86-'88	Balanced, '86-'88	Unbalanced, '84-'88	Unbalanced, '84-'88
TIME	0.058 (0.043)	0.047^{**} (0.001)	0.049 (0.048)	0.026^{**} (0.002)	0.085 (0.050)	−0.002 (0.029)	0.008 (0.028)
LOI	−0.093 (0.082)	-0.066^{**} (0.010)	−0.007 (0.064)	-0.073^{**} (0.008)	−0.561 (0.422)	−0.068 (0.052)	-0.153^{*} (0.065)
HM	−0.004 (0.008)	0.001^{**} (0.0004)	−0.006 (0.009)	-0.006^{**} (0.0004)	0.001 (0.011)	−0.012 (0.006)	−0.012 (0.006)
DCH6	-2.357^{*} (0.254)	-1.353^{**} (0.163)	-2.297^{**} (0.265)	-1.623^{**} (0.157)	-2.405^{**} (0.266)	-2.680^{**} (0.178)	-2.484^{**} (0.184)
IEM	0.193 (0.557)	0.006 (0.029)	0.262 (0.540)	0.345^{**} (0.030)	0.855 (0.967)	0.416 (0.292)	0.361 (0.356)
NN[c]		(6,95)	(8,65)	(6,135)	(14,120)		(5,50)
Function	21.65				18.70	69.11	58.66

[a] fully: the moments conditional on (LOI,HM,DCH6,IEM) are used in estimation
[b] partly: only the moments conditional on (HM,DCH6) are used in estimation
[c] (a,b): number of nearest neighbours used estimating D and Ω, respectively
* significant at the 5% level
** significant at the 1% level

Appendix

We briefly discuss the assumptions needed for Theorem 1 and 2, drawn from Newey (1993). Finally, some additional computational details of applying Theorem 2 are discussed.

Assumptions for Theorem 1

The main part of assumption 4.1 is the regularity condition that $E\{B(x_1,x_2)\Omega(x_1,x_2)B(x_1,x_2)'\}$ is nonsingular. Assumption 4.2 of Newey (1993) is not required because we do not apply GMM, but the two step procedure (11). Assumption 4.3 deals with properties of the first round estimator, $\hat{\beta}_H$ in our case, which can be checked partly. The crucial identification part of this assumption is that $E\{\Delta x\rho(y_1,y_2,\beta'\Delta x)\}=0$ is uniquely satisfied at the true value of β. This is proven in Honoré (1992). The condition that the first stage estimator should be based on a GMM type of objective function is not satisfied here, but lemma A.1 of Newey (1993) can be easily adapted such that theorem 1 still goes through for the first stage estimator used here. Assumption 4.4 contains moment and smoothness conditions related to $\|\rho\|^4$, $\|\partial\rho/\partial\beta\|^4$ and $\|\partial^2\rho/\partial\beta\partial\beta'\|$. The conditions in terms of the second order derivatives of ρ are satisfied here because $\partial^2\rho/\partial\beta\partial\beta'=0$, see (3). Assumption 4.5 of Newey (1993) is not needed because D is estimated nonparametrically. The only additional assumption in theorem 1 of Newey (1993) concerns the rate at which the number of nearest neighbours used in estimation tends to infinity as N tends to infinity.

Additional assumptions for Theorem 2

Apart from some regularity conditions, assumption 5.1 contains an assumption on the existence of $E\{\|D(x,\eta)\|^{2a/(a-2)+\delta}\}= E\{\|\Delta x\|^{2a/(a-2)+\delta}\}$ for some $a>2$, $\delta>0$, and an assumption on $(\eta-\eta_0)$. The latter is not needed here since we do not use η. The condition on Q and $\hat{Q}$ in assumption 5.1, is satisfied because we choose $Q=\hat{Q}=I$.

Assumptions 5.3 and 5.5 are verified for our specific application. We aim at approximating $B(x_1,x_2) = -\Delta x F(x_1,x_2)$, with the real valued function $F(x_1,x_2)$ given by (12). The function $F(x_1,x_2)$ is approximated using a polynomial base in elements of x_1 and x_2. More formally, let $a_K(x_1,x_2)=[a_{1K}(x_1,x_2),..,a_{KK}(x_1,x_2)]'$ represent the elements of the polynomial base. We then approximate $F(x_1,x_2)$ by $\hat{\gamma}'a_K(x_1,x_2)$, where $\hat{\gamma}$ still has to be determined (see below). Checking that Newey's assumptions 5.3 and 5.5 (that imply assumptions 5.2 and 5.4) are satisfied is easy here: Choose $a_{kK}(x_1,x_2)=p_k(x_1,x_2)$ [with $\tau_j(x_j)=x_j$]. (x_1,x_2) contains at least one continuously distributed component (LOI in our empirical example) with density assumed to be bounded away from zero on $(0,\infty)$. The elements in Δx are not linearly dependent, implying that the smallest eigenvalue of

$\{\Delta x \Delta x'\}\Omega(x_1,x_2)$ is bounded away from zero. By choosing the degree of the approximating polynomial increasing in K, J(K)=K and L_K=I, assumption 5.3 is satisfied except for the boundedness of the τ_j. The latter is not a problem because boundedness can be relaxed without affecting the results (Newey, 1993, p. 440). Assumption 5.5 is also easy to check, since R is only one dimensional here. With $a_{kK}(x_1,x_2)=p_k(x_1,x_2)$ and J=K, we can choose $\gamma_j=\gamma_{1j}$, j=1,..,K, so that assumption 5.5 is satisfied.

Computation of series approximations

Two problems remain after having chosen which elements in (x_1,x_2) form the polynomial base: how to estimate γ and how to determine K, the number of terms in the series approximation. Both are addressed by Newey (1993). Estimation of γ (for given K) is based on the same intuition as in nearest neighbours: approximate the efficient scores as good as possible using a minimum mean-square error criterion. This leads to an explicit expression for γ. Replacing expectations by sample averages then yields $\hat{\gamma}$.

To obtain K, a cross-validation criterion is used in which the $\hat{\gamma}$ is calculated N times, leaving out the i-th observation (i=1,..,N). This is repeated for several values of K. The optimal K minimizes the difference between the estimated and the true scores. Because the true scores consist of conditional expectations that are not observed, these are replaced by their nonparametric estimates, and the estimated scores are replaced by the dependent variables (as in cross-validation). Given γ and K, estimating the asymptotic covariance matrix of the estimator is straightforward.

Note that (12) implies $F(x_1,x_2)\geq 0$. For the approximation to F, this is indeed the case for the first model in table 6. The other two models in table 6 led to negative estimates for $F(x_{i1},x_{i2})$ for 56 and 70 observations, respectively. This is an additional reason why the first model is referred to as the 'best' series approximation model. Avoiding this problem is also a reason to focus on the nearest neighbour estimation when we consider more than two waves.