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

Erwin Charlier\textsuperscript{a,b,*}, Bertrand Melenberg\textsuperscript{a}, Arthur van Soest\textsuperscript{a}

\textsuperscript{a}Tilburg University, Department of Econometrics and CentER, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands
\textsuperscript{b}ABP Investments, Fixed Income Europe, P.O. Box 2889, 6401 DJ, Heerlen, The Netherlands

Received 1 November 1995; received in revised form 1 January 1999; accepted 1 April 1999

Abstract

Honore has introduced a semiparametric estimator for the censored regression panel data model with fixed individual effects. Newey has shown how to obtain efficient estimators under a given conditional moment restriction. We apply Newey's approach to obtain a two-step GMM estimator which is more efficient than the Honore estimator. We compare this estimator to the Honore estimator and to parametric estimators including Chamberlain's quasi-fixed effects estimator in a Monte Carlo experiment. We also extend the Honore estimator and the two-step GMM estimator to the case of a balanced or unbalanced panel of more than two waves. We apply the estimators to an empirical example concerning earnings of married females, using data from the Dutch Socio-Economic Panel. © 2000 Elsevier Science S.A. All rights reserved.

\textit{JEL classification:} C14; C23; C24

\textit{Keywords:} Panel data; Censored regression; Semiparametric efficiency; Unbalanced panel

*Corresponding author. Tel.: +31-13-66-91-11; fax: +31-13-66-30-72.
E-mail address: e.charlier@kub.nl (E. Charlier)
1. Introduction

This paper considers estimation of the censored regression panel data model with individual effects. This model has important applications in micro-econometrics. The seminal example is the labor supply model of Heckman and MaCurdy (1980), where nonparticipation leads to censoring at zero, and where the individual effects have a clear economic interpretation in a life cycle context. Other examples are Udry (1995), who analyzes grain sales and purchases of rural households in Nigeria, and Alderman et al. (1995) and Udry (1996), who analyze various types of labour inputs and manure input in agricultural production.

Two types of estimators for this model can be distinguished. Estimators which require a parametric specification of the model are discussed by Chamberlain (1984). These models have the advantage that not only the parameters can be estimated, but also quantities such as marginal effects of covariates on the observed censored variable, which may be more interesting than the parameters themselves from a policy point of view. On the other hand, they have the drawback that estimates of the parameters of such effects will, in general, be asymptotically biased if the parametric model is misspecified. To overcome this problem, Honoré (1992) has introduced a semiparametric version of the model, which avoids assumptions on the distributions of individual effects or error terms. He derives consistent estimators for the parameters of this model, but does not address estimation of the marginal effects.

In this paper, we introduce a new semiparametric estimator, which aims at improving the efficiency of one of the Honoré (1992) estimators. Honoré's estimator is based upon (unconditional) moment restrictions derived from a conditional moment restriction. Following Newey (1993), we nonparametrically estimate the optimal moment restrictions for Honoré's conditional moment restriction, and construct a two-step GMM estimator which will be asymptotically more efficient than the Honoré estimator. Like Honoré's estimator, our estimator is easy to compute. It is efficient in the class of estimators based upon the given conditional moment restriction, but does not attain the semiparametric efficiency bound for Honoré's semiparametric model, since this model leads to many more conditional moment restrictions, which our estimator does not exploit.

First, we consider the case of two panel waves. In a Monte Carlo experiment, we compare the two-step GMM estimator with Honoré's estimator and with some parametric estimators, including the Chamberlain (1984) estimator allowing for quasi-fixed effects. For the latter, we do not only look at parameter estimates, but also analyze the estimates of the marginal effects of changes in the covariates on the observed censored dependent variable. We then generalize the estimator to the case of more than two waves, for balanced as well as unbalanced panels. Finally, we compare the semiparametric and parametric estimates in an empirical example. 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 unsystematic earnings variations mainly reflect changes in labour supply, it is natural to add fixed effects (Heckman and MaCurdy, 1980).

The remainder of the paper is structured as follows. In Section 2, we present the model and discuss merits and drawbacks of existing parametric and semiparametric estimators. Section 3 introduces our two step GMM estimator and its properties for the case of two time periods. In Section 4, we compare semiparametric and parametric estimators in a Monte Carlo experiment. Section 5 considers the empirical application for two time periods. In Section 6, the GMM-estimator is extended to panel data with more than two waves, either balanced or unbalanced. In Section 7, it is applied to the same empirical example, using five waves of SEP data, from 1984 to 1988. This section also presents the economic interpretation of the results. Section 8 concludes.

2. Parametric and semiparametric models and estimators

The censored regression model for panel data with individual effects is given by

\[ y_{it}^* = x_i + \beta_0' x_{it} + u_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T, \]

\[ y_{it} = \max\{0, y_{it}^*\}. \]

Here \( i \) denotes the individual and \( t \) is the time period, \( y_{it}^* \) is a latent variable, \( y_{it} \) is the observed dependent variable, \( x_{it} \) is a vector of covariates, \( x_i \) is the individual effect, \( u_{it} \) is an error term and \( \beta_0 \) is a vector of unknown parameters to be estimated. We observe \( (y_{it}, x_{it}) \). We are interested in asymptotic results for \( N \to \infty \), but for fixed \( T \). We assume independence across individuals, but not necessarily over time. We discuss various models with different assumptions and corresponding estimators.

In models with random effects, \( x_i \) is assumed to be independent of \( x_{it} = (x_{i1}, \ldots, x_{iT})' \). Several parametric random effects models and corresponding estimators for \( \beta_0 \) have been proposed in the literature. For example, \( x_i \sim N(0, \sigma_x^2) \) and \( u_{it} = (u_{i1}, \ldots, u_{iT})' \sim N(0, \Sigma) \), with \( \Sigma = \sigma_u^2 I_T \) and \( I_T \) the \( T \times T \) identity matrix, yields the specification of equicorrelation of Heckman and Willis (1976). Here Maximum Likelihood (ML) can be applied. This requires numerical integration in one dimension. If milder restrictions on \( \Sigma \) are imposed, estimators can be based on \( T - 1 \) dimensional numerical integration or simulation (such as simulated ML or simulated moments; see Gourieroux and Monfort, 1993).
If the assumptions on the distributions of error terms and individual effects are satisfied, the estimators are consistent and asymptotically normal. If the assumptions of normality of \((u_i, z_i)\) or independence between \((u_i, z_i)\) and \(x_i\) are not satisfied, the estimators will generally be inconsistent (see, for example, 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 \(z_i\) conditional on \(x_i\) are imposed. The \(z_i\) are then usually considered as nuisance parameters. In the second category, some restrictions on the distribution of \(z_i\) are imposed, which do not exclude dependence between \(z_i\) and \(x_i\).

In models in the first category, it is usually assumed that \(u_{its}, i = 1, \ldots, N, t = 1, \ldots, T\), are i.i.d. and independent of \(x_i\). Since the \(z_i\) are parameters, models in this category suffer from the incidental parameter problem, see Neyman and Scott (1948). ML estimates will generally be inconsistent. For a specific distribution of the \(u_{its}\), \(\beta_0\) could be estimated up to scale using Chamberlain’s conditional logit estimator, but this approach does not work in general.

An example of a model in the second category is given by Chamberlain (1984). It assumes that \(x_i = \alpha_0 x_i + w_i\), with \(\alpha_0\) an unknown vector of (nuisance) parameters, \(w_i \sim N(0, \sigma_w^2)\), \(u_i \sim N(0, \Sigma)\) without restrictions on \(\Sigma\), and \(w_i, u_i\) and \(x_i\) independent. Chamberlain proposes a two stage procedure to estimate \(\beta_0\). First the reduced-form model for each cross-section is estimated, ignoring restrictions on the parameters across time. The second step is minimum distance, to take account of these restrictions. This model allows for a specific form of correlation between \(z_i\) and \(x_i\), but it retains the assumption of normality of \(w_i\) and \(u_i\). For \(\alpha_0 = 0\), it simplifies to a random effects model.

The main goal of semiparametric estimation is to avoid the distributional and independence assumptions discussed above, and to construct estimators of \(\beta_0\) which are consistent under more general assumptions. Honoré (1992) derives various semiparametric estimators for the case \(T = 2\). He provides two representations of the identification assumption. We use the representation stated in Honoré (1992, Assumption E.3, footnote 6). The starting point is the following basic assumption (with the subscript \(i\) suppressed from now on):

\[(A1) \; (\text{Conditional exchangeability assumption}).\]

The distribution of \((u_1, u_2)\), conditional on \((z, x_1, x_2)\), is absolutely continuous and \(u_1\) and \(u_2\) are conditionally interchangeable (i.e., have conditional density \(f\), with \(f(u_1, u_2|z, x_1, x_2) = f(u_2, u_1|z, x_1, x_2)\) for all \((u_1, u_2)\) and \((z, x_1, x_2)\)).

\footnote{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; Chamberlain, 1984).}
Assumption (A1) allows for nonnormality and dependence between the errors and \((x, x_1, x_2)\), and imposes no restrictions on the distribution of \(x\) conditional on \((x_1, x_2)\). In this sense it is more general than the assumptions needed by Chamberlain (1984). On the other hand, the Chamberlain (1984) model is not fully nested in that of Honoré (1992), since Chamberlain (1984) does not impose that the conditional distributions of \(u_1\) and \(u_2\) have the same variance.

Honoré derives two conditional moment restrictions (CMRs) from (A1). He then constructs unconditional moment restrictions (UMRs) from these CMRs. The value of the (generic) parameter vector \(\beta\) which satisfies the empirical counterpart of these UMRs provides a consistent estimator of \(\beta_0\). The UMRs are chosen in such a way that these empirical counterparts of the UMRs are the first-order conditions of minimizing a strictly convex objective function with respect to \(\beta\), implying that the estimates are easy to compute with a local search algorithm.

Each CMR yields its own estimator for \(\beta_0\); Honoré does not combine the two CMRs. The two estimators share the property that, in the UMRs or the objective function, \((\beta, x_1, x_2)\) appears only as \(\beta'(x_1 - x_2)\). This implies that estimation hinges on variation in \(\Delta x = x_1 - x_2\). The coefficients of time-invariant regressors are not identified.

We will use Honoré’s second estimator, for which the corresponding objective function is everywhere continuously differentiable and twice differentiable in all but a finite number of points (the UMRs are given in (2.5) in Honoré (1992)). This makes it straightforward to derive the limit distribution of this estimator and to estimate its covariance matrix. The corresponding CMR will be referred to as the ‘smooth’ CMR.

There are various strategies to construct more efficient estimators than this Honoré estimator. The first would be to construct a semiparametrically efficient estimator based upon the efficient scores corresponding to (A1). Estimation of the efficient scores appears to be hard in general, however. Honoré (1993) needs a specific distributional assumption concerning \((u_1 + x, u_2 + x)\), conditional on \((x_1, x_2)\). This approach can thus not be applied without making specific additional assumptions, and we will not use it.

A second strategy is suggested by Newey (1991). Following Chamberlain (1987), this starts with noting that conditional exchangeability leads to infinitely many CMRs and infinitely many UMRs. The idea is then to let the number of CMRs used in estimation grow to infinity at an appropriate rate as \(N\) tends to infinity. Newey shows that this approach leads to an estimator which attains the semiparametric efficiency bound for (A1). In finite samples, however, this approach requires many choices: which of the infinitely many CMRs should be used, and which functions of the covariates should be used to form UMRs. We compare some Monte Carlo results for one and two conditional moment restrictions in Section 4. The Monte Carlo evidence suggests that increasing the
number of CMRs used in estimation does not automatically lead to a large increase in efficiency, unless the data set is very large.

We will focus on the easier approach based on Newey (1993). Starting point is the smooth CMR of Honoré (1992). This CMR is used to construct optimal UMRs, on which a GMM estimator is based. Thus, we do not aim at attaining the semiparametric efficiency bound for (A1). Our estimator will attain the efficiency bound for the class of models satisfying this single smooth CMR. This class may be larger than the class of models satisfying the conditional exchangeability Assumption (A1), since (A1) implies more CMRs.

3. Identification, consistency, efficiency, and GMM estimation

Let $T = 2, y = (y_1, y_2), x = (x_1, x_2), \text{ and } \Delta x = x_1 - x_2$. In the remainder we assume that Honoré’s conditions (Assumption (A1) and regularity conditions given in Honoré (1992)) are satisfied. The assumptions lead to infinitely many CMRs which can be presented compactly as in Honoré and Powell (1994). To do this, define

$$e_{12}(\beta) = \max\{x + u_1, -\beta' x_2, -\beta' x_1\} = \max\{y_1 - \beta' \Delta x, 0\} - \beta' x_2,$$

$$e_{21}(\beta) = \max\{x + u_2, -\beta' x_1, -\beta' x_2\} = \max\{y_2 + \beta' \Delta x, 0\} - \beta' x_1.$$

Then, under the exchangeability Assumption (A1),

$$e_{12}(\beta_0) - e_{21}(\beta_0) = \max\{y_1 - \beta'_0 \Delta x, 0\} - \max\{y_2 + \beta'_0 \Delta x, 0\} + \beta'_0 \Delta x$$

$$= \rho(y, \beta'_0 \Delta x) \quad (2)$$

is distributed symmetrically around zero conditional on $x$. This implies

$$E\{\xi(e_{12}(\beta_0) - e_{21}(\beta_0))|x\} = 0 \quad (3)$$

for any odd function $\xi$. In this section we restrict attention to the smooth CMR based on $\xi(a) = a$, used by Honoré (1992):

$$E\{\rho(y, \beta'_0 \Delta x)\} = 0. \quad (4)$$

In Section 4 we will present some results for other choices for $\xi(\cdot)$. As shown by Honoré, CMR (4) identifies $\beta_0$ if and only if $E\{1(P(y_1 > 0, y_2 > 0 | x') > 0) \Delta x \Delta x'\}$ has full rank. This excludes time constant regressors, whose effects will be picked up by the fixed effects.

CMR (4) implies that, for any function $A(x)$,

$$E\{A(x)\rho(y, \beta'_0 \Delta x)\} = 0. \quad (5)$$
For a given choice for $A(x)$, UMRs (5) can be used to apply GMM. A condition for consistency of the GMM estimator is that $b_0$ is the only value of $b$ which satisfies (5). This is difficult to prove in general. Honore (1992) solves the problem for this case: he chooses $A(x) = \Delta x$, and constructs a strictly convex objective function, whose first-order derivative is the sample analogue of the UMRs. This guarantees identification of the censored regression model with UMRs (5), and, since (5) is implied by (4), also of the model defined by CMR (4). It also guarantees consistency of the estimator obtained by minimizing the strictly convex function. We denote the Honore (1992) estimator for $b_0$ based on $A(x) = \Delta x$ by $\hat{b}_H$. An estimator based on (5) for some arbitrary choice of $A(x)$ is denoted by $\hat{b}_K$. The limit distribution of $\hat{b}_K$ is given by

$$\sqrt{N}(\hat{b} - b_0) \rightarrow dN(0, G^{-1}V G^{-1}),$$

where

$$G = E \left\{ A(x) \frac{\partial \rho(y, \beta_0 \Delta x)}{\partial \beta} \right\},$$

$$V = E \left\{ A(x) \rho(y, \beta_0 \Delta x) \rho'(y, \beta_0 \Delta x) A(x)' \right\}.$$  

For an arbitrary choice of $A(x)$, including $A(x) = \Delta x$, $\hat{b}$ is generally not efficient. The semiparametric efficiency bound using the information provided by CMR (4) can be attained by using an optimal choice of instruments $A(x)$. Newey (1993) shows that the optimal choice of instruments $A(x)$ is $B(x) = D(x) \Omega(x)^{-1}$, where, using (2),

$$D(x) = E \left\{ \frac{\partial \rho(y, \beta_0 \Delta x)}{\partial \beta} \right\} x = -\Delta x' E \{1(-y_2 < \beta_0 \Delta x < y_1) | x \},$$

$$\Omega(x) = E \{ \rho(y, \beta_0 \Delta x) \rho(y, \beta_0 \Delta x)' | x \}.$$  

As shown by Chamberlain (1987), the efficient GMM estimator is not only asymptotically efficient in the class of GMM estimators, but also in the wider class of all consistent and asymptotically normal (regular) estimators that only use the conditional moment restriction. The components of $B(x) \rho(y, \beta_0 \Delta x)$ can be interpreted as the efficient scores for CMR (4).

The optimal instruments $B(x)$ are generally unobserved. Newey shows that they may be replaced by consistent (nonparametric) estimates $\hat{B}(x)$, without affecting the limit distribution of the resulting GMM estimator. For computational convenience we will not compute the exact GMM estimator. Instead, we use $\hat{b}_H$ as the starting point and go one Newton–Raphson step towards efficient GMM. This yields the following estimator $\tilde{b}$, which is
asymptotically equivalent to efficient GMM (where $\Sigma$ denotes summation over the $N$ observations).

$$\hat{\beta} = \hat{\beta}_H - \left[ \sum \hat{B}(x) \frac{\partial \rho(y, \hat{\beta}_H \Delta x)}{\partial \beta'} \right]^{-1} \sum \hat{B}(x) \rho(y, \hat{\beta}_H \Delta x).$$  \hspace{1cm} (10)

This approach is computationally convenient: $\hat{\beta}_H$ is easy to obtain due to strict convexity of the objective function it minimizes, and the second step requires no numerical optimization.

To apply (10), we need $\hat{B}(x)$. Newey proposes to use nearest neighbors or series approximation.

3.1. Nearest neighbors estimation

We separately estimate the conditional expectations $\Omega(x)$ in (8) and $D(x)$ in (9) nonparametrically, with $\beta_0$ replaced by $\hat{\beta}_H$. The following theorem now follows from Newey (1993).

**Theorem 1 (Nearest neighbors).** If the conditions stated in Assumptions 4.1, 4.3, 4.4 and Theorem 1 of Newey (1993) are satisfied, then

$$\sqrt{N}(\hat{\beta} - \beta_0) \rightarrow^d N(0, A) \quad \text{where} \quad A = (E \{D(x)'\Omega(x)^{-1}D(x)\})^{-1}. \hspace{1cm} (11)$$

A consistent estimator for $A$ is given by

$$\hat{A} = \left[ \frac{1}{N} \sum \hat{D}(x)'\hat{\Omega}(x)^{-1}\hat{D}(x) \right]^{-1}. \hspace{1cm} (12)$$

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 terms of our application in the appendix. Here, we only remark that some of Newey’s assumptions are stronger than those of Honoré (1992) in the sense that higher (eighth) order moments of the variables $y$ and $x$ have to exist, implying that the improvement in asymptotic efficiency comes with a certain cost.

Nearest neighbors estimates of the conditional expectations $D(x)$ and $\Omega(x)$ are constructed as weighted averages of the $K$ values at the observations $i$ where $x_i$ is closest to $x$. This requires the choice of a norm (determining the distance function), the number of nearest neighbors $K$, and the weights. We use two norms: $\|x\|_1 = (x'S_{xx}^{-1}x)^{1/2}$ (norm 1), where $S_{xx}$ is the sample covariance matrix of $x$, and $\|x\|_2 = (x'\Delta^{-1}x)^{1/2}$ (norm 2), where $\Delta$ is the diagonal matrix with the
sample variances of the components of \( x \) on the diagonal. The first norm is invariant to (nonsingular) linear transformations of \( x \). The second (proposed by Newey) is only invariant to the scale of \( x \). We use all three choices for the weights given in Robinson (1987) (uniform, triangular and quartic).

For both norms and all three weights, Newey’s heuristic suggestion for determining the optimal number \( K \) of nearest neighbors failed to work in our empirical application. The value of the cross-validation objective function was decreasing in the number of nearest neighbors, and thus would lead to very large \( K \). Instead, we use cross-validation, separately for \( D \) and \( \Omega \).

3.2. Series approximation

Using (8) and (9), \( B(x) \) can be written as \(- \Delta x F(x)\). The function \( F(x) \) can be approximated by a series expansion \( \Sigma_{k} \gamma_k a_k(x) \), where \( a_k(k = 1, \ldots, K) \) is a set of polynomials satisfying the spanning condition that the linear combinations can approximate \( F(x) \) arbitrarily close as \( K \to \infty \). Newey (1993, Section 5) explains how to choose the \( \gamma_k(k = 1, \ldots, K) \) for given \( K \) and \( a_k \), and provides a heuristic rule for choosing the smoothing parameter \( K \). The details are intuitively less clear than for nearest neighbors estimation and we present them in the appendix. The results of Newey (1993) imply the following theorem.

**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

\[
\sqrt{N}(\hat{\beta} - \beta_0) \to ^d N(0, \Lambda) \quad \text{where} \quad \Lambda = (E\{D(x)' \Omega(x)^{-1} D(x)\})^{-1}. \tag{13}
\]

A consistent estimator for \( \Lambda \) is given by

\[
\hat{\Lambda} = \left[ \frac{1}{N} \sum \tilde{B}(x) \rho(y, \hat{\beta}_H \Delta x) \rho(y, \hat{\beta}_H \Delta x)' \tilde{B}(x) \right]^{-1}. \tag{14}
\]

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

The nearest neighbors procedure is intuitively easier to understand than the series approximations, since it relies on two nonparametric regressions. In the Monte Carlo simulation we shall focus on nearest neighbors. In the empirical application, however, we also apply the series approximation procedure.

---

2 Let \( m \) be the number of nearest neighbors. Uniform weights give all neighbors equal weight \( 1/m \), triangular weights give weight \((m - j + 1)/\lfloor 1/2m(m + 1) \rfloor \) to the \( j \)th nearest neighbor, \( j = 1, \ldots, m \), and quartic weights give weight \([m^2 - (j - 1)^2]/\lfloor m^2 - (m - 1)(2m - 1)/6 \rfloor \) to the \( j \)th nearest neighbor, \( j = 1, \ldots, m \).
4. Monte Carlo experiments

To analyze how well the two-step GMM estimator can perform in practice, a small Monte Carlo experiment is conducted, with a set-up similar to that in Honoré (1992). Two covariates are included. Normally distributed explanatory variables might result in too optimistic conclusions from Monte Carlo’s (Chesher, 1995). Instead, we use independent chi-squared covariates, as Honoré does in most specifications. We also allow for correlation between covariates and fixed effects in the same way as Honoré. Heteroskedasticity in the error terms is incorporated as in specification 5 of Honoré (1992). To be precise, with $x_t = (x_{1t}, x_{2t})$, $t = 1, 2$, we take $x_{1t} = z + \eta_t$, and the random variables $z, \eta_1, \eta_2, x_{21}$ and $x_{22}$ are independent and follow standardized $\chi_2^2$ distributions (with mean zero and variance one). Conditional on $z$, $u_1$ and $u_2$ are independent $N(0, \frac{1}{2} + \frac{1}{2}z^2)$. The true value of the parameter vector is $\beta_0 = (1, 1)'$.

We compare three estimators for $\beta_0$: the Honoré (1992) estimator and the efficient two step GMM estimator introduced above (both using only CMR (3) with $\xi(a) = a$), and the same type of efficient two-step GMM estimator based upon using both $\xi(a) = a$ and $\xi(a) = a^21(a > 0) - a^21(a \leq 0)$ in (3). In the nonparametric estimation of $D(x)$ and $\Omega(x)$ we used nearest neighbors with uniform weights and norm 1 (invariant to linear transformations of $x$).

In Table 1, we present the results for 1000 replications and different sample sizes. The Table reports the estimated bias and root mean squared error (RMSE), the root mean squared error implied by the asymptotic theory (ARMSE), the first and third quartile (LQ and UQ), the median absolute error (MAE), and the median absolute error predicted by the asymptotic distribution (AMAE).

For the Honoré estimator, the asymptotic variance is substantially smaller than the true variance for sample size $N = 200$, but not for the larger sample sizes. For our GMM estimator based upon the same CMR, the same problem also occurs for $N = 500$. For $N = 200$ and $N = 500$, the asymptotic RMSE is much smaller than the true finite sample RMSE, and the actual RMSE is larger than that of the Honoré estimator. Only for $N = 5000$, the GMM estimator behaves in accordance with its asymptotic approximation and clearly outperforms its inefficient counterpart: the RMSE is about half as large as that of the Honoré estimator. In terms of mean absolute errors (MAE), GMM already performs better than the Honoré estimator for smaller samples. The reason is that, for the smaller samples, some replications lead to extreme GMM estimates which get a much larger weight in the RMSE criterion than in the MAE criterion.

The quartiles LQ and UQ show that the distribution of the Honoré estimator is skewed to the right in small samples, as in the Monte Carlo study in Honoré (1992). The efficient GMM estimator is skewed to the left. The bias is positive in the Honoré estimates and negative for the efficient GMM estimates.
Table 1
Monte Carlo simulation. Honoré (1992) estimates and efficient GMM estimates with nearest neighbors, uniform weights, norm 1

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>Bias</th>
<th>RMSE</th>
<th>ARMSE</th>
<th>LQ</th>
<th>Median</th>
<th>UQ</th>
<th>MAE</th>
<th>AMAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>0.075</td>
<td>0.399</td>
<td>0.302</td>
<td>0.863</td>
<td>1.026</td>
<td>1.231</td>
<td>0.176</td>
<td>0.204</td>
</tr>
<tr>
<td>b2</td>
<td>1.000</td>
<td>0.046</td>
<td>0.373</td>
<td>0.293</td>
<td>0.842</td>
<td>0.991</td>
<td>1.180</td>
<td>0.165</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>-0.075</td>
<td>0.916</td>
<td>0.205</td>
<td>0.795</td>
<td>0.952</td>
<td>1.123</td>
<td>0.175</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>-0.062</td>
<td>0.794</td>
<td>0.173</td>
<td>0.813</td>
<td>0.952</td>
<td>1.097</td>
<td>0.155</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>2.43</td>
<td>71.323</td>
<td>0.127</td>
<td>0.655</td>
<td>1.012</td>
<td>1.468</td>
<td>0.390</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>1.177</td>
<td>25.625</td>
<td>0.110</td>
<td>0.624</td>
<td>0.975</td>
<td>1.382</td>
<td>0.380</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>0.033</td>
<td>0.204</td>
<td>0.186</td>
<td>0.902</td>
<td>1.020</td>
<td>0.130</td>
<td>0.114</td>
<td>0.125</td>
</tr>
<tr>
<td>b2</td>
<td>1.000</td>
<td>0.026</td>
<td>0.211</td>
<td>0.186</td>
<td>0.892</td>
<td>1.000</td>
<td>1.120</td>
<td>0.113</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>-0.018</td>
<td>0.232</td>
<td>0.111</td>
<td>0.908</td>
<td>0.988</td>
<td>1.063</td>
<td>0.081</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 5000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>0.030</td>
<td>0.236</td>
<td>0.094</td>
<td>0.902</td>
<td>0.981</td>
<td>1.057</td>
<td>0.080</td>
<td>0.063</td>
</tr>
<tr>
<td>b2</td>
<td>1.000</td>
<td>0.063</td>
<td>2.308</td>
<td>0.060</td>
<td>0.854</td>
<td>0.980</td>
<td>1.109</td>
<td>0.126</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>0.007</td>
<td>0.060</td>
<td>0.058</td>
<td>0.962</td>
<td>1.004</td>
<td>1.046</td>
<td>0.041</td>
<td>0.039</td>
</tr>
<tr>
<td>b2</td>
<td>1.000</td>
<td>0.005</td>
<td>0.060</td>
<td>0.058</td>
<td>0.963</td>
<td>1.001</td>
<td>1.041</td>
<td>0.040</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 5000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>0.000</td>
<td>0.029</td>
<td>0.033</td>
<td>0.981</td>
<td>0.999</td>
<td>1.020</td>
<td>0.019</td>
<td>0.023</td>
</tr>
<tr>
<td>b2</td>
<td>1.000</td>
<td>-0.001</td>
<td>0.028</td>
<td>0.029</td>
<td>0.981</td>
<td>0.998</td>
<td>1.018</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>1.000</td>
<td>-0.001</td>
<td>0.028</td>
<td>0.023</td>
<td>0.982</td>
<td>1.000</td>
<td>1.017</td>
<td>0.018</td>
<td>0.016</td>
</tr>
<tr>
<td>b2</td>
<td>1.000</td>
<td>-0.002</td>
<td>0.026</td>
<td>0.020</td>
<td>0.980</td>
<td>0.997</td>
<td>1.015</td>
<td>0.018</td>
<td>0.014</td>
</tr>
</tbody>
</table>
The third estimator should indicate the efficiency gain of using two instead of one CMR. For $N = 200$ and $N = 500$, the bias of this estimator is relatively large and the asymptotic approximation is inaccurate (ARMSE is much smaller than RMSE, and AMAE is much smaller than MAE). For $N = 5000$, this estimator outperforms the estimator based upon a single CMR, but the efficiency gain is quite small, in terms of RMSE as well as MAE.

In spite of the limitations of this Monte Carlo set up, we would conclude that the estimator based on efficient GMM using one CMR performs quite well provided that the sample size is large enough. Using an additional CMR did not work in small samples and hardly helped to improve efficiency for $N = 5000$.

The semiparametric model allows for estimating $b_0$ but not for estimating, for example, $E\{y_1 | x\}$ or the effects of changes in $x$ on $E\{y_1 | x\}$, which are often more interesting in practice, from a policy point of view. Such quantities can be estimated in a model which completely specifies the conditional distribution of the individual effects and error terms, such as the Chamberlain (1984) model. This requires more assumptions than the semiparametric model and, therefore, can easily be misspecified. If, in spite of misspecification, the model leads to reasonable estimates of the quantities of interest, it will still be more useful for practical purposes than the semiparametric model. This makes it worthwhile to investigate how large the bias on the policy relevant parameters can be when a misspecified Chamberlain model is used. We analyze this using the same Monte Carlo setup. We have computed the Chamberlain estimates for the same samples as before. Moreover, since the DGP used above differs from the Chamberlain model in various respects, we have also carried out the simulations for DGPs which are closer to the Chamberlain model. We focus on sample size $N = 500$.

The results are presented in Table 2. We consider the mean $E\{y_1 | x\}$ and the marginal effects (keeping the fixed effect constant, cf. Chamberlain, 1984, pp. 1273–1274), $E\partial/\partial x_{j1}[E\{y_1 | x, z\}] | x (j = 1, 2)$ at $x = 0$ (the mean of $x$). For five different DGPs, we present the three true values, the mean of the Chamberlain estimates for 500 Monte Carlo replications, and the 10th and 90th percentiles in these 500 replications. The first panel refers to the Chamberlain DGP (see Table 2 for details). In this case the Chamberlain model is not misspecified, and the Chamberlain estimates perform quite well. Mean values and marginal effects are close to their true values, and the true values are always between the 10th and 90th percentile of the Chamberlain estimates. This shows that 500 observations is enough to get reasonable finite sample behaviour of the Chamberlain estimator, if the model is correctly specified.

The second DGP differs from the first in the sense that the errors $u_i$ are heteroskedastic ($u_i | x, z \sim N(0, \delta^2(1 + x_i^2))$, as in specification 6 of Honoré). The performance of the Chamberlain estimator (not accounting for the heteroskedasticity) is not as good as for the first DGP, but still quite reasonable. All
Table 2
Estimates for the mean $E\{y_1 \mid x\}$ and the marginal effects (ME$_j$), based on the Chamberlain estimates in 500 Monte Carlo replications and for 5 different DGPs

<table>
<thead>
<tr>
<th>DGP</th>
<th>Quantity of interest$^*$</th>
<th>True value</th>
<th>Estimate based on Chamberlain model</th>
<th>10th and 90th percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Chamberlain</td>
<td>$E{y_1 \mid x}$</td>
<td>0.85</td>
<td>0.85</td>
<td>0.72–0.98</td>
</tr>
<tr>
<td></td>
<td>ME$_1$</td>
<td>0.64</td>
<td>0.65</td>
<td>0.55–0.75</td>
</tr>
<tr>
<td></td>
<td>ME$_2$</td>
<td>0.64</td>
<td>0.64</td>
<td>0.50–0.79</td>
</tr>
<tr>
<td>2. See DGP 1 but now $u_{it} \mid x_{2it} \sim N(0, \frac{1}{2}(1 + x_{2it}^2))$</td>
<td>$E{y_1 \mid x}$</td>
<td>0.78</td>
<td>0.89</td>
<td>0.73–1.09</td>
</tr>
<tr>
<td></td>
<td>ME$_1$</td>
<td>0.66</td>
<td>0.63</td>
<td>0.49–0.77</td>
</tr>
<tr>
<td></td>
<td>ME$_2$</td>
<td>0.66</td>
<td>0.53</td>
<td>0.18–0.81</td>
</tr>
<tr>
<td>3. DGP of Table 1 but now $u_{it} \sim N(0, 1)$ (z standardized chi-square)</td>
<td>$E{y_1 \mid x}$</td>
<td>0.43</td>
<td>0.43</td>
<td>0.35–0.51</td>
</tr>
<tr>
<td></td>
<td>ME$_1$</td>
<td>0.58</td>
<td>0.44</td>
<td>0.37–0.51</td>
</tr>
<tr>
<td></td>
<td>ME$_2$</td>
<td>0.58</td>
<td>0.45</td>
<td>0.41–0.56</td>
</tr>
<tr>
<td>4. DGP of Table 1 but now $z_i \sim N(0, 1)$ $u_{it} \mid z \sim N(0, \frac{1}{2}(1 + z_i^2))$</td>
<td>$E{y_1 \mid x}$</td>
<td>0.34</td>
<td>0.48</td>
<td>0.36–0.59</td>
</tr>
<tr>
<td></td>
<td>ME$_1$</td>
<td>0.65</td>
<td>0.47</td>
<td>0.38–0.55</td>
</tr>
<tr>
<td></td>
<td>ME$_2$</td>
<td>0.65</td>
<td>0.50</td>
<td>0.41–0.59</td>
</tr>
<tr>
<td>5. DGP of Table 1</td>
<td>$E{y_1 \mid x}$</td>
<td>0.30</td>
<td>0.53</td>
<td>0.27–1.16</td>
</tr>
<tr>
<td></td>
<td>ME$_1$</td>
<td>0.59</td>
<td>0.43</td>
<td>0.30–0.57</td>
</tr>
<tr>
<td></td>
<td>ME$_2$</td>
<td>0.59</td>
<td>0.45</td>
<td>0.30–0.62</td>
</tr>
</tbody>
</table>

$^*$All quantities of interest are evaluated at $x = 0$, the mean of $x$. Marginal effects are computed as $ME_j = E(\partial\bar{y}_j/\partial x_{ij} \mid E\{y_1 \mid x, z\} \mid x) = E\{y_1 \mid x, z\} \mid x = 0$ (the mean of $x$), $j = 1, 2$.

three true values lie between the 10th and 90th percentile of the Chamberlain estimates. The estimate of the marginal effect of the second covariate – the one driving the heteroskedasticity – is very inaccurate, in accordance with the large variation in the estimates of $\beta_2$ (not reported).

In the third DGP, we have used Honoré’s specification of fixed effects and covariates, as in Table 1. This implies that the conditional expectation of the fixed effects is no longer linear in the covariates, and the distribution of the fixed effects conditional upon the covariates is nonnormal, so that the Chamberlain model is misspecified. In DGP 3, this is the only source of misspecification – the error terms $u_t$ are normal and homoskedastic. For this DGP, the Chamberlain model is able to estimate the conditional mean of $y_t$ quite well, but not the marginal effects. These are underpredicted, and their true values exceed the 90th percentile of the distribution of the Chamberlain estimates. The source of this problem is that the estimates of $P(y_t > 0 \mid x)$ are too low (while there is only a small bias on $\beta$).
In DGP 4, the marginal distribution of the fixed effects is standard normal, but the error terms are heteroskedastic in the same way as in the setup for Table 1. The marginal effects are underpredicted and they are similar to the previous case. However, the conditional mean of $y_i$ is overestimated in the Chamberlain model. Finally, DGP 5 is the same as in Table 1 and combines the various sources of misspecification of the Chamberlain model. The Chamberlain estimates have a high variance, reflected in the large interval between 10th and 90th percentile of the parameter estimates, the conditional mean and the marginal effects. In particular, the estimate of the conditional mean is very inaccurate. Although the mean estimate deviates substantially from the true value, the true value is between the 10th and 90th percentile of the estimates. As in DGPs 3 and 4, the marginal effects are underestimated, and, for both covariates, their true value is near the 90th percentile of the 500 Chamberlain estimates.

These results imply that the conclusions on the marginal effects on the basis of the Chamberlain model can be biased if this model is misspecified. The size of this problem varies with the type of misspecification and may also be specific to our chosen Monte Carlo setup. In practical examples, misspecification of the Chamberlain model may be much less of a problem than for our DGPs, and the advantages of the full specification of the Chamberlain model could still outweigh the drawback of misspecification bias.

Alternatively, we could use the semiparametric model to give an approximation of the marginal effects of interest. Under the additional assumption that $u_t$ and $x_t$ are independent, it is easy to show that $E\{\partial/\partial x_t[E\{y_t \mid x, z]\}|x\} = \beta_0 P(x + \beta_0 x_t + u_t > 0 \mid x)$. Since the distributions of $x$ and $u$ are not specified, the probability in this expression cannot be computed, even if $\beta_0$ were known. But it can be approximated by the sample fraction of positive values of $y$, or estimated nonparametrically at the given value of $x$. We will use the former in Section 7 to interpret the results of our empirical application.

5. Empirical application

We want to explain earnings of married Dutch women in the age group 18–65. Earnings are positive if the female works and zero if she does not, and are thus censored at zero. A static microeconomic model leading to a censored regression cross-section equation for earnings is presented in Stoker (1992). 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 individual effects) can easily be obtained as in Heckman and MacCurdy (1980) or MacCurdy (1981). Since the human capital variables (education level, potential experience or age) do not
vary over time or vary over time in a systematic way, the coefficients in a fixed effects model should be interpreted as labor supply responses.\(^3\)

We use data of the Dutch Socio-Economic Panel (SEP) of the years 1984–1988. Earnings are measured in each wave for all respondents. The natural logarithm of after tax earnings + 1 is our 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, which has a negligible effect on the positive weekly earnings values.

Our explanatory variables include the logarithm of weekly other family income (LOI; it includes the husband’s earnings; again, + 1 is added to account for zeros), the number of hours per week that the husband works (HM), and a dummy indicating whether the husband works (IEM). A dummy indicating whether the family contains children younger than six (DCH6) captures the effect of household composition; preliminary results indicated that other variables related to children were not significant. Also included are calendar time (TIME), year of birth (YOB), and the woman’s education level (EDF). Data on actual experience are not available.

Table 3 contains the variable definitions and sample statistics. In each of the five waves, about 35% of the married females has a job. Comparing consecutive years, 31% work in both years, 3% switch from not working to working and 3.5% switch from working to not working. Comparing waves which are two years apart, these percentages are 28%, 6%, and 6%. For three years difference, the percentages are 25.5%, 7.5%, and 8%, and for four years, they are 25%, 9%, and 10%, respectively. These figures suggest that accounting for censoring is important.

In this section, we only use the 1987 and 1988 waves. We focus on the differences between the various estimates and on tests of the various specifications. The economic interpretation will be addressed in Section 7. Our results are based upon 2278 women who are in both waves. We start with the Chamberlain (1984) model. The vector of regressors includes a constant term and the variables TIME, LOI, HM, DCH6, IEM, YOB, and EDF. To identify the model, the variables that are time-invariant or change linearly over time (YOB, EDF, TIME) are not included in the quasi-fixed effects. The estimated parameters for these variables can partly reflect fixed effects.

---

\(^3\)This raises the question why we do not estimate a model for hours worked instead of earnings. In some waves of the 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. Moreover, hours worked are measured per week and show large spikes at 20 and 40 h (see Van Soest, 1995, for example). A censored regression model is not appropriate to deal with this.
Table 3
Variable definitions and sample statistics, 10,976 observations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINC *</td>
<td>Log after tax earnings of the female (Dfl per week)</td>
<td>5.37</td>
<td>0.75</td>
<td>2.37</td>
<td>7.17</td>
</tr>
<tr>
<td>TIME</td>
<td>Time (in years after 1900)</td>
<td>86.22</td>
<td>1.35</td>
<td>84</td>
<td>88</td>
</tr>
<tr>
<td>LOI</td>
<td>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)</td>
<td>6.24</td>
<td>0.88</td>
<td>0</td>
<td>9.26</td>
</tr>
<tr>
<td>HM</td>
<td>Male’s number of hours worked per week</td>
<td>34.93</td>
<td>17.39</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td>DCH6</td>
<td>Dummy, indicating whether the family contains one or more children with an age less than 6 yr (DCH6 = 1) or not (DCH6 = 0)</td>
<td>0.30</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IEM</td>
<td>Dummy, IEM = 1 if the husband works, IEM = 0 otherwise</td>
<td>0.84</td>
<td>0.37</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>YOB</td>
<td>Year of birth of the female</td>
<td>1947.79</td>
<td>10.96</td>
<td>1921</td>
<td>1969</td>
</tr>
<tr>
<td>EDF</td>
<td>Education level of the female (from 1: primary school only, to 5: university level)</td>
<td>2.31</td>
<td>0.97</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

*Based on 3837 positive observations only.

5.1. Chamberlain estimates

The estimates are presented in the third column of Table 4. For comparison, the results of the random effect model (with quasi-fixed effects independent of regressors) are presented in the second column. Some coefficients in the quasi-fixed effect are significant, and the random effects specification is rejected against the fixed effect specification by a Wald test. The parameter estimates for the time varying regressors have the same sign but differ in size: the random effects model overestimates the effect of children (DCH6) and employment of the husband (IEM).

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 \( \text{Var}(w_i + u_i) = \exp(\lambda'x_{it}) \). Both assumptions were rejected, implying that the first round estimates and thus also the second round estimates of \( \beta_0 \) may be inconsistent, and that semiparametric estimation of
Table 4
Estimation results Chamberlain model and Honoré estimates (\(T = 2\)) (standard errors in parentheses)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Random effects estimates</th>
<th>Fixed effects estimates*</th>
<th>Honoré estimates</th>
<th>Linear model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONST</td>
<td>–385.415d (\text{(26.214)})</td>
<td>–390.462d (\text{(27.242)})</td>
<td>–0.052 (\text{(0.078)})</td>
<td>–0.676d (\text{(0.013)})</td>
</tr>
<tr>
<td>TIME</td>
<td>–0.179 (\text{(0.111)})</td>
<td>–0.156 (\text{(0.106)})</td>
<td>–0.212 (\text{(0.149)})</td>
<td>–0.699d (\text{(0.013)})</td>
</tr>
<tr>
<td>LOI</td>
<td>–0.228 (\text{(0.130)})</td>
<td>–0.336d (\text{(0.013)})</td>
<td>–0.031c (\text{(0.014)})</td>
<td>–0.005c (\text{(0.002)})</td>
</tr>
<tr>
<td>HM</td>
<td>–0.052d (\text{(0.014)})</td>
<td>–0.030c (\text{(0.013)})</td>
<td>–0.031c (\text{(0.014)})</td>
<td>–0.005c (\text{(0.002)})</td>
</tr>
<tr>
<td>DCH6</td>
<td>–4.863d (\text{(0.284)})</td>
<td>–1.973d (\text{(0.467)})</td>
<td>–1.813d (\text{(0.356)})</td>
<td>0.153d (\text{(0.038)})</td>
</tr>
<tr>
<td>IEM</td>
<td>3.500d (\text{(0.721)})</td>
<td>0.457 (\text{(0.663)})</td>
<td>0.526 (\text{(0.728)})</td>
<td>0.429 (\text{(0.076)})</td>
</tr>
<tr>
<td>YOB</td>
<td>0.204d (\text{(0.013)})</td>
<td>0.206d (\text{(0.013)})</td>
<td>0.726 (\text{(0.013)})</td>
<td>0.429 (\text{(0.076)})</td>
</tr>
<tr>
<td>EDF</td>
<td>1.336d (\text{(0.130)})</td>
<td>1.342d (\text{(0.130)})</td>
<td>(\beta_0) (\text{(1.044)})</td>
<td>(\text{(1.042)})</td>
</tr>
<tr>
<td>(\hat{\beta}_1)</td>
<td>4.899b (\text{(1.044)})</td>
<td>4.900b (\text{(1.042)})</td>
<td>(\beta_1) (\text{(1.068)})</td>
<td>(\text{(1.072)})</td>
</tr>
<tr>
<td>Obj. function</td>
<td>8.981</td>
<td>10.53</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Here \(\hat{\beta}_0\) is an estimate for \(\text{Var}(\tau_i + u_{it}), t = 1, 2\).

*For the fixed effect, the coefficient in \(a_1\) related to DCH6 and in \(a_2\) related to 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.

bSignificant at the 5% level.

cSignificant at the 1% level.

\(\beta_0\) may be useful. The overidentifying restrictions are tested in the second step, by comparing the objective function value with the critical value of a \(\chi^2\) distribution. They are rejected at the 5% level (though this test is not valid if the first round estimators are inconsistent).

5.2. Honoré estimates

The estimates following Honoré (1992), based on (5) with \(A(x) = \Delta x\), are presented in the fourth column of Table 4. Since \(\beta_0\) affects \(\rho\) only through \(\beta_0 \Delta x\), the coefficients of YOB and EDF are not identified. The estimates for HM,
DCH6 and IEM are similar to those in the Chamberlain model, with similar significance levels. The time trend remains insignificant. The effect of other family income is negative, but its significance level has dropped. The low significance levels might be due to the fact that the Honoré estimator is not efficient.

To show the importance of taking account of the censoring, we present the OLS estimates on the first differences in the fifth column of Table 4. It is clear that these estimates are very different from all others. For example, the OLS estimate of the parameter related to DCH6 is significantly positive.

A specification test for the fixed effects model (1) with the conditional exchangeability assumption (A1), is a moment test based upon the smooth CMR for the truncated model, discarding observations with \( y_1 = 0 \) or \( y_2 = 0 \) (see Eq. (2.3) in Honoré, 1992). Not conditioning on \( x \) leads to one UMR, which can be used to construct a method of moments test following Newey (1985). The test statistic is evaluated at the Honoré (1992) estimate for \( \beta_0 \). Under the null of no misspecification, the test statistic is asymptotically chi-squared distributed with one degree of freedom. The null was rejected at the 5% level. An interpretation of this result is that the data do not support the specific model assumptions (1) and (A1), although they may still support the weaker CMR assumption (4) used for estimation. Because of this, we only consider estimators based upon (4) and do not use more CMRs.

5.3. GMM with nearest neighbors

For applying efficient GMM, observations with \( y_1 = y_2 = 0 \) are discarded in nearest neighbors estimation, since they contribute zero to \( \rho(y, \beta' \Delta x) \), whatever the value of \( \beta \). This reduces computing time substantially, reducing the dataset to 938 observations. YOB and EDF could, in principle, be included in the instruments, although their coefficients in \( \beta \) are not identified. Including them may change the weights in nearest neighbors estimation, but would also increase the dimension of the nonparametric regressions. Therefore, we do not include them.

For both norms and all three weights, and for \( D \) as well as \( \Omega \) (see Section 4), the cross-validation criterion function appeared to be U-shaped. The optimal numbers of neighbors varied from 5 to 7 for \( D \), and from 46 to 68 for \( \Omega \). Using these numbers of neighbors and performing one Newton–Raphson step starting from \( \tilde{\beta}_{H} \), led to the results in Table 5. Reported standard errors are based on (12). For norm 1, the parameter estimates of DCH6 and HM are significant at the 5% level. For norm 2, only the parameter estimate of DCH6 is significant. The choice of weights does not affect the sign of the parameters and has a modest effect on the significance levels. The significant negative estimates for DCH6 and HM are quite robust. The estimates of TIME, HM and IEM are rather different for the two norms, but the differences are insignificant.
Table 5
Nearest neighbors estimates (standard errors in parentheses)

<table>
<thead>
<tr>
<th>Norm</th>
<th>Weights</th>
<th>Uniform (benchmark model)</th>
<th>Triangular</th>
<th>Quartic (5,50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(5,50)^a</td>
<td>(7,68)^a</td>
<td>(7,66)^a</td>
</tr>
<tr>
<td>Invariant to linear transforms</td>
<td>TIME</td>
<td>0.063 (0.104)</td>
<td>0.058 (0.058)</td>
<td>0.048 (0.058)</td>
</tr>
<tr>
<td></td>
<td>LOI</td>
<td>-0.237^b (0.013)</td>
<td>-0.214 (0.108)</td>
<td>-0.203 (0.107)</td>
</tr>
<tr>
<td></td>
<td>HM</td>
<td>-0.031^b (0.031)</td>
<td>-0.031^b (0.013)</td>
<td>-0.030^b (0.014)</td>
</tr>
<tr>
<td></td>
<td>DCH6</td>
<td>-2.007^c (0.342)</td>
<td>-1.957^c (0.349)</td>
<td>-1.983^c (0.344)</td>
</tr>
<tr>
<td></td>
<td>IEM</td>
<td>0.505 (0.654)</td>
<td>0.580 (0.672)</td>
<td>0.503 (0.684)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5,46)^a</td>
<td>(5,54)^a</td>
<td>(5,50)^a</td>
</tr>
<tr>
<td>Invariant to multiplication (cf. Newey, 1993)</td>
<td>TIME</td>
<td>-0.236** (0.056)</td>
<td>-0.040 (0.056)</td>
<td>-0.059 (0.056)</td>
</tr>
<tr>
<td></td>
<td>LOI</td>
<td>-0.161 (0.118)</td>
<td>-0.201 (0.110)</td>
<td>-0.189 (0.115)</td>
</tr>
<tr>
<td></td>
<td>HM</td>
<td>-0.067^c (0.014)</td>
<td>-0.049^c (0.013)</td>
<td>-0.052^c (0.014)</td>
</tr>
<tr>
<td></td>
<td>DCH6</td>
<td>-1.871^c (0.338)</td>
<td>-1.835^c (0.350)</td>
<td>-1.850^c (0.349)</td>
</tr>
<tr>
<td></td>
<td>IEM</td>
<td>1.864^c (0.660)</td>
<td>1.114 (0.657)</td>
<td>1.214 (0.662)</td>
</tr>
</tbody>
</table>

^a Signiﬁcant at the 5% level.
^b Significant at the 1% level.
We investigated the sensitivity of the results for the numbers of nearest neighbors. Since the results with norm 1 are robust to the choice of weights, we focus on the model with norm 1 and uniform weights (the benchmark model). The results are presented in Table 6. Varying the number of nearest neighbors for $D$ or $\Omega$ does not substantially change the estimates or their standard errors. The largest variation is found for the parameters of TIME and IEM, but these parameters always remain insignificant.

The standard errors in the benchmark model in Table 5 are smaller than their Honoré counterparts in Table 4, though some of the differences are quite small. Significance levels have increased, and the parameter of LOI is now significant. The significant parameters have the same sign and are of similar magnitude for both estimators. A Hausman-type test for misspecification is performed, based upon comparing the two sets of estimates. The null hypothesis of no misspecification was not rejected at the 5% level. This result was obtained for all specifications in Table 5.

5.4. GMM with series approximation

The results for GMM with series approximations are presented in Table 7 (see Appendix for computational details). Columns 2–4 contain the results for different choices of the polynomials used in the series expansions. Standard errors are based on (14). They tend to be smallest in column 4, where most terms are used in the series approximation. The estimates of some parameters are rather sensitive to the choice of polynomials. HM and DCH6 are significant and negative in all three cases. The results in the lefthand panel gave the lowest value for Newey’s cross-validation criterion. As in the nearest neighbors case, the Hausman-type specification test does not reject the null of a correct specification at the 5% level.

The major difference between series approximation and nearest neighbor results is that LOI is now insignificant. Estimated standard errors are not systematically lower or higher for either technique. The standard errors tend to be smaller than those of the Honoré estimates in Table 4, but there are exceptions.

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 will assume throughout that whether or not an individual is observed in a given wave is

\[ V\{\sqrt{N(\hat{\beta} - \beta)}\} \]

This test requires a positive definite estimate for $V\{\sqrt{N(\hat{\beta} - \beta)}\}$. We follow the standard approach: (10) implies that $\sqrt{N(\hat{\beta} - \beta)} = \sqrt{C_{1N}}$, with $\sqrt{C_{1N}} \rightarrow p C$, $v_{N} \rightarrow ^{d} N(0, \Sigma)$ as $N \rightarrow \infty$. Obtaining a positive definite estimator $\hat{\Sigma}$ for $\Sigma$ is straightforward. Subsequently, $\hat{C}_{\Sigma} \hat{C}^{\prime}$ is a consistent positive (semi-) definite estimator for $V\{\sqrt{N(\hat{\beta} - \beta)}\}$. 

Table 6
Sensitivity of the results for the number of nearest neighbors (uniform weights)\(^d\)

<table>
<thead>
<tr>
<th>Norm</th>
<th>(3,50)(^a)</th>
<th>(7,50)(^b)</th>
<th>(5,45)(^b)</th>
<th>(5,55)(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invariant to linear transform</td>
<td>Time</td>
<td>0.055 (0.057)</td>
<td>0.065 (0.057)</td>
<td>0.103 (0.057)</td>
</tr>
<tr>
<td></td>
<td>Locally</td>
<td>-0.263(^c) (0.092)</td>
<td>-0.226(^b) (0.108)</td>
<td>-0.244(^b) (0.108)</td>
</tr>
<tr>
<td></td>
<td>Homogenously</td>
<td>-0.030(^b) (0.012)</td>
<td>-0.029(^b) (0.014)</td>
<td>-0.031(^b) (0.013)</td>
</tr>
<tr>
<td></td>
<td>DCH6</td>
<td>2.007(^c) (0.334)</td>
<td>1.997(^c) (0.337)</td>
<td>1.993(^c) (0.351)</td>
</tr>
<tr>
<td></td>
<td>IEM</td>
<td>0.785 (0.605)</td>
<td>0.482 (0.692)</td>
<td>0.502 (0.661)</td>
</tr>
</tbody>
</table>

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

Table 7
Efficient GMM estimates based on optimal number of terms in the series approximation\(^d\)

<table>
<thead>
<tr>
<th>Base(^a)</th>
<th>IEM87,IEM88,IEM87*IEM88</th>
<th>HM87,IEM87,HM88,IEM88</th>
<th>HM87,DCH687,HM88,DCH688,HM87<em>HM87,HM87</em>DCH687,HM87<em>HM88,HM87</em>DCH688,DCH687<em>HM88,DCH687</em>HM88<em>HM88,DCH688,HM88</em>HM88,DCH688</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>-0.048 (0.079)</td>
<td>-0.047 (0.068)</td>
<td>0.060 (0.056)</td>
</tr>
<tr>
<td>LOI</td>
<td>-0.119 (0.099)</td>
<td>-0.037 (0.095)</td>
<td>-0.075 (0.072)</td>
</tr>
<tr>
<td>HM</td>
<td>-0.029(^b) (0.012)</td>
<td>-0.029(^c) (0.011)</td>
<td>-0.024(^b) (0.012)</td>
</tr>
<tr>
<td>DCH6</td>
<td>-1.836(^c) (0.209)</td>
<td>-1.919(^c) (0.183)</td>
<td>-2.897(^c) (0.171)</td>
</tr>
<tr>
<td>IEM</td>
<td>0.132 (0.807)</td>
<td>0.046 (0.789)</td>
<td>-0.308 (0.757)</td>
</tr>
</tbody>
</table>

\(^a\) A constant term was always included in the base.
\(^b\) Significant at the 5% level.
\(^c\) Significant at the 1% level.
\(^d\) The cells contain the parameter estimates and standard errors based on (4).
wave is not related to the error terms \( u_t \), i.e., we will not address the possibility of attrition or selection bias. We first look at the balanced panel. The basic idea is to combine the conditional moment restrictions in (4) for each pair of panel waves. A sufficient assumption for this, together with regularity conditions similar to those for two waves, is the following generalization of Honore’s exchangeability condition:

\[
(A2) \quad \text{For all } s, t \in \{1, \ldots, T\}, s \neq t, \text{ the distribution of } (u_s, u_t), \text{ conditional on } (z, x) = (z, x'_1, \ldots, x'_T), \text{ is absolutely continuous, and } u_s \text{ and } u_t \text{ are exchangeable conditional on } (z, x).^5
\]

This assumption is rather general and allows for many correlation structures between the random errors \( u_t \). For example, it is less restrictive than the assumption of complete exchangeability, that, conditional on \((z, x)\), \( u = (u_1, \ldots, u_T) \) has the same distribution as \( (u_{\pi(1)}, \ldots, u_{\pi(T)}) \) for any permutation \( \pi \). For example, the latter does not allow for first order autocorrelation in the \( u_t \), while (A2) does.

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

\[
\rho_{st}(\beta) = \rho(y_s, y_t, \beta' \Delta x_{st}),
\]

where \( \rho \) is defined in (2). Then, for all \( 1 \leq s < t \leq T \),

\[
E\{\rho_{st}(\beta_0) \mid x\} = 0.
\]

These CMRs can be stacked into one vector defined as

\[
\rho(\beta) = [\rho_{12}(\beta), \rho_{13}(\beta), \ldots, \rho_{1T}(\beta), \rho_{21}(\beta), \ldots, \rho_{T-1,T}(\beta)]'.
\]

For any \( A(x) \), this leads to the UMR

\[
E\{A(x) \rho(\beta_0)\} = 0.
\]

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

\[
D(x) = E\left\{ \frac{\partial \rho(\beta_0)}{\partial \beta'} \mid x \right\}, \quad \Omega(x) = E\{\rho(\beta_0) \rho(\beta_0)' \mid x\}.
\]

Estimation of the optimal instruments requires a preliminary estimator for \( \beta_0 \). Honoré (1992) suggests to construct such an estimator on the basis of

\[
A(x) = \begin{bmatrix}
\Delta x_{12} & 0 & 0 & \ldots & 0 \\
0 & \Delta x_{13} & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & \ldots & \ldots & \Delta x_{T-1,T}
\end{bmatrix}.
\]

---

^5 For some of our estimators, it is sufficient to impose the slightly weaker condition of exchangeability conditional upon \( z, x_s \) and \( x_t \) instead of upon \( z, x'_1, \ldots, x'_T \).
This is the estimator which minimizes the equally weighted sum of the criterion functions of the Honoré (1992) estimates for all pairs of waves. Combining (18) and (20), more moments than parameters are used in estimation, so, for example, GMM with the optimal weighting matrix can be used. This requires estimating the optimal weighting matrix. For this, a consistent preliminary estimator for $\beta_0$ can be constructed giving equal weights to the moments $\Delta x_{st} \rho_{st}$. This is convenient since the estimator can be obtained by minimizing a strictly convex objective function. Given these preliminary estimates, the optimal weighting matrix can be estimated and one Newton–Raphson step towards the solution of the optimal GMM estimator based on (18) and (20) can be performed. We refer to this estimator, which is asymptotically equivalent to GMM 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 $\beta_0$ can be used as a starting point to perform efficient GMM with the optimal choice for $A(x)$, i.e. $B(x)$. As in Section 3, we will use the asymptotically equivalent estimator which goes one Newton–Raphson step towards the solution of the efficient GMM minimization problem. We refer to this 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 not very small, as is the case in our empirical example.

Alternatively, we can use that $E_{\rho_0}(\beta_0 | x_{st}, x_t) = 0$ for each $1 \leq s < t \leq T$ and use the two step GMM estimator for $T = 2$ introduced in Section 3 for each combination $(s, t), 1 \leq s < t \leq T$. To restrict the estimates for $\beta_0$ to be the same for each combination $(s, t)$, the final step in estimation is then Asymptotic Least Squares (ALS) (see, for example, Kodde et al., 1990). This procedure, which we will refer to as the ALS estimator, might asymptotically be less efficient than two step efficient GMM, but is easier from a practical point of view. Moreover, it can also be applied to unbalanced panels.

Using the balanced subpanel with complete information for all five waves and with at least one nonzero observation on $y_t$ leads to only 243 observations which is too few for our high dimensional nonparametric regression. Therefore, we applied the estimators to the balanced subpanel for the years 1986–1988 ($T = 3$). The dataset (with $y_{it} > 0$ for at least one $t$) consists of 823 individuals. We use norm 1 and nearest neighbors with uniform weights. Cross-validation was used to determine the optimal numbers of nearest neighbors for $D$ and $\Omega$ (for ALS, to limit the computational burden, the optimal numbers are computed for 1987 and 1988 and used for all pairs of waves).

The Honoré estimates are presented in the second column of Table 8. Only the dummy for young children (DCH6) is significant. The results for the other two estimators using all covariates but TIME to compute the distances needed for nearest neighbors, are presented in Columns 3 and 4. Although their signs are the same, the magnitudes and standard errors for the ALS and GMM estimates are rather different. The GMM estimates are all significant but one.
Table 8
Estimation results with more than two waves (standard errors in parentheses)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Honoré estimator</th>
<th>Efficient GMM, fully</th>
<th>Two stage ALS, fully</th>
<th>Efficient GMM, partly</th>
<th>Two stage ALS, partly</th>
<th>Honoré estimator</th>
<th>Two stage ALS, fully</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>0.058 (0.043)</td>
<td>0.047* (0.048)</td>
<td>0.049 (0.048)</td>
<td>0.026* (0.002)</td>
<td>0.085 (0.050)</td>
<td>0.002 (0.029)</td>
<td>-0.008 (0.028)</td>
</tr>
<tr>
<td>LOI</td>
<td>-0.093 (0.082)</td>
<td>-0.066* (0.064)</td>
<td>-0.007 (0.008)</td>
<td>-0.073* (0.022)</td>
<td>-0.561 (0.422)</td>
<td>-0.068 (0.052)</td>
<td>-0.153 (0.065)</td>
</tr>
<tr>
<td>HM</td>
<td>-0.004 (0.008)</td>
<td>0.001* (0.009)</td>
<td>-0.006 (0.004)</td>
<td>-0.006* (0.011)</td>
<td>0.001 (0.011)</td>
<td>-0.012 (0.006)</td>
<td>-0.012 (0.006)</td>
</tr>
<tr>
<td>DCH6</td>
<td>-2.357 (0.254)</td>
<td>-1.353* (0.163)</td>
<td>-2.297* (0.265)</td>
<td>-1.623* (0.157)</td>
<td>-2.405 (0.266)</td>
<td>-2.680* (0.178)</td>
<td>-2.484* (0.184)</td>
</tr>
<tr>
<td>IEM</td>
<td>0.193 (0.557)</td>
<td>0.006 (0.029)</td>
<td>0.262 (0.540)</td>
<td>0.345* (0.030)</td>
<td>0.855 (0.967)</td>
<td>0.416 (0.292)</td>
<td>0.361 (0.356)</td>
</tr>
<tr>
<td>NN</td>
<td>(6,95) (6,5)</td>
<td>(8,65) (8,65)</td>
<td>(6,135) (6,135)</td>
<td>(14,120) (14,120)</td>
<td>(28,99) (28,99)</td>
<td>(5,50) (5,50)</td>
<td>(5,50) (5,50)</td>
</tr>
<tr>
<td>Objective function</td>
<td>21.65</td>
<td>16.15</td>
<td>18.70</td>
<td>69.11</td>
<td>58.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Fully: the moments conditional on (LOI,HM,DCH6,IEM) are used in estimation.

*Partly: only the moments conditional on (HM,DCH6) are used in estimation.

*(a, b): number of nearest neighbors used estimating D and Ω, respectively.

*Significant at the 5% level.

*Significant at the 1% level.
However, for ALS only DCH6 is significant. Compared to the results based on 1987–1988 (Tables 5 and 6), the effect of the husband’s hours worked (HM) has disappeared.

We have only 823 observations, but the GMM estimates use a twelve-dimensional nonparametric regression. To avoid this problem of dimensionality, we also present (in Columns 5 and 6 of Table 8) estimates based on conditioning only on all periods’ values for HM and DCH6, the two most important explanatory variables according to Section 5. This reduces the dimension of the nonparametric regression from 12 to 6. It leads to some substantial changes in the parameter estimates for GMM. All ALS estimates but one remain insignificant. The significant negative sign of DCH6 remains.

The objective function value (bottom row of Table 8) can be used to perform a test on overidentifying restrictions in the Honore estimates and the two-stage ALS estimates. The hypothesis of no misspecification could not be rejected at the 5% level for the two-stage ALS estimates in Column 4. For the Honore estimates and the two-stage ALS estimates in Column 6, this hypothesis is rejected at the 5% level, but not at the 1% level. We also performed a Hausman test based upon the difference between GMM estimates and Honore estimates. The null of no misspecification was rejected at the 1% level.

6.2. 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 $x = (x_1, \ldots, x_T)'$ (no selection or attrition bias). We then have

$$E_{c_{st} \rho_{st} (\beta_0)}[x_1, \ldots, x_T] = 0, \text{ for all } s, t, \text{ with } 1 \leq s < t \leq T. \quad (21)$$

Since we do not observe $x_1, \ldots, x_T$ for all individuals, we use the weaker CMR

$$E_{c_{st} \rho_{st} (\beta_0)}[x_s, x_t] = 0, \text{ for all } s, t, \text{ with } 1 \leq s < t \leq T. \quad (22)$$

We apply the two waves estimation procedure for each $(s, t)$ separately and use ALS to estimate $\beta_0$. To limit the computational burden we determine the smoothing parameters for $(s, t) = (1, 2)$ 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 $(x_s, x_t)$ except for TIME are included in the distance calculations needed for nearest neighbors. The optimal numbers of nearest neighbors are the same as in Section 5.

The Honore estimates are presented in Column 7 of Table 8. Again, DCH6 is the only significant variable. Compared to the balanced sub-panel 1986–1988,
standard errors have decreased. The ALS estimates (Column 8) are similar to the Honore estimates, the main difference is that other income (LOI) is also significant at the 5% level. Most ALS standard errors are larger than the Honore standard errors, in contrast to what asymptotic theory predicts. At the 5% level, the test on overidentifying restrictions results in rejecting the hypothesis of a correct specification for the Honore estimate ($\chi^2_{45;0.05} = 60.61$) but not for two stage ALS ($58.67 < \chi^2_{45;0.05}$).

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 4 already suggest that fixed effects are important, since the random effects model is rejected against the quasi-fixed effects alternative. Fixed effects in the labor 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 disentangled.

In the fixed effect models, only the variables of the time varying regressors are identified. These mainly refer to the labor supply decision. From the results we conclude that, ceteris paribus, the presence of a child less than 6 yr old has a strong negative effect on the woman’s labor supply. The magnitude of the effect is much smaller in the fixed effects model than in the random effects model. Based upon the approximation discussed at the end of Section 4, the effect of a young child would be a reduction of average log earnings of about $-1.7$ (parameter estimate $-4.86$, times employment rate $0.35$) in the random effects model. In all fixed effects models it would be about $-0.7$. This is the most robust finding in the paper. The negative effect is in line with the common finding in the female labor supply literature.

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 Column 8 of Table 8, the (geometric) average of women’s earnings ($+1$) would fall by about $0.05\%$ if their husbands’ net income would increase by $1\%$. In a standard life cycle model without uncertainty, the elasticity would be zero, because family consumption would be smoothed for changes in family income. Our results suggest that changes in other family income may partly be unanticipated, and will lead to an adjustment of the family’s permanent income.

We find some evidence suggesting that, ceteris paribus, the number of hours the husband works has a negative effect on the wife’s labor supply. The effect of a $1\%$ rise of all husbands’ hours would be about $-0.13\%$, though 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, but this was never significant.

We find that the joint impact of time and year of birth (or 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.

Our main findings are interpreted in terms of women’s labor supply behavior. This raises the question whether an equation for hours of work instead of earnings would lead to different results. We have already explained above why our data are better equipped for analyzing earnings rather than hours worked. Still, we have estimated the same type of model for hours worked. This led to the same economic conclusions as those given above.

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, 1996). Apparently, fixed effects may make a large difference here. On the other hand, tests including more time periods often led to the conclusion that the censored regression fixed effects model should be rejected.

8. Conclusions

We have considered estimators for the censored regression model, compared them in a Monte Carlo experiment, and applied them to panel data on earnings of married females. We have focused on the semiparametric estimator for models with fixed effects designed by Honoré (1992), and efficient GMM estimators based upon this estimator, following Newey (1993). For two panel waves, Monte Carlo results suggest that the GMM technique works well in practice, though many observations are needed before efficiency gains compared to the Honoré estimator are obtained. We have also compared the semiparametric estimators with the parametric Chamberlain (1984) estimator, which models the fixed effect in terms of a linear combination of covariates and an error term, and assumes normality and homoskedasticity. Such an estimator has the advantage that the model is fully specified such that the model can be used to compute effects other than the slope coefficients, such as marginal effects of covariates on the observed dependent variable. Monte Carlo results, however, suggest that estimates of such effects can be biased if the Chamberlain model is misspecified.

For more than two panel waves, we have compared the estimator proposed by Honoré (1992), an efficient GMM estimator for a balanced panel, and
Asymptotic Least Squares estimators using either the balanced subpanel or the complete unbalanced panel.

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

The semiparametric estimators are fairly easy to compute, though the efficient GMM estimator requires the choice of smoothing parameters. Our sensitivity analysis for a panel with two waves shows that the results are not very sensitive to the choice of these parameters. Still, our results are somewhat mixed. Where the efficient GMM estimators should asymptotically be more efficient than Honoré’s estimator, they do not always lead to unambiguously smaller estimated standard errors.

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. In our Monte Carlo experiment, using a second conditional moment only leads to a small efficiency gain. In our empirical example Hausman specification tests suggest that the conditional moment restriction used by Honoré is valid but that the assumption of conditional exchangeability might not hold.

In principle, the GMM framework used here could also be extended to selection models. Kyriazidou (1997) 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.

Acknowledgements

We thank Bo Honoré, Marno Verbeek, Myoung-Jae Lee, Gerard Pfann, an associate editor and two referees for helpful comments. 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.

Appendix A

We briefly discuss the assumptions needed for Theorems 1 and 2, drawn from Newey (1993), for our empirical application, and present some computational details for applying Theorem 2.

A.1. Assumptions for Theorem 1

In Newey’s notation, write $z = (y, x')$ and $\rho(z, \beta)$ for $\rho(y, \beta' \Delta x)$ in (2). Assumption 4.1 of Newey (1993) requires that with probability one $\rho(z, \beta)$ is
continuous on the interior of a compact set, continuously differentiable on a neighborhood \( \mathcal{E} \) of \( \beta_0 \), that \( \rho(z, \beta) \) and \( \partial \rho / \partial \beta \) are bounded by a function \( d(y_1, y_2, x) \), with \( \mathbb{E}\{d^2(y_1, y_2, x)\} < \infty \), uniformly over \( \beta \), and the regularity condition that \( \mathbb{E}\{B(x)\Omega(x)B(x)\}' \) is nonsingular. For continuous differentiability we need that \( 1(- y_2 < \beta' \Delta x < y_1) \) is continuous in \( \beta \) with probability one. In our application, \( y_1 \) and \( y_2 \) are the log(female’s earnings + 1), and thus mixed discrete-continuous (zero or positive), \( x \) contains LOI, the log(other income + 1), and HM, hours worked by the husband, which are also mixed discrete-continuous. However, notice that female’s income and other income or hours worked by the husband generally will not simultaneously be equal to zero, so that the set at which \( 1(- y_2 < \beta' \Delta x < y_1) \) is discontinuous will be negligible. Alternatively, one could modify Newey’s approach by basing it upon Pakes and Pollard (1989), similar to Honoré (1992), in which case one will need as requirement continuity of \( P(- y_2 < \beta' \Delta x < y_1 | x) \). Assumption 4.2 of Newey (1993) is not required, because we do not apply GMM, but the two step procedure. Newey (1993)'s 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 \( \mathbb{E}\{\Delta x \rho(y_1, y_2, \beta' \Delta x)\} = 0 \) is uniquely satisfied at \( \beta_0 \). 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 of Newey (1993) requires uniform boundedness in \( L^2 \) norm of \( \| \rho(z, \beta) \|^4 \), \( \| \partial \rho(z, \beta) / \partial \beta \|^4 \), and \( \| \partial^2 \rho / \partial \beta \partial \beta' \| \), and a Lipschitz condition on \( \| \partial^2 \rho / \partial \beta \partial \beta' \| \) (all on a neighborhood \( \mathcal{E} \) of \( \beta_0 \)). From (2), it follows that this condition will be satisfied if eighth order moments of \( z \) exist. Thus the required assumptions are stronger than those required by Honoré (1992). 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 neighbors used in estimation tends to infinity as \( N \) tends to infinity.

**A.2. Additional assumptions for Theorem 2**

Apart from some regularity conditions, Assumption 5.1 of Newey (1993) contains an assumption on the existence of \( \mathbb{E}\{\|D(x,y)\|^{2a/(a-2)+\delta}\} = \mathbb{E}\{\|\Delta x\|^{2a/(a-2)+\delta}\} \) for some \( a > 2, \delta > 0 \). 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 \( B(x) = - \Delta x F(x) \), with the real valued function \( F(x) \) given by (23) below. The function \( F(x) \) is approximated using a polynomial base in elements of \( x \). More formally, let \( a_k(x) = [a_{1k}(x), \ldots, a_{kk}(x)]' \) represent the
elements of the polynomial base. We then approximate \( F(x) \) by \( \hat{\gamma} a_k(x) \), where \( \hat{\gamma} \) still has to be determined (see below). Checking that Newey’s assumptions 5.3 and 5.5 (that imply his Assumptions 5.2 and 5.4) are satisfied is easy here: Choose \( a_k(x) = p_k(x) \) [with \( \tau_j(x_j) = x_j \)]. \( x \) should contain at least one continuously distributed component with density assumed to be bounded away from zero on \((0, \infty)\). From a practical viewpoint, we can let LOI do the job. We tried estimation both with LOI included and excluded. The reported results are without LOI. Excluding LOI only changed the parameter estimates related to LOI and IEM. The elements in \( \Delta x \) are not linearly dependent, implying that the smallest eigenvalue of \( M' \Delta x \) is bounded away from zero. By choosing the degree of the approximating polynomial increasing in \( K \), \( J(K) = K \) and \( J' = I \), Assumption 5.3 is satisfied except for the boundedness of the \( \tau_j \). The latter is not a problem because boundedness can be relaxed without affecting the results (Newey, 1993, p. 440). Assumption 5.5 of Newey (1993) is also easy to check, since \( R \) is only one dimensional here. With \( a_{kk}(x) = p_k(x) \) and \( J = K \), we can choose \( \gamma_j = \gamma_{1p}, j = 1,...,K \), so that Assumption 5.5 is satisfied.

**A.3. Computation of series approximations**

In the series approximation of the optimal instruments we follow the procedure described in Newey (1993, pp. 435–438). Here, we discuss the series approximation in more detail. In the censored regression panel data model we can write \( B(x) = - \Delta x F(x) \), with

\[
F(x) = \mathbb{E}\{1(- y_2 < \beta'_0 \Delta x < y_1) \mid x\}\mathbb{E}\{[\rho(y_1, y_2, \beta'_0 \Delta x)]^2 \mid x\}^{-1}. \tag{23}
\]

In this case, instead of having to approximate a vector of functions of \( x \), we only have to approximate the real valued function \( F(x) \). This function is approximated by a series expansion \( \sum_k \gamma_k a_{kk}(x) \), as discussed above. For the functions \( a_{kk}(x) \) we use polynomials in elements of \( x \). For the moment, assume that the number of terms in the series approximation (i.e., \( K \)) is given. We now need to estimate \( \gamma = (\gamma_1, \ldots, \gamma_K)' \). Newey (1993, p. 436) gives an infeasible estimator \( \hat{\gamma} \) that minimizes the mean-square distance between the function \( B(x) \) and the series approximation \( - \Delta x[\sum_k \gamma_k a_{kk}(x)] \). In this distance function, a positive definite matrix \( Q \) appears. For simplicity and because no good guidelines for choosing \( Q \) are available, we choose \( Q \) equal to the identity matrix. The resulting estimator \( \hat{\gamma} \) is infeasible because it depends on, among other things, \( B(x) \), which is not observed. Therefore, the estimator \( \hat{\gamma} \) is rewritten in such a way that constructing a feasible estimator is easy. The resulting estimator \( \tilde{\gamma} \) is presented in Newey (1993, Eq. (5.3)), and \( \gamma \), the feasible equivalent of this estimator, is obtained by replacing the expectations by sample averages and \( \beta_0 \) by \( \hat{\beta}_0 \). Then \( \tilde{\gamma} \) is used to approximate the optimal instruments \( B(x) \) by \( - \Delta x[\sum_k \gamma_k a_{kk}(x)] \), which, in turn, are used to compute the efficient estimator \( \tilde{\beta} \).
Remains the problem of how to choose $K$, the number of terms in the series approximation. We follow Newey (1993, pp. 437, 438), where leave-one-out cross-validation is used to determine $K$. The idea is again to minimize the mean-square distance between the function $B(x)$ and the series approximation, as in the construction of $\hat{\gamma}$. The only difference is that $\gamma$ is estimated for each individual $i$, using only the sample leaving out the $i$th observation. The cross-validation objective function again depends on $B(x)$ which leads to the same problems as mentioned in estimating $\gamma$. The same solution is used again, so the cross-validation objective function is rewritten in such a way that it can easily be estimated, see Newey (1993, p. 438). For each individual this gives an estimate for the function $B(x_i)$. However, $B(x_i)$ is not observed. After rewriting the criterion function as in Newey (1993, p. 438) a feasible cross-validation criterion function is derived that is used to determine $K$.

Note that (23) implies $F(x) \geq 0$. For the approximation to $F$, this is indeed the case for the first model in Table 7. The other two models in Table 7 led to negative estimates for $F(x_i)$ 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 neighbors estimation when we consider more than two waves.

References


