Two-step estimation of panel data models with censored endogenous variables and selection bias

Vella, F.; Verbeek, M.J.C.M.

Published in:
Journal of Econometrics

Publication date:
1999

Citation for published version (APA):
Vella, F., & Verbeek, M. J. C. M. (1999). Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics*, *90*(2), 239-263.

ELSEVIER Journal of Econometrics 90 (1999) 239–263

JOURNAL OF Econometrics

Two-step estimation of panel data models with censored endogenous variables and selection bias

Francis Vella[a], Marno Verbeek[b,*]

[a] *Rutgers University, Department of Economics, New Jersey Hall, New Brunswick, NJ 08903-5055, USA and I.F.S.*

[b] *Center for Economic Studies, K.U. Leuven, Naamsestraat 69, B-3000 Leuven, Belgium and Tilburg University, Tilburg, The Netherlands*

Received 1 February 1996; received in revised form 1 March 1998

Abstract

This paper presents some two-step estimators for a wide range of parametric panel data models with censored endogenous variables and sample selection bias. Our approach is to derive estimates of the unobserved heterogeneity responsible for the endogeneity/selection bias to include as additional explanatory variables in the primary equation. These are obtained through a decomposition of the reduced form residuals. The panel nature of the data allows adjustment, and testing, for two forms of endogeneity and/or sample selection bias. Furthermore, it incorporates roles for dynamics and state dependence in the reduced form. Finally, we provide an empirical illustration which features our procedure and highlights the ability to test several of the underlying assumptions. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C23; C33; C34

Keywords: Two-step estimation; Panel data; Endogenous regressors; Sample selection; Conditional maximum likelihood

1. Introduction

Despite the frequent use of panel data in empirical work there are few suitable estimators for panel data models with sample selection, truncation and limited dependent variables. While maximum likelihood can, under appropriate

* Corresponding author. E-mail: marno.verbeek@econ.kuleuven.ac.be

PII: S0304-4076(98)00043-8

distributional assumptions, provide consistent estimators, its empirical use is hampered by computational complexities such as local maxima and multi-dimensional integrals. This paper proposes some two-step estimators, for a range of parametric panel data models, which avoid these problems. The models comprise a primary equation with an endogeneous explanatory variable, or selection bias, and a reduced form for the endogenous explanator or selection process. Following Heckman (1979), we argue that endogeneity and sample selection bias result from the failure to account for unobserved heterogeneity in the primary equation. We derive estimates of this heterogeneity from the reduced form residuals, to include as additional explanatory variables.

We examine two classes of models. The first is characterized by a primary equation with an uncensored dependent variable and an endogenous censored explanator. This case includes the sample selection model. Given estimates of the reduced form parameters, the primary equation parameters can be estimated by ordinary least squares, based upon a conditional expectation, and we refer to this as *conditional moment estimation*. The second class features a primary equation with a censored dependent variable and an uncensored endogenous explanatory variable. As the primary equation is estimated by maximum likelihood, where the likelihood function corresponds to the conditional density given the endogenous explanatory variable, we assign it an interpretation of *conditional maximum likelihood* (Smith and Blundell, 1986).

Many panel data estimators assume that the endogeneity or selection bias is due to time-invariant individual effects (see, for example, Hausman and Taylor, 1981; Amemiya and MaCurdy, 1986; Honoré, 1992, Honoré, 1993).[1] We, however, also incorporate endogeneity/selectivity through an individual time specific component. This extends cross-sectional estimators (see, for example, Heckman, 1978, Heckman, 1979; Smith and Blundell, 1986; Rivers and Vuong, 1988; Vella, 1993) by separating the individual effects from these individual specific/time effects. We also capture state dependence in the process generating the endogeneity/selection bias. Our approach also encompasses existing panel data procedures for sample selection and attrition bias (see, for example, Ridder, 1990; Nijman and Verbeek, 1992).

Two-step procedures are generally inefficient (see, for example, Newey, 1987) and thus the attraction of our approach, in contrast to maximum likelihood, is its relative computational ease. In some instances, however, our two-step estimator is asymptotically efficient within a limited information framework (LIML). Our method provides initial consistent estimators for a LIML

[1] A recent exception is Kyriazidou (1997) who provides a semi-parametric estimator for the panel data sample selection model in which the selection bias also operates through time-varying effects.

approach so that asymptotically efficient estimators can be obtained in one iteration.[2]

The following two sections consider conditional moment and conditional maximum likelihood estimation, respectively. Section 4 presents an empirical example, featuring the wage–hours relationship, which illustrates a non-conventional form of sample selection bias where the endogenous explanatory variable, which is also the basis of the selection rule, enters the conditional mean non-linearly. This empirical example illustrates our procedure and highlights how many of our assumptions can be tested. Concluding comments are contained in Section 5.

2. Conditional moment estimation

Consider the following model where the parameters of Eq. (1) are of primary focus while Eq. (2) is the reduced form for the explanatory variable which is endogenous and/or the basis of the selection rule. The censoring and selection rules are in Eqs. (3) and (4):

$$y_{it}^* = m_1(x_{it}, z_{it}; \theta_1) + \mu_i + \eta_{it}, \tag{1}$$

$$z_{it}^* = m_2(x_{it}, z_{i,t-1}; \theta_2) + \alpha_i + v_{it}, \tag{2}$$

$$z_{it} = h(z_{it}^*; \theta_3), \tag{3}$$

$$\begin{aligned} y_{it} &= y_{it}^* \quad \text{if } g_t(z_{i1}, \dots, z_{iT}) = 1, \\ &= 0 \text{ (unobserved)} \quad \text{if } g_t(z_{i1}, \dots, z_{iT}) = 0, \end{aligned} \tag{4}$$

where i indexes individuals ($i = 1, \dots, N$) and t indexes time ($t = 1, \dots, T$); y_{it}^* and z_{it}^* are latent endogenous variables with observed counterparts y_{it} and z_{it}; m_1 and m_2 denote general functions characterized by the unknown parameters in θ_1 and θ_2, respectively. This allows a non-linear relationship between z_{it} and y_{it}^*. The mapping from the latent to the observed form is through the censoring functions h and g_t noting that the former may depend on the unknown parameter vector θ_3. The function g_t indicates that y_{it} is only observed for certain values of $z_{i1}, \dots, z_{iT}$. This includes sample selection where y_{it} is only observed if, for example, $z_{it} = 1$ or, alternatively in the balanced subsample case, if $z_{i1} = \cdots = z_{iT} = 1$.

The equations' errors comprise random individual effects, μ_i and α_i, and random individual specific time effects, η_{it} and v_{it}, which are assumed to be

[2] A potentially computationally attractive alternative is simulate maximum likelihood, in which the integrals in the log-likelihood function are replaced by simulators (see Gourieroux and Monfort, 1993). Nevertheless, the use of consistent starting values reduces computational costs substantially.

independent across individuals. Denote $\varepsilon_{it} = \mu_i + \eta_{it}$; $u_{it} = \alpha_i + v_{it}$; and u_i as the T vector of u_{it}'s for individual i. Writing $X_i = [x_{i1}, \dots, x_{iT}]'$, we assume

$$u_i|_{X_i} \sim N.I.D.(0, \sigma_\alpha^2 \iota\iota' + \sigma_v^2 I), \tag{5}$$

$$\mathrm{E}\{\varepsilon_{it} \mid X_i, u_i\} = \tau_1 u_{it} + \tau_2 \bar{u}_i, \tag{6}$$

where ι is a T vector of ones; $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$; and τ_1 and τ_2 are unknown constants. Eq. (5) imposes normality and a strict error components structure on the reduced form and excludes any form of autocorrelation in v_{it}. Eq. (6) allows for heteroskedasticity and autocorrelation in η_{it} but imposes the strict exogeneity of x_{it}.

The model features a potential role for state dependence in the reduced form. This ensures that the error components do not incorrectly capture the dynamics which should be attributed to lagged dependent variables. Unfortunately, we are unable to incorporate dynamics operating through the lagged dependent variable in the primary equation.[3] To estimate this model we first consider condition (1) on the T vector of outcomes, $z_i = (z_{i1}, \dots, z_{iT})'$, and $g_t(z_i) = 1$, to get

$$\mathrm{E}\{y_{it} \mid X_i, z_{i0}, z_i\} = m_1(x_{it}, z_{it}, \theta_1) + \mathrm{E}\{\varepsilon_{it} \mid X_i, z_{i0}, z_i\}. \tag{7}$$

If m_2 does not depend on $z_{i,t-1}$, and sample selection only depends on the current z_{it}, it is possible to condition upon z_{it} only, rather than z_i, and base estimators on the corresponding conditional moments (see Wooldridge, 1995). In this case z_{i0} drops from the conditioning set.

Eq. (6) implies that the conditional expectation of ε_{it} in Eq. (7) is a linear function of the conditional expectation of u_{it}. To derive this latter expectation we use:

$$\mathrm{E}\{u_{it} \mid X_i, z_{i0}, z_i\} = \int [\alpha_i + \mathrm{E}\{v_{it} \mid X_i, z_{i0}, z_i, \alpha_i\}] f(\alpha_i \mid X_i, z_{i0}, z_i)\, \mathrm{d}\alpha_i, \tag{8}$$

where $f(\alpha_i \mid X_i, z_{i0}, z_i)$ denotes the conditional density of α_i. The conditional expectation of v_{it} given X_i, z_{i0}, z_i and α_i is the cross-sectional generalized residual (see Gourieroux et al., 1987) from Eq. (2) as, conditional on α_i, the errors from Eq. (2) are independent across observations. The conditional distribution of α_i can be derived using the result[4]

$$f(\alpha_i \mid X_i, z_{i0}, z_i) = \frac{f(z_i, z_{i0} \mid X_i, \alpha_i) f(\alpha_i)}{f(z_i, z_{i0} \mid X_i)}, \tag{9}$$

[3] Arellano et al. (1997) present an estimator that allows for the lagged *latent* dependent variables to enter both equations linearly. This approach is restricted to models with tobit types of censoring.

[4] We use $f(.)$ as generic notation for any density/mass function.

where we have used that α_i is independent of X_i and

$$f(z_i, z_{i0} \mid X_i) = \int f(z_i, z_{i0} \mid X_i, \alpha_i) f(\alpha_i)\, \mathrm{d}\alpha_i \tag{10}$$

is the likelihood contribution of individual i in Eq. (2). Finally,

$$f(z_i, z_{i0} \mid X_i, \alpha_i) = \left[\prod_{t=1}^{T} f(z_{it} \mid X_i, z_{i,t-1}, \alpha_i) \right] f(z_{i0} \mid X_i, \alpha_i), \tag{11}$$

where $f(z_{it} \mid X_i, z_{i,t-1}, \alpha_i)$ has the form of the likelihood contribution in the cross-sectional case.

If we assume that $f(z_{i0} \mid X_i, \alpha_i)$ does not depend upon α_i and any of the other error components, z_{i0} is strictly exogenous and $f(z_{i0} \mid X_i, \alpha_i) = f(z_{i0} \mid X_i)$. Thus we can condition upon z_{i0} in Eqs. (10) and (11) and obtain valid inferences neglecting its distribution. In general, however, we require an expression for the distribution of the initial value conditional on the exogenous variables and α_i. For most applications, the only practical solution will be a methodology, suggested by Heckman (1981) for the random effects probit model, in which the reduced form for z_{i0} is approximated using all pre-sample information on the exogenous variables. Apart from its dependence upon the unobserved component α_i, the specification of $f(z_{i0} | X_i, \alpha_i)$ can be tested separately from the rest of the model (see Labeaga, 1996).

The unknown parameters in Eq. (2) can be estimated consistently by maximum likelihood under the usual regularity conditions. With these values we can estimate Eq. (8), which requires an expression for the likelihood contribution in an i.i.d. context, the corresponding generalized residual, and the numerical evaluation of two one-dimensional integrals.[5] The estimate of Eq. (8) and its average over time provide two additional terms to be included in the primary equation, with coefficients τ_1 and τ_2. These parameters can be estimated jointly with θ_1 in the second step from conditional moment restrictions such as least squares based on Eq. (7). Under the null hypothesis of no endogeneity $\tau_1 = \tau_2 = 0$ standard errors can be computed in the usual way. Consequently, the standard Wald test of the significance of the additional terms is an endogeneity test. In general, standard errors should be adjusted for heteroskedasticity, autocorrelation, and for the estimation of the correction terms. This is discussed in Appendix A. In the sample selection model the second step is estimated using the subsample based upon $g_t(z_i) = 1$ and the t-statistics on the correction terms provide tests for sample selection bias.

[5] Evaluation of these integrals is straightforward using, for example, results from Butler and Moffitt (1982).

The model, in theory, can be estimated by maximum likelihood if we complement Eqs. (5) and (6) with additional distributional assumptions. If all error components are assumed to be homoskedastic and jointly normal, excluding autocorrelation in the time-varying components, it follows that Eq. (6) holds with $\tau_1 = \sigma_{\eta v}/\sigma_\eta^2$ and $\tau_2 = T(\sigma_{\mu\alpha} - \sigma_{\eta v}\sigma_\alpha^2/\sigma_\eta^2)/(\sigma_v^2 + T\sigma_\alpha^2)$, where $\sigma_{\eta v}$ and $\sigma_{\mu\alpha}$ denote the covariances between the corresponding terms. This shows that τ_2 is non-zero even when the individual effects α_i and μ_i are uncorrelated. In contrast, the two-step approach readily allows for heteroskedasticity and autocorrelation in the primary equation. Moreover, the assumption in Eq. (6) can easily be relaxed to, for example,

$$\mathrm{E}\{\varepsilon_{it}\,|\,u_i\} = \lambda_{1t}u_{i1} + \lambda_{2t}u_{i2} + \cdots + \lambda_{Tt}u_{iT}. \tag{12}$$

By altering Eq. (6) our approach can be extended to multiple endogenous regressors or sample selection rules. With two endogenous variables, $z_{1,it}$ and $z_{2,it}$, say, with reduced form errors $u_{1,it}$ and $u_{2,it}$, respectively, Eq. (6) is replaced by

$$\mathrm{E}\{\varepsilon_{it}\,|\,u_{1,i}, u_{2,i}\} = \tau_{11}u_{1,it} + \tau_{12}\bar{u}_{1,i} + \tau_{21}u_{2,it} + \tau_{22}\bar{u}_{2,i}. \tag{13}$$

Computation of the generalized residuals, however, now requires the evaluation of $\mathrm{E}\{u_{j,it}\,|\,X_i, z_{1,i}, z_{2,i}\}$ for $j = 1, 2$. Unless $z_{1,i}$ and $z_{2,i}$ are independent, conditional upon X_i, the required expressions are different from those obtained from Eqs. (8) and (9) and generally involve multi-dimensional numerical integration.

Many of the above assumptions can be tested empirically. While relaxing normality in the reduced form is computationally difficult, except in special cases for the h function, it is possible to test for departures from normality. It is also possible to test for serial correlation and heteroskedasticity using conditional moment tests. Also, we assume that the variables in X_i are strictly exogenous. This assumption excludes lagged dependent variables in the primary equation as well as feedback from lagged values of y to current x's. If components of X_i are not strictly exogenous they should be included in z and excluded from the reduced form. The conditional moment restriction given in Eq. (7) would typically then be changed to

$$\mathrm{E}\{(y_{it} - \mathrm{E}\{y_{it}\,|\,X_i, z_i, z_{i0}\})\zeta(x_{it}, z_{it})\} = 0 \tag{14}$$

for a vectorial instrument function ζ, depending upon the functional form of m_1. Consequently, strict exogeneity of x_{it} could be tested through, for example, the following moment condition:

$$\mathrm{E}\{(y_{it} - \mathrm{E}\{y_{it}\,|\,X_i, z_i, z_{i0}\})x_{i,t+1}\} = 0. \tag{15}$$

That is, if the last period's values of y influence current x's the strict exogeneity condition would be violated.

An important issue is identification. As in the standard non-linear regression model we require that the matrix of first derivatives of m_j with respect to θ_j, evaluated at the observation points, has full column rank ($j = 1, 2$). Then, non-linearities in m_1 or m_2, or a non-linear mapping from the reduced form variables to the correction terms will identify the model. However, in instances where we do not wish to rely on such an identification scheme we need an exclusion restriction in m_1 for every endogenous variable. If the endogeneity is restricted to the time-invariant component ($\tau_1 = 0$) no exclusion restrictions are required as, in the spirit of Hausman and Taylor (1981), the time variation in the exogenous regressors is exploited.

The contribution of our approach is the following. First, it extends many available cross-sectional estimators (see, for example Heckman, 1978; Heckman, 1979; Vella, 1993) by exploiting the panel nature of the data to isolate the form of heterogeneity responsible for the endogeneity/selection bias. Second, while our treatment of the sample selection model is not new (see Ridder, 1990; Nijman and Verbeek, 1992), we generalize it to a wide range of selection rules. We also extend the panel data dummy endogenous regressor model in Vella and Verbeek (1998) by allowing for other forms of censored endogenous regressors. Finally, our estimator represents first attempt to introduce dynamics into panel data models involving general forms of sample selection and/or censored endogenous regressors. This is particularly important as it exploits the panel nature of the data and enables isolation of the individual effects from state dependence.

Our analysis also indicates how the estimation procedures proposed by Wooldridge (1995) for the sample selection model, can be applied to more general specifications. For example, one can estimate the primary equation while allowing for fixed individual effects. This relaxes Eq. (6) to

$$E\{\varepsilon_{it} \,|\, X_i, u_i\} = \tau u_{it} + \xi_i, \tag{16}$$

for some ξ_i which may be a function of X_i. If $E\{u_{it} \,|\, X_i, z_{i0}, z_i\} = u_{it}$, as for tobit-type sample selection rules, the resulting two-step estimator is essentially the same as that in Wooldridge (1995).[6] Wooldridge also considers assumptions of the form

$$E\{\varepsilon_{it} \,|\, X_i, u_{it}\} = \tau_t u_{it} + X_i \psi. \tag{17}$$

As the conditioning set in Eq. (17) is smaller than in Eq. (16), u_{it} is no longer strictly exogenous in the second step and using a fixed effects or GLS estimator is generally inconsistent (for fixed T). Thus a correlated random effects approach (see Chamberlain, 1980) is required where the second step would include all

[6] Wooldridge only imposes that $u_{it} \,|\, X_i \sim N.I.D.(0, \sigma_t^2)$ for each t, leaving the temporal dependence in u_{it} unrestricted. The reduced form is estimated cross-section by cross-section. Clearly, this does not allow for dynamics in the reduced form specification.

exogenous variables from all periods as additional regressors. While Wooldridge restricts attention to sample selection models with tobit- or probit-type censoring, the approach would be more generally valid. For general h functions, $\mathrm{E}\{u_{it} \mid X_i, z_{it}\}$ is the cross-sectional generalized residual from the reduced form model, which can be included as additional regressor. Note that the assumption in Eq. (17) excludes cases where attention is restricted to a balanced sub-panel of the original data (the conditioning set is larger than z_{it}), and when lagged dependent variables appear in the reduced form.

Arellano et al. (1997) also discuss estimation of panel data models which allow for lagged latent dependent variables in both equations, though only in a linear additive way. They derive orthogonality conditions, based on first-differencing the primary equation, that can be exploited in estimation provided there exist subsets of the data for which the latent variables are observed. Obviously, this seriously restricts the form of the censoring function h, and the functions m_1 and m_2.

3. Conditional maximum-likelihood estimation

We now consider where the dependent variable in the primary equation is censored while the endogenous explanatory variable is fully observed. The model has the following general form:

$$y_{it}^* = m_1(x_{it}, z_{it}; \theta_1) + \mu_i + \eta_{it}, \tag{18}$$

$$z_{it}^* = m_2(x_{it}, z_{i,t-1}; \theta_2) + \alpha_i + v_{it}, \tag{19}$$

$$y_{it} = h(y_{it}^*; \theta_3), \tag{20}$$

$$z_{it} = z_{it}^*, \tag{21}$$

where h is a function mapping the latent y_{it}^* into the observed y_{it}. The assumptions on the unobservables are stronger than before and given by

$$\left.\begin{pmatrix} \mu_i \iota + \eta_i \\ \alpha_i \iota + v_i \end{pmatrix}\right|_{X_i} \sim N.I.D.\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\mu^2 \iota\iota' + \sigma_\eta^2 I & \sigma_{\mu\alpha}\iota\iota' + \sigma_{\eta v} I \\ & \sigma_\alpha^2 \iota\iota' + \sigma_v^2 I \end{pmatrix}\right) \tag{22}$$

which implies Eq. (6) with $\tau_1 = \sigma_{\eta v}/\sigma_\eta^2$ and $\tau_2 = T(\sigma_{\mu\alpha} - \sigma_{\eta v}\sigma_\alpha^2/\sigma_\eta^2)/(\sigma_v^2/T\sigma_\alpha^2)$.

When $y_{it} \neq y_{it}^*$ the model's parameters cannot be estimated by exploiting the conditional expectations of the random components only and the conditional moment approach is not applicable. However, as the endogenous explanator is fully observed the conditional distribution of the error terms in Eq. (18) given z_i is still normal and the error components structure is preserved. Thus we can estimate the model in Eqs. (18) and (20) conditional on the estimated parameters from Eq. (19) using the random effects likelihood function, after making appropriate adjustments for the mean and noting that the variances now reflect the conditional variances.

Write the joint density of $y_i = (y_{i1}, \dots, y_{iT})'$ and z_i given X_i as[7]

$$f(y_i \,|\, z_i, X_i; \theta^{(1)}, \theta^{(2)}) f(z_i \,|\, X_i; \theta^{(2)}), \tag{23}$$

where $\theta^{(1)}$ denotes $(\theta_1, \theta_3, \sigma_\mu^2, \sigma_\eta^2, \sigma_{\mu\alpha}, \sigma_{\eta v})$ and $\theta^{(2)}$ denotes $(\theta_2, \sigma_\alpha^2, \sigma_v^2)$. We first estimate $\theta^{(2)}$ by maximizing the marginal likelihood function of the z_i's. Subsequently, the conditional likelihood function

$$\prod_i f(y_i \,|\, z_i, X_i; \theta^{(1)}, \hat{\theta}^{(2)}) \tag{24}$$

is maximized with respect to $\theta^{(1)}$. The conditional distribution of y_i given z_i is multivariate normal with an error components structure. The conditional expectation can be derived directly from Eq. (6), substituting $u_{it} = z_{it} - m_2(x_{it}, z_{i,t-1}; \theta_2)$, while the covariance structure corresponds to that of $v_{1i} + v_{2,it}$, where v_{1i} and $v_{2,it}$ are zero mean normal variables with zero covariance and variances

$$\sigma_1^2 := V\{v_{1i}\} = \sigma_\eta^2 - \sigma_{\eta v}^2 \sigma_v^{-2}, \tag{25}$$

$$\sigma_2^2 := V\{v_{2,it}\} = \sigma_\mu^2 - \frac{T\sigma_{\mu\alpha}^2 \sigma_v^2 + 2\sigma_{\mu\alpha}\sigma_{\eta v}\sigma_v^2 - \sigma_{\eta v}^2 \sigma_\alpha^2}{\sigma_v^2(\sigma_v^2 + T\sigma_\alpha^2)}. \tag{26}$$

These follow from straightforward matrix manipulations and show that the error components structure is preserved and the conditional likelihood function of Eqs. (18) and (20) has the same form as the marginal likelihood function without endogenous explanatory variables.[8] The asymptotic variance of the conditional ML estimator, and the efficiency loss compared to the LIML estimator, is provided in Appendix A.

The conditional maximum likelihood estimator can be easily extended to account for multiple endogenous variables. For two endogenous regressors, for example, this implies the imposition of Eq. (13), the conditional expectation of which is easily obtained as all endogenous regressors are continuously observed. Even if the reduced form errors of the endogenous regressors are correlated,

[7] When Eq. (19) is dynamic with an exogenous initial value z_{i0}, Eq. (23) is valid if z_{i0} is included in X_i. When the initial value is endogenous, we need to include z_{i0} in z_i.

[8] The algebraic manipulations are simplified if σ_1^2 and σ_2^2 replace the unconditional variances σ_η^2 and σ_μ^2 in $\theta^{(1)}$. In this case, estimates for the latter two variances are easily obtained in a third step from the estimates from the first stage for σ_v^2 and σ_α^2, and the estimated covariances from the mean function, using the equalities in Eqs. (25) and (26).

provided they are characterized by an error components structure it can be shown that the conditional distribution of $\mu_i + \eta_{it}$ also has an error components structure. Time-specific heteroskedasticity in η_{it} does not affect the conditional expectations and can be incorporated by having σ_1^2 vary over time. The model can also be estimated, along the lines suggested above, over subsets of the data chosen on the basis of z_{it}.

In general the conditional maximum-likelihood estimator cannot be employed when $z_{it} \neq z_{it}^*$. Thus the family of sample selection models considered in the conditional moment section cannot be estimated by conditional maximum likelihood. One exception is when the primary equation is estimated over the subsample of individuals that have $z_{is} = z_{is}^*$, for all $s = 1, \dots, T$.[9] This follows from the result that the error components structure is preserved when the reduced form dependent variables are observed.

Once again it is important to consider identification. In these models there is no non-linearity induced in the correction terms, but the non-linearity of m_1 or m_2 will identify the model. In the linear case, or if one does not want to rely on non-linearities for identification, exclusion restrictions are required. More explicitly, for each endogenous explanatory variable we need one exclusion restriction in the primary equation, unless, as before, the endogeneity can be restricted to be related to the time-invariant components only ($\sigma_{\eta v} = 0$).

This conditional maximum-likelihood approach generalizes the cross-sectional estimators of Smith and Blundell (1986) and Rivers and Vuong (1988) to panel data. As in the previous section the use of panel data allows the separation of the error into two components and the incorporation of dynamics in the reduced form. By excluding dynamics and assuming that there is a single-error component one obtains the models considered by Smith and Blundell (1986) and Rivers and Vuong (1988) as special cases.

4. Empirical example

We now provide an empirical example estimating the impact of weekly hours worked on the offered hourly wage rate while accounting for the endogeneity of hours. This issue has attracted attention in the labor economics literature (see, for example, Moffitt, 1984; and Biddle and Zarkin, 1989). The model has the form

$$w_{it} = x'_{1,it}\beta_1 + x'_{2,it}\beta_2 + m(\text{hours}_{it}; \beta_3) + \mu_i + \eta_{it}, \tag{27}$$

$$\text{hours}_{it}^* = x'_{3,it}\theta_1 + \text{hours}_{i,t-1}\theta_2 + \alpha_i + v_{it}, \tag{28}$$

[9] A similar argument is exploited in Arellano et al. (1997).

$$\text{hours}_{it} = \text{hours}^*_{it} \quad \text{if hours}^*_{it} > 0, \tag{29}$$

$$\text{hours}_{it} = 0,\ w_{it} \text{ not observed if hours}^*_{it} \leqslant 0,$$

where w_{it} represents the log of the hourly wage of individual i in time period t; $x_{1,it}$ and $x_{3,it}$ are variables representing individual characteristics; $x_{2,it}$ are characteristics of the individual's work place; hours$^*_{it}$ and hours$_{it}$ represent desired and observed number of hours worked, respectively; m denotes a polynomial of known length with unknown coefficients β_3, while $\beta' = (\beta'_1, \beta'_2, \beta'_3)$ and $\theta' = (\theta'_1, \theta'_2)$ are parameters to be estimated. The assumptions on the unobservables are given in Eqs. (5) and (6) above. This example highlights several aspects of our approach. First, as the wage equation is estimated over the sub-sample reporting positive hours it illustrates our ability to control for selection bias. Second, the inclusion of the correction terms enables the identification of the form of the endogeneity while their inclusion also corrects for the endogeneity of experience via past labor supply. Finally, the inclusion of a lagged dependent variable in the reduced form isolates the role of dynamics and state dependence.

Note that the model is not easily estimated by existing procedures. First, as the wage equation is only estimated over those reporting positive hours it is not possible to use instrumental variables. Second, as the reduced form includes a role for dynamics it is not appropriate to employ cross-sectional procedures nor apply panel data estimators which ignore the presence of state dependence. Third, as the primary equation is non-linear in hours, the approach of Arellano et al. (1997) cannot be employed. Fourth, even with stronger distributional assumptions (as given in Eq. (22)), non-linearities due to both the sample selection problem and the non-linear appearance of hours in the wage equation, substantially complicate analytical expressions for the joint likelihood function (see Davidson and Mackinnon, 1993, Section 18.7). As two-dimensional numerical integrals are involved, maximum likelihood does not appear to be a viable approach. Moreover, the number of parameters in our most general specification equals 79, so that the risk of computational problems will be high for any one-step procedure.

To estimate the model we employ data for young females from the National Longitudinal Survey (Youth Sample) for the period 1980–1987. For the period examined there were 2300 observations for each of the eight years. From these 18,400 total observations there were 12,039 observations reporting positive hours of work in a given period.[10] The summary statistics are reported in Table 1.

[10] Only 481 women report positive hours of work over all eight years.

Table 1
Variable definitions and means

SCHOOL	Years of schooling	11.91	1.55	11.67
HISP	Dummy, 1 if hispanic	0.126	0.331	0.132
BLACK	Dummy, 1 if black	0.189	0.189	0.208
AGE	Age in years	24.08	2.73	23.89
MAR	Dummy, 1 if married	0.386	0.487	0.415
WAGE	Log of real hourly wage	1.285	0.640	—
EXPER	Actual exp. in yr.	5.22	2.45	4.95
EX2	Actual exp. squared	33.27	29.09	30.33
HLTH	Dummy, 1 if health disability	0.038	0.191	0.047
HOURS	Hours of work per week	33.98	11.82	—
UNION	Dummy, 1 if wage set by collective bargaining	0.168	0.374	—
RUR	Dummy, 1 if rural area	0.181	0.385	0.192
S	Dummy, 1 if lives in South	0.370	0.483	0.370
NC	Dummy, 1 if Northern Central	0.240	0.427	0.248
NE	Dummy, 1 if North East	0.204	0.403	0.187
Industry dummies				
PUB	Public sector	0.048	0.215	—
AG	Agricultural	0.008	0.087	—
MIN	Mining	0.003	0.057	—
CON	Construction	0.008	0.087	—
MAN	Manufacturing	0.159	0.366	—
TRA	Transportation	0.040	0.195	—
TRAD	Trade	0.237	0.425	—
FIN	Finance	0.097	0.296	—
BUS	Business and repair service	0.045	0.206	—
PER	Personal service	0.061	0.240	—
ENT	Entertainment	0.013	0.115	—
PRO	Professional and related services	0.281	0.450	—

Note: Averages in column 3 are based on observations over 2300 women over eight years (1980–1987); averages and standard deviations in columns 1 and 2 are for the subsample of observations with positive hours only (12,039 cases).

We estimate three variants of the random effects tobit model in Eq. (28). To highlight the differences which may arise from misspecifying the dynamics we estimate the model with and without the lagged dependent variable. Given the presence of the individual effects α_i it is unlikely that one can validly assume that hours worked in the first period are truly exogenous. Therefore, the dynamic model is estimated either treating initial hours worked as exogenous, which ignores the initial conditions problem, or modelling the initial value in the spirit of Heckman (1981). For this last approach, we approximated the reduced form for hours_{io} by a tobit function using all pre-sample information on the exogenous variables.[11] The results of all three specifications are presented in Table 2. Column 1 represents the specification without the lagged dependent variable while Columns 2 and 3 include it. The specification does not include any work-related variables as these characteristics are not recorded for the non-participants and their inclusion, while assigning values of zero for the non-participants, would produce a spurious statistical relationship between these characteristics and hours of work.

Several points are worth noting. First, the parameter estimates in all columns are generally in keeping with expectations and most variables have a statistically significant impact on hours worked. Second, the impact of lagged hours of work, in Columns 2 and 3, is highly significant and indicates the presence of 'state' dependence. The inclusion of this variable has a relatively small impact on the coefficients of the other variables although it generally reduces their magnitude. This is expected as their effect now partially operates through lagged hours. Third, the coefficients on the time dummies indicate an increasing trend in hours worked over the period examined in both specifications. Comparing the columns with the exogenous [A] and endogenous [B] treatment of the initial value, we see that exogeneity of initial hours of work is strongly rejected. The parameter ξ measures to what extent the individual-specific heterogeneity enters the initial equation's error, given by $\xi\alpha_i + v_{i0}$, and does not differ significantly from unity. We found no statistical evidence that ξ varies with age.[12] Despite the rejection of exogeneity, the estimates of the tobit coefficients do not appear to be very sensitive with respect to the treatment of the initial condition. A final point is related to the contribution of the error components to the total variance. The model which omits a lagged dependent variable attributes approximately 39% of the total variance to the individual effects. The specifications in columns 2 and 3, where the contribution of the individual effects reduces to approximately 13 percent, indicate that this may be due to the failure to account for dynamics.

We noted above that many of the assumptions of the model are testable. The first we consider is normality of the two error components. We test this via Lagrange Multiplier tests using a parameterization suggested by Ruud (1984). Unfortunately, these tests reject normality of both the individual specific component and the idiosyncratic component for each of the specifications.[13] This is

[11] More details of treatment of the initial value are provided in Appendix B. The estimation results for the initial period are not presented here.

[12] A likelihood ratio test against the alternative that ξ is a linear function of age produced a test statistic of 0.02.

[13] The LM test statistics (χ^2_2) for testing normality of α_i and η_{it} in Column 3 are 119.52 and 535.28, respectively. For Column 2 they are 244.64 and 436.60, while the corresponding values for the Column 1 specification are 297.6 and 814.5.

Table 2
Maximum likelihood estimation results hours of work equation

Variable	Static model	Dynamic model [A]	Dynamic model [B]
Dummy 1981	1.816	4.549	4.483
	(0.777)	(0.646)	(0.650)
Dummy 1982	1.697	3.457	3.436
	(0.845)	(0.736)	(0.739)
Dummy 1983	3.807	5.505	5.476
	(1.017)	(0.817)	(0.822)
Dummy 1984	6.561	7.205	7.238
	(1.241)	(0.923)	(0.931)
Dummy 1985	8.592	8.029	8.124
	(1.475)	(1.036)	(1.047)
Dummy 1986	11.849	10.300	10.418
	(1.679)	(1.121)	(1.136)
Dummy 1987	14.313	11.016	11.188
	(1.922)	(1.245)	(1.268)
Age	13.368	6.073	6.297
	(0.796)	(0.988)	(0.982)
Age-squared	− 0.291	− 0.139	− 0.143
	(0.015)	(0.020)	(0.020)
Married	− 5.682	− 4.737	− 4.946
	(0.384)	(0.398)	(0.396)
Black	− 6.419	− 4.016	− 4.154
	(0.940)	(0.578)	(0.591)
Hispanic	− 0.510	− 0.349	− 0.350
	(1.113)	(0.679)	(0.694)
Rural	− 1.893	− 1.289	− 1.434
	(0.534)	(0.495)	(0.493)
Health	− 3.870	− 4.760	− 4.783
	(0.285)	(0.752)	(0.749)
School	4.172	2.481	2.546
	(0.285)	(0.168)	(0.172)
Lagged hours	—	0.616	0.594
		(0.014)	(0.013)
σ_α^2	259.09	55.406	60.81
	(12.36)	(4.53)	(4.74)
σ_v^2	402.87	398.32	395.51
	(4.93)	(5.78)	(5.69)
ξ		0	0.939
			(0.099)
Loglikelihood	− 59,856.13	− 58,731.91	
with initial period		− 66,074.19	− 66,019.24

Note: Models also include regional dummies. Dynamic Model [A] assumes exogenous initial labor supply, while [B] employs a specification for the initial distribution.

not surprising given the large number of observations. However, as this example is primarily illustrative we continue despite this rejection.

Before focusing on the primary equation consider our exclusion restrictions. First, given the form of the correction terms we are unable to estimate the model unless we exclude at least one variable from the wage equation which appears in the hours equation. The variables we employ are the health measure and lagged hours. Health is excluded on the basis that while health will influence one's ability to seek employment it is increasingly difficult for employers to offer different wage levels on the basis of an individual's health. The lagged hours variable is excluded on the basis that we believe that the variable has no direct effect but rather operates through its impact on current hours. As this implies that the model is overidentified, we can test this latter restriction.

We estimate the wage equation for the subsample of working women, specifying m as a fourth order polynomial. The order of the polynomial is guided by the results of Newey et al. (1998) which provides a semi-parametric estimate of the wage/hours profile. Using this result as a starting point we were unable to identify a dominating alternative specification. The OLS results reported in Column 1 of Table 3, without corrections for the potential endogeneity of hours, reveal a significant non-linear relationship between the number of weekly hours worked and the weekly wage rate. The remaining coefficients in this column are reasonable in sign and magnitude although the coefficient reflecting the returns to schooling is rather high. As it is well known that wage inequality increased over this period it is likely that the error terms are heteroskedastic. The estimated standard errors in Table 3 allow for unknown forms of time-related heteroskedasticity.

Column 2 of Table 3 reports the results from estimating the wage equation while correcting for the endogeneity of hours and sample selection through the inclusion of correction terms based upon the static hours model. The results in Columns 3 and 4 incorporate a role for dynamics in the reduced form hours equation, treating the initial value as exogeneous or endogeneous, respectively. The inclusion of the correction terms simultaneously accounts for the endogeneity of the participation decision and hours worked. As the decision to not participate corresponds to a zero value for hours_{it}, the inclusion of the correction terms also account for the selection bias from estimating over the subsample of workers.

A number of features are worth noting from a comparison of the unadjusted estimates, in Column 1, to those from our preferred specification in Column 4.[14] Both correction terms are statistically significant indicating that the two forms

[14] We refer to the specification in Column 4 as preferred in spite of the evidence of non-normality in the reduced form. We choose this specification on the basis of the statistically significant role of lagged hours and the strong rejection of exogeneity of the initial value.

Table 3
Wage equation results

Method	OLS	OLS	OLS	OLS
Correction	No	Yes	Yes	Yes
Hours equation	—	Static	Dynamic	Dynamic
Initial value	—	—	Exogenous	Endogenous
Married	− 0.018	0.043	0.017	0.025
	(0.011)	(0.018)	(0.011)	(0.011)
Black	− 0.147	− 0.103	− 0.133	− 0.124
	(0.015)	(0.022)	(0.015)	(0.015)
Hispanic	− 0.029	− 0.024	− 0.024	− 0.023
	(0.017)	(0.017)	(0.017)	(0.017)
Rural	− 0.116	− 0.102	− 0.111	− 0.109
	(0.014)	(0.015)	(0.014)	(0.04)
School	0.096	0.064	0.080	0.076
	(0.004)	(0.010)	(0.004)	(0.004)
Union	0.116	0.120	0.118	0.118
	(0.013)	(0.013)	(0.013)	(0.013)
Exper	0.037	0.009	− 0.009	− 0.010
	(0.009)	(0.009)	(0.009)	(0.009)
Exper2	− 0.0010	− 0.0005	0.0004	0.0006
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Hours/100	− 5.062	− 4.130	− 4.474	− 4.334
	(1.378)	(1.379)	(1.360)	(1.360)
$(\text{Hours}/100)^2$	31.925	30.509	30.296	30.206
	(6.118)	(6.017)	(6.000)	(6.006)
$(\text{Hours}/100)^3$	− 61.275	− 58.647	− 58.780	− 58.672
	(10.986)	(10.755)	(10.710)	(10.720)
$(\text{Hours}/100)^4$	34.586	33.383	33.742	33.688
	(6.835)	(6.658)	(6.621)	(6.627)
u_{it}	—	− 0.0121	− 0.0084	− 0.0095
		(0.0022)	(0.0006)	(0.0006)
$\bar{u}_i$	—	− 0.0016	0.0145	0.0109
		(0.0087)	(0.0018)	(0.0016)
adj R^2	0.249	0.271	0.279	0.279
Observations	12,039	12,039	12,039	12,039

Note: All specifications also include regional, industry and time dummies.

of endogeneity/selectivity are present. The positive coefficient on the individual effect indicates that the time-invariant unobserved individual effects which increase the number of hours worked also increase the wage level. In constrast, the time varying effects generating the simultaneity of wages to hours appears to increase hours and decrease wages. Note, however, that as the reduced form is estimated by tobit it is possible that the unobserved effects are affecting the probability of participation in addition to the number of hours worked. The inclusion of the correction terms induces a number of changes in the results. First, the time effects, which are not reported, appear to be much stronger in this adjusted equation in comparison to the unadjusted equation. However, this increased time effect is partially due to the reduced experience effects. Second, the inclusion of the correction terms substantially reduces the impact of schooling from almost 10% to a more reasonable 7.6%. Finally, the hours/wage profile appears to have noticeably changed from that implied in the Column 1.

For our preferred specification in column 4 we tested the assumption in Eq. (6) against several alternatives. In particular, we tested whether τ_1 and τ_2 were time-invariant. Both tests resulting in a marginal rejection without noticeable changes in the estimated coefficients of the model. Further, we tested Eq. (6) against a restricted version of Eq. (12) which specifies

$$E\{\varepsilon_{it}\,|\,u_i\} = \tau_1 u_{it} + \lambda_1 u_{i1} + \cdots + \lambda_T u_{iT}.$$

The Wald test statistic for the hypothesis $\lambda_1 = \cdots = \lambda_T$ resulted in a value of 60.52, which is highly significant for a χ^2 distribution with seven degrees of freedom. Despite this rejection, the estimated coefficients were only marginally different from those for our preferred specification in column 4 of Table 3 and we choose not to report them.[15]

Human capital accumulation suggests that lagged hours of work positively affect current wages. As our model conditions upon the entire vector of labor supply, this hypothesis can be directly tested by our two-step procedure by including lagged hours as additional explanatory variables in columns 1 to 4. This resulted in t-ratios of 15.3, 11.2, -0.18 and -3.16, respectively.[16] These results indicate that ignoring the endogeneity of hours or incorrectly imposing a static hours equation, lead to a spurious finding of a statistically significant human capital accumulation effect. Apparently, the effect of lagged hours upon current wages operates through its effect on current hours. This provides support for our exclusion restriction.

Our approach assumes strict exogeneity of the explanatory variables. For time-varying variables this can be tested by including values of the particular exogeneous variable from other time periods in the mean function, and testing for a zero coefficient. This is restricted to time-varying variables as those which are time invariant are imposed to be uncorrelated with the error components in estimation. For the current application union status is a natural candidate of a time-varying regressor which may violate the strict exogeneity assumption.

[15] The estimated coefficients for the hours variables in this specification were -4.339, 30.282, -58.870, 33.831, respectively, implying a very similar wage-hours profile.

[16] The estimated coefficients ranged from -0.004 to 0.006 indicating that the lagged hours effect was small.

We tested its strict exogeneity by including next period's union status in the wage equation. The coefficient on this lead value of the variable had a t-statistic of 7.29 leading one to reject that union status is strictly exogenous. Another candidate is the schooling variable but as this is time invariant in our sample we cannot test its exogeneity. Recall that we allow experience to be endogenous through lagged hours and that age and age-squared are included in the reduced form.

It is also useful to contrast our preferred estimates to those in Column 2 to highlight the effect of including a lagged dependent variable in the reduced form. First, the failure to account for dynamics in Column 2 has led to the inclusion of the incorrect correction terms. This has resulted in an incorrect inference regarding the presence of individual effects. Second, the inclusion of dynamics in the reduced form has resulted in more precise estimates for a number of the regressors. Third, while the estimates across the two columns are generally similar the coefficient capturing the returns to schooling is quite different. This coefficient is expected to be affected by the treatment of the individual effects as it is likely to be sensitive to unobserved ability. Also note that the estimate of this coefficient, in our preferred specification, is notably more precise than its counterpart in Column 2. Finally, despite the differences across Columns 2 and 4 the coefficients on the hours variables appear to be similar.

To examine the effect of weekly hours worked on the wage rate we employ the wage equation estimates to plot the implied relationships. We do this by plotting the hours effect, relative to the sample average of 34 h per week in 1987, against hours worked. The four plots corresponding to the results from columns 1–4 are shown in Fig. 1. Several important points are worth noting. First, all of the profiles indicate that the hourly wage rate is responsive to changes in the hours of weekly work. For each of the profiles the wage rate appears to either initially slightly increase or slightly decrease before showing a notable increase in the range of around 15 to approximately 42 to 45 h. The point at which the wage rate stops increasing depends on the specification. After wages peak they tend to show a dramatic decrease. Note, however, that the profile at higher hours is dominated by relatively few observations.[17] Second, for the values of hours at which the majority of the observations are located, the unadjusted figure clearly reveals less response in wages to changes in hours worked. This suggests that the impact of endogeneity is reducing the estimated hours effect on wages. Third, the adjusted profiles peak to the right of that for the unadjusted equation. While the unadjusted results suggest that hours peak in the 35–40 range the adjusted

[17] The hours distribution in the sample of workers is as follows: 0–10 h (4.4%); 10–20 h (10.7%); 20–30 h (15.0%); 30–40 h (49.9%); 40–50 h (15.6%); 50 + h (4.4%).

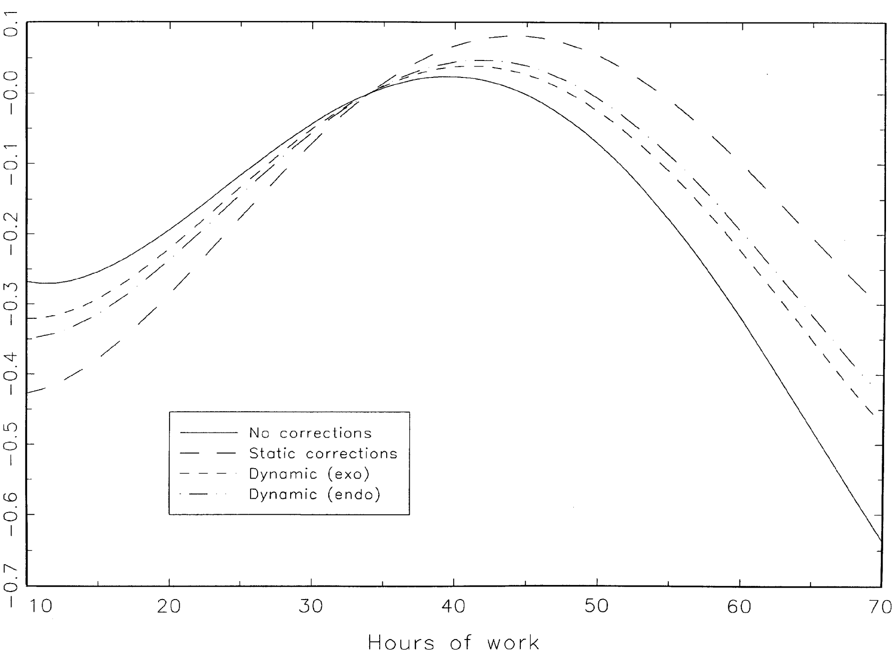

Fig. 1. The Estimated effect of hours on wages, relative to average of 34 hours.

results indicate that the overtime effects are stronger and lead to wages peaking at approximately 42 h per week, for the dynamic specifications, and at approximately 45 h for the static model. Fig. 1 reveals that the profiles for the two dynamic specifications are located between the OLS and adjusted static model. This highlights the importance of incorporating dynamics. In the absence of the lagged dependent variable we are assigning too much importance to the error components. This results in overadjusting for endogeneity. Finally, consider the implied increase in wages over the interval revealed by our preferred specification to feature an increasing effect of hours. Over the hours interval 10–41, comprising 82% of the data, the unadjusted estimates imply a wage increase of 34%. The corresponding increase implied by our preferred specification is 43%, while correcting with a static labor supply equation produces a number of almost 65%.

Ideally we would estimate the model by maximum likelihood to explore the efficiency loss that occurs in this particular example by employing our two step approach. Unfortunately, the complicated form of the likelihood function made such a comparison impractical. While this failure to make efficiency comparisons is disappointing it highlights a major attraction of our approach, namely the ability to estimate models which would otherwise not be estimable. We also considered other methods of obtaining asymptotically efficient standard errors in order to make comparisons. However, due to the complicated nature of the model we were unable to do so. The one piece of information we are able to provide is how much the standard errors increased in the second step by having to estimate the parameters in the first step. We found that, depending on the specification employed, the standard errors increased by up to 10%.

5. Concluding remarks

This paper presents a two-step approach to estimating panel data models with censored endogenous variables and sample selection. In contrast to maximum likelihood estimation our procedure is computationally simple because only one-dimensional numerical integration is required, while a closed-form solution for the second step estimator is available. The cost is a loss of efficiency, which partially depends upon the magnitude of the covariances responsible for the endogeneity/selectivity. Our approach can handle a large variety of models. In particular it allows for truncation, censoring and sample selection. In addition, it is straightforward to extend the current framework to switching regressions models with endogenous switching and observed regimes. Moreover, the two-step method identifies an additional number of parameters that may have economic appeal, including two potential sources of endogeneity/selectivity and state dependence. Direct test for endogeneity/selectivity are also provided.

Acknowledgements

This paper was partially written while the authors were visitors in the Department of Economics, Research School of Social Sciences and the Department of Statistics, The Faculties at the Australian National University, Canberra, and while Vella was visiting the CentER for Economic Research at Tilburg University. An earlier version of this paper was circulated under the title 'Estimating and Testing Simultaneous Equations Panel Data Models with Censored Endogenous Variables'. Helpful comments by Bertrand Melenberg, Robin Sickles, Jeffrey Wooldridge and seminar participants at the University of Groningen, Australian National University, Texas A&M University, Southern Methodist University, Nuffield College, Tilburg University and the London School of Economics are gratefully acknowledged. We are also grateful for detailed comments from Richard Blundell and two anonymous referees. We alone are responsible for any remaining errors.

Appendix A. Covariance matrix estimation

First consider conditional moment estimation. The estimator is $\sqrt{N}$ consistent and asymptotically normal under weak regularity conditions. The asymptotic covariance matrix can be obtained using the results in Newey (1984). Let $\theta^{(2)}$ denote the parameter vector from the reduced form (2) and (3) and let its $\sqrt{N}$ consistent estimator be given by $\hat{\theta}^{(2)}$ with asymptotic covariance matrix V_2. Furthermore, let $\theta^{(1)}$ denote the vector of parameters characterizing the conditional mean in the second step, i.e:

$$\theta^{(1)} = (\theta_1', \tau')',$$

with $\tau = (\tau_1, \tau_2)'$. In vector notation, write

$$y_i = m_1(x_i, z_i; \theta_1) + Z_i(\theta^{(2)})\tau + e_i,$$

where $Z_i(\theta^{(2)})$ denotes the generated regressors and where e_i is an error term vector corresponding to the conditional distribution of ε_i. Let Ω_i denote the variance of e_i for individual i and let

$$G_i = \frac{\partial m_1(x_i, z_i; \theta_1)}{\partial \theta_1}.$$

Defining:

$$M_N = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}\{[G_i\ Z_i(\theta^{(2)})]'[G_i\ Z_i(\theta^{(2)})]\},$$

$$V_N = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}\{[G_i\ Z_i(\theta^{(2)})]'\Omega_i[G_i\ Z_i(\theta^{(2)})]\},$$

$$D_N = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}\left\{[G_i\ Z_i(\theta^{(2)})]'\frac{\partial Z_i(\theta^{(2)})\theta_2^{(1)}}{\partial \theta^{(2)\prime}}\right\},$$

the asymptotic covariance matrix of the second step estimator for $\theta^{(1)}$ is given by

$$V_1 = \lim_{N\to\infty} M_N^{-1}(V_N + D_N V_2 D_N')M_N^{-1},$$

which can be consistently estimated by replacing expectations with sample moments and unknown parameters by their estimators. The second part within brackets for V_1 is due to the generated regressors problem and equals zero if $\tau = 0$. A heteroskedasticity- and autocorrelation-consistent estimator for V_N is given by

$$\hat{V}_N = \frac{1}{N}\sum_{i=1}^{N} [\hat{G}_i\ Z_i(\hat{\theta}^{(2)})]'\hat{e}_i\hat{e}_i'[\hat{G}_i\ Z_i(\hat{\theta}^{(2)})],$$

where $\hat{e}_i$ is the T-dimensional vector of residuals.

Finally, if conditioning upon z_i implies that certain observations are excluded from the second stage estimation, the dimensions of all vectors and matrices should be adjusted to include only those observations used in estimation.

Now consider conditional maximum likelihood estimation. Let $\hat{\theta}^{(2)}$ denote the (marginal) ML estimator for $\theta^{(2)}$ with asymptotic covariance matrix V_2 and $\hat{\theta}^{(1)}$ the conditional ML estimator obtained from maximizing Eq. (24). Let f_1 denote the conditional density of y_i given z_i. Define

$$F_{11} = \mathrm{E}\left\{-\frac{\partial^2 \ln f_1}{\partial\theta^{(1)}\partial\theta^{(1)'}}\right\}, \qquad F_{12} = \mathrm{E}\left\{\frac{\partial^2 \ln f_1}{\partial\theta^{(1)}\partial\theta^{(2)'}}\right\}.$$

Then, it is easily verified from Taylor expansions that

$$\sqrt{N}(\hat{\theta}^{(1)} - \theta^{(1)}) = F_{11}^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial \ln f_1}{\partial\theta^{(1)}} + F_{12}\sqrt{N}(\hat{\theta}^{(2)} - \theta^{(2)})\right] + \mathrm{o_p}(1).$$

Using the result that the two terms in square brackets are asymptotically independent (see Pierce, 1982; Parke, 1986) it follows that the conditional ML estimator is asymptotically normal with covariance matrix:

$$V_1 = F_{11}^{-1}[F_{11} + F_{12}V_2F_{12}']F_{11}^{-1}.$$

F_{11}^{-1} is estimated by standard ML programs. In two cases $\hat{\theta}^{(1)}$ attains the Cramér–Rao lower bound. The first corresponds to the null hypothesis of exogeneity ($\sigma_{\mu\alpha} = \sigma_{\eta v} = 0$) and is characterized by $F_{12} = 0$. The second case is given by

$$J = F_{22} - F_{12}'F_{11}^{-1}F_{12} = 0,$$

where

$$F_{22} = \mathrm{E}\left\{-\frac{\partial^2 \ln f_1}{\partial\theta^{(2)}\partial\theta^{(2)'}}\right\}.$$

Note that J summarizes all information on $\theta^{(2)}$ contained in the conditional distribution f_1. We refer to $J = 0$ as the exactly identified case (see Rivers and Vuong, 1988, for an example). In general, the efficiency loss due to the two-step nature is given by

$$F_{11}^{-1}F_{12}(V_2 - [V_2^{-1} + J]^{-1})F'_{12}F_{11}^{-1}.$$

Appendix B. Treatment of initial conditions

As the hours process for different individuals may have different starting dates, and as some of the explanatory variables are likely to be non-stationary, it is not tractable to derive an exact and estimable form of the marginal distribution of hours_{i0} consistent with the conditional model in Eq. (28) (see Heckman, 1981). Our first approach ignores this problem and assumes that the initial value hours_{i0} can be treated as exogenous, i.e. independent of all error components in the model (including the unobserved heterogeneity (α_i) in the hours process for $t = 1$ onwards). In this case, all inferences are conditional upon hours_{i0}, the marginal distribution of which is not needed for consistency. In our second approach we approximate the distribution of hours_{i0} by a reduced form tobit using presample (1979) information. With different starting dates, as indicated by different levels of age or experience, it can be expected that the tobit parameters are heterogeneous with respect to these two variables. Similarly, the error terms are likely to exhibit heteroskedasticity. The specification of this reduced form is determined, and empirically tested, independently of the rest of the model.

Starting from a linear additive specification with the explanatory variables from Table 2, we experimented with interaction terms and heteroskedasticity. We arrived at a specification for initial hours that included, besides the exogenous variables from Table 2, age-cubed and age interacted with years schooling and the two racial dummies. For this specification there was no statistical evidence of omitted interaction terms or heteroskedasticity related to age. Normality, however, was rejected. This is similar to the finding for the other periods. In the next step, this specification for the initial period was estimated jointly with the dynamic model for the other periods. This requires one to specify how the error terms are related. We chose to allow for different error variances for this initial period. In particular, we modelled the initial error as $\xi\alpha_i + v_{i0}$, where v_{i0} is i.i.d. normal with unrestricted variance and ξ is an unknown parameter.

References

Amemiya, T, MaCurdy, T, 1986. Instrumental-variable estimation of an error-components model. Econometrica 54, 869–881.

Arellano, M., Bover, O., Labeaga, J.M., 1997. Autoregressive models with sample selectivity for panel data. Working paper 9706, CEMFI, Madrid.

Biddle, J, Zarkin, G, 1989. Choice among wage-hours packages: An empirical investigation of male labor supply. Journal of Labor Economics 7, 415–437.
Butler, J, Moffitt, R, 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. Econometrica 50, 761–764.
Chamberlain, G, 1980. Analysis of covariance with qualitative data. Review of Economic Studies 47, 225–238.
Davidson, R, MacKinnon, J.G, 1993. Estimation and Inference in Econometrics. Oxford University Press, Oxford.
Gourieroux, C, Monfort, A, 1993. Simulation-based inference. A survey with special reference to panel data models. Journal of Econometrics 59, 5–33.
Gourieroux, C, Monfort, A, Renault, E, Trognon, A, 1987. Generalized residuals. Journal of Econometrics 34, 5–32.
Hausman, J, Taylor, W, 1981. Panel data and unobservable individual effects. Econometrica 49, 1377–1398.
Heckman J.J., 1978. Dummy endogenous variables in a simultaneous equation system. Econometrica 931–959.
Heckman, J.J, 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.
Heckman, J.J., 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete-time-discrete-data stochastic process. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.
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. Orthogonality conditions for tobit models with fixed effects and lagged dependent variables. Journal of Econometrics 59, 35–61.
Kyriazidou, E, 1997. Estimation of a panel data sample selection model. Econometrica 65, 1335–1364.
Labeaga, J.M., 1996. A dynamic panel data model with limited-dependent variables: An application to the demand for tobacco. Working paper, Universitat Pompeu Fabra.
Moffitt, R, 1984. The estimation of a joint wage-hours labor supply model. Journal of Labor Economics 2, 550–566.
Newey, W, 1984. A method of moments interpretation of sequential estimators. Economics Letters 14, 201–206.
Newey, W, 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. Journal of Econometrics 36, 231–250.
Newey, W., Powell, J., Vella, F., 1998. Nonparametric instrumental variables estimation of triangular simultaneous equations models. Econometrica, forthcoming.
Nijman, T, Verbeek, M, 1992. Nonresponse in panel data: The impact on estimates of a life cycle consumption function. Journal of Applied Econometrics 7, 243–257.
Parke, W.R, 1986. Pseudo maximum likelihood estimation: The asymptotic distribution. Annals of Statistics 14, 355–357.
Pierce, D.A, 1982. The asymptotic effect of substituting estimators for parameters in certain types of statistics. Annals of Statistics 10, 475–478.
Ridder, G., 1990. Attrition in multi-wave panel data. In: Hartog, J., Ridder, G., Theeuwes, J. (Eds.), Panel Data and Labor Market Studies. Elsevier, North-Holland, Amsterdam.
Rivers, D, Vuong, Q, 1988. Limited information estimators and exogeneity tests for simultaneous probit models. Journal of Econometrics 39, 347–366.
Ruud, P.A, 1984. Tests for specification in econometrics. Econometric Reviews 3, 221–242.
Smith, R.J, Blundell, R, 1986. An exogeneity test for a simultaneous equation Tobit model with an application to labor supply. Econometrica 54, 679–685.

Vella, F, 1993. A simple estimator for simultaneous models with censored endogenous regressors. International Economic Review 34, 441–457.
Vella, F., Verbeek, M., 1998. Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. Journal of Applied Econometrics, 13, 163–168.
Wooldridge, J, 1995. Selection corrections for panel data models under conditional mean independence assumptions. Journal of Econometrics 68, 115–132.