INVESTIGACIONES ECONOMICAS, vol. XXI(2), 1997, 141-165

ESTIMATING DYNAMIC LIMITED DEPENDENT VARIABLE MODELS FROM PANEL DATA

OLYMPIA BOVER
Banco de España

MANUEL ARELLANO
CEMFI

We propose a simple two-step within-groups estimator for limited dependent variable models, which may include lags of the dependent variable, other endogenous explanatory variables, and unobservable individual effects. The models that we present are extensions of the random effects probit model of Chamberlain (1984), and have application in the analysis of binary choice, linear regression subject to censoring, and other models with endogenous selectivity. The estimator is based on reduced form predictions of the latent endogenous variables. We also show how to obtain, in one more step, chi-squared test statistics of the overidentifying restrictions, and linear GMM estimators that are asymptotically efficient. (JEL C23*)*

1. Introduction

In this article we consider the problem of estimating a limited dependent variable (LDV) model from panel data, which may include lags of the dependent variable, other endogenous explanatory variables, and unobservable individual effects. The models that we present are extensions of the random effects probit model of Chamberlain (1984), and have application in the analysis of binary choice, linear regression subject to censoring, and models with endogenous selectivity.

An earlier version of this paper previously circulated under the title "Estimating Dynamic Limited Dependent Variable Models from Panel Data with an Application to Female Labour Supply", Nuffield College, Oxford, February 1988. We are grateful to two anonymous referees and the editor for helpful comments on this work. All remaining errors are our own.

We propose a simple within-groups estimator which uses reduced form predicted values of the dependent variable. It can be regarded as a member of Chamberlain's class of random effects minimum distance estimators, and as such it is consistent and asymptotically normal for a fixed number of periods in the absence of misspecification. It also provides a convenient framework for the estimation of semiparametric random effects models in which some of the distributional assumptions implicit in the basic models are relaxed. However, the within-groups estimator is not asymptotically efficient within the minimum distance class, since it implicitly uses a non-optimal weighting matrix. In this regard, we also show how to obtain, in one more step, chi-squared test statistics of the overidentifying restrictions, and linear GMM estimators that are asymptotically efficient.

In the models considered in this paper, variables are either endogenous or exogenous. Lags of the dependent variable are treated as endogenous variables, since we do not restrict the pattern of serial correlation in the errors. Models which contain predetermined variables (through assumptions of absence of autocorrelation, or some other form of sequential conditioning over time) are outside the scope of the present paper.

The random effects models considered here are attractive because they are sufficiently flexible to make it possible the estimation of various nonlinear models of empirical interest subject to permanent unobservable effects.[1] The disadvantage is, of course, that they rely on an explicit specification of the reduced form.

The paper is organized as follows. Section 2 presents the two-step within-groups estimator for the basic model with only exogenous explanatory variables. In addition, we show that the same idea can be applied in the case of a binary choice model with time-series heteroskedasticity. Sections 3 and 4 extend the results, respectively, to a dynamic specification and to endogenous explanatory variables. Section 5 considers asymptotically efficient linear GMM estimation

[1] Labeaga (1990, Ch. 4) applies some of the models and the two-step estimator proposed in this paper to his empirical analysis of the demand for tobacco in Spain using household unbalanced panel data.

and specification testing. Finally, Section 6 contains some concluding remarks.

2. A within-groups estimator for random effects LDV models

2.1. The model and the estimator

We begin by considering a static random effects LDV model of the form

$$y_{it}^* = x_{it}^{'}\beta + \eta_i + v_{it} \qquad (t = 1, ..., T;\ i = 1, ..., N) \qquad [1]$$

where x_{it} is a $k \times 1$ vector of exogenous variables such that

$$E(v_{it} \mid x_{i1}, ..., x_{iT}, \eta_i) = 0$$

and η_i is an unobservable individual effect potentially correlated with x_{it}. y_{it}^* is a latent dependent variable which is not directly observable. We observe instead y_{it} which is some function of y_{it}^*. In the Tobit model $y_{it} = \max(y_{it}^*, 0)$ while in the binary choice model $y_{it} = 1(y_{it}^* > 0)$, where $1(A)$ is the indicator function of the event A. In a generalized selectivity model $y_{it} = 1(I_{it} > 0)y_{it}^*$ where I_{it} is some stochastic index determining whether y_{it} is zero or equal to y_{it}^*. The result of not observing y_{it}^* is that the parameter vector in its conditional mean, β, will not be identified in the absence of additional assumptions concerning the conditional distribution of the error terms. In an obvious notation, the T equations in model [1] can be written as

$$y_i^* = X_i\beta + \eta_i\iota + v_i \qquad (i = 1, ..., N) \qquad [2]$$

where ι is a $T \times 1$ vector of ones.

Following Chamberlain (1984), we parameterize the expectation of η_i conditional on the values of the exogenous variables. Suppose that

$$E\left(\eta_i \mid x_{i1}, ..., x_{iT}\right) = \lambda_0 + \lambda_1' x_{i1} + ... + \lambda_T' x_{iT} + \lambda_*' r_i \qquad [3]$$

where r_i is a vector of variables that includes nonlinear terms in the x_{it}'s. Therefore, letting z_i be the $m \times 1$ vector $(m \geq Tk)$ $z_i = (x'_{i1}...x'_{iT}\ r'_i)'$, the reduced form of the model is given by

$$y_i^* = \Pi z_i + \varepsilon_i \qquad (i = 1, ..., N). \tag{4}$$

The estimators below will be sensitive to the specification of the conditional distribution of η_i. Notice that the conditional expectation of η_i can be approximated to any degree by a polynomial expansion. We would expect that a linear specification, possibly with the addition of quadratic or cubic terms, may often provide a reasonably good approximation. In any event, the reduced form [4] can be tested to some extent against functional misspecification.

If we transform the variables in [2] into deviations from time means, the η_i's are eliminated. Letting $y_i^+ = Qy_i^*$, $X_i^+ = QX_i$ and $v_i^+ = Qv_i$, where Q is the deviations from time means operator $Q = I_T - \iota\iota'/T$:

$$y_i^+ = X_i^+\beta + v_i^+.$$

If y_i^* is directly observed, the OLS regression of y_i^+ on X_i^+ gives us the within-groups estimator of β. However, even if y_i^* is not directly observed, the following expression for the restrictions

$$X_i^+\beta = Q\Pi z_i \qquad (i = 1, ..., N) \tag{5}$$

implies that

$$\beta = \left(\sum_{i=1}^{N} X_i^{+\prime} X_i^+\right)^{-1} \sum_{i=1}^{N} X_i^{+\prime}\Pi z_i. \tag{6}$$

This suggests to estimate β by replacing Π in [6] by a consistent estimator $\widehat{\Pi}$. That is, if we let $\widehat{y}_i$ be a consistent reduced form predictor of y_i^*,

$$\widehat{y}_i = \widehat{\Pi} z_i,$$

we consider as an estimator of β the within-groups regression of $\widehat{y}_i$ on X_i:

$$\widehat{\beta} = \left(\sum_{i=1}^{N} X_i^{+\prime} X_i^+\right)^{-1} \sum_{i=1}^{N} X_i^{+\prime}\widehat{y}_i^+ \tag{7}$$

where $\widehat{y}_i^+ = Q\widehat{y}_i$.

If $\widehat{\Pi}$ is a consistent and asymptotically normal estimator of Π, then $\widehat{\beta}$ can easily be shown to be also consistent and asymptotically normal. Subtracting β from [7] we can write

$$\left(\sum_i X_i^{+\prime} X_i^+\right)(\widehat{\beta}-\beta) = \sum_i X_i^{+\prime}\left(\widehat{y}_i^+ - X_i^+\beta\right) = \sum_i X_i^{+\prime}(\widehat{\Pi}-\Pi)z_i$$

or

$$\widehat{\beta}-\beta = \left(\sum_i X_i^{+\prime} X_i^+\right)^{-1} \sum_i \left(X_i^+ \quad z_i\right)' vec(\widehat{\Pi}-\Pi). \qquad [8]$$

That is, since $\widehat{\beta}$ is linear in $vec(\widehat{\Pi})$, provided the latter is asymptotically normal, the asymptotic normality of the former follows from Cramer's transformation theorem.[2] Assuming that

$$\sqrt{N} vec(\widehat{\Pi}-\Pi) \underset{a}{\sim} N(0, V),$$

the asymptotic variance of $\widehat{\beta}$ can be consistently estimated as[3]

$$\widehat{AVAR}(\widehat{\beta}) = \left(\sum_i X_i^{+\prime} X_i^+\right)^{-1} M'\widehat{V}M \left(\sum_i X_i^{+\prime} X_i^+\right)^{-1} \qquad [9]$$

where $M = \sum_i \left(X_i^+ \quad z_i\right)$ and $\widehat{V}$ is a consistent estimator of V.

Notice that if we evaluate expression [7] at the OLS regression coefficient matrix of y_i on z_i

$$\widetilde{\Pi}_{OLS} = \left(\sum_i y_i z_i'\right)\left(\sum_i z_i z_i'\right)^{-1},$$

we obtain the actual within-groups estimated coefficients in the regression of the observed endogenous variable y_{it} on x_{it}:

$$\widetilde{\beta} = \left(\sum_i X_i^{+\prime} X_i^+\right)^{-1} \sum_i X_i^{+\prime}\widetilde{\Pi}_{OLS} z_i = \left(\sum_i X_i^{+\prime} X_i^+\right)^{-1} \sum_i X_i^{+\prime} y_i.$$

[2]For any matrix A, $vec(A)$ is obtained by stacking the rows of A.

[3]$\widehat{AVAR}(\widehat{\beta})$ denotes a consistent estimate of the variance of the asymptotic distribution of $\sqrt{N}(\widehat{\beta}-\beta)$.

In general, such estimator will not be consistent for β because $\widetilde{\Pi}_{OLS}$ is not a consistent estimate for Π in LDV models. The connection, however, illustrates the fact that the ordinary within-groups estimator can also be regarded as a random effects estimator of the type given in [7].

2.2. Estimating the reduced form

We now turn to consider the problem of obtaining a consistent estimate $\widehat{\Pi}$ and an estimate of its asymptotic variance matrix $\widehat{V}$. For this purpose, it is convenient to provide separate discussion for binary choice, censored (or Tobit) regression, and models with selectivity.

Binary choice

In the binary choice model $y_{it} = 1(y^*_{it} > 0)$. The simplest probit specification is based on the assumption that each of the errors of equation [4] are independent of z_i, and follow a normal distribution with a constant variance $\varepsilon_{it} \mid z_i \sim N(0, \sigma^2)$. Using $\sigma^2 = 1$ as a normalization, we then have that

$$Pr(y_{it} = 1 \mid z_i) = \Phi(\pi'_t z_i)$$

where $\Phi(.)$ is the $N(0,1)$ cdf and π_t is the t-th row of Π. Although the components of ε_i will be correlated in general, separate ML probit estimates of the π_t for each period are consistent and asymptotically normal.

A less restrictive probit model can be obtained allowing for time-series heteroskedasticity (cf. Chamberlain, 1984, pp. 1270-1274). In such case, we assume $\varepsilon_{it} \mid z_i \sim N(0, \sigma^2_t)$, so that $Pr(y_{it} = 1 \mid z_i) = \Phi(\pi^{*\prime}_t z_i)$ with $\pi^*_t = \pi_t / \sigma_t$. As before, some normalization must be chosen. For example, using $\sigma^2_1 = 1$ as the normalization restriction, notice that period by period probit estimates of the reduced form for $t = 2, ..., T$ will be consistent for π^*_t but not for π_t. However, it is still possible to obtain a linear within-groups estimator for the probit model with unequal variances based on reduced form estimates of the π^*_t. Using equation [5] we have

$$X^+_i \beta = Q \Lambda \Pi^* z_i \qquad [11]$$

where $\Lambda = diag(\sigma_t)$. Letting d_{ti} be a $T \times 1$ vector with one in the t-th position and zero elsewhere, this can be written as[4]

$$W_i^{+}\delta = Qd_{1i}(\pi_1^{*\prime} z_i) \qquad [12]$$

where $\delta = (\beta', \sigma_2, ..., \sigma_T)'$, $W_i = (X_i \vdots - (\pi_2^{*\prime} z_i)d_{2i} \vdots ... \vdots - (\pi_T^{*\prime} z_i)d_{Ti})$, and $W_i^{+} = QW_i$.

An implication is that

$$\delta = \left(\sum_{i=1}^{N} W_i^{+\prime} W_i^{+}\right)^{-1} \sum_{i=1}^{N} W_i^{+\prime} d_{1i}(\pi_1^{*\prime} z_i). \qquad [13]$$

As before, this suggests estimating δ by replacing the π_t^* in expression [13] by their period-specific probit estimates:

$$\widehat{\delta} = \left(\sum_{i=1}^{N} \widehat{W}_i^{+\prime} \widehat{W}_i^{+}\right)^{-1} \sum_{i=1}^{N} \widehat{W}_i^{+\prime} d_{1i}(\widehat{\pi}_1^{*\prime} z_i) \qquad [14]$$

where $\widehat{W}_i^{+}$ is as W_i^{+} but using the estimated π_t^*.

Notice that subtracting δ from [14] we can write

$$\begin{aligned}
\left(\sum_i \widehat{W}_i^{+\prime} \widehat{W}_i^{+}\right)\left(\widehat{\delta} - \delta\right) &= \sum_i \widehat{W}_i^{+\prime} Q \left[d_{1i}\left(\widehat{\pi}_1^{*\prime} z_i\right) - \widehat{W}_i \delta\right] \\
&= \sum_i \widehat{W}_i^{+\prime} \left(Q\Lambda\widehat{\Pi}^* z_i - Q\Lambda\Pi^* z_i\right) \\
&= \sum_i \widehat{W}_i^{+\prime} \Lambda \left(\widehat{\Pi}^* - \Pi^*\right) z_i.
\end{aligned}$$

As a consequence, the equation error can be written as

$$\sqrt{N}(\widehat{\delta} - \delta) = \left(\sum_i \widehat{W}_i^{+\prime} \widehat{W}_i^{+}\right)^{-1} \left(\sum_i \widehat{W}_i^{+\prime} \Lambda \quad z_i'\right) \sqrt{N} vec \left(\widehat{\Pi}^* - \Pi^*\right)$$

which suggests that again the consistency and asymptotic normality of $\widehat{\delta}$ follows from the consistency and asymptotic normality of $\widehat{\Pi}^*$.

[4]We are using the fact that $\Lambda\Pi^* z_i = \sum_{t=1}^{T} \sigma_t(\pi_t^{*\prime} z_i)d_{ti}$ together with the normalization $\sigma_1 = 1$.

Finally, the binary choice model can be further generalized by relaxing the assumption of normality. Suppose that z_i contains at least a continuous element, and that $\varepsilon_{it} \mid z_i$ is distributed independent of z_i with a continuously differentiable unknown cdf F_t. Let now re-define $\pi_t^* = \pi_t / \parallel \pi_t \parallel$ exclusive of the constant term, which would be subsumed in F_t. The semiparametric ML estimator of Klein and Spady (1993) or the least-squares estimator of Ichimura (1993) can be used to obtain consistent and asymptotically normal estimates of π_t^*, from which estimates of β and the relative scales $\parallel \pi_t \parallel$ can be developed along the lines of the previous specification.

Censored (or Tobit) models

In the top censored regression model with known censoring point c (as, for example, in the case of top-coded wages), we have $y_{it} = min(y_{it}^*, c)$. In this case, assuming that $\varepsilon_{it} \mid z_i \sim N(0, \sigma_t^2)$, separate ML estimates of each of the T rows of Π are consistent and asymptotically normal. Here the scale parameters σ_t are separately identified, so the problem discussed above for probit models does not arise. On the other hand, in view of the well known lack of robustness of the Tobit estimator to heteroskedasticity and non-normality, the ε_i's can alternatively be maintained to be just independent errors with symmetric distributions. Under theses circumstances, each row of Π can be consistently estimated using, for example, the trimmed least squares method proposed by Powell (1986).

Models with selectivity

In model with selectivity, where y_{it}^* is a theoretical construct (for example, "desired labour supply" or "reservation wages") as opposed to an actual variable subject to censoring, the selection mechanism is often found not to be governed by y_{it}^* itself. Mroz (1987), for example, presented evidence that this type of misspecification may have serious consequences in a model of women's hours of work, and Blundell, Ham and Meghir (1987) also rejected the Tobit model in favour of a double hurdle specification (see Heckman, 1993, for a survey of the literature). It is also possible to extend semiparametric methods to a generalized sample selection model where $y_{it} = 1(I_{it} > 0)y_{it}^*$ (see, for example, Newey, Powell and Walker, 1990, and the references cited there). However, in the more standard context, if we specify an index model of the form

$$I_{it} = \gamma_t' z_i + \nu_{it}$$

and assume that ν_{it} has a known parametric distribution, Heckman (1979)'s lambda-corrected least squares estimators of separate rows of Π are consistent.

2.3. Estimating the asymptotic variance matrix

Since for simplicity we restrict our attention to single equation estimators of [4], for any particular choice of model and estimator, $\widehat{V}$ can be calculated as follows. Let $\widehat{\pi}_t'$ be an estimate of the t-th row of Π defined to minimize a differentiable criterion[5]

$$s_t = \sum_{i=1}^{N} s_{it}\left(y_{it}, z_i, \pi_t\right)$$

(for example, s_t can represent (minus) a Tobit log-likelihood), so that $\widehat{\pi} = vec(\widehat{\Pi})$ minimizes $s(\pi) = \sum_{t=1}^{T} s_t$. Subject to suitable regularity conditions, a first order expansion of $\partial s(\widehat{\pi})/\partial \pi$ about the true value of π gives

$$\left(-\frac{1}{N} diag\left(\frac{\partial^2 s_t}{\partial \pi_t \partial \pi_t'}\right)\right)\sqrt{N}(\widehat{\pi}-\pi) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} \partial s_{i1}/\partial \pi_1 \\ \vdots \\ \partial s_{iT}/\partial \pi_T \end{pmatrix} + o_p(1)$$

which suggests an estimate $\widehat{V}$ of the form

$$\widehat{V} = \widehat{H}^{-1}\widehat{\Psi}\widehat{H}^{-1} \qquad [15]$$

where $\widehat{H} = diag\left(N^{-1}\partial^2 \widehat{s}_t/\partial \pi_t \partial \pi_t'\right)$ and $\widehat{\Psi} = N^{-1}\sum_{i=1}^{N}\left\{\frac{\partial \widehat{s}_{it}}{\partial \pi_t}\cdot\frac{\partial \widehat{s}_{is}}{\partial \pi_s'}\right\}$.

[5]If $\widehat{\pi}_t$ is a trimmed least-squares or a least absolute deviations estimator then the corresponding s_t is not differentiable. Nevertheless, the argument below can be generalized to accommodate asymptotically normal estimators from non-differentiable criteria.

Turning to efficiency issues, since $\widehat{\beta}$ will be usually based on an inefficient reduced form prediction of $y_i^*, \widehat{\beta}$ itself will be inefficient.[6] However, in general $\widehat{\beta}$ will also be inefficient relative to the optimal minimum distance estimator of β based on $\widehat{\Pi}$. This is because $\widehat{\beta}$ can be regarded as a transformed minimum distance estimator which uses a non-optimal norm (see the Appendix for the details). The advantages of $\widehat{\beta}$ are that it is simpler to compute and has a straightforward interpretation. Nevertheless, in Section 5 we show how to obtain linear GMM estimates that are asymptotically efficient relative to a given $\widehat{\Pi}$, and specification tests of the overidentifying restrictions.

3. A dynamic specification

The previous discussion has a straightforward extension to a dynamic specification, since the within-groups estimator based on the same reduced form predictions of the dependent variable remains consistent and asymptotically normal when lags of y_{it}^* are amongst the explanatory variables in the equation.

The use of lags of y_{it}^* as opposed to lags of y_{it} is the natural choice in models with censoring. For example, in analysing the dynamics of wages with top coded wage data y_{it}^* would typically be the process of interest. In binary choice models the situation is rather different, however, since we would condition on past states by conditioning on lags of y_{it}. By conditioning on lags of y_{it}^* instead, one is specifying distributed lagged effects of past exogenous variables and past errors on the current choices (see Heckman, 1981, for a description of these two models). So both types of models are potentially interesting in applications. Nevertheless, models which specify lags of y_{it} would not in general be compatible with linear reduced forms, as assumed here, and therefore they are not considered in this paper (Arellano and Carrasco, 1996, and Honoré and Kyriazidou, 1997, consider binary choice models with state dependence and individual effects).

It should also be mentioned that in the models considered in this pa-

[6]Single equation estimates of Π will in general be inefficient because they ignore the dependence among the components of ε_i. Taking into account such dependence would typically require the use of simulation based estimators (see Hajivassiliou and Ruud, 1994, and references cited there).

per the serial correlation in v_{it} is left unrestricted, and so we treat the lags of y_{it}^* as endogenous variables. If, for example, we assumed v_{it} to be serially independent, values of y_{it}^* lagged two periods or more would be predetermined variables in the equation in first differences. Such situation would generate a type of identifying restrictions that are distinct from those arising from the presence of strictly exogenous variables analyzed here (Arellano and Bond, 1991, considered linear models of this kind, and Arellano, Bover and Labeaga, 1997, considered similar models subject to censoring).

We consider the equation

$$y_{it}^* = \alpha y_{i(t-1)}^* + x_{it}'\beta + \eta_i + v_{it} = w_{it}^{*\prime}\delta + \eta_i + v_{it} \qquad [16]$$

where $w_{it}^* = \left(y_{i(t-1)}^* \vdots x_{it}'\right)'$ and $\delta = (\alpha \vdots \beta')'$. T time periods are observed ($T \geqslant 3$), and as above we assume

$$E(\eta_i \mid z_i) = \lambda' z_i.$$

In addition we assume

$$E(y_{i1}^* \mid z_i) = \mu' z_i$$

so that the reduced form of the model is also given by [4].

The set of $(T-1)$ equations in [16] can be written as

$$(I_0 - \alpha L) y_i^* = X_i \beta + \eta_i \iota + v_i \qquad [17]$$

where I_0 is the $(T-1) \times T$ trim operator $I_0 = (0 \vdots I_{T-1})$, L is the $(T-1) \times T$ lag operator $L = (I_{T-1} \vdots 0)$, X_i is now of order $(T-1) \times k$, and ι and v_i are $(T-1) \times 1$ vectors. Again the individual effects can be eliminated transforming [17] into deviations from time means. Letting Q now be the $(T-1)$ within-groups operator and $B = I_0 - \alpha L$ we have

$$QBy_i^* = X_i^+ \beta + v_i^+.$$

Comparing this equation with the reduced form equation [4] pre-multiplied by QB, we can write the restrictions in the form

$$X_i^{+}\beta = QB\Pi z_i \qquad (i = 1, ..., N). \tag{18}$$

Letting $W_i = (L\Pi z_i \vdots X_i)$ and $W_i^{+} = QW_i$, this can be rewritten as

$$W_i^{+}\delta = QI_0\Pi z_i$$

which implies

$$\delta = \left(\sum_i W_i^{+\prime} W_i^{+}\right)^{-1} \sum_i W_i^{+\prime} I_0 \Pi z_i. \tag{19}$$

Again this suggests estimating δ by replacing Π in [19] by a consistent estimator $\widehat{\Pi}$:

$$\widehat{\delta} = \left(\sum_i \widehat{W}_i^{+\prime} \widehat{W}_i^{+}\right)^{-1} \sum_i \widehat{W}_i^{+\prime} \widehat{y}_{i0}^{+} \tag{20}$$

where $\widehat{y}_{i0} = I_0\widehat{\Pi}z_i, \widehat{y}_{i(-1)} = L\widehat{\Pi}z_i$ and $\widehat{W}_i = (\widehat{y}_{i(-1)} \vdots X_i)$, with the $(+)$ symbols denoting within-groups transformations.

Subtracting δ from [20] we can write

$$\left(\sum_i \widehat{W}_i^{+\prime} \widehat{W}_i^{+}\right)(\widehat{\delta} - \delta) = \sum_i \widehat{W}_i^{+\prime}\left(\widehat{y}_{i0}^{+} - \widehat{W}_i^{+}\delta\right)$$

$$\sum_i \widehat{W}_i^{+\prime}\left(QB\widehat{y}_i - X_i^{+}\beta\right) = \sum_i \widehat{W}_i^{+\prime} B(\widehat{\Pi} - \Pi)z_i$$

or

$$\widehat{\delta} - \delta = \left(\sum_i \widehat{W}_i^{+\prime} \widehat{W}_i^{+}\right)^{-1} \sum_i \left(\widehat{W}_i^{+} \quad z_i\right)^{\prime}(B \quad I_m)\, vec(\widehat{\Pi} - \Pi). \tag{21}$$

In this case, $\widehat{\delta}$ is not linear in $\widehat{\Pi}$ but it is still true that

$$\sqrt{N}(\widehat{\delta} - \delta) =$$
$$\left(\sum_i W_i^{+\prime} W_i^{+}\right)^{-1} \sum_i \left(W_i^{+} \quad z_i\right)^{\prime}(B \quad I_m)\sqrt{N} vec(\widehat{\Pi} - \Pi) + o_p(1)$$

since $p\lim N^{-1} \sum_i \left(\widehat{W}_i^{+\prime}\widehat{W}_i^{+} - W_i^{+\prime}W_i^{+}\right) = 0$ and

$p\lim N^{-1} \sum_i \left[\left(\widehat{W}_i^{+} \quad z_i\right) - \left(W_i^{+} \quad z_i\right)\right] = 0$, so that the consistency and asymptotic normality of $\widehat{\delta}$ follows from the consistency and asymptotic normality of $\widehat{\Pi}$.

The asymptotic variance of $\widehat{\delta}$ can be consistenly estimated as

$$\widehat{AVAR}(\widehat{\delta}) = \left(\sum_i \widehat{W}_i^{+\prime}\widehat{W}_i^{+}\right)^{-1} \widehat{M}'\widehat{V}^{*}\widehat{M} \left(\sum_i \widehat{W}_i^{+\prime}\widehat{W}_i^{+}\right)^{-1} \qquad [22]$$

where $\widehat{M} = \sum_i \left(\widehat{W}_i^{+} \quad z_i\right)$, $\widehat{V}^{*} = \left(\widehat{B} \quad I_m\right) \widehat{V} \left(\widehat{B}' \quad I_m\right)$ and $\widehat{B} = I_0 - \widehat{\alpha}L$.

Our previous comments on efficiency also apply to $\widehat{\delta}$. $\widehat{\delta}$ can be regarded as the minimizer of a transformed minimum distance criterion which uses in general a non-optimal norm, and it is therefore inefficient relative to the optimal MD estimator of δ based on $\widehat{\Pi}$ (see Appendix).

The robustness of $\widehat{\delta}$ depends directly on the robustness of $\widehat{y}_{it}$. In particular, note that $\widehat{\delta}$ is in all cases robust to arbitrary forms of serial correlation in the errors, since no restrictions are placed in the covariances between the components of ε_i when estimating the rows of Π.

To summarize, note that the $\widehat{y}_{it}$ need only be calculated once, and from then on they can be used as our data on the dependent variable to estimating alternative models using the within-groups procedure. Of course, the same is true for $\widehat{V}$. Generally, an attractive feature of methods of the Chamberlain type is a convenient separation between specification searches at the level of the reduced form and at the level of the structural equation. That is, functional form, distributional and observability assumptions can be tested in the reduced form until statistically satisfactory $\widehat{y}_{it}$'s are available, and concentrate on the equation of interest thereafter.

4. Endogenous explanatory variables

Finally we consider a model with endogenous explanatory variables.

For simplicity of presentation a static case with only one endogenous explanatory variable is described. Let

$$y^*_{1it} = \gamma y^*_{2it} + x'_{it}\beta + \eta_i + v_{it} = w^{*\prime}_{it}\delta + \eta_i + v_{it} \qquad [23]$$

where now $w^*_{it} = \left(y^*_{2it} \vdots x'_{it}\right)'$ and $\delta = (\gamma \vdots \beta')'$. The endogenous variable y^*_{2it} may or may not be subject to some censoring rule. In any event we asume

$$E(\eta_i \mid z_i) = \lambda' z_i$$

and

$$E(y^*_{2i} \mid z_i) = \Pi_2 z_i$$

where z_i will now typically include some time-varying outside instrumental variables in addition to functions of the x_{it}. The complete reduced form is given by

$$y^*_i = \Pi z_i + \varepsilon_i \qquad [24]$$

where $y^*_i = \left(y^{*\prime}_{1i} \vdots y^{*\prime}_{2i}\right)'$ is $2T \times 1$ and $\Pi = (\Pi'_1 \vdots \Pi'_2)'$ is $2T \times m$. The set of T equations in [23] can be written as

$$\left(I_T \vdots -\gamma I_T\right) y^*_i = X_i\beta + \eta_i\iota + v_i. \qquad [25]$$

In addition, multiplying through by Q to eliminate the individual effects, and letting $C = \left(I_T \vdots -\gamma I_T\right)$:

$$QCy^*_i = X^+_i\beta + v^+_i. \qquad [26]$$

Pre-multiplying [24] by QC, and comparing with [26], the restrictions can be written as

$$X^+_i\beta = QC\Pi z_i = Q\Pi_1 z_i - \gamma Q\Pi_2 z_i$$

or

$$W^+_i\delta = Q\Pi_1 z_i$$

where $W_i = \left(\Pi_2 z_i \vdots X_i\right)$. Once again we estimate δ by using predicted values of y_{1i}^* and y_{2i}^*:

$$\widehat{\delta} = \left(\sum_i \widehat{W}_i^{+\prime} \widehat{W}_i^{+}\right)^{-1} \sum_i \widehat{W}_i^{+\prime} \widehat{y}_{1i}^{+} \qquad [27]$$

where $\widehat{W}_i = \left(\widehat{y}_{2i} \vdots X_i\right)$, $\widehat{y}_{2i} = \widehat{\Pi}_2 z_i$ and $\widehat{y}_{1i} = \widehat{\Pi}_1 z_i$. If y_{2it}^* is directly observable a valid choice for $\widehat{\Pi}_2$ is the OLS estimate of Π_2.

The discussion concerning the asymptotic distribution of $\widehat{\delta}$ in this case, parallels the one for the dynamic model. Equation [22] remains a valid expression for an estimate of the asymptotic variance of $\widehat{\delta}$ as given in [27], except that now $\widehat{V}^*$ is defined to be

$$\widehat{V}^* = \left(\widehat{C} \quad I_m\right) \widehat{V} \left(\widehat{C}' \quad I_m\right),$$

$\widehat{V}$ is an estimate of the $2Tm \times 2Tm$ variance matrix of $vec(\widehat{\Pi})$, and $\widehat{C} = \left(I_T \vdots -\widehat{\gamma} I_T\right)$.

5. GMM estimation and testing

The within-groups (WG) estimators presented in the previous sections are simple to calculate but, as we pointed out, are inefficient relative to the optimal minimum distance estimator of β based on $\widehat{\Pi}$. Another disadvantage of the WG estimates is that it is not straightforward to obtain from them a chi-squared test statistic of the overidentifying restrictions.[7] Nevertheless, it is still possible to obtain linear GMM asymptotically efficient estimators (relative to $\widehat{\Pi}$) and test statistics in one more step, which do not require the specification of the nonlinear constraints in Π or the estimation of the nuisance parameters λ.

Let us consider the following model that combines the previous specifications

$$y_{1it}^* = \gamma y_{2it}^* + \alpha y_{1i(t-1)}^* + x_{it}'\beta + \eta_i + v_{it} = w_{it}^{*\prime}\delta + \eta_i + v_{it} \qquad [28]$$

[7] The results by Newey (1985) can, nevertheless, be applied to this context to obtain an asymptotic chi-squared statistic.

where now $w_{it}^* = (y_{2it}^* \vdots y_{1i(t-1)}^* \vdots x_{it}')'$ and $\delta = (\gamma \vdots \alpha \vdots \beta')'$.

Since the within-groups equation errors are uncorrelated to the conditioning variables z_i, we can write

$$E[Z_i' Q(y_{1i0}^* - W_i^* \delta)] = 0 \qquad [29]$$

where $Z_i = (I \quad z_i'), Q$ is the $(T-1)$ within-groups operator, $y_{1i0}^* = (y_{1i2}^*, ..., y_{1iT}^*)'$ and $W_i^* = (w_{i2}^*, ..., w_{iT}^*)'$. Moreover, using the law of iterated expectations

$$E\{Z_i' Q[E(y_{1i0}^* \mid z_i) - E(W_i^* \mid z_i)\delta]\} = 0 \qquad [30]$$

where $E(W_i^* \mid z_i) = W_i = (I_0 \Pi_2 z_i \vdots L\Pi_1 z_i \vdots X_i)$ and $E(y_{1i0}^* \mid z_i) = I_0 \Pi_1 z_i$.

This suggests to consider GMM estimators of δ based on the sample orthogonality conditions:

$$b_N(\delta) = \frac{1}{N} \sum_{i=1}^{N} Z_i' (\widehat{y}_{1i0}^{+} - \widehat{W}_i^{+} \delta) \qquad [31]$$

where $\widehat{y}_{1i0} = I_0 \widehat{\Pi}_1 z_i$, $\widehat{W}_i = (\widehat{y}_{2i0} \vdots \widehat{y}_{1i(-1)} \vdots X_i)$, $\widehat{y}_{2i0} = I_o \widehat{\Pi}_2 z_i, \widehat{y}_{1i(-1)} = L\widehat{\Pi}_1 z_i$, and as before the $(+)$ symbols denote within-groups transformations.

A GMM estimator of δ based on $b_N(\delta)$ takes the form

$$\widetilde{\delta}_A = \left[\left(\sum_i \widehat{W}_i^{+\prime} Z_i\right) A_N \left(\sum_i Z_i' \widehat{W}_i^{+}\right)\right]^{-1} \left(\sum_i \widehat{W}_i^{+\prime} Z_i\right) A_N \left(\sum_i Z_i' \widehat{y}_{1i0}^{+}\right) \qquad [32]$$

where A_N is a weighting matrix. With $A_N = \left(\sum_i Z_i' Z_i\right)^{-1}$, $\widetilde{\delta}_A$ coincides with the WG estimator:

$$\widehat{\delta} = \left(\sum_i \widehat{W}_i^{+\prime} \widehat{W}_i^{+}\right)^{-1} \sum_i \widehat{W}_i^{+\prime} \widehat{y}_{1i0}^{+}. \qquad [33]$$

This numerical equivalence results from the fact that the columns in $\widehat{W}_i^+$ are linear combinations of those in Z_i.

In order to obtain the large sample distribution of $\widetilde{\delta}_A$ for an arbitrary A_N, we require an expression for the asymptotic variance of $b_N(\delta)$. This follows from noticing that $b_N(\delta)$ can be expressed as a transformation of $vec(\widehat{\Pi} - \Pi)$ (the Appendix contains a similar discussion, but conducted in terms of transformed minimum distance criteria). Specifically, we have

$$b_N(\delta) = \frac{1}{N}\sum_{i=1}^{N}(I \quad z_i)(Q\Gamma\widehat{\Pi}z_i - QX_i\beta) \qquad [34]$$

where $\Gamma = (I_0 - \alpha L \vdots - \gamma I_0)$.

Using the fact that $QX_i\beta = Q\Gamma\Pi z_i$:

$$\begin{aligned} b_N(\delta) &= \frac{1}{N}\sum_{i=1}^{N}(I \quad z_i)[Q\Gamma(\widehat{\Pi} - \Pi)z_i] \qquad [35] \\ &= (Q \quad \frac{1}{N}\sum_i z_i z_i')(\Gamma \quad I_m)vec(\widehat{\Pi} - \Pi). \end{aligned}$$

Therefore,

$$\sqrt{N}b_N(\delta) \underset{a}{\sim} N\left(0, \overline{Q}E(Z_i'Z_i)V^*E(Z_i'Z_i)\overline{Q}'\right) \qquad [36]$$

where $\overline{Q} = (Q \quad I_m)$, and $V^* = (\Gamma \quad I_m)V(\Gamma' \quad I_m)$.

Hence, a consistent estimate of the asymptotic variance of $\widetilde{\delta}_A$ is given by

$$\widehat{AVAR}(\widetilde{\delta}_A) =$$

$$\left(M_{zw}'A_N M_{zw}\right)^{-1} M_{zw}'A_N \left(\overline{Q}M_{zz}\widehat{V}^*M_{zz}\overline{Q}'\right) A_N M_{zw} \left(M_{zw}'A_N M_{zw}\right)^{-1} \qquad [37]$$

where $M_{zw} = \sum_i Z_i'\widehat{W}_i^+$, $M_{zz} = \sum_i Z_i'Z_i$ and $\widehat{V}^*$ is a consistent estimate of V^*. When $A_N = M_{zz}^{-1}$ this expression blows down to the asymptotic variance estimates for each of the versions of model [28] discussed in the previous sections.

In the previous discussion the within-groups operator Q can be replaced by any $(T-2)\times(T-1)$ matrix K of rank $(T-2)$ such that $K\iota = 0$, since then $Q = K'(KK')^{-1}K$. Natural candidates are the first difference operator, the orthogonal deviations operator, or the first $(T-2)$ rows of the within groups operator (cf. Arellano and Bover, 1995). For any such K matrix, a generic GMM estimator takes the form:

$$\widetilde{\delta}_A =$$

$$\left[\left(\sum_i \widehat{W}_i' K' Z_i\right) A_N \left(\sum_i Z_i' K \widehat{W}_i\right)\right]^{-1} \left(\sum_i \widehat{W}_i' K' Z_i\right) A_N \left(\sum_i Z_i' K \widehat{y}_{1i0}\right). \tag{38}$$

With $A_N = (\sum_i Z_i' K K' Z_i)^{-1}$, $\widehat{\delta}_A$ is numerically the same as the within-groups estimator [33]. The formula [37] remains a valid expression for $\widehat{AVAR}(\widetilde{\delta}_A)$ provided we re-define M_{zw} and Q as $M_{zw} = \sum_i Z_i' K \widehat{W}_i$ and $\overline{Q} = (K \quad I_m)$.

We can now turn to consider efficient estimation relative to $\widehat{\Pi}$. From standard GMM theory, we know that an optimal choice of A_N is given by a consistent estimate of the inverse of the covariance matrix of the orthogonality conditions. Notice that if the within-groups operator is used there are m redundant moment conditions in [31], with the result that their covariance matrix is singular. It is still possible to construct an optimal estimator using a generalized inverse of the covariance matrix of the within-groups moments. This problem does not arise, however, if a transformation K of the type discussed above is used (e.g. first differences). An estimator $\widetilde{\delta}_V$ of the form given in [38] with weighting matrix $A_N = \widehat{V}_b^{-1}$ where

$$\widehat{V}_b = (K \quad I_m) M_{zz} \widehat{V}^* M_{zz} (K' \quad I_m) \tag{39}$$

is asymptotically efficient. Moreover, since the moments eliminated by the transformation to $vec(\widehat{\Pi} - \Pi)$ implicit in $b_N(\delta)$ are unrestricted,

$\widehat{\delta}_V$ is also asymptotically equivalent to the optimal minimum distance estimator of $\widehat{\delta}$ based on $\widehat{\Pi}$.

Finally, the minimized estimation criterion for $\widehat{\delta}_V$ provides a test statistic of the overidentifying restrictions that has an asymptotic chi-squared distribution with $[m(T-2)-(k+2)]$ degrees of freedom under the null of lack of misspecification:

$$S = N\left(\sum_{i=1}^{N}\widehat{u}_i' K' Z_i\right)\widehat{V}_b^{-1}\left(\sum_{i=1}^{N} Z_i' K \widehat{u}_i\right) \underset{a}{\sim} \chi^2_{m(T-2)-(k+2)} \qquad [40]$$

where $\widehat{u}_i = \widehat{y}_{1i0} - \widehat{W_i}\widehat{\delta}_V$.

6. Concluding remarks

This paper has considered various extensions of the random effects probit model of Chamberlain (1984), which include Tobit and other sample selection models with dynamics in the latent endogenous variable and endogenous regressors. We show that all these models can be estimated using a relatively simple two-step within-group method based on estimated reduced form predictions of the latent endogenous variables. The method is not difficult to implement, and its application can be expected to be most promising when based on robust estimates of the reduced form of the type recently developed in the cross-sectional literature on selection models. We also show how to obtain chi-squared specification tests and linear GMM estimators in one more step, that are asymptotically efficient relative to the minimum distance class. The drawbacks of this approach are the same as for Chamberlain's probit model. Namely, that it requires the availability of strictly exogenous variables, and relies on a specification of the conditional distribution of the effects.

The latter assumption can be relaxed somewhat along the lines of Newey (1994) who assumes a nonparametric conditional expectation for the effects. Newey's probit model could be easily extended to incorporate the kind of dynamics and endogenous regressors considered in this paper. Finally, an alternative to the random effects approach spoused in this paper is to consider a fixed effects approach as in the work of Honoré (1992 and 1993) and Honoré and Kyriazidou (1997). The advantage of the fixed effects approach is that it leaves the distribution of the effects unrestricted, but often at the expense of unavoidable lack of flexibility in specifying the structural model of interest.

Appendix

Minimum distance criteria of the random effects LDV estimators

A minimum distance (MD) estimator $\widehat{\theta}$ minimizes a criterion function of the form

$$[vec\left(\widehat{\Pi}-\Pi(\theta)\right)]'\Psi^{-1}vec\left(\widehat{\Pi}-\Pi(\theta)\right). \qquad \text{[A.1]}$$

The optimal choice for the weighting matrix Ψ is $\widehat{V}$, a consistent estimate of the asymptotic variance matrix of $vec(\widehat{\Pi})$. For the static model of Section 2, $\theta=(\beta' \vdots \lambda')'$ while for the dynamic specification $\theta=(\delta' \vdots \lambda' \vdots \mu')'$.

However, the original distance function [A.1] can be considerably simplified without efficiency loss by using the following two properties in MD estimation. Firstly, if K is a nonsingular matrix of dimension T which may or may not depend on θ, the minimizer of the transformed criterion

$$\left[vec(K\widehat{\Pi}-K\Pi)\right]'\Psi_*^{-1}vec(K\widehat{\Pi}-K\Pi) \qquad \text{[A.2]}$$

where $\Psi_*=(\widetilde{K} \quad I_m)\Psi(\widetilde{K}' \quad I_m)$ and $\widetilde{K}$ is such that $p\lim\widetilde{K}=K$, is asymptotically equivalent to $\widehat{\theta}$ (see for example Newey, 1987). Secondly, if some of the coefficients of $K\Pi$ are unrestricted they can be concentrated out, thus obtaining a distance function which depends on a smaller set of parameters.

For the static model let us consider the following non-singular difference transformation:

$$D^*=\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix}=\begin{pmatrix} 1 & 0 & \cdots & 0 \\ & D & & \end{pmatrix}$$

where D is the $(T-1)\times T$ difference operator, $D=I_0-L$. Note that the last $(T-1)$ rows of $D^*\Pi$ only depend on β and that the first row is a transformation of λ which can be concentrated out to obtain the following criterion for β

$$(vec(D\widehat{\Pi} - D\Pi))'\Psi_0^{-1} vec(D\widehat{\Pi} - D\Pi) \qquad [A.3]$$

where

$$\Psi_0 = (D \quad I_m)\Psi(D' \quad I_m).$$

Letting H be a $(T-1)m \times k$ $0-1$ matrix such that $vec(D\Pi) = H\beta$, the minimizer of [A.3] can be obtained as

$$\widetilde{\beta} = \left(H'\Psi_0^{-1}H\right)^{-1} H'\Psi_0^{-1} vec(D\widehat{\Pi}). \qquad [A.4]$$

The efficient estimator relative to $\widehat{\Pi}$ sets $\Psi = \widehat{V}$. However, we can now show that the within-groups estimator $\widehat{\beta}$ given in [7] is an estimator of the form of $\widetilde{\beta}$ which sets $\Psi = \left(I_T \quad \sum_i z_i z_i'\right)^{-1}$. For this choice of Ψ the criterion in [A.3] becomes

$$\begin{aligned}
&\left[vec(D\widehat{\Pi} - D\Pi)\right]' \left[(DD')^{-1} \quad \sum_i z_i z_i'\right] vec(D\widehat{\Pi} - D\Pi) \\
= \;& \sum_i tr(\widehat{\Pi} - \Pi)'D'(DD')^{-1}D(\widehat{\Pi} - \Pi)z_i z_i' \\
= \;& \sum_i z_i'(\widehat{\Pi} - \Pi)'Q(\widehat{\Pi} - \Pi)z_i = \sum_i \left(\widehat{y}_i^+ - X_i^+\beta\right)' \left(\widehat{y}_i^+ - X_i^+\beta\right)
\end{aligned}$$

since $Q = D'(DD')^{-1}D$ and $Q\Pi z_i = X_i^+\beta$. Therefore $\widehat{\beta}$ can only be efficient with respect to $\widehat{\Pi}$ if a multiple of $\left(I_T \quad \sum_i z_i z_i'\right)^{-1}$ is consistent for V.

In a more explicit way, if we let $\mathrm{E}(\eta_i \mid z_i) = \lambda' z_i$ and let r_i to be of order $p \times 1$:

$$\Pi = \left(I_T \quad \beta' \vdots 0\right) + \iota\lambda'$$

where 0 is a $T \times p$ matrix of zeros. Then since $Q\iota = 0$:

$$Q\Pi z_i = Q\left(I_T \quad \beta' \vdots 0\right)\begin{pmatrix} vecX_i \\ r_i \end{pmatrix} = Q\left(I_T \quad \beta'\right) vecX_i = QX_i\beta.$$

Turning to the dynamic model, we consider the non-singular transformation D^*B^* where B^* is given by

$$B^* = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ & B & & \end{pmatrix}.$$

Note that the last $(T-2)$ rows of $D^*B^*\Pi$ only depend on β and that the first two rows are transformations of λ and μ which can be concentrated out to obtain the following criterion for α and β[8]

$$\left[vec(DB\widehat{\Pi} - DB\Pi)\right]' \Psi_0^{-1} vec(DB\widehat{\Pi} - DB\Pi) \qquad [A.5]$$

where now D is the $(T-2) \times (T-1)$ difference operator and

$$\Psi_0 = (DB \quad I_m)\, \Psi\, (B'D' \quad I_m)\,.$$

Letting $\widehat{H}$ be the $(T-2)m \times (k+1)$ matrix such that $vec(\alpha DL\widehat{\Pi} + DB\Pi) = \widehat{H}\delta$ ($\widehat{H}$ is linear in $\widehat{\Pi}$), the minimizer of [A.5] can be written as

$$\widetilde{\delta} = \left(\widehat{H}'\Psi_0^{-1}\widehat{H}\right)^{-1} \widehat{H}'\Psi_0^{-1} vec\left(DI_0\widehat{\Pi}\right). \qquad [A.6]$$

Since Ψ_0 depends on α through B, the calculation of the efficient estimator will require in general a preliminary consistent estimate of α. However if we choose

$$\Psi = \left(B^{*\prime}B^* \quad \sum_i z_i z_i'\right)^{-1}$$

we obtain the within-groups estimator $\widehat{\delta}$ given in [20] which can be calculated in one step. For this choice of Ψ the criterion [A.5] becomes

[8]For the dynamic model $\Pi = B^{*-1}\Gamma$ where $\Gamma = \left(\mu \vdots \Gamma_1'\right)'$ and Γ_1 is the $(T-1) \times m$ matrix

$$\Gamma_1 = \left(0 \vdots I_{T-1} \quad \beta' \vdots 0\right) + \iota\lambda'.$$

$$\left[vec\left(DB\widehat{\Pi} - DB\Pi\right)\right]' \left[\left(DB(B^{*\prime}B^{*})^{-1}B'D'\right)^{-1} \quad \sum_i z_i z_i'\right] vec(DB\widehat{\Pi} - DB\Pi)$$

$$= \sum_i z_i'(B\widehat{\Pi} - B\Pi)'Q(B\widehat{\Pi} - B\Pi)z_i$$

$$= \sum_i \left(\widehat{y}_{i0}^{+} - \alpha\widehat{y}_{i(-1)}^{+} - X_i^{+}\beta\right)' \left(\widehat{y}_{i0}^{+} - \alpha\widehat{y}_{i(-1)}^{+} - X_i^{+}\beta\right)$$

since $DB(B^{*\prime}B^{*})^{-1}B'D' = DD'$ and QBΠ$z_i = X_i^{+}\beta$. Note that $\widehat{\delta}$ would be efficient when $y_{it} = y_{it}^{*}$, $\widehat{\Pi}$ is the unrestricted OLS estimator and the u_{it} are white noise iid errors.

References

Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies* 58, pp. 277-297.

Arellano, M. and O. Bover (1995): "Another look at the instrumental variable estimation of error-components models", *Journal of Econometrics* 68, pp. 29-51.

Arellano, M. and R. Carrasco (1996): "Binary choice panel data models with predetermined variables", Working Paper 9618, CEMFI, Madrid.

Arellano, M., O. Bover and J.M. Labeaga (1997): "Autoregressive models with sample selectivity for panel data", in C. Hsiao, K. Lahiri, L.-F. Lee and H. Pesaran (eds.), *Analysis of Panels and Limited Dependent Variable Models,* Cambridge University Press, forthcoming.

Blundell, R., J.C. Ham and C. Meghir (1987): "Unemployment and female labour supply", *Economic Journal* 97, pp. 44-64.

Chamberlain, G. (1984): "Panel data", in Z. Griliches and M.D. Intrilligator (eds.), *Handbook of Econometrics*, vol. II, Elsevier Science.

Hajivassiliou, V.A. and P.A. Ruud (1994): "Classical estimation methods for LDV models using simulation", in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, North-Holland.

Heckman, J.J. (1979): "Sample selection bias as a specification error", *Econometrica* 47, pp. 153-161.

Heckman, J.J. (1981): "Statistical models for discrete panel data", in C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Mass.

Heckman, J.J. (1993): "What has been learned about labour supply in the past twenty years?", *AEA Papers and Proceedings* 83, pp. 116-121.

Honoré, B. (1992): "Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects", *Econometrica* 60, pp. 533-565.

Honoré, B. (1993): "Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables", *Journal of Econometrics* 59, pp. 35-61.

Honoré, B. and E. Kyriazidou (1997): "Panel data discrete choice models with lagged dependent variables", unpublished.

Ichimura, H. (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models", *Journal of Econometrics* 58, pp. 71-120.

Klein, R.W. and R.H. Spady (1993): "An efficient semiparametric estimator for binary response models", *Econometrica* 61, pp. 387-421.

Labeaga, J.M. (1990): "A microeconometric analysis of the demand for tobacco in Spain", unpublished Ph.D. Thesis, Universitat Autonoma de Barcelona.

Mroz, T.A. (1987): "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions", *Econometrica* 55, pp. 765-799.

Newey, W. (1985): "Generalized method of moments specification tests", *Journal of Econometrics* 29, pp. 229-256.

Newey, W. (1987): "Efficient estimation of limited dependent variable models with endogenous explanatory variables", *Journal of Econometrics* 36, pp. 231-250.

Newey, W. (1994): "The asymptotic variance of semiparametric estimators", *Econometrica* 62, pp. 1349-1382.

Newey, W., J.L. Powell and J. R. Walker (1990): "Semiparametric estimation of selection models: Some empirical results", *AEA Papers and Proceedings* 80, pp. 324-328.

Powell, J.L. (1986): "Symmetrically trimmed least squares estimation for Tobit models", *Econometrica* 54, pp. 1435-1460.

Resumen

En este trabajo proponemos un estimador intra-grupos en dos etapas para modelos con variable dependiente limitada, que pueden incluir retardos de la variable dependiente, otras variables explicativas endógenas y efectos individuales inobservables. Los modelos que presentamos son extensiones del modelo probit con efectos aleatorios de Chamberlain (1984) y tienen aplicación en el análisis de elección discreta, regresión lineal censurada y otros modelos con selección endógena. El estimador se basa en predicciones de forma reducida de las variables endógenas latentes. También mostramos cómo obtener, en una etapa adicional, contrastes ji-cuadrado de las restricciones de sobreidentificación y estimadores lineales del método generalizado de momentos que son asintóticamente eficientes.