Panel Data Models with Predetermined Instruments

Manuel Arellano*

CEMFI, Madrid

March 1995

Very Preliminary and Incomplete Draft

*This paper has been written for presentation as the *Review of Economic Studies* Lecture at the Royal Economic Society Conference at the University of Kent in March 1995. I am grateful to Marco Becht, Olympia Bover and Enrique Sentana for helpful discussions of the material contained in this lecture.

1. Introduction

Time series econometricians have long recognized the distinction between strictly exogenous variables and predetermined variables as a fundamental one in the specification of empirical models. This concern has not been so prominent in the analysis of micro panel data where a sizeable part of the existing work has concentrated on models with just strictly exogenous variables. This situation partly reflects the way in which the literature developed. One strand of this literature found its original motivation in the desire of exploiting panel data for controlling unobserved time-invariant heterogeneity in cross-sectional models. Another strand was interested in panel data as a way to disentangle components of variance and to estimate transition probabilities among states. Papers in either of these two veins have produced various linear and nonlinear "fixed effects" and "random effects" models in which feedback from dependent variables to explanatory variables is typically absent.

Hausman and Taylor (1981) and Amemiya and MaCurdy (1986) among others discussed linear models of this kind. Bhargava and Sargan (1983) considered a linear model which included a lagged dependent variable but required a strictly exogenous variable for identification. Strict exogeneity is also required in the multiplicative models considered by Wooldridge (1990) and Chamberlain (1992a), and in the discrete choice models presented by Chamberlain (1980), Manski (1987) and Newey (1994). Honoré (1992) deals with Tobit and truncated models under strict exogeneity, and while Honoré (1990) extends the model to include a lagged dependent variable, he also requires a strictly exogenous regressor.

A third strand of the literature has studied autoregressive models with individual effects, and, more generally, models with lagged dependent variables.

The work in this area has originated some of the techniques that are customarily applied in panel data models with predetermined variables. These methods are applicable to linear models with additive effects, and have more recently been extended to multiplicative models, which are a generalisation of the exponential regressions used with Poisson-like count data. However, much fewer results are available on dynamic discrete choice models and other non-linear models of interest in microeconometrics.

Autoregressive linear models or models with predetermined variables have been considered by Anderson and Hsiao (1982), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), and Keane and Runkle (1992). Multiplicative models without strictly exogenous variables are treated by Chamberlain (1992b, 1993) and Wooldridge (1991). Discrete choice models with state dependence and heterogeneity have been considered by Heckman (1981a, 1981b), Card and Sullivan (1988), Narendranathan and Elias (1990), and Moon and Stotsky (1993), among others. Censored autoregressive models are studied by Arellano, Bover and Labeaga (1992).

The interaction between unobserved heterogeneity and dynamics in short panels poses new and difficult problems that are absent from time series models. These issues are of importance since with micro data one is typically more interested in the identification and estimation of individual agent's structural responses than in forecasting exercises.

The purpose of this paper is twofold. The first aim is to review recent work on linear and multiplicative panel data models with predetermined variables. The second objective is to show the applicability of some of the insights from this literature for developing useful nonlinear discrete choice models without the strict exogeneity assumption. Throughout, most of the discussion is conducted using simple first-order autoregressive models with individual effects. Although this particular model is not necessarily an interesting one in applied work, it illustrates in a simple setting most of the issues concerning identification and inference that appear in the present context.

According to the definition of Sims (1972), a variable x is strictly exogenous relative to y if

$$E^*(y_t \mid x^T) = E^*(y_t \mid x^t)$$

where E^* denotes a best linear predictor and $x^t = (x_1..x_t)'$. This is well known to be equivalent to the statement that y does not Granger-cause x (cf. Granger (1969)) in the sense that[1]

$$E^*(x_{t+1} \mid x^t, y^t) = E^*(x_{t+1} \mid x^t)$$

With panel data we can define strict exogeneity conditional on an unobserved individual effect η:

$$E^*(y_t \mid x^T, \eta) = E^*(y_t \mid x^t, \eta) \tag{1.1}$$

The Sims' equivalence result based on linear predictors extends to this definition[2]. However, unlike the linear predictors definition, a conditional independence

[1] If linear projections are replaced by conditional distributions the equivalence does not hold and it turns out that the definition of Sims is weaker than Granger's definition. As shown by Chamberlain (1982), conditional Granger non-causality is equivalent to the stronger Sims' condition given by:

$$f(y_t \mid x^T, y^{t-1}) = f(y_t \mid x^t, y^{t-1})$$

[2] Namely, letting $x^{(t+1)T} = (x_{t+1}, .., x_T)'$ if we have

$$E^*(y_t \mid x^T, \eta) = \beta_t' x^t + \delta_t' x^{(t+1)T} + \gamma_t \eta$$

and

$$E^*(x_{t+1} \mid x^t, y^t, \eta) = \psi_t' x^t + \phi_t' y_t^t + \zeta_t \eta$$

it turns out that the restrictions $\delta_t = 0$ and $\phi_t = 0$ are equivalent. This result is due to Chamberlain (1984), and it motivated the analysis in Holtz-Eakin, Newey and Rosen (1988) definition of strict exogeneity given an individual effect is not restrictive, in the sense that there always exists a random variable η such that the condition is satisfied (cf. Chamberlain (1984)). This lack of identification result implies that a test of strict exogeneity given an individual effect will necessarily be a joint test involving a (semi) parametric specification of the conditional distribution.

Lack of control of individual heterogeneity could result in a *spurious* rejection of strict exogeneity, and so a definition of strict exogeneity based on (1.1) is a useful extension of the standard concept to panel data. It also provides a richer framework for the empirical analysis of the time series properties of large micro panels. However, there are many instances in which for theoretical or empirical reasons one is concerned with models exhibiting *true* lack of strict exogeneity after controlling for individual heterogeneity.

This paper considers three types of models with such property. Firstly, linear panel data models for N individuals observed T consecutive time periods of the form

$$y_{it} = \beta x_{it} + \eta_i + v_{it} \qquad (i = 1, ..., N;\ t = 1, ..., T) \tag{1.2}$$

Secondly, multiplicative models exemplified by the exponential regression

$$y_{it} = \exp(\beta \mathbf{x}_{it} + \eta_i) + v_{it} \tag{1.3}$$

and thirdly, binary choice models specified as

$$y_{it} = \mathbf{1}(\beta x_{it} + \eta_i + v_{it} \geq 0) \tag{1.4}$$

where $\mathbf{1}(.)$ denotes the indicator function. An assumption in common to the three models is that the error term v_{it} is mean independent of x_i^t but not of future values of x:

who were concerned with the estimation of the coefficients ψ_t, ϕ_t and ζ_t ,and the testing of the restrictions $\phi_t = 0$.

$$E(v_{it} \mid x_i^t) = 0 \tag{1.5}$$

This assumption allows for unspecified dynamic feedback from y to x, which would be ruled out by the more restrictive strict exogeneity condition

$$E(v_{it} \mid x_i^T) = 0$$

Examples of the previous models include Euler equations for household consumption (cf. Zeldes (1989) and Runkle (1991)) or for company investment (cf. Bond and Meghir (1994)), in which variables in the agents' information sets are uncorrelated with current and future idiosyncratic shocks but not with past shocks.

Another example is the effect of children in female labour force participation decisions. In this context, assuming that children are strictly exogenous is much stronger than the assumption of predeterminedness, since it would require us to maintain that labour supply plans have no effect on fertility decisions at any point in the life cycle (see Browning (1992, p. 1462)). This is a particularly difficult case since participation equations are usually non-linear discrete choice models.

Finally, a last example that is associated with the use of multiplicative models, concerns the relationship between number of patents by a firm in a given year and R&D expenditures (eg. Hausman, Hall and Griliches (1984), Montalvo (1993) and Blundell, Griffith and van Reenen (1993)). It is quite plausible for company decisions on R&D expenditures to depend on the number of patents awarded in previous years.

In all these cases feedback effects from lagged dependent variables (or lagged errors) to current and future values of the explanatory variables cannot be ruled out. The result is that the identification arrangements and the estimation techniques that are useful with strictly exogenous variables break down. In linear models, regressions in first differences or in deviations from means are no more free of bias and could induce larger biases than the regressions in levels (cf. Nickell (1981)). In nonlinear models identification becomes more difficult (see for example the nonidentification results of Chamberlain (1993) for multiplicative models with more than one unobserved effect).

The paper is organized as follows. Section 2 reviews some of the results on linear panel data models and related estimation problems. Section 3.1 presents a similar background discussion for multiplicative models, while Section 3.2 sketches an application of such models to a conditional variance specification with unobserved effects. Section 4 develops semiparametric binary choice models with predetermined variables. Both identification and estimation issues are considered. Special attention is paid to a simple first-order probit autoregressive model with semiparametric individual effects. Finally, Section 5 contains the conclusions of the paper.

2. Linear Models

2.1. Autoregressive Models

Suppose that a random sample of individual time-series of size T $\{y_i^T, i = 1, .., N\}$ is available. We have in mind a typical microeconometric panel where T is small and N is large, so that we shall rely on cross-sectional large sample approximations of estimators and test statistics keeping T fixed.[3]

Let the second-order moment matrix of y_i^T be $E(y_i^T y_i^{T\prime}) = \Omega = \{\omega_{ts}\}$. We assume that the joint distribution of y_i^T and the unobservable time-invariant effect η_i satisfies the following assumption:

[3]This Section follows a similar discussion in Alonso-Borrego and Arellano (1994)

Assumption A

$$y_{it} = \gamma + \alpha y_{i(t-1)} + \eta_i + v_{it} \qquad (t = 2, .., T) \tag{2.1}$$

$$E(v_{it} \mid y_i^{t-1}) = 0 \tag{2.2}$$

where $E(\eta_i) = 0, E(v_{it}^2) = \sigma_t^2$ and $E(\eta_i^2) = \sigma_\eta^2$.

Notice that since equation (2.1) includes a constant term, it is not restrictive to assume that η_i has zero mean. However, in general $E(\eta_i \mid y_i^T)$ will be a function of y_i^T. Moreover, the dependence between η_i and v_{it} is not restricted by the Assumption. Another remark is that Assumption A does not rule out the possibility of conditional heteroskedasticity, since $E(v_{it}^2 \mid y_i^{t-1})$ need not coincide with σ_t^2.

Mean independence of v_{it} with respect to y_i^{t-1} implies that Δv_{it} is mean independent of y_i^{t-2}. This is the basic insight here since Δv_{it} does not depend on η_i. Following Arellano and Bond (1991), Assumption A implies $(T-2)(T-1)/2$ linear moment restrictions of the form

$$E\left[y_i^{t-2}\left(\Delta y_{it} - \alpha \Delta y_{i(t-1)}\right)\right] = 0 \qquad (t = 3, .., T) \tag{2.3}$$

This just says that errors in first differences are uncorrelated with variables lagged two periods or more, which therefore are valid instruments in the estimation of α provided their correlation with Δy_{it-1} does not vanish.[4]

The restrictions above can also be represented as constraints on the elements of Ω. Multiplying (2.1) by y_{is} for $s < t$, and taking expectations gives:

$$\omega_{ts} = \alpha \omega_{(t-1)s} + c_s \qquad (t = 2, .., T;\ s = 1, .., t-1) \tag{2.4}$$

[4]Notice that autocorrelation in the v_{it} is not ruled out since $E(v_{it} v_{i(t-1)}) = -E(v_{it}\eta_i)$ need not be zero. It is nevertheless the case that $E(\Delta v_{it} \mid \Delta v_{i2} ... \Delta v_{i(t-2)}) = 0$

where $c_s = E\left[y_{is}\left(\gamma + \eta_i\right)\right]$. This means that, given Assumption A, the $T(T+1)/2$ different elements of Ω can be written as functions of the $2T$x1 parameter vector

$$\theta = \left(\alpha, c_1, .., c_{T-1}, \omega_{11}, .., \omega_{TT}\right)'.$$

We call this moment structure Model 1. Since the moment restrictions in (2.3) are linear in α, they can be used as the basis for a linear GMM estimator of the type discussed by Arellano and Bond (1991). [5]

The orthogonality conditions (2.3) are the only restrictions implied by Assumption A on the second-order moments of the data. However, they are not the only restrictions available since (2.2) also implies that nonlinear functions of y_i^{t-2} are uncorrelated with Δv_{it}. The semiparametric efficiency bound for this model can be obtained from the more general results in Chamberlain (1994). Estimators whose asymptotic variance attain the bound can be developed using nonparametric estimates of the optimal instruments along the lines of Newey (1990). One reason why estimators based on (2.3) may not be fully efficient asymptotically is that the dependence between η_i and y_i^T may be nonlinear. Another reason would be unaccounted conditional heteroskedasticity.

Notice that with $T = 3$ the parameters (α, c_1, c_2) are just-identified as functions of the elements of Ω:

$$\alpha = \left(\omega_{21} - \omega_{11}\right)^{-1}\left(\omega_{31} - \omega_{21}\right) \tag{2.5}$$

$$c_1 = \omega_{21} - \alpha\omega_{11} \tag{2.6}$$

$$c_2 = \omega_{32} - \alpha\omega_{22} \tag{2.7}$$

[5] The discussion has ignored the intercept coefficient γ. Notice the moment conditions $\gamma = E\left(y_{it}\right) - \alpha E\left(y_{i(t-1)}\right)$ which can be used to determine γ.

With $T \geq 4$ there will be over-identifying restrictions that can be enforced and tested by minimum distance, pseudo maximum likelihood, or GMM methods.

Assumption A specified a first-order autoregressive process with an individual specific level $(\gamma + \eta_i)$ and a common autoregressive coefficient α. An alternative model would specify a homogeneous intercept and heterogeneity in the autoregressive behaviour:

$$y_{it} = \gamma + (\alpha + \eta_i)\, y_{i(t-1)} + v_{it} \tag{2.8}$$

This is a potentially useful model if one is interested in allowing for agent specific adjustment cost functions, as for example in labour demand models (eg. see the analysis in Pesaran and Smith (1995)). The average autoregressive coefficient α and the intercept γ can be determined in a way similar to the previous case since the mean independence of v_{it} relative to y_i^{t-1} implies

$$E(y_{it} y_{i(t-1)}^{-1} \mid y_i^{t-1}) = \gamma y_{i(t-1)}^{-1} + \alpha + \eta_i \tag{2.9}$$

However, a model with heterogeneity in both intercept and slope coefficients in general would not be identified with fixed T (cf. Chamberlain (1993)).

Another remark on Assumption A concerns the nature of the shocks. The errors v_{it} are idiosyncratic shocks that are assumed to have conditional and unconditional cross-sectional zero mean at each point in time. However, if v_{it} contains aggregate shocks (eg. the inflation rate) its cross-sectional mean will not be zero in general. This suggests an extension of the basic specification in which the intercept γ is allowed to vary over time:

$$y_{it} = \gamma_t + \alpha y_{i(t-1)} + \eta_i + v_{it} \tag{2.10}$$

Nevertheless, this extension does not essentially alter the discussion in this section and so it will not be pursued further.

Model 1 is attractive because it is based on minimal assumptions. The autoregressive coefficient α is exclusively identified through the basic mechanism that it is supposed to capture. However, we may be willing to impose additional structure in the problem if this conforms to a priori beliefs. One possibility is to assume that the errors v_{it} are mean independent of the individual effect η_i given y_i^{t-1}. This will often be a reasonable assumption if, for example, the v_{it} are interpreted as innovations that are independent of variables in the agents' information set. In such case, even if η_i is not observable to the econometrician, being time-invariant it is likely to be known to the individual . This situation motivates Assumption B:

Assumption B

$$E\left(v_{it} \mid y_i^{t-1}, \eta_i\right) = 0 \tag{2.11}$$

Note that Assumption B is more restrictive than Assumption A. When $T \geq 4$, Assumption B implies the following additional $T-3$ moment restrictions

$$E\left[\left(y_{it} - \alpha y_{i(t-1)}\right)\left(\Delta y_{i(t-1)} - \alpha \Delta y_{i(t-2)}\right)\right] = 0 \qquad (t = 4, .., T) \tag{2.12}$$

In effect, we can write

$$E\left[\left(y_{it} - \gamma - \alpha y_{i(t-1)} - \eta_i\right)\left(\Delta y_{i(t-1)} - \alpha \Delta y_{i(t-2)}\right)\right] = 0$$

and since $E\left[(\gamma + \eta_i)\, \Delta v_{i(t-1)}\right] = 0$ the result follows. GMM estimators of α that exploit these restrictions in addition to those in (2.3) have been considered by Ahn and Schmidt (1995). An alternative representation of the restrictions in (2.12) is in terms of a recursion of the coefficients c_t introduced in (2.4). Multiplying (2.1) by $(\gamma + \eta_i)$ and taking expectations gives:

$$c_t = \alpha c_{t-1} + \phi \qquad (t = 2, .., T) \tag{2.13}$$

where $\phi = \gamma^2 + \sigma_\eta^2 = E\left[(\gamma + \eta_i)^2\right]$, so that $c_1 ... c_{T-1}$ can be written in terms of c_1 and ϕ. This gives rise to Model 2 in which Ω depends on the $(T+3)$x1 parameter vector

$$\theta = (\alpha, \phi, c_1, \omega_{11}, .., \omega_{TT})'.$$

Notice that with $T = 3$ Assumption B does not imply further restrictions in Ω with the result that α remains just identified relative to the second-order moments. One can solve for ϕ in terms of α, c_1 and c_2:

$$\phi = (\omega_{32} - \omega_{21}) - \alpha (\omega_{22} - \omega_{11}). \tag{2.14}$$

Other forms of additional structure than can be imposed are various versions of mean or variance stationarity conditions. Assumption C specifies a particularly useful mean stationarity condition.

Assumption C

$$E\left(y_{it} - y_{i(t-1)} \mid \eta_i\right) = 0 \qquad (t = 2, .., T) \tag{2.15}$$

This assumption requires the change in y_{it} to be mean independent of the individual effect η_i.

Notice that in combination with Assumption B, Assumption C implies

$$E(y_{it} \mid \eta_i) = \gamma + \alpha E(y_{i(t-1)} \mid \eta_i) + \eta_i \tag{2.16}$$

so that if $E(y_{it} \mid \eta_i)$ is constant it would be the case that

$$E(y_{it} \mid \eta_i) = (\gamma + \eta_i) / (1 - \alpha) \tag{2.17}$$

and $E(y_{it}) = \gamma/(1-\alpha)$.

Relative to Assumption A and Model 1, Assumption C adds the following $(T-2)$ moment restrictions on Ω:

$$E\left[\left(y_{it}-\gamma-\alpha y_{i(t-1)}\right)\Delta y_{i(t-1)}\right]=0 \quad (t=3,..,T) \tag{2.18}$$

which were proposed by Arellano and Bover (1995), who developed a linear GMM estimator of α on the basis of (2.3) and (2.18). However, relative to Assumption B and Model 2, Assumption C only adds one moment restriction which can be written as

$$E\left[\left(y_{i3}-\alpha y_{i2}\right)\Delta y_{i2}\right]=0 \tag{2.19}$$

In terms of the parameters c_t, since

$$c_t-c_{t-1}=E\left[\left(\gamma+\eta_i\right)\left(y_{it}-y_{i(t-1)}\right)\right] \tag{2.20}$$

the implication of Assumption C is that $c_1=...=c_{T-1}$ if we move from Model 1, or that $c_1=\phi/\left(1-\alpha\right)$ if we move from Model 2. This gives rise to Model 3 in which Ω depends on the $(T+2)$x1 parameter vector

$$\theta=\left(\alpha,c_1,\omega_{11},..,\omega_{TT}\right)'.$$

Notice that with $T=3$ α is overidentified under Assumption C. Now α will also satisfy

$$\alpha=\left(\omega_{22}-\omega_{21}\right)^{-1}\left(\omega_{32}-\omega_{31}\right) \tag{2.21}$$

In the presence of aggregate shocks the condition in Assumption C would be replaced by

$$E\left(y_{it}-y_{i(t-1)}\mid\eta_i\right)=\mu_t-\mu_{t-1} \tag{2.22}$$

where $E(y_{it})=\mu_t$. The orthogonality conditions in (2.18) remain valid in this case with the addition of a time-varying intercept.

The basic specification can be restricted further in various ways. For example, we could consider time series homoskedasticity of the form $E\left(v_{it}^2\right)=\sigma^2$ for

$(t = 2, .., T)$ and stationarity of the variance of the initial conditions. The combination of these assumptions with Models 2 or 3 would give rise to additional models, some of which have been discussed in detail in the paper by Ahn and Schmidt (1995). We illustrate the identification content of homoskedasticity by considering the following second-order moment stationarity assumption.

Assumption D

$$E(y_{it}^2) = \omega_{11} \qquad (t = 1, .., T) \tag{2.23}$$

Notice that given Assumption A

$$E\left(\Delta y_{i(t+1)} \Delta y_{it}\right) = -(1-\alpha)\left(\omega_{tt} - \alpha\omega_{(t-1)(t-1)}\right) + (c_t - \alpha c_{t-1}) \tag{2.24}$$

$$E\left[\left(y_{it} - y_{i(t-1)}\right)^2\right] = \omega_{tt} + (1-2\alpha)\,\omega_{(t-1)(t-1)} - 2c_{t-1} \tag{2.25}$$

In addition, under Assumptions C and D it follows that

$$\frac{E\left(\Delta y_{i(t+1)} \Delta y_{it}\right)}{E\left[(\Delta y_{it})^2\right]} = -\frac{(1-\alpha)}{2} \tag{2.26}$$

This is a well known expression for the bias of the least squares regression in first-differences under homoskedasticity, which can be expressed as the orthogonality conditions

$$E\left\{\Delta y_{it}\left[\left(2y_{i(t+1)} - y_{it} - y_{i(t-1)}\right) - \alpha\Delta y_{it}\right]\right\} = 0 \qquad (t = 2, .., T-1) \tag{2.27}$$

With $T = 3$ this implies that α would also satisfy

$$\alpha = (\omega_{22} + \omega_{11} - 2\omega_{21})^{-1}\left[2\left(\omega_{32} - \omega_{31}\right) + \omega_{11} - \omega_{22}\right] \tag{2.28}$$

2.2. Estimation

We discuss estimation issues in the context of a more general linear specification of the form

$$y_{it} = \beta' x_{it} + \eta_i + v_{it} \tag{2.29}$$

$$E(v_{it} \mid z_i^t) = 0 \tag{2.30}$$

where x_{it} is a kx1 vector of possibly endogenous explanatory variables and z_{it} is a px1 vector of instrumental variables. The previous model is a special case with $x_{it} = z_{it} = \left(1, y_{i(t-1)}\right)'$.

The model implies orthogonality conditions of the form

$$E\left[z_i^{t-1}\left(\Delta y_{it} - \beta' \Delta x_{it}\right)\right] = 0 \qquad (t = 2, .., T) \tag{2.31}$$

on which estimation of β is usually based. A GMM estimator of β based on (2.31) is given by

$$\widehat{\beta}_{GMM} = \left(\Delta X' Z A_N Z' \Delta X\right)^{-1} \Delta X' Z A_N Z' \Delta y \tag{2.32}$$

where $\Delta y = (\Delta y_1' .. \Delta y_N')', \Delta X = (\Delta X_1' .. \Delta X_N')', Z = (Z_1' .. Z_N')'$, and $\Delta y_i = (\Delta y_{i2} .. \Delta y_{iT})', \Delta X_i = (\Delta x_{i2} .. \Delta x_{iT})', Z_i = diag\left(z_i^1 .. z_i^{T-1}\right)'$. The weighting matrix A_N is a consistent estimate of the inverse of the covariance matrix of the orthogonality conditions given by

$$V_0 = E\left(Z_i' \Delta v_i \Delta v_i' Z_i\right) \tag{2.33}$$

where $\Delta v_i = \Delta y_i - \beta' \Delta X_i$.

In the remainder of this section we discuss some alternative estimators of β, which are asymptotically equivalent to the GMM estimator but might exhibit better properties in finite samples. These issues are of importance because ordinary instrumental variable estimators can be largely biased towards OLS when the instruments are poorly correlated with the endogenous variables (see Bound, Jaeger and Baker (1993)). Moreover, the correlation between estimated weight matrices A_N and orthogonality conditions can also be responsible for biases in the estimation by GMM of coefficients (eg. Altonji and Segal (1993)) and of standard errors (eg. Arellano and Bond (1991)). We nevertheless wish to draw a careful distinction between the choice of models and the choice of estimation methods. Section 1.1. dealt with alternative autoregressive models that are all amenable of estimation by asymptotically equivalent pseudo-maximum likelihood, minimum distance and GMM methods. This Section deals with alternative estimatos of the same generic model represented by (2.31).

Poor Instruments and Symmetrically Normalized Estimators

Under standard regularity conditions, for sufficiently large N and fixed T the distribution of $\widehat{\beta}_{GMM}$ is approximately multivariate normal with mean β and a covariance matrix than can be consistently estimated as

$$\widehat{Var}(\widehat{\beta}_{GMM}) = (\Delta X' Z A_N Z' \Delta X)^{-1} \tag{2.34}$$

However, as documented in the literature on the finite sample properties of instrumental variables estimators (see Phillips (1983)), this approximation may be very inaccurate even for very large samples when the instruments are poor. Specifically, the two-stage least squares (2SLS) estimator is biased towards the ordinary least squares (OLS) estimator and in the completely unidentified case it converges to the same probability limit as OLS. On the other hand, the limited-information maximum likelihood (LIML) estimator is centred at the parameter value, although its distribution has thicker tails than the distribution of 2SLS.

Since the number of instruments increases with T, the model in (2.29) and (2.30) generates many overidentifying restrictions even for moderate values of

T. However, often the quality of the instruments is poor given that it is usually difficult to predict variables in first differences on the basis of past values of other variables. Motivated by this concern, Alonso-Borrego and Arellano (1994) considered symmetrically normalized GMM estimators which are invariant to normalization and can be expected to have smaller finite sample biases than ordinary GMM estimators. The basis for this conjecture is the result by Hillier (1990) who showed that the alternative normalization rules adopted by LIML and 2SLS are at the root of their different sampling behaviour. Hillier proved that the symmetrically normalized 2SLS estimator has essentially similar finite sample properties to those of the LIML estimator. Symmetrically normalized 2SLS estimators are of interest because, unlike LIML, they are GMM estimators based on structural form orthogonality conditions and therefore they can be readily extended to the nonstandard IV situations that are of interest in panel data models with predetermined variables, while relying on standard GMM asymptotic theory.

When all the variables Δx_{it} are endogenous, a symmetrically normalized GMM estimator of β is given by

$$\widehat{\beta}_{SNM} = (\Delta X' Z A_N Z' \Delta X - \widehat{\lambda} I)^{-1} \Delta X' Z A_N Z' \Delta y \tag{2.35}$$

where $\widehat{\lambda} = \min eigen(W' Z A_N Z' W)$, with $W = (\Delta y, \Delta X)$. The two estimators $\widehat{\beta}_{GMM}$ and $\widehat{\beta}_{SNM}$ are asymptotically equivalent but their finite sample behaviour can be very different. Indeed, the Monte Carlo results of Alonso-Borrego and Arellano for first-order autoregressive models show that GMM estimates can exhibit large biases when the instruments are poor, while the symmetrically normalized estimators remain essentially unbiased. However, the symmetrically normalized estimators always have a larger standard deviation than the ordinary estimators, although the differences are small except in the almost unidentified cases.

Avoiding Estimated Weight Matrices: Maximum Empirical Likelihood Estimators

An undesirable feature of GMM estimators is that the optimal weighting matrix A_N needs to be estimated from preliminary estimators of the sample orthogonality conditions using the same data that are employed in the calculation of the estimator. This fact can be responsible for distortions in the finite sample properties of the GMM estimators and in the estimators of their asymptotic variances.

An alternative method that does not require weight matrix estimates and yet achieves the same asymptotic efficiency as optimal GMM estimators is provided by maximum empirical likelihood estimation (cf. Back and Brown (1993),Qin and Lawless (1994) and Imbens (1993)).

The maximum empirical likelihood estimator $\widehat{\beta}_{MEL}$ maximizes a multinomial pseudo likelihood function subject to the orthogonality conditions. The procedure works because the multinomial pseudo likelihood (or empirical likelihood) is not restrictive. The estimator $\widehat{\beta}_{MEL}$ is given by

$$\widehat{\beta}_{MEL} = \arg\min \sum_{i=1}^{N} \log \left[1 + \sum_{t=1}^{T} \delta_t(\beta)' z_i^{t-1} (\Delta y_{it} - \beta' \Delta x_{it}) \right] \tag{2.36}$$

where the $\delta_t(\beta)$ are implicit functions obtained as the solution to the following system of equations for a given value of β:

$$\sum_{i=1}^{N} \left[1 + \sum_{t=1}^{T} \delta_t' z_i^{t-1} (\Delta y_{it} - \beta' \Delta x_{it}) \right]^{-1} z_i^{t-1} (\Delta y_{it} - \beta' \Delta x_{it}) = 0 \tag{2.37}$$

A test of overidentifying restrictions is provided by the pseudo likelihood ratio statistic

$$w = 2 \sum_{i=1}^{N} \log \left[1 + \sum_{t=1}^{T} \widehat{\delta}_t' z_i^{t-1} (\Delta y_{it} - \widehat{\beta}_{MEL}' \Delta x_{it}) \right] \tag{2.38}$$

which is asymptotically distributed as χ^2 with as many degrees of freedom as the number of overidentifying restrictions.

3. Multiplicative Models

3.1. Exponential Regression for Count Data

In this section we consider models with multiplicative as opposed to additive individual effects. An example of this situation is the model with heterogeneous autoregressive coefficient of the previous Section.

Another example is a life-cycle model of consumption with a household specific rate of time preference δ_i (see Zeldes (1989) and Examples 5.1 in Wooldridge (1991)). Let the utility of household consumption c_{it} at time t be given by

$$u_t(c_{it}) = \frac{c_{it}^{1-\alpha}}{1-\alpha} e^{x_{it}\beta + \eta_i} \tag{3.1}$$

where x_{it} denotes a vector of family composition variables and η_i is an unobserved effect. In the absence of liquidity constraints the Euler equation is

$$E\left[\left(\frac{c_{it}}{c_{it-1}}\right)^{-\alpha}(1+r_{it}) \mid z_i^{t-1}\right] = (1+\delta_i)e^{\Delta x_{it}\beta} \tag{3.2}$$

where the family information set z_i^{t-1} includes Δx_{it} and lagged values of c and r.

However, a leading case which derives its motivation from the literature on Poisson models is an exponential regression for count data of the form

$$E(y_{it} \mid x_i^t, \eta_i) = \exp(\beta' x_{it} + \eta_i) \tag{3.3}$$

The exponential specification is chosen to ensure that the conditional mean is always non negative. With count data a log-linear regression is not a feasible alternative since a fraction of the observations of y_{it} will be zeroes.

Clearly, in such cases first-differencing does not eliminate the unobservable effects, but since the effects are multiplicative there are simple alternative transformations that can be used to construct orthogonality conditions. Generalizing the previous specification:

$$y_{it} = g_t(z_i^t, \beta)\eta_i + v_{it} \tag{3.4}$$

$$E(v_{it} \mid z_i^t) \tag{3.5}$$

where $g_{it} = g_t(z_i^t, \beta)$ is a function of predetermined variables and unknown parameters. Diving by g_{it} and first differencing the resulting equation, we obtain

$$y_{i(t-1)} - y_{it} g_{it}^{-1} g_{i(t-1)} = v_{it}^* \tag{3.6}$$

and

$$E(v_{it}^* \mid z_i^{t-1}) = 0 \tag{3.7}$$

Any function of z_i^{t-1} will be uncorrelated with v_{it}^* and therefore can be used as an instrument in the estimation of the parameters in $g_{it}^{-1} g_{i(t-1)}$. This kind of transformation has been suggested by Chamberlain (1992b) and Wooldridge (1991). Notice that the use of this transformation does not require us to condition on η_i. However, it does require g_t to be a function of predetermined variables as opposed to endogenous variables.

If the analysis is conditional on η_i, that is, if (3.5) is replaced by

$$E(v_{it} \mid z_i^t, \eta_i) = 0$$

then the model admits an alternative representation in terms of a multiplicative error:

$$y_{it} = g_{it}(z_i^t, \beta) \eta_i \varepsilon_{it} \tag{3.8}$$

$$E(\varepsilon_{it} \mid z_i^t, \eta_i) = 1 \tag{3.9}$$

Sometimes, the error ε_{it} may have a more straightforward interpretation than v_{it}.

Asymptotically efficient estimation of β on the basis of (3.6) is not straightforward since arbitrary choices of instruments will be suboptimal in general (see Chamberlain (1993) for a thorough discussion of this problem and derivation of semiparametric efficiency bounds). Here we illustrate the problem by means of a simple autoregressive model with T=3. Models with T>3 are more difficult to analyse because the correlation between equations needs to be taken into account (cf. Chamberlain (1993)). Let the model be:

$$y_{it} = \exp(\gamma + \alpha y_{i(t-1)} + \eta_i) + v_{it} \qquad (t = 1, 2, 3) \tag{3.10}$$

The transformed equation (3.6) in this case reduces to:

$$E\left[y_{i2} - y_{i3}\exp(-\alpha \Delta y_{i2}) \mid y_{i1}\right] = 0 \tag{3.11}$$

If y_{i1} has a finite support consisting of r different known values $\phi_1 ... \phi_r$, the conditional moment restriction (3.11) is equivalent to the r unconditional moment restrictions:

$$E\left[d_{ij}(y_{i2} - y_{i3}e^{-\alpha \Delta y_{i2}})\right] = 0 \qquad (j = 1, .., r) \tag{3.12}$$

where $d_{ij} = \mathbf{1}(y_{i1} = \phi_j)$. The conditions in (3.12) could be used to obtain asymptotically efficient two-step GMM or maximum empirical likelihood estimates of α.

More generally, the efficient non-feasible instrumental variable for this problem is given by

$$z_i = \left[E(v_{i3}^{*2} \mid y_{i1})\right]^{-1} E(y_{i3}\Delta y_{i2}e^{-\alpha \Delta y_{i2}} \mid y_{i1}) \tag{3.13}$$

in the sense that a method of moments estimator of α based on

$$E\left[z_i(y_{i2} - y_{i3}e^{\alpha \Delta y_{i2}})\right] = 0 \tag{3.14}$$

would achieve the semiparametric efficiency bound. Obviously, since z_i is unknown, any feasible estimator of this kind would have to rely on an unrestricted estimate of z_i taking into account the discrete nature of the data.[6]

Here we have assumed that interest concentrates on the conditional mean of an integer-valued random variable. If the analysis were also concerned with the estimation of conditional probabilities, likelihood models would be required. For example, Brännäs (1994) discusses a useful first-order autoregressive generalized Poisson model.

[6]The asymptotic properties of estimators of nonparametric regressions containing discrete regressors have been recently discussed by Delgado and Mora (1995).

3.2. Conditional Variance with Individual Effects

We now consider a linear autoregressive process of the type discussed in Section 2 but assuming that not only the conditional mean but also the conditional variance is of interest. An example is an intertemporal model of savings in which the conditional variance of household income plays a role due to precautionary motives (cf. Blundell and Stoker (1994)).

In the analysis of conditional variances with data on individuals it is important to take into account unobserved heterogeneity since failure to do so could result in spurious dynamics in the second moments.

Let us consider a linear conditional mean and a multiplicative conditional variance given by:

$$E(y_{it} \mid y_i^{t-1}, \eta_i) = \gamma + \alpha y_{i(t-1)} + \eta_i \tag{3.15}$$

$$Var(y_{it} \mid y_i^{t-1}, \eta_i) \equiv E(v_{it}^2 \mid y_i^{t-1}, \eta_i) = \sigma_t^2(y_i^{t-1}, \theta)\eta_i^2 \tag{3.16}$$

where $v_{it} = y_{it} - \gamma - \alpha y_{i(t-1)} - \eta_i$. Let us also define the error $u_{it} = y_{it} - \gamma - \alpha y_{i(t-1)}$.

Notice that since we are conditioning on η_i and $E(u_{it} \mid y_i^{t-1}, \eta_i) = \eta_i$ we have:

$$E(v_{it}^2 \mid y_i^{t-1}, \eta_i) = E(u_{it}^2 \mid y_i^{t-1}, \eta_i) - \eta_i^2 \tag{3.17}$$

Therefore

$$E(u_{it}^2 \mid y_i^{t-1}, \eta_i) = \left[1 + \sigma_t^2(y_i^{t-1}, \theta)\right] \eta_i^2 \tag{3.18}$$

In view of the discussion in the previous sections, the model implies the following conditional moment restrictions in terms of observable variables and parameters:

$$E(\Delta y_{it} - \alpha \Delta y_{i(t-1)} \mid y_i^{t-2}) = 0 \tag{3.19}$$

$$E\left(u_{i(t-1)}^2 - u_{it}^2(1 + \sigma_{it}^2)^{-1}(1 + \sigma_{i(t-1)}^2) \mid y_i^{t-2}\right) = 0 \tag{3.20}$$

where $\sigma_{it} = \sigma_t(y_i^{t-1}, \theta)$.

A possible specification for σ_{it}^2 of the ARCH type is given by

$$\sigma_{it}^2 = \theta_0 + \theta_1 u_{i(t-1)}^2 \tag{3.21}$$

Alternatively, we could consider an exponential ARCH specification with asymmetric response of the form:

$$\sigma_{it}^2 = \exp\left(\theta_0 + \theta_1 u_{i(t-1)} + \theta_2 u_{i(t-1)}^2\right) \tag{3.22}$$

For a particular parameterizarion of σ_{it}^2, joint GMM estimates of α and θ can be obtained using orthogonality conditions derived from (3.19) and (3.20). Another possibility is to obtain estimates of θ conditional on a previous estimate of α from which residuals $\widehat{u}_{it}^2$ could be constructed.

A number of hypothesis can be tested within this framework. A test for conditional homoskedasticity allowing for unobserved cross-sectional heterogeneity is a test for the constancy of the ratio $(1+\sigma_{i(t-1)}^2)/(1+\sigma_{it}^2)$. On the other hand, a test for the absence of individual effects in variances could be based on testing the validity of orthogonality conditions of the form:

$$E\left[y_i^{t-1}(u_{it}^2 - 1 - \sigma_t^2(y_i^{t-1}, \theta))\right] = 0 \tag{3.23}$$

4. Binary Choice Models

4.1. Autoregressive Probit with Individual Effects

Suppose that y_{it} is a 0-1 variable. Let us consider the following model

$$y_{it} = \mathbf{1}(\gamma_t + \alpha y_{i(t-1)} + \eta_i + v_{it} \geq 0) \qquad (t = 2, .., T) \tag{4.1}$$

The basic motivation for this model is to facilitate the distinction between unobserved heterogeneity and state dependence in the analysis of binary-state discrete-time processes. One example is the analysis of sequences of employment and unemployment states, where a substantive question is whether or not unemployment causes future unemployment (cf. Heckman (1981c), Card and Sullivan (1988), and Narendranathan and Elias (1990); Card and Sullivan use these models for measuring the effect of training programs on employment and unemployment probabilities). Another example is the analysis of a housing quality indicator over time as in the work by Moon and Stotsky (1993). Moon and Stotsky consider the effect of rent control on a two state housing condition variable (sound and unsound) allowing for state dependence and unobserved heterogeneity.

Treating the sample values of η_i as parameters to be estimated (fixed-effects) when N is large but T is fixed in general produces inconsistent estimates of α and γ_t , due to the problem of incidental parameters (cf. Chamberlain (1980) and Heckman (1981a, 1981b)). So, the point of departure for a number of studies in the literature has been the following conditional independence assumption

$$\Pr(-v_{it} \leq \xi \mid y_i^{t-1}, \eta_i) = \Pr(-v_{it} \leq \xi) \equiv F_t(\xi) \tag{4.2}$$

together with a parametric specification of $F_t(\xi)$. Thus,

$$\Pr(y_{it} = 1 \mid y_i^{t-1}, \eta_i) = F_t(r_t + \alpha y_{i(t-1)} + \eta_i) \equiv F_t(y_{i(t-1)}, \eta_i) \tag{4.3}$$

When F_t is the logistic distribution and the model only includes strictly exogenous regressors, the sample mean of y_{it} provides a sufficient statistic for η_i. Therefore, the likelihood function conditional on $\Sigma_t y_{it}$ does not depend on η_i, thus providing a basis for consistent inferences on the parameters of interest (see Chamberlain (1980)). However, the conditional likelihood approach cannot be extended to the logit model with state dependence. Indeed, as shown by Card and Sullivan (1988, Appendix), in the presence of state dependence the minimal sufficient statistic for η_i is in general the entire data vector for individual i.

We are thus led to consider a specification of the distributions of $\eta_i \mid y_{i1} = 1$ and $\eta_i \mid y_{i1} = 0$. Notice the following relationships:

$$\Pr\left[(y_{i1}...y_{iT}) = (\phi_1...\phi_T)\right] = \Pr(y_{i1} = \phi_1) \Pr\left[(y_{i2}...y_{iT}) = (\phi_2...\phi_T) \mid y_{i1} = \phi_1\right]$$

$$= \Pr(y_{i1} = \phi_1) \int \Pr\left[(y_{i2}...y_{iT}) = (\phi_2...\phi_T) \mid y_{i1} = \phi_1, \eta_i\right] dG(\eta_i \mid y_{i1} = \phi_1)$$

$$= \Pr(y_{i1} = \phi_1) \int \prod_{t=2}^{T} F_t(y_{i(t-1)}, \eta_i)^{y_{it}} \left[1 - F_t(y_{i(t-1)}, \eta_i)\right]^{(1-y_{it})} dG(\eta_i \mid y_{i1} = \phi_1) \tag{4.4}$$

where G is the conditional distribution of η_i given $y_{i1} = \phi_1$, and $\phi_t \in \{0, 1\}, (t = 1, ..., T)$.

Note that by specifying the conditional distributions of η_i given y_{i1} as opposed to the marginal distribution of η_i, we allow for dependence between y_{i1} and η_i, while leaving the initial conditions of the process unrestricted.

Moon and Stotsky (1993), following related work by Heckman and Singer (1984) for semiparametric duration models, specified G as an unrestricted discrete distribution with finite support, given by m mass points $e_1, ..., e_m$. In this case expression (4.4) becomes

$$\Pr(y_{i1} = \phi_1) \sum_{l=1}^{m} \prod_{t=2}^{T} F_t(y_{i(t-1)}, e_l)^{y_{it}} \left[1 - F_t(y_{i(t-1)}, e_l)\right]^{(1-y_{it})} \Pr(\eta_i = e_l \mid y_{i1} = \phi_1) \tag{4.5}$$

For a given value of m, and a logistic specification for F_t, Moon and Stotsky obtained maximum likelihood estimates of the parameters of interest in their housing quality model together with the mass points $e_1, .., e_m$ and their conditional probabilities $\Pr(\eta_i = e_l \mid y_{i1} = \phi_1)$.

An Alternative Random Effects Model with Unrestricted Conditional Means

Let us consider equation (4.1) together with the assumption:

$$\eta_i + v_{it} \mid y_i^{t-1} \sim N\left(E(\eta_i \mid y_i^{t-1}), \sigma_t^2\right) \tag{4.6}$$

The sequence of conditional means $\{E(\eta_i \mid y_i^s), s = 1, .., T-1\}$ is left unrestricted except for the fact that they are linked by the law of iterated expectations:

$$E(\eta_i \mid y_i^{t-1}) = E\left(E(\eta_i \mid y_i^t) \mid y_i^{t-1}\right) \tag{4.7}$$

The conditional probabilities specified by the model are

$$\Pr\left(y_{it} = 1 \mid y_i^{t-1}\right) = \Phi\left(\frac{\gamma_t + \alpha y_{i(t-1)} + E\left(\eta_i \mid y_i^{t-1}\right)}{\sigma_t}\right) \tag{4.8}$$

where $\Phi(.)$is the standard normal cdf. Obviously, the assumption of normality is unessential and could be replaced by any other parametric assumption like the logistic distribution.

$\Pr(y_{i1} = 1)$ is left unrestricted and just adds one parameter to the full likelihood function of the data. The other parameters are γ_t, α and the collection of $E(\eta_i \mid y_i^{t-1})$ that, as will be seen below, are identified up to scale with T>3, together with the relative scales.

Note that contrary to the fixed effects model, there is no incidental parameters problem here since we are estimating $E(\eta_i \mid y_i^{t-1})$ as opposed to the η_i themselves. Indeed, the motivation here is to develop a framework that, while avoiding the problem of initial conditions and the incidental parameters problem when T is small and fixed, can be easily extended to more general binary choice models with predetermined variables (see below).

It is useful to relate the present model to a model in which (4.6) is replaced by the assumption of a mass point distribution for $\eta_i \mid y_{i1}$ together with

$$v_{it} \mid y_i^{t-1}, \eta_i \sim N(0, \omega_t^2) \tag{4.9}$$

so that

$$\Pr\left(y_{it} = 1 \mid y_i^{t-1}, \eta_i\right) = \Phi\left(\frac{\gamma_t + \alpha y_{i(t-1)} + \eta_i}{\omega_t}\right) \tag{4.10}$$

In the latter model it is possible to argue that the distribution of $\eta_i \mid y_{i1}$ could be approximated arbitrarily well if we allow m to increase as N increases, although the model assumes that η_i is conditionally independent of v_{it} given y_i^{t-1}. On the other hand, model (4.1)-(4.6) leaves $E(\eta_i \mid y_i^{t-1})$ unrestricted but it may implicitly restrict other features of the distribution of $\eta_i \mid y_i^{t-1}$ by assuming that the distribution of $\eta_i + v_{it} \mid y_i^{t-1}$ is parametric. However, in model (4.1)-(4.6) η_i and v_{it} are not required to be conditionally independent.

It is of some interest to compare this situation with the linear case discussed in Section 2 where the model

$$E\left(y_{it} \mid y_i^{t-1}, \eta_i\right) = \gamma_t + \alpha y_{i(t-1)} + \eta_i$$

was a specialization of the less restrictive specification

$$E\left(y_{it} \mid y_i^{t-1}\right) = \gamma_t + \alpha y_{i(t-1)} + E\left(\eta_i \mid y_i^{t-1}\right)$$

Note that if for *some* t

$$\eta_i \mid y_i^{t-1} \sim N\left(E\left(\eta_i \mid y_i^{t-1}\right), \sigma_{\eta t}^2\right) \tag{4.11}$$

it can be easily shown that this assumption together with (4.9) implies an expression identical to (4.8) with $\sigma_t^2 = \sigma_{\eta t}^2 + \omega_t^2$. It would thus appear that assumptions (4.6) and (4.9) are connected through (4.11). However, if $\eta_i \mid y_i^{t-1}$ is normal, since

y_i^{t-1} is a sequence of binary variables, it follows that $\eta_i \mid y_i^{t-2}$ cannot be normal unless $\Pr(y_{i(t-1)} = 1 \mid y_i^{t-2})$ is one or zero. In fact, the distribution of $\eta_i \mid y_i^{t-2}$ would be a normal mixture and therefore an expression of the form of (4.8) could not hold for $\Pr(y_{i(t-1)} = 1 \mid y_i^{t-2})$.

The tx1 random vector y_i^t has a multivariate Bernouilli distribution, and takes 2^t different values ϕ_j^t $(j = 1, .., 2^t)$. Similarly, y_i^{t-1} takes on $2^{(t-1)}$ different values $\phi_j^{t-1}(j = 1, .., 2^{t-1})$. As a matter of notational convenience we order the ϕ_j^t such that for t>1:

$$\phi_j^t = \begin{cases} (\phi_j^{t-1}, 1) & if \quad j = 1, ..., 2^{(t-1)} \\ (\phi_j^{t-1}, 0) & if \quad j = 2^{(t-1)} + 1, ..., 2^t \end{cases} \tag{4.12}$$

Let us denote

$$p_{tj}^{t-1} = \Pr\left(y_{it} = 1 \mid y_i^{t-1} = \phi_j^{t-1}\right) \equiv h_t(\phi_j^{t-1}) \qquad (j = 1, .., 2^{(t-1)}) \tag{4.13}$$

and

$$\psi_j^{t-1} = E\left(\eta_i \mid y_i^{t-1} = \phi_j^{t-1}\right) \qquad (j = 1, .., 2^{(t-1)}) \tag{4.14}$$

Therefore we have

$$p_{tj}^{t-1} = \Phi\left(\frac{\gamma_t + \alpha\phi_j^{[t-1]} + \psi_j^{t-1}}{\sigma_t}\right) \tag{4.15}$$

where $\phi_j^{[t-1]}$ denotes the last element of the vector ϕ_j^{t-1}. By the law of iterated expectations we also have

$$\psi_j^{t-1} = \psi_j^t p_{tj}^{t-1} + \psi_{2^{(t-1)}+j}^t (1 - p_{tj}^{t-1}) \qquad (j = 1, .., 2^{(t-1)}; t = 2, .., T-1) \tag{4.16}$$

Moreover, since the model includes a constant term, it is not restrictive to assume that $E(\eta_i) = 0$. Therefore:

$$E(\eta_i) = E(\eta_i \mid y_{i1} = 1)\Pr(y_{i1} = 1) + E(\eta_i \mid y_{i1} = 0)\Pr(y_{i1} = 0) = 0$$

or

$$\psi_1^1 p_1 + \psi_2^1 (1 - p_1) = 0 \tag{4.17}$$

where $p_1 = \Pr(y_{i1} = 1)$ (The notation is $\psi_j^1 = E(\eta_i \mid y_{i1} = \phi_j^1)$, $(j = 1, 2)$ with $\phi_1^1 = 1$ and $\phi_2^1 = 0$).

The number of reduced form parameters p_{tj}^{t-1} is $(2 + 2^2 + ... + 2^{T-1})$, and with the addition of p_1 gives a total number of cell probabilities of $\Sigma_{j=0}^{T-1} 2^j$ (eg. with $T = 6$ there would be 63 coefficients).

The coefficients can be estimated up to scale. Using σ_2 as the scale, we can estimate γ_t/σ_2, α/σ_2, ψ_j^{t-1}/σ_2 and the relative scales σ_2/σ_t. We shall use $\sigma_2 = 1$ as the normalization restriction.

The structural parameters are α, $\gamma_2 ... \gamma_T$, $\sigma_3 ... \sigma_T$ and the ψ_j^{t-1}. The number of ψ_j^{t-1} parameters is $(2^1 + 2^2 + ... + 2^{T-1})$, although they are subject to restrictions.[7] In conclusion, the total number of orthogonality conditions is

$$r = 2\Sigma_{j=1}^{T-2} 2^j + 2^{T-1} + 1 \tag{4.18}$$

while the number of parameters to be estimated is

$$k = 2(T-1) + \Sigma_{j=1}^{T-1} 2^j \tag{4.19}$$

Hence the number of overidentifying restrictions is

$$r - k = \Sigma_{j=1}^{T-2} 2^j - 2T + 3 \tag{4.20}$$

Identification of α up to scale requires that at least $T \geq 4$. With $T = 3$, α would only be identified under homoskedasticity. Indeed, setting $\sigma_2 = \sigma_3 = 1$ and γ_t constant, a straightforward calculation reveals that

$$\alpha = \frac{\Phi^{-1}(\pi_1) - \pi_1 \Phi^{-1}(\pi_{11})}{(1 - \pi_1)} - \Phi^{-1}(\pi_{10}) \tag{4.21}$$

where

$$\pi_1 = \Pr(y_{i2} = 1 \mid y_{i1} = 1)$$

[7] We could alternatively say that the required free parameters are

$$\psi_j^{T-1} = E(\eta_i \mid y_i^{T-1} = \phi_j^{T-1}) \quad (j = 1, ..., 2^{T-1})$$

since the remaining ψ_j^{t-1} are functions of those.

$$\pi_{11} = \Pr(y_{i3} = 1 \mid y_{i1} = 1, y_{i2} = 1)$$

$$\pi_{10} = \Pr(y_{i3} = 1 \mid y_{i1} = 1, y_{i2} = 0)$$

Minimum Distance and Maximum Likelihood Estimation

Let us define the variables

$$d_{ij}^{t} = \mathbf{1}(y_i^t = \phi_j^t) \tag{4.22}$$

Then the unrestricted maximum likelihood estimator of p_{tj}^{t-1} is given by

$$\widehat{p}_{tj}^{t-1} = \frac{1}{\Sigma_{i=1}^{N} d_{ij}^{t-1}} \Sigma_{i=1}^{N} y_{it} d_{ij}^{t-1} \quad (t = 2, ..., T; \quad j = 1, ..., 2^{(t-1)}) \tag{4.23}$$

Similarly, for p_1 we have

$$\widehat{p}_1 = \frac{1}{N} \Sigma_{i=1}^{N} y_{i1} \tag{4.24}$$

We can form the vector

$$g(\widehat{p}, \theta) = \begin{pmatrix} \psi_1^1 \widehat{p}_1 + \psi_2^1 (1 - \widehat{p}_1) \\ \widehat{p}_{tj}^{t-1} - \Phi\left(\frac{\gamma_t + \alpha \phi_j^{[t-1]} + \psi_j^{t-1}}{\sigma_t}\right) \\ \psi_j^{t-1} - \psi_j^t \widehat{p}_{tj}^{t-1} - \psi_{2^{(t-1)}+j}^t (1 - \widehat{p}_{tj}^{t-1}) \\ \vdots \end{pmatrix} \tag{4.25}$$

The vector of functions $g(\widehat{p}, \theta)$ includes the terms for all j and t. The vector $\widehat{p}$ contains the $\widehat{p}_{tj}^{t-1}$ and $\widehat{p}_1$, while θ contains all the parameters to be estimated. A minimum distance (MD) estimator of θ solves

$$\widehat{\theta} = \arg\min_{\theta} g(\widehat{p}, \theta)' A_N g(\widehat{p}, \theta) \tag{4.26}$$

where A_N is a consistent estimate of the inverse of the covariance matrix of $g(\widehat{p}, \theta)$.

As an alternative to the MD procedure, the model can be estimated by maximum likelihood. The log-likelihood is maximized as a function of θ subject to the restrictions (4.16). The difference in this case is that in (4.16), the p_{tj}^{t-1} represent the probabilities (4.15) implied by the model, as opposed to being substituted by their unrestricted estimates.

GMM Estimation

The following simpler method avoids the joint estimation of the parameters of interest with the nuisance coefficients ψ_j^{t-1}.

By inverting equation (4.8) we obtain

$$\sigma_t \Phi^{-1}[h_t(y_i^{t-1})] = \gamma_t + \alpha y_{i(t-1)} + E(\eta_i \mid y_i^{t-1}) \tag{4.27}$$

First-differencing this equation we have:

$$\sigma_t \Phi^{-1}[h_t(y_i^{t-1})] - \sigma_{t-1}\Phi^{-1}[h_{t-1}(y_i^{t-2})] - \Delta\gamma_t - \alpha \Delta y_{i(t-1)} = \varepsilon_{it} \tag{4.28}$$

where

$$\varepsilon_{it} = E(\eta_i \mid y_i^{t-1}) - E(\eta_i \mid y_i^{t-2}) \tag{4.29}$$

Therefore

$$E(\varepsilon_i \mid y_i^{t-2}) = 0 \tag{4.30}$$

Notice that

$$h_t(y_i^{t-1}) = \sum_{j=1}^{2^{(t-1)}} p_{tj}^{t-1} d_{ij}^{t-1} \tag{4.31}$$

Moreover, the conditional moment restriction (4.30) is equivalent to the following unconditional moments (see Chamberlain (1987, p. 308)):

$$E(d_{ij}^{t-2}\varepsilon_{it}) = 0 \qquad (j = 1, ..., 2^{t-2}) \tag{4.32}$$

or:

$$E\left\{ d_j^{t-2} \left[\sigma_t \Phi^{-1} \left(\sum_{j=1}^{2^{(t-1)}} p_{tj}^{t-1} d_{ij}^{t-1} \right) - \sigma_{t-1}\Phi^{-1} \left(\sum_{j=1}^{2^{(t-2)}} p_{(t-1)j}^{t-2} d_{ij}^{t-2} \right) - \Delta\gamma_t - \alpha\Delta y_{i(t-1)} \right] \right\} = 0 \tag{4.33}$$

The orthogonality conditions corresponding to the p_{tj}^{t-1} are

$$E[d_{ij}^{t-1}(y_{it} - p_{tj}^{t-1})] = 0 \qquad (j = 1, ..., 2^{(t-1)}) \tag{4.34}$$

$$E[d_{ij}^{t-1}(y_{i(t-1)} - p_{(t-1)j}^{t-2})] = 0 \qquad (j = 1, ..., 2^{(t-2)}) \tag{4.35}$$

The complete set of moment conditions can be used to obtain joint estimates of the p_{tj}^{t-1} and the coefficients of interest. However, since the former are unrestricted moments there is no efficiency loss (as far as the estimation of the parameters of interest is concerned) in replacing in the first set of orthogonality conditions (4.33) unrestricted estimates of the p_{tj}^{t-1} and the $p_{(t-1)j}^{t-2}$.

Letting

$$\widehat{h}_t(y_i^{t-1}) = \sum_{j=1}^{2^{(t-1)}} \widehat{p}_{tj}^{t-1} d_{ij}^{t-1}, \tag{4.36}$$

a two-step GMM method can be based on the sample orthogonality conditions:

$$\frac{1}{N}\sum_{i=1}^{N} d_{ij}^{t-2}(\sigma_t \Phi^{-1}[\widehat{h}_t(y_i^{t-1})] - \sigma_{t-1}\Phi^{-1}[\widehat{h}_t(y_i^{t-2})] - \Delta\gamma_t - \alpha\Delta y_{i(t-1)}) \tag{4.37}$$

$$(j = 1, ..., 2^{(t-2)};\ t = 3, ..., T)$$

yielding asymptotically efficient estimates of α, $\Delta\gamma_t$ and σ_t subject to the normalization restriction $\sigma_2 = 1$. Since y_i^T has a finite support the model is fully parametric and the asymptotic distribution of the estimators can be obtained using standard GMM asymptotic theory.

4.2. Probit Model with Continuous Predetermined Variables

We now consider a model of the form

$$y_{it} = \mathbf{1}(\beta' x_{it} + \eta_i + v_{it} \geq 0) \tag{4.38}$$

$$\eta_i + v_{it} \mid x_i^t \sim N(E(\eta_i \mid x_i^t), \sigma_t^2) \tag{4.39}$$

where x_{it} is a vector of continuous predetermined random variables.[8]

Therefore

$$\Pr(y_{it} = 1 \mid x_i^t) \equiv h_t(x_i^t) = \Phi\left(\frac{\beta' x_{it} + E(\eta_i \mid x_i^t)}{\sigma_t}\right) \tag{4.40}$$

[8]In fact, some of the x's could be discrete (for example, if x_{it} includes $y_{i(t-1)}$ or other dummy variables). The implications of such situation will be discussed below.

We can proceed as in the previous subsection to obtain

$$\sigma_t \Phi^{-1}[h_t(x_i^t)] - \sigma_{t-1}\Phi^{-1}[h_{t-1}(x_i^{t-1})] - \beta' \Delta x_{it} = \varepsilon_{it} \tag{4.41}$$

where

$$\varepsilon_{it} = E(\eta_i \mid x_i^t) - E(\eta_i \mid x_i^{t-1})$$

so that

$$E(\varepsilon_{it} \mid x_i^{t-1}) = 0 \tag{4.42}$$

The transformation (4.41) is similar to the one employed by Newey (1994a) for a probit model with strictly exogenous variables. In the strictly exogenous case, the error term ε_{it} does not appear since there is no sequential updating of the conditional expectations of the individual effects.

Contrary to the purely autoregressive case we now rely on nonparametric kernel estimators $\widehat{h}_2(x_i^2), ..., \widehat{h}_T(x_i^T)$, in order to construct orthogonality conditions for β up to scale and the relative standard deviations.

Let define

$$\widehat{\psi}_{it}(\theta) = x_i^{t-1}\{\sigma_t \Phi^{-1}[\widehat{h}_t(x_i^t)] - \sigma_{t-1}\Phi^{-1}[\widehat{h}_{t-1}(x_i^{t-1})] - \beta' \Delta x_{it}\} \tag{4.43}$$

where $\theta = (\beta', \sigma_3, ..., \sigma_T)'$ and we are using $\sigma_2 = 1$ as the normalizing restriction.

Let the sample orthogonality conditions be given by

$$\widehat{b}_N(\theta) = \frac{1}{N}\Sigma_{i=1}^{N}(\widehat{\psi}_{i2}(\theta)', ..., \widehat{\psi}_{iT}(\theta)')' \tag{4.44}$$

A semiparametric two-step GMM estimator of θ solves

$$\widehat{\theta} = \arg\min \widehat{b}_N(\theta)' A_N \widehat{b}_N(\theta) \tag{4.45}$$

where A_N is a weight matrix.

Under appropriate regularity conditions (see Newey (1994a) and Newey & McFadden (1995)):

$$\sqrt{N}\widehat{b}_N(\theta) \xrightarrow{d} N(0, V_0) \tag{4.46}$$

with

$$V_0 = E[(\psi_i(\theta) + a_i)(\psi_i(\theta) + a_i)'] \tag{4.47}$$

where $\psi_i(\theta) = (\psi_{i2}(\theta)'...\psi_{iT}(\theta)')'$, $\psi_{it}(\theta) = x_i^{t-1}\varepsilon_{it}$, and a_i is an adjustment term that takes into account the fact that the $h_t(x_i^t)$ have been replaced by nonparametric estimates.

Following Newey (1994b), V_0 can be consistently estimated by mean of

$$\widehat{V} = \frac{1}{N}\Sigma_{i=1}^N(\widehat{\psi}_i + \widehat{a}_i)(\widehat{\psi}_i + \widehat{a}_i)' \tag{4.48}$$

where

$$\widehat{a}_i = \frac{1}{N}\Sigma_{s=2}^T \Sigma_{j=1}^N \left(\frac{\partial \widehat{\psi}_j}{\partial h_s}\right) y_{is} K_s(x_j^s - x_i^s) \tag{4.49}$$

and $K(.)$ is the kernel used in the estimation of the $h_t(x_i^t)$.

Finally, a consistent estimate of the asymptotic variance matrix of $\widehat{\theta}$ is given by

$$(\widehat{D}' A_N \widehat{D})^{-1} \widehat{D}' A_N \widehat{V} A_N \widehat{D} (\widehat{D}' A_N \widehat{D})^{-1} \tag{4.50}$$

where

$$\widehat{D} = \frac{1}{N}\Sigma_{i=1}^N \frac{\partial \widehat{\psi}_i}{\partial \theta'} \tag{4.51}$$

If model (4.38) is augmented to include $y_{i(t-1)}$:

$$y_{it} = \mathbf{1}(\alpha y_{i(t-1)} + \beta' x_{it} + \eta_i + v_{it} \geq 0) \tag{4.52}$$

$$\eta_i + v_{it} \mid x_i^t, y_i^{t-1} \sim N(E(\eta_i \mid x_i^t, y_i^{t-1}), \sigma_t^2) \tag{4.53}$$

we have

$$h_t(x_i^t, y_i^{t-1}) \equiv \Pr(y_{it} = 1 \mid x_i^t, y_i^{t-1}) = \sum_{j=1}^{2^{(t-1)}} h_t(x_i^t, d_{ij}^{t-1}) d_{ij}^{t-1} \tag{4.54}$$

where d_{ij}^{t-1} is defined in (4.22). Thus, in this case, instead of one $t + (t-1)$ dimensional nonparametric function, we have $2^{(t-1)}$ functions of dimension t, each of which would be estimated by kernel methods. A similar procedure would also apply to other discrete predetermined variables with finite support.

Another extension is a model where individual effects are interacted with time effects given by

$$y_{it} = \mathbf{1}(\beta' x_{it} + \eta_i \delta_t + v_{it} \geq 0) \tag{4.55}$$

and

$$\eta_i \delta_t + v_{it} \mid x_i^t \sim N(E(\eta_i \mid x_i^t)\delta_t, \sigma_t^2) \tag{4.56}$$

In this model, estimation can be based on the transformation

$$\sigma_t \Phi^{-1}[h_t(x_i^t)] - r_t \sigma_{t-1} \Phi^{-1}[h_{t-1}(x_i^{t-1})] = \beta' x_{it} - r_t \beta' x_{i(t-1)} + \varepsilon_{it}^* \tag{4.57}$$

where $r_t = \delta_t / \delta_{(t-1)}$ and $E(\varepsilon_{it}^* \mid x_i^{t-1}) = 0$.

An example of this situation is provided by the study of macroeconomic effects on dividends policy (see for example the paper by Gertler and Hubbard (1993), who test the hypothesis that equity provides firms with a cushion against aggregate fluctuations). Suppose that $y_{it} = 1$ if the change in dividends between periods t and $(t-1)$ is positive, and $y_{it} = 0$ otherwise. The variables x_{it} would represent the changes in predetermined firm variables like earnings per share, and δ_t would represent macro and tax shocks that are allowed to have firm-specific effects measured by η_i.

References

[1] Ahn, S. and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data", *Journal of Econometrics*, forthcoming.

[2] Alonso-Borrego, C. and M. Arellano (1994): "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data", unpublished manuscript, CEMFI.

[3] Altonji, J.G. and L. Segal (1993): "Small Sample Bias in GMM Estimation of Covariance Structures", unpublished manuscript, Northwestern University.

[4] Amemiya, T. and T.E. MaCurdy (1986): "Instrumental-Variable Estimation of an Error-Components Model", *Econometrica*, 54, 869-881.

[5] Anderson, T.W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, 18, 47-82.

[6] 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, 277-297.

[7] Arellano, M. and O. Bover (1995): "Another Look at the Instrumental-Variable Estimation of Error-Components Models", *Journal of Econometrics*, forthcoming.

[8] Arellano, M., O. Bover and J.M. Labeaga (1992): "Autoregressive Models with Sample Selectivity for Panel Data", unpublished manuscript, CEMFI.

[9] Back, K. and D.P. Brown (1993): "Implied Probabilities in GMM Estimators", *Econometrica*, 61, 971-975.

[10] Bhargava, A. and J.D. Sargan ((1983): " Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods", *Econometrica*, 51, 1635-1659.

[11] Blundell, R., R. Griffith and J. van Reenen (1993): "Knowledge Stocks, Persistent Innovation and Market Dominance: Evidence from a Panel of British Manufacturing Firms", IFS Working Paper W93/19, London.

[12] Blundell, R. and T.M. Stoker (1994): "Consumption and the Timing of Income Risk", IFS Working Paper W94/9, London.

[13] Bond, S. and C. Meghir (1994): "Dynamic Investment Models and the Firm's Financial Policy", *Review of Economic Studies*, 61, 197-222.

[14] Bound, J., D.A. Jaeger and R. Baker (1993): " The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables", NBER Technical Working Paper 137.

[15] Brännäs, K. (1994): "Estimation and Testing in Integer-Valued AR(1) Models", Umeå Economic Studies 335, University of Umeå.

[16] Browning, M. (1992): "Children and Household Economic Behavior", *Journal of Economic Literature*, 30, 1434-1475.

[17] Card, D. and D. Sullivan (1988) : "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment", *Econometrica*, 56, 497-530.

[18] Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47, 225-238.

[19] Chamberlain, G. (1982): "The General Equivalence of Granger and Sims Causality", *Econometrica*, 50, 569-581.

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

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

[22] Chamberlain, G. (1992a): "Efficiency Bounds for Semiparametric Regression", *Econometrica*, 60, 567-596.

[23] Chamberlain, G. (1992b): "Comment: Sequential Moment Restrictions in Panel Data", *Journal of Business and Economic Statistics*, 10, 20-26.

[24] Chamberlain, G. (1993): "Feedback in Panel Data Models", unpublished manuscript, Harvard University.

[25] Delgado, M.A. and J. Mora (1995): "Nonparametric and Semiparametric Estimation with Discrete Regressors", *Econometrica*, forthcoming.

[26] Gertler, M. and R.G. Hubbard (1993): "Corporate Financial Policy, Taxation, and Macroeconomic Risk", *RAND Journal of Economics*, 24, 286-303.

[27] Granger, C.W.J. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica*, 37, 424-438.

[28] Hausman, J.A. and W.E. Taylor (1981): "Panel Data and Unobservable Individual Effects", *Econometrica*, 49, 1377-1398.

[29] Hausman, J., B. Hall and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica*, 51, 783-798.

[30] Heckman, J.J. (1981a): "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.

[31] Heckman, J.J. (1981b): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process", in C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Mass.

[32] Heckman, J.J. (1981c): "Heterogeneity and State Dependence", in S. Rosen (ed.) *Studies in Labor Markets*, NBER, University of Chicago Press.

[33] Heckman, J.J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 52, 271-320.

[34] Hillier, G.H. (1990): "On Normalization of Structural Equations: Properties of Direction Estimators", *Econometrica*, 58, 1181-1194.

[35] Holtz-Eakin, D., W. Newey and H.S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data", *Econometrica*, 56, 1371-1395.

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

[37] Honoré, B. (1990): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables", unpublished manuscript, Northwestern University.

[38] Imbens, G.W. (1993): "A New Approach to Generalized Method of Moments Estimation", unpublished manuscript, Harvard University.

[39] Keane, M. and D. Runkle (1992): " On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous", *Journal of Business and Economic Statistics*, 10, 1-9.

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

[41] Montalvo, J.G. (1993): "Patents and R&D at the Firm Level: A New Look", *Revista Española de Economía*, Número Monográfico, 67-81.

[42] Moon, C.-G. and J.G. Stostky (1993): "The Effect of Rent Control on Housing Quality Change: A Longitudinal Analysis", *Journal of Political Economy*, 101, 1114-1148.

[43] Narendranathan, W. and P. Elias (1990): "Influences of Past History on the Incidence of Youth Unemployment: Empirical Findings for the U.K.", Warwick Economic Research Papers 369, University of Warwick.

[44] Newey, W. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models", *Econometrica*, 58, 809-837.

[45] Newey, W. (1994a): "The Asymptotic Variance of Semiparametric Estimators", *Econometrica*, 62, 1349-1382.

[46] Newey, W. (1994b): "Kernel Estimation of Partial Means and a General Variance Estimator", *Econometric Theory*, 10, 233-253.

[47] Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics*, IV, Ch. 36.

[48] Nickell, S. (1981): "Bises in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1417-1426.

[49] Pesaran, M.H. and R. Smith (1995): "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels", *Journal of Econometrics*, forthcoming.

[50] Phillips, P.C.B. (1993): "Exact Small Sample Theory in the Simultaneous Equations Model", in Griliches, Z. and M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. I, North-Holland, Amsterdam, Ch. 8.

[51] Qin, J. and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations", *The Annals of Statistics*, 22, 300-325.

[52] Runkle, D.E. (1991): "Liquidity Constraints and the Permanent Income Hypothesis: Evidence from Panel Data", *Journal of Monetary Economics*, 97, 73-98.

[53] Sims, C.A. (1972): "Money, Income, and Causality", *The American Economic Review*, 62, 540-552.

[54] Wooldridge, J. (1990): "Distribution-Free Estimation of Some Nonlinear Panel Data Models", unpublished manuscript, Massachussets Institute of Technology.

[55] Wooldridge, J. (1991): "Multiplicative panel Data Models Without the Strict Exogeneity Assumption", unpublished manuscript, Massachusetts Institute of Technology.

[56] Zeldes, S.P. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation", *Journal of Political Economy*, 97, 305-346.