Autoregressive Models with Sample Selectivity for Panel Data*

Manuel Arellano[†], Olympia Bover[‡], and José M. Labeaga[§]

Revised: May 1997

Abstract

The purpose of this paper is to formulate procedures for the analysis of the time series behaviour of micro panel data subject to censoring. We assume an autoregressive model with random effects for a latent variable which is only partly observed due to a selection mechanism. Our methods are based on the observation that the subsamples which only include individuals without censored past observations are exogenously selected for the purpose of estimating features of the distribution of the censored endogenous variable conditional on its past. We apply these methods to analyze the dynamics of female labour supply and wages using PSID data.

*An earlier version of this paper was presented at the Mannheim Meeting on Labour Markets, 19-22 March 1992, and at the European Meeting of the Econometric Society, Brussels, August 1992. We are grateful to two anonymous referees for helpful comments on this work. All remaining errors are our own.

[†]CEMFI, Madrid.

[‡]Research Department, Bank of Spain.

[§]Universitat Pompeu Fabra and UNED.

1 Introduction

Recent studies have developed econometric procedures for the analysis of the time series properties of panel data sets consisting of large numbers of short individual time series (eg. Anderson and Hsiao (1981), Chamberlain (1984), Holtz-Eakin, Newey and Rosen (1988), and Arellano and Bond (1991)). The analysis is typically based on empirical autoregressive equations including time and individual effects, and possibly observed time-varying exogenous variables. Individual effects are removed by differencing and lagged variables are used as instruments in order to retrieve consistent estimators of the autoregressive coefficients of the levels equation. Alternatively, one could choose moving average processes and components of variance to model the autocovariance matrix of the data in first differences, using methods of moments estimation and testing as well (as done, for example, by Abowd and Card (1989)). In either case, the motivation for this type of analysis with micro data is often to establish a mapping between the observed dynamic interactions and those implied by a theoretical model, or at least to test particular time series implications of such model.

The purpose of this paper is to formulate procedures for the analysis of the time series behaviour of panel data subject to censoring. We apply these methods to analyse the dynamics of female labour supply and wages using PSID data. We follow the standard latent variable approach to models with selectivity and assume a linear autoregressive model for a latent variable which is only partly observed due to a selection mechanism.

These models arise as a natural limited-dependent-variable extension of similar linear models, and may be a representation of the reduced form of interesting structural models. In this regard, it is important to distinguish an interest in the dynamics of the censored variable given the selection rule, from a concern with the dynamics of the selection process. For example, in terms of our application, we are interested in the time series behaviour of female labour supply and wages conditional on participation and individual effects. If the focus were on the dynamics of participation, it would be important to model dependence on past states as well as unobserved heterogeneity (see Heckman (1981) for a menu of alternative models), In effect, the models we consider are strictly speaking models of selectivity in the sense that qualitative choice models are not covered, since at least some values of the latent variables (not just their sign) must be observed.

The paper is organized as follows. Section 2 presents the model and compares our assumptions with those typically made for linear models. Section 3 discusses methods of parameter estimation and testing. The basic method of estimation can be regarded as an application of the asymptotic least squares procedures of Gourieroux, Monfort and Trognon (1985). Section 4 contains the application to female labour supply and wages using two samples from the Michigan database. Finally, Section 5 presents some concluding remarks.

2 The Model

We begin by considering a first-order autoregression for a scalar latent variable y^*_{it} including an individual effect η_i. The index i denotes cross-sectional observations and t refers to time periods. Specifically, we have

$$y^*_{it} = \alpha y^*_{i(t-1)} + \eta_i + v_{it} \qquad \mid \alpha \mid < 1 \tag{1}$$

with

$$E(v_{it} \mid y^*_{i1}, ..., y^*_{i(t-1)}) = 0.$$

The variable y^*_{it} is observed subject to endogenous selection. We use the notation y_{it} for the observed variable, and the sample consists of N independently distributed individual time series of length T. Throughout, T is small and N is large. This framework will include truncated, and Type I and Type II Tobit censored autoregressive models (using the terminology of Amemiya (1985), see below).

Even in the absence of selection, equation (1) presents the problem that the permanent effect η_i is unobserved. However, the equation error in first differences satisfies

$$E(\Delta y^*_{it} - \alpha \Delta y^*_{i(t-1)} \mid y^*_{i1}, ..., y^*_{i(t-2)}) = 0 \tag{2}$$

which implies moment restrictions on the joint distribution of $(y^*_{i1}, ..., y^*_{iT})$, but marginal with respect to η_i. In particular we have the following $(T-2)(T-1)/2$ orthogonality conditions:

$$E\left[y^*_{i(t-j)}(\Delta y^*_{it} - \alpha \Delta y^*_{i(t-1)})\right] = 0 \qquad (j = 2, ..., (t-1); t = 3, ..., T) \tag{3}$$

which are the basis for instrumental variables inferences in the linear model without selectivity.

For later use, we notice that the orthogonality conditions (3) can also be written in terms of the coefficients of the best linear predictors of y^*_{it} and $y^*_{i(t-1)}$ given $(y^*_{i1}, ..., y^*_{i(t-2)})$. Letting

$$\pi_{t-1} = \left[E\left(x_{i(t-2)}x'_{i(t-2)}\right)\right]^{-1} E\left(x_{i(t-2)}y^*_{i(t-1)}\right) \tag{4}$$

$$p_t = \left[E\left(x_{i(t-2)}x'_{i(t-2)}\right)\right]^{-1} E\left(x_{i(t-2)}y^*_{it}\right) \tag{5}$$

where $x_{i(t-2)} = (y^*_{i1}, ..., y^*_{i(t-2)})'$, the orthogonality conditions in (3) can be written as

$$(p_t - \pi_{t-1}) = \alpha(\pi_{t-1} - q_{t-2}) \qquad (t = 3, ..., T) \tag{6}$$

where q_{t-2} is a $(t-2) \times 1$ vector which has one in the last position and zero elsewhere. Clearly, the coefficients p_t are related to the π_t and the relation is given by

$$p_t = (I_{t-2} : \pi_{t-1})\pi_t \tag{7}$$

This approach is attractive because it places no restrictions on the distribution of the effects given the observed conditioning variables. However, it cannot be directly used in our case since we only observe sample moments conditional on selection. That is, we do not observe sample counterparts of the population regression coefficients π_t.

In fact, the selection model is unidentified in the absence of additional prior restrictions on the distribution of the latent variable. If T where sufficiently large, we could choose to place restrictions on the conditional distribution of y^*_{it} while treating the realizations of η_i as parameters to be estimated. Honoré (1992) presents a static Type I Tobit model with fixed effects together with a consistent and asymptotically normal estimator for that model with fixed T and large N. Honoré's estimator places no restrictions on the distribution of η_i given the exogenous variables. In return, he requires the distribution of the errors given the exogenous variables and the effects to be fully stationary, hence ruling out time series heteroskedasticity. In a similar vein, Honoré (1993) gives moment conditions that do not depend on η_i for a Type I Tobit model with a lagged dependent variable, strictly exogenous variables, and stationary and serially uncorrelated errors.

Here we achieve identification by placing restrictions on the conditional distribution of the latent variables y_{it}^* given $y_{i1}^*, ..., y_{i(t-1)}^*$ but not η_i. Firstly, we specify the mean of $y_{it}^* \mid y_{i1}^*, ..., y_{i(t-1)}^*$ as an unrestricted (non-Markovian) linear regression. That is, we assume that this mean coincides with the corresponding linear projection (which, for example, would be the case if the y_{it}^* were jointly normally distributed). This amounts to specifying the mean of the effects given $y_{i1}^*, ..., y_{i(T-1)}^*$ and so we assume some knowledge of the conditional distribution of η_i. In doing this we follow the work of Chamberlain (1984). Secondly, additional features of the distribution of $y_{it}^* \mid y_{i1}^* ... y_{i(t-1)}^*$ will be specified to overcome the selection problem, using methods existing in the literature. A benefit of this approach is that we can consider Type I and Type II censored models within the same framework. Another advantage is that nonstationary errors (like errors with time series heteroskedasticity) are not ruled out.

In general we have

$$E\left(y_{it}^* \mid y_{i1}^*, ..., y_{i(t-1)}^*\right) = \alpha y_{i(t-1)}^* + E\left(\eta_i \mid y_{i1}^*, ..., y_{i(t-1)}^*\right) \tag{8}$$

and we assume

$$E\left(y_{it}^{*}\mid y_{i1}^{*},...,y_{i(t-1)}^{*}\right) = \pi_{t1}y_{i1}^{*}+...+\pi_{t(t-1)}y_{i(t-1)}^{*}=\pi_{t}'x_{i(t-1)} \quad (t = 2,...,T) \tag{9}$$

which implies that

$$E\left(\eta_{i}\mid y_{i1}^{*},...,y_{i(T-1)}^{*}\right)=\lambda_{1}y_{i1}^{*}+...+\lambda_{T-1}y_{i(T-1)}^{*} \tag{10}$$

Notice that given (8) and (9), all the conditional expectations $E(\eta_i \mid x_{it})$ are linear and their coefficients are functions of $\lambda_1, ..., \lambda_{T-1}$ and α. Using the law of iterated expectations we have

$$\begin{aligned} E(\eta_i \mid x_{it}) &= E(\lambda_1 y_{i1}^{*}+...+\lambda_{T-1}y_{i(T-1)}^{*}\mid x_{it}) \\ &= \sum_{k=1}^{t}\lambda_k y_{ik}^{*}+\sum_{j=1}^{T-t-1}\lambda_{t+j}E(y_{i(t+j)}^{*}\mid x_{it}) \end{aligned}$$

and for $j \geq 2$

$$E(y_{i(t+j)}^{*}\mid x_{it})=\pi_{t+j}'\prod_{s=1}^{j-1}(I_{t+j-s-1}\vdots\pi_{t+j-s})'x_{it}.$$

The coefficients π_t are nonlinear functions of α and the λ's, with the latter being nuisance parameters. For example, with $T=3$ α is uniquely determined given the π's. In this case we have

$$E(y_{i2}^{*}\mid y_{i1}^{*})=\pi_{21}y_{i1}^{*}$$

$$E(y_{i3}^{*}\mid y_{i1}^{*},y_{i2}^{*})=\pi_{31}y_{i1}^{*}+\pi_{32}y_{i2}^{*}$$

and

$$\pi_{21}=(\alpha+\lambda_1)/(1-\lambda_2), \qquad \pi_{31}=\lambda_1,\ \ \pi_{32}=\alpha+\lambda_2$$

Solving for α we obtain

$$\alpha = \frac{(\pi_{31} + \pi_{21}\pi_{32}) - \pi_{21}}{\pi_{21} - 1} = \frac{p_{31} - \pi_{21}}{\pi_{21} - 1} = \frac{E(y_{i1}^* \Delta y_{i3}^*)}{E(y_{i1}^* \Delta y_{i2}^*)}$$

The expression on the right hand side is the population counterpart of the Anderson-Hsiao (1981) instrumental variables estimator used in linear models. With $T > 3$ there are $(1/2)(T-2)(T-1)-1$ overidentifying restrictions given the π_t's. Notice that with $T = 3$ a second-order autoregression with individual effects would not be identified. Here we assume that although T is small, it is sufficiently large to avoid problems of lag truncation.

When $\alpha > 0$, the dependence of y_{it}^* on both $y_{i(t-1)}^*$ and η_i generates positive autocorrelation on y_{it}^*. Having assumed that the reduced form autoregression (9) is a linear one, the structure of the model apportions the overall serial correlation in y_{it}^* between the autoregressive and the permanent components.

In the censored sample selection model, the observed variable y_{it} is given by

$$y_{it} = d_{it} y_{it}^* \tag{11}$$

where d_{it} is a binary selection indicator. In the Type I model d_{it} takes the form

$$d_{it} = 1(y_{it}^* > 0) \tag{12}$$

where $1(A)$ denotes an indicator function of the event A, while in the Type II model we have

$$d_{it} = 1(\gamma_t' w_{it} + \varepsilon_{it} > 0) \tag{13}$$

where ε_{it} is an unobserved error term and w_{it} is a vector of variables which includes $x_{i(t-1)}$, but may also contain other variables known on a priori grounds to be independent of $y_{it}^* \mid x_{i(t-1)}$. In this sense, predictors of the individual effects η_i would be excluded. Finally, in the truncated model, y_{it} consists of observations from the distribution of y_{it}^* conditional on $y_{it}^* > 0$.

Although most of the discussion on estimation methods will be conducted in terms of the first-order scalar autoregression presented above, the analysis is intended to cover the following p-th order vector autoregression

$$y_{it}^* = \delta_t + \sum_{j=1}^{p} A_j y_{i(t-j)}^* + \eta_i + v_{it} \tag{14}$$

$$E(v_{it} \mid y_{i1}^*, ..., y_{i(t-1)}^*) = 0$$

where y_{it}^* is a $g \times 1$ vector of (at least partly) latent variables, η_i is a $g \times 1$ vector of individual effects and δ_t is a vector of time effects treated as parameters to be estimated (in the empirical section we consider a bivariate model with $p = 2$ and time effects). The first-order scalar autoregression without time effects is notationally much simpler to work with and yet does not miss any essential aspect of the more general vector problem. Another remark is that our framework is consistent with, but does not require the stronger assumption $E(v_{it} \mid y_{i1}^*, ..., y_{i(t-1)}^*, \eta_i) = 0$.

3 Estimation and Hypothesis Testing

A. *Estimating the Reduced Form*

We begin by considering the estimation of the set of $(T-1)$ equations

$$E(y^*_{it} \mid y^*_{i1}, ..., y^*_{i(t-1)}) = \pi'_t x_{i(t-1)} \qquad (t = 2, ..., T) \tag{15}$$

in the case where the selection mechanism is censored Type I, so that $d_{it} = 1(y^*_{it} > 0)$. Let $h_{i(t-1)}$ be the indicator function of the event $(y^*_{i1} > 0, ...y^*_{i(t-1)} > 0)$. The coefficient vector π_t will be estimated using the subsample with $h_{i(t-1)} = 1$, so that each estimated π_t will be based on a different subsample. Notice that these subsamples are exogenously selected for the purpose of estimating π_t. The choice of estimator will depend on the assumptions we make about the distribution of $y^*_{it} \mid x_{i(t-1)}$. We give the details for a fully parametric normal model, but the same ideas can be applied to any asymptotically normal semiparametric method (like the trimmed least squares estimator due to Powell (1986), which is a popular semiparametric alternative that we employ in the empirical application, and is described in Appendix B). Our analysis can also accommodate exogenous variables with some straightforward modifications, which are discussed in Appendix C.

Assuming that

$$y^*_{it} \mid x_{i(t-1)} \sim N(\pi'_t x_{i(t-1)}, \sigma^2_t) \tag{16}$$

we can choose $\widehat{\theta}_t = (\widehat{\pi}'_t, \widehat{\sigma}_t)'$ to maximize

$$L_t = \sum_{i=1}^{N} h_{i(t-1)} \left[d_{it} \ln \frac{1}{\sigma_t} \phi \left(\frac{y_{it} - \pi'_t x_{i(t-1)}}{\sigma_t} \right) + \right. \tag{17}$$

$$(1-d_{it})\ln\Phi\left(\frac{-\pi_t' x_{i(t-1)}}{\sigma_t}\right)\Bigg] = \sum_{i=1}^{N}\ell_{it}(\pi_t,\sigma_t) \qquad (t=2,...,T)$$

where $\phi(.)$ and $\Phi(.)$ are, respectively, the *pdf* and the *cdf* for a standard normal variable.

The resulting stacked vector of estimates $\widehat{\pi} = (\widehat{\pi}_2'...\widehat{\pi}_T')'$ can be regarded as maximising the criterion function

$$L(\theta) = \sum_{t=2}^{T} L_t(\theta_t) \tag{18}$$

Thus, $\widehat{\pi}$ is not a full maximum likelihood estimator, since $L(\theta)$ does not take into account the correlation between variables corresponding to different time periods.

Subject to standard regularity conditions, a first order expansion of $\partial L(\widehat{\theta})/\partial\theta$ about the true value of θ gives

$$\left(-\frac{1}{N}diag\left\{\frac{\partial^2 L_t}{\partial\theta_t\partial\theta_t'}\right\}\right)\sqrt{N}(\widehat{\theta}-\theta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix}\partial\ell_{i2}/\partial\theta_2\\ \vdots \\ \partial\ell_{iT}/\partial\theta_T\end{pmatrix} + o_p(1) \tag{19}$$

from which a joint limiting normal distribution for $\sqrt{N}(\widehat{\theta}-\theta)$ can be obtained. A consistent estimator of the asymptotic covariance matrix of $\sqrt{N}(\widehat{\theta}-\theta)$ is given by

$$\widehat{V}_\theta = \widehat{H}_\theta^{-1}\widehat{\Psi}_\theta\widehat{H}_\theta^{-1} \tag{20}$$

where

$$\widehat{H}_\theta = diag\{N^{-1}\partial^2\widehat{L}_t/\partial\theta_t\partial\theta_t'\}$$

and

$$\widehat{\Psi}_\theta = N^{-1}\sum_{i=1}^{N}\left\{\frac{\widehat{\partial\ell}_{it}}{\partial\theta_t}\cdot\frac{\widehat{\partial\ell}_{is}}{\partial\theta_s'}\right\}$$

where $\widehat{\ell}_{it} = \ell_{it}(\widehat{\pi}_t, \widehat{\sigma}_t)$ and $\widehat{L}_t = \sum_{i=1}^{N} \widehat{\ell}_{it}$.

The previous method for the Type I Tobit model illustrates the particularities involved in the estimation of the system of equations (15). In the case of the Type II Tobit model, the parametric method most frequently used in practice is Heckman's two-step estimator (see Heckman (1979)) which can be applied to (15), equation by equation, on the basis of subsamples with $h_{i(t-1)} = 1$, where now

$$h_{i(t-1)} = 1(d_{i1} = 1, ..., d_{i(t-1)} = 1).$$

In such a case let us redefine $\widehat{\theta}_t = (\widehat{\pi}_t', \widehat{\phi}_t)'$ to minimize

$$S_t = \sum_{i=1}^{N} h_{it}[y_{it} - \pi_t' x_{i(t-1)} - \phi_t \lambda(\widehat{\gamma}_t' w_{it})]^2 = \sum_{i=1}^{N} s_{it}(\theta_t, \widehat{\gamma}_t) \qquad (21)$$

where $\lambda(.) = \phi(.)/\Phi(.)$, and $\widehat{\gamma}_t$ are probit estimates of γ_t in (13) using the subsample with $h_{i(t-1)} = 1$. That is, $\widehat{\gamma}_t$ maximizes

$$\begin{aligned} L_{pt} &= \sum_{i=1}^{N} h_{i(t-1)} \left[d_{it} \ln \Phi_{it} + (1 - d_{it}) \ln(1 - \Phi_{it})\right] \qquad (22) \\ &= \sum_{i=1}^{N} \ell_{pit}(\gamma_t) \end{aligned}$$

with $\Phi_{it} = \Phi(\gamma_t' w_{it})$. Using similar arguments as above we can obtain

$$H_t \sqrt{N}(\widehat{\theta}_t - \theta_t) = B_t \frac{1}{\sqrt{N}} \sum_{i=1}^{N} m_{it} + o_p(1) \qquad (t = 2, ..., T) \qquad (23)$$

where

$$m_{it} = \begin{pmatrix} \partial s_{it}/\partial \theta_t \\ \partial \ell_{pit}/\partial \gamma_t \end{pmatrix}$$

$$B_t = \left[I \vdots - \left(\frac{\partial^2 S_t}{\partial \theta_t \partial \gamma_t'} \right) \left(\frac{\partial^2 L_{pt}}{\partial \gamma_t \partial \gamma_t'} \right)^{-1} \right]$$

and

$$H_t = \frac{1}{N}\frac{\partial^2 S_t}{\partial\theta_t \partial\theta_t'}$$

From such expressions one can obtain a joint limiting normal distribution for $\sqrt{N}(\widehat{\theta} - \theta)$. A consistent estimator of the (t, s) block of the asymptotic covariance matrix of $\sqrt{N}(\widehat{\theta} - \theta)$ is given by

$$\widehat{Cov}(\widehat{\theta}_t, \widehat{\theta}_s) = \widehat{H}_t^{-1}\widehat{B}_t\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{m}_{it}\widehat{m}_{is}'\right)\widehat{B}_s'\widehat{H}_s^{-1} \tag{24}$$

where the symbols are as before but replacing true parameters by their estimated values.

There are also available asymptotically normal semiparametric two-step alternatives to Heckman's estimator, like the series estimator of Newey (1988) and the weighted kernel estimator of Powell (1989), both of which can also be applied to our context (see also Newey, Powell and Walker (1990)).

B. *Asymptotic Least Squares Estimation*

We turn to consider the estimation of the autoregressive coefficient α. Given consistent and asymptotically normal estimates of π_t, it would be possible to obtain joint minimum distance (MD) estimates of α and the λ's (see Chamberlain (1982)). However, the π_t are highly nonlinear functions of these parameters and, moreover, the λ's are not parameters of direct interest. For these reasons, it is more convenient to exploit the instrumental variables restrictions in the form of the relationships between α and the π_t given in (6). Stacking the equations we have

$$f(\alpha, \pi) = (p - \pi^1) - \alpha(\pi^1 - q) = 0 \tag{25}$$

where $p = (p_3'...p_T')', \pi^1 = (\pi_2'...\pi_{T-1}')'$ and $q = (q_1'...q_{T-2}')'$.

The vector p consists of functions of $\pi = (\pi_2' ... \pi_T')'$ of the form given in (7).

Given $T > 3$ and some consistent and asymptotically normal estimator $\widehat{\pi}$ together with a consistent estimator of its asymptotic covariance matrix $\widehat{V}_\pi$, say, an asymptotic least squares (ALS) estimator of α is given by

$$\widehat{\alpha} = \arg\min_{\alpha} f(\alpha, \widehat{\pi})' A_N f(\alpha, \widehat{\pi}) = \frac{(\widehat{\pi}^1 - q)' A_N (\widehat{p} - \widehat{\pi}^1)}{(\widehat{\pi}^1 - q)' A_N (\widehat{\pi}^1 - q)} \tag{26}$$

where $\widehat{p} = p(\widehat{\pi})$ and A_N is a weighting matrix (see Gourieroux, Monfort and Trognon (1985), and Gourieroux and Monfort (1995, 9.1)). The optimal choice of A_N for given $\widehat{\pi}$ is $\widehat{V}_r^{-1}$, which corresponds to the inverse of a consistent estimate of the asymptotic covariance matrix of $f(\alpha, \widehat{\pi})$:

$$V_r = Q V_\pi Q' \tag{27}$$

where Q is a $(1/2)(T-1)(T-2) \times (1/2)(T-1)T$ matrix given by

$$Q = \frac{\partial f(\alpha, \pi)}{\partial \pi'} = (I \otimes \pi^0)' \frac{\partial vecC}{\partial \pi'} + C \frac{\partial \pi^0}{\partial \pi'} - (1+\alpha) \frac{\partial \pi^1}{\partial \pi'}$$

and

$$C = diag[(1 : \pi_2), ..., (I_{T-2} : \pi_{T-1})]$$

$$\pi^0 = (\pi_3' ... \pi_T')'$$

The estimated variance $\widehat{V}_r$ can be obtained replacing V_π by $\widehat{V_\pi}$ in (27), and evaluating Q at $\widehat{\pi}$ and a preliminary consistent estimate of α. The optimal ALS estimator of α based on $f(\alpha, \widehat{\pi})$ is asymptotically equivalent to the optimal MD estimator of α based on $\widehat{\pi} - \pi(\alpha, \lambda_1, ..., \lambda_{T-1})$ and solved jointly with the λ's (see Alonso-Borrego and Arellano (1996)).

A consistent estimate of the asymptotic variance of $\sqrt{N}(\widehat{\alpha} - \alpha)$ for arbitrary A_N is given by

$$a\widehat{var}(\widehat{\alpha}) = \frac{(\widehat{\pi}^1 - q)' A_N \widehat{V}_r A_N (\widehat{\pi}^1 - q)}{[(\widehat{\pi}^1 - q)' A_N (\widehat{\pi}^1 - q)]^2} \tag{28}$$

C. *Estimates based on orthogonal deviations*

As an alternative to the moment conditions for the errors in first differences given in (3), we can use similar moments for the errors in forward orthogonal deviations (see Arellano and Bover (1995)). Contrary to the first differenced errors, the errors in orthogonal deviations are free from serial correlation if the original errors are not autocorrelated. Namely, we have

$$E\left[x_{i(t-1)}\left(\widetilde{y}^*_{it} - \alpha \widetilde{y}^*_{it(-1)}\right)\right] = 0 \qquad (t = 2, ..., T-1) \tag{29}$$

where

$$\widetilde{y}^*_{it} = c_t[y^*_{it} - \frac{1}{(T-t)}(y^*_{i(t+1)} + ... + y^*_{iT})]$$

$$\widetilde{y}^*_{it(-1)} = c_t[y^*_{it-1} - \frac{1}{(T-t)}(y^*_{it} + ... + y^*_{i(T-1)})]$$

and $c_t^2 = (T-t)/(T-t+1)$. As shown by Arellano and Bover (1995), in linear models the two sets of moments produce the same optimal GMM estimates, but this will not be the case here in general. Contrary to the linear case, in our context there are no natural one-step estimators that are optimal under certain assumptions. Thus we may expect preliminary consistent estimates based on first-differences (and the two-step estimates based on them) to show a different behaviour from those based on orthogonal deviations.

The moment conditions (29) translate into the following restrictions among linear projection coefficients and the parameter α:

$$\begin{aligned} g_t(\alpha, \pi) = {} & [\pi_{t|t-1} - \frac{1}{(T-t)}(\pi_{t+1|t-1} + ... + \pi_{T|t-1})] \\ & -\alpha[q_{t-1} - \frac{1}{(T-t)}(\pi_{t|t-1} + ... + \pi_{T-1|t-1})] = 0 \end{aligned} \tag{30}$$

where

$$\pi_{t+j|t} = [E(x_{it}x'_{it})]^{-1}E(x_{it}y^*_{i(t+j)})$$

so that in the previous notation $\pi_t = \pi_{t|t-1}$ and $p_t = \pi_{t|t-2}$. As before, the coefficients $\pi_{t+j|t}$ are linked by the law of iterated projections and can all be expressed as functions of $(\pi_2 ... \pi_T)$.[1] Stacking the g_t functions the analysis can proceed as in the first-difference case developed above.

D. *Testing the Overidentifying Restrictions*

When $T > 3$, there are $(1/2)(T-1)(T-2)-1$ overidentifying restrictions implied by the model which can be tested. The testing of these constraints is facilitated by the fact that the minimized optimal ALS criterion multiplied by the sample size has a limiting chi-squared distribution with degrees of freedom equal to the number of overidentifying restrictions (a proof of this result can be found in Gourieroux and Monfort (1989, vol. 2, p. 175, and 1995, vol. 2, p.154) and a similar method is proposed by Szroeter (1983)). Thus, the test statistic is given by

$$S = Nf(\widehat{\alpha}, \widehat{\pi})' \widehat{V}_r^{-1} f(\widehat{\alpha}, \widehat{\pi}) \tag{31}$$

This test statistic can be regarded as an extension to sample selection models of the Sargan specification tests of overidentifying restrictions for linear panel data GMM estimators considered by Arellano and Bond (1991) (cf. Sargan (1958 and 1988)). On the same lines, it is also possible to consider extensions of Sargan difference tests in order to discriminate between nested hypotheses.

[1] Specifically, we have

$$\pi_{t|t-j} = (I_{t-j} : \pi_{t-j+1|t-j} \ : ... : \ \pi_{t-1|t-j})\pi_{t|t-1}$$

E. *Consistent OLS Estimation Using Predicted Differences*

Calculation of the optimal ALS estimator of α requires a preliminary consistent estimator in order to obtain $\widehat{V}_r$. The following estimator can be computed in one step (given the $\widehat{\pi}_t$) and has a straightforward interpretation.

Let us define

$$\Delta\widehat{y}^*_{it|t-1} = x'_{i(t-1)}(\widehat{\pi}_t - q_{t-1}) \tag{32a}$$

$$\Delta\widehat{y}^*_{it|t-2} = x'_{i(t-2)}(\widehat{p}_t - \widehat{\pi}_{t-1}) \tag{32b}$$

Then we consider the OLS regression of $\Delta\widehat{y}^*_{it|t-2}$ on $\Delta\widehat{y}^*_{i(t-1)|t-2}$ for all periods and individuals with $h_{i(T-2)} = 1$:

$$\widetilde{\alpha} = \frac{\sum_{i=1}^N h_{i(T-2)} \sum_{t=3}^T \Delta\widehat{y}^*_{i(t-1)|t-2}\Delta\widehat{y}^*_{it|t-2}}{\sum_{i=1}^N h_{i(T-2)} \sum_{t=3}^T (\Delta\widehat{y}^*_{i(t-1)|t-2})^2} \tag{33}$$

Simple algebra reveals that

$$\widetilde{\alpha} = \frac{\sum_{t=3}^T (\widehat{\pi}_{t-1} - q_{t-2})' M_{t-2} (\widehat{p}_t - \widehat{\pi}_{t-1})}{\sum_{t=3}^T (\widehat{\pi}_{t-1} - q_{t-2})' M_{t-2} (\widehat{\pi}_{t-1} - q_{t-2})} \tag{34}$$

where $M_{t-2} = \sum_{i=1}^N h_{i(T-2)} x_{i(t-2)} x'_{i(t-2)}$. Therefore, $\widetilde{\alpha}$ is a non-optimal ALS estimator of the form given in (26) with weighting matrix given by

$$A_N = diag(M_1, ..., M_{T-2})$$

and estimated asymptotic covariance matrix of the form given by (28). Clearly, an alternative consistent OLS estimator can be calculated along the same lines using predicted orthogonal deviations as opposed to first-differences.

4 An Application to Female Labour Supply and Wages

We estimate separate autoregressive equations for annual hours of work and log wages (average hourly earnings) for two different samples of the Panel Study of Income Dynamics (PSID), covering the periods 1970-1976 and 1978-1984. In both cases the observations correspond to prime-age, white, married women from the random PSID sub-sample that were continuously married to the same husband, and who were 30-65 years old in 1968 (for the first sample) or 1976 (for the second). After selecting the samples with the criteria above, and removing some inconsistencies and non-respondents, we had 660 women available in the first sample, and 804 in the second.

The starting point for each dataset is the following second-order bivariate autoregression

$$\begin{aligned} h_{it}^{*} &= a_t^h + b_1^h h_{i(t-1)}^{*} + b_2^h h_{i(t-2)}^{*} + c_1^h \ln w_{i(t-1)}^{*} + \\ &\quad c_2^h \ln w_{i(t-2)}^{*} + \eta_i^h + v_{it}^h \end{aligned} \tag{35a}$$

$$\begin{aligned} \ln w_{it}^{*} = a_t^w + b_1^w h_{i(t-1)}^{*} + b_2^w h_{i(t-2)}^{*} + c_1^w \ln w_{i(t-1)}^{*} + \\ c_2^w \ln w_{i(t-2)}^{*} + \eta_i^w + v_{it}^w \qquad (t = 3, ..., 7) \end{aligned} \tag{35b}$$

where h_{it}^{*} is the supply of hours of work for individual i in period t and $\ln w_{it}^{*}$ is the natural log of the wage of individual i in period t. The variables η_i^h and η_i^w are individual effects, and a_t^h and a_t^w are time effects.

Inferences will be based on the following conditional moment restrictions

$$E(\Delta v_{i4}^h \quad | \quad h_{i1}^{*}, h_{i2}^{*}, \ln w_{i1}^{*}, \ln w_{i2}^{*}) = 0 \tag{36}$$

$$E(\Delta v_{i5}^h \quad | \quad h_{i1}^*, h_{i2}^*, h_{i3}^*, \ln w_{i1}^*, \ln w_{i2}^*, \ln w_{i3}^*) = 0$$
$$E(\Delta v_{i6}^h \quad | \quad h_{i1}^*, ..., h_{i4}^*, \ln w_{i1}^*, ..., \ln w_{i4}^*) = 0$$
$$E(\Delta v_{i7}^h \quad | \quad h_{i1}^*, ..., h_{i5}^*, \ln w_{i1}^*, ..., \ln w_{i5}^*) = 0$$

and similarly for first-differenced log wage errors (actually, we shall use errors in orthogonal deviations for which similar conditions hold). Both h_{it}^* and w_{it}^* are subject to censoring with a common selection mechanism. The unconditional non-participation rates for all individuals and time periods are around 50 percent in the first sample and 40 percent in the second. However, conditional non-participation rates for the sequence of sub-samples on which inferences will be based are much lower as can be seen from Table 1. The frequency of non-participants is under 10 percent for the four subsamples corresponding to the period 1970-76, and even lower for those of the more recent period. This suggests that LDV estimates of the linear projection coefficients π_t will have a small bias whatever the truth of the linear conditional expectation assumption and of the specification of the selection mechanism. Some additional descriptive information on the two datasets is provided on Table A1 in the Appendix.

Tables 2 and 3 contain results for the hours and wage equations, respectively. To the basic autoregressive equations we have added two children variables which are treated as predetermined variables in the estimation (a dummy for a child less than 6 years-old and another for a child between 6 and 9). All the results we present include these children dummies, but their exclusion does not alter the observed dynamics of hours and wages in our data.

All the results reported in both tables are optimal ALS estimates based on moment conditions in orthogonal deviations. The preliminary consistent estimates are OLS using predicted differences in orthogonal deviations. The differences among the columns are in the way the reduced form coefficients are estimated, or in the number of moment conditions used. Columns labelled OLS present optimal ALS estimates based on OLS estimates of the reduced form coefficients for the sequentially censored subsamples. Thus, these estimates do not correct for selectivity, but given the low conditional non-participation rates in the samples we would not expect them to differ substantially from those with selectivity corrections. The estimates shown in the remaining two columns are based on Heckman's estimates of the reduced form coefficients. Those in the second columns of Tables 2 and 3 use the same moments as the ones in the first columns, and as expected the differences between the two sets of estimates are small. Finally, the estimates in the third columns are also based on Heckman's estimates of the reduced form, but only use instrumental variables dated t-3 or less. The motivation for these estimates arises from a concern with measurement errors in observed wages, which would invalidate the moment conditions given in (36). However, if the measurement error is not serially correlated, backdating the conditioning variables one period the mean independence of the first-differenced errors is restored. Tables A2 and A3 in the Appendix show some additional ALS estimates of the hours and wage equations based on Powell's (1986) symmetrically censored LS estimates of the reduced form, and also Tobit estimates with and without allowance for measurement errors in wages.

Starting with the results for the hours equations, there are some significant differences between the estimates with and without measurement error corrections which, nevertheless, are not signaled by the test statistics of overidentifying restrictions. There appears to be an increase of the effect of past hours of work on current hours in the second period, which is consistent with the notion of a trend towards a steadier involvement with the labour market for those who participate. The estimated effects of lagged wages on hours of work are not robust, and lack a clear pattern. In the first sample the effect is stronger when allowing for measurement error in wages, but this situation is reversed in the second sample. Lastly, the empirical coefficients of the children dummies have the expected sign, but their magnitude becomes consistently smaller when moving from the first panel to the second.

Turning to the wage equations, the estimates in this case exhibit larger differences between the two periods. Firstly, there is a positive effect of lagged wages (net of individual effects) in the first period, which disappears altogether -or becomes even negative- in the second period. Secondly, there is a positive effect of lagged hours on wages whose size doubles from the first period to the second. The change in the effect of lagged wages suggests higher occupational mobility, while the change in the effect of lagged hours points to higher returns to experience. Finally, in no case have the children dummies a significant effect on wages.

Table 4 presents alternative estimates without individual effects for comparison. The reported estimates are pooled OLS for each of the two samples of participants in previous periods (Table A4 presents Tobit estimates for the same models and data). As expected, the wage equations without permanent effects show a stronger autoregressive pattern for lagged wages. It is, however, noticeable the change in the pattern of serial dependence in wages between the two panels that is broadly consistent with the results found for the models with individual effects. Notice that these equations include education and age variables, whose effects are captured by individual and year effects in the equations in orthogonal deviations.

We now turn to interpret the previous empirical autoregressions in terms of a life-cycle labour supply framework. Let us consider the following labour supply equation

$$h_{it}^{*} = \mu_{it} + \beta \ln w_{it}^{*} + \beta \ln \lambda_{it} \tag{37}$$

where μ_{it} reflects variation in preferences due to individual and time specific factors, λ_{it} is the marginal utility of wealth and the parameter β divided by h_{it}^{*} represents the intertemporal substitution elasticity. Browning, Deaton and Irish (1985, pp. 521-2) obtain the profit function from which this equation can be derived.[2] They also show (among other authors, see also Heckman and MaCurdy (1980) and MaCurdy (1981)) that the first difference of $\ln \lambda_{it}$ can be approximated by a time effect plus a serially uncorrelated innovation ξ_{it}. The innovation ξ_{it} will be correlated with current wages but uncorrelated to all lagged variables in the individual's information set.

We assume that μ_{it} can be represented as the sum of time and individual effects, the effects of children, and a disturbance term v_{it}. Given the observed autoregressive behaviour of hours of work and the fact that in model (37)

[2] Equation (37) assumes that labour supply and goods are additive within periods. The introduction of an extra term of the form ${w_{it}^{*}}^{-1/2}$ to relax this assumption (as suggested by Browning et al. (1995)) would make impossible the direct mapping with our VAR. The estimates of the effect of this additional variable reported by Browning et al. for cohorts male labour supply never turned out to be significantly different from zero.

this can only be rationalized through serial correlation in v_{it}, we specify

$$v_{it} = \rho v_{i(t-1)} + \varepsilon_{it}.$$

Excluding children dummies for simplicity of presentation, the labour supply equation in first differences can be written in the form

$$\Delta h^*_{it} = \Delta\delta_t + \rho\Delta h^*_{i(t-1)} + \beta\Delta\ln w^*_{it} - \rho\beta\Delta\ln w^*_{i(t-1)} + (\xi_{it} - \rho\xi_{i(t-1)} + \Delta\varepsilon_{it}) \quad (38)$$

Let us denote a simplified model for the change in log wages excluding second-order lags as

$$\Delta\ln w^*_{it} = \Delta a^w_t + c\Delta\ln w^*_{i(t-1)} + b\Delta h^*_{i(t-1)} + \Delta v^w_{it} \quad (39)$$

Now combining equations (38) and (39), the life-cycle labour supply model implies that the process for hours follows

$$\Delta h^*_{it} = \Delta a^h_t + m\Delta h^*_{i(t-1)} + \gamma\Delta\ln w^*_{i(t-1)} + \Delta v^h_{it} \quad (40)$$

where

$$\begin{aligned} m &= \rho + \beta b \quad (41)\\ \gamma &= \beta(c-\rho)\\ a^h_t &= \delta_t + \beta a^w_t\\ \Delta v^h_{it} &= \xi_{it} - \rho\xi_{i(t-1)} + \Delta\varepsilon_{it} + \beta\Delta v^w_{it} \end{aligned}$$

Therefore, under the previous interpretation the coefficient on lagged wages in the autoregressive hours equation can be regarded as an estimate of $\beta(c-\rho)$, where β divided by h^*_{it} gives the intertemporal labour supply elasticity. However, given the lack of robustness of the estimated effects of lagged wages on hours of work, we may expect the implied estimates of β to be very imprecise (indeed if $c - \rho = 0$ the parameter β would be unidentified). This result is similar to the finding of Abowd and Card (1989) for men labour supply using moving average representations, and suggests that the dynamics of hours and wages in these data sets contains little information on intertemporal labour supply responses.

5 Concluding Remarks

The methods developed in this paper are based on the observation that the subsamples which only include individuals without censored past observations (those with $h_{i(t-1)} = 1$) are exogenously selected for the purpose of estimating features of the distribution of y_{it}^* conditional on its past. In the application to female labour supply and wages presented in Section 3, it turns out that most of the selectivity due to censoring is accounted by the permanent effects, and our methods make precise the sense in which this is so. For other applications, however, these procedures may retain very few or no observations with $h_{i(t-1)} = 1$ for the larger values of t. In practice, such problem could be addressed by considering distributions that are conditional on the more recent observations only. In effect, if the linearity of the conditional expectation of y_{it}^* given $(y_{i1}^*, ..., y_{it-1}^*)$ holds, we might expect this to hold for any time sequence since the initial observation is often arbitrary. In such case, we could rely on the linearity of the conditional expectation of y_{it}^* given $y_{i(t-1)}^*, ..., y_{i(t-s)}^*$ for any t and s in devising asymptotic least squares estimates of the parameters of interest. This may create a trade off between the number of moment restrictions being used and the actual sample size, which remains to be explored. Future work will also have to address ways of relaxing some of the distributional asumptions, and consider ways of introducing stationarity restrictions.

References

[1] Abowd, J.M. and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes", *Econometrica*, 57, 411-445.

[2] Alonso-Borrego, C. and M. Arellano (1996): "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data", Working Paper 9612, CEMFI, Madrid.

[3] Anderson, T.W. and C. Hsiao (1981): "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606.

[4] Amemiya, T. (1985): *Advanced Econometrics*, Blackwell, Oxford.

[5] Arellano, M. and S.R. 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.

[6] Arellano, M. and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models", *Journal of Econometrics*, 68, 29-51.

[7] Browning, M., A. Deaton and M. Irish (1985): "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle", *Econometrica*, 53, 503-543.

[8] Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data", *Journal of Econometrics*, 18, 5-46.

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

[10] Gourieroux, C., A. Monfort and A. Trognon (1985): "Moindres Carrés Asymptotiques", *Annales de l'Inséé*, 58, 91-122.

[11] Gourieroux, C. and A. Monfort (1989): *Statistique et Modèles Econométriques*, Vol. 2, Économica, Paris.

[12] Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models*, Vols. 1 and 2, Cambridge University Press.

[13] Heckman, J.J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-161.

[14] 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.

[15] Heckman, J.J. and T.E. MaCurdy (1980): "A Lyfe-Cycle Model of Female Labour Supply", *Review of Economic Studies*, 47, 47-74.

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

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

[18] Honoré, B. (1993): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables", *Journal of Econometrics*, 59, 35-61.

[19] Huber, P.J. (1967): "The Behaviour of Maximum Likelihood Estimates Under Nonstandard Conditions", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221-233.

[20] MaCurdy, T.E. (1981): "An Empirical Model of Labor Supply in a Life-Cycle Setting", *Journal of Political Economy*, 89, 1059-1085.

[21] Newey, W.K. (1988): "Two Step Series Estimation of Sample Selection Models", unpublished discussion paper, Princeton University.

[22] Newey, W.K., J.L. Powell and J. R. Walker (1990): "Semiparametric Estimation of Selection Models: Some Empirical Results", *The American Economic Review Papers and Proceedings*, 80,324-328.

[23] Powell, J.L. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models", *Econometrica,* 54, 1435-1460.

[24] Powel, J.L. (1989): "Semiparametric Estimation of Censored Selection Models", mimeo.

[25] Sargan, J.D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables", *Econometrica*, 26, 393-415.

[26] Sargan, J.D. (1988): "Testing for Misspecification after Estimating Using Instrumental Variables", in E. Maasoumi (ed.): *Contributions to Econometrics: John Denis Sargan*, Vol. 1, Cambridge University Press.

[27] Szroeter, J. (1983): "Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions", *Econometrica*, 51, 335-353.

TABLE 1

Conditional Participation Frequencies

Subsample	Size	No. of non-participants	Percentage
First Sample, 1970-76, $N = 660$ (Period 1 = 1970)			
$h_1 > 0, h_2 > 0$	$N_1 = 303$	$no(h_3 = 0 \mid h_1 > 0, h_2 > 0) = 30$	9.9
$h_1 > 0, ..., h_3 > 0$	$N_2 = 273$	$no(h_4 = 0 \mid h_1 > 0, ..., h_3 > 0) = 13$	4.8
$h_1 > 0, ..., h_4 > 0$	$N_3 = 260$	$no(h_5 = 0 \mid h_1 > 0, ..., h_4 > 0) = 16$	6.2
$h_1 > 0, ..., h_5 > 0$	$N_4 = 244$	$no(h_6 = 0 \mid h_1 > 0, ..., h_5 > 0) = 18$	7.4
$h_1 > 0, ..., h_6 > 0$	$N_5 = 226$	$no(h_7 = 0 \mid h_1 > 0, ..., h_6 > 0) = 16$	7.1
Second Sample, 1978-84, $N = 804$ (Period 1=1978)			
$h_1 > 0, h_2 > 0$	$N_1 = 438$	$no(h_3 = 0 \mid h_1 > 0, h_2 > 0) = 25$	5.7
$h_1 > 0, ..., h_3 > 0$	$N_2 = 413$	$no(h_4 = 0 \mid h_1 > 0, ..., h_3 > 0) = 27$	6.5
$h_1 > 0, ..., h_4 > 0$	$N_3 = 386$	$no(h_5 = 0 \mid h_1 > 0, ..., h_4 > 0) = 25$	6.5
$h_1 > 0, ..., h_5 > 0$	$N_4 = 361$	$no(h_6 = 0 \mid h_1 > 0, ..., h_5 > 0) = 23$	6.4
$h_1 > 0, ..., h_6 > 0$	$N_5 = 338$	$no(h_7 = 0 \mid h_1 > 0, ..., h_6 > 0) = 19$	5.6

TABLE 2
Hours Equations
(Optimal ALS Estimates Using Orthogonal Deviations)

	OLS Reduced Form Projections	LS with Select. Correction	Measurement Error Correction
	Panel A: Sample 1970-76		
h_{t-1}	0.167	0.188	0.310
	(3.38)	(3.10)	(3.74)
h_{t-2}	0.126	0.120	0.145
	(2.95)	(2.32)	(2.30)
$\ln w_{t-1}$	20.37	83.01	287.59
	(0.48)	(1.41)	(4.88)
$\ln w_{t-2}$	94.76	123.84	85.40
	(2.81)	(3.18)	(2.29)
D_{1t}	−312.46	−310.26	−362.02
	(−3.99)	(−3.73)	(−3.51)
D_{2t}	−43.16	−41.95	−98.45
	(−0.87)	(−0.76)	(−1.30)
S	82.68(58)	63.47(58)	56.00(42)
	Panel B: Sample 1978-84		
h_{t-1}	0.466	0.379	0.337
	(11.2)	(6.46)	(5.29)
h_{t-2}	−0.051	−0.072	0.002
	(−1.47)	(−2.08)	(0.52)
$\ln w_{t-1}$	253.56	232.72	−13.01
	(4.91)	(4.20)	(−0.22)
$\ln w_{t-2}$	−20.44	−28.52	−104.67
	(−0.58)	(−0.77)	(−3.07)
D_{1t}	−250.07	−209.30	−152.67
	(−3.85)	(−2.84)	(−1.43)
D_{2t}	−16.62	7.97	−246.94
	(−0.29)	(0.14)	(−4.38)
S	54.61(58)	56.09(58)	37.73(42)

Notes to Table 2

(i) Time dummies are included in all equations.
(ii) Figures in parentheses are t-ratios.
(iii) $D_{1t} = 1$ if at least one child less than 6 years old is present.
$D_{2t} = 1$ if at least one child older than 5 and younger than 10 years old is present.
(iv) S is the chi-squared test statistic of overidentifying restrictions.
(v) Estimates in cols. (1) and (2) use variables dated $t-2$ and less as instruments, while the estimates in col (2) use only variables dated at most $t-3$.
(vi) Estimates in cols. (2) and (3) use Heckman's estimates of the reduced form.

TABLE 3
Wage Equations
(Optimal ALS Estimates Using Orthogonal Deviations)

	OLS Reduced Form Projections	LS with Selectivity Correction	Measurement Error Correction
	Panel A: Sample 1970-76		
h_{t-1}	0.00027	0.00020	0.00022
	(7.15)	(3.87)	(4.48)
h_{t-2}	0.00006	0.00006	−0.00025
	(1.78)	(1.60)	(−5.90)
$\ln w_{t-1}$	0.140	0.210	0.576
	(2.92)	(3.72)	(23.0)
$\ln w_{t-2}$	0.078	0.079	−0.281
	(2.72)	(1.84)	(−10.7)
D_{1t}	−0.015	−0.014	−0.071
	(−0.41)	(−0.34)	(−1.13)
D_{2t}	0.149	0.118	0.009
	(4.48)	(3.15)	(0.17)
S	76.98(58)	80.37(58)	56.02(42)
	Panel B: Sample 1978-84		
h_{t-1}	0.00040	0.00043	0.00036
	(11.9)	(10.3)	(8.30)
h_{t-2}	0.00012	0.00012	0.00012
	(5.56)	(4.79)	(3.82)
$\ln w_{t-1}$	−0.167	−0.089	−0.172
	(−3.99)	(−1.81)	(−4.45)
$\ln w_{t-2}$	−0.106	−0.092	−0.132
	(−5.54)	(−4.07)	(−6.19)
D_{1t}	0.020	0.005	0.269
	(0.46)	(0.11)	(5.08)
D_{2t}	−0.038	−0.051	0.053
	(−1.00)	(−1.33)	(1.35)
S	48.57(58)	44.77(58)	32.52(42)

See Notes to Table 2

TABLE 4
Levels Equations without Individual Effects
OLS estimates for the sample of participants in previous periods

Panel A: Sample 1970-76 (1306 observations)

	Hours equations		Wage equations	
h_{t-1}	0.663	0.659	0.00029	0.00029
	(13.9)	(13.9)	(7.74)	(7.76)
h_{t-2}	0.127	0.131	-0.00011	-0.00010
	(2.75)	(2.87)	(-2.89)	(-2.63)
$\ln w_{t-1}$	205.46	205.69	0.514	0.484
	(4.62)	(4.45)	(10.3)	(9.66)
$\ln w_{t-2}$	-42.59	-34.50	0.239	0.206
	(-0.92)	(-0.70)	(4.56)	(3.94)
D_{1t}	–	-191.39	–	-0.104
		(-1.79)		(-1.41)
D_{2t}	–	-61.66	–	-0.011
		(-1.09)		(-0.24)
Education	–	-14.15	–	0.040
		(-1.13)		(3.44)
Age	–	-5.35	–	-0.0035
		(-2.89)		(-2.52)

TABLE 4 (continued)

Levels Equations without Individual Effects
OLS estimates for the sample of participants in previous periods

Panel B: Sample 1978-84 (1936 observations)

	Hours equations		Wage equations	
h_{t-1}	0.542	0.539	0.0002	0.0002
	(16.0)	(16.2)	(5.25)	(5.92)
h_{t-2}	0.218	0.215	−0.00007	−0.000002
	(6.59)	(6.62)	(−0.22)	(−0.08)
$\ln w_{t-1}$	4.20	−0.073	0.349	0.325
	(0.12)	(−0.002)	(8.52)	(8.24)
$\ln w_{t-2}$	39.99	33.36	0.360	0.329
	(1.00)	(1.07)	(8.67)	(8.03)
D_{1t}	–	−268.89	–	−0.363
		(−4.19)		(−5.27)
D_{2t}	–	−19.28	–	0.032
		(−0.28)		(0.51)
Education	–	9.37	–	0.066
		(1.02)		(6.74)
Age	–	−3.96	–	0.0004
		(−2.60)		(0.29)

Notes:
(i) Time dummies are included in all equations.
(ii) Figures in parentheses are t-ratios robust to heteroskedasticity.
(iii) $D_{1t} = 1$ if at least one child less than 6 years old is present.
$D_{2t} = 1$ if at least one child older than 5 and younger than 10 years old is present.

TABLE A1
Descriptive Statistics

	Full sample		Sample of participants in previous periods	
	Mean	St. Dev.	Mean	St. Dev.
	Panel A: Sample 1970-76			
Participation	0.506	0.500	0.928	0.257
Hours	672.7	839.4	1379.6	732.9
Wages	1.847	2.899	3.837	3.060
Partic. (1972)	0.483	0.500	0.901	0.299
Hours (1972)	650.4	852.5	1319.6	808.6
Wages (1972)	1.571	2.422	3.102	2.413
Partic. (1976)	0.515	0.500	0.929	0.257
Hours (1976)	682.0	827.7	1426.0	687.9
Wages (1976)	2.379	3.771	4.570	3.282
D_1	0.096	0.295	0.038	0.192
D_2	0.115	0.319	0.084	0.278
Sample size	4620($= 660 \times 7$)		1306	
	Panel B: Sample 1978-84			
Participation	0.615	0.487	0.939	0.238
Hours	849.9	871.2	1433.2	719.5
Wages	3.637	5.094	6.367	5.961
Partic. (1980)	0.651	0.477	0.943	0.232
Hours (1980)	883.2	870.8	1397.1	726.9
Wages (1980)	3.422	4.185	5.135	4.491
Partic. (1984)	0.610	0.488	0.950	0.219
Hours (1984)	857.3	888.4	1486.3	722.3
Wages (1984)	4.561	7.214	8.040	8.990
D_1	0.082	0.274	0.069	0.253
D_2	0.078	0.269	0.084	0.278
Sample size	5628($= 804 \times 7$)		1936	

Notes to Table A1:

(i) Wages are average hourly earnings.
(ii) $D_1 = 1$ if at least one child less than 6 years old is present.
(iii)$D_2 = 1$ if at least one child older than 5 and younger than 10 years old is present.
(iv) The "samples of participants in previous periods" includes participants and non participants conditional on participation in the previous years available in the data. They effectively combine the sequentially censored subsamples described in Table 1.

TABLE A2
Hours Equations
(Optimal ALS Estimates Using Orthogonal Deviations)

	Tobit Reduced Form Projections	SCLS Reduced Form Projections	Measurement Error Correction
	Panel A: Sample 1970-76		
h_{t-1}	0.423	0.179	0.248
	(6.34)	(3.89)	(2.85)
h_{t-2}	−0.001	0.155	0.167
	(−0.26)	(3.74)	(2.99)
$\ln w_{t-1}$	−101.19	9.82	−94.44
	(−1.35)	(0.28)	(−1.13)
$\ln w_{t-2}$	−45.35	83.53	−87.84
	(−1.04)	(2.48)	(−1.69)
D_{1t}	−302.53	−312.68	−236.47
	(−4.61)	(−4.56)	(−2.74)
D_{2t}	−14.46	−60.09	−229.18
	(−0.20)	(−1.18)	(−2.44)
S	93.22 (58)	91.24 (58)	74.02(42)
	Panel B: Sample 1978-84		
h_{t-1}	0.219	0.478	0.394
	(3.78)	(11.1)	(6.49)
h_{t-2}	−0.071	−0.055	0.043
	(−2.31)	(−1.71)	(0.83)
$\ln w_{t-1}$	141.34	207.13	−255.38
	(3.02)	(4.14)	(−4.76)
$\ln w_{t-2}$	−119.07	−32.50	−166.13
	(−4.41)	(−0.98)	(−6.15)
D_{1t}	−294.65	−309.36	−247.40
	(−3.70)	(−4.84)	(−2.32)
D_{2t}	25.32	−29.51	−138.49
	(0.42)	(−0.54)	(−1.64)
S	48.83(58)	74.64(58)	45.98(42)

Notes to Table A2

(i) Time dummies are included in all equations.
(ii) Figures in parentheses are t-ratios.
(iii) $D_{1t} = 1$ if at least one child less than 6 years old is present.
$D_{2t} = 1$ if at least one child older than 5 and younger than 10 years old is present.
(iv) S is the chi-squared test statistic of overidentifying restrictions.
(v) Estimates in cols. (1) and (2) use variables dated $t - 2$ and less as instruments, while the estimates in col (3) use only variables dated at most $t - 3$.
(vi) Estimates in cols. (1) and (3) use Tobit estimates of the reduced form.
(vii) Estimates in col. (2) use Powell's (1986) symmetrically censored least squares (SCLS) estimates of the reduced form.

TABLE A3
Wage Equations
(Optimal ALS Estimates Using Orthogonal Deviations)

	Tobit Reduced Form Projections	SCLS Reduced Form Projections	Measurement Error Correction
	Panel A: Sample 1970-76		
h_{t-1}	0.0004	-0.00002	0.0004
	(8.07)	(-0.44)	(7.65)
h_{t-2}	0.0001	-0.0002	0.00015
	(2.70)	(-1.20)	(3.17)
$\ln w_{t-1}$	0.237	0.282	0.301
	(3.89)	(2.17)	(5.79)
$\ln w_{t-2}$	0.051	0.119	0.022
	(1.54)	(1.85)	(0.71)
D_{1t}	-0.063	1.218	-0.410
	(-0.48)	(7.45)	(-3.16)
D_{2t}	0.143	0.525	0.139
	(2.31)	(9.75)	(2.63)
S	53.67(58)	59.05(58)	48.20(42)
	Panel B: Sample 1978-84		
h_{t-1}	0.00033	0.00029	0.00034
	(5.93)	(8.67)	(5.46)
h_{t-2}	0.00009	0.00007	0.00011
	(2.87)	(3.47)	(2.16)
$\ln w_{t-1}$	0.067	-0.217	-0.197
	(1.75)	(-5.07)	(-3.77)
$\ln w_{t-2}$	0.025	-0.098	-0.161
	(1.10)	(-4.17)	(-5.43)
D_{1t}	-0.352	-0.410	-0.464
	(-4.42)	(-13.1)	(-4.94)
D_{2t}	0.122	-0.107	-0.032
	(1.60)	(3.14)	(-0.32)
S	53.35(58)	60.11(58)	52.83(42)

See Notes to Table A2

TABLE A4
Levels Equations without Individual Effects
Tobit estimates for the sample of participants in previous periods

Panel A: Sample 1970-76 (1306 observations)

	Hours equations		Wage equations	
h_{t-1}	0.717	0.714	0.00034	0.00034
	(13.4)	(13.4)	(7.73)	(7.75)
h_{t-2}	0.101	0.105	-0.00014	-0.00013
	(1.97)	(2.08)	(-3.14)	(-2.92)
$\ln w_{t-1}$	245.70	247.36	0.560	0.529
	(4.77)	(4.62)	(10.1)	(9.47)
$\ln w_{t-2}$	-63.23	-54.93	0.225	0.194
	(-1.23)	(-1.02)	(3.97)	(3.45)
D_{1t}	–	-240.01	–	-0.139
		(-2.03)		(-1.69)
D_{2t}	–	-58.30	–	-0.012
		(-0.98)		(-0.23)
Education	–	-16.00	–	0.038
		(-1.20)		(3.05)
Age	–	-5.96	–	-0.0039
		(-2.99)		(-2.30)

TABLE A4 (continued)

Levels Equations without Individual Effects
Tobit estimates for the sample of participants in previous periods

Panel B: Sample 1978-84 (1936 observations)

	Hours equations		Wage equations	
h_{t-1}	0.569	0.566	0.0002	0.0002
	(15.8)	(15.9)	(5.45)	(6.09)
h_{t-2}	0.217	0.213	−0.00001	−0.000008
	(6.18)	(6.18)	(−0.32)	(−0.20)
$\ln w_{t-1}$	12.09	7.24	0.362	0.336
	(0.32)	(0.19)	(8.19)	(7.91)
$\ln w_{t-2}$	28.32	30.57	0.364	0.333
	(0.83)	(0.90)	(8.22)	(7.64)
D_{1t}	–	−302.06	–	0.032
		(−4.30)		(0.50)
D_{2t}	–	−14.82	–	0.0002
		(−0.21)		(0.12)
Education	–	11.17	–	0.068
		(1.15)		(6.53)
Age	–	−4.21	–	0.004
		(−2.63)		(0.10)

Notes:

(i) Time dummies are included in all equations.
(ii) Figures in parentheses are t-ratios.
(iii) $D_{1t} = 1$ if at least one child less than 6 years old is present.
$D_{2t} = 1$ if at least one child older than 5 and younger than 10 years old is present.

Appendix B: Type I Tobit with Symmetric Trimming

Assuming that $y^*_{it} \mid x_{i(t-1)}$ has a symmetric distribution with heteroskedasticity of unknown form, we can estimate consistently the π_t using Powell's (1986) symmetrically censored least squares method (SCLS). The SCLS estimator $\widetilde{\pi}_t$ solves the iteration

$$\begin{aligned} \widetilde{\pi}_t^{r+1} &= \left(\sum_{i=1}^{N} h_{i(t-1)}\varphi_{it}(\widetilde{\pi}_t^r)x_{i(t-1)}x'_{i(t-1)}\right)^{-1} \\ &\quad \sum_{i=1}^{N} h_{i(t-1)}\varphi_{it}(\widetilde{\pi}_t^r)x_{i(t-1)}\min\{y_{it}, 2x'_{i(t-1)}\widetilde{\pi}_t^r\} \end{aligned} \tag{B.1}$$

where $\varphi_{it}(\widetilde{\pi}_t^r) = 1(x'_{i(t-1)}\widetilde{\pi}_t^r > 0)$ for $r \geq 1$ and $\varphi_{it}(\widetilde{\pi}_t^0) = 1$ (that is, the initial value is the OLS estimator).

In this case, since the SCLS estimation criterion is not differentiable we cannot use the argument of Section 3 for normal Tobit or Heckman's estimator in order to obtain the asymptotic covariance matrix of $\widetilde{\pi} = (\widetilde{\pi}_2 ... \widetilde{\pi}'_T)'$. However, Powell (1986) following the approach of Huber (1967) obtains an asymptotic relation (see equation (A.15) of his paper) which combined for our (T-1) equations can be written as

$$(diag\{C_t\})\sqrt{N}(\widetilde{\pi} - \pi) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} \psi_{i2}(\pi_2) \\ \vdots \\ \psi_{iT}(\pi_T) \end{pmatrix} + o_p(1) \tag{B.2}$$

where

$$C_t = \frac{1}{N}\sum_{i=1}^{N} E\left[h_{i(t-1)}1(0 < y_{it} < 2x'_{i(t-1)}\pi_t)x_{i(t-1)}x'_{i(t-1)}\right]$$

and

$$\psi_{it}(\pi_t) = 1(x'_{i(t-1)}\pi_t > 0)v_{it}x_{i(t-1)}$$

$$v_{it} = \min\{y_{it}, 2x'_{i(t-1)}\pi_t\} - x'_{i(t-1)}\pi_t$$

From this expression a joint limiting normal distribution for $\sqrt{N}(\widetilde{\pi} - \pi)$ can be obtained. A consistent estimator of the asymptotic covariance matrix of $\sqrt{N}(\widetilde{\pi} - \pi)$ is given by

$$\widetilde{V}_\pi = (diag\{\widetilde{C}_t^{-1}\})\widetilde{\Psi}_\pi(diag\{\widetilde{C}_t^{-1}\}) \tag{B.3}$$

where $\widetilde{C}_t$ is a "natural" estimator of C_t and $\widetilde{\Psi}_\pi$ has the following block structure:

$$\widetilde{\Psi}_r = \frac{1}{N}\sum_{i=1}^{N}\{\psi_{it}(\widetilde{\pi}_t)\psi'_{is}(\widetilde{\pi}_s)\}$$

Appendix C: Models with Exogenous Variables

The analysis for autoregressive models can accommodate exogenous variables in a straightforward manner. Suppose that model (1) is extended to include a strictly exogenous variable z_{it}:

$$y_{it}^{*} = \alpha y_{i(t-1)}^{*} + \beta z_{it} + \eta_i + v_{it} \tag{C.1}$$

such that

$$E(v_{it} \mid y_{i1}^{*}, ..., y_{i(t-1)}^{*}, z_{i1}, ..., z_{iT}) = 0 \tag{C.2}$$

This assumption implies that

$$E(x_{i(t-2)} \Delta v_{it}) = 0 \qquad (t = 3, ..., T) \tag{C.3}$$

or

$$(p_t - \pi_{t-1}) = \alpha(\pi_{t-1} - q_{t-2}) + \beta h_t \qquad (t = 3, ..., T) \tag{C.4}$$

where $x_{i(t-2)}$ has been re-defined as

$$x_{i(t-2)} = (y_{i1}^{*}, ..., y_{i(t-2)}^{*}, z_{i1}, ..., z_{iT})' \tag{C.5}$$

and p_t, π_{t-1} and q_{t-2} are also re-defined accordingly. h_t is a selection vector of known constants such that

$$h_t = [E(x_{i(t-2)} x_{i(t-2)}')]^{-1} E(x_{i(t-2)} \Delta z_{it}) \tag{C.6}$$

The discussion in the main text will apply provided we can assume

$$E(y_{it}^{*} \mid x_{i(t-1)}) = \pi_t' x_{i(t-1)} \tag{C.7}$$

together with sufficient distributional assumptions about the distribution of $y_{it}^{*} \mid x_{i(t-1)}$ in order to identify its mean in the presence of selectivity.