

# On the estimation of a fixed effects model with selectivity bias

Marno Verbeek \*

*Tilburg University, 5000 LE Tilburg, The Netherlands*

Received 6 February 1990

Accepted 21 March 1990

In case of sample selectivity the maximum likelihood estimator of the parameters in a model with fixed effects will not be consistent when the number of time periods is small. In this paper, we present a transformation to eliminate the fixed individual effects and show that the corresponding marginal maximum likelihood estimator is computationally feasible and can be used to estimate the remaining parameters consistently even if the number of time periods is finite.

## 1. Introduction

A commonly used model to analyze individual behavior using a panel data set is the following linear model:

$$y_{it} = X_{it}\beta + \alpha_i + \epsilon_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, N, \quad (1)$$

where  $X_{it}$  is a row vector of exogenous variables,  $\beta$  a column vector of unknown parameters of interest and the  $\alpha_i$  are unobserved individual specific effects. Because it is well known that in many applications these effects may be correlated with the explanatory variables [see Mundlak (1961) for a classical example of this topic] treating them as i.i.d. errors will usually lead to inconsistent estimators. A convenient way to circumvent this problem is to treat the  $\alpha_i$  as fixed unknown parameters and to estimate them either implicitly or explicitly.

It is assumed that observations for  $y_{it}$  are only available if an unobserved variable  $r_{it}^*$  is nonnegative, for which we assume

$$r_{it}^* = Z_{it}\gamma + \xi_i + \eta_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, N, \quad (2)$$

where  $Z_{it}$  is a row vector of predetermined variables, possibly containing (partly) the same variables as  $X_{it}$ , and  $\xi_i$  is an individual specific component of the error term. The (observed) indicator variable

\* Financial support by the Netherlands Organization for Scientific Research (N.W.O.) is gratefully acknowledged. I would like to thank Peter Kooreman, Peter C.B. Phillips and Arthur van Soest for helpful comments. I am particularly grateful to Theo Nijman and Bertrand Melenberg for several stimulating discussions and helpful comments. Any remaining errors are my own.

$r_{it}$  is defined as  $r_{it} = 1$  if  $r_{it}^* \geq 0$  and  $r_{it} = 0$  if  $r_{it}^* < 0$ . Letting  $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})'$ ,  $\eta_i = (\eta_{i1}, \dots, \eta_{iT})'$  and  $\iota_T = (1, \dots, 1)'$  of dimension  $T$ , we assume that the error terms in (1) and (2) are normally distributed according to

$$\begin{pmatrix} \epsilon_i \\ \eta_i + \iota_T \xi_i \end{pmatrix} \sim N \left( 0, \begin{pmatrix} \sigma_\epsilon^2 I_T & \\ \rho \sigma_\epsilon \sigma_\eta I_T & \sigma_\eta^2 I_T + \sigma_\xi^2 \iota_T \iota_T' \end{pmatrix} \right), \quad (3)$$

so there is sample selectivity if  $\rho \neq 0$ . For identification purposes  $\sigma_\eta^2 + \sigma_\xi^2$  is usually set to one.

In a recent application by Keane, Moffitt and Runkle (1988) this model is used to describe the development of real wages over the business cycle; the endogenous variable  $y_{it}$  is the logarithm of the wage rate and  $r_{it}$  is a dummy variable if individual  $i$  is employed at time  $t$ . In other applications [Hausman and Wise (1979), Ridder (1990)] eq. (2) is used to explain the probability of attrition of an individual from the panel. In all cases there is sample selectivity if the error terms in the two equations are correlated.

## 2. Estimation of the model

The standard fixed effects (or 'within') estimator of  $\beta$  in (1) which ignores the selectivity problem is inconsistent if both  $\rho \neq 0$  and  $Z_{it}\gamma$  varies with  $t$  [cf. Verbeek and Nijman (1990)]. An obvious alternative is to use the maximum likelihood estimator incorporating selectivity, as done in Keane et al. (1988). This is a straightforward extension of the method of Hausman and Wise (1979), but instead of treating the  $\alpha_i$  as random errors we treat them as fixed unknown parameters. However, the fixed effects  $\alpha_i$  cannot be estimated consistently when the number of periods that individual  $i$  is observed ( $T_i$ , say) is small and this inconsistency is transmitted to the other coefficient estimators in models with limited dependent variables [see, e.g., Chamberlain (1980)]. In our model this inconsistency occurs so long as  $\rho \neq 0$ . Although Heckman (1981b) has provided some Monte Carlo evidence that the bias is fairly small in a fixed effects probit model with  $T = 8$ , it is not clear to what extent his results hold for the present model. In addition, one has to optimize the likelihood function with respect to a large number of parameters, which is computationally unattractive.

The standard solution to this incidental parameters problem is to condition the likelihood upon some (minimal) sufficient statistics for the incidental parameters resulting in a conditional likelihood function which is independent of the incidental parameters [Andersen (1970), Chamberlain (1980)]. However, in general there is no guarantee that these sufficient statistics exist. In the fixed effects model (1) with no selectivity bias ( $\rho = 0$ ) minimal sufficient statistics for  $\alpha_i$  are  $\bar{y}_i$ , the observed individual averages of  $y_{it}$  [see e.g. Chamberlain (1980)]. Unfortunately, in the model with selectivity  $\bar{y}_i$  is no longer a sufficient statistic for  $\alpha_i$  and conditioning the likelihood function does not eliminate the fixed effects, unless, again,  $\rho = 0$ .

Therefore, one has to look for an alternative way to overcome the incidental parameters problem, which is provided by transforming the data in such a way that the individual effects are eliminated and maximizing the likelihood of the transformed data. This can be seen as an application of marginal maximum likelihood [Kalbfleisch and Sprott (1970), Gourieroux and Monfort (1989, p. 208)] since (in general) only the likelihood of part of the original data is used. Well known effective transformations for the standard fixed effects model [eq. (1) with  $\rho = 0$ ] are taking deviations from individual means (the 'within' transformation) and taking first differences. It appears that the 'within' transformation, i.e. taking deviations from observed individual means, works equally well in

the model with selectivity bias, since it eliminates the incidental parameters ( $\alpha_i$ ) and thus yields a consistent estimator which is asymptotically normal.

The marginal likelihood contribution of individual  $i$  is given by

$$\ell_i^m = P\{r_{it}^* \geq 0, t \in \mathcal{T}_i, r_{it}^* < 0, t \notin \mathcal{T}_i \mid \tilde{y}_i\} f(\tilde{y}_i), \tag{4}$$

where

$$\mathcal{T}_i = \{t \in \{1, \dots, T\}; r_{it} = 1\}, \quad i = 1, \dots, N,$$

is the set of time indices for the periods in which individual  $i$  is observed and denotes deviations from the observed individual means, i.e.

$$\tilde{y}_{it} = y_{it} - \frac{1}{T_i} \sum_{t \in \mathcal{T}_i} y_{it} = y_{it} - \bar{y}_i, \quad t \in \mathcal{T}_i, \tag{5}$$

where  $T_i$  denotes the number of elements in  $\mathcal{T}_i$ ;  $\tilde{y}_i$  denotes the  $T_i$ -vector of observed  $\tilde{y}_{it}$ 's. Since (4) does not involve  $\alpha_i$  the incidental parameters problem is solved and maximizing the marginal likelihood function (the product of all  $\ell_i^m$ ) will lead to consistent and asymptotically normally distributed estimators for  $\beta$ ,  $\gamma$ ,  $\rho$ ,  $\sigma_\eta^2$  and  $\sigma_\epsilon^2$ .

At first sight the computation of the marginal maximum likelihood estimator seems to require numerical integration over  $T$  dimensions because of the  $T$ -variate conditional probabilities in (4), which is only feasible for small enough  $T$ . Fortunately however, it can be shown that the dimension of numerical integration can be reduced to two due to the special structure of the conditional distribution of the error term in the probit equation, which is identical to the distribution of

$$v_{it} + u_{i1} + r_{it}u_{i2}, \tag{6}$$

where  $v_{it}$ ,  $u_{i1}$  and  $u_{i2}$  are uncorrelated error terms with

$$E\{u_{i1}\} = E\{u_{i2}\} = 0, \quad E\{v_{it}\} = (\rho\sigma_\eta/\sigma_\epsilon)r_{it}(\tilde{y}_{it} - \tilde{X}_{it}\beta) = \Delta_{it},$$

and

$$V\{v_{it}\} = \sigma_\eta^2(1 - r_{it}\rho^2), \quad V\{u_{i1}\} = \sigma_\xi^2 \text{ and } V\{u_{i2}\} = \rho^2\sigma_\eta^2/T_i. \tag{7}$$

So just like in the random effects case analyzed by Ridder (1990) the conditional probit error term has an error components structure, which can be used to reduce the dimension of integration. In particular

$$P\{r_{it}^* \geq 0, t \in \mathcal{T}_i, r_{it}^* < 0, t \notin \mathcal{T}_i \mid \tilde{y}_i\} = \int \int \prod_{t=1}^T \Phi\left((2r_{it} - 1) \frac{Z_{it}\gamma + \Delta_{it} + u_{i1} + r_{it}u_{i2}}{\sqrt{(\sigma_\eta^2(1 - r_{it}\rho^2))}}\right) f(u_{i1}, u_{i2}) du_{i1} du_{i2}, \tag{8}$$

where  $f(\cdot)$  is the (normal) density function of  $u_1$  and  $u_2$  and  $\Phi$  is the standard normal distribution function. If individual  $i$  is observed in all periods ( $r_i = \iota_T$ ) or if there is no individual effect in the probit error term ( $\sigma_\xi^2 = 0$ ), this simplifies further to a single integral.

### 3. Concluding remarks

In this paper we have proposed a marginal maximum likelihood estimator that can be used to obtain consistent estimators of the parameters in a fixed effects model if there is sample selectivity. This estimator is consistent if the number of individuals ( $N$ ) tends to infinity, even if the number of time periods ( $T$ ) is finite, and does not require numerical optimization with respect to the fixed effects. Just like in standard cases, Wald, Lagrange Multiplier and likelihood ratio tests can be performed using the marginal likelihood function.

The marginal maximum likelihood estimator proposed above can be generalized in a number of ways. First, the normality assumption of the individual effect in the probit equation can be replaced by any other assumption concerning the distribution of  $\xi_i$ , including semi-parametric ones [cf. Keane et al. (1988)]. More general autocorrelation patterns of the probit error term can also be allowed, although computational tractability will usually require that  $T$  is small (because of the  $T$ -variate numerical integrals). Additionally, the strict exogeneity of the  $X_{it}$  variables required for the within transformation can be relaxed to predeterminedness if an alternative transformation is performed, for example the one proposed by Arellano (1988). Finally, if the initial conditions problem is properly taken into account,  $Z_{it}$  in eq. (2) may also contain the lagged dummy variable  $r_{i,t-1}$  without affecting consistency, so that state dependence and unobserved heterogeneity [cf. Heckman (1981a)] can be distinguished.

### References

- Andersen, E.B., 1970, Asymptotic properties of conditional maximum likelihood estimation, *Journal of the Royal Statistical Society, Series B*, 32, 283–301.
- Arellano, M., 1988, An alternative transformation for fixed effects models with predetermined variables, *Applied Economics Discussion paper 57*, (University of Oxford, Oxford).
- Chamberlain, G., 1980, Analysis of covariance with qualitative data, *Review of Economic Studies* 47, 225–238.
- Gourieroux, C. and A. Monfort, 1989, *Statistique et Modèles Économétriques*, Vol. 1 (Economica, Paris).
- Hausman, J.A. and D. Wise, 1979, Attrition bias in experimental and panel data: The Gary income maintenance experiment, *Econometrica* 47, 455–473.
- 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) 114–178.
- 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) 179–195.
- Kalbfleisch, J.D. and D.A. Sprott, 1970, Application of likelihood methods to models involving large numbers of parameters, *Journal of the Royal Statistical Society, Series B*, 32, 175–208.
- Keane, M., R. Moffitt and D. Runkle, 1988, Real wages over the business cycle: Estimating the impact of heterogeneity with micro data, *Journal of Political Economy* 96, 1232–12666.
- Mundlak, Y., 1961, Empirical production free of management bias, *Journal of Farm Economics* 43, 44–56.
- Ridder, G., 1990, Attrition in multi-wave panel data, in: J. Hartog, G. Ridder and J. Theeuwes, eds., *Panel data and labor market studies* (North-Holland, Amsterdam).
- Verbeek, M. and Th.E. Nijman, 1990, Testing for selectivity bias in panel data models, Mimeo. (Tilburg University, Tilburg).