State dependence in work-related training participation among British employees: A comparison of different random effects probit estimators.

Panos, Sousounis

Keele University

July 2008

State dependence in work-related training participation among British employees: A comparison of different random effects probit estimators.

Panos Sousounis†
Keele University
Centre for Economic Research

[This version: July 2008]

Abstract

This paper compares three different estimation approaches for the random effects dynamic panel data model, under the probit assumption on the distribution of the errors. These three approaches are attributed to Heckman (1981), Wooldridge (2005) and Orme (2001). The results are then compared with those obtained from generalised method of moments (GMM) estimators of a dynamic linear probability model, namely the Arellano and Bond (1991) and Blundell and Bond (1998) estimators. A model of work-related training participation for British employees is estimated using individual level data covering the period 1991-1997 from the British Household Panel Survey. This evaluation adds to the existing body of empirical evidence on the performance of these estimators using real data, which supplements the conclusions from simulation studies. The results suggest that for the dynamic random effects probit model the performance of no one estimator is superior to the others. GMM estimation of a dynamic LPM of training participation suggests that the random effects estimators are not sensitive to the distributional assumptions of the unobserved effect.

Keywords: state dependence; training; dynamic panel data models
JEL codes: C23, C25

---

† I am grateful to Gauthier Lanot for helpful comments. Address for correspondence: School of Economic and Management Studies, Keele University, Keele, ST5 5BG, UK; email: p.sousounis@econ.keele.ac.uk

1 Introduction

This paper evaluates different approaches to estimating dynamic binary-choice panel data models and distinguishing between true state dependence from spurious serial correlation. With experimental data, random assignment ensures that past experience is uncorrelated with the unobserved component and thus any observed persistence can be attributed to true state dependence. However, experimental data is rarely available in economic analyses and thus researchers have to rely on econometric techniques for non-experimental data to uncover the true effects of past experience on current period behaviour.

In the typical AR(1) model that allows for unobserved heterogeneity, if sampling does not coincide with the start of the process, and assuming that unobserved heterogeneity and state dependence are present in the pre-sample history, misspecifying the initial conditions will result in biased estimates of the parameters of interest i.e. the magnitude of state dependence as captured by the coefficient of the lagged term will not provide a true representation of its effect. Even though there is a well studied number of consistent and efficient estimators for the linear dynamic panel data model (see, among others, Hsiao, 1986, Baltagi, 1995 and Arellano and Honorè, 2000, for a review), the nonlinear case is considerably less developed.

There is relatively limited empirical evidence on the performance of different estimators for the nonlinear dynamic panel data model with most coming from simulation studies (Honorè and Kyriazidou, 2000). Evaluations with real world data are provided by Chay and Hyslop (2000) in their study of welfare participation and female labour force participation in the US and Stewart (2005) who studies the dynamics of unemployment and low-wage employment amongst the British workforce.

The first part of this paper compares three different approaches in estimating a random effects dynamic panel data model, under the probit assumption on the distribution of the errors, attributed to Heckman (1981), Wooldridge (2005) and Orme (2001). These estimators are compared with the rather naïve approach of treating the initial conditions as exogenous. In the second part, generalised method of moments (GMM) estimators are applied and compared in the context of a dynamic linear probability model. GMM estimators do not require distributional assumption for the unobserved effect and have thus become popular particularly in cases where the panel consists of a few time periods.

A model of work-related training participation for British employees is estimated using individual level data covering the period 1991-1997 from the British Household Panel Survey. This evaluation adds to the existing body of empirical evidence on the performance of these estimators using real data, which supplements the conclusions from simulation studies which may be sensitive to design.

The remaining of the paper is organised as follows. Section two presents the three different approaches to tackling the initial conditions problem proposed in the literature. Sections three and four discuss the data used and present the econometric estimates respectively. Section five discusses GMM estimation in the context of a dynamic limited probability model. Sections six and seven present the empirical model and different GMM estimators proposed in the literature respectively. Finally, section eight present estimates from the applied GMM estimators while section nine summarises and concludes.

2 Dynamic Random Effects Probit Models

This section considers three estimators for the dynamic random effects probit model which allows for state dependence and unobserved heterogeneity. The treatment of initial conditions is crucial in such models since misspecification will result in an inflated parameter of the lagged dependent variable term which measures the magnitude of past experience. Furthermore, possible unobserved heterogeneity could also overstate the effect of state dependence in work-related training if unaccounted for.

In modelling state dependence in the incidence of work-related training among British employees, the analysis begins with the specification of a general dynamic unobserved effects model of the form

$$P(y_{it}|y_{it-1},\dots,y_{i0},\boldsymbol{x}_i,\alpha_i) = \Phi(\boldsymbol{x}_{it}\boldsymbol{\beta} + \rho y_{it-1} + \alpha_i), \tag{1}$$

where $\boldsymbol{x}_{it}$ is a vector of strictly exogenous explanatory variables conditional on the unobserved effect. Under this formulation, the response probability of a positive outcome depends on the unobserved effect and past (one period) experience. It is further assumed that the dynamics are correctly specified i.e. one period lag is sufficient to allow the conditioning set to include all relevant past information. Testing the hypothesis of a non zero ρ is equivalent to testing the presence of true state dependence, having controlled for the unobserved heterogeneity. The above model can also be expressed as

$$f(y_1, y_2, \dots, y_T|y_0, \boldsymbol{x}, \alpha; \boldsymbol{\delta}) = \prod_{t=1}^{T} f(y_t|y_{t-1}, \dots, y_1, y_0, \boldsymbol{x}_t, \alpha; \boldsymbol{\delta})$$

$$= \prod_{t=1}^{T} \Phi(\boldsymbol{x}_t\boldsymbol{\beta} + \rho y_{t-1} + \alpha)^{y_t} [1 - \Phi(\boldsymbol{x}_t\boldsymbol{\beta} + \rho y_{t-1} + \alpha)^{1-y_t}. \tag{2}$$

Since T is fixed, a consistent log-likelihood function cannot be derived from this density (Wooldridge, 2002). The unobserved effect needs to be integrated out before estimation can progress. The need to integrate out the unobserved effect evokes the question of how the initial observation is to be treated.

A dynamic reduced form model of work-related training participation is specified as

$$y_{it} = 1[x_{it}\beta' + \gamma y_{it-1} + v_{it} > 0] \;\; i = 1, \dots, N; t = 2, \dots, T, \tag{3}$$

where y_{it}, a binary outcome variable, denotes participation in some form of work-related training ($y_{it} = 1$) in the current period, x_{it} is a vector of explanatory variables and $v_{it} \sim N(0, \sigma_u^2)$. The subscript i indexes individuals and t time periods. N is large and T is small and fixed, so asymptotics are on N alone.

The errors, v_{it}, are assumed independently and identically distributed. However, if we assume that the unobservable individual-specific heterogeneity is time invariant, the error term can be decomposed to, $v_{it} = \alpha_i + u_{it}$, which will be correlated over time due to the unobserved individual-specific time-invariant effect.

Estimating a standard uncorrelated random effects model implicitly assumes zero correlation between the unobserved effect and the set of explanatory variables. However, this assumption is most likely not to hold in this context. Consider that the unobserved effect captures an individual's motivation. In this case it is reasonable to expect it to be correlated with at least some of the elements of the set of explanatory variables, for example the educational qualifications variable. The assumption of no correlation between α_i and x_{it} is relaxed following Mundlak (1978) and Chamberlain's (1984) suggestion that the regression function of α_i is linear in the means of all the time varying covariates and it can therefore be written as,

$$\alpha_i = \delta_0 + \bar{x}_{it}\delta' + e_i,$$

where $e_i \sim N(0, \sigma_e^2)$ and independent of x_{it} and u_{it} for all i and t. Hence, model (3) can be rewritten as (the constant has been absorbed into β)

$$y_{it} = 1[x_{it}\beta' + \gamma y_{it-1} + \bar{x}_{it}\delta' + e_i + u_{it} > 0] \;\; i = 1, \dots, N; t = 2, \dots, T. \quad (4)$$

This formulation implies that, $\lambda = corr(v_{it}, v_{it-1}) = \sigma_e^2/(\sigma_e^2 + \sigma_u^2)$ for $t = 2, \dots, T$.

A further assumption about the relationship of the initial observations, y_{i0}, and the unobserved effect is needed for consistent estimation of (4). If the initial conditions are assumed exogenous, the likelihood decomposes and any standard random effects probit program can be used. However, if the initial conditions are correlated with the unobserved effect, as would be expected in the current context, this method will overestimate the effect of state dependence.

2.1 Heckman's estimator

The approach proposed by Heckman (1981) involves the specification of a reduced form equation for the initial conditions of the form

$$y_{i0} = 1[z_{i0}'\zeta + \xi_i > 0].$$

The vector z_{i0} includes all variables relevant to period zero in addition to some exogenous pre-sample variables and the vector of means, $\bar{x}_i$, $var(\xi_i) = \sigma_\xi^2$ and $corr(e_i, \xi_i) = \rho$. To account for a possible non-zero ρ, the error ξ_i is decomposed to $\xi_i = \theta e_i + u_{i0}$ with $e_i \perp u_{i0}$ and it is further assumed that u_{i0} satisfies the same distributional assumptions as u_{it} for $t \geq 2$. The period zero linear reduced form equation is then,

$$y_{i0} = 1[z_{i0}'\zeta + \theta e_i + u_{i0} > 0].$$

Under the normalisation $\sigma_u^2 = 1$, the joint probability for individual i given the unobserved time-invariant effect e_i, is

$$\Phi[(z_{i0}'\zeta + \theta e_i)(2y_{i0} - 1)] \prod_{t=2}^{T} \Phi[(\gamma y_{it-1} + x_{it}'\beta + e_i)(2y_{it} - 1)].$$

For a random sample, the likelihood function is thus

$$\prod_i \int_{e^*} \left\{ \Phi[(z_{i0}'\zeta + \theta\sigma_\alpha e^*)(2y_{i0} - 1)] \prod_{t=2}^{T} \Phi[(\gamma y_{it-1} + x_{it}'\beta + \sigma_e e^*)(2y_{it} - 1)] \right\} dF(e^*),$$

where F is the distribution function of $e^* = e/\sigma_e$ and $\sigma_e = \sqrt{\lambda/(1-\lambda)}$ due to the normalisation used. If it is further assumed that the unobserved effect is normally distributed, the integral over e^* can be evaluated using Gauss-Hermite quadrature (Stewart, 2006).

2.2 Wooldridge's Conditional ML estimator

Wooldridge (2005) proposes a conditional maximum likelihood estimator that considers the distribution conditional on the initial period observations and exogenous covariates. In effect, Wooldridge suggests modelling the density $(y_{i1}, \dots, y_{iT})$ conditional on (y_{i0}, x_i) instead of the density $(y_{i0}, \dots, y_{iT})$ conditional on x_i. The advantage of the resulting estimator is that it may be implemented using standard econometric software and is computationally inexpensive in contrast to the Heckman estimator which requires special software to be written.

Instead of specifying a model for the initial conditions given observed covariates and the unobserved effect, a model is specified for the unobserved effect given observed covariates and the initial conditions. First assume that

$$e_i = e_0 + e_1 y_{i0} + \epsilon_i.$$

Substituting into (4) gives

$$y_{it} = 1[x_{it}\beta' + \gamma y_{it-1} + e_0 + e_1 y_{i0} + \bar{x}_{it}\delta + \epsilon_i + u_{it} > 0],$$

where the Mundlak specification has been incorporated. Wooldridge further suggests that one may allow for a more flexible conditional mean in the analysis by including various interactions of the initial period observations with the means of the time-varying covariates.

2.3 Orme's estimator

Orme (2001) proposes a two step estimator for the dynamic random effects model. It is an approximation for small values of ρ and follows from Heckman's standard sample selection correction method. He proposes to incorporate a correction term in the conditional model to account for the correlation between the unobserved heterogeneity and the initial observations.

In Orme's (2001) estimator, a reduced form equation for the initial observation as in Heckman's procedure needs also be specified. The specification for a non-zero ρ (in terms of orthogonal error components) is different to the Heckman case though and has the form

$$e_i = \psi\xi_i + \omega_i,$$

in which, by construction, $\xi_i \perp \omega_i$, $\psi = \rho\sigma_e/\sigma_\xi$ and $var(\omega_i) = \sigma_e^2(1-\rho^2)$. If we then substitute to model (4), we get

$$y_{it} = 1[x'_{it}\beta + \gamma y_{it-1} + \psi\xi_i + \omega_i + u_{it} > 0].$$

Orme argues that if (ξ_i, ω_i) follow a bivariate normal distribution, $E(\omega_i|y_{i0}) = 0$ but $E(\xi_i|y_{i0}) = h_i$ where by construction,

$$h_i = \frac{(2y_{i0}-1)\phi(\boldsymbol{x_i\beta})}{\Phi[(2y_{i0}-1)\boldsymbol{x_i\beta}]}.$$

Assuming that u_{it} is orthogonal to the regressors, if e_i is replaced by the conditional expectation h_i, ω_i will be the random component in a standard random effects probit model of the form

$$y_{it} = 1(x'_{it}\beta + \gamma y_{it-1} + \psi h_i + \omega_i + u_{it} > 0),$$

estimable by standard econometric software. Orme (2001) argues that the estimator despite being local to zero performs well for 'large' values of ρ as well.

3 The data

Data from the first seven waves of the British Household Panel Survey (BHPS hereafter), a longitudinal survey of randomly selected households in Great Britain, is used. The interviews for the first wave of the BHPS were conducted between September and December 1991 and annually thereafter[1]. The sample comprises an unbalanced panel and includes men and women of working

[1] For more details see Taylor, M.F et al (2006).

age who are present and in employment as employees in the first wave (1991) but who may subsequently drop out of the sample as a result of missing information, attrition or having moved out of scope[2]. Thus the sample excludes self-employed individuals, the unemployed, those in full time education and members of the armed forces[3].

Over the sample period, the BHPS contains two variables that relate to an individual's participation in training during the twelve months prior to the interview date. The first of these variables records the incidence of formal on-the-job training undertaken as part of the individual's present employment[4] whilst the second question records any other education or training that was undertaken that enhances skills for current or future employment. The training referred to in this latter respect is, at least potentially, work-related, excluding any education or training undertaken as a pastime, hobby or solely for general interest. In the current analysis we combine responses from both questions in our definition of work-related training[5].

In modelling work-related training participation I include a set of variables that reflect individual characteristics such as age, indicators of prior educational attainment, race, and occupation, employment and employer characteristics such as job permanency, part-time, full-time status, hierarchical position within the firm, trade union presence and firm size together with an indicator of training history. In addition, a set of variables recording past information, including the socio-economic and personal characteristics of the respondent's father and pre-sample information on the respondent, is utilised in the estimation of the reduced-form equation for the initial conditions for the Orme and Heckman estimators.

4 Random effects probit estimates

Estimates of the random effects probit model for the probability of work-related training participation using the Heckman, Wooldridge and Orme estimators are given in Table 1 columns 2, 3 and 4 respectively. The vector of regressors includes the listed variables plus regional and year dummies. The models also contain means over time for

[2] Individuals were not allowed to enter or re-enter the sample after the first wave.
[3] The estimation was carried out for a balanced sample as well (individuals for whom complete BHPS histories are available) with the main findings reported here remaining unaltered.

[4] In Wave One, only the employed were asked this. At Wave Two, this was extended to all currently working. The scope of the question was widened to include education or training courses.
[5] The relevant questions in the 1991 BHPS are D23 and E17.

each time-varying variable (following Mundlak's suggestion as specified above). The corresponding pooled probit model (without random effects) estimated on the same sample is given in the first column for comparison.

The dynamic random effects probit model and the pooled probit model involve different normalisations (Arulampalam, 1998). The random effects probit estimates are normalised on $\sigma_u^2 = 1$, while the pooled probit estimates are normalised on $\sigma_v^2 = 1$. Thus random effects probit models provide an estimate of γ/σ_u while pooled probit models an estimate of γ/σ_v. For comparison, the former model needs to be multiplied by an estimate of $\sigma_u/\sigma_v = \sqrt{1-\lambda}$. Table 1a presents scaled coefficient estimates.

Exogeneity of the initial conditions in the random effects model can be tested by a simple significance test under the null of $\theta = 0$ for the Heckman estimator and by $y_0 = 0$ and $h_i = 0$ for the Wooldridge and Orme estimators respectively. It is clear the exogeneity hypothesis is strongly rejected in these models.

The estimates from all three estimators are very similar. The coefficient of the lagged dependent variable is positive and highly significant indicating strong persistence effects in the incidence of work-related training. Assuming the initial conditions as exogenous overstates the effect of state dependence as is obvious from the rather inflated coefficient of the pooled probit model (without random effects). In contrast, the magnitude of the coefficient is almost halved for the rest of the estimators employed.

For the Heckman estimator, parental variables and pre-sample variables related to educational characteristics are used as instruments. Specifically, dummy variables for father's broad socio-economic class, when the respondent was aged 14 together with indicator variables for the father being in a managerial profession or not as well as for father being in gainful employment as opposed to being deceased are used in addition to a variable recording whether the respondent attended a public school or otherwise. This set of variables is also included in the first period probit model of the Orme estimator.

[table 1 here]

Table 1 Dynamic Random Effects Probit Models for Training Probability†.

Variable Name	**Pooled Probit** [1]	**Heckman Estimator** [2]	**Wooldridge Estimator** [3]	**Orme Estimator** [4]
Lagged dependent variable				
Trained *t-1*	0.848 (35.7)	0.415 (12.8)	0.447 (15.5)	0.431 (14.1)
Personal characteristics				
Sex (Female)	0.063 (2.24)	0.068 (1.71)	0.075 (2.22)	0.041 (1.14)
Age	0.000 (0.09)	-0.009 (0.77)	-0.001 (0.16)	-0.001 (0.13)
Age ^{2}	0.000 (0.83)	0.000 (0.00)	0.000 (0.62)	0.000 (0.77)
Race (white)	0.112 (1.57)	0.165 (1.53)	0.123 (1.40)	0.173 (1.81)
Marital Status (single)	-0.051 (1.75)	-0.065 (1.65)	-0.047 (1.42)	-0.060 (1.67)
Social class				
Professional occupation	0.222 (2.37)	0.294 (2.50)	0.191 (1.87)	0.241 (2.19)
Managerial & Technical occupation	0.311 (3.78)	0.377 (3.60)	0.279 (3.11)	0.304 (3.12)
Skilled non-manual occupation	0.225 (2.83)	0.344 (3.39)	0.211 (2.42)	0.231 (2.45)
Skilled manual occupation	0.146 (1.79)	0.244 (2.36)	0.159 (1.79)	0.143 (1.50)
Partly skilled occupation	0.012 (0.15)	0.111 (1.08)	0.012 (0.14)	-0.004 (0.05)
Highest Educational Qualification				
Higher degree	0.293 (3.48)	0.381 (3.15)	0.324 (3.18)	0.354 (3.23)
First degree	0.499 (9.13)	0.635 (8.13)	0.547 (8.22)	0.623 (8.64)
Teaching qf.	0.700 (8.67)	0.876 (7.69)	0.729 (7.55)	0.850 (8.18)
Other higher qf.	0.483 (10.9)	0.602 (9.52)	0.523 (9.82)	0.603 (10.3)
Nursing qf	0.233 (2.56)	0.494 (3.91)	0.287 (2.67)	0.473 (3.96)
GCE A levels	0.253 (5.21)	0.294 (4.22)	0.271 (4.63)	0.346 (5.41)
GCE O levels or equivalent	0.154 (3.66)	0.178 (2.97)	0.168 (3.32)	0.210 (3.78)
Commercial qf / No O levels	0.085 (1.13)	0.058 (0.57)	0.105 (1.15)	0.165 (1.68)
CSE Grade 2-5 / Scottish Grd	0.015 (0.22)	0.014 (0.15)	0.026 (0.33)	0.034 (0.39)
Apprentiship	0.174 (1.66)	0.181 (1.17)	0.149 (1.18)	0.181 (1.33)
Other qualifications	-0.378 (2.14)	-0.474 (1.81)	-0.401 (1.98)	-0.378 (1.76)
Characteristics of current job/employer				
Private Sector	-0.166 (3.77)	-0.159 (2.84)	-0.183 (3.75)	-0.175 (3.33)
Permanent position	0.167 (2.75)	0.118 (1.53)	0.188 (3.00)	0.201 (2.94)
Working Part Time	-0.198 (5.73)	-0.249 (5.50)	-0.214 (5.54)	-0.216 (5.13)
Trade union coverage in the workplace	0.152 (5.41)	0.178 (4.89)	0.141 (4.44)	0.172 (5.05)
Managerial position	0.125 (3.42)	0.109 (2.44)	0.123 (3.10)	0.110 (2.61)
Supervisor/foreman	0.115 93.58)	0.103 (2.72)	0.104 (3.08)	0.105 (2.90)
Size of employing organization (manpower)				
More than 25 / 50 to 99 (small)	0.035 (0.99)	0.056 (1.30)	0.040 (1.05)	0.041 (1.01)
100 to 499 (medium)	0.071 (2.39)	0.089 (2.38)	0.075 (2.28)	0.088 (2.50)
500 or more (large)	0.152 (4.45)	0.139 (3.24)	0.148 (3.90)	0.136 (3.35)
Auxiliary parameters				

Intercept	-1.800 (5.19)	-1.032 (0.01)	-1.485 (4.18)	-1.321 (3.41)
θ		0.961 (8.19)		
y_0			0.340 (10.7)	
h_i				0.248 (10.4)
σ_u			0.509 (20.3)	0.517 (19.4)
λ		0.243 (12.5)	0.206 (12.8)	0.210 (12.3)
Log likelihood	-8140.29	-7315.48	-8001.49	-6946.68
NT	14647	18270	14647	12645
N	2441	3045	2768	2373
Pred. Prob. p_0'	0.415	0.421	0.417	0.423
Pred. Prob. p_1'	0.682	0.563	0.565	0.566
APE: $p_1' - p_0'$	0.267	0.142	0.148	0.143
PPR: p_1'/p_o'	1.64	1.33	1.35	1.33

Notes:
1. All models contain dummy variables for region and industrial classification.
2. All models contain yearly dummies and means of time-varying covariates.
3. t-ratios in parentheses.
4. p_0', p_1' = predicted proabbilitties of training participation at t given non-participation and participation at time $t-1$ respectively.
5. APE = Average Partial Effect, PPR = Predicted Probability Ratio
† Estimation was carried out in stata© 9.2.

Table 1 provides estimates of the predicted probabilities together with the average partial effects (APE), $p_1 - p_0$, and the predicted probability ratios (PPR), p_1/p_0. The partial effect of y_{it-1} on the $P(y_{it} = 1)$ is calculated based on the calculation of a counter-factual outcome probability assuming y_{t-1} fixed at the two alternate states evaluated at $x_{it} = \bar{x}$ following Stewart (2006),

$$p_1 = \frac{1}{N}\sum_{i=1}^{N} \Phi\left[(\bar{x}\beta + \gamma)\sqrt{(1-\lambda)}\right]$$

and

$$p_0 = \frac{1}{N}\sum_{i=1}^{N} \Phi\left[(\bar{x}\beta)\sqrt{(1-\lambda)}\right].$$

All three estimators provide strong support to the proposition of serial dependence in the incident of work-related training. The effects of the rest of the covariates are in line with the findings of other studies in the literature. When the initial conditions are assumed exogenous, the random effects variance is restricted to zero, implying that there is no unobserved heterogeneity in participation probabilities and all observed serial persistence is due to γ and β. The estimate of state dependence in this case is substantial ($\gamma = 0.84$). However, this estimate will overstate state dependence if the unobserved individual specific effect influences the sample initial conditions. Columns 2, 3 and 4 of Table 1 present estimates of models allowing for a endogenous initial conditions by approximating y_{i0} with a flexible reduced form equation. The results change substantially and the state dependence estimate is less than halved ($\gamma \approx 0.43$). The estimate of $Var(\alpha_i)$ implies that approximately 20% of the total error variance is attributable to unobserved heterogeneity.

The choice between estimators is usually based on relative performance and ease of implementation. The three estimators for the dynamic random effects probit model considered here produce very similar results suggesting that none dominates the others. The estimators proposed by Wooldridge and Orme are easier to implement with standard software compared to Heckman's approach. However, Stewart (2005) suggests a routine for implementing the Heckman estimator with standard software alleviating any problems of special software development. This places all three estimators on an equal footing and reduces the choice to a matter of preference between the different approximations for the treatment of the initial observations.

5 GMM estimation of a DLP model of training.

The dynamic random effects probit models considered so far necessitate an auxiliary distributional assumption on the individual-specific unobserved effect. In this section, Generalised Method of Moments (GMM) estimators (Hansen, 1982), in the context of a dynamic linear probability (DLP) model that does not require such assumptions, are considered. Such estimators have been labelled semi-parametric since they are non-parametric for the unobserved individual-specific effects.

In non-linear panel data models where $N \to \infty$ but T is small or fixed, maximum likelihood estimation relies on some very restrictive assumptions about the distribution of the error term and can be computationally burdensome - controlling for potential serial correlation in the error terms, involves n-dimensional numerical integration, which for panels with $T \geq 3$ may not be feasible, subject to the choice of probability density function (Breitung and Lechner, 1995).

GMM techniques are widely applied to the estimation of dynamic linear panel data models (Avery, Hansen and Hotz, 1983, Holtz-Eakin *et al*, 1989, Arellano and Bond, 1991). The GMM estimators have been proven popular since they do not require the analyst to explicitly specify the covariance structure of the errors and are considerably less demanding in terms of computational effort. Here I apply GMM estimators proposed in the literature, which utilise instruments originating within the dataset, to a dynamic LPM for participation in work-related training.

In the linear probability model (LPM), unless the range of the regressors set, $\mathbf{x}$, is severely restricted, the estimates will not provide an adequate description of the underlying population response probability since for some values of the explanatory variables the fitted values will fall outside the unit interval. More, the linear probability model implies that the response probability will always change by the same amount following a unit increase in any one covariate, x_i, *ceteris paribus,* regardless of the initial values of the covariate. This is clearly counterintuitive since it implies that the probability could eventually increase beyond one or decrease below zero following enough changes in x_i. The LPM is therefore best considered as an approximation of the population response probability.

Conveniently though, the linear probability model approximates the response probability for common values of the covariates. In addition, it appears to provide 'good' estimates of the partial effects on the outcome probability near the centre of the distribution of $\mathbf{x}$. The 'goodness' of these estimates is usually assessed via a comparison with the estimates from nonlinear estimation techniques such as probit and/or logistic regressions. If interest lies with the partial effect of a covariate on the response probability averaged across $\mathbf{x}$, then the predicted values that lie outside the unit interval may diminish in importance and thus the LPM may indeed provide 'good' estimates.

6 The empirical model

The general model of the data-generating process is specified as

$$y_{it} = x'_{it}\beta + \rho y_{it-1} + u_{it},\ i = 1, \dots, N \text{ and } t = 2, \dots T, \tag{5}$$

where $u_{it} = \alpha_i + \varepsilon_{it}$, with $E(\alpha_i) = 0$, $E(\varepsilon_{it}) = 0$, $E(\alpha_i \varepsilon_{it}) = 0$ and $E(\varepsilon_{it}\varepsilon_{is}) = 0\ \forall\ t \neq s$. Differencing removes the individual-specific unobserved effects and the model can be written as

$$y_{it} - y_{it-1} = \beta(x_{it} - x_{it-1}) + \rho(y_{it-1} - y_{it-2}) + u_{it} - u_{it-1} \Leftrightarrow$$

$$\Delta y_{it} = \Delta x'_{it}\beta + \rho\Delta y_{it-1} + \Delta\varepsilon_{it}, i = 1, \dots, N \text{ and } t = 2, \dots T. \tag{6}$$

The problem with applying OLS to (5) is that y_{it-1} is endogenous to the fixed effects in the error term, resulting in biased estimates. The first differenced transformation proposed in the literature (model 6) does not overcome this problem either. Even if ε_{it} are serially independent, Δy_{it-1} and $\Delta\varepsilon_{it}$ will be correlated since the y_{it-1} term in $\Delta y_{it-1} = y_{it-1} - y_{it-2}$ correlates with $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$. Similarly, any predetermined regressors (not strictly exogenous) could be rendered endogenous due to their potential relation to ε_{it-1} (Roodman, 2006).

The first-difference transformation has a further drawback in the sense that it exacerbates the problem of non-response in unbalanced panels. If y_{it} is missing, then $y_{it} - y_{it-1}$ and/or any $y_{is} - y_{it}$, $\forall\ s > t$, cannot be defined and will also be missing.

Arellano and Bover (1995) propose to overcome this potential source of difficulty by subtracting the average of all future observations from the current observation instead of subtracting the previous observation from the current one. That way a 'differenced' value can almost always be obtained and given that lagged observations are not used to transform the variables as in the first-differenced case, they can serve as instruments. For variable x the transformation formula is hence

$$x_{it}^* = c_{it}\left(x_{it} - \frac{1}{T_{it}}\sum_{s>t} x_{is}\right),$$

(7)

where T_{it} is all future available observations and $c_{it} = \sqrt{T_{it}/(T_{it}+1)}$ is a scale factor. This transformation, also referred to as "(forward) orthogonal deviations", allows the x_{it} to retain their properties i.e. if x_{it} are independently distributed, x_{it}^* will be too. The choice of scale factor also assures that if x_{it} is identically in addition to independently distributed, x_{it}^* will again retain the property. This is not achieved through differencing since the transformation induces correlation between successive error terms even if there is no correlation in the levels. This can be seen by the mathematical relationship of $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ to $\Delta\varepsilon_{it-1} = \varepsilon_{it-1} - \varepsilon_{it-2}$ through the common ε_{it-1} term (Roodman, 2006).

7 Applying GMM

7.1 Arellano-Bond (1991) estimator

Optimal GMM requires first, the estimation of the covariance matrix of the transformed errors, Ω^*, which hinges on the assumption that ε_{it} are i.i.d. Secondly, Ω^* is proxied by the robust estimates of

$$\widehat{\Omega}^* = \hat{E}_i\hat{E}_i' = \begin{bmatrix} \hat{e}_{i1}^2 & \hat{e}_{i1}\hat{e}_{i2} & \cdots & \hat{e}_{i1}\hat{e}_{iT} \\ \hat{e}_{i2}\hat{e}_{i1} & \hat{e}_{i2}^2 & \cdots & \hat{e}_{i2}\hat{e}_{iT} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{e}_{iT}\hat{e}_{i1} & \cdots & \cdots & \hat{e}_{iT}^2 \end{bmatrix}. \quad (8)$$

The block-diagonal matrix $\widehat{\Omega}$ allows for arbitrary patterns of covariance within individuals but not across them. The inclusion of time dummies in the estimation removes any time shocks from the errors and is therefore advised. As long as the $\hat{e}$ are derived from a consistent estimate of β, a GMM estimator derived from them will be asymptotically efficient (Arellano, 2003).

The resulting estimator is known as the classic Arellano-Bond (1991) difference GMM estimator for dynamic panels. This estimator is more efficient than the Anderson-Hsiao estimator, which instruments Δy_{it-1} with either y_{it-2} or Δy_{it-2}. As noted, before, the consistency of all these estimators hinges on ε_{it} being serially uncorrelated.

7.2 Blundell-Bond (1998) estimator

Arellano and Bond (1991) carry out simulation analysis for the performance of the one and two step difference GMM estimator and find that the latter outperforms the OLS, within groups and Anderson-Hsiao, both in differences and levels, estimators. Blundell and Bond (1998) also conduct Monte Carlo simulations to assess the performance of the different GMM estimators. They find that the performance of the difference GMM is limited when the autoregressive process is (or is close to be) a random walk. The authors argue this to be due to the potentially limited explanatory power of past level states on future changes.

Blundell and Bond (1998) propose a different approach, which involves the construction of an additional set of instruments. Instead of transforming the set of regressors to rid them of the fixed effects, they propose to apply the differencing transformation to the set of instruments thus making them exogenous to the fixed effects. This is valid under the assumption that changes in any instrumenting variable z are not correlated with the fixed effects i.e. $E(\Delta z_{it}\alpha_i) = 0$ for all i and t. Another way of putting this is to say that $E(z_{it}\alpha_i)$ is constant over time (Roodman, 2006).

If this holds true then Δx_{it-1} could act as a valid instrument for the variables in levels since

$$E(\Delta z_{it-1}u_{it}) = E(\Delta z_{it-1}\alpha_i) + E(z_{it-1}\varepsilon_{it}) - E(z_{it-2}\varepsilon_{it}) = 0.$$

The rationale behind the Blundell and Bond (1998) estimator is to instrument levels with differences whereas the Arellano and Bond (1991) estimator instruments differences with levels. If the process resembles a random walk, previous changes may be more informative of current (or future) states than previous states would be of current (or future) changes.

As noted earlier, the Blundell-Bond estimator requires one additional assumption, $E(\Delta z_{it}\alpha_i) = 0$ for all i and t, which is in effect a stationarity assumption. If Δy_{it-1} acts as an instrument for y_{it-1} and since both Δy_{it-1} and u_{it} contain α_i, the proposition that the instrument is orthogonal to the error, $E(\Delta y_{\mathrm{it}-1}u_{\mathrm{it}}) = 0$, is not straightforward. Blundell and Bond suggest that it is possible if the data generating process is of a form that allows the fixed effects and the autoregressive process (as determined by the coefficient on the lagged dependent variable) to offset each other in expectation (Roodman, 2006, p.29).

Blundell and Bond (1998) propose the following procedure. The first step involves the creation of a 'new' dataset. Apply the chosen (differences or orthogonal deviations) transformation to the data. Then, combine the transformed and levels observations into one dataset. Setting up the data in this way does not cause any confusion to the applied estimation techniques since both the transformed and levels data are characterised by the same linear functional form.

The second stage involves the construction of the instruments. This is done in a fashion similar to the data format. The appropriately strictly exogenous variable acting as an instrument is transformed (through differencing) and a column vector of instruments is created by again combining the transformed and untransformed observations and imposing the moment condition $\sum z_{it}^{*}\hat{e}_{it}^{*} + \sum z_{it}\hat{e}_{it} = 0$, where z is the instrument and $\hat{e}$ the empirical errors and z^{*} and $\hat{e}^{*}$ denote transformed quantities.

Next, the Arellano-Bond instruments i.e. instruments in levels, are set to zero for levels observations and the transformed instruments are set to zero for the transformed observations. This results in a GMM-style instrument matrix, which could potentially include a full set of differenced instruments for the levels equation using all available lags. However, most of these would not result in further efficiency gains since they would be mathematically redundant as Roodman (2006, p.31) shows.

The estimator proposed by Blundell and Bond (1998) is also referred to as system-GMM since it combines observations and instruments both transformed and in levels.

8 GMM estimates

As already mentioned the random effects probit estimators are potentially sensitive to the auxiliary distributional assumption of the individual-specific unobserved effect. Further investigation of this issue could be provided by GMM estimates of a DLP model as described in the previous sections. The random effects formulation provides efficiency gains if the auxiliary distributional assumption is not violated (Stewart, 2005). GMM estimators of the fixed effects model are efficiency-wise inferior, however, do not require an assumptions about the distribution of the unobserved heterogeneity. A comparison of the two sets of results provides a way of assessing the validity of the auxiliary distributional assumption, namely that of normality in this case.

Table 2 presents estimates of the DLP model using different estimators and OLS estimates for comparison. Columns 2 and 3 present the Arellano-Bond optimal-GMM and Blundell-Bond system-GMM estimators respectively using only lagged training participation variables as GMM-style instruments. For the Arellano-Bond estimator, the one-step estimates are presented following the recommendation of Doornik, Arellano and Bond (1999). The two-step estimates and their (corrected) standard errors are very similar to the one-step estimates.

Column 1 of Table 2 presents OLS estimates of the DLP model comparable to column 1 of Table 1; the results are similar once put in a comparable basis. The lagged training coefficient from the OLS regression (0.305) is not much different to the APE for the pooled probit estimator (0.267). The estimated coefficients for the dynamic training term are not large, easing any weak-instrument concerns. The models pass the Arellano-Bond second order autocorrelation test and the Hansen and Sargan tests of over-identifying restrictions.

[table 2 here]

Table 2 GMM estimates

Variable Name	OLS [1]		Arellano-Bond (One-step diff.) [2]		Blundell-Bond (system-GMM) [3]	
Lagged dependent variable						
Trained *t-1*	0.3059	(39.0)	0.1516	(10.1)	0.1323	(7.32)
Personal characteristics						
Sex (Female)	0.0209	(2.34)			0.0214	(1.81)
Age	-0.0011	(0.40)	1.0042	(22.7)	-0.0020	(0.57)
Age ^{2}	0.0000	(0.54)	0.0000	(0.36)	0.0000	(0.35)
Race (white)	0.0323	(1.45)			0.0417	(1.45)
Marital Status (single)	-0.0159	(1.73)	-0.0252	(0.84)	-0.0197	(1.71)
Social class						
Professional occupation	0.0663	(2.32)	-0.0771	(1.19)	0.0929	(2.73)
Managerial & Technical occupation	0.0945	(3.86)	-0.0036	(0.06)	0.1211	(4.24)
Skilled non-manual occupation	0.0641	(2.75)	0.0118	(0.21)	0.0838	(3.14)
Skilled manual occupation	0.0379	(1.58)	0.0195	(0.37)	0.0460	(1.68)
Partly skilled occupation	-0.0024	(0.10)	-0.0062	(0.12)	0.0013	(0.05)
Highest Educational Qual.						
Higher degree	0.0870	(3.17)	0.4330	(2.12)	0.1091	(2.77)
First degree	0.1583	(9.22)	0.4451	(2.80)	0.1975	(8.53)
Teaching qf.	0.2181	(8.83)	0.3626	(2.60)	0.2751	(9.30)
Other higher qf.	0.1519	(11.0)	0.4580	(4.40)	0.1903	(10.1)
Nursing qf	0.0705	(2.42)	0.3974	(2.83)	0.0933	(2.25)
GCE A levels	0.0701	(4.64)	0.3991	(3.68)	0.0898	(4.41)
GCE O levels or equivalent	0.0365	(2.85)	0.4004	(3.88)	0.0446	(2.79)
Commercial qf / No O levels	0.0153	(0.66)	0.2896	(1.27)	0.0134	(0.45)
CSE Grade 2-5	-0.0073	(0.36)	0.3938	(1.99)	-0.0112	(0.49)
Apprenticeship	0.0439	(1.36)	0.1531	(0.46)	0.0625	(1.86)
Other qualifications	-0.0937	(2.03)	0.4297	(3.98)	-0.1137	(3.83)
Characteristics of current job/employer						
Private Sector	-0.0545	(3.86)	-0.0542	(1.50)	-0.0691	(3.92)
Permanent position	0.0535	(2.81)	0.0589	(1.73)	0.0653	(2.92)
Working Part Time	-0.0635	(5.87)	-0.1104	(4.28)	-0.0756	(5.60)
Trade union coverage	0.0497	(5.53)	0.0220	(0.94)	0.0645	(5.63)
Managerial position	0.0409	(3.47)	0.0465	(1.91)	0.0510	(3.45)
Supervisor/foreman	0.0360	(3.48)	0.0199	(1.07)	0.0458	(3.67)
Size of employing organization (manpower)						
More than 25 / 50 to 99 (small)	0.0107	(0.94)	0.0696	(3.23)	0.0101	(0.76)
100 to 499 (medium)	0.0218	(2.29)	0.0575	(2.56)	0.0236	(1.98)
500 or more (large)	0.0496	(4.50)	0.0637	(2.50)	0.0585	(4.14)
Intercept	0.0737	(1.07)			0.1004	(1.15)
AR(1)	-5.59		-30.25		-33.37	
AR(2)	10.17		2.04		2.01	
Sargan (χ^2_d)			17.42		37.16	

Hansen (χ_d^2)		15.75	35.02
Degrees of freedom (d)		14	19
NT	14647	11563	14647
Pred. Prob. p_0'	0.6600	0.5001	0.6613
Pred. Prob. p_1'	0.7611	0.5046	0.7074
APE: $p_1' - p_0'$	0.1011	0.0045	0.0461
PPR: p_1'/p_o'	1.1531	1.0089	1.0697

8.1 Testing for autocorrelation

Testing the validity of the instruments (moment conditions) employed in the estimation of dynamic panel data models is done by means of the GMM test of overidentifying restrictions commonly associated with Sargan (1958) and Hansen (1982)[6]. The Sargan test of overidentifying restrictions essentially involves regressing the residuals obtained from the IV regression on all exogenous variables (instruments and controls) and recording the coefficient of determination. The test statistic is constructed as $S = nR^2$, where n is the number of observations. Under the null hypothesis that all instruments are exogenous, the test statistic S is χ^2 distributed with degrees of freedom equal to $m - k$, where $m - k$ is the number of instruments minus the number of endogenous variables respectively.

Arellano and Bond (1991) consider the case when the idiosyncratic error term ε_{it} suffers from autocorrelation. If ε_{it} are serially correlated of order one, then y_{it-2} is endogenous to ε_{it-1} in the differenced error term, $\Delta u_{it} = \varepsilon_{it} - \varepsilon_{i-1}$, and hence invalid as instrument. The composite error, u_{it}, will of course be correlated via the unobserved effect but the estimators are designed to account for such autocorrelation. If order one serial correlation is proven, then lags of order three and higher can only be used. If higher order serial correlation is present, then even higher order lags need be utilised. The problem this causes is obvious and it may even prove impossible to overcome in very short panels.

[6] For a description of the test procedure in the dynamic panel data model see Arellano and Bond (1991). Altonji and Segal (1996) find that the Sargan test has poor size properties for panels with large N and small/fixed T.

Testing for first-order serial correlation in the levels involves checking for second-order serial correlation in differences since this is expected to unveil correlation between ε_{it-1} in $\Delta\varepsilon_{it}$ and ε_{it-2} in $\Delta\varepsilon_{it-2}$. First order serial correlation is expected in differences since $\Delta\varepsilon_{it}$ shares term ε_{it-1} with $\Delta\varepsilon_{it-1}$ and thus finding evidence of that provides no new information. Consequently, the general approach for testing the presence of serial correlation of order h in levels, is to look for serial correlation of order $h+1$ in differences. If the data transformation was in orthogonal deviations, the test of autocorrelation in the idiosyncratic error ε_{it} should still be carried out in differences since all residuals in deviations are mathematically interrelated. The Arellano-Bond test of autocorrelation is applicable to any GMM regression on panel data provided none of the regressors depend on future errors and the errors are not correlated across individuals (Roodman, 2006).

9 Conclusion

This paper presented evidence on the performance of three different estimators for the dynamic random effects probit model proposed in the literature, namely, the Heckman (1981), Wooldrgidge (2005) and Orme (2001) estimators. It did so by estimating a model for the determinants of work-related participation amongst British employees. The estimates were then compared to those from a dynamic limited probability model using GMM techniques, namely the estimators proposed by Arellano and Bond (1991) and Blundell and Bond (1998).

The results suggest that for the dynamic random effects probit model the performance of no one estimator is superior to the others and given the recent development of a routine that allows implementation of the Heckman estimator with standard software by Stewart (2005), the choice is entirely up to the analyst's preferences as to the approximation method for the initial conditions. GMM estimation of a dynamic LPM of training participation suggests that the random effects estimators are not sensitive to the distributional assumptions of the unobserved effect.

References

ARELLANO, M., and S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 58, 277-297.

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

ARELLANO, M., and B. E. HONORE (2000): "Panel Data Models: Some Recent Developments," *CEMFI Working Paper No. 0016*.

ARULAMPALAM, W. (1998): "A Note on Estimated Coefficients in Random Effects Probit Models," *The Warwick Economics Research Paper Series (TWERPS)*.

AVERY, R. B., L. P. HANSEN, and V. J. HOTZ (1983): " Multiperiod Probit Models and Orthogonality Condition Estimation," *International Economic Review 24, pp. 21-35.*

BALTAGI, B. (1995): *Econometric Analysis of Panel Data*. New York: John Wiley and Sons.

BLUNDELL, R., and S. BOND (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-143.

BREITUNG, J., and M. LECHNER (1995): "Gmm-Estimation of Nonlinear Models on Panel Data," *Sonderforschungsbereich 373, Humboldt Universitaet Berlin*.

CHAMBERLAIN, G. (1984): "Panel Data," *Handbook of Econometrics, S. Griliches and M. Intriligator (eds)*, Amsterdam, North-Holland, 1247-318.

CHAY, K. Y., and D. R. HYSLOP (2000): "Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence Using Aletrnative Approaches," *University of California at Berkeley, Working Paper Series*.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

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, 114-78.

HOLTZ-EAKIN, D., W. K. NEWEY, and H. S. ROSEN (1989): "Implementing Causality Tests with Panel Data, with an Example from Localpublic Finance," *NBER Technical Working Papers, 0048*.

HONORE, B. E., and E. KYRIAZIDOU (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839-874.

— (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839-874.

HSIAO, C. (1986): *Analysis of Panel Data*. Cambridge UK: Cambridge University Press.

MUNDLAK, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69-85.

ORME, C. D. (2001): "The Initial Conditions Problem and Two-Step Estimation in Discrete Panel Data Models," *mimeo*, University of Manchester.

ROODMAN, D. (2006): "How to Do Xtabond2: An Introduction To "Difference" And "System" Gmm in Stata," *Working Papers*, Center for Global Development.

STEWART, M. B. (2006): "-Redprob- a Stata Program for the Heckman Estimator of the Random Effects Dynamic Probit Model," *mimeo, University of Warwick*.

— (2007): "The Interrelated Dynamics of Unemployemnt and Low-Wage Employment," *Journal of Applied Econometrics*, 22, 511-531.

WOOLDRIDGE, J. (2002): "Econometric Analysis of Cross Section and Panel Data," *The*

MIT Press.

WOOLDRIDGE, J. M. (1995): "Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions," *Journal of Econometrics*, 68, 115-132.

— (2005): "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, 20, 39-54.