Estimating Nonlinear Models with Multiple Fixed Effects: A Computational Note*

Manuel Arellano
CEMFI

Laura Hospido
Bank of Spain - Research Division

This draft: January 2008
First draft: March 2006

Abstract

In this paper we consider estimation of nonlinear panel data models that include multiple individual fixed effects. Estimation of these models is complicated both by the difficulty of estimating models with possibly thousands of coefficients and also by the incidental parameters problem; that is, noisy estimates of the fixed effects when the time dimension is short contaminate the estimates of the common parameters due to the nonlinearity of the problem. We propose a bias corrected likelihood estimator which can exploit the additivity of the effects for numerical optimization, thereby avoiding the calculation of estimates of the effects for given values of the common parameters. We exhibit the performance of this new estimator in simulations.

JEL Code: C23.

Keywords: Panel data, non-linear models, multiple fixed effects, incidental parameters, bias reduction.

*We thank Stéphane Bonhomme and seminar participants at the XXXII Simposio de Análisis Económico in Granada, for useful comments. All errors are our responsability.

1. Introduction

In a typical nonlinear micropanel data model with fixed effects there are hundreds or thousands of individual coefficients to estimate together with a relatively small number of common parameters. A well known computational simplification in the linear model is to obtain first the maximum likelihood (ML) estimates of the common parameters from a regression on the data in deviations from individual means, and secondly retrieve ML estimates of the effects from averaged residuals one by one. A similar computational simplification is available for Newton-Raphson and related algorithms for nonlinear fixed effects models, which exploits the block-diagonal structure of the Hessian. This simplification has been discussed in Hall (1978), Chamberlain (1980), and Greene (2004) for nonlinear models with a scalar fixed effect. The first purpose of this work is to show how to use an iterated algorithm of this type in a nonlinear model with multiple fixed effects.

As first noted by Neyman and Scott (1948), when the time series dimension T is small relative to the cross-sectional dimension n, ML estimates of the common parameters can be severely biased, specially in dynamic models. This Incidental Parameters problem arises because the unobserved individual characteristics are replaced by noisy estimates, which bias estimates of model parameters. In particular, the bias of the MLE is of order $1/T$. In some special cases it is possible to obtain fixed T large n consistent estimators of certain common parameters, but these situations are more the exception than the rule. Alternatively, a number of additional approaches have been proposed to obtain approximately unbiased estimators as opposed to estimators with no bias at all[1]. One of these approaches consists of estimation from a bias corrected objective function relative to some target criterion[2]. In this paper we discuss the application of computationally efficient

[1] See Arellano and Hahn (2006a) for a review of this literature on bias-adjusted estimation methods for nonlinear panel data models with fixed effects.

[2] See Pace and Salvan (2005) for adjustments of this type for a generic concentrated likelihood with independent observations, Arellano and Hahn (2006a) for static nonlinear panel models and Arellano and Hahn (2006b) and Hospido (2007) for the dynamic case.

algorithms to modified concentrated likelihoods of this type to obtain estimators without bias to order $1/T$ in nonlinear panel models with multiple fixed effects.

The paper is organized as follows. Section 2 introduces the model and notation. Section 3 explains how the iterated algorithm works. Section 4 discusses its application to bias corrected concentrated likelihoods. Section 5 presents some simulation results. Finally, Section 6 concludes. Detailed derivations are given in the Appendix.

2. Model and Notation

Let us consider the following model for the joint density of T random vectors conditioned on initial observations, strictly exogenous variables, and fixed effects:

$$f\left(y_{i1},...,y_{iT} \mid y_{i0},x_{i1},...,x_{iT},\alpha_{i0}\right)=\prod_{t=1}^{T} f\left(y_{it} \mid y_{i(t-1)},x_{it},\alpha_{i0},\theta_0\right)$$

where θ_0 is a vector of common parameters and α_{i0} is a vector of fixed effects. We observe the random sample $\{y_{i0},...,y_{iT},x_{i0},...,x_{iT}\}_{i=1}^{n}$ and we denote $\alpha_0 = (\alpha'_{10},...,\alpha'_{n0})'$ and $\delta_0=(\theta'_0,\alpha'_0)'$. Let the log likelihood of one observation be

$$\ell_{it}\left(\theta,\alpha_i\right)=\ln f\left(y_{it} \mid y_{i(t-1)},x_{it},\alpha_i,\theta\right)$$

and let $\ell_i\left(\theta,\alpha_i\right)=\sum_{t=1}^{T}\ell_{it}\left(\theta,\alpha_i\right)$.

3. Efficient Newton-Raphson iteration

Let us consider the estimator

$$\begin{pmatrix} \widehat{\theta} \\ \widehat{\alpha} \end{pmatrix} = \arg\max_{\theta,\alpha}\sum_{i=1}^{n}\ell_i\left(\theta,\alpha_i\right)$$

and let first and second derivatives be denoted by

$$\begin{aligned} d_{\theta i} &= \frac{\partial \ell_i\left(\theta,\alpha_i\right)}{\partial\theta}, \quad d_{\alpha i}=\frac{\partial \ell_i\left(\theta,\alpha_i\right)}{\partial\alpha_i} \\ H_{\theta\theta i} &= \frac{\partial^2 \ell_i\left(\theta,\alpha_i\right)}{\partial\theta\partial\theta'}, \quad H_{\alpha\alpha i}=\frac{\partial^2 \ell_i\left(\theta,\alpha_i\right)}{\partial\alpha_i\partial\alpha'_i}, \quad H_{\theta\alpha i}=\frac{\partial^2 \ell_i\left(\theta,\alpha_i\right)}{\partial\theta\partial\alpha'_i}. \end{aligned}$$

The Kth step of the iteration of a computationally efficient algorithm for obtaining $\widehat{\theta}$ and $\widehat{\alpha}$ takes the form

$$\theta_{[K]} - \theta_{[K-1]} = -\left[\sum_{i=1}^{n}\left(H_{\theta\theta i} - H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}H_{\alpha\theta i}\right)\right]^{-1}\sum_{i=1}^{n}\left(d_{\theta i} - H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}d_{\alpha i}\right) \quad (3.1)$$

$$\alpha_{i[K]} - \alpha_{i[K-1]} = -H_{\alpha\alpha i}^{-1}\left[d_{\alpha i} + H_{\alpha\theta i}\left(\theta_{[K]} - \theta_{[K-1]}\right)\right], \quad (i = 1, ..., n) \quad (3.2)$$

where all derivatives are evaluated at $\theta_{[K-1]}$ and $\alpha_{i[K-1]}$.

This result can be easily proved using partitioned inverse formulae (a detailed derivation is in the Appendix A). It is a standard result in nonlinear estimation of models with many group effects.[3]

4. Adjusted Concentrated Likelihood

When T is short we may be interested to consider an estimator that maximizes a bias corrected concentrated likelihood of the type reviewed in Arellano and Hahn (2006a):

$$\widehat{\theta}_c = \arg\max_{\theta}\sum_{i=1}^{n}\left[\ell_i\left(\theta, \widehat{\alpha}_i\left(\theta\right)\right) + \beta_i\left(\theta, \widehat{\alpha}_i\left(\theta\right)\right)\right]$$

where

$$\widehat{\alpha}_i\left(\theta\right) = \arg\max_{\alpha}\ell_i\left(\theta, \alpha\right)$$

and $\beta_i\left(\theta, \alpha_i\right)$ is an adjustment term.

As long as the adjustment term depends on α, the iterated algorithm discussed above cannot be directly used for estimating $\widehat{\theta}_c$. Note that

$$\begin{pmatrix} \widehat{\theta}_c \\ \widehat{\alpha}_c \end{pmatrix} = \arg\max_{\theta,\alpha}\sum_{i=1}^{n}\left[\ell_i\left(\theta, \alpha_i\right) + \beta_i\left(\theta, \widehat{\alpha}_i\left(\theta\right)\right)\right]$$

where $\widehat{\alpha}_c = \widehat{\alpha}\left(\widehat{\theta}_c\right)$. Thus, if we use the analysis of covariance algorithm discussed in the previous section we still need to calculate $\widehat{\alpha}_i\left(\theta\right)$ for given values of θ.

[3] An alternative Gauss–Newton algorithm which leads to a regression-based iteration is discussed in Appendix B.

An Alternative, Computationally Effective Estimator Alternatively, we can consider an estimator of the form

$$\begin{pmatrix} \widetilde{\theta} \\ \widetilde{\alpha} \end{pmatrix} = \arg\max_{\theta,\alpha} \sum_{i=1}^{n} \left[\ell_i \left(\theta, \alpha_i \right) + \beta_i \left(\theta, \alpha_i \right) \right]$$

for which the iterated algorithm can be used. This is equivalent to:

$$\widetilde{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \left[\ell_i \left(\theta, \widetilde{\alpha}_i \left(\theta \right) \right) + \beta_i \left(\theta, \widetilde{\alpha}_i \left(\theta \right) \right) \right]$$

where

$$\widetilde{\alpha}_i \left(\theta \right) = \arg\max_{\alpha} \left[\ell_i \left(\theta, \alpha \right) + \beta_i \left(\theta, \alpha_i \right) \right].$$

The statistic $\widetilde{\alpha}_i \left(\theta \right)$ can be regarded as a Bayesian estimator that uses $e^{\beta_i(\theta,\alpha_i)}$ as the prior distribution of α_i for a given value of θ. Thus, under general conditions, $\widetilde{\alpha}_i \left(\theta \right)$ will be asymptotically equivalent to $\widehat{\alpha}_i \left(\theta \right)$, and $\widetilde{\theta}$ will have similar (bias reducing) properties as $\widehat{\theta}$ (see Severini, 1998, section 4, for a discussion on the use of adjusted concentrated likelihoods using alternative estimates of nuisance parameters).

It appears that $\widetilde{\theta}$ is not only computationally convenient, but it may also exhibit improved finite sample properties in certain situations due to the replacement of $\widehat{\alpha}_i \left(\theta \right)$ by $\widetilde{\alpha}_i \left(\theta \right)$.

Bias Expression The form of the approximate bias (Arellano and Hahn, 2006a) is

$$\beta_i \left(\theta \right) \approx \frac{1}{2} tr \left(H_i^{-1} \left(\theta \right) \Upsilon_i \left(\theta \right) \right), \tag{4.1}$$

where

$$H_i \left(\theta \right) = - \lim_{T \to \infty} E \left[\frac{\partial^2 \ell_i \left(\theta, \bar{\alpha}_i \left(\theta \right) \right)}{\partial \alpha_i \partial \alpha_i'} \right],$$

$$\Upsilon_i \left(\theta \right) = \lim_{T \to \infty} TE \left[\frac{\partial \ell_i \left(\theta, \bar{\alpha}_i \left(\theta \right) \right)}{\partial \alpha_i} \cdot \frac{\partial \ell_i \left(\theta, \bar{\alpha}_i \left(\theta \right) \right)}{\partial \alpha_i'} \right],$$

and

$$\bar{\alpha}_i \left(\theta \right) \equiv \arg\max_{\alpha} \lim_{T \to \infty} E \left[\frac{1}{T} \sum\nolimits_{t=1}^{T} \ell_{it} \left(\theta, \alpha \right) \right].$$

Adjustment Terms Given (4.1), for estimation we consider the following adjustment term

$$\beta_i(\theta, \alpha_i) \approx \frac{1}{2} tr\left(H_i^{-1}(\theta, \alpha_i)\Upsilon_i(\theta, \alpha_i)\right),$$

with

$$H_i(\theta, \alpha_i) = -\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial^2 \ell_{it}(\theta, \alpha)}{\partial \alpha_i \partial \alpha_i'}\right].$$

For $\Upsilon_i(\cdot)$, in static models, we can also use its observed sample counterpart (Arellano and Hahn, 2006a):

$$\Upsilon_i(\theta, \alpha_i) = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial \ell_{it}(\theta, \alpha)}{\partial \alpha_i} \cdot \frac{\partial \ell_{it}(\theta, \alpha)}{\partial \alpha_i'}\right].$$

However, in dynamic models, using the observed quantities evaluated at $\hat{\alpha}_i(\theta)$ or $\tilde{\alpha}_i(\theta)$ will not work. Instead, we use a trimmed version (Arellano and Hahn, 2006b):

$$\begin{aligned}
\Upsilon_i(\theta, \alpha_i) &= \Omega_0 + \sum_{l=1}^{r}(\Omega_l + \Omega_l'), \\
\Omega_l &= \frac{1}{T-l}\sum_{t=l+1}^{T}\left(1 - \frac{l}{r+1}\right)\frac{\partial \ell_{it}(\theta, \alpha)}{\partial \alpha_i} \cdot \frac{\partial \ell_{it-l}(\theta, \alpha)}{\partial \alpha_i'}.
\end{aligned}$$

Jackknife-corrected likelihood The jackknife-corrected log-likelihood (Dhaene, Jochmans and Thuysbaert, 2006) is defined as

$$\dot{\ell}_{1/2}(\theta) \equiv 2\ell(\theta) - \bar{\ell}_{1/2}(\theta),$$

where

$$\ell(\theta) = \sum_{i=1}^{n} \ell_i(\theta, \hat{\alpha}_i(\theta))$$

and $\bar{\ell}_{1/2}(\theta)$ is the average of two half-panel concentrated log-likelihood functions. So, the maximum jackknife-corrected likelihood estimator is given by

$$\widehat{\theta}_c = \arg\max_{\theta} \dot{\ell}_{1/2}(\theta).$$

Alternatively, we can consider an estimator of the form

$$\widetilde{\theta} = \arg\max_{\theta} \left[2 \sum_{i=1}^{n} \ell_i \left(\theta, \tilde{\alpha}_i \left(\theta \right) \right) - \tilde{\ell}_{1/2} \left(\theta \right) \right]$$

where

$$\widetilde{\alpha}_i \left(\theta \right) = \arg\max_{\alpha} \left[2 \sum_{t=1}^{T} \ell_{it} \left(\theta, \alpha \right) - \left(\frac{\sum_{t=1}^{T/2} \ell_{it} \left(\theta, \alpha \right) + \sum_{t=(T/2)+1}^{T} \ell_{it} \left(\theta, \alpha \right)}{2} \right) \right],$$

and $\tilde{\ell}_{1/2} \left(\theta \right)$ is the average of the corresponding two half-panel log-likelihood functions.

5. Monte Carlo Study

We consider four probit examples, keeping the simulation design as consistent as possible across models[4]. Thus,

$$y_{it} = \mathbf{1} \left[w_{it} + \epsilon_{it} > 0 \right]; \quad (t = 1, \ldots, T; i = 1, \ldots, n)$$

where $\epsilon_{it} \sim N \left(0, 1 \right)$, $T = \{6, 8, 10, 12, 20\}$, and $n = 1,000$.

- Static model with scalar fixed effects:

$$w_{it} = \eta_i + \delta d_{it} + \beta x_{it}; \quad \left(\theta = \left[\beta, \delta \right]' ; \alpha_i = \eta_i \right),$$

- Static model with multiple fixed effects:

$$w_{it} = \eta_i + \delta_i d_{it} + \beta x_{it}; \quad \left(\theta = \beta; \alpha_i = \left[\eta_i, \delta_i \right]' \right),$$

[4] Other studies, that consider nonlinear designs with scalar fixed effects (Bester and Hansen (2005), Carro (2006) and Fernández-Val (2005)), show that the bias in the ML estimator is similar in magnitude for the logit and the probit models and that bias corrections also perform similarly. Here, we focus on probit designs and extend the analysis to consider multiple fixed effects.

- Dynamic model with scalar fixed effects:

$$w_{it} = \eta_i + \delta y_{it-1} + \beta x_{it}; \quad \left(\theta = [\beta, \delta]'; \alpha_i = \eta_i\right),$$

- Dynamic model with multiple fixed effects:

$$w_{it} = \eta_i + \delta_i y_{it-1} + \beta x_{it}; \quad \left(\theta = \beta; \alpha_i = [\eta_i, \delta_i]'\right).$$

Design 1: Static Probit with Scalar Fixed Effects.

$$y_{it} = \mathbf{1}\left[\eta_i + \delta d_{it} + \beta x_{it} + \epsilon_{it} > 0\right]; \; \epsilon_{it} \sim N(0,1)$$

d_{it}	$\mathbf{1}[x_{it} + h_{it} > 0]$
h_{it}	$N(0,1)$
x_{it}	$N(0,1)$
δ_0	0.5
β_0	1
η_i	0, $\forall i$

Table 1. Static Probit with Scalar Fixed Effects for different values of T.

		$T = 6$			$T = 8$			$T = 12$		
		Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
MLE	$\hat{\beta}$	0.302	0.302	0.306	0.191	0.190	0.195	0.115	0.113	0.117
	$\hat{\delta}$	0.148	0.151	0.161	0.094	0.094	0.106	0.061	0.063	0.070
MMLE1	$\hat{\beta}$	0.161	0.162	0.167	0.074	0.073	0.082	0.029	0.027	0.036
	$\hat{\delta}$	0.081	0.085	0.099	0.039	0.042	0.059	0.020	0.028	0.038
MMLE2	$\hat{\beta}$	0.161	0.161	0.167	0.074	0.073	0.082	0.029	0.027	0.036
	$\hat{\delta}$	0.081	0.844	0.099	0.039	0.042	0.059	0.020	0.028	0.038
MMLE3	$\hat{\beta}$	0.116	0.114	0.124	0.042	0.041	0.055	0.009	0.015	0.024
	$\hat{\delta}$	0.066	0.065	0.088	0.032	0.038	0.055	0.013	0.024	0.034
MMLE4	$\hat{\beta}$	0.116	0.114	0.114	0.042	0.041	0.055	0.009	0.015	0.024
	$\hat{\delta}$	0.066	0.065	0.088	0.032	0.038	0.055	0.013	0.024	0.034

Note: 100 simulations. Bias = mean bias. MAE = median absolute error. RMSE = root mean squared error. MLE: maximum likelihood estimator. MMLE1: corrected concentrated MLE. MMLE2: corrected efficient MLE. MMLE3: split-jackknife concentrated MLE. MMLE4: split-jackknife efficient MLE.

Design 2: Static Probit with Multiple Fixed Effects.

$$y_{it} = \mathbf{1}\left[\eta_i + \delta_i d_{it} + \beta x_{it} + \epsilon_{it} > 0\right];\ \epsilon_{it} \sim N(0,1)$$

d_{it}	$\mathbf{1}[x_{it} + h_{it} > 0]$
h_{it}	$N(0,1)$
x_{it}	$N(0,1)$
β_0	1
η_i	0, $\forall i$
δ_i	0.5, $\forall i$

Table 2. Static Probit with Multiple Fixed Effects for different values of T.

		$T = 6$	$T = 8$	$T = 12$	$T = 20$
MLE	Bias	0.563	0.415	0.259	0.142
	MAE	0.556	0.413	0.255	0.142
	RMSE	0.568	0.418	0.261	0.143
MMLE1	Bias	0.433	0.280	0.137	0.051
	MAE	0.429	0.280	0.134	0.051
	RMSE	0.438	0.283	0.141	0.054
MMLE2	Bias	0.429	0.288	0.135	0.050
	MAE	0.429	0.280	0.134	0.051
	RMSE	0.435	0.280	0.139	0.053
MMLE3	Bias	0.446	0.281	0.125	0.036
	MAE	0.443	0.277	0.125	0.037
	RMSE	0.451	0.285	0.129	0.041
MMLE4	Bias	0.446	0.281	0.127	0.036
	MAE	0.441	0.277	0.125	0.038
	RMSE	0.451	0.285	0.131	0.042

Note: 100 simulations. Bias = mean bias. MAE = median absolute error. RMSE = root mean squared error. MLE: maximum likelihood estimator. MMLE1: corrected concentrated MLE. MMLE2: corrected efficient MLE. MMLE3: split-jackknife concentrated MLE. MMLE4: split-jackknife efficient MLE.

Design 3: Dynamic Probit with Scalar Fixed Effects.

$$y_{it} = \mathbf{1}\left[\eta_i + \delta y_{it-1} + \beta x_{it} + \epsilon_{it} > 0\right];\ \epsilon_{it} \sim N(0,1)$$

y_{i0}	$\mathbf{1}[\eta_i + \beta x_{i0} + \epsilon_{i0} > 0]$
x_{it}	$N(0,1)$
δ_0	0.5
β_0	1
η_i	$0, \forall i$

Table 3. Dynamic Probit with Scalar Fixed Effects for different values of T.

		$T=6$			$T=8$			$T=12$		
		Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
MLE	$\hat{\beta}$	0.258	0.260	0.261	0.185	0.185	0.189	0.106	0.107	0.108
	$\hat{\delta}$	−0.461	0.460	0.465	−0.320	0.317	0.323	−0.206	0.205	0.208
MMLE1	$\hat{\beta}$	0.182	0.183	0.186	0.104	0.104	0.110	0.040	0.041	0.045
	$\hat{\delta}$	−0.215	0.214	0.221	−0.127	0.125	0.133	−0.068	0.066	0.073
MMLE2	$\hat{\beta}$	0.183	0.184	0.187	0.105	0.105	0.111	0.041	0.041	0.046
	$\hat{\delta}$	−0.215	0.214	0.221	−0.128	0.126	0.133	−0.068	0.066	0.074
MMLE3	$\hat{\beta}$	0.101	0.101	0.107	0.047	0.045	0.057	0.010	0.016	0.023
	$\hat{\delta}$	−0.178	0.178	0.185	−0.096	0.093	0.104	−0.039	0.037	0.049
MMLE4	$\hat{\beta}$	0.101	0.101	0.107	0.047	0.045	0.057	0.010	0.016	0.023
	$\hat{\delta}$	−0.178	0.178	0.185	−0.096	0.093	0.104	−0.039	0.037	0.049

Note: 100 simulations. Bias = mean bias. MAE = median absolute error. RMSE = root mean squared error. MLE: maximum likelihood estimator. MMLE1: corrected concentrated MLE. MMLE2: corrected efficient MLE. MMLE3: split-jackknife concentrated MLE. MMLE4: split-jackknife efficient MLE.

Design 4: Dynamic Probit with Multiple Fixed Effects.

$$y_{it} = \mathbf{1}\left[\eta_i + \delta_i y_{it-1} + \beta x_{it} + \epsilon_{it} > 0\right];\ \epsilon_{it} \sim N(0,1)$$

y_{i0}	$\mathbf{1}[\eta_i + \beta x_{i0} + \epsilon_{i0} > 0]$
x_{it}	$N(0,1)$
β_0	1
η_i	$0,\ \forall i$
δ_i	$0.5,\ \forall i$

Table 4. Dynamic Probit with Multiple Fixed Effects for different values of T.

		$T=6$	$T=8$	$T=12$	$T=20$
MLE	Bias	0.543	0.405	0.250	0.131
	MAE	0.545	0.403	0.250	0.130
	RMSE	0.548	0.408	0.251	0.132
MMLE1	Bias	0.461	0.301	0.145	0.051
	MAE	0.467	0.295	0.145	0.050
	RMSE	0.466	0.305	0.147	0.053
MMLE2	Bias	0.457	0.299	0.144	0.051
	MAE	0.467	0.295	0.145	0.050
	RMSE	0.462	0.303	0.146	0.053
MMLE3	Bias	0.419	0.257	0.108	0.028
	MAE	0.422	0.256	0.107	0.027
	RMSE	0.425	0.262	0.110	0.032
MMLE4	Bias	0.419	0.257	0.108	0.028
	MAE	0.422	0.256	0.107	0.027
	RMSE	0.425	0.262	0.1105	0.032

Note: 100 simulations. Bias = mean bias. MAE = median absolute error. RMSE = root mean squared error. MLE: maximum likelihood estimator. MMLE1: corrected concentrated MLE. MMLE2: corrected efficient MLE. MMLE3: split-jackknife concentrated MLE. MMLE4: split-jackknife efficient MLE.

6. Conclusions

In this paper we consider estimation of nonlinear panel data models that include multiple individual fixed effects. Estimation of these models is complicated both by the difficulty of estimating models with possibly thousands of coefficients and also by the incidental parameters problem; that is, noisy estimates of the fixed effects when the time dimension is short contaminates the estimates of the common parameters due to the nonlinearity of the problem. We show how to use an iterated algorithm which simplifies estimation in a nonlinear model with multiple fixed effects and we also discuss its application to bias corrected concentrated likelihoods.

Simulations show that the estimator proposed is not only computationally convenient but it is also as good as others in a variety of probit designs. Different adjustments of the likelihood function result in bias corrected estimators that perform comparably to other bias corrections proposed in the literature. We can think in many microeconometric applications that use nonlinear panel data models. The results of the paper suggest that bias corrected estimates will be very useful in relevant empirical settings given the sample sizes of the panels more often used by researchers and, moreover, because they allow us to introduce more individual heterogeneity to address endogeneity concerns in a robust way.

References

Arellano, M. and J. Hahn (2006a): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments". In: R. Blundell, W. Newey, and T. Persson (eds.): *Advances in Economics and Econometrics, Ninth World Congress*, Cambridge University Press, forthcoming.

Arellano, M. and J. Hahn (2006b): "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects", unpublished manuscript.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974): "Estimation and Inference in Nonlinear Structural Models", *Annals of Economic and*

Social Measurement, 3/4, 653–665.

Bester, C.A., and C. Hansen (2005): "A Penalty Function Approach to Bias Reduction in Non-linear Panel Models with Fixed Effects", unpublished manuscript.

Carro, J. M. (2006): "Estimating Dynamic Panel Data Discrete Choice Models", *Journal of Econometrics,* forthcoming.

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

Dhaene, G., K. Jochmans, and B. Thuysbaert (2006): "Split-panel jackknife estimation of fixed effects models", unpublished manuscript.

Fernández-Val, I. (2005): "Estimation of Structural Parameters and Marginal Effects in Binary Choice Panel Data Models with Fixed Effects", unpublished manuscript.

Greene, W. (2004): "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects", *The Econometrics Journal*, 7, 98–119.

Hall, B. H. (1978): "A General Framework for Time Series–Cross Section Estimation", *Annales de l'INSEE*, 30–31, 177–202.

Hospido, L. (2007): "Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages", unpublished manuscript.

Neyman, J. and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations", *Econometrica*, 16, 1-32.

Pace, L. and A. Salvan (2005): "Adjustments of the profile likelihood from a new perspective", *Journal of Statistical Planning and Inference*, forthcoming.

Severini, T. A. (1998): "Likelihood Functions for Inference in the Presence of a Nuisance Parameter", *Biometrika* 85, 507–522.

A. Newton-Raphson iteration

The Kth step of the Newton-Raphson iteration takes the form

$$\delta_K = \delta_{K-1} - \left(\frac{\partial^2 L\left(\delta_{K-1}\right)}{\partial\delta\partial\delta'}\right)^{-1} \frac{\partial L\left(\delta_{K-1}\right)}{\partial\delta},$$

or for shortness

$$\Delta\delta = -\left(\frac{\partial^2 L}{\partial\delta\partial\delta'}\right)^{-1} \frac{\partial L}{\partial\delta}$$

where $L\left(\delta\right) = \sum_{i=1}^{n} \ell_i\left(\theta, \alpha_i\right)$ and

$$\frac{\partial L}{\partial\delta} = \begin{pmatrix} \frac{\partial L}{\partial\theta} \\ \frac{\partial L}{\partial\alpha_1} \\ \vdots \\ \frac{\partial L}{\partial\alpha_n} \end{pmatrix} = \sum_{t=1}^{T} \begin{pmatrix} \sum_{i=1}^{n} \frac{\partial\ell_{it}(\theta,\alpha_i)}{\partial\theta} \\ \frac{\partial\ell_{1t}(\theta,\alpha_1)}{\partial\alpha_1} \\ \vdots \\ \frac{\partial\ell_{nt}(\theta,\alpha_n)}{\partial\alpha_n} \end{pmatrix} = \begin{pmatrix} d_\theta \\ d_\alpha \end{pmatrix}$$

$$\frac{\partial^2 L}{\partial\delta\partial\delta'} = \sum_{t=1}^{T} \begin{pmatrix} \sum_{i=1}^{n} \frac{\partial^2\ell_{it}(\theta,\alpha_i)}{\partial\theta\partial\theta'} & \frac{\partial^2\ell_{1t}(\theta,\alpha_1)}{\partial\theta\partial\alpha_1'} & \cdots & \frac{\partial^2\ell_{nt}(\theta,\alpha_n)}{\partial\theta\partial\alpha_n'} \\ \frac{\partial^2\ell_{1t}(\theta,\alpha_1)}{\partial\alpha_1\partial\theta'} & \frac{\partial^2\ell_{1t}(\theta,\alpha_1)}{\partial\alpha_1\partial\alpha_1'} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2\ell_{nt}(\theta,\alpha_n)}{\partial\alpha_n\partial\theta'} & 0 & \cdots & \frac{\partial^2\ell_{nt}(\theta,\alpha_n)}{\partial\alpha_n\partial\alpha_n'} \end{pmatrix} = \begin{pmatrix} H_{\theta\theta} & H_{\theta\alpha} \\ H_{\theta\alpha}' & H_{\alpha\alpha} \end{pmatrix}$$

and

$$d_\alpha = \begin{pmatrix} d_{\alpha 1} \\ \vdots \\ d_{\alpha n} \end{pmatrix}, \quad H_{\alpha\alpha} = \begin{pmatrix} H_{\alpha\alpha 1} & & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & H_{\alpha\alpha n} \end{pmatrix}, \quad H_{\theta\alpha} = \begin{pmatrix} H_{\theta\alpha 1} & \cdots & H_{\theta\alpha n} \end{pmatrix}$$

so that $d_\theta = \sum_{i=1}^{n} d_{\theta i}$ and $H_{\theta\theta} = \sum_{i=1}^{n} H_{\theta\theta i}$, and

$$H_{\theta\alpha}H_{\alpha\alpha}^{-1} = \begin{pmatrix} H_{\theta\alpha 1}H_{\alpha\alpha 1}^{-1} & \cdots & H_{\theta\alpha n}H_{\alpha\alpha n}^{-1} \end{pmatrix}$$

$$H_{\theta\alpha}H_{\alpha\alpha}^{-1}H_{\alpha\theta} = \sum_{i=1}^{n} H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}H_{\alpha\theta i}$$

Letting

$$\begin{pmatrix} H_{\theta\theta} & H_{\theta\alpha} \\ H_{\theta\alpha}' & H_{\alpha\alpha} \end{pmatrix}^{-1} = \begin{pmatrix} H^{\theta\theta} & H^{\theta\alpha} \\ H^{\theta\alpha\prime} & H^{\alpha\alpha} \end{pmatrix}$$

where

$$\begin{aligned}
H^{\theta\theta} &= \left(H_{\theta\theta} - H_{\theta\alpha}H_{\alpha\alpha}^{-1}H_{\alpha\theta}\right)^{-1} \\
H^{\theta\alpha} &= -H^{\theta\theta}H_{\theta\alpha}H_{\alpha\alpha}^{-1} \\
H^{\alpha\alpha} &= H_{\alpha\alpha}^{-1} + H_{\alpha\alpha}^{-1}H_{\alpha\theta}H^{\theta\theta}H_{\theta\alpha}H_{\alpha\alpha}^{-1}
\end{aligned}$$

the partitioned formula gives:

$$\begin{pmatrix} \Delta\theta \\ \Delta\alpha \end{pmatrix} = -\begin{pmatrix} H^{\theta\theta} & H^{\theta\alpha} \\ H^{\theta\alpha\prime} & H^{\alpha\alpha} \end{pmatrix} \begin{pmatrix} d_\theta \\ d_\alpha \end{pmatrix}$$

or

$$\begin{aligned}
\Delta\theta &= -\left(H^{\theta\theta}d_\theta + H^{\theta\alpha}d_\alpha\right) \\
\Delta\alpha &= -\left(H^{\theta\alpha\prime}d_\theta + H^{\alpha\alpha}d_\alpha\right).
\end{aligned}$$

We have

$$\begin{aligned}
H^{\theta\theta} &= \left[\sum_{i=1}^{n}\left(H_{\theta\theta i} - H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}H_{\alpha\theta i}\right)\right]^{-1} \\
H^{\theta\alpha}d_\alpha &= -H^{\theta\theta}H_{\theta\alpha}H_{\alpha\alpha}^{-1}d_\alpha = -H^{\theta\theta}\sum_{i=1}^{n}H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}d_{\alpha i}
\end{aligned}$$

and

$$\begin{aligned}
-\Delta\theta &= H^{\theta\theta}d_\theta + H^{\theta\alpha}d_\alpha = H^{\theta\theta}d_\theta - H^{\theta\theta}\sum_{i=1}^{n}H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}d_{\alpha i} \\
&= H^{\theta\theta}\left(d_\theta - \sum_{i=1}^{n}H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}d_{\alpha i}\right)
\end{aligned}$$

so that

$$\Delta\theta = -\left[\sum_{i=1}^{n}\left(H_{\theta\theta i} - H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}H_{\alpha\theta i}\right)\right]^{-1}\sum_{i=1}^{n}\left(d_{\theta i} - H_{\theta\alpha i}H_{\alpha\alpha i}^{-1}d_{\alpha i}\right).$$

Similarly, we have[5]

$$\Delta\alpha = -H_{\alpha\alpha}^{-1}\left(d_\alpha + H_{\alpha\theta}\Delta\theta\right)$$

so that

$$\Delta\alpha_i = -H_{\alpha\alpha i}^{-1}\left(d_{\alpha i} + H_{\alpha\theta i}\Delta\theta\right), \qquad (i = 1, ..., n)$$

B. A regression-based iteration

Alternatively, we may consider a Gauss–Newton approach after enforcing block diagonality. The motivation is the same as in Berndt, Hall, Hall, and Hausman (1974) in that the nonzero components of the Hessian are approximated by outer product terms. The advantages of this procedure are that it only requires first derivatives and that it leads to a regression-based iteration.

Let us introduce the notation

$$\begin{aligned} d_{\theta it} &= \frac{\partial \ell_{it}(\theta,\alpha_i)}{\partial\theta}, \qquad d_{\alpha it} = \frac{\partial \ell_{it}(\theta,\alpha_i)}{\partial\alpha_i} \\ \Psi_{\theta\theta i} &= \sum_{t=1}^{T} d_{\theta it}d_{\theta it}', \qquad \Psi_{\alpha\alpha i} = \sum_{t=1}^{T} d_{\alpha it}d_{\alpha it}', \qquad \Psi_{\theta\alpha i} = \sum_{t=1}^{T} d_{\theta it}d_{\alpha it}'. \end{aligned}$$

The Kth step of the iteration of the Gauss–Newton algorithm for obtaining $\widehat{\theta}$ and $\widehat{\alpha}$ takes the form

$$\theta_{[K]} - \theta_{[K-1]} = -\left[\sum_{i=1}^{n}\left(\Psi_{\theta\theta i} - \Psi_{\theta\alpha i}\Psi_{\alpha\alpha i}^{-1}\Psi_{\alpha\theta i}\right)\right]^{-1}\sum_{i=1}^{n}\left(d_{\theta i} - \Psi_{\theta\alpha i}\Psi_{\alpha\alpha i}^{-1}d_{\alpha i}\right)$$

$$\alpha_{i[K]} - \alpha_{i[K-1]} = -\Psi_{\alpha\alpha i}^{-1}\left[d_{\alpha i} + \Psi_{\alpha\theta i}\left(\theta_{[K]} - \theta_{[K-1]}\right)\right], \qquad (i = 1, ..., n)$$

[5]Note that

$$\begin{aligned} -\Delta\alpha &= H^{\theta\alpha\prime}d_\theta + H^{\alpha\alpha}d_\alpha \\ &= -H_{\alpha\alpha}^{-1}H_{\alpha\theta}H^{\theta\theta}d_\theta + \left(H_{\alpha\alpha}^{-1} + H_{\alpha\alpha}^{-1}H_{\alpha\theta}H^{\theta\theta}H_{\theta\alpha}H_{\alpha\alpha}^{-1}\right)d_\alpha \\ &= H_{\alpha\alpha}^{-1}\left(d_\alpha + H_{\alpha\theta}H^{\theta\theta}H_{\theta\alpha}H_{\alpha\alpha}^{-1}d_\alpha - H_{\alpha\theta}H^{\theta\theta}d_\theta\right) \\ &= H_{\alpha\alpha}^{-1}\left(d_\alpha - H_{\alpha\theta}H^{\theta\alpha}d_\alpha - H_{\alpha\theta}H^{\theta\theta}d_\theta\right) \\ &= H_{\alpha\alpha}^{-1}\left[d_\alpha - H_{\alpha\theta}\left(H^{\theta\alpha}d_\alpha + H^{\theta\theta}d_\theta\right)\right] \\ &= H_{\alpha\alpha}^{-1}\left(d_\alpha + H_{\alpha\theta}\Delta\theta\right). \end{aligned}$$

where all derivatives are evaluated at $\theta_{[K-1]}$ and $\alpha_{i[K-1]}$.

Thus,

$$\theta_{[K]} - \theta_{[K-1]} = -\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{d}_{\theta it}\widetilde{d}'_{\theta it}\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{d}_{\theta it}$$

where

$$\widetilde{d}_{\theta it} = d_{\theta it} - \widetilde{\Pi}_i d_{\alpha it}$$

and

$$\widetilde{\Pi}_i = \Psi_{\theta\alpha i}\Psi_{\alpha\alpha i}^{-1},$$

so that the $\widetilde{d}_{\theta it}$ are the residuals of individual-specific regressions of $d_{\theta it}$ on $d_{\alpha it}$. Next, $\theta_{[K]} - \theta_{[K-1]}$ can be calculated as a pooled regression of minus one on $\widetilde{d}_{\theta it}$. Finally, $\alpha_{i[K]} - \alpha_{i[K-1]}$ can be obtained as a regression of $-\left[1 + d'_{\theta it}\left(\theta_{[K]} - \theta_{[K-1]}\right)\right]$ on $d_{\alpha it}$:

$$\alpha_{i[K]} - \alpha_{i[K-1]} = -\left(\sum_{t=1}^{T} d_{\alpha it}d'_{\alpha it}\right)^{-1}\sum_{t=1}^{T} d_{\alpha it}\left[1 + d'_{\theta it}\left(\theta_{[K]} - \theta_{[K-1]}\right)\right], \quad (i = 1, ..., n).$$