Working Paper Series
Department of Economics
University of Verona

Autocorrelation and masked heterogeneity in panel data models estimated by maximum likelihood

Giorgio Calzolari, Laura Magazzini

WP Number: 53 February 2009

ISSN: 2036-2919 (paper), 2036-4679 (online)

Autocorrelation and masked heterogeneity in panel data models estimated by maximum likelihood

Giorgio Calzolari* Laura Magazzini†

This version: March 22, 2011

Abstract

In a panel data model with random effects, when autocorrelation in the error is considered, (Gaussian) maximum likelihood estimation produces a dramatically large number of corner solutions: the variance of the random effect appears (incorrectly) to be zero, and a larger autocorrelation is (incorrectly) assigned to the idiosyncratic component. Thus heterogeneity could (incorrectly) be lost in applications to panel data with customarily available time dimension, even in a correctly specified model. The problem occurs in linear as well as nonlinear models. This paper aims at pointing out how serious this problem can be (largely neglected by the panel data literature). A set of Monte Carlo experiments is conducted to highlight its relevance, and we explain this unpleasant effect showing that, along a direction, the expected log-likelihood is nearly flat. We also provide two examples of applications with corner solutions.
Keywords: Panel data, autocorrelation, random effects, maximum likelihood, expected log-likelihood.

1 Introduction

Panel datasets comprise time series of cross-section pertaining to the same unit, or, put it differently, repeated cross-section at different points in time over the same sample. As pooling cross-section and time series data, a panel dataset would allow to jointly model individual heterogeneity and autocorrelation over time. Over the

*Department of Statistics, Università di Firenze, <calzolar@ds.unifi.it>
†Department of Economics, Università di Verona, <laura.magazzini@univr.it>

cross-sectional dimension, the error terms of different units can be thought of as mutually independent and heteroskedastic. When looking at data along the time dimension, residuals belonging to the same unit are expected to be autocorrelated (Kmenta, 1986 [24, ch. 12]; Wansbeek, 1992 [34]).

Different approaches have been proposed in the literature. For example, Kmenta (1986) [24, ch. 12] proposes a cross-sectional heteroskedastic and timewise autoregressive (AR) model, allowing both for heteroskedasticity among units, and autocorrelation over time. Most often, empirical analysis relies on the error component model, made popular by Balestra and Nerlove in 1966 [1] (see Baltagi, 1986 [2]; Baltagi *et al.*, 2008 [5]). This approach explicitly decomposes the error term as the sum of three different elements: (i) one component that varies over units but is constant over time, capturing individual heterogeneity; (ii) one component that varies over time but is constant across units, to model economywide variations not explicitly included in the regressors;[1] (iii) the "standard" regression residual that varies over both dimensions (idiosyncratic component). Even if "workhorse" in panel data analysis, this structure allows poor modeling of dynamics (Wansbeek, 1992 [34]). The idiosyncratic component is customarily assumed to be uncorrelated over time, therefore implying that the correlation between the error terms of the same unit at two different points in time is constant, independently of the time lag between the two time periods. This assumption might not be appropriate in the case of economic relationships, where the effect of an unobserved shock this period is often expected to wipe out as the time lag increases. Ignoring this pattern could be troublesome. Troubles are more severe in non-linear panel data models, where misspecification of the error structure can affect coefficient estimates (Stern, 1997 [32, fn. 25]). The problem is less severe in linear models, where coefficients can be consistently estimated in any case, and "robust" formulas can be applied for the computation of standard errors. In both cases, however, it could be interesting for the researcher to distinguish transitory (idiosyncratic) from permanent (individual) error component; the latter can be particularly interesting when heterogeneity is at the core of the analysis.

The purpose of this paper is to show an unpleasant behavior of the Gaussian maximum likelihood estimates for linear panel data models with random effects and autocorrelated idiosyncratic noise. The problem can easily invalidate results of applications. We deal with a "correctly specified" Gaussian framework, with various combinations of the three error term parameters (the variance of the individual random effect, the variance of the idiosyncratic noise, the autoregression parameter). Monte Carlo simulation of the data, followed by maximum likelihood estimation, very often produces a zero estimate of the variance of the individual

[1] As the time dimension in panel dataset is typically short, this component is often accommodated by the use of dummy variables identifying the different time periods.

random effect parameter (corner solution): heterogeneity, even of a sizable amount, can be “masked”. Correspondingly, there is a shift of the other parameters, whose distributions are no more scattered only around their “true” values. Bi-modal distributions become the rule, rather than an exception.

The problem occurs in a dramatically large number of cases when the time dimension is very small, no matter what combination of parameters is adopted (balanced or completely unbalanced values of the two variances, small or large values of the autoregression parameter). Only enormously large numbers (unreasonable) of individuals can compensate very small time dimensions.

For moderately small time dimensions (T) the problem is still relevant when the autoregression parameter is large, for any combination of the two variance parameters. For a large value of the autoregression parameter and a variance of the individual random effect smaller than the idiosyncratic variance, the problem still occurs even for moderately large values of T.

For any combination of the parameters, the problem becomes less important when the time dimension enlarges and disappears for very large T. It is quite relevant for the N and T dimensions used in practical applications.

Whether the model is linear or non-linear, the problem seems to have been neglected by the literature. On the contrary, earlier work of the authors in the context of the Tobit model evidenced the existence of corner estimation for the variance of the indivual effect (Calzolari and Magazzini, 2008 [10]). Being clear that the difficulty arises from the error structure and not from the non-linearity, we focus in this paper on the linear case, where some analytical explanations can be provided, besides Monte Carlo evidence. We show that this unpleasant effect is driven by the shape of the expected log-likelihood, as it is nearly flat over a (large) subset of the admissible parameter space. A non-linear application will anyway be discussed.

The next section provides a description of the model of interest and of the related literature. Section 3 presents a Monte Carlo study in order to highlight the behavior of the (Gaussian) maximum likelihood estimator in typical panel data samples. Section 4 analyzes the expected log likelihood function (with the score and Hessian matrix). Section 5 provides two empirical examples where the problem presented in this paper arises in applications. Section 6 summarizes our results.

2 Notation and related literature

The data generating process on the dependent variable y_{it} is expressed as a linear function of a set of independent variables, x_{it}, and an error term, ν_{it}:

$$y_{it} = x'_{it}\beta + \nu_{it} \tag{1}$$

where i denotes the unit (individual, household, firm, country, ...), and the index t denotes time, with $i = 1, ..., N; t = 1, ..., T$.

Kmenta (1986) [24, ch.12] combines the assumption of heteroskedasticity, usual in a cross-section context, with the assumption of autocorrelation over time, typical of time series modeling. In its simplest form, Kmenta suggests to specify:

$$\nu_{it} = \rho\nu_{it-1} + w_{it} \tag{2}$$

where $w_{it} \sim N(0, \sigma^2_{w(i)})$. Different units in the sample are assumed to be independent.[2] The model is estimated by correcting for serial correlation in the first step and heteroskedasticity in the second step.

More commonly in the literature, the error components approach decomposes ν_{it} into three independent terms:

$$\nu_{it} = \alpha_i + \lambda_t + e_{it} \tag{3}$$

where α_i is the individual effect, representing all the time-invariant characteristics of unit i (unobserved or unobservable heterogeneity), λ_t is the time effect, representing all the characteristics of time t, invariant across all the cross-sectional units in the sample, and e_{it} is a random term that varies over time and individuals (idiosyncratic noise). The error term λ_t is not considered in the analysis, since it can be easily accounted for in a typical (short-T) panel setting by inserting time-dummies in the regression. Thus, we are considering

$$\nu_{it} = \alpha_i + e_{it} \tag{4}$$

In standard settings, the error term e_{it} is assumed to be serially uncorrelated. All the correlation in ν_{it} over time is ascribed to the presence of the individual effect. As a result, the autocorrelation in ν_{it} is constant over time, independently of the time lag between the two observations considered: $E[\nu_{it}\nu_{is}] = \sigma^2_\alpha$ for each $s \neq t, t = 1, ..., T$. This assumption is not suited to situations where the effect of unobserved variables varies systematically over time, as in the case of serially correlated omitted variables or transitory variables whose effect lasts more than one period. In order to take into account these variables and provide a more general autocorrelation scheme, autocorrelation can be considered among e_{it} for

[2] Kmenta also proposes more complex formulations, where the autocorrelation parameter ρ is allowed to vary for each unit in the sample, and some correlation may exist between the units in the sample (Kmenta, 1986 [24, ch. 12]).

the same individual.[3] In this paper we will consider disturbances generated by an AR(1) process:[4]

$$e_{it} = \rho e_{it-1} + w_{it} \tag{5}$$

with w_{it} homoskedastic, uncorrelated, and with mean zero. Normality of all the error terms is assumed.[5]

As a result of these assumptions, the variance-covariance matrix of the error term ν_{it} has the following structure:

$$E[\nu_{it}\nu_{js}] = \begin{cases} \sigma_\alpha^2 + \sigma_e^2 & \text{if } i = j, t = s \\ \sigma_\alpha^2 + \rho^{|t-s|}\sigma_e^2 & \text{if } i = j, t \neq s \\ 0 & \text{if } i \neq j \end{cases} \tag{6}$$

If the explanatory variables x_{it} are strictly exogenous, there is no effect of the x_{it} on the problems we are considering in this paper. We can therefore focus on the error terms parameters only.[6] Assuming the error terms follow a Gaussian distribution, the likelihood depends only on the variance-covariance matrix, thus only on the three parameters σ_α^2, σ_e^2, ρ.

Autocorrelated disturbances in panel data linear models have been considered for the first time in longitudinal studies of wages and earnings (David, 1971 [12]; Hause, 1977 [17]; Lillard and Willis, 1978 [26]; Lillard and Weiss, 1979 [25]; MaCurdy, 1982 [27]; Bhargava *et al.*, 1982 [9]; Berry *et al.*, 1998 [8]).

Lillard and Willis (1978) [26] estimate an earning function with permanent and serially correlated transitory components due to both measured and unmeasured variables on data drawn by the US National Science Foundation's Register of Technical and Scientific Personnel. Their sample includes a balanced panel of 4,330 Ph.D. observed at two-years intervals over the decade 1960-70 ($T = 6$). As

[3]Recent research in linear models with random effects considers serial correlation in λ_t (Karlsson and Skoglund, 2004 [22]).

[4]More general error structures are also considered in the literature. In the linear case, MaCurdy (1982) [27] considers an $ARMA(p, q)$ specification; Hause (1977) [17] considers heteroskedastic innovations and time-varying AR parameters. In the non-linear (probit) case Inkman (2000) [21] considers the AR(1) process with heterosckedasticity.

[5]Without need of the normality assumption, Baltagi and Li (1991) [3] discuss GLS estimation of panel data models with serially correlated error component disturbances.

[6] Of course, $T \geq 3$. With $T = 2$, the covariance matrix would only contain two *different* elements ($\sigma_\alpha^2 + \sigma_e^2$ on the diagonal, and $\sigma_\alpha^2 + \sigma_e^2\rho$ off-diagonal), from which it is impossible to identify separately the three parameters of the model σ_α^2, σ_e^2 and ρ. $T = 3$ is the smallest possible time dimension that leads to three different elements in the variance-covariance matrix, making identification possible. Of course the situation becomes even better for larger values of T, where the presence of higher powers of ρ increases the number of different elements in the matrix. However, a first glance at equation (6) suggests that when $\rho^{|t-s|}$ is near 1, variances and covariances will be hardly distinguishable. Thus, it is reasonable to expect troubles in estimation when ρ is near 1 and/or T is small (then, $|t - s|$ is always very small).

transitory effects and permanent effects have different economic implications, the paper aims at separating the permanent and transitory elements of earnings development. As the sample covariance between the error terms at two different points in time is smaller for longer time lags, an AR specification of the idiosyncratic component is considered in the analysis. In order to estimate the model parameters, the authors first apply OLS to the data pooled over individuals and years, and then estimate the variance components and the autocorrelation parameter by applying maximum likelihood to the OLS residuals. This two-step approach is asymptotically equivalent to quasi-maximum likelihood procedure, robust to failure of the normality hypothesis (MaCurdy, 1982 [27]). Maximum likelihood estimation applied to OLS residuals for the identification of the components of the covariance matrix is also exploited in Lillard and Weiss (1979) [25], who draw data from the University of Michigan Panel Study of Income Dynamics (PSID). Data about 1,114 male head of households observed over 7 years are used for the estimation of the parameters of an empirical earning function with permanent and serially correlated transitory component due to both measured and unmeasured variables. The estimates are then employed to track the probability of time sequences of poverty states.

More recently, the one-way error component model with AR disturbances has been applied to study the impact of plant closing on the mean and variance of log earnings (Berry *et al.*, 1998 [8]). Data are drawn from the PSID over the period 1975-81 ($T = 7$).

Theoretical investigation has mainly focused on the development of testing procedures for identifying the presence of serial correlation in the disturbances.

Within the fixed effect framework, Bhargava *et al.* (1982) [9] develop a Durbin-Watson type statistics to test the model for serial independence and test the random walk hypothesis. An earning function is estimated using data from the Michigan Survey of Income Dynamics to illustrate the uses of the tests and the proposed estimation procedures.

A locally best invariant test for zero first order autocorrelation is provided by Baltagi and Wu (1999) [7], who also deal with unequally spaced panels.

Within the random effect framework, a series of LM statistics for testing the presence of serial correlation and individual effect is devised by Baltagi and Li [4]. The proposed LM statistics are invariant to the form of first order serial correlation, i.e. they can be applied both to AR(1) and MA(1) processes. Size and power of the tests are studied by means of Monte Carlo simulation for various combination of the autocorrelation parameter and ratio of the individual effect variance over the amount of the composite error variance. Different sample sizes are considered, changing both N and T, but always with $T \geq 10$.

Wooldridge (2002) [35, §10.6] proposes a new test based on the first differences of pooled OLS residuals. If e_{it} is serially uncorrelated then we expect $\text{Corr}(e_{it}, e_{it-1}) = -0.5$. The test is easily implemented and requires little assumptions with respect to available options (see also Drukker, 2003 [13]).

If the assumption of spherical disturbances for e_{it} is violated, as it is in our AR(1) setting, the ordinary formulae for estimating coefficient variances will lead to inconsistent standard errors. Although estimators of the variance exist that are robust to the presence of heteroskedasticity and serial correlation of arbitrary form for fixed T and large N, they have two main limitations. First, robust estimation can have poor finite sample properties as it requires the estimation of $T(T-1)/2$ parameters (Wooldridge, 2002 [35, §10.4]); then, it is not possible to distinguish between the two sources of variability. Finally, the AR(1) specification is customarily employed in order to tackle the issue of autocorrelation in the idiosyncratic component of non-linear panel data models (see, e.g., Roberts and Tybout, 1997 [31]; Contoyannis *et al.*, 2004 [11]). It has also been used as a benchmark for simulation-based estimators (Hajivassiliou, 1994 [14]; Keane, 1994 [23]).[7]

3 Monte Carlo Experiment

In this section, we present a Monte Carlo experiment that study the finite sample properties of the maximum likelihood estimator applied to panel data model (1). However, the presence of the exogenous variables has no impact on the troublesome effect that we want to evidence in this paper. The model is set as follows:

$$y_{it} = \nu_{it} = \alpha_i + e_{it}$$

where $\alpha_i \sim IN(0, \sigma_\alpha^2)$, and $e_{it} = \rho e_{i,t-1} + w_{it}$ with $w_{it} \sim IN(0, \sigma_e^2 \times (1-\rho^2))$. Therefore, our analytical and Monte Carlo analysis will focus only on the estimation of the three parameters σ_α^2, σ_e^2, ρ. Empirical applications, in Section 5, will obviously include exogenous explanatory variables.

By means of a wide Monte Carlo experiment, we shall show that, whatever the values of the two variance parameters, balanced ($\sigma_\alpha^2 = 0.5$ and $\sigma_e^2 = 0.5$), or quite unbalanced in either way ($\sigma_\alpha^2 = 0.1$ and $\sigma_e^2 = 0.9$, or viceversa $\sigma_\alpha^2 = 0.9$ and $\sigma_e^2 = 0.1$), corner solutions for the maxima can occur, providing an estimate $\hat{\sigma}_\alpha^2 = 0.0$; correspondingly, the estimate of σ_e^2 will compensate the zero, giving a value around the sum of the two "true" variances; the estimate of ρ will be close to a weighted average between the "true" value and one. The sample size chosen for the experiments resembles the values of N and T encountered in practical

[7] This paper has been inspired by the poor performance of the indirect inference approach in the context of the panel data Tobit model with random effects and AR(1) disturbances (Calzolari and Magazzini, 2008 [10]). The poor performance of the simulation estimator can be ascribed to the large incidence of corner solutions.

applications. In particular, we consider $N = 100, 1000$ units, and $T = 3, 5, 10, 20$ time periods.

Computations have been performed using own ad-hoc programs in Fortran 77 (double precision).

		σ_α^2 (with $\sigma_\alpha^2 + \sigma_e^2 = 1$)					
	ρ	0.1	0.5	0.9	0.1	0.5	0.9
$T = 3$	0.0	17.0%	0%	0%	0%	0%	0%
	0.5	39.0%	12.0%	2.0%	18.0%	0%	0%
	0.9	45.0%	40.0%	34.0%	45.0%	25.0%	5.0%
$T = 5$	0.0	1.0%	0%	0%	0%	0%	0%
	0.5	22.0%	0%	0%	0.6%	0%	0%
	0.9	43.0%	25.0%	16.0%	41.0%	7.5%	0%
$T = 10$	0.0	0%	0%	0%	0%	0%	0%
	0.5	5.8%	0%	0%	0%	0%	0%
	0.9	40.1%	3.3%	0%	17.4%	0%	0%
$T = 20$	0.0	0%	0%	0%	0%	0%	0%
	0.5	0.3%	0%	0%	0%	0%	0%
	0.9	25.6%	0.4%	0%	2.0%	0%	0%
		$N = 100$			$N = 1000$		

Table 1: Share of corner solutions ($\hat{\sigma}_\alpha^2 = 0$). 10000 Monte Carlo replications

Corner solutions occur more often when ρ is large and T is small. This confirms the intuition that troubles would arise in the case when ρ is near 1 and/or T is small (variances and covariances in equation (6) are hardly distinguishable; see footnote 6). However, few cases occur also when the "true" autoregression parameter is zero (but only if T, N and σ_α^2 are small). Some cases can be found even with the combination of large T and large N (but only when σ_α^2 is small).

The typical histograms of the three parameters will show: a probability "mass" at zero for $\hat{\sigma}_\alpha^2$, and a partial histogram around the "true" value; a bimodal distribution for each of the other two parameters. As an example, we report the Monte Carlo distribution of parameters for two different set ups (Figures 1 and 2). Inference would really be problematic (according to Table 1, there would be cases much worse than the two represented in the Figures)!

From the viewpoint of the empirical economist, a corner solution, $\hat{\sigma}_\alpha^2 = 0$, would mislead the conclusion of the analysis: individual heterogeneity does not appear.

A different parameterization, which avoids restrictions, can provide a different picture of the bimodal distribution.[8] Concentrating out one of the variance parameters, then defining $\gamma = \log(\sigma_\alpha^2/\sigma_\nu^2)$ and $\omega = \rho/\sqrt{1-\rho^2}$, the same results of Figure 1 appear as in Figure 3. Large negative values of the $\hat{\gamma}$ parameter correspond, in practice, to the corner solutions of Figure 1, while the autoregression parameter $\hat{\omega}$ has a clear bimodal distribution as it was for the autoregression parameter $\hat{\rho}$.

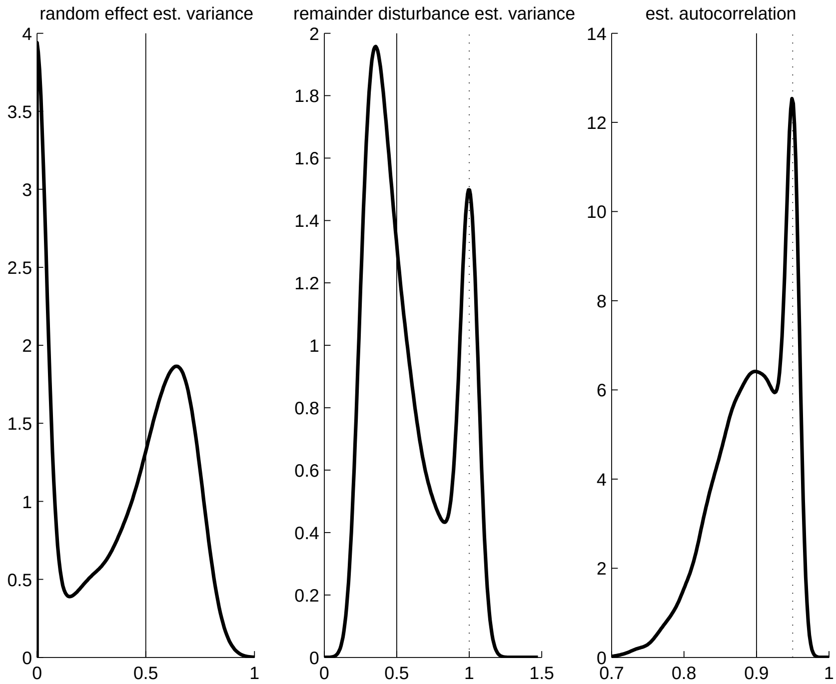

Figure 1: Monte Carlo distribution of estimated parameters: $T = 3$, $N = 1000$, $\sigma_\alpha^2 = \sigma_e^2 = 0.5$, and $\rho = 0.9$

[8]This parameterization has been explicitly proposed by one of the Referees of the paper that was submitted to Empirical Economics. We gratefully acknowledge this very useful suggestion.

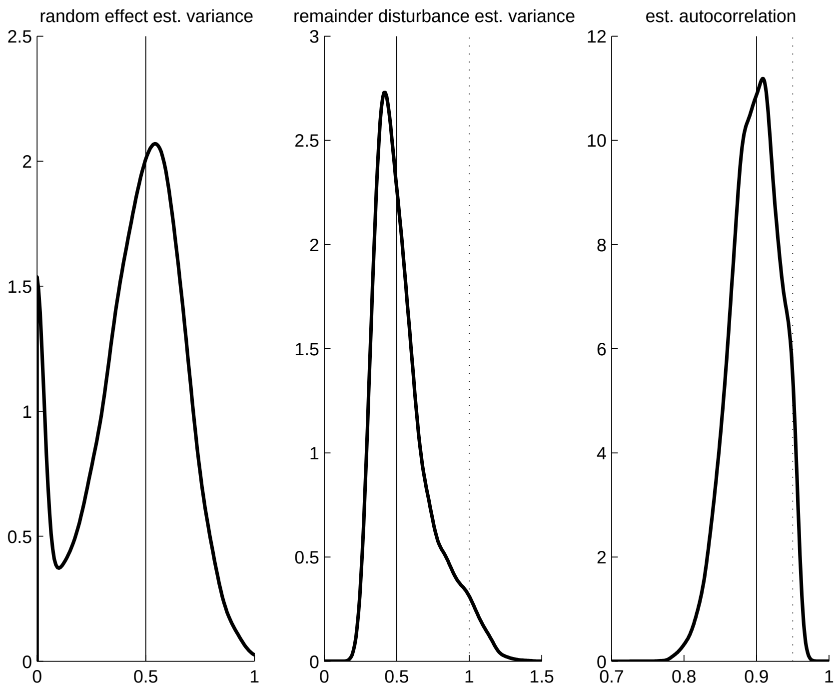

Figure 2: Monte Carlo distribution of estimated parameters: $T = 10$, $N = 100$, $\sigma_\alpha^2 = \sigma_e^2 = 0.5$, and $\rho = 0.9$

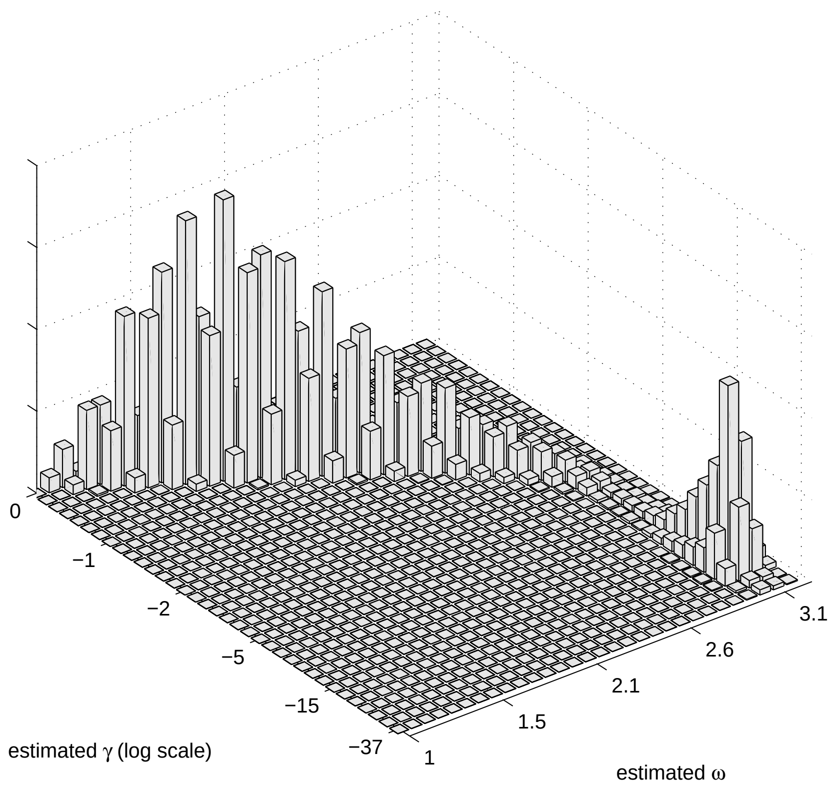

Figure 3: Monte Carlo distribution of estimated parameters under the alternative parametrization (set up as in Figure 1)

4 Numerical considerations on the expected log-likelihood

Ignoring the exogenous explanatory variables in model (1), the log-likelihood is

$$\log L(\nu,\theta) = \sum_{i=1}^{N} \log L(\nu_i,\theta) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log||B|| - \frac{1}{2}\sum_{i=1}^{N}(\nu_i' B^{-1}\nu_i) \quad (7)$$

where $\theta = [\sigma_\alpha^2, \sigma_e^2, \rho]'$ is the vector of parameters, and B is the $(T \times T)$ covariance matrix of the vector of error terms ν_i (4, 5). As from (6), B is therefore the sum of a matrix whose elements are all σ_α^2, say $\sigma_\alpha^2 \iota\iota'$ (where ι is the $T \times 1$ vector whose elements are all $= 1$) and the covariance matrix of the AR(1) process (5) with the well known Toepliz structure: $B = \sigma_\alpha^2 \iota\iota' + \sigma_e^2 A$ with $\{A_{ij}\} = \rho^{|i-j|}$.

Remembering that the particular structure of matrix A leads to a simple expression of the determinant $|A| = (1-\rho^2)^{T-1}$ and of the inverse (a tridiagonal matrix), the computation of the determinant and inverse of matrix B greatly benefits from the use of Woodbury (or Sherman-Morrison) matrix inversion lemma (e.g. Householder 1964 [20, §5.1]), both in terms of computation time as well as (more important) computation accuracy: no matrix inversion is involved in the formulas, so numerical correctness does not depend on the routines that perform numerical inversions. If one is not scared by the complexity of the equations (see Appendix A for details), their use guarantees the accuracy that sometimes we need when suspicion arises about the numerical value of a gradient at the maximum: is it zero or is it not zero? We shall frequently meet gradients whose norm is less than 10^{-6}, and are not zeroes (not to mention cases where 10^{-11} have to be discarded; any "standard" optimization program would achieve convergence at such points, and this might produce misleading results).

Let the vector ν_i be produced by a multivariate normal distribution with zero mean and "true" covariance matrix $R = \psi_\alpha^2 \iota\iota' + \psi_e^2 C$ with $\{C_{ij}\} = \phi^{|i-j|}$.

Taking expectation of equation (7), the expected value of the log-likelihood computed at any value of $\theta = [\sigma_\alpha^2, \sigma_e^2, \rho]'$ is

$$E[\log L(\nu,\theta)] = N\ E[\log L(\nu_i,\theta)] = -\frac{N}{2}\left[\log(2\pi) + \log||B|| + tr(B^{-1}R)\right] \quad (8)$$

The expression above has a unique maximum at the "true" parameter values: $\sigma_\alpha^2 = \psi_\alpha^2$, $\sigma_e^2 = \psi_e^2$, $\rho = \phi$ (so that, $B = R$),[9] thus ensuring, in principle, identification (Newey and McFadden, 1994 [30, p.2124]). However, along a particular direction, this expected log-likelihood is nearly flat, making it difficult to distinguish between different sets of parameters.

[9] The result follows from the so called "information inequality"; for instance Newey and McFadden (1994) [30, p.2124].

For example, for error terms produced by "true" parameter values $\psi_\alpha^2 = 0.5$, $\psi_e^2 = 0.5$, $\phi = 0.5$ and a "large" time dimension $T = 20$, comparing the expected log-likelihood at the "true" values $\sigma_\alpha^2 = 0.5$, $\sigma_e^2 = 0.5$, $\rho = 0.5$ with the value at $\sigma_\alpha^2 = 0.0$, $\sigma_e^2 = 1.0$, $\rho = 0.75$ the difference is about 50%; so the two points are easily distinguishable. But for $T = 5$ the difference is 6%, and for $T = 3$ (the smallest possible value for identification) the difference is only a bit more than 1%.

When the autocorrelation parameter is large (e.g. $\rho = 0.9$ as in Table 2), these differences become much smaller, and the risk of confusing the different points in numerical computations becomes not negligible.

In Table 2, we first compute, for several values of T (till the smallest possible value which is $T = 3$), the expected log-likelihood at the "true" values of the parameters 0.5, 0.5, 0.9, where it attains its maximum. We then "constrain" σ_α^2 at different values between the true value (0.5) and 0, and for each of them we maximize the expected log-likelihood with respect to the other two parameters. In Table 2 we display the absolute difference between each constrained maximum and the absolute maximum (that might be relevant for hypotheses testing), as well as the relative difference (that might be relevant when a poor computational accuracy does not guarantee to discriminate between the two). The constrained

σ_α^2	σ_e^2	ρ	$T = 20$ diff.	$T = 10$ diff.	$T = 5$ diff.	$T = 3$ diff.
0.4	near 0.6	near 0.91	$N \times .00325$	$N \times .00106$	$N \times .00021$	$N \times .00004$
0.2	near 0.8	near 0.93	$N \times .02339$	$N \times .00590$	$N \times .00107$	$N \times .00018$
0.0	1.0	0.95	$N \times .04403$	$N \times .01061$	$N \times .00191$	$N \times .00033$
			%diff.	%diff.	%diff.	%diff.
0.4	near 0.6	near 0.91	0.03%	0.02%	0.01%	0.004%
0.2	near 0.8	near 0.93	0.20%	0.11%	0.05%	0.02%
0.0	1.0	0.95	0.40%	0.19%	0.10%	0.04%

Table 2: Constrained maximization at σ_α^2 *versus* absolute maximum at the true values $\psi_\alpha^2 = 0.5, \psi_e^2 = 0.5, \phi = 0.90$

maxima for σ_α^2 fixed at 0.4 or 0.2 are at values of the other two parameters always near the values displayed in the tables, when T varies from 20 to 3. However, when σ_α^2 is fixed at zero, always the constrained maximum is attained for a σ_e^2 equal to the sum of the two "true" variances, and ρ at a weighted average between the "true" value and 1.0 (thus, respectively, 1.0 and 0.95 in the tables).

Observing each case in greater detail, one would find that, varying σ_α^2 slowly from the "true" value to zero, the constrained maximum decreases monotonically:

faster, if T is large, very very slowly if T is small. The difference between the maximum and the value at zero is only $N \times 0.00033$ when $T = 3$, thus one would need a value of N of several thousands to discriminate between the two maxima with some kind of criterion (e.g. likelihood ratio). Moreover, in relative terms, the two maxima differ only by 0.04%: an optimization algorithm must be very accurately applied to be able to discriminate between the two. And what just described is a comparison at the two extremes of the interval for σ^2_α; differences with respect to intermediate values would be even more difficult to appreciate (0.004% between the constrained maxima at $\sigma^2_\alpha = 0.5$ and $\sigma^2_\alpha = 0.4$).

A look at first and second derivatives of the expected log-likelihood (formulas in Appendix B) shows that, starting from three first derivatives (expected score) equal zero at the maximum, the derivative with respect to σ^2_α is no more zero at the constrained maxima, but nevertheless it is very small even when departing far away from the true value, if T is small. For example, still in the case of the Table 2, when T has the smallest value ($T = 3$), the norm of the (average) expected score (sum of the squared first derivatives) quickly grows from zero to a value 5×10^{-7} as soon as we constrain σ^2_α to a value slightly smaller than the "true" 0.5; then this norm varies very little (say between 5×10^{-7} and 3×10^{-7}) when σ^2_α is constrained to values smaller and smaller, till $\sigma^2_\alpha = 0$. A stopping rule of an optimization algorithm based on such a norm might easily fail and stop at a false maximum.[10]

Second order derivatives could be helpful at this point. The three eigenvalues of the Hessian matrix have the correct sign (negative) when derivatives are computed at the "true" value (or absolute maximum). But one of them is very small in absolute value. Still for the example above, still for the most critical case of $T = 3$, the smallest (in absolute value) eigenvalue is -0.0052 (the largest is -103.8). When σ^2_α is constrained to zero, two eigenvalues remain negative, while the smallest (in absolute value) is positive ($+0.00033$). Thus, if the numbers were "exact", we might conclude that $\sigma^2_\alpha = 0$ does not provide a maximum, but something close to a saddle point. But could we really trust in the change of sign of the smallest eigenvalue from -0.0052 to $+0.00033$ when a matrix is so ill conditioned? Maybe we can trust because all derivatives, till second order, are supposed to be very

[10]Let's play even further with numbers and joke on a problem that might be serious! For the same values of the variances just considered in the example, but with an autoregression parameter 0.99, if the time dimension is very small ($T = 3$), the expected log-likelihoods at the "true" parameter values 0.5, 0.5, 0.99 and false values 0.0, 1.0, 0.995 differ approximately by 0.0001%. Between the "true" parameter values and the false values 0.4, 0.6, 0.99167, the difference is about 0.00001%; in absolute value, the difference is $N \times 3.5 \times 10^{-7}$; discrimination between "true" and false maxima by means of some criterion (e.g. likelihood ratio) might require a panel with some 10^7 individuals! In none of the points the norm (sum of the squared elements of the average expected score) is greater than 3×10^{-11}.

accurate, having been computed analytically and without using any "numerical" routine for matrix inversion, thanks to the use of the Woodbury (or Sherman-Morrison) matrix inversion lemma; for sure, there would be no warranty of this kind if derivatives were computed numerically, as it usually happens. Notice, en passing, that the smallest eigenvalue would remain with the correct negative sign till some distance from the absolute maximum: at $\sigma_\alpha^2 = 0.3$ the norm is about 5×10^{-7} and all eigenvalues are negative. Easily, an "ordinary" stopping rule might incorrectly find this as a satisfactory maximum; only some "pathologically tight" convergence criteria, as we have applied in our computations, can detect the existence of the problem, avoiding the serious risk of misleading false maxima.

A conclusion of this section seems therefore as follows: the model is identified, but the identification rules are very close to be "numerically" violated for small values of T. When passing from "expected log-likelihoods" (8) to log-likelihoods computed from a sample (7), "swaps" might easily occur. It will be enough for the sample log-likelihood around the "true" parameter values to be just a bit smaller than its expectation, and/or sample log-likelihood computed at parameter values far away from the "true" to be just a bit higher, and maximum likelihood estimates would fall in a wrong place. Inverting the expected Hessian at the true parameter values provides the asymptotic variance-covariance matrix of the maximum likelihood parameters. A large number of corner solutions is accompanied by a large value of the (asymptotic) standard deviation of the maximum likelihood estimated $\hat{\sigma}_\alpha^2$.

The consequences on estimation could be severe, like those evidenced in Section 3.

5 Empirical applications

In this Section, we will use the method of maximum likelihood to estimate models with random effect and autocorrelation in the idiosyncratic component. We are aware of the fact that in the case of linear models other "robust" options are available for the autocorrelation problem. So we also provide an empirical application on a non-linear model, where maximum likelihood is a common estimation technique.

5.1 A Cobb-Douglas production function

First, we focus on the data analyzed by Baltagi and Pinnoi (1995) [6] (see also Munnel, 1990 [29] and Holtz-Eakin, 1994 [19]). The research question that inspired the three papers aimed at capturing the effect of public infrastructures on the performance of the private economy. An "augmented" Cobb-Douglas production function is considered, where public capital stock is included among the regressors along with private capital and employment.

The panel dataset consists of 48 US states observed over the period 1970–1986 ($T = 17$). Data include information about gross state product Y, labor input L and private capital stock K, as well as public capital stock PK.[11] Information about state unemployment rate $Unemp$ is also provided, and it will be included in the regression to capture the effect of the business cycle at the state level.[12]

The analysis is limited to "Model 1" in Baltagi and Pinnoi (1995) [6], specified as:[13]

$$y_{it} = \beta_0 + \beta_1 l_{it} + \beta_2 k_{it} + \beta_3 Unemp_{it} + \nu_{it} \tag{9}$$

where small caps indicates log of the variables and $\nu_{it} = \alpha_i + e_{it}$ and we allow for serial correlation in e_{it}.

Results are reported in Table 3.[14]

Column 1 reports the estimated coefficients by means of the FGLS estimator (standard random effect approach). At the bottom of the Table, we report the Hausman statistic comparing the random effects and fixed effects specification of the model. The statistic does not allow to reject the null of no correlation between the individual effect and the variables included on the right hand side of the model (p-value = .4733). According to the estimates of the variances, 84% of the variance in the error term is due to the individual effect α_i.

In order to understand whether the "standard" random effect specification is suited to the data, we report three tests for the autocorrelation of the residuals,

[11] The original dataset also reports information about the components of public capital: highways and streets, water and sewer facilities, and other public buildings and structures.

[12] Please refer to Baltagi and Pinnoi (1995) [6] for details on data and measurement issues.

[13] Baltagi and Pinnoi (1995) [6] consider three different models: "Model 1" considered here, and two different "augmented" Cobb-Douglas production functions, where output is a function of (i) private inputs (capital and labor) and public capital stock; (ii) private inputs (capital and labor) and the components of public capital stock (highways and streets, water and sewer facilities, and other public buildings and structures). We do not report results on the other models for two main reasons. First, when public capital is included in the regression, the Hausman test comparing the random effects and fixed effects specifications of the model (Hausman, 1978 [18]), would lead us to reject the random effect specification (at the 5% level of significance). Second, Baltagi and Pinnoi (1995) [6] points to the possibility of measurement errors in public infrastructure, leading to correlation between PK and the idiosyncratic error component. The authors solve this issue by using an instrumental variable (IV) approach. The interest in this paper is about the random effect specification of the model, we therefore focus on the model where only private inputs are considered in the analysis. The IV estimator is also considered by Hotz-Eakin (1994) [19] who raises the issue of simultaneous determination of output, capital and labor. We will not pursue this issue further in this paper and assume that the variables are strictly exogenous.

[14] Standard errors are deliberately omitted from the Table, as interest is in the estimation of the two variance components and of the autocorrelation parameter.

	RE FGLS	RE+AR(1) MLE
l_{it}	.7340	.9407
k_{it}	.3094	.0835
$Unemp_{it}$	-.0061	-.0043
$constant$	2.160	3.102
σ^2_α	.0077	.1e-16
σ^2_e	.0015	.0200
ρ	–	.9881
Hausman stat.	2.51	
Bhargava *et al.* stat.	.3870	
Baltagi-Wu statistic	.6480	
Wooldridge statistic	128.9	

Table 3: Maximum Likelihood Estimator

namely the Bhargava *et al.* (1982) [9] modified Durbin-Watson statistics, the Baltagi and Wu (1999) [7] test, and the Wooldridge (2002) [35, §10.6] test.[15]

The results show that the null of lack of autocorrelation has to be rejected. The Bhargava *et al.* statistic and the Baltagi-Wu statistic are generalizations of the Durbin-Watson statistic to panel data. Under the null hypothesis of no correlation in the idiosyncratic error component, the value of the test is expected to be around 2. In our case the value is close to zero, suggesting the presence of autocorrelation in the idiosyncratic error component. On the contrary the Wooldridge statistic has a F-distribution. The p-value of the tests lead us to reject the null of no autocorrelation.

Therefore we propose the application of a random effect model with autocorrelation in the idiosyncratic component, and the method of maximum likelihood is applied for the estimation of the parameters.[16]

As a non-negativity constraint is imposed when estimating the variances by MLE, the estimated value of σ^2_α is practically zero. Deeper investigation in the value that maximizes the likelihood function shows us that the score function with respect to σ^2_α is negative (while it is zero with respect to σ^2_e and ρ), i.e.

[15]Computations of the tests are performed using STATA 9. The Wooldridge (2002) [35, §10.6] test is computed by using the `xtserial` command (Drukker, 2003 [13]).

[16]As already said, a more robust option than MLE is available for estimation in linear model (as it is the case in this example). Nonetheless, the paper aims at pointing out a problem of the MLE in this context, largely employed for the estimation of non-linear models. However, the problem is not merely driven by the non-linearity of the model, rather (as shown in the previous section) is related to the shape of the log-likelihood function.

the likelihood function is increasing as σ_α^2 decreases (the unconstrained maximum would be on the negative segment of the real line). The Hessian matrix is negative definite.

5.2 Union membership

The second empirical application is based on the data used by Vella and Verbeek (1998) [33]. The sample is drawn from the National Longitudinal Survey (Youth Data) and comprises 545 full-time working males followed over the period 1980-1987. For additional details on the dataset, please refer to Vella and Verbeek (1998) [33].

We focus on the determinants of union membership. The dependent variable U_{it} is a 0/1 variable, indicating whether individual i joined union at time t. Vella and Verbeek (1998) [33] consider a dynamic random effect probit model, where U_{it-1} is included among the regressors. We take a different approach and consider a static model with AR(1) idiosyncratic noise:

$$U_{it}^* = x_{it}'\beta + \nu_{it}$$

with $U_{it} = 1$ if $U_{it}^* > 0$ (0 otherwise), $\nu_{it} = \alpha_i + e_{it}$ and $e_{it} = \rho e_{it-1} + w_{it}$ (with w_{it} white noise). A normal distribution is assumed both for α_i and e_{it}, and we normalize $\sigma_e^2 = 1$.

To obtain an estimate of the parameters, we use the method of simulated maximum likelihood. The Geweke-Hajivassiliou-Keane (GHK) simulator is used for the evaluation of normal probabilities.[17]

Results are reported in Table 4. A corner solution arises when a subset of the original sample is considered, namely Hispanics and Afro-Americans (H-AA, $N = 148$), and we also restrict the time window and consider $T = 5$.

Note that when σ_α^2 is estimated to be zero, the estimated ρ is biased towards one. Assuming as "good" the estimated values when all time periods are considered for the selected subset of individuals (in column II), we could predict the value of the autocorrelation coefficient that will be estimated when a corner solution arises. As discussed in Section 4, it would be a weighted average between 1 and the "good" value: this gives a value of the autocorrelation coefficient equal to .7076, close to values which are in fact estimated in column (III) and (IV).

[17]See e.g. Hajivassiliou and McFadden, 1998 [16] for details. The GHK simulator is based on the algorithm by Moeltner and Layton (2002) [28] and Hajivassiliou (1997) [15].

	(I)	(II)	(III)	(IV)
Sample	All obs.	H-AA only	H-AA only	H-AA only
Time window	1980–87	1980–87	1980–84	1980–84
LogExper	-.0885	-.2295	.5020	.0490
School	-.0151	-.0246	.0806	.0553
Mar	.1725	.1715	-.0470	-.0083
Black	.8142	–	–	–
Hisp	.4076	.0313	.0269	-.0552
Rur	.0149	-.0715	-.3009	-.0899
Hlth	-.0098	.0545	.6346	.3006
NE	.0972	.2206	.0538	-.0394
S	.0032	.0704	.1090	-.1349
NC	.1946	1.100	.7879	.6987
Constant	-.8215	.0076	-1.886	-1.012
Industry dummies	yes	yes	yes	yes
Occupational dummies	no	no	no	yes
Time dummies	yes	yes	yes	yes
σ_α^2	1.560	1.045	.9e-22	.0919
ρ	.5942	.4021	.7353	.6888

Table 4: The determinants of union membership, RE probit with AR(1) idiosyncratic noise

6 Summary

This paper has shown a severe problem that arises in the case of Gaussian maximum likelihood estimation of the random effect panel data model with AR(1) idiosyncratic noise.

By means of a wide Monte Carlo experiment, we show that corner solutions (zero estimate of the variance of the individual random effect) are encountered in a large fraction of the experiments. As a consequence, individual heterogeneity (even strong) is not properly accounted for by the estimation results. When this is the case, the other parameters of the model shift, and the distributions are no more scattered only around their "true" values, rather they are bi-modal. The problem is less relevant when T is large (but we show an empirical application where the problem occurs). When T is small, the problem would disappear only if the sample involves an (unreasonably) enormous number of units. For the N and T dimensions used in practical applications, the problem is relevant.

Some explanations was found considering that the expected log-likelihood is nearly flat along a particular direction.

Two empirical applications show that the problem is likely to appear in practice, both in the linear and non-linear context.

Acknowledgments

We gratefully acknowledge comments and suggestions from the Editor and three anonymous Referees. Earlier versions of the paper were presented at the Third Italian Congress of Econometrics and Empirical Economics, Ancona, January 30-31, 2009; and at the 15th Conference on Panel Data, Bonn, July 3-5, 2009 with title "Poor identification and estimation problems in panel data models with random effects and autocorrelated errors". We also gratefully acknowledge comments and suggestions from E. Battistin, G. Fiorentini, D. Lubian, F. Mealli, G. Millo, P. Paruolo, T. Proietti, C. Rampichini, A. Rossi, A. Sembenelli, E. Sentana, and M. Wagner, but we retain full responsibility for the contents of this paper.

A Matrix Inversion

Given the particular structure of matrix B in equation (7), the computation of the determinant and inverse of matrix B greatly benefits from the use of Woodbury (or Sherman-Morrison) matrix inversion lemma (e.g., Householder 1964 [20, §5.1]), both in terms of computation time as well as (more important) computation accuracy. Its application (tedious but straightforward) gives

$$|B| = \left\{ \left(1-\rho^2\right)^{T-1} + \frac{\sigma_\alpha^2 \left(1-\rho^2\right)^{T-1}}{\sigma_e^2(1+\rho)} \left[T-(T-2)\rho\right] \right\} \left(\sigma_e^2\right)^T \tag{10}$$

$$B^{-1} = \frac{1}{\sigma_e^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho & 0 & ... & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & ... & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & ... & -\rho & 1 \end{bmatrix}$$

$$- \frac{\sigma_\alpha^2}{\sigma_e^2(1+\rho)\left\{\sigma_e^2(1+\rho) + [T-(T-2)\rho]\sigma_\alpha^2\right\}}$$

$$\begin{bmatrix} 1 & (1-\rho) & (1-\rho) & (1-\rho) & ... & (1-\rho) & 1 \\ (1-\rho) & (1-\rho)^2 & (1-\rho)^2 & (1-\rho)^2 & ... & (1-\rho)^2 & (1-\rho) \\ (1-\rho) & (1-\rho)^2 & (1-\rho)^2 & (1-\rho)^2 & ... & (1-\rho)^2 & (1-\rho) \\ ... & ... & ... & ... & ... & ... & ... \\ (1-\rho) & (1-\rho)^2 & (1-\rho)^2 & (1-\rho)^2 & ... & (1-\rho)^2 & (1-\rho) \\ (1-\rho) & (1-\rho)^2 & (1-\rho)^2 & (1-\rho)^2 & ... & (1-\rho)^2 & (1-\rho) \\ 1 & (1-\rho) & (1-\rho) & (1-\rho) & ... & (1-\rho) & 1 \end{bmatrix} \tag{11}$$

Notice that there is no matrix inversion in the last two equations, so numerical correctness does not depend on routines that perform numerical inversions.

B Expected score and Hessian

Recall that

$$\frac{\partial \log ||B||}{\partial B} = B^{-1} \qquad\qquad \frac{\partial B^{-1}}{\partial \theta_j} = -B^{-1}\frac{\partial B}{\partial \theta_j}B^{-1} \tag{12}$$

(where θ_j is one of the three elements of the parameters vector $\theta = [\sigma_\alpha^2, \sigma_e^2, \rho]')$. We then observe that derivative of the Toepliz matrix A with respect to ρ ($D = \partial A/\partial \rho$) is the matrix whose h,k-th element is $|h-k|\rho^{|h-k|-1}$.

The expected value of the score can thus be written in closed form:

$$E\left[\frac{\partial \log L(\nu,\theta)}{\partial \theta}\right] = \frac{N}{2} \begin{bmatrix} -\iota'(B^{-1})\iota & + & \iota'(B^{-1}RB^{-1})\iota \\ -tr(B^{-1}A) & + & tr(B^{-1}AB^{-1}R) \\ -tr(B^{-1}D)\sigma_e^2 & + & tr(B^{-1}DB^{-1}R)\sigma_e^2 \end{bmatrix} \tag{13}$$

We now indicate with F the $T \times T$ matrix $F = \partial D / \partial \rho = \partial^2 A / \partial \rho^2$, whose h, k-th element is $|h-k|(|h-k|-1)\rho^{|h-k|-2}$. Some tedious but straightforward algebra allows to compute in closed form the second derivatives of the expected log-likelihood (the 3×3 expected Hessian matrix)

$$H = E\left[\frac{\partial^2 \log L(\nu, \theta)}{\partial \theta \partial \theta'}\right] = \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} \\ h_{2,1} & h_{2,2} & h_{2,3} \\ h_{3,1} & h_{3,2} & h_{3,3} \end{bmatrix} \tag{14}$$

where

$$h_{1,1} = E\left[\frac{\partial^2 \log L(\nu, \theta)}{\partial (\sigma_\alpha^2)^2}\right] = \frac{N}{2}\left[\iota' B^{-1} \iota \iota' B^{-1} \iota - 2\iota' B^{-1} R B^{-1} \iota \iota' B^{-1} \iota\right]$$

$$h_{2,1} = h_{1,2} = E\left[\frac{\partial^2 \log L(\nu, \theta)}{\partial \sigma_\alpha^2 \partial \sigma_e^2}\right] = \frac{N}{2} tr\left[B^{-1} \iota \iota' B^{-1} A - 2B^{-1} \iota \iota' B^{-1} A B^{-1} R\right]$$

$$h_{3,1} = h_{1,3} = E\left[\frac{\partial^2 \log L(\nu, \theta)}{\partial \sigma_\alpha^2 \partial \rho}\right] = \frac{N}{2} tr\left[B^{-1} \iota \iota' B^{-1} D - 2B^{-1} \iota \iota' B^{-1} D B^{-1} R\right] \sigma_e^2$$

$$h_{2,2} = E\left[\frac{\partial^2 \log L(\nu, \theta)}{\partial (\sigma_e^2)^2}\right] = \frac{N}{2} tr\left[B^{-1} A B^{-1} A - 2B^{-1} A B^{-1} A B^{-1} R\right]$$

$$h_{3,2} = h_{2,3} = E\left[\frac{\partial^2 \log L(\nu, \theta)}{\partial \sigma_e^2 \partial \rho}\right]$$

$$= \frac{N}{2} tr\left[B^{-1} D B^{-1} A - 2B^{-1} D B^{-1} A B^{-1} R\right] \sigma_e^2 + tr\left[B^{-1} D B^{-1} R - B^{-1} D\right]$$

$$h_{3,3} = E\left[\frac{\partial^2 \log L(\nu, \theta)}{\partial \rho^2}\right]$$

$$= \frac{N}{2} tr\left[B^{-1} D B^{-1} D \sigma_e^2 - 2B^{-1} D B^{-1} D B^{-1} R \sigma_e^2 + B^{-1} F B^{-1} R - B^{-1} F\right] \sigma_e^2$$

References

[1] Balestra, P. and Nerlove, M.: 1966, Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas, *Econometrica* **34**, 585–612.

[2] Baltagi, B.H.: 1986, Pooling Under Misspecification: Some Monte Carlo Evidence on the Kmenta and the Error Components Techniques, *Econometric Theory* **2**, 429–440.

[3] Baltagi, B.H. and Li, Q.: 1991, A Transformation That Will Circumvent the Problem of Autocorrelation in an Error-Component Model, *Journal of Econometrics* **48**(3), 385–393.

[4] Baltagi, B.H. and Li, Q.: 1995, Testing AR(1) against MA(1) Disturbances in an Error Component Model, *Journal of Econometrics* **68**(1), 133–151.

[5] Baltagi, B.H., Matyas, L. and Sevestre, P.: 2008, Error Component Models, in Matyas, P. and Sevestre, P. (eds.), *The Econometrics of Panel Data*, Berlin: Springer-Verlag.

[6] Baltagi, B.H. and Pinnoi, N.: 1995, Public Capital Stock and State Productivity Growth: Further Evidence from an Error Component Model, *Empirical Economics* 20, 351–359.

[7] Baltagi, B.H. and Wu, P.X.: 1999, Unequally Spaced Panel Data Regressions with AR(1) Disturbances, *Econometric Theory* 15, 814–823.

[8] Berry, S., Gottschalk, P. and Wissoker, D.: 1988, An Error Components Model of the Impact of Plant Closing on Earnings., *Review of Economics & Statistics* **70**(4), 701–07.

[9] Bhargava, A., Franzini, L. and Narendranathan, W.: 1982, Serial Correlation and the Fixed Effects Model, *The Review of Economic Studies* **49**(4), 533–549.

[10] Calzolari, G. and Magazzini, L.: 2008, Estimating Tobit Models for Panel Data with Autocorrelated Errors, *Paper presented at the European Meeting of the Econometric Society*, Milan, 27-31 August, mimeo.

[11] Contoyannis, P., Jones, A.M. and Rice, N.: 2004, Simulation-based Inference in Dynamic Panel Probit Models: An Application to Health, *Empirical Economics* **29**, 49–77.

[12] David, M.: 1971, Lifetime Income Variability and Income Profiles, *Proceedings of the Annual Meeting of the American Statistical Association*, pp. 285–92.

[13] Drukker, D.M.: 2003, Testing for Serial Correlation in Linear Panel-Data Models, *The Stata Journal* 3(2), 168–177.

[14] Hajivassiliou, V.A.: 1994, A Simulation Estimation Analysis of the External Debt Crises of Developing Countries, *Journal of Applied Econometrics* **9**, 109–131.

[15] Hajivassiliou, V.A.: 1997, Some Practical Issue in Maximum Simulated Likelihood, Working Paper, London School of Economics.

[16] Hajivassiliou, V.A. and McFadden, D.L.: 1998, The Method of Simulated Scores for the Estimation of LDV Models, *Econometrica* 66(4), 863–896.

[17] Hause, J.: 1980, The Fine Structure of Earnings and the On-the-Job Training Hypothesis, *Econometrica* **48**(4), 1013–1029.

[18] Hausman, J.A.: 1978, Specification Tests in Econometrics, *Econometrica* **46**(6), 1251–1271.

[19] Hotz-Eakin, D.: 1994, Public Sector Capital and The Productivity Puzzle, *Review of Economics and Statistics* **76**, 12–21.

[20] Householder, A.S.: 1964, *The Theory of Matrices In Numerical Analysis*, New York: Dover Publications, Inc.

[21] Inkman, J.: 2000, Misspecified Heteroskedasticity in the Panel Probit Model: A Small Sample Comparison of GMM and SML Estimators, *Journal of Econometrics* **97**, 227–259.

[22] Karlsson, S. and Skoglund, J.: 2004, Maximum-likelihood Based Inference in the Two-way Random Effects Model with Serially Correlated Time Effects, *Empirical Economics* **29**(1), 79–88.

[23] Keane, M.P.: 1994, A Computationally Practical Simulation Estimator for Panel Data, *Econometrica* **62**(1), 95–116.

[24] Kmenta, J.: 1986, *Elements of Econometrics* (2nd ed.), New York: Macmillan Publishing Company.

[25] Lillard, L. and Weiss, Y.: 1979, Components of Variation in Panel Earnings Data: American Scientists 1960-1970, *Econometrica* **47**(2), 437–54.

[26] Lillard, L. and Willis, R.: 1978, Dynamic Aspects of Earning Mobility, *Econometrica* **46**(5), 985–1012.

[27] MaCurdy, T.: 1982, The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis, *Journal of Econometrics* **18**(1), 83–114.

[28] Moeltner, K. and Layton, D.F.: 2002, A Censored Random Coefficients Model for Pooled Survey Data With Application to the Estimation of Power Outage Costs, *The Review of Economics and Statistics* 84(3), 552–561.

[29] Munnel, A.: 1990, How Does Public Infrastructure Affect Regional Economic Performance?, *New England Economic Review*, 11-32.

[30] Newey, W.K. and McFadden, D.: 1994, Large Sample and Hypothesis Testing, in Engle, R.F. and McFadden, D.L. (eds.), *Handbook of Econometrics*, vol. 4, 2111–2245.

[31] Roberts, M.J. and Tybout, J.R.: 1997, The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs, *American Economic Review* 87(4), 545–564.

[32] Stern, S.: 1997, Simulation-Based Estimation, *Journal of Economic Literature* **35**(4), 2006–2039.

[33] Vella, F. and Verbeek, M.: 1998, Whose Wages Do Union Raise? A Dynamic Model of Unionism and Wage Rate Determination For Young Men, *Journal of Applied Econometrics* 13(2), 163–183.

[34] Wansbeek, T.: 1992, Transformation for Panel Data when the Disturbances are Autocorrelated, *Structural Change and Economic Dynamics* **3**(2), 375–384.

[35] Wooldridge, J.: 2002, *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.