View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Specification and estimation of random effects models with serial correlation of general form

Jimmy Skoglund* and Sune Karlsson†
Department of Economic statistics,
Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden

Stockholm School of Economics
SSE/EFI Working Paper Series in Economics and Finance
No. 433

February 2001

Abstract

This paper is concerned with maximum likelihood based inference in random effects models with serial correlation. Allowing for individual effects we introduce serial correlation of general form in the time effects as well as the idiosyncratic errors. A straightforward maximum likelihood estimator is derived and a coherent model selection strategy is suggested for determining the orders of serial correlation as well as the importance of time and individual effects. The methods are applied to the estimation of a production function for the Japanese chemical industry using a sample of 72 firms observed during 1968-1987. Empirically, our focus is on measuring the returns to scale and technical change for the industry.

Keywords: Panel data; serial correlation; random effects.
JEL codes: C12; C13; C23

*Jimmy.Skoglund@hhs.se
†Sune.Karlsson@hhs.se

1 Introduction

In the econometric analysis of panel data unobserved heterogeneity is typically handled by including fixed or random effects in the model. In the fixed effects model the individual and time effects are assumed to be fixed parameters to be estimated and in general correlated with the regressors. In this case the differences across individuals and time can be captured by differences in the constant term. In the random effects approach the individual and time effects are assumed to be stochastic and uncorrelated with the regressors. The random effects formulation allows for the inclusion of time-invariant or individual-invariant explanatory variables and for the number of parameters to be reduced to only two, the mean and the variance. While the choice of random effects has the advantage of providing many degrees of freedom, it also complicates the treatment of two estimation problems, that is heteroskedasticity and serial correlation. Mazodier and Trognon (1978) generalized the one-way model with individual effects to the case where the individual effects are heteroskedastic. An alternative heteroskedastic model keeps the individual effects homoskedastic while allowing for heteroskedasticity in idiosyncratic error terms or allows them both to be heteroskedastic, see Randolph (1988).

This paper is concerned with the second problem, namely, serial correlation. As in the heteroskedastic case serial correlation can be introduced in two distinct ways. First, through serially correlated idiosyncratic errors and secondly through serially correlated time effects. Serial correlation in the idiosyncratic errors introduces a time series type of correlation at the individual level whereas serial correlation in the time effects introduces the empirically plausible phenomena that some of the factors driving the unobserved time specific heterogeneity are serially correlated. Examples of such factors include business cycles, oil price shocks and economic policies that persist during several time periods. Lillard and Willis (1978), Baltagi and Li (1991, 1994), King (1986), Magnus and Woodland (1988), Karlsson and Skoglund (2000) and Skoglund and Karlsson (2001), among many others, contain further discussion of serially correlated period effects and/or serially correlated idiosyncratic errors.

The paper is organized as follows. In section 2 we present the model and discuss its relation to models previously suggested in the literature. Section 3 is concerned with estimation and inference issues. A maximum likelihood estimator, feasible in the presence of a large individuals dimension, is derived and the estimation problem is discussed. In this section we also consider model selection procedures for determining the orders of serial correlation as well as the significance of time and individual effects. Section 4 contains an application of the proposed model and the associated model selection procedures to the estimation of a production function for the Japanese chemical industry. Section 5 concludes.

2 The model

Consider the panel data regression model

$$y_{it} = \mathbf{z}'_{it}\boldsymbol{\delta} + \varepsilon_{it} \quad i = 1, \ldots, N; \; t = 1, \ldots, T \tag{1}$$

where $\boldsymbol{\delta}$ is a $k \times 1$ vector of regression coefficients including the intercept and in addition $\mathbf{z}_{it}$ may contain time invariant or individual invariant explanatory variables. The error term, ε_{it} follows a two-way random effects model, see Baltagi (1995, ch. 3),

$$\varepsilon_{it} = \mu_i + \lambda_t + v_{it} \tag{2}$$

where $\mu_i \sim iid\left(0, \sigma_\mu^2\right)$ denotes the unobservable individual effect, λ_t denotes the unobservable time effect and v_{it} is the idiosyncratic error term. Following Revankar (1979), Karlsson and Skoglund (2000) introduce serial correlation in the time effects via an AR(1),

$$\lambda_t = \rho_\lambda \lambda_{t-1} + u_t, \tag{3}$$

or MA(1),

$$\lambda_t = u_t + \theta_\lambda u_{t-1}$$

process for λ_t. In addition one can not rule out the possibility that the idiosyncratic errors are serially correlated as well. In an asymptotic analysis Skoglund and Karlsson (2001) introduce serial correlation in the time effects and idiosyncratic errors via an AR(1) process for λ_t (3), and v_{it}

$$v_{it} = \rho_v v_{it-1} + e_{it} \tag{4}$$

In practice serial correlation need however not be restricted to AR(1) processes, or MA(1) processes for that matter. Viable alternatives include the AR(2) or MA(2) or even the general ARMA(p, q) specification. Consequently we adopt a general approach, allowing both the time effects and the idiosyncratic errors to have an arbitrary serial correlation form.

The present model does away with an arbitrary restriction on the time independence of period effects and idiosyncratic errors which is commonly encountered in applications. But more importantly it does so in the framework of the two-way model (1,2). This is in contrast to previous empirical and theoretical work on random effects models with serial correlation which focus on one-way models. That is, the one-way model with individual effects,

$$\varepsilon_{it} = \mu_i + v_{it} \tag{5}$$

and serially correlated v_{it}, or the one-way model with time effects,

$$\varepsilon_{it} = \lambda_t + v_{it} \tag{6}$$

and serially correlated λ_t and/or v_{it}. Lillard and Willis (1978) consider a first-order autoregression in the one-way model with individual effects (5) whereas Baltagi and Li (1991, 1994) consider AR(2), AR(4) or MA(q) processes as well and Galbraith and Zinde-Walsh (1995) allow for general ARMA(p,q) disturbances in a semi-parametric framework. King (1986) consider serially correlated time effects and independent idiosyncratic errors in (6) whereas Magnus and Woodland (1988) consider both serially correlated time effects and serially correlated idiosyncratic errors.

The two-way model with serially correlated λ_t and v_{it} nests both of these models since we have neither imposed the auxiliary assumption of no time effects in (5) nor the auxiliary assumption of no individual effects in (6). Indeed the existence of such effects should be part of a hypothesis to be tested and not an assumption.

3 Estimation and model specification

3.1 Likelihood

In matrix form we can write the model (1,2) as

$$\begin{aligned} \mathbf{y} &= \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \mathbf{Z}_\mu \boldsymbol{\mu} + \mathbf{Z}_\lambda \boldsymbol{\lambda} + \boldsymbol{\nu} \end{aligned}$$

where $\mathbf{Z}_\mu = (I_N \otimes \boldsymbol{\iota}_T)$, $\mathbf{Z}_\lambda = (\boldsymbol{\iota}_N \otimes \mathbf{I}_T)$, $\boldsymbol{\mu}' = (\mu_{1,}, ..., \mu_N)$, $\boldsymbol{\lambda}' = (\lambda_1, ..., \lambda_T)$ and $\boldsymbol{\iota}_N$ is a vector of ones of dimension N.

Under the assumption that μ_i, λ_t and v_{it} are independent of each other and the explanatory variables we obtain the covariance matrix of the combined error term as

$$\begin{aligned} \boldsymbol{\Sigma} &= E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{Z}_\mu E(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{Z}_\mu' + \mathbf{Z}_\lambda E(\boldsymbol{\lambda}\boldsymbol{\lambda}')\mathbf{Z}_\lambda' + E(\boldsymbol{\nu}\boldsymbol{\nu}') \\ &= \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_u^2(\mathbf{J}_N \otimes \boldsymbol{\Psi}_\lambda) + \sigma_e^2(\mathbf{I}_N \otimes \boldsymbol{\Psi}_v) \end{aligned}$$

where $\mathbf{J}_T = \boldsymbol{\iota}_T\boldsymbol{\iota}_T'$ a $T \times T$ matrix of ones and $\sigma_u^2\boldsymbol{\Psi}_\lambda$ is the covariance matrix of λ and $\sigma_e^2\boldsymbol{\Psi}_v$ is covariance matrix of v. Both $\sigma_u^2\boldsymbol{\Psi}_\lambda$ and $\sigma_e^2\boldsymbol{\Psi}_v$ may be the covariance matrix of any stationary and strictly invertible ARMA(p,q) process.

Maximum likelihood estimation requires a specific distributional choice and throughout we will maintain the assumption that $\mu_i \sim N\left(0,\sigma_\mu^2\right)$, $u_t \sim N\left(0,\sigma_u^2\right)$ and $e_{it} \sim N\left(0,\sigma_e^2\right)$. However, maximum likelihood also requires evaluation of the inverse and determinant of the $NT \times NT$ matrix $\mathbf{\Sigma}$. Direct inversion of $\mathbf{\Sigma}$ is clearly impractical even for panels of moderate size and the usual spectral decomposition "tricks" employed in the panel data literature are not directly applicable here. Our method of solution is to reduce the amount of numerical computation necessary. As in Karlsson and Skoglund (2000) and Skoglund and Karlsson (2001) this is accomplished by using elementary results on inverses and determinants involving sums (Dhrymes (1984, p. 39-40)). This yields,

$$\begin{aligned}\mathbf{\Sigma}^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\boldsymbol{\iota}_N \otimes \mathbf{I}_T)[\sigma_u^{-2}\mathbf{\Psi}_\lambda^{-1} + N\mathbf{A}^*]^{-1}(\boldsymbol{\iota}_N^{'} \otimes \mathbf{I}_T)\mathbf{A}^{-1} \\ &= \mathbf{I}_N \otimes \mathbf{A}^* - (\boldsymbol{\iota}_N \otimes \mathbf{A}^*)[\sigma_u^{-2}\mathbf{\Psi}_\lambda^{-1} + N\mathbf{A}^*]^{-1}(\boldsymbol{\iota}_N^{'} \otimes \mathbf{A}^*) \\ &= \mathbf{I}_N \otimes \mathbf{A}^* - \sigma_u^2(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\left[\mathbf{I}_T + N\sigma_u^2\mathbf{\Psi}_\lambda\mathbf{A}^*\right]^{-1}\mathbf{\Psi}_\lambda(\boldsymbol{\iota}_N^{'} \otimes \mathbf{A}^*)\end{aligned} \tag{7}$$

and

$$|\mathbf{\Sigma}| = |\mathbf{A}^*|^{-N}\left|\mathbf{I}_T + N\sigma_u^2\mathbf{\Psi}_\lambda\mathbf{A}^*\right| \tag{8}$$

where $\mathbf{E}_N = \mathbf{I}_N - \overline{\mathbf{J}}_N$, $\overline{\mathbf{J}}_N = \mathbf{J}_N/N$, $\mathbf{A}^* = \left(\sigma_\mu^2\mathbf{J}_T + \sigma_e^2\mathbf{\Psi}_v\right)^{-1}$ and $\mathbf{J}_N = \boldsymbol{\iota}_N\boldsymbol{\iota}_N'$ is a $N \times N$ matrix of ones.

The present form of the inverse and determinant of $\mathbf{\Sigma}$ has reduced the amount of numerical computation to the $T \times T$ matrices $\mathbf{A}^*$ and $\mathbf{I}_T + N\sigma_u^2\mathbf{\Psi}_\lambda\mathbf{A}^*$. This is useful since in a typical panel the individuals dimension is large whereas the time dimension is small. In addition we can obtain further simplification in some special cases of interest. The matrix $\mathbf{A}^*$ is readily recognized as the T-dimensional part of the inverse variance matrix of the one-way model with individual effects and serially correlated v_{it}. Baltagi and Li (1991), extending the Wansbeek and Kapteyn (1982, 1983) "trick", show how to obtain a spectral decomposition of this matrix in case of an AR(1), AR(2) or AR(4) process for v_{it} and Baltagi and Li (1994) contain an extension to the MA(q) case. As pointed out in Baltagi and Li (1991) we can however obtain a spectral decomposition as long as there exists a simple known matrix $\mathbf{C}$ such that the transformation $(\mathbf{I}_N \otimes \mathbf{C})\boldsymbol{\nu}$ has mean zero and variance $\sigma_e^2\mathbf{I}_{NT}$. A leading special case is of course an AR(1) process for v_{it} (4) where $\mathbf{A}^* = \mathbf{C}'\left(\sigma_\alpha^{-2}\overline{\mathbf{J}}_T^\alpha + \sigma_e^{-2}\overline{\mathbf{E}}_T^\alpha\right)\mathbf{C}$ with $\mathbf{C}$ the Prais-Winsten transformation matrix for an AR(1) process, $\sigma_\alpha^2 = d^2\sigma_\mu^2\left(1-\rho_v\right)^2 + \sigma_e^2$, $\overline{\mathbf{J}}_T^\alpha = \boldsymbol{\iota}_T^\alpha\boldsymbol{\iota}_T^{\alpha\prime}/d^2$, $\boldsymbol{\iota}_T^{\alpha\prime} = \left(\alpha, \boldsymbol{\iota}_{T-1}'\right)$ and $\overline{\mathbf{E}}_T^\alpha = \mathbf{I}_T - \overline{\mathbf{J}}_T^\alpha$ with $d^2 = \boldsymbol{\iota}_T^{\alpha\prime}\boldsymbol{\iota}_T^\alpha = \alpha^2 + (T-1)$, $\alpha = \sqrt{\left(1+\rho_v\right)/\left(1-\rho_v\right)}$.

The references given above contain further details about the transformation and an extension to more general time series processes.

Finding a spectral decomposition of the matrix $\mathbf{I}_T + N\sigma_u^2 \mathbf{\Psi}_\lambda \mathbf{A}^*$ is however more difficult. In the special case of $\mathbf{\Psi}_\lambda = \mathbf{\Psi}_v$ we can use the method of Baltagi and Li (1991, 1994) since

$$\mathbf{I}_T + N\sigma_u^2 \mathbf{\Psi}_\lambda \mathbf{A}^* = \mathbf{L}\mathbf{A}^*$$

where $\mathbf{L} = \left(N\sigma_u^2 + \sigma_e^2\right) \mathbf{\Psi}_v + \sigma_\mu^2 \mathbf{J}_T$ is of the same form as the inverse of $\mathbf{A}^*$. This, of course, includes the standard two-way model where $\mathbf{\Psi}_\lambda = \mathbf{\Psi}_v = \mathbf{I}_T$ and for which the inverse and determinant of $\mathbf{\Sigma}$ reduces to

$$\begin{aligned} \mathbf{\Sigma}^{-1} &= \overline{\mathbf{E}}_N \otimes \left(\frac{1}{\sigma_v^2} \overline{\mathbf{E}}_T + \frac{1}{\sigma_1^2} \overline{\mathbf{J}}_T \right) + \overline{\mathbf{J}}_N \otimes \left(\frac{1}{\sigma_v^2 + N\sigma_\lambda^2} \overline{\mathbf{E}}_T + \frac{1}{\sigma_2^2} \overline{\mathbf{J}}_T \right) \\ |\mathbf{\Sigma}| &= \left(\sigma_v^2\right)^{NT} \left(\frac{\sigma_v^2}{\sigma_1^2} \right)^{-(N-1)} \left(\frac{\sigma_v^2}{\sigma_2^2 - T\sigma_\mu^2} \right)^{-(T-1)} \left(\frac{\sigma_v^2}{\sigma_2^2} \right)^{-1} \end{aligned}$$

with $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$, $\sigma_2^2 = \sigma_1^2 + N\sigma_\lambda^2$. In the case of serially correlated time effects and/or $\mathbf{\Psi}_\lambda \neq \mathbf{\Psi}_v$ the inverse of $\mathbf{I}_T + N\sigma_u^2 \mathbf{\Psi}_\lambda \mathbf{A}^*$ as well as its determinant can be computed numerically. For the modest time series dimensions common in panel data applications this is both speedy and accurate.

Using the results in (7) and (8) we can write the log-likelihood as

$$\begin{aligned} l(\theta) &= -\frac{TN}{2} \ln 2\pi + \frac{(N-1)}{2} \ln |\mathbf{A}^*| \\ &\quad -\frac{1}{2} \ln \left| \mathbf{I}_T + N\sigma_u^2 \mathbf{\Psi}_\lambda \mathbf{A}^* \right| - \frac{1}{2} \boldsymbol{\varepsilon}' \left(\mathbf{I}_N \otimes \mathbf{A}^* \right) \boldsymbol{\varepsilon} \\ &\quad + \frac{\sigma_u^2}{2} \boldsymbol{\varepsilon}' (\boldsymbol{\iota}_N \otimes \mathbf{A}^*) \left[\mathbf{I}_T + N\sigma_u^2 \mathbf{\Psi}_\lambda \mathbf{A}^* \right]^{-1} \mathbf{\Psi}_\lambda (\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \boldsymbol{\varepsilon} \end{aligned} \tag{9}$$

where $\theta = (\boldsymbol{\delta}', \boldsymbol{\gamma})$ and $\boldsymbol{\gamma}$ is the vector of covariance parameters including the serial correlation parameters of λ_t and v_{it}. When a spectral decomposition of $\mathbf{A}^*$ is available $\ln |\mathbf{A}^*|$ simplifies accordingly. For example, in the AR(1) case

$$\ln |\mathbf{A}^*| = -(T-1) \ln \sigma_e^2 + \ln \left(1 - \rho_v^2\right) - \ln \sigma_\alpha^2$$

In practice iterative methods are used to obtain the maximum likelihood estimate, say $\widehat{\theta}$, and these methods require us to supply the first derivatives of the log-likelihood as well. In our experience numerical derivatives perform poorly, especially if there is serial correlation in λ_t and/or v_{it}, leaving analytical derivatives the preferred choice. The score vector for the mean parameters, $\boldsymbol{\delta}$ is straightforward to obtain and following Hartley and Rao

(1967) or Hemmerle and Hartley (1973) the elements of the score vector for the variance parameters, $\boldsymbol{\gamma}$ are obtained as

$$\frac{\partial l(\theta)}{\partial \gamma_i} = -\frac{1}{2}\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i}) + \frac{1}{2}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}$$

Variance estimates can be based on either a numerical approximation to the hessian matrix or the information matrix. The elements of the information matrix are computed as (see Harville (1977))

$$\mathcal{I}_{\gamma_i\gamma_j} = \frac{1}{2}\,\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_i})\boldsymbol{\Sigma}^{-1}(\frac{\partial \boldsymbol{\Sigma}}{\partial \gamma_j})]$$

Note that if $\sigma_u^2 = 0$ in (9) it reduces to the log-likelihood of the one-way model with individual effects and serially correlated v_{it} whereas if $\sigma_\mu^2 = 0$, $\mathbf{A}^* = \sigma_e^{-2}\boldsymbol{\Psi}_v^{-1}$ and (9) reduces to the log-likelihood of the one-way model with serially correlated λ_t and v_{it}. For estimation purposes there are therefore no reason to employ strict inequality restrictions, $\left(\sigma_\mu^2 > 0, \sigma_u^2 > 0\right)$, weak inequalities, $\left(\sigma_\mu^2 \geq 0, \sigma_u^2 \geq 0\right)$, are sufficient.

3.2 Model selection

Model selection can be based on either hypothesis tests or model selection criteria, e.g., the AIC criterion of Akaike (1974) and the BIC criterion of Schwarz (1978), or possibly a mixture of the two approaches. This section is concerned with the hypothesis tests approach and our focus is on hypothesis tests on the variance parameters, $\boldsymbol{\gamma}$. More specifically, in the framework of the model (1,2) we propose simple methods and a straightforward strategy for determining the orders of serial correlation in the time effects and idiosyncratic errors as well as the significance of individual and time effects.

3.2.1 Determining the orders of serial correlation in λ_t and v_{it}

For obvious reasons a strategy for determining the orders of serial correlation in both λ_t and v_{it} might be expected to encounter serious difficulties. In particular, we may not know if the test for serial correlation in λ_t rejects the null due to misspecification of serial correlation in v_{it} and vice versa. A first step towards resolving could be to try to establish the presence or absence of a local robustness property, that the tests lack local asymptotic power against serial correlation in the other component. Considered alone such a local property would, however, be of rather limited value. This is because in practice misspecifications are global in nature, typically rendering the variance-covariance matrix estimator employed in the test-statistic inconsistent. A much more useful situation would emerge if we could establish that the test for serial correlation in λ_t or v_{it} that ignores the misspecification is in some sense equivalent to the test that takes the serial correlation into account and/or the test that employs a robust variance-covariance matrix estimator. Indeed this turns out to be the case here. More specifically, we have the following situation which we state for the LM (score) test although we expect similar results to hold for the other classical tests (Wald and LR) as well.

Property 1 *Denote by ξ_{LM_λ} the LM test for serial correlation in λ_t that takes into account the (global) serial correlation in v_{it} and let $\widetilde{\xi}_{LM_\lambda}$ be the corresponding LM test that fails to take into account the serial correlation in v_{it} but employs a robust variance-covariance matrix estimator. Finally, denote by $\xi^*_{LM_\lambda}$ the LM test that fails to take into account the serial correlation in v_{it} and does not employ a robust variance-covariance matrix estimator. Using similar notation for the corresponding LM tests for serial correlation in v_{it} i.e. ξ_{LM_v}, $\widetilde{\xi}_{LM_v}$ and $\xi^*_{LM_v}$ respectively we have*

$$\left|\xi_{LM_\lambda} - \widetilde{\xi}_{LM_\lambda}\right| \xrightarrow{p} 0 \text{ and } \left|\xi^*_{LM_\lambda} - \xi_{LM_\lambda}\right| \xrightarrow{p} 0 \text{ if } \left(\frac{\sqrt{T}}{N^{\frac{3}{2}}}\right) \to 0$$

and,

$$\left|\xi_{LM_v} - \widetilde{\xi}_{LM_v}\right| \xrightarrow{p} 0 \text{ and } \left|\xi^*_{LM_v} - \xi_{LM_v}\right| \xrightarrow{p} 0 \text{ if } \left(\frac{\sqrt{T}}{N}\right) \to 0$$

where the different rates of convergence are due to the different probabilistic orders of the serial correlation parameters of λ_t and v_{it} respectively.

A sketch of the proof of this property is given in appendix A. It is perhaps worth pointing out that the local result alluded to above, i.e. lack of local asymptotic power against serial correlation in the other component, holds as well. In addition this result does not require conditions on the relative rate of convergence of N and T. The present result is a global one. It has reduced the problem of determining the orders of serial correlation in both λ_t and v_{it} to a procedure employed for models with only one serially correlated error component. That is, one can keep v_{it} *iid* when testing for serial correlation in λ_t and one can keep λ_t *iid* when testing for serial correlation in v_{it}. This is so because the LM test for serial correlation in λ_t or v_{it} keeping v_{it} and λ_t *iid* respectively is asymptotically equivalent to the LM test taking the serial correlation into account and the LM test employing a robust variance-covariance matrix estimator. The usefulness of this property in practice depends entirely on the small-sample performance of the LM test that ignores the misspecification of serial correlation. To evaluate the performance we conducted a limited Monte-Carlo experiment with an AR(1) process for λ_t and v_{it} and with $T = 20$, $N = 70$. The results where encouraging, we observed no significant difference between the test that takes the serial correlation in the other component into account and the test that ignores the serial correlation in the other component. The situation is thus quite similar when testing for serial correlation in λ_t and v_{it}. In both cases we can ignore serial correlation in the other component. The only difference of importance is the different asymptotics for the test-statistics. Parameters involving λ_t are $\sqrt{T}$ consistent and $\xi^*_{LM_\lambda}$ converges to a χ^2 at the rate $\sqrt{T}$ whereas parameters involving v_{it} are $\sqrt{NT}$ consistent and $\xi^*_{LM_v}$ converges to a χ^2 at the rate $\sqrt{NT}$. Due to the similarity of the testing situation we discuss tests for serial correlation in general terms.

Testing ARMA(p, q) **against ARMA**$(p + r, q)$ LM testing is quite attractive when general ARMA(p, q) models are considered. Primarily because we do not need to estimate the alternative, but also because we can be less specific about the alternative we are testing against. That is, the well-known property that the LM test of the null of ARMA(p, q) against ARMA$(p + r, q)$ is identical to the LM test against ARMA$(p, q + r)$ holds here as well. The drawback is that it is not clear which alternative to choose when the null is rejected.

To discriminate between ARMA$(p + r, q)$ and ARMA$(p, q + r)$ processes with LM tests we need to consider non-nested hypotheses. For example, in the case of deciding between AR(1) or MA(1) a test of the hypothesis that the process is AR(1) amounts to testing the null hypothesis of AR(1) in the ARMA(1, 1) specification. Correspondingly, testing the null that the process is MA(1) amounts to testing the null hypothesis of MA(1) in the ARMA(1, 1) specification. Karlsson and Skoglund (2000) consider non-nested LM tests for discriminating between an AR(1) or MA(1) process for λ_t in an extensive Monte-Carlo experiment. This procedure worked well for large sample sizes but for small sample sizes (small T) and/or small values of the AR or MA parameters test results are frequently inconclusive. A decision can then be based on information criteria or a comparison of the p-values of the tests.

Pure AR models In practice attention is frequently based on pure AR models. In this case the order of serial correlation in λ_t or v_{it} can be determined with sequential hypothesis tests. That is, we test the following nested sequence for λ_t (v_{it})

$$\begin{aligned} H_0 &: \lambda_t \sim iid \\ H_1 &: \lambda_t \sim AR(1) \\ H_2 &: \lambda_t \sim AR(2) \\ & \quad \cdots \\ H_p &: \lambda_t \sim AR(p) \end{aligned}$$

As is well-known tests of hypotheses in such a sequence of nested hypotheses has a very interesting property. Asymptotically, under H_0, the test of H_0 against H_1 is independent of the test of H_1 against H_2, both of these are independent of the test of H_2 against H_3 and so on. This has the useful implication that we can compute the overall asymptotic significance level and that the test of H_0 against H_2 is equal to the sum of the tests of H_0 against H_1 and H_1 against H_2.

3.2.2 Testing the null hypothesis of no individual or time effects

Having decided on the orders of serial correlation we also want to test the null hypothesis of the one-way model with time effects, $H_0 : \sigma_\mu^2 = 0$, and the null hypothesis of the one-way model with individual effects, $H_0 : \sigma_u^2 = 0$[1]. Both of these hypotheses involve a parameter on the boundary of the parameter space. Tests that require estimation of the alternative when a parameter is on the boundary will in general not have a χ^2 distribution under the null. The breakdown of conventional theory for these tests reflect the fact that they involve the unrestricted maximum likelihood estimator for which the continuity of the asymptotic distribution is violated if a parameter is allowed to be on the boundary. In contrast, the LM test is not affected by the fact that a parameter lies on the boundary, see Godfrey (1988, sec. 3.5.2) and the references therein.

The null hypothesis of the one-way model with time effects is uncomplicated. On the other hand the null hypothesis of the one-way model with individual effects is complicated if λ_t is serially correlated. This is so because under the null hypothesis the serial correlation parameter(s) of λ_t are not identified, so even the LM test have a non-standard distribution. This

[1] The local asymptotic power results of Bera, Sosa-Escudero and Yoon (2001) in the one-way model with individual effects indicate that it is important to specify the serial correlation correctly when testing for random effects. However their results seem to apply only for large N, i.e. when holding T fix.

"nuisance parameter" problem is treated in Davies (1977, 1987) and more recently by Andrews and Ploberger (1994) and Hansen (1996). Hansen (1996) suggests a bootstrap procedure to simulate the asymptotic distribution of the LR test and Hansen (1999) contains an application to treshold effects in the one-way model with fixed individual effects.

Similar to Hansen we consider a bootstrap procedure to obtain an estimate of the asymptotic p-value. Andersson and Karlsson (1999) evaluates several algorithms for bootstrapping random effects models in the context of bootstrap tests on the regression parameters. They find little difference between non-parametric and parametric procedures, assuming normality, even in the presence of non-normality. Since the parametric bootstrap is straightforward to implement we use a slight modification, taking account of serial correlation in v_{it}, of the parametric bootstrap of Andersson and Karlsson (1999). See Efron and Tibshirani (1993) for a general discussion of the bootstrap and Davidson and MacKinnon (1999) for bootstrap testing in particular.

The bootstrap procedure we suggest consists of the following steps:

1. Estimate under the null hypothesis of the one-way model with individual effects to obtain an estimated Data Generating Process (DGP) with parameters $\widehat{\boldsymbol{\delta}}_0, \widehat{\boldsymbol{\gamma}}_0$ under the null hypothesis.

2. Generate bootstrap samples from the DGP. That is, generate the bootstrap sample, $\varepsilon_{it}^* = \mu_i^* + v_{it}^*$ with $\mu_i^* \sim N\left(0, \widehat{\sigma}_{\mu,0}^2\right)$, v_{it}^* an ARMA(p,q) process with innovation $e_{it}^* \sim N\left(0, \widehat{\sigma}_{e,0}^2\right)$ and parameters from the null model and create $y_{it}^* = \mathbf{z}_{it}'\widehat{\boldsymbol{\delta}}_0 + \varepsilon_{it}^*$.

3. Calculate the test-statistic using the bootstrap sample.

4. Repeat steps 2 and 3 B times

5. Calculate the percentage of draws for which the simulated statistic exceeds the actual. This gives the bootstrap estimate of the p-value.

The implementation of the above procedure with either LR or Wald tests might be quite time consuming since they both require estimation of the alternative. But more seriously, due to the breakdown of continuity in the asymptotic distribution of the unrestricted estimator, one can suspect that the bootstrap procedure suggested above does not yield a consistent estimate of the asymptotic p-value with either of these tests, see Andrews (2000). Hence, one should consider the LM test in step 3 above. However, under the null hypothesis the score and information matrix depends on the serial correlation parameter(s) of λ_t and so the LM test is not computable in the presence of these unknown nuisance parameters. As in Davies (1977, 1987) we consider a supremum LM test, where the supremum term indicates that we are taking the supremum with respect to the serial correlation parameters of λ_t as they vary over the parameter space. Denoting by Ξ the parameter space of the serial correlation parameters of λ_t the supremum LM test is computed as

$$\xi_{\sup LM} = \sup_{\Xi} \left(\left. \frac{\partial l}{\partial \sigma_u^2} \right|_{\sigma_u^2=0} \right)' \mathcal{I}^{\sigma_u^2 \sigma_u^2} \left(\left. \frac{\partial l}{\partial \sigma_u^2} \right|_{\sigma_u^2=0} \right) \tag{10}$$

where $\mathcal{I}^{\sigma_u^2 \sigma_u^2}$ is obtained from the generalized inverse variance matrix evaluated at the null hypothesis. The generalized inverse is needed here since the information is singular under the null hypothesis. Computationally this corresponds to inverting the positive-definite submatrix of the information, obtained by discarding the elements of information that belongs to the unidentified parameters. In addition by the standard (asymptotic) block-diagonality between the mean and variance parameters it is sufficient to only consider the block of the information matrix for the variance parameters.

3.2.3 A model specification strategy

To summarize, the considerations above leads to the following sequence of specification tests for two-way random effects models where both λ_t and v_{it} are allowed to be serially correlated:

1. While keeping v_{it} *iid*, test for serial correlation and determine the order of the ARMA process for λ_t.
2. While keeping λ_t *iid*, test for serial correlation and determine the order of the ARMA process for v_{it}.
3. Conditionally on the chosen orders for λ_t and v_{it} test for the presence of individual effects.
4. Conditionally on the chosen orders for λ_t and v_{it} and the outcome of the test for individual effects test for the presence of time effects.

Table 1 Summary statistics

Variable	Mean	Std. deviation	Minimum	Maximum
Y	87653.00	124621.01	2750.00	810887.00
K	70771.66	107797.26	1153.00	646335.00
L	12417.50	14123.84	732.00	90798.00
M	69362.31	103496.98	1812.00	663145.00

4 Application

4.1 The model and data

In this section we apply the proposed methods to the estimation of a production function using a sample of 72 Japanese chemical industries observed annually over the period 1968-1987. In the econometric analysis of production functions one is, naturally, concerned with serial correlation. The data contain information on output (Y) and inputs, labor (L), capital (K) and material (M) used[2]. Summary statistics of the input and output quantities are given in Table 1.

To approximate the unknown production function we consider the transcendental logarithmic specification (Christensen, Jorgenson and Lau (1973))

$$\ln y_{it} = \alpha + \sum_{j=1}^{k} \beta_j \ln X_{jit} + \frac{1}{2} \sum_{s=1}^{k} \sum_{j=1}^{k} \phi_{sj} \ln X_{sit} \ln X_{jit} + \varepsilon_{it} \qquad (11)$$

where $\phi_{sj} = \phi_{js}$. This function is quadratic in the logarithms of the variables and reduces to the familiar Cobb-Douglas case if $\phi_{sj} = 0$ for all s, j.

Estimates of the returns to scale are obtained from the sums of the logarithmic derivatives with respect to the inputs, $X_j, j = 1, \ldots, k$. The technology exhibits constant returns to scale if the sum is unity and increasing and decreasing returns to scale if the sum is above or below unity respectively. By including time as a component of the input vector we can also obtain estimates of the rate of technical change. The rate of technical change can further be decomposed into neutral technical change (no interaction with the inputs) and non-neutral technical change (interaction with the inputs).

[2] See Kumbakhar, Nakamura and Heshmati (2000) for a detailed description of the data.

4.2 Specification of the error components

Table 2 gives results from the LM-tests for serial correlation. For comparison Table 4 reports the AIC and BIC criteria for some of the estimated models[3]. First we test for serial correlation in λ_t while keeping v_{it} *iid*. We reject the null of no serial correlation in λ_t at the 12% level which is not overwhelming evidence in favor of serial correlation in λ_t. The test is however consistent in T alone and the simulation evidence of Karlsson and Skoglund (2000) show that for small values of T we can expect low power from this test. This motivates us to consider an AR(1) or MA(1) process for λ_t. The LM test of the null hypothesis of AR(1) against ARMA(1, 1) yields a p-value of 0.07 whereas the LM test of the null hypothesis of MA(1) against ARMA(1, 1) yields a p-value of 0.63. The obtained p-values thus suggest that an MA(1) process is appropriate. Next we consider tests for serial correlation in the idiosyncratic errors, v_{it} while keeping λ_t *iid*. Table 2 shows that we strongly reject the null of *iid* v_{it} in favor of an AR(1) or MA(1) process. Unfortunately we experienced convergence problems with the MA(1) model for v_{it} and hence the LM test of the null of MA(1) against ARMA(1, 1) could not be computed. However the LM test of the null of AR(1) against ARMA(1, 1) is not rejected at reasonable levels. Since the MA(1) process cannot match the moments of an AR(1) process for high absolute values of the AR(1) coefficient the convergence problems with the MA model is not too surprising in view of the estimated AR(1) coefficient of approximately 0.75. Hence, we conclude that an AR(1) process is sufficient to capture the serial correlation in v_{it}[4].

Having decided on the orders of serial correlation, that is λ_t is MA(1) and v_{it} is AR(1), we proceed to consider the significance of individual and time effects. Table 3 shows that the LM-test of the null of no individual effects, $H_0 : \sigma_\mu^2 = 0$, rejects the null at the 1% level and the bootstrap p-value of the supremum LM-test of the null of no time effects, $H_0 : \sigma_u^2 = 0$, indicates that the time effects cannot be rejected at any significance level. The bootstrap

[3] All the models are estimated with analytical derivatives using the Newton algorithm. Variance constraints are imposed as $\sigma_e^2 > 0$ as well as $\sigma_\mu^2 \geq 0$ and/or $\sigma_u^2 \geq 0$. AR parameters are restricted according to the stationarity condition whereas MA parameters are not restricted. Instead estimates that do not satisfy the invertibility condition are mapped back to the invertibility region. In case variance estimates are used they are based on the information matrix.

[4] For comparison the LM test of the null hypothesis of *iid* λ_t against AR(1) with v_{it} an AR(1) yields a test statistic of 2.605 and the LM test of the null of *iid* v_{it} against AR(1) with λ_t an AR(1) yields a test statistic of 540.4. The values of these test statistics are very close to the corresponding values of the test statistics for the case when v_{it} is *iid* and λ_t is *iid* respectively. Similar results are obtained for the other tests in Table 2 as well, showing that the simple procedure for determining serial correlation in both λ_t and v_{it} works very well in practice.

Table 2 LM tests of serial correlation

Null	Alternative	Statistic	p-value
v_{it}, λ_t iid	v_{it} iid, λ_t $AR(1)$	2.458	0.117
v_{it} iid, λ_t $AR(1)$	v_{it} iid, λ_t $ARMA(1,1)$	3.184	0.074
v_{it} iid, λ_t $MA(1)$	v_{it} iid, λ_t $ARMA(1,1)$	0.225	0.635
v_{it}, λ_t iid	v_{it} $AR(1), \lambda_t$ iid	540.5	0.000
v_{it} $AR(1), \lambda_t$ iid	v_{it} $ARMA(1,1), \lambda_t$ iid	0.977	0.323

Table 3 LM tests of no individual or time effects for the two-way model, λ_t MA(1) and v_{it} AR(1)

Null	Alternative	Statistic	p-value
$\sigma_\mu^2 = 0$	$\sigma_\mu^2 \neq 0$	7.932	0.004
$\sigma_u^2 = 0$	$\sigma_u^2 \neq 0$	5731^a	$< 0.001^b$

[a]supremum LM test.
[b]bootstrap p-value, $B = 399$.

p-value of the supremum LM-test is based on $B = 399$ bootstrap replicates.

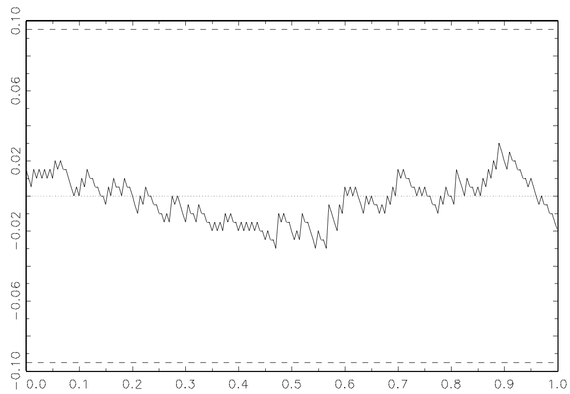
To evaluate the size properties of the bootstrap test we conducted a small Monte-Carlo experiment with the null model as the DGP. Figure 1 plots the size discrepancy (estimated size minus nominal size) against nominal size obtained from 200 Monte-Carlo replicates, using $B = 99$ for the bootstrap p-values, together with 95% Kolmogornov-Smirnov "confidence bands". Inspection of the Figure shows that the size properties of the supremum LM-test are very good.

Overall, the specification tests suggest that the appropriate random effects specification is two-way with λ_t an MA(1) and v_{it} an AR(1) process. From Table 4 this choice is supported by the AIC criterion whereas the BIC criterion prefers the one-way model with serially uncorrelated time effects and v_{it} an AR(1). The BIC criterion, although consistent, is however well-known to underestimate the true parametrization in finite samples.

Table 4 Model selection criteria

Model	AIC	BIC
One-way(λ_t, v_{it}), v_{it}, λ_t iid	-2.5181	-2.4559
One-way(λ_t, v_{it}), v_{it} iid, λ_t $AR(1)$	-2.5181	-2.4521
One-way(λ_t, v_{it}), v_{it} iid, λ_t $MA(1)$	-2.5201	-2.4522
One-way(λ_t, v_{it}), v_{it} $AR(1), \lambda_t$ iid	-3.2902	-3.2248
One-way(λ_t, v_{it}), v_{it}, λ_t $AR(1)$	-3.2909	-3.2214
One-way(λ_t, v_{it}), v_{it} $AR(1), \lambda_t$ $MA(1)$	-3.2930	-3.2234
One-way(λ_t, v_{it}), v_{it} $AR(1), \lambda_t$ $AR(2)$	-3.2911	-3.2200
One-way(λ_t, v_{it}), v_{it} $AR(2), \lambda_t$ $AR(1)$	-3.2907	-3.2185
One-way(λ_t, v_{it}), v_{it}, λ_t $AR(2)$	-3.2917	-3.2167
One-way(μ_i, v_{it}), v_{it} iid	-2.6220	-2.5597
One-way(μ_i, v_{it}), v_{it} $AR(1)$	-2.9981	-2.9322
One-way(μ_i, v_{it}), v_{it} $AR(2)$	-3.0083	-2.9388
Two-way, v_{it}, λ_t iid	-2.8198	-2.7538
Two-way, v_{it} iid, λ_t $AR(1)$	-2.8200	-2.7505
Two-way, v_{it} iid, λ_t $MA(1)$	-2.8220	-2.7524
Two-way, v_{it} $AR(1), \lambda_t$ iid	-3.2936	-3.2240
Two-way, v_{it}, λ_t $AR(1)$	-3.2939	-3.2207
Two-way, v_{it} $AR(1), \lambda_t$ $MA(1)$	-3.2960	-3.2227
Two-way, v_{it} $AR(1), \lambda_t$ $AR(2)$	-3.2944	-3.2175
Two-way, v_{it} $AR(2), \lambda_t$ $AR(1)$	-3.2932	-3.2163
Two-way, v_{it}, λ_t $AR(2)$	-3.2938	-3.2133

Figure 1 Size discrepancy of the supremum LM test



4.3 Elasticities and returns to scale

Table 5 gives the overall mean of the input elasticities and returns to scale and Figure 2a-2d plots the elasticities and returns to scale over time. The elasticity of output with respect to capital, reflecting percent changes in output due to one percent change in capital, is 0.086. It is interpreted as one percent change in capital will result in 0.086% change in value added for given, labor, material and technology. The corresponding labor and material elasticities are 0.155 and 0.754 respectively. All three elasticities are statistically significant at the 1% level and are of expected (positive) sign. The sum of input elasticities is 0.995, indicating on average constant returns to scale. Looking at the temporal patterns of input elasticities (Figure 2) we find that the elasticity of capital and labor are declining over time, indicating development of labor and capital input saving technologies. The fluctuations in the elasticity of capital in the beginning of the sample might be a consequence of the oil crisis of 1968 and 1973, resulting in increased capital intensity to introduce material (oil) saving technologies. The changes in capital are reflected in the development of the elasticity of material. The material input is constantly increasing over time, reflecting increasing share of cost associated with the raw oil input in the chemical industry. The returns to scale changes abruptly from increasing to decreasing returns to scale in the beginning of the sample but is quite stable after the oil crisis. Although Figure 2d provides a dramatic picture we should keep in mind that the fluctuations are contained in a narrow band and the returns to scale is never significantly different from unity.

Figure 2 Elasticities, returns to scale and technical change

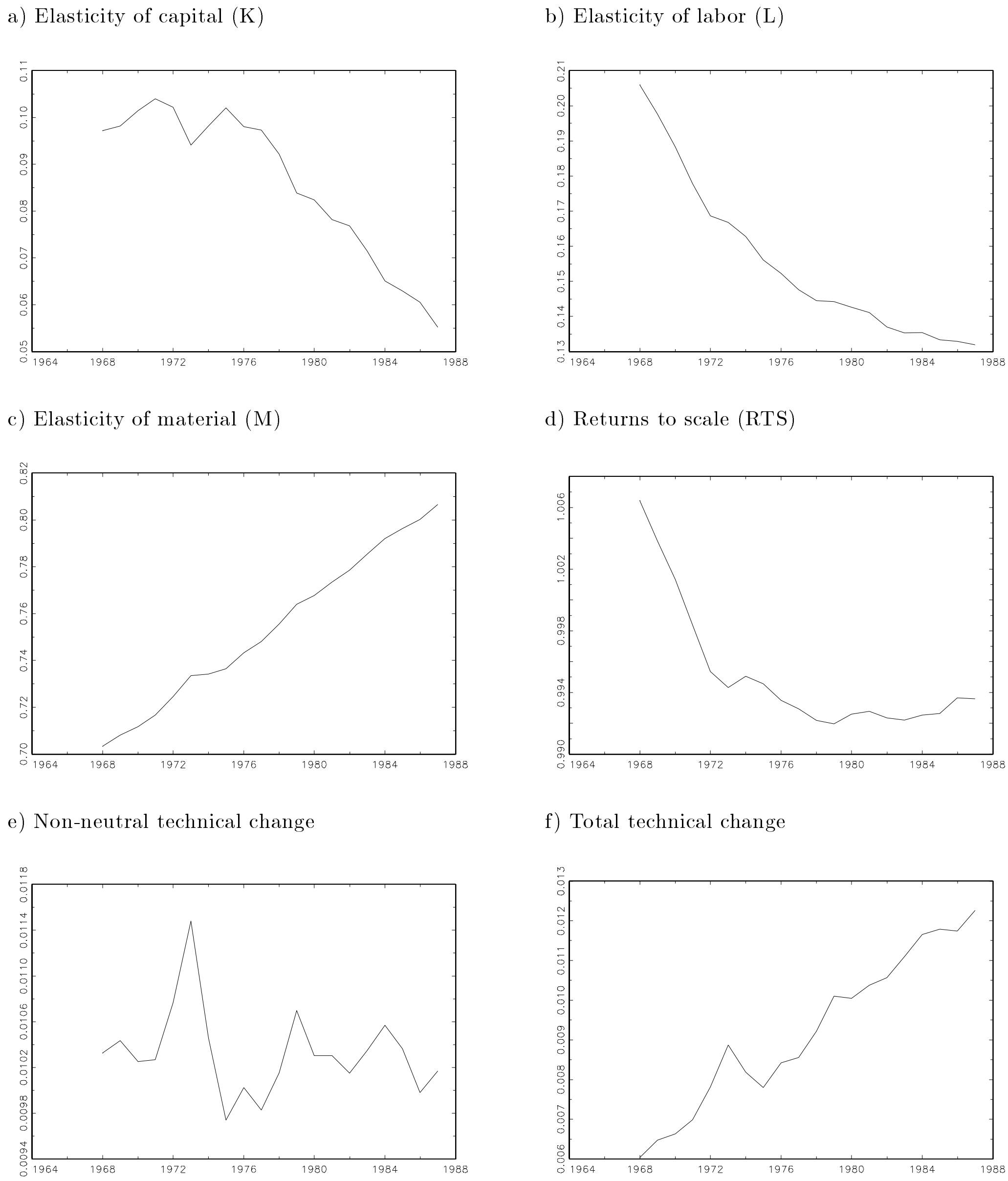


Table 5 Elasticities, returns to scale and technical change

Variable	Elasticity	Standard err.	t-stat
K	0.0861	0.0084	10.250
L	0.1552	0.0121	12.826
M	0.7537	0.0089	84.685
RTS	0.9951	0.0049	203.08
TC	0.0092	0.0017	5.4117

4.4 Technical change

The last row of Table 5 gives the overall mean of total technical change and Figure 2e and 2f plots non-neutral and total technical change over time. The average rate of technical change is 0.9% per annum with contributions from the neutral and non-neutral components being -0.01% and 0.1% respectively. The time patterns of non-neutral and total technical change reflect changes in technology due to the oil crises, indicating substitution among inputs. As a consequence of the affine neutral component we do not observe major changes or any technical regress during the post oil crises period.

5 Conclusions

The purpose of this paper has been to provide a framework for specification and estimation of two-way random effects models with serial correlation in general form for both the time effects and idiosyncratic errors.

In addition to providing a straightforward maximum likelihood estimator we have considered a model selection strategy for determining the orders of serial correlation as well as the significance of time and individual effects.

By relying on large sample theory results it has been possible to reduce the potential complexity of determining the order of serial correlation in both time effects and idiosyncratic errors to a standard procedure suitable for two-way models with only one serially correlated error component.

Conditional on the appropriate orders of serial correlation we considered an LM test of the null of no individual effects as well as an LM test of the null of no time effects. The LM test of the null of no time effects typically have non-standard distribution and we have suggested a simple bootstrap procedure to obtain an estimate of the p-value of the test.

An application to the estimation of a production function for Japanese chemical firms has illustrated the proposed methods.

Acknowledgement *We thank Professor Shinichiro Nakamura for providing us with the data set and Almas Heshmati for helpful comments.*

A Proof of property 1

This appendix contains a sketch of the proof of property 1. To avoid unneccesary complication and to be able to cut down on details by referring to the results of Skoglund and Karlsson (2001) we concentrate on the AR(1) case for both λ_t and v_{it} (with serial correlation parameter ρ_λ for λ_t and serial correlation parameter ρ_v for v_{it}).

To introduce some notation let $l(\boldsymbol{\tau}, 0, r)$ denote the data generating process, where $\boldsymbol{\tau}=\left(\sigma_\mu^2, \sigma_e^2, \sigma_u^2\right)$ and r is interior to a compact subset of $(-1,1)$[5]. Now let $l(\boldsymbol{\tau}, \rho, 0)$ be the model under consideration. Then r represents a misspecification of the serial correlation in v_{it} or λ_t. That is, $r = \rho_v$ if there is misspecification of serial correlation in v_{it} and $r = \rho_\lambda$ if there is misspecification of serial correlation in λ_t where, of course, $\rho = \rho_v$ if $r = \rho_\lambda$ and vice versa. We are interested in the properties of the LM test under the null hypothesis, $H_0 : \rho = 0$, and for this purpose it is useful to consider the behavior of the score vector evaluated under the null hypothesis.

By a mean-value expansion of the score vector of the (joint) log-likelihood, $l(\boldsymbol{\tau}, \rho, r)$ we have

$$\mathbf{F}^{-1}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}|_{\widetilde{\gamma}}\right) = \mathbf{F}^{-1}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}|_{\gamma_0}\right) - \left[-\mathbf{F}^{-1}\left(\frac{\partial^2 l(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'}|_{\overline{\gamma}}\right)\mathbf{F}^{-1}\right]\mathbf{F}\left(\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\right)$$

where $\boldsymbol{\gamma} = (\boldsymbol{\tau}, \rho, r)$, $\widetilde{\boldsymbol{\gamma}} = (\widetilde{\boldsymbol{\tau}}, 0, 0)$, $\boldsymbol{\gamma}_0 = (\boldsymbol{\tau}, 0, r)$, $\overline{\boldsymbol{\gamma}}$ a mean value and $\mathbf{F}$ is an appropriate scaling matrix (see Skoglund and Karlsson (2001) for details). To introduce some notation for the information matrix we write $E\left[-\mathbf{F}^{-1}\left(\frac{\partial^2 l(\gamma)}{\partial\gamma\partial\gamma'}|_{\overline{\gamma}}\right)\mathbf{F}^{-1}\right] = \mathcal{I}(\boldsymbol{\gamma})$ in partitioned form as,

$$\mathcal{I}(\boldsymbol{\gamma}) = \begin{bmatrix} \mathcal{I}_{\tau\tau} & \mathcal{I}_{\tau\rho} & \mathcal{I}_{\tau r} \\ \mathcal{I}_{\rho\tau} & \mathcal{I}_{\rho\rho} & \mathcal{I}_{\rho r} \\ \mathcal{I}_{r\tau} & \mathcal{I}_{r\rho} & \mathcal{I}_{rr} \end{bmatrix}$$

and solve for $s^{-\frac{1}{2}}\left(\frac{\partial l(\gamma)}{\partial \rho}|_{\widetilde{\gamma}}\right)$ in the expansion above, where s is an index obtained from the scaling matrix, $\mathbf{F}$. After some manipulation we can then write

$$s^{-\frac{1}{2}}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \rho}|_{\widetilde{\gamma}}\right) = s^{-\frac{1}{2}}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \rho}|_{\gamma_0}\right) + \mathcal{I}_{\rho\tau}\mathcal{I}_{\tau\tau}^{-1}\mathbf{F}_\tau^{-1}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \tau}|_{\gamma_0}\right) + \mathcal{I}_{\rho r,\tau} rc \quad (12)$$

where c is also an index obtained from the scaling matrix, $\mathbf{F}_\tau$ is the subscaling matrix for $\boldsymbol{\tau}$ and $\mathcal{I}_{\rho r,\tau}$ denotes the matrix $\mathcal{I}_{\rho r} - \mathcal{I}_{\rho\tau}\mathcal{I}_{\tau\tau}^{-1}\mathcal{I}_{\tau r}$. For example, if $r = \rho_v$ and $\rho = \rho_\lambda$ then we obtain $s = T$, $c = \sqrt{NT}$ from the matrix

[5] By theorem 2 of Skoglund and Karlsson (2000) we may ignore the mean parameters $\boldsymbol{\delta}$ for simplicity.

F and (12) becomes

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \rho_\lambda}|_{\tilde{\gamma}}\right) \\
=\ &\frac{1}{\sqrt{T}}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \rho_\lambda}|_{\gamma_0}\right) + \mathcal{I}_{\rho_\lambda \boldsymbol{\tau}}\mathcal{I}_{\boldsymbol{\tau\tau}}^{-1}\mathbf{F}_{\boldsymbol{\tau}}^{-1}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\tau}}|_{\gamma_0}\right) + \mathcal{I}_{\rho_\lambda\rho_v,\boldsymbol{\tau}}\rho_v\sqrt{NT}
\end{aligned}
$$

whereas if $r = \rho_\lambda, \rho = \rho_v$ we obtain $s = NT$, $c = \sqrt{T}$ and,

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \rho_v}|_{\tilde{\gamma}}\right) \\
=\ &\frac{1}{\sqrt{NT}}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \rho_v}|_{\gamma_0}\right) + \mathcal{I}_{\rho_v \boldsymbol{\tau}}\mathcal{I}_{\boldsymbol{\tau\tau}}^{-1}\mathbf{F}_{\boldsymbol{\tau}}^{-1}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\tau}}|_{\gamma_0}\right) + \mathcal{I}_{\rho_v\rho_\lambda,\boldsymbol{\tau}}\rho_\lambda\sqrt{T}
\end{aligned}
$$

In contrast, the LM test with no misspecification ($r = 0$) is based on the score vector

$$
s^{-\frac{1}{2}}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \rho}|_{\tilde{\gamma}}\right) = s^{-\frac{1}{2}}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \rho}|_{\gamma_0}\right) + \mathcal{I}_{\rho\boldsymbol{\tau}}\mathcal{I}_{\boldsymbol{\tau\tau}}^{-1}\mathbf{F}_{\boldsymbol{\tau}}^{-1}\left(\frac{\partial l(\boldsymbol{\gamma})}{\partial \boldsymbol{\tau}}|_{\gamma_0}\right) \tag{13}
$$

To investigate if the score equations (12) and (13) are in some sense equivalent it therefore suffices to consider the behavior of the term, $\mathcal{I}_{\rho_\lambda\rho_v,\boldsymbol{\tau}}\rho_v\sqrt{NT}$ and the term, $\mathcal{I}_{\rho_v\rho_\lambda,\boldsymbol{\tau}}\rho_\lambda\sqrt{T}$. By using the limit results of Skoglund and Karlsson (2001) it is straightforward to show that $\mathcal{I}_{\rho_\lambda\rho_v,\boldsymbol{\tau}}$ and $\mathcal{I}_{\rho_v\rho_\lambda,\boldsymbol{\tau}}$ shrink towards zero at the rate $N^{-\frac{3}{2}}$. Hence, $\mathcal{I}_{\rho_\lambda\rho_v,\boldsymbol{\tau}}\rho_v\sqrt{NT}$ shrinks to zero if $\left(\frac{\sqrt{T}}{N}\right) \to 0$ and $\mathcal{I}_{\rho_v\rho_\lambda,\boldsymbol{\tau}}\rho_\lambda\sqrt{T}$ shrinks to zero if $\left(\frac{\sqrt{T}}{N^{\frac{3}{2}}}\right) \to 0$. By using similar reasoning, and essentially the same limits as above, one can show that the score equation (13) is for large N equivalent to the score equation that takes into account the serial correlation. Now, considering the quadratic form of the LM test, we are done if we can show that the information matrix equality holds for the relevant block of the variance matrix employed in the LM test. A first step in the proof of this would be to show that the limit of the negative of the expected hessian matrix (suitably normalized of course) is block-diagonal between the parameters $\left(\sigma_\mu^2, \sigma_e^2, \rho_v\right)$ and the parameters $\left(\sigma_u^2, \rho_\lambda\right)$ as $N \to \infty$. This is accomplished by theorem 2 of Skoglund and Karlsson (2001). In addition theorem 2 of Skoglund and Karlsson (2001) shows that for large N the information of the parameters $\left(\sigma_\mu^2, \sigma_e^2, \rho_v\right)$ does not depend on the parameters $\left(\sigma_u^2, \rho_\lambda\right)$ and vice versa. This is in fact the key to the result since, for example, misspecification of serial correlation in v_{it} does not change the probabilistic order of the variance of the score and hence not the large N limit of the variance of the score for the block of time-specific parameters, $(\sigma_u^2, \rho_\lambda)$. Hence, rendering the information matrix equality valid for this block. We omit the details of this result since it is mainly algebraic, using the limit results of Skoglund and Karlsson (2001). Combining what we have obtained so far gives the results in property 1.

References

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**, 716–723.

Andersson, M. and Karlsson, S. (1999), Bootstrapping error component models, SSE/EFI Working Paper Series in Economics and Finance 304, Stockholm School of Economics. Forthcoming in computational statistics.

Andrews, D. W. K. (2000), 'Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space', *Econometrica* **68**, 399–405.

Andrews, D. W. K. and Ploberger, W. (1994), 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica* **62**, 1383–1414.

Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, John Wiley & Sons.

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**, 385–393.

Baltagi, B. H. and Li, Q. (1994), 'Estimating errror component models with general MA(q) disturbances', *Econometric Theory* **10**, 396–408.

Bera, A. K., Sosa-Escudero, W. and Yoon, M. (2001), 'Tests for the error component model in the presence of local misspecification', *Journal of Econometrics* **101**, 1–23.

Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1973), 'Transcendental logarithmic production frontiers', *The Review of Economics and Statistics* **55**, 28–45.

Davidson, R. and MacKinnon, J. G. (1999), 'The size distortion of bootstrap tests', *Econometric Theory* **15**, 361–376.

Davies, R. B. (1977), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* **64**, 247–254.

Davies, R. B. (1987), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* **74**, 33–43.

Dhrymes, P. J. (1984), *Mathematics for Econometrics*, Springer Verlag.

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman and Hall.

Galbraith, J. W. and Zinde-Walsh, V. (1995), 'Transforming the error-components model for estimation with general ARMA disturbances', *Journal of Econometrics* **66**, 349–355.

Godfrey, L. G. (1988), *Misspecification Tests in Econometrics*, Cambridge university press.

Hansen, B. E. (1996), 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica* **64**, 413–430.

Hansen, B. E. (1999), 'Treshold effects in non-dynamic panels: Estimation, testing, and inference', *Journal of Econometrics* **93**, 345–368.

Hartley, H. O. and Rao, J. N. K. (1967), 'Maximum likelihood estimation for the mixed analysis of variance model', *Biometrika* **54**, 93–108.

Harville, D. A. (1977), 'Maximum likelihood approaches to variance component estimation and to related problems', *Journal of the American Statistical Association* **72**, 320–340.

Hemmerle, W. J. and Hartley, H. O. (1973), 'Computing maximum likelihood estimates for the mixed a.o.v. model using the w-transformation', *Technometrics* **15**, 819–831.

Karlsson, S. and Skoglund, J. (2000), Maximum-likelihood based inference in the two-way random effects model with serially correlated time effects, SSE/EFI Working Paper Series in Economics and Finance 383, Stockholm School of Economics.

King, M. L. (1986), 'Efficient estimation and testing of regressions with a serially correlated error component.', *Journal of Quantitative Economics* **2**, 231–247.

Kumbakhar, S. C., Nakamura, S. and Heshmati, A. (2000), 'Estimation of firm-specific technological bias, technical change and total factor productivity growth: A dual approach', *Econometric Reviews* **19(4)**, 493–515.

Lillard, L. A. and Willis, R. J. (1978), 'Dynamic aspects of earnings mobility', *Econometrica* **46**, 985–1012.

Magnus, J. R. and Woodland, A. D. (1988), 'On the maximum likelihood estimation of multivariate regression models containing serially correlated error components', *International Economic Review* **29**, 707–725.

Mazodier, P. and Trognon, A. (1978), 'Heteroskedasticity and stratification in error components models', *Annales de l'INSEE* **30-31**, 451–482.

Randolph, W. C. (1988), 'A transformation for heteroskedastic error components regression models', *Economics Letters* **27**, 349–354.

Revankar, N. S. (1979), 'Error component models with serially correlated time effects', *Journal of the Indian Statistical Association* **17**, 137–160.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**, 461–464.

Skoglund, J. and Karlsson, S. (2001), Asymptotic properties of the maximum likelihood estimator of random effects models with serial correlation, SSE/EFI Working Paper Series in Economics and Finance 432, Stockholm School of Economics.

Wansbeek, T. J. and Kapteyn, A. (1982), 'A simple way to obtain the spectral decomposition of variance components models for balanced data', *Communications in Statistics* **A11**, 2105–2112.

Wansbeek, T. J. and Kapteyn, A. (1983), 'A note on spectral decomposition and maximum likelihood estimation of ANOVA models with balanced data', *Statistics and Probability Letters* **1**, 213–215.