Maximum-likelihood based inference in the two-way random effects model with serially correlated time effects

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

January 29, 2000

Abstract

This paper considers maximum likelihood estimation and inference in the two-way random effects model with serial correlation. We derive a straightforward maximum likelihood estimator when the time-specific component follow an AR(1) or MA(1) process. The estimator is easily generalized to arbitrary stationary and strictly invertible ARMA processes. Furthermore we derive tests of the null hypothesis of no serial correlation as well as tests for discriminating between the AR(1) and MA(1) specifications. A Monte-Carlo experiment evaluates the finite-sample properties of the estimators and test-statistics

1 Introduction

Following the influential work of Lillard and Willis (1978) there has been a continued interest in error component models which allow for dynamics in the form of a serially correlated error component. As in Lillard and Willis, Anderson and Hsiao (1982), MaCurdy (1982) and Baltagi and Li (1991, 1994) consider a one-way error component model with individual specific effects and serially correlated idiosyncratic errors. King (1986) studies a one-way model with correlated time specific effects and independent idiosyncratic errors whereas Magnus and Woodland (1988) consider a multivariate panel data model where both the time specific effects and the idiosyncratic errors are correlated. See Baltagi (1995, ch. 4) for a review of the literature.

In this paper we consider the two way random effects model with serially correlated time specific effects. That is, the serially correlated component is common to individuals and can be taken to represent common or macro effects not accounted for by the explanatory variables. More specifically, the model of interest is

$$\begin{aligned} y_{it} &= \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it} \\ \varepsilon_{it} &= \mu_i + \lambda_t + \nu_{it} \end{aligned} \tag{1}$$

with λ_t an AR(1),

$$\lambda_t = \rho\lambda_{t-1} + u_t, \tag{2}$$

or MA(1),

$$\lambda_t = u_t + \theta u_{t-1}, \tag{3}$$

process. Revankar (1979) studied this model and gave a rather cumbersome two-step estimator for the special case where λ_t follows an AR(1) process. We offer a computationally straightforward maximum likelihood estimator which is easily generalized to arbitrary stationary and strictly invertible ARMA processes for λ_t. In addition we consider the model selection problem and give tests for autocorrelation in λ_t as well as tests that allow us to discriminate between the autoregressive and moving average specifications.

The organization of the paper is as follows. Section 2 presents the maximum likelihood estimator of the model. Section 3 derives the specification tests. Section 4 contains results from a Monte-Carlo experiment and section 5 concludes.

2 The Maximum-likelihood estimator

In matrix form the two way model (1) is written 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 $\mathbf{Z}_\mu = (I_N \otimes \boldsymbol{\iota}_T)$, $\mathbf{Z}_\lambda = (\boldsymbol{\iota}_N \otimes \mathbf{I}_T)$, $\mathbf{Z} = [\boldsymbol{\iota}_{NT}, \mathbf{X}]$, $\boldsymbol{\delta} = [\alpha, \boldsymbol{\beta}']'$, $\boldsymbol{\mu}' = (\mu_1, ..., \mu_N)$, $\boldsymbol{\lambda}' = (\lambda_1, ..., \lambda_T)$ and $\boldsymbol{\iota}_N$ is a vector of ones of dimension N. Throughout we will maintain the assumption that $\nu_{it} \sim N(0, \sigma_\nu^2)$, $\mu_i \sim N(0, \sigma_\mu^2)$, $u_t \sim N(0, \sigma_u^2)$ independent of each other and $\mathbf{X}$. In addition we assume that $\rho, \theta \in (-1, 1)$ that is the AR process (2) is stationary and the MA process (3) is strictly invertible.

The covariance matrix of the combined error term is given by

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

where $\mathbf{J}_T = \boldsymbol{\iota}_T\boldsymbol{\iota}_T'$ a $T \times T$ matrix of ones and Ψ is the covariance matrix of (2) or (3) with unit innovation variance. When the distinction between the two types of processes is important we will refer to the covariance matrix of the AR(1) process as Ψ_ρ and the covariance matrix of the MA(1) process as Ψ_θ. More generally Ψ can be the covariance matrix of an arbitrary stationary and strictly invertible ARMA(p, q) process.

Direct inversion of Σ 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. For maximum likelihood estimation to be practical convenient expressions for Σ^{-1} and $|\Sigma|$ must be found. To this end, let $\mathbf{E}_T = \mathbf{I}_T - \overline{\mathbf{J}}_T$, $\overline{\mathbf{J}}_T = \mathbf{J}_T/T$, $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$ and

$$\begin{aligned}\mathbf{A} &= \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_\nu^2(\mathbf{I}_N \otimes \mathbf{I}_T) = \mathbf{I}_N \otimes \left(\sigma_\mu^2\mathbf{J}_T + \sigma_\nu^2\mathbf{I}_T\right) \\ &= \mathbf{I}_N \otimes \left(\sigma_1^2\overline{\mathbf{J}}_T + \sigma_\nu^2\mathbf{E}_T\right)\end{aligned}$$

be the covariance matrix of the one-way model with individual specific effects. We can then write

$$\Sigma = \mathbf{A} + \boldsymbol{\sigma}_u^2(\boldsymbol{\iota}_N \otimes \mathbf{I}_T)\Psi(\boldsymbol{\iota}_N' \otimes \mathbf{I}_T)$$

Using a well known result from matrix algebra

$$\begin{aligned}\Sigma^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\boldsymbol{\iota}_N \otimes \mathbf{I}_T)[\sigma_u^{-2}\Psi^{-1} + (\boldsymbol{\iota}_N' \otimes \mathbf{I}_T)\mathbf{A}^{-1}(\boldsymbol{\iota}_N \otimes \mathbf{I}_T)]^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{I}_T)\mathbf{A}^{-1} \\ &= \mathbf{I}_N \otimes \mathbf{A}^* - (\boldsymbol{\iota}_N \otimes \mathbf{A}^*)[\sigma_u^{-2}\Psi^{-1} + N\mathbf{A}^*]^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \qquad (4) \\ &= \mathbf{I}_N \otimes \mathbf{A}^* - \sigma_u^2(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\left[\mathbf{I}_T + N\sigma_u^2\Psi\mathbf{A}^*\right]^{-1}\Psi(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)\end{aligned}$$

where

$$\mathbf{A}^{-1} = \mathbf{I}_N \otimes \left(\frac{1}{\sigma_v^2}\mathbf{E}_T + \frac{1}{\sigma_1^2}\overline{\mathbf{J}}_T\right) = \mathbf{I}_N \otimes \mathbf{A}^*$$

We obtain the determinant of Σ in a similar fashion as

$$\begin{aligned}|\Sigma| &= |\mathbf{A}^*|^{-N}|\sigma_u^2\Psi||\sigma_u^{-2}\Psi^{-1} + N\mathbf{A}^*| \qquad (5) \\ &= \sigma_\nu^{2N(T-1)}\sigma_1^{2N}\left|\mathbf{I}_T + N\sigma_u^2\Psi\mathbf{A}^*\right|\end{aligned}$$

Using these results we have the log-likelihood function as

$$
\begin{aligned}
l(\boldsymbol{\delta},\boldsymbol{\gamma}) \;=\; & -\frac{TN}{2}\ln 2\pi-\frac{N(T-1)}{2}\ln\sigma_v^2-\frac{N}{2}\ln\left(T\sigma_\mu^2+\sigma_v^2\right) \\
& -\frac{1}{2}\boldsymbol{\varepsilon}'\left(\mathbf{I}_N\otimes\mathbf{A}^*\right)\boldsymbol{\varepsilon}-\frac{1}{2}\ln\left|\mathbf{I}_T+N\sigma_u^2\Psi\mathbf{A}^*\right| \\
& +\frac{\sigma_u^2}{2}\boldsymbol{\varepsilon}'(\boldsymbol{\iota}_N\otimes\mathbf{A}^*)\left[\mathbf{I}_T+N\sigma_u^2\Psi\mathbf{A}^*\right]^{-1}\Psi(\boldsymbol{\iota}_N'\otimes\mathbf{A}^*)\boldsymbol{\varepsilon}
\end{aligned}
\tag{6}
$$

where $\boldsymbol{\gamma}$ is the vector of covariance parameters, $\left(\sigma_\mu^2,\sigma_\nu^2,\sigma_u^2,\rho\right)$ for (2) and $\left(\sigma_\mu^2,\sigma_\nu^2,\sigma_u^2,\theta\right)$ for (3). This only requires numerical calculation of the determinant and inverse of the $T\times T$ matrix $\mathbf{I}_T+N\sigma_u^2\Psi\mathbf{A}^*$ which for the modest time series dimensions common in panel data applications is both speedy and accurate[1]. The theorem below shows that the models (1, 2) and (1, 3) are locally identified in the sense of Rothenberg (1971). The proof is given in appendix C.

Theorem 1 *Assume that* $-1<\tau<1$ *where* $\tau=\rho$ *or* $\tau=\theta$*, and* $0<\sigma_\mu^2,\sigma_v^2,\sigma_u^2\leq C<\infty$ *for some finite constant* C*. The dynamic two way random effect models (1, 2) and (1, 3) are then locally identified in the sense of Rothenberg (1971) when* $N,T\geq 2$*.*

The elements of the score are given in appendix A.1 and the information matrix in appendix A.2. The use of an analytic score is strongly suggested in applications since numerical derivatives performed poorly. Variance estimates can be based on either a numerical approximation to the Hessian matrix or the information matrix given in the appendix.

3 Specification tests

3.1 Testing for autocorrelation in λ_t

To derive an LM-statistic to test the null hypothesis $H_0:\rho=0$ against $\rho\neq 0$ in the AR(1) specification, we need the score and the information matrix evaluated at the two-way model with $\lambda_t=u_t\sim N(0,\sigma_u^2)$. The information matrix and the relevant element of the score vector evaluated under the null hypothesis are obtained from appendix A.2 and A.1 respectively by setting

[1] If an analytic inverse and determinant is available for Ψ it is more convenient to work with $\sigma_u^{-2}\Psi^{-1}+N\mathbf{A}^*$ (line 2 of (4) 1 of (5)) since the computations are much more efficient for symmetric positive definite matrices than for general matrices.

$\Psi = \mathbf{I}_T$ and $\mathbf{L} = \mathbf{G}$, where $\mathbf{G}$ is a bidiagonal matrix with bidiagonal elements all equal to one. The LM-test is computed as

$$\xi_1 = (\frac{\partial l}{\partial \rho}|_{\rho=0})' \mathcal{I}^{4,4} (\frac{\partial l}{\partial \rho}|_{\rho=0}) \tag{7}$$

where $\mathcal{I}^{4,4}$ is the $(4,4)$ element of the inverse information matrix for the variance parameters, $\mathcal{I}_{\gamma,\gamma}$, evaluated at the null hypothesis. Since the information matrix is block-diagonal between δ and γ it is sufficient to obtain this block.

Inspection of the score vector for the MA(1) model shows that $\frac{\partial l}{\partial \theta}|_{\theta=0} = \frac{\partial l}{\partial \rho}|_{\rho=0}$. It follows that (7) is also the LM-test against an MA(1) alternative.

The hypothesis of no autocorrelation can, of course, also be tested using Wald or LR-tests. In addition to requiring the use of slightly more complicated estimators, these tests require the choice of a specific alternative. In general we expect Wald or LR-tests against the correct alternative to have more power than the LM-test and the Wald or LR-tests against the wrong alternative to have lower power than the LM-test.

3.2 Testing AR(1) vs. MA(1)

Having rejected the null of no serial correlation using one of the tests discussed in the previous section, the next step is to decide wether to model λ_t as an AR or MA process. In this section we develop formal tests which allow us to discriminate between the AR(1) and MA(1) specifications. Testing is complicated by the hypotheses being non-nested and test results will frequently be inconclusive. Model choice can then be based on less formal criteria, such as comparison of p-values or information criteria. Note that in the case of AR(1) vs. MA(1), the choice of information criteria to use is irrelevant since they all boil down to a simple comparison of the likelihoods for the two specifications.

In order to develop formal tests we nest the two hypothesis in the comprehensive ARMA(1,1) specification for λ_t. Since estimation of the comprehensive model is complicated we do not consider Wald or LR-tests and concentrate on LM-tests. The test of the hypothesis that the true process for λ_t is AR(1) then corresponds to testing $H_0 : \theta = 0$ in the ARMA(1,1) specification. We will refer to this test as the LM-AR test. Correspondingly, testing the null that the true process for λ_t is MA(1) is equivalent to testing $H_0 : \rho = 0$ in the ARMA(1,1) specification. We refer to this test as the LM-MA test.

Using the standard block diagonality between regression and variance parameters we have the test statistic for $H_0 : \tau = 0$ as

$$\xi_{2,\tau} = (\frac{\partial l}{\partial \tau}|_{\tau=0})' \mathcal{I}^{\tau\tau} (\frac{\partial l}{\partial \tau}|_{\tau=0})$$

where τ is θ if the null hypothesis is AR(1) and ρ if the null hypothesis is MA(1) and $\mathcal{I}^{\tau\tau}$ is the appropriate element of the inverse information matrix for the variance parameters, evaluated under the null hypothesis. The elements of the score and the information matrix evaluated under the null hypothesis are given in appendix B.

The LM-tests are relatively complicated and as an alternative we consider two tests which can be computed using only the within estimates of the standard two-way model. These tests are based on the same ideas as the BGT tests of Baltagi and Li (1995), to test implications of the process for λ_t being AR(1) or MA(1).

Let $\widehat{\lambda}_t$ be the dummy variable estimates of λ_t. Then

$$\widehat{\zeta}_j = \frac{1}{T} \sum_{t=j+1}^{T} \widehat{\lambda}_t \widehat{\lambda}_{t-j}$$

is a consistent estimator of $\zeta_j = \mathrm{cov}(\lambda_t, \lambda_{t-j})$. Under the null of MA(1) we have $\zeta_2 = 0$ and $\sqrt{T}\left(\widehat{\zeta}_2 - \zeta_2\right) \xrightarrow{d} N\left(0, \zeta_0^2 + 2\zeta_1^2\right)$ under H_0 and the normality assumption. An asymptotically $N(0,1)$ test statistic for the null of MA(1) is thus given by

$$\xi_3 = \sqrt{T} \frac{\widehat{\zeta}_2}{\sqrt{\widehat{\zeta}_0^2 + 2\widehat{\zeta}_1^2}} \tag{8}$$

Under the alternative of AR(1), $\zeta_2 > 0$ and we reject in the right tail only in order to maximize power. We refer to the test (8) as the BGT-MA test.

Let $\eta_j = \mathrm{corr}\left(\lambda_t \lambda_{t-j}\right)$, under the null hypothesis of an AR(1) process

$$\eta_2 - (\eta_1)^2 = 0$$

whereas under the alternative of an MA(1) process $\eta_2 = 0$. The test statistic

$$\xi_4 = \sqrt{T}(\widehat{\eta}_2 - (\widehat{\eta}_1)^2)/(1 - \widehat{\eta}_2) \tag{9}$$

is asymptotically $N(0,1)$ under the null hypothesis and we reject in the left tail in order to maximize power against MA(1). We refer to the test (9) as the BGT-AR test. To get a test for which size approaches zero asymptotically we may also accept the null hypothesis if $\widehat{\eta}_1 > \frac{1}{2} + \frac{1}{\sqrt{T}}$, see Baltagi and Li (1995).

4 Monte-Carlo study

4.1 Design

We generate data from the two way model

$$\begin{aligned} y_{it} &= \alpha + \beta x_{it} + \varepsilon_{it} \\ \varepsilon_{it} &= \mu_i + \lambda_t + v_{it} \end{aligned}$$

where $\alpha = 0$, $\beta = 1$ and with λ_t an AR(1) (2) or MA(1) (3) process. The regressors, x_{it} are generated as

$$x_{it} = 0.6 x_{it-1} + \eta_{it}$$

where η_{it} is *iid* $N(0,1)$ and is held constant over the replicates of y_{it}. The variance parameters takes the values as $\sigma_\mu^2, \sigma_v^2 = (1/6, 2/6, 3/6, 4/6)$ and $\sigma_\lambda^2 = (1 - \sigma_\mu^2 - \sigma_v^2)$ for feasible combinations of σ_μ^2 and σ_v^2. That is $\sigma_u^2 = \sigma_\lambda^2 (1 - \rho^2)$ for the AR(1) specification and $\sigma_u^2 = \frac{\sigma_\lambda^2}{1+\theta^2}$ for the MA(1) specification. This choice of variance parameters holds the explanatory power of the model constant with an R^2 of 0.6. Finally ρ, θ takes the values $(-0.8, -0.4, 0, 0.4, 0.8)$. For each combination of parameter values we generate $10,000$ samples of $N = (10, 20)$ and $T = (25, 50)$. Normal μ_i, u_t, ε_{it} and η_{it} are obtained from the normal random number generator in GAUSS and initial values of the AR(1) process are obtained from the stationary distribution of λ_t.

Due to the large amount of output from the simulation experiment it is necessary to conserve on space. We only present results for the sample sizes $N = 10, T = 25$ and $N = 20, T = 50$. A full set of results can be obtained from the authors upon request.

4.2 Parameter estimates

The bias of parameters are small and the only potentially troublesome parameter to estimate is θ. The estimated variance of θ is very large for estimates close to one, which comes from the fact that the information matrix is singular at $|\theta| = 1$. Restricting $|\theta|$ below one led to serious convergence problems. Instead estimates above one in absolute value are transformed back to the invertibility region. The near singularity of the information matrix close to the invertibility boundary is however still reflected in the poor performance of the information matrix estimate of the variance.

We experienced some convergence problems with the MA(1) model when the true model was AR(1) with $|\rho| = 0.8$[2]. This is not too surprising since

[2] A replicate was dropped from the simulation if convergence was not achieved after 100 iterations. This reduces the effective number of replicates to between $8,467$ and $10,000$.

the MA(1) model cannot match the moments of the AR(1) process for high values of $|\rho|$.

Figure 1a shows the empirical distributions of parameters for $N = 10, T = 25$ in the MA(1) model when true model is MA(1) with $\theta = 0.8$. Figure 1b depicts the corresponding case for the AR(1) model when true model is AR(1) with $\rho = 0.8$. Normal densities with the same means and variances as the empirical distributions are superimposed. Pictures for negative values of ρ and θ are similar and corresponding pictures for $N = 20, T = 50$ improves on the negative skewness of the empirical distributions of variance parameters as well as centering the empirical distributions of ρ and θ around their true values.

4.3 Hypothesis tests

In each replicate we compute the LM-test of the null of no serial correlation as well as the LR and Wald-tests of the null of no MA(1) or AR(1). The Wald-tests are computed using a numerical approximation to the Hessian. Wald-tests based on the information matrix given in the appendix A.2 failed in the MA(1) model for $|\theta| = 0.8$ due to near singularity of the information. In addition we compute the tests for discriminating between the two specifications i.e. the LM-MA, LM-AR as well as the BGT-MA and BGT-AR tests.

In reporting our Monte-Carlo results for the test-statistics we use the graphical methods advocated by Davidson and McKinnon (1998). The size discrepancy graphs plot the difference between estimated size and nominal size against the nominal size of the tests. The size-power graphs plot power against the nominal size of the tests.

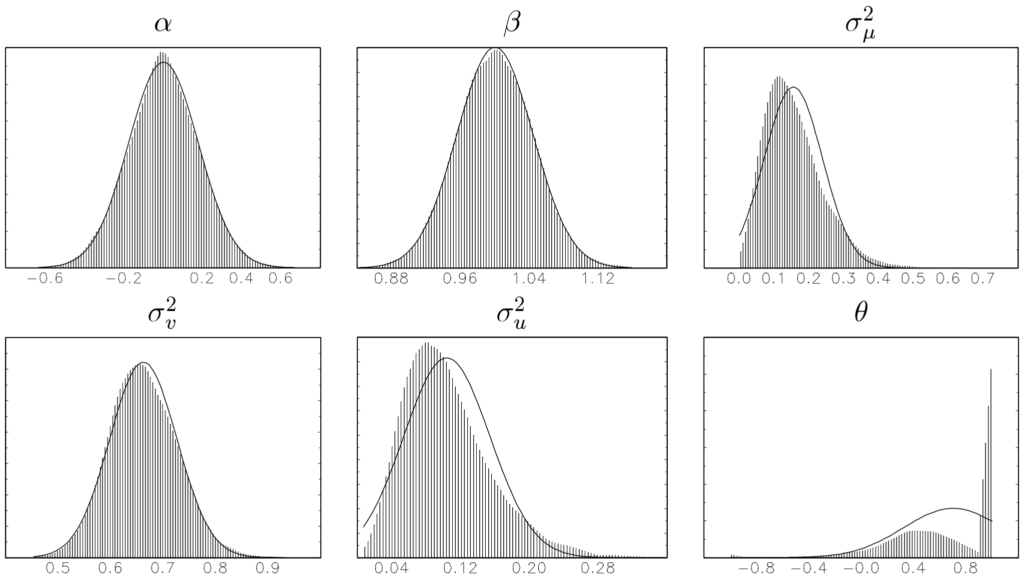
4.3.1 Tests of the null hypothesis of no serial correlation

Size Figure 2 shows the nominal size (x–axis) and size discrepancy (y-axis) with 95% Kolmogornov-Smirnov "confidence bands" for the LR, Wald and LM-tests[3]. For $N = 20, T = 50$ the size properties are very good for the LM-test and the Wald and LR-tests against an AR(1) alternative (Figure 2a-2c). When testing against an MA(1) alternative the LR and, especially the Wald test suffer from size distortion and are sensitive to the choice of variance parameters (Figure 2d-2e).

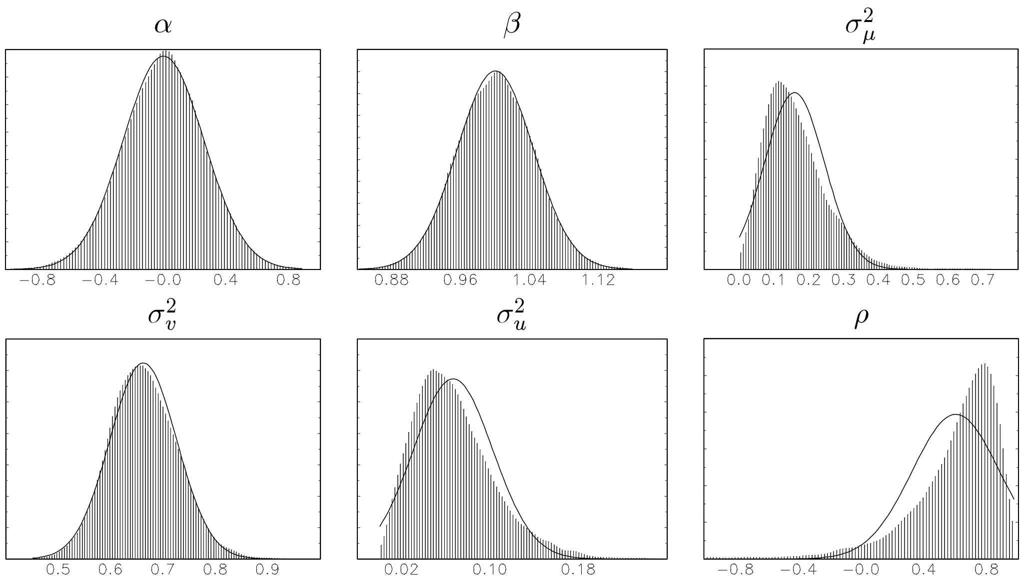
[3] In the graphs we refer to the parameter values of σ_μ^2, σ_v^2 as mi, $i = 1, .., 4$ and vj, $j = 1, .., 4$ respectively. For example m1v4 refer to $\sigma_\mu^2 = 1/6, \sigma_v^2 = 4/6$ and m2v1 refer to $\sigma_\mu^2 = 2/6, \sigma_v^2 = 1/6$.

Figure 1 Empirical distributions of parameters, $\sigma_\mu^2 = 1/6$ and $\sigma_v^2 = 4/6$, N=10, T=25

a) MA(1), $\theta = 0.8$



b) AR(1), $\rho = 0.8$



For $N = 10, T = 25$ the Wald and LR-tests against the alternative of MA(1) have serious size problems and are sensitive to the choice of variance parameters. Due to the serious size problems with these tests they will not be considered further. This is in contrast to the LM-test and Wald and LR-tests against an AR(1) alternative which performs reasonably well even for the smaller sample sizes (Figure 2f-2h).

Power Since power results for negative and positive values of ρ and θ are similar, we only report results for positive values of ρ and θ. For $N = 20, T = 50$ the LR-test typically has the highest power, but power differences are not large. Figure 3a-3c shows the nominal size (x-axis) and power (y-axis) for the LM, Wald and LR-tests in the AR(1) model with $\rho = 0.4$ and Figure 3d-3e shows the size and power of the LM-test in the MA(1) model. The picture is similar for the Wald and LR-tests. The tests are relatively insensitive to the choice of variance parameters, though a small reduction in power is achieved by decreasing σ_λ^2 (increasing $\sigma_\mu^2 + \sigma_v^2$), which is not surprising since a low σ_λ^2 makes it harder to detect the AR(1) or MA(1) structure. Furthermore for fix σ_λ^2 power is decreasing with increasing σ_v^2. Comparing Figure 3c and 3d it appears that the LM-test has lower power against MA(1) than AR(1) alternatives. It should however be kept in mind that the AR(1) process with a high value of $|\rho|$ is more persistent than the MA(1) process with $\theta = \rho$ and we would expect more power against the AR(1) process due to it being further away from the null hypothesis.

In the case of $N = 10, T = 25$ power is obviously lower, but it is also more sensitive to the choice of variance parameters. As for $N = 20, T = 50$ the LR-test typically has the highest power and power in the AR(1) model is larger. Still, the power differences between the tests and the models are relatively small. Figure 3f shows the size and power of the LM-test in the AR(1) model with $\rho = 0.4$.

4.3.2 Tests for discriminating between the AR(1) and MA(1) specifications

Size of BGT-AR and LM-AR Figure 4a-4d shows the size discrepancy of the BGT-AR and LM-AR tests for negative values of ρ. For $N = 20, T = 50$ the BGT-AR test is undersized at usual significance levels and the size is also sensitive to the choice of variance parameters[4]. A low σ_λ^2 with a relatively large σ_v^2 makes the BGT-AR test more undersized. For $|\rho| = 0.4$

[4] All references in the text and in the graphs refer to the unadjusted BGT-AR test i.e. the statistic (9) without size adjustment.

Figure 2 Size discrepancy of tests of no serial correlation

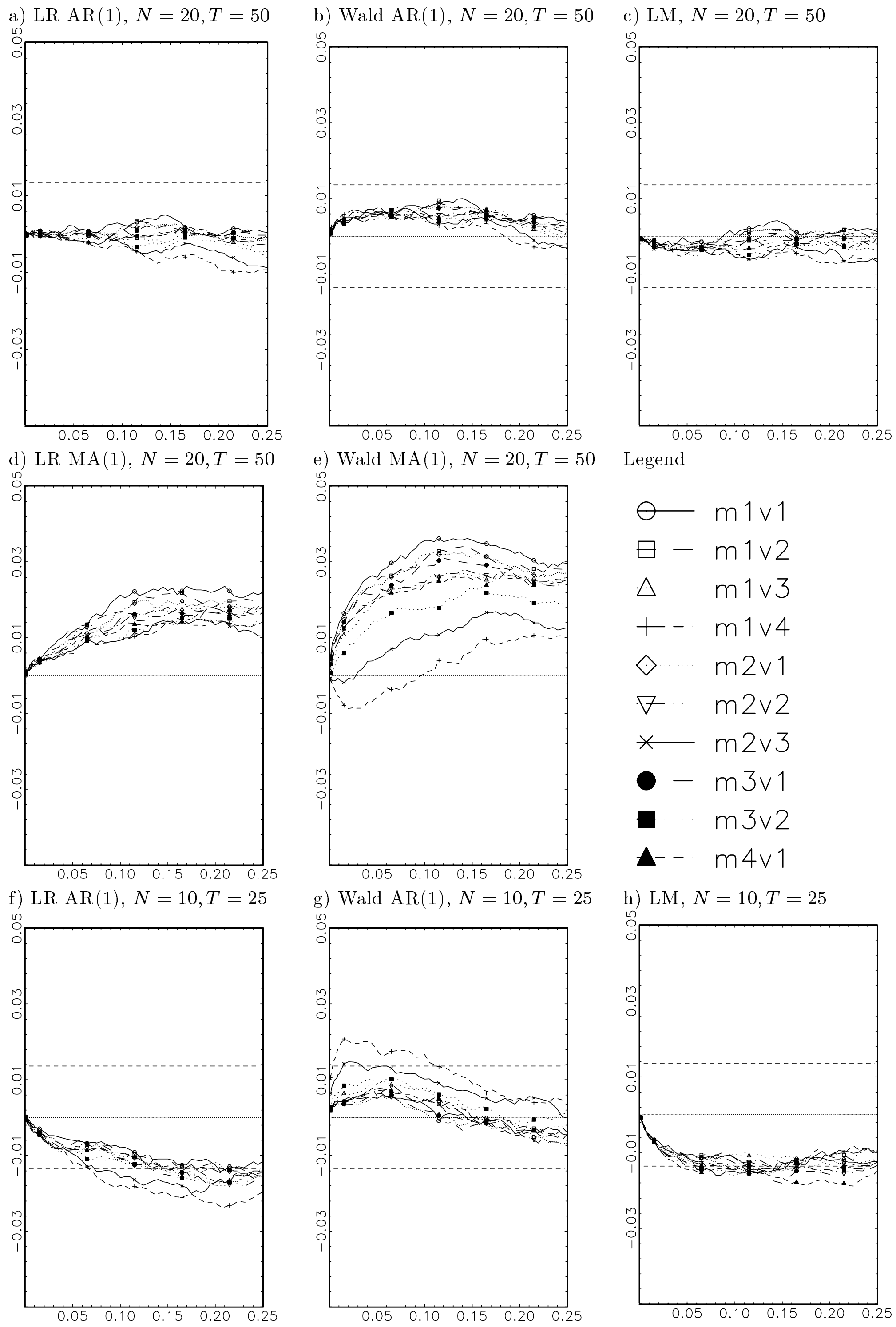


Figure 3 Power of tests of the null hypothesis of no AR(1)

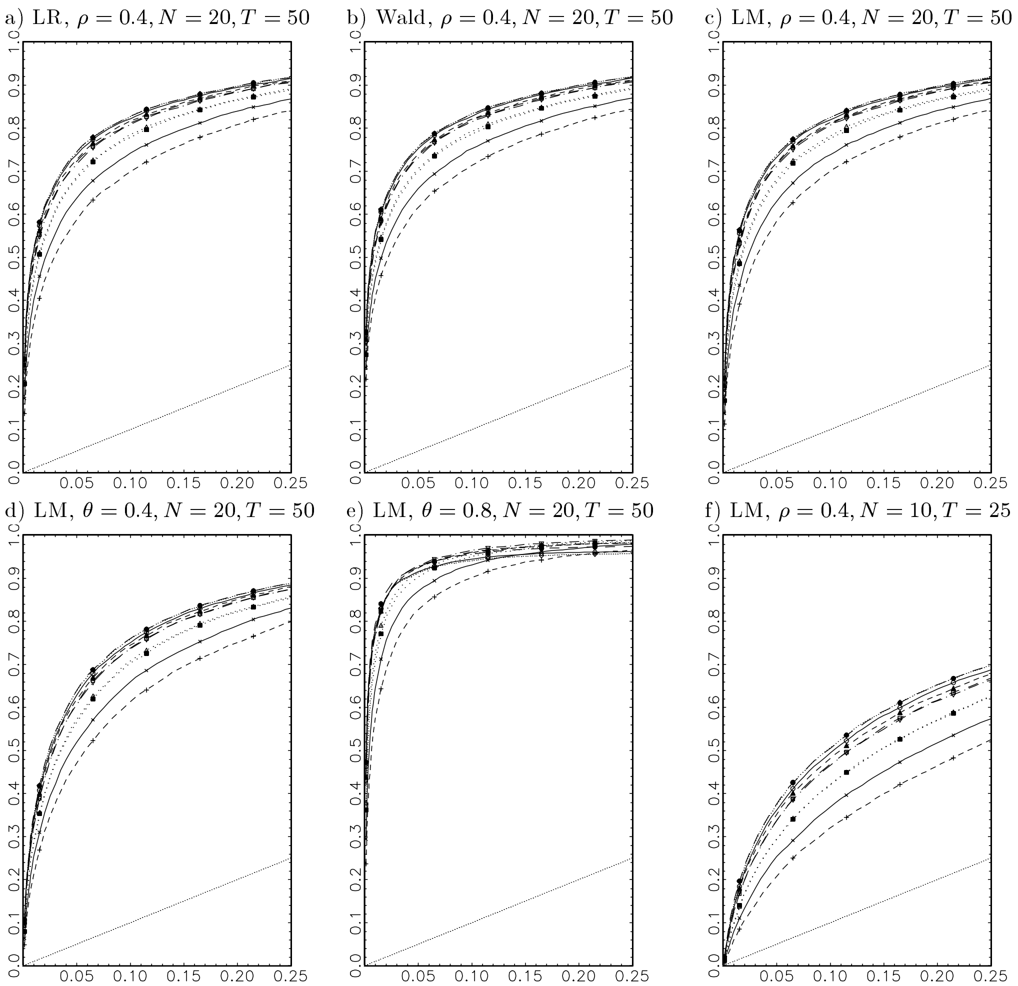
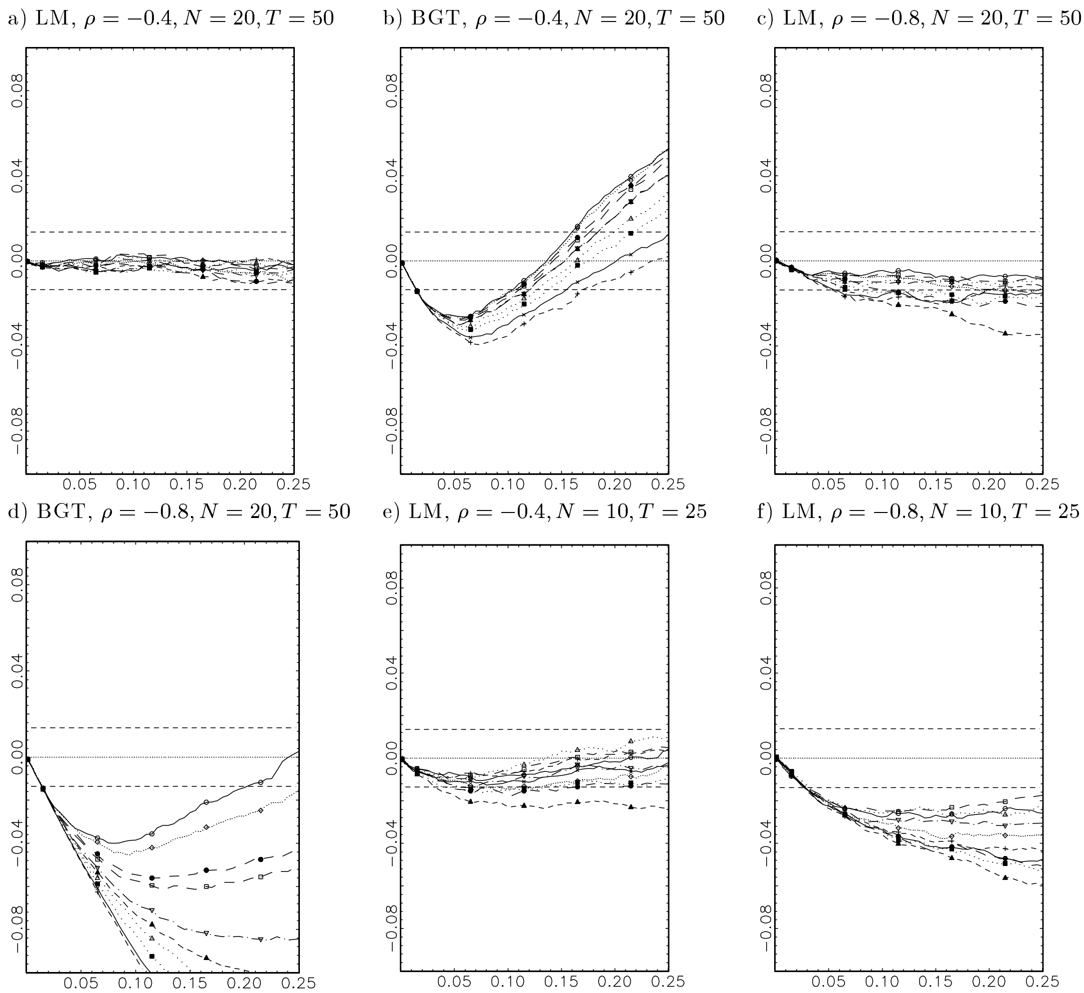


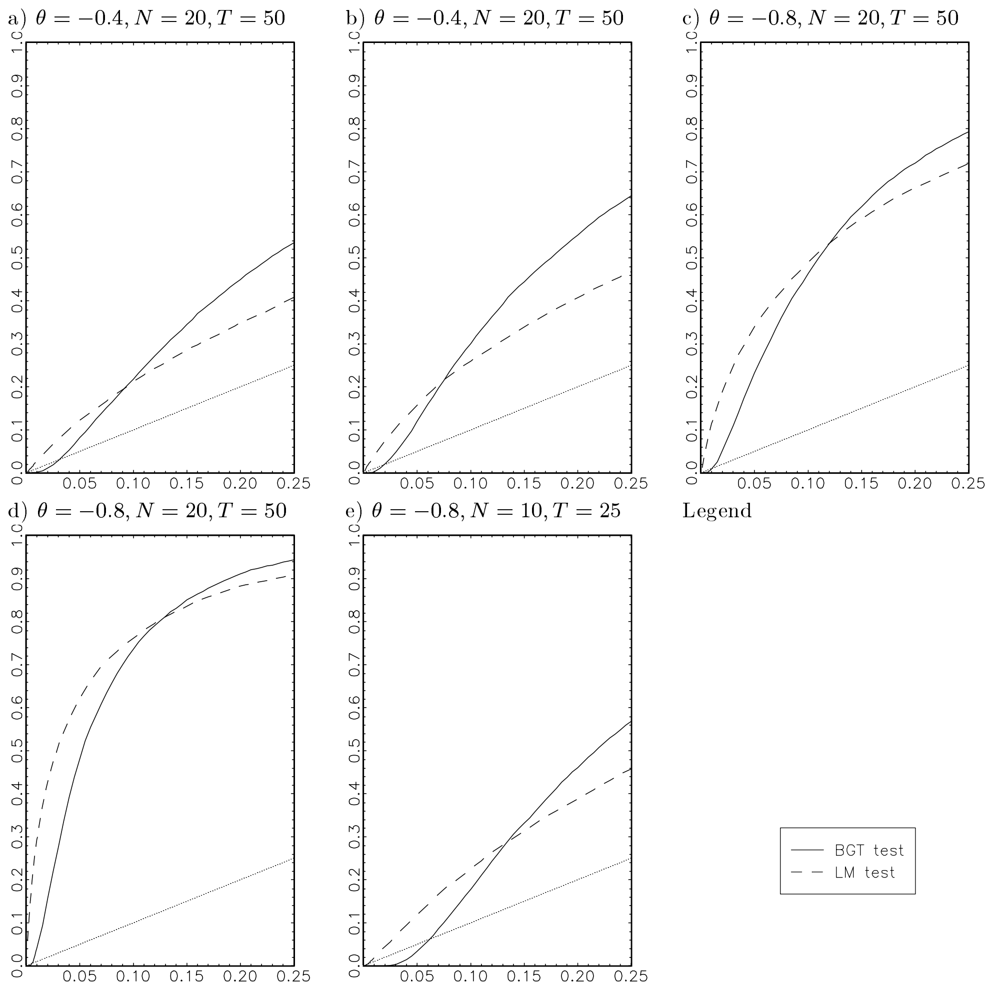
Figure 4 Size discrepancy of BGT-AR and LM-AR



the LM-AR test has correct size, and the size is insensitive to the variance parametrization. For $|\rho| = 0.8$ the LM-AR test is slightly undersized but still performs much better than the BGT-AR test.

Figure 4e-4f shows the size discrepancy of the LM-AR test for $N = 10, T = 25$ with $\rho = -0.4$ and $\rho = -0.8$. The picture is similar for positive values of ρ. The size properties of the BGT-AR test has not changed much for these smaller sample sizes. The LM-AR test is now undersized for $|\rho| = 0.4$ as well, but not by much. For $|\rho| = 0.8$ the size problem is more serious, but not as severe as for the BGT-AR test.

Figure 5 Power of BGT-AR and LM-AR



Power of BGT-AR and LM-AR For $N = 20, T = 50$ Figure 5a-5d compares the power functions for negative θ. Figure 5a and 5c with variance parameters $\sigma_\mu^2 = 1/6, \sigma_v^2 = 4/6$ and Figure 5b and 5d with $\sigma_\mu^2 = 2/6$ and $\sigma_v^2 = 1/6$. The LM-AR test is typically more powerful than the BGT-AR test at usual significance levels. In fact the power curves cross and the crossing point moves to the right with decreasing σ_λ^2. Similar to the tests of the null of no autocorrelation power is generally reduced for a low σ_λ^2 and high σ_v^2.

Figure 5e compares the power functions for $\theta = -0.8$ and $N = 10, T = 25$ with $\sigma_\mu^2 = 1/6$ and $\sigma_v^2 = 4/6$. The power of the LM-AR test is still higher than the BGT-AR test at usual significance levels. For $|\theta| = 0.4$ we have no useful power with either of these tests.

Size of BGT-MA and LM-MA Figure 6a-6d shows the size discrepancy of the BGT-MA and LM-MA tests for $N = 20, T = 50$. The size of the BGT-MA test is insensitive to the choice of variance parameters, however it is undersized with the more severe cases occurring for positive θ. Given the sign of θ, size is also unaffected by $|\theta| = 0.4$ or $|\theta| = 0.8$. The LM-MA test is also undersized but not by as much as the BGT-MA test, on the other hand it is slightly more sensitive to the variance parametrization for $|\theta| = 0.8$. The LM-MA test also has better size properties for negative θ.

For $N = 10, T = 25$ Figure 6e-6f shows the size discrepancy of the LM-MA test for positive values of θ. The size of the LM-MA test is quite sensitive to the choice of variance parameters, and undersized. The BGT-MA test continues to be insensitive to the choice of variance parameters. The size distortion is however still greater than for the LM-MA test.

Power of BGT-MA and LM-MA Figure 7a-7d shows the size-power curves for $N = 20, T = 50$ with $\sigma_\mu^2 = 1/6$ and $\sigma_v^2 = 4/6$. The power of the BGT-MA test is typically higher than the power of the LM-MA test at usual significance levels. At lower significance levels the power of the LM-MA test is higher and the crossing point of the power curves depends on σ_λ^2, specifically the crossing point moves to the right with decreasing σ_λ^2 as for the BGT-AR and LM-AR tests. Furthermore $|\rho| = 0.8$ is needed to get large power with either of these tests.

For $N = 10, T = 25$ the relative power properties are similar to the $N = 20, T = 50$ case, except that the crossing point of power curves occurs at higher significance levels in these smaller sample sizes. Figure 7e illustrates the crossing point for $\rho = -0.8$ with variance parameters $\sigma_\mu^2 = 1/6$ and $\sigma_v^2 = 4/6$. The LM-MA and BGT-MA tests have power equal to size at usual significance levels for $|\rho| = 0.4$.

4.4 Model selection

In the previous section we saw that for small sample sizes (small T) and/or small values of $|\rho|$ and $|\theta|$ test results for discrimination may very well be inconclusive. If a decision is needed we may have to resort to information criteria or discrimination based on p-values of the tests. Furthermore some researchers advocate the use of information criteria for model choice rather than hypothesis tests, see for example Granger, King and White (1995).

In this section we briefly consider the small-sample properties of model selection criteria for (i) the two-way model with λ_t an AR(1) or MA(1) process and (ii) overall model selection criteria for choosing between the standard two-way model and the two-way models (1, 2) and (1, 3). In the first case

Figure 6 Size discrepancy of BGT-MA and LM-MA

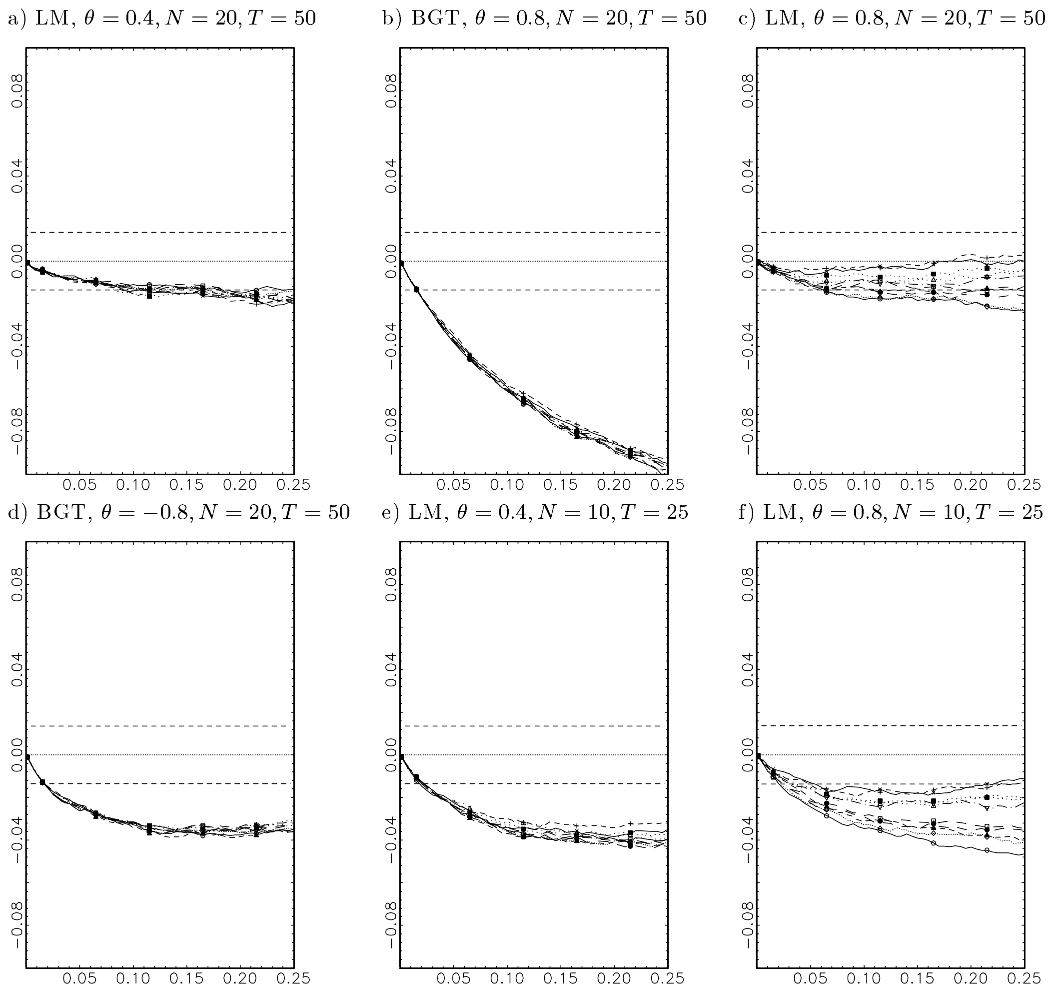


Figure 7 Power of BGT-MA and LM-MA

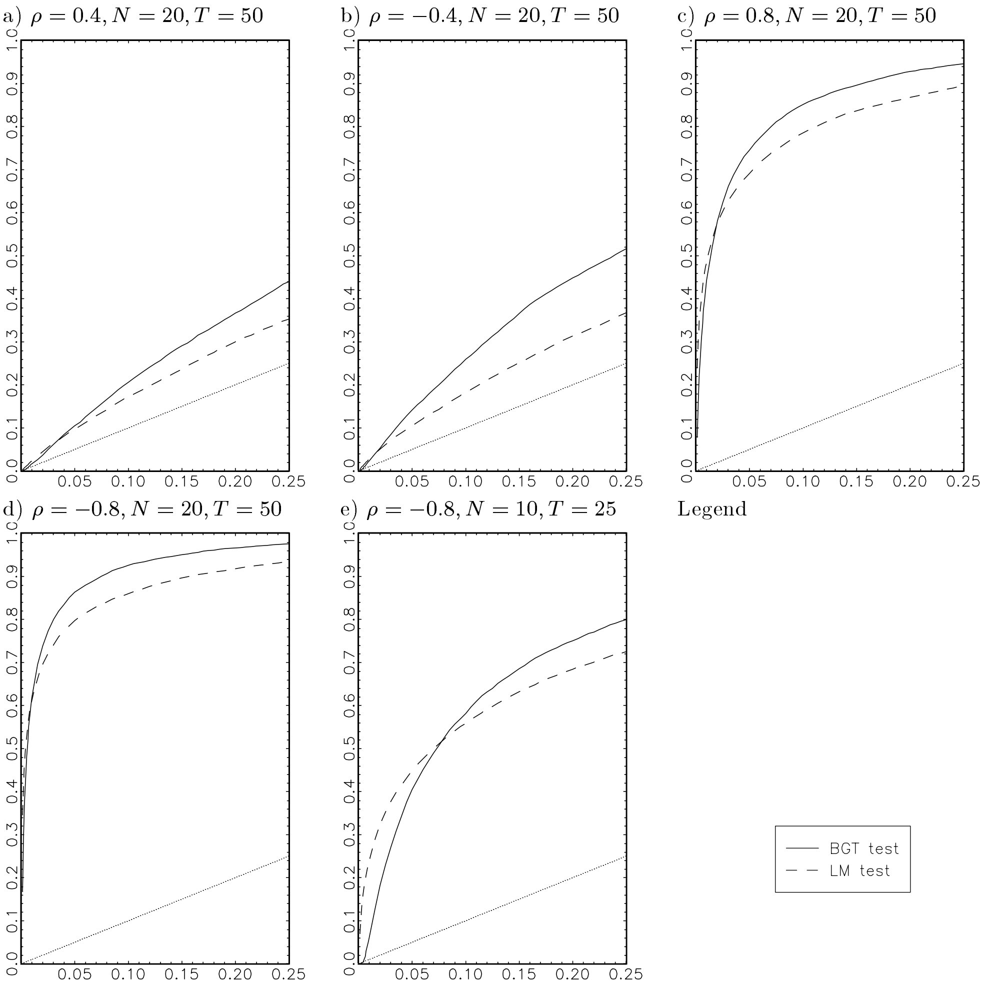


Table 1 Frequencies of correct classification of the AR(1) or MA(1) model, $\sigma_\mu^2 = 2/6$, $\sigma_v^2 = 2/6$, N=20, T=50

Model	LL	LM-p	LM-tests
$\rho = -0.8$	0.98	0.98	0.83
$\rho = -0.4$	0.70	0.68	0.14
$\rho = 0.4$	0.68	0.67	0.13
$\rho = 0.8$	0.96	0.96	0.77
$\theta = -0.8$	0.94	0.91	0.51
$\theta = -0.4$	0.66	0.65	0.14
$\theta = 0.4$	0.69	0.67	0.15
$\theta = 0.8$	0.96	0.93	0.59

the choice of model selection criteria to use is irrelevant and model choice can simply be based on a comparison of likelihoods of the two specifications, or p-values of the discriminating tests. In the second case the choice of model selection criteria matters and we consider the AIC criterion of Akaike (1974) and the BIC criterion of Schwarz (1978). These two criteria are compared to a hypothesis testing/p-value approach based on the LM-tests. In the first step we apply the LM-test of the null of no autocorrelation. If the null is not rejected at 5% significance level the standard two-way model is favored. If the null is rejected discrimination of the AR(1) and MA(1) process is based on the p-values of the discriminating LM-tests. We refer to this as the LM/LM-p strategy.

4.4.1 Discriminating between the AR(1) and MA(1) specifications

Let LL and LM-p denote discrimination based on comparing the log-likelihoods and p-values of the LM-tests respectively. We do not consider discrimination based on the p-values of the BGT-tests due to their disappointing size properties. In what follows discrimination with the LL criteria and the LM-p strategy is conditional on the LM-test of the null of no autocorrelation rejecting the null at the 5% level.

Table 1 shows the frequencies of correct classification of the AR(1) or MA(1) model for $N = 20, T = 50$ with $\sigma_\mu^2 = 2/6, \sigma_v^2 = 2/6$. For comparison we also include the frequencies of correct classification with the discriminating LM-tests, based on the 5% significance level.

For $N = 20, T = 50$ the LL criteria and the LM-p strategy are insensi-

Table 2 Classification frequencies for the standard two-way model (2-way), AR(1) and MA(1) models, $\sigma^2_\mu = 2/6$, $\sigma^2_v = 2/6$, N=20, T=50

Model	AIC			BIC			LM/LM-p		
	2-way	AR(1)	MA(1)	2-way	AR(1)	MA(1)	2-way	AR(1)	MA(1)
$\rho = -0.8$	0	0.98	0.02	0	0.98	0.02	0	0.98	0.02
$\rho = -0.4$	0.09	0.60	0.31	0.43	0.40	0.17	0.25	0.51	0.24
$\rho = 0.4$	0.13	0.56	0.31	0.49	0.35	0.16	0.29	0.48	0.23
$\rho = 0.8$	0	0.96	0.04	0.01	0.96	0.03	0	0.96	0.04
$\rho, \theta = 0$	0.81	0.08	0.11	0.99	0	0.01	0.95	0.03	0.02
$\theta = -0.8$	0	0.06	0.94	0.04	0.06	0.90	0.07	0.08	0.85
$\theta = -0.4$	0.12	0.25	0.63	0.52	0.15	0.33	0.40	0.22	0.40
$\theta = 0.4$	0.13	0.23	0.64	0.54	0.13	0.33	0.40	0.20	0.40
$\theta = 0.8$	0	0.04	0.96	0.03	0.04	0.93	0.07	0.06	0.87

tive to the choice of variance parameters. The LL criterion performs slightly better than the LM-p strategy. The rather low frequencies of correct classification for the LM-tests are mainly due to a large inconclusive region and illustrates the need to resort to the LL criteria or LM-p strategy if a decision must be made.

Corresponding frequencies for $N = 10, T = 25$ are obviously lower, but also more sensitive to variance parametrization. For example, the frequencies of correct classification with the LL criteria and LM-p strategy are only slightly above 0.5 for some variance parametrizations (low σ^2_λ and high σ^2_v) with a small $|\rho|$ or $|\theta|$.

4.4.2 Overall model selection

As for the LL criteria and LM-p strategy considered above the AIC and BIC criteria and the LM/LM-p strategy are more or less sensitive to variance parametrization. Generally the performance deteriorate with decreasing σ^2_λ and increasing σ^2_v.

Table 2 shows the classification frequencies for the standard two-way model (2-way), AR(1) and MA(1) models for $N = 20, T = 50$ with $\sigma^2_\mu = 2/6, \sigma^2_v = 2/6$.

BIC favors the standard two-way model whereas AIC favors the AR(1) or MA(1) model. This behavior is expected since the BIC criterion penalize extra parameters harder than AIC. The LM/LM-p strategy is typically intermediate to AIC and BIC in performance.

For $N = 10, T = 25$ frequencies of correct classification of the AR(1) and MA(1) models are lower, but the relative performance of the AIC and BIC criteria and the LM/LM-p strategy is similar to the $N = 20, T = 50$ case.

5 Conclusions

In this paper we have derived a straightforward maximum likelihood estimator of the two-way model with a serially correlated time-specific effect. In addition we have considered specification tests as well as various model selection strategies.

When testing for the null of no serial correlation we recommend the LM, Wald (based on Hessian) and LR-tests against AR(1) since they have the best size properties. Furthermore the power loss to the corresponding Wald and LR-tests against MA(1) is small. In practice the LM test may be preferred since it is simple to compute, requiring only estimation under the null hypothesis of the standard two-way model.

To discriminate between the AR(1) and MA(1) process we have considered LM-tests as well as tests requiring only the within estimates of the standard two-way model. The LM-AR test typically performs better than the BGT-AR test. The size of the LM-AR test is not so sensitive to the choice of variance parameters as the BGT-AR test and the LM-AR test has the highest power at usual significance levels. In contrast the BGT-MA test is less sensitive to variance parametrization than the LM-MA test and typically has the highest power at usual significance levels. We can however not recommend the BGT-MA test due to its disappointing size properties. Large values of $|\rho|$ or $|\theta|$ are needed for discrimination with these tests and test results may very well be inconclusive. One possible way to "split the tie" is to simply compare likelihoods or p-values of tests. Of these the likelihood comparison performs best.

Model selection can also be used to discriminate between the standard two-way model and the two-way model with λ_t and AR(1) or MA(1) process. We have considered model selection based on the AIC and BIC criterions as well as an LM/LM–p strategy. The AIC criterion performs best when AR(1) or MA(1) is the true process. BIC favors the standard two-way model and the LM/LM-p strategy is typically intermediate in performance. When the standard two-way model is the true model the ranking is reversed.

A Score and Information

A.1 The score vector

This appendix derives the elements of the score vector for the models (1, 2) and (1, 3). For the regression parameters we have the standard result

$$\frac{\partial l}{\partial \boldsymbol{\delta}} = \mathbf{Z}'\Sigma^{-1}\boldsymbol{\varepsilon}$$

and for the variance parameters the score is given by

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

where $\boldsymbol{\gamma} = \left(\sigma_\mu^2, \sigma_\nu^2, \sigma_u^2, \rho\right)'$ for (2) and $\left(\sigma_\mu^2, \sigma_\nu^2, \sigma_u^2, \theta\right)'$ for (3).

For σ_μ^2 we have

$$\begin{aligned}
\operatorname{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \sigma_\mu^2}\right) &= \operatorname{tr}\left(\Sigma^{-1}\left(\mathbf{I}_N \otimes \mathbf{J}_T\right)\right) \\
&= \operatorname{tr}\left(\mathbf{I}_N \otimes \mathbf{A}^*\mathbf{J}_T\right) - \operatorname{tr}\left[\left(\boldsymbol{\iota}_N \otimes \mathbf{A}^*\right)\mathbf{B}^{-1}\left(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*\mathbf{J}_T\right)\right] \\
&= \frac{1}{\sigma_1^2}\operatorname{tr}\left(\mathbf{I}_N \otimes \mathbf{J}_T\right) - \frac{N}{\sigma_1^4}\operatorname{tr}\left[\mathbf{B}^{-1}\mathbf{J}_T\right] \\
&= \frac{NT}{\sigma_1^2} - \frac{N}{\sigma_1^4}\boldsymbol{\iota}_T'\mathbf{B}^{-1}\boldsymbol{\iota}_T
\end{aligned}$$

where $B = \sigma_u^{-2}\Psi^{-1} + NA^*$, $B^{-1} = \sigma_u^2\left(\mathbf{I}_T + N\sigma_u^2\Psi\mathbf{A}^*\right)^{-1}\Psi$ and

$$\begin{aligned}
\Sigma^{-1}\frac{\partial \Sigma}{\partial \sigma_\mu^2}\Sigma^{-1} &= \Sigma^{-1}(\mathbf{I}_N \otimes \mathbf{J}_T)\Sigma^{-1} \\
&= \mathbf{I}_N \otimes \mathbf{A}^*\mathbf{J}_T\mathbf{A}^* - (\boldsymbol{\iota}_N \otimes \mathbf{A}^*\mathbf{J}_T\mathbf{A}^*)\mathbf{B}^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \\
&\quad -(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\mathbf{B}^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*\mathbf{J}_T\mathbf{A}^*) \\
&\quad +(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\mathbf{B}^{-1}(\boldsymbol{\iota}_N'\boldsymbol{\iota}_N \otimes \mathbf{A}^*\mathbf{J}_T\mathbf{A}^*)\mathbf{B}^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \\
&= \frac{1}{\sigma_1^4}(\mathbf{I}_N \otimes \mathbf{J}_T) - \frac{1}{\sigma_1^4}(\boldsymbol{\iota}_N \otimes \mathbf{J}_T)\mathbf{B}^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \\
&\quad -\frac{1}{\sigma_1^4}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)\mathbf{B}^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{J}_T) + \frac{N}{\sigma_1^4}(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\mathbf{B}^{-1}\mathbf{J}_T\mathbf{B}^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*).
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{\partial l}{\partial \sigma_\mu^2} &= -\frac{NT}{2\sigma_1^2} + \frac{N}{2\sigma_1^4}\boldsymbol{\iota}_T'\mathbf{B}^{-1}\boldsymbol{\iota}_T \\
&\quad +\frac{1}{2\sigma_1^4}\textstyle\sum_{i=1}^N\sum_{t=1}^T\varepsilon_{it}^2 - \frac{1}{\sigma_1^4}\widetilde{\boldsymbol{\varepsilon}}'\mathbf{B}^{-1}\overline{\boldsymbol{\varepsilon}} + \frac{N}{2\sigma_1^4}\widetilde{\boldsymbol{\varepsilon}}'\mathbf{B}^{-1}\mathbf{J}_T\mathbf{B}^{-1}\widetilde{\boldsymbol{\varepsilon}}
\end{aligned}$$

where $\overline{\varepsilon} = (\iota'_N \otimes J_T)\varepsilon$ and $\widetilde{\boldsymbol{\varepsilon}} = (\iota'_N \otimes A^*)\varepsilon$.

For σ^2_ν we have

$$\begin{aligned}
\operatorname{tr}(\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma^2_v}) &= \operatorname{tr}(\Sigma^{-1}) = N\operatorname{tr}(\mathbf{A}^*) - N\operatorname{tr}\left[(\mathbf{A}^*)^2\mathbf{B}^{-1}\right] \\
&= \frac{N}{\sigma^2_1} + \frac{N(T-1)}{\sigma^2_v} - \left(\frac{N}{T\sigma^4_1} - \frac{N}{T\sigma^4_v}\right)\boldsymbol{\iota}'_T\mathbf{B}^{-1}\boldsymbol{\iota}_T - \frac{N}{\sigma^4_v}\operatorname{tr}\mathbf{B}^{-1}, \\
\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma^2_v}\Sigma^{-1} &= \Sigma^{-2} = \mathbf{I}_N \otimes (\mathbf{A}^*)^2 - \mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}^{-1}(\mathbf{A}^*)^2 \\
&\quad -\mathbf{J}_N \otimes (\mathbf{A}^*)^2\mathbf{B}^{-1}\mathbf{A}^* + N\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}^{-1}(\mathbf{A}^*)^2\mathbf{B}^{-1}\mathbf{A}^*
\end{aligned}$$

with

$$\begin{aligned}
\frac{\partial l}{\partial\sigma^2_\nu} &= -\frac{N}{2\sigma^2_1} - \frac{N(T-1)}{2\sigma^2_v} + \frac{1}{2}\left(\frac{N}{T\sigma^4_1} - \frac{N}{T\sigma^4_v}\right)\boldsymbol{\iota}'_T\mathbf{B}^{-1}\boldsymbol{\iota}_T + \frac{N}{2\sigma^4_v}\operatorname{tr}\mathbf{B}^{-1} \\
&\quad +\frac{1}{2}\boldsymbol{\varepsilon}'\left[\mathbf{I}_N \otimes (\mathbf{A}^*)^2\right]\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{*\prime}\widetilde{\boldsymbol{\varepsilon}} + \frac{N}{2}\boldsymbol{\varepsilon}^{*\prime}\boldsymbol{\varepsilon}^*
\end{aligned}$$

where $\varepsilon^* = (\boldsymbol{\iota}'_N \otimes \mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*)\,\varepsilon$.

For σ^2_u we have

$$\begin{aligned}
\operatorname{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma^2_u}\right) &= \operatorname{tr}(\Sigma^{-1}(\mathbf{J}_N \otimes \Psi)) = N\operatorname{tr}(\mathbf{A}^*\Psi) - N^2\operatorname{tr}(\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{B}^{-1}), \\
\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma^2_u}\Sigma^{-1} &= \Sigma^{-1}(\mathbf{J}_N \otimes \Psi)\Sigma^{-1} = \mathbf{J}_N \otimes \mathbf{A}^*\Psi\mathbf{A}^* - N(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*) \\
&\quad -N(\mathbf{J}_N \otimes \mathbf{A}^*\Psi\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*) + N^2(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial l}{\partial\sigma^2_u} &= -\frac{N}{2}\operatorname{tr}(\mathbf{A}^*\Psi) + \frac{N^2}{2}\operatorname{tr}(\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{B}^{-1}) \\
&\quad +\frac{1}{2}\widetilde{\boldsymbol{\varepsilon}}'\boldsymbol{\varepsilon}^\Psi - N\boldsymbol{\varepsilon}^{*\prime}\boldsymbol{\varepsilon}^\Psi + \frac{N^2}{2}\boldsymbol{\varepsilon}^{*\prime}(\mathbf{I}_N \otimes \Psi)\,\boldsymbol{\varepsilon}^*
\end{aligned}$$

where $\varepsilon^\Psi = (\boldsymbol{\iota}'_N \otimes \Psi\mathbf{A}^*)\,\varepsilon$.

For the fourth and last variance parameter, γ_4, we have

$$\frac{\partial\Sigma}{\partial\gamma_4} = \sigma^2_u\left(\mathbf{J}_N \otimes \frac{\partial\Psi}{\partial\gamma_4}\right) = \sigma^2_u\,(\mathbf{J}_N \otimes \mathbf{L})$$

with $L_\rho = \frac{\partial\Psi_\rho}{\partial\rho} = \frac{2\rho}{(1-\rho^2)^2}\Psi + \frac{1}{1-\rho^2}D$ where D is a band matrix with zeros on the main diagonal and $i\rho^{i-1}$ on the i^{th} subdiagonal for the AR(1) specification

(2) and $L_\theta = \frac{\partial \Psi_\theta}{\partial \theta}$ a bidiagonal matrix with 2θ on the main diagonal and ones on the subdiagonals for the MA(1) specification (3). This gives

$$\begin{aligned}\frac{\partial l}{\partial \gamma_4} &= \sigma_u^2(-\frac{N}{2}\operatorname{tr}(\mathbf{A}^*\mathbf{L}) + \frac{N^2}{2}\operatorname{tr}(\mathbf{A}^*\mathbf{L}\mathbf{A}^*\mathbf{B}^{-1}) \\ &\quad + \frac{1}{2}\widetilde{\boldsymbol{\varepsilon}}'\boldsymbol{\varepsilon}^L - N\boldsymbol{\varepsilon}^{*\prime}\boldsymbol{\varepsilon}^L + \frac{N^2}{2}\boldsymbol{\varepsilon}^{*\prime}\left(\mathbf{I}_N \otimes \mathbf{L}\right)\boldsymbol{\varepsilon}^*)\end{aligned}$$

with the appropriate L matrix and $\varepsilon^L = \left(\boldsymbol{\iota}_N' \otimes \mathbf{L}\mathbf{A}^*\right)\varepsilon$.

A.2 The information matrix

This appendix derives the elements of the information matrix for the models (1, 2) and (1, 3). For the first element we have the result

$$\mathcal{I}_{\delta,\delta} = \mathbf{Z}'\Sigma^{-1}\mathbf{Z}$$

and for the elements $I_{\delta,\gamma}$ we have the familiar block-diagonality result

$$E[-\frac{\partial}{\partial \gamma}\mathbf{Z}'\Sigma^{-1}\boldsymbol{\varepsilon}] = E[\mathbf{Z}'\Sigma^{-1}\frac{\partial \Sigma}{\partial \gamma}\Sigma^{-1}\boldsymbol{\varepsilon}] = 0$$

The elements of the information matrix for the γ parameters are obtained as

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

We have for the I_{γ_1,γ_j} elements

$$\begin{aligned}\mathcal{I}_{\gamma_1,\gamma_1} &= \frac{1}{2}\operatorname{tr}[\Sigma^{-1}(\frac{\partial \Sigma}{\partial \gamma_1})\Sigma^{-1}(\frac{\partial \Sigma}{\partial \gamma_1})] \\ &= \frac{1}{2(\sigma_1^2)^2}[NT^2 - \frac{2NT}{\sigma_1^2}\left(\boldsymbol{\iota}_T'\mathbf{B}^{-1}\boldsymbol{\iota}_T\right) + \frac{N^2}{(\sigma_1^2)^2}\left(\boldsymbol{\iota}_T'\mathbf{B}^{-1}\boldsymbol{\iota}_T\right)]\end{aligned}$$

where B^{-1} is defined in appendix A.1.

$$\begin{aligned}\mathcal{I}_{\gamma_1,\gamma_3} &= \frac{1}{2}\operatorname{tr}[\Sigma^{-1}(\frac{\partial \Sigma}{\partial \gamma_1})\Sigma^{-1}(\frac{\partial \Sigma}{\partial \gamma_3})] \\ &= \frac{1}{2(\sigma_1^2)^2}[N\left(\boldsymbol{\iota}_T'\Psi\boldsymbol{\iota}_T\right) - 2N^2\left(\boldsymbol{\iota}_T'\mathbf{B}^{-1}\mathbf{A}^*\Psi\boldsymbol{\iota}_T\right) \\ &\quad + N^3\left(\boldsymbol{\iota}_T'\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T\right)]\end{aligned}$$

$$
\begin{aligned}
\mathcal{I}_{\gamma_1,\gamma_4} &= \frac{1}{2}\,\mathrm{tr}[\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_1})\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_4})] \\
&= \frac{\sigma_u^2}{2(\sigma_1^2)^2}[N\,(\boldsymbol{\iota}_T'\mathbf{L}\boldsymbol{\iota}_T) - 2N^2\left(\boldsymbol{\iota}_T'\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\boldsymbol{\iota}_T\right) \\
&\quad + N^3\left(\boldsymbol{\iota}_T'\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T\right)]
\end{aligned}
$$

where L is L_ρ or L_θ defined in appendix A.1. For the relevant I_{γ_2,γ_j} elements

$$
\begin{aligned}
\mathcal{I}_{\gamma_2,\gamma_2} &= \frac{1}{2}\,\mathrm{tr}[\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_2})\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_2})] \\
&= \frac{1}{2}[N\,\mathrm{tr}(\mathbf{A}^*)^2 - 2N\,\mathrm{tr}(\mathbf{B}^{-1}(\mathbf{A}^*)^3) \\
&\quad + N^2\,\mathrm{tr}(\mathbf{B}^{-1}(\mathbf{A}^*)^2\mathbf{B}^{-1}\,(\mathbf{A}^*)^2)]
\end{aligned}
$$

where the r:th matrix power denote multiplication of the matrix with itself r times.

$$
\begin{aligned}
\mathcal{I}_{\gamma_2,\gamma_3} &= \frac{1}{2}\,\mathrm{tr}[\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_2})\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_3})] \\
&= \frac{1}{2}[N\,\mathrm{tr}(\mathbf{A}^*\Psi\mathbf{A}^*) - 2N^2\,\mathrm{tr}(\mathbf{B}^{-1}(\mathbf{A}^*)^2\Psi\mathbf{A}^*) \\
&\quad + N^3\,\mathrm{tr}(\mathbf{B}^{-1}(\mathbf{A}^*)^2\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*)]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{I}_{\gamma_2,\gamma_4} &= \frac{1}{2}\,\mathrm{tr}[\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_2})\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_4})] \\
&= \frac{\sigma_u^2}{2}[N\,\mathrm{tr}(\mathbf{A}^*\mathbf{L}\mathbf{A}^*) - 2N^2\,\mathrm{tr}(\mathbf{B}^{-1}(\mathbf{A}^*)^2\mathbf{L}\mathbf{A}^*) \\
&\quad + N^3\,\mathrm{tr}(\mathbf{B}^{-1}(\mathbf{A}^*)^2\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*)]
\end{aligned}
$$

Finally for the elements involving γ_3, γ_4 we have

$$
\begin{aligned}
\mathcal{I}_{\gamma_3,\gamma_3} &= \frac{1}{2}\,\mathrm{tr}[\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_3})\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_3})] \\
&= \frac{1}{2}[N^2\,\mathrm{tr}((\mathbf{A}^*\Psi)^2) - 2N^3\,\mathrm{tr}(\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*\Psi\mathbf{A}^*) \\
&\quad + N^4\,\mathrm{tr}((\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*)^2)]
\end{aligned}
$$

$$\mathcal{I}_{\gamma_3,\gamma_4} = \frac{1}{2}\operatorname{tr}[\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_3})\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_4})]$$
$$= \frac{\sigma_u^2}{2}[N^2\operatorname{tr}(\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{L}) - 2N^3\operatorname{tr}(\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}) + N^4\operatorname{tr}(\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*)]$$

$$\mathcal{I}_{\gamma_4,\gamma_4} = \frac{1}{2}\operatorname{tr}[\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_4})\Sigma^{-1}(\frac{\partial\Sigma}{\partial\gamma_4})]$$
$$= \frac{\sigma_u^4}{2}[N^2\operatorname{tr}((\mathbf{A}^*\mathbf{L})^2) - 2N^3\operatorname{tr}(\mathbf{B}^{-1}(\mathbf{A}^*\mathbf{L})^2\mathbf{A}^*) + N^4\operatorname{tr}((\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*)^2)]$$

B The LM-test against ARMA(1,1)

This appendix derives the score and information matrix for the LM test against ARMA(1,1). Under ARMA(1,1) disturbances we have the covariance matrix as

$$\Sigma_1 = \mathbf{A} + \boldsymbol{\sigma}_u^2(\boldsymbol{\iota}_N \otimes \mathbf{I}_T)\Gamma(\boldsymbol{\iota}_N' \otimes \mathbf{I}_T)$$

with inverse

$$\Sigma_1^{-1} = \mathbf{I}_N \otimes \mathbf{A}^* - (\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\mathbf{B}^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)$$

where $B_1^{-1} = \sigma_u^2\left(\mathbf{I}_T + N\sigma_u^2\Gamma\mathbf{A}^*\right)^{-1}\Gamma$ and Γ is covariance matrix of ARMA(1,1) process with elements $\Gamma_{tt} = \frac{1+\theta^2+2\theta\rho}{1-\rho^2}$ and $\Gamma_{ts} = \frac{(\rho+\theta)(1+\rho\theta)\rho^{|t-s|-1}}{1-\rho^2}$ for $t \neq s$. To derive the LM-test we need the score and information matrix evaluated under the null hypothesis $\theta = 0$ or $\rho = 0$. The score needed is given by

$$\frac{\partial l}{\partial\tau}|_{\tau=0} = -\frac{1}{2}\operatorname{tr}(\Sigma_{1,\tau=0}^{-1}(\frac{\partial\Sigma_1}{\partial\tau}|_{\tau=0})) + \frac{1}{2}\boldsymbol{\varepsilon}'\Sigma_{1,\tau=0}^{-1}(\frac{\partial\Sigma_1}{\partial\tau}|_{\tau=0})\Sigma_{1,\tau=0}^{-1}\boldsymbol{\varepsilon}$$

where $\tau = \theta$ or $\tau = \rho$ and

$$\Sigma_{1,\tau=0}^{-1} = \Sigma^{-1} = \mathbf{I}_N \otimes \mathbf{A}^* - \sigma_u^2(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\left[\mathbf{I}_T + N\sigma_u^2\Psi\mathbf{A}^*\right]^{-1}\Psi(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)$$

where Ψ is given by Ψ_ρ if $\tau = \theta$ and Ψ_θ if $\tau = \rho$. Also $\frac{\partial \Sigma_1}{\partial \tau}|_{\tau=0} = \sigma_u^2 \left(\mathbf{J}_N \otimes \frac{\partial \Gamma}{\partial \tau} \right)|_{\tau=0} = \sigma_u^2 \left(\mathbf{J}_N \otimes \mathbf{K}_\tau \right)$ where K_θ has $\frac{2\rho}{1-\rho^2}$ on the main diagonal and $\frac{(\rho^2+1)\rho^{|t-s|-1}}{1-\rho^2}$ on the off-diagonal elements and K_ρ has 2θ on the main diagonal, $1+\theta^2$ on the subdiagonal and θ on the subsubdiagonal. We get

$$\begin{aligned}
\operatorname{tr}(\Sigma^{-1}\frac{\partial \Sigma_1}{\partial \tau}|_{\tau=0}) &= \operatorname{tr}(\Sigma^{-1}\sigma_u^2 \left(\mathbf{J}_N \otimes \mathbf{K}_\tau \right)) \\
&= N\sigma_u^2 \operatorname{tr}(\mathbf{A}^*\mathbf{K}_\tau) - N^2\sigma_u^2 \operatorname{tr}(\mathbf{A}^*\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}) \\
\Sigma^{-1}\frac{\partial \Sigma_1}{\partial \tau}|_{\tau=0}\Sigma^{-1} &= \Sigma^{-1}\sigma_u^2 \left(\mathbf{J}_N \otimes \mathbf{K}_\tau \right) \Sigma^{-1} \\
&= \sigma_u^2((\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{K}_\tau\mathbf{A}^*) - N(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*) \\
&\quad -N(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{K}_\tau\mathbf{A}^*) + N^2(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*))
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial l}{\partial \tau}|_{\tau=0} &= \frac{\sigma_u^2}{2}(-N \operatorname{tr}(\mathbf{A}^*\mathbf{K}_\tau) + N^2 \operatorname{tr}(\mathbf{A}^*\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}) \\
&\quad + \widetilde{\boldsymbol{\varepsilon}}'\boldsymbol{\varepsilon}^{K_\tau} - N\boldsymbol{\varepsilon}^{*\prime}\boldsymbol{\varepsilon}^{K_\tau} + N^2\boldsymbol{\varepsilon}^{*\prime}\left(\mathbf{I}_N \otimes \mathbf{K}_\tau\right)\boldsymbol{\varepsilon}^*)
\end{aligned}$$

where $\widetilde{\boldsymbol{\varepsilon}}$, ε^* are defined in appendix A.1 and $\varepsilon^{K_\tau} = \left(\boldsymbol{\iota}_N' \otimes \mathbf{K}_\tau\mathbf{A}^*\right)\varepsilon$. The information matrix evaluated under the null hypothesis is obtained as

$$\mathcal{I}_{\gamma_i,\gamma_j} = \frac{1}{2}\operatorname{tr}[\Sigma^{-1}(\frac{\partial \Sigma_1}{\partial \gamma_i}|_{\tau=0})\Sigma^{-1}(\frac{\partial \Sigma_1}{\partial \gamma_j}|_{\tau=0})]$$

where γ now is defined as $\gamma = (\sigma_\mu^2, \sigma_\nu^2, \sigma_u^2, \rho, \theta)'$. By noting that $(\frac{\partial \Sigma_1}{\partial \gamma_i}|_{\tau=0}) = \frac{\partial \Sigma}{\partial \gamma_i}$ for elements not involving τ the only elements needed apart from those derived in appendix A.2 are those containing τ. We have

$$\begin{aligned}
\mathcal{I}_{\gamma_1,\tau} &= \frac{1}{2}tr[\Sigma^{-1}(\frac{\partial \Sigma_1}{\partial \gamma_1}|_{\tau=0})\Sigma^{-1}(\frac{\partial \Sigma_1}{\partial \tau}|_{\tau=0})] \\
&= \frac{\sigma_u^2}{2(\sigma_1^2)^2}[N\left(\boldsymbol{\iota}_T'\mathbf{K}_\tau\boldsymbol{\iota}_T\right) - 2N^2\left(\boldsymbol{\iota}_T'\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T\right) \\
&\quad + N^3\left(\boldsymbol{\iota}_T'\mathbf{A}^*\mathbf{B}^{-1}\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T\right)]
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}_{\gamma_2,\tau} &= \frac{1}{2}tr[\Sigma^{-1}(\frac{\partial \Sigma_1}{\partial \gamma_2}|_{\tau=0})\Sigma^{-1}(\frac{\partial \Sigma_1}{\partial \tau}|_{\tau=0})] \\
&= \frac{\sigma_u^2}{2}[N \operatorname{tr}((\mathbf{A}^*)^2\mathbf{K}_\tau) - 2N^2 \operatorname{tr}(\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}(\mathbf{A}^*)^2) \\
&\quad + N^3 \operatorname{tr}(\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}(\mathbf{A}^*)^2\mathbf{B}^{-1}\mathbf{A}^*)]
\end{aligned}$$

$$
\begin{aligned}
\mathcal{I}_{\gamma_3,\tau} &= \frac{1}{2}tr[\Sigma^{-1}(\frac{\partial\Sigma_1}{\partial\gamma_3}|_{\tau=0})\Sigma^{-1}(\frac{\partial\Sigma_1}{\partial\tau}|_{\tau=0})] \\
&= \frac{\sigma_u^2}{2}[N^2\,\mathrm{tr}(\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{K}_\tau) - 2N^3\,\mathrm{tr}(\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*) \\
&\quad + N^4\,\mathrm{tr}(\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\Psi\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*)]
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{I}_{\gamma_4,\tau} &= \frac{1}{2}tr[\Sigma^{-1}(\frac{\partial\Sigma_1}{\partial\gamma_4}|_{\tau=0})\Sigma^{-1}(\frac{\partial\Sigma_1}{\partial\tau}|_{\tau=0})] \\
&= \frac{\sigma_u^4}{2}[N^2\,\mathrm{tr}(\mathbf{A}^*\mathbf{L}\mathbf{A}^*\mathbf{K}_\tau) - 2N^3\,\mathrm{tr}(\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*) \\
&\quad + N^4\,\mathrm{tr}(\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*)]
\end{aligned}
$$

where L is L_ρ for $\tau = \theta$ and L_θ for $\tau = \rho$

$$
\begin{aligned}
\mathcal{I}_{\tau,\tau} &= \frac{1}{2}tr[\Sigma^{-1}(\frac{\partial\Sigma_1}{\partial\tau}|_{\tau=0})\Sigma^{-1}(\frac{\partial\Sigma_1}{\partial\tau}|_{\tau=0})] \\
&= \frac{\sigma_u^4}{2}[N^2\,\mathrm{tr}((\mathbf{A}^*\mathbf{K}_\tau)^2) - 2N^3\,\mathrm{tr}(\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{K}_\tau\mathbf{A}^*) \\
&\quad + N^4\,\mathrm{tr}((\mathbf{K}_\tau\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*)^2)]
\end{aligned}
$$

C Proof of theorem 1

It is trivial to show that the information matrix is block-diagonal

$$
\mathcal{I} = \begin{pmatrix} \mathcal{I}_\delta & \\ & \mathcal{I}_\gamma \end{pmatrix}
$$

and that I_δ is of full rank under standard assumptions on the explanatory variables. The information matrix of the variance parameters is for the i, j element

$$
\begin{aligned}
&\frac{1}{2}\,\mathrm{tr}\left[\Sigma^{-1}\left(\frac{\partial\Sigma}{\partial\gamma_i}\right)\Sigma^{-1}\left(\frac{\partial\Sigma}{\partial\gamma_j}\right)\right] \\
=\ &\frac{1}{2}\left[\mathrm{vec}\left(\frac{\partial\Sigma}{\partial\gamma_i}\right)'\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)\mathrm{vec}\left(\frac{\partial\Sigma}{\partial\gamma_j}\right)\right]
\end{aligned}
$$

The conditions on the γ parameters ensures that Σ is of full rank. That I is of full column rank then follows from the full column rank of

$$
\mathbf{W} = \left[\mathrm{vec}\left(\frac{\partial\Sigma}{\partial\sigma_\mu^2}\right), \mathrm{vec}\left(\frac{\partial\Sigma}{\partial\sigma_v^2}\right), \mathrm{vec}\left(\frac{\partial\Sigma}{\partial\sigma_u^2}\right), \mathrm{vec}\left(\frac{\partial\Sigma}{\partial\tau}\right)\right]
$$

Suppose there exists a vector $a \neq 0$ $s.t. Wa = 0$, then this must also hold for the submatrix W^* consisting of rows $1, 2, T+1$ and $T+2$ of W. For $\tau = \rho$

$$\mathbf{W}^* = \begin{pmatrix} 1 & 1 & \frac{1}{1-\rho^2} & \frac{2\sigma_u^2\rho}{(1-\rho^2)^2} \\ 1 & 0 & \frac{\rho}{1-\rho^2} & \frac{2\sigma_u^2\rho}{(1-\rho^2)^2} + \frac{\sigma_u^2}{1-\rho^2} \\ 0 & 0 & \frac{1}{1-\rho^2} & \frac{2\sigma_u^2\rho}{(1-\rho^2)^2} \\ 0 & 0 & \frac{\rho}{1-\rho^2} & \frac{2\sigma_u^2\rho}{(1-\rho^2)^2} + \frac{\sigma_u^2}{1-\rho^2} \end{pmatrix}$$

For $W^*a = 0$ it is clear that we must have $a_1 = a_2 = 0$ and that a_3 and a_4 are determined by rows 3 and 4. When $\rho = 0, W^*a = 0$ iff $a_3 = a_4 = 0$ as well giving a contradiction. For $\rho \neq 0$ we normalize a_4 to 1 and use row 3 to obtain $a_3 = \frac{-2\sigma_u^2\rho}{(1-\rho^2)^2}$. Substituting into row 4 yields

$$\frac{-2\sigma_u^2\rho}{(1-\rho^2)^2} + \frac{2\sigma_u^2\rho}{(1-\rho^2)^2} + \frac{\sigma_u^2}{1-\rho^2} > 0$$

which again contradicts the premise. The proof is similar for $\tau = \theta$.

References

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

Anderson, T. W. and Hsiao, C. (1982), 'Formulation and estimation of dynamic models using panel data', *Journal of Econometrics* **18**, 47–82.

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.

Baltagi, B. H. and Li, Q. (1995), 'Testing AR(1) against MA(1) disturbances in an error component model', *Journal of Econometrics* **68**, 133–151.

Davidson, R. and McKinnon, J. (1998), 'Graphical methods for investigating the size and power of hypothesis tests', *The Manchester school* **66**, 1–26.

Granger, C., King, M. L. and White, H. (1995), 'Testing economic theories and the use of model selection criteria', *Journal of Econometrics* **67**, 173–187.

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

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

MaCurdy, T. E. (1982), 'The use of time series processes to model the error structure of earnings in longitudinal data analysis', *Journal of Econometrics* **18**, 83–114.

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.

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

Rothenberg, T. J. (1971), 'Identification in parametric models', *Econometrica* **39**, 577–592.

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