Essays on
Random Effects Models and GARCH

Jimmy Skoglund

AKADEMISK AVHANDLING

Som för avläggande av filosofie
doktorsexamen vid Handelshögskolan i
Stockholm
framläggs för offentlig granskning
fredagen den 23 mars 2001, kl. 13.15
i sal Torsten Handelshögskolan,
Sveavägen 65

Essays on Random Effects Models and GARCH

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

EFI Mission
EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution which works independently of economic, political and sectional interests. It conducts theoretical and empirical research in management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics.
EFI selects its projects based on the need for theoretical or practical development of a research domain, on methodological interests, and on the generality of a problem.

Research Organization
The research activities are organized in nineteen Research Centers within eight Research Areas. Center Directors are professors at the Stockholm School of Economics.

ORGANIZATION AND MANAGEMENT	
Management and Organisation; (A)	Prof Sven-Erik Sjöstrand
Center for Ethics and Economics; (CEE)	Adj Prof Hans de Geer
Public Management; (F)	Prof Nils Brunsson
Information Management; (I)	Prof Mats Lundeberg
Center for People and Organization (PMO)	Acting Prof Jan Löwstedt
Center for Innovation and Operations Management; (T)	Prof Christer Karlsson
ECONOMIC PSYCHOLOGY	
Center for Risk Research; (CFR)	Prof Lennart Sjöberg
Economic Psychology; (P)	Prof Lennart Sjöberg
MARKETING	
Center for Information and Communication Research; (CIC)	Adj Prof Bertil Thorngren
Center for Consumer Marketing; (CCM)	Associate Prof Magnus Söderlund
Marketing, Distribution and Industrial Dynamics; (D)	Prof Lars-Gunnar Mattsson
ACCOUNTING, CONTROL AND CORPORATE FINANCE	
Accounting and Managerial Finance; (B)	Prof Lars Östman
Managerial Economics; (C)	Prof Peter Jennergren
FINANCE	
Finance; (FI)	Prof Clas Bergström
ECONOMICS	
Center for Health Economics; (CHE)	Prof Bengt Jönsson
International Economics and Geography; (IEG)	Prof Mats Lundahl
Economics; (S)	Prof Lars Bergman
ECONOMICS STATISTICS	
Economic Statistics; (ES)	Prof Anders Westlund
LAW	
Law; (RV)	Prof Erik Nerep

Chairman of the Board: Prof Sven-Erik Sjöstrand. *Director:* Associate Prof Bo Sellstedt,

Adress
EFI, Box 6501, S-113 83 Stockholm, Sweden • Internet: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

Essays on Random Effects Models and GARCH

Jimmy Skoglund

STOCKHOLM SCHOOL OF ECONOMICS
EFI, THE ECONOMIC RESEARCH INSTITUTE

Dissertation for the degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics, 2001

ISBN NR 91-7258-553-6

Keywords:
Random effects models, Maximum likelihood estimation,
Hypothesis tests, GARCH models, Efficient GMM estimation

Distributed by:
EFI, Stockholm School of Economics,
Box 6501, S-113 83 Stockholm, Sweden

Printed by:
Elanders Gotab, Stockholm 2001

Contents

Preface v

Introduction and summary vii

I Random Effects Models 1

1 Maximum likelihood based inference in the two-way random effects model with serially correlated time effects 3

1.1 Introduction 3

1.2 The maximum likelihood estimator 4

1.3 Specification tests 6

1.3.1 Testing for autocorrelation in λ_t 6

1.3.2 Testing AR(1) vs. MA(1) 7

1.4 Monte-Carlo study 9

1.4.1 Design 9

1.4.2 Parameter estimates 9

1.4.3 Hypothesis tests 10

1.4.4 Model selection 19

1.5 Conclusions 22

A Score and Information 25

A.1 The score vector 25

A.2 The information matrix 27

B The LM-test against ARMA(1,1) 31

C Proof of theorem 1 35

2 Asymptotic properties of the maximum likelihood estimator of random effects models with serial correlation 39
2.1 Introduction 39
2.2 The comprehensive specification 41
2.3 Asymptotic properties 42
2.3.1 Assumptions 42
2.3.2 Consistency 45
2.3.3 Asymptotic normality 47
2.3.4 Misspecification 54
2.4 Final remarks 55

A Score and Information 57
A.1 The score vector 57
A.2 The information matrix 59

B Proofs 63

3 Specification and estimation of random effects models with serial correlation of general form 89
3.1 Introduction 89
3.2 The model 90
3.3 Estimation and model specification 92
3.3.1 Likelihood 92
3.3.2 Model selection 95
3.4 Application 100
3.4.1 The model and data 100
3.4.2 Specification of the error components 101
3.4.3 Elasticities and returns to scale 102
3.4.4 Technical change 105
3.5 Conclusions 107

A Proof of property 1 109

II GARCH 117

4 A simple efficient GMM estimator of GARCH models 119
4.1 Introduction 119
4.2 A GARCH(1,1) conditional variance process 120
4.2.1 The efficient GMM estimator 120
4.2.2 Asymptotic properties 122

4.2.3 Finite-sample properties 124
4.3 Extension to models with a conditional mean 127
4.3.1 The GARCH(1,1)-M regression model 127
4.3.2 Asymptotic efficiency comparison 128
4.3.3 Specification tests 130
4.4 Empirical illustration 131
4.5 Final remarks 135

A Proofs **137**

Preface

The work on this thesis started a few years ago when I entered the Ph.D. program at Stockholm School of Economics. Of course, at that time I did not have a clue what the thesis was finally going to be about. A course in financial econometrics (held by Professor Timo Teräsvirta) spurred my interest in GARCH models which subsequently led to the fourth essay in this thesis. It was my thesis supervisor Sune Karlsson that finally directed my attention towards random effects models of which the main part of this thesis is about.

During the work on this thesis I have been fortunate to receive support and guidance from a number of people. I take this opportunity to express my sincere gratitude to my family, colleagues and friends for your never failing support.

There are a few persons to whom I would like to express my particular gratitude. First of all, my wife, Ann-Louise and our three daughters, Izabel, Ellen and Nelly for enduring an absent minded and pre-occupied husband and father. Secondly, my thesis supervisor Sune Karlsson. Being a co-author on the first three essays his influence on this thesis is self-evident.

Financial support from the Swedish Research Council for the Humanities and the Social Sciences (HSFR) is gratefully acknowledged.

Stockholm, February 2001

Jimmy Skoglund

Introduction and summary

This thesis consists of four essays, three in the field of random effects models and one in the field of GARCH. Although these fields are not directly related they reflect my diverse interests in econometrics. This chapter gives a brief introduction to random effects models and GARCH models as well as a summary of the essays.

In many cases data is available on a cross section of households, countries, firms etc. over several time periods. In this situation a plausible specification may be a random effects model

$$y_{it} = \alpha + X_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$

where $i = 1, \ldots, N$ denotes the cross section dimension whereas $t = 1, \ldots, T$ denotes the time dimension. The parameter α is a scalar, $\boldsymbol{\beta}$ is a $k \times 1$ vector and X_{it} is the $it : th$ observation on the k explanatory variables. The error term, ε_{it} is a sum of components. In the one-way model with individual effects,

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

whereas in the one-way model with time effects,

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

where μ_i is the individual specific effect, λ_t is the time specific effect and v_{it} is the idiosyncratic error term. A model which nests both of these one-way models is obtained by letting the error term be a sum of three components. That is,

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

This model is referred to in the literature as the two-way random effects model.

The first essay in this thesis, "Maximum likelihood based inference in the two-way random effects model with serially correlated time effects", 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 also easily generalized to allow for arbitrary stationary and strictly invertible ARMA processes. In addition we consider the model selection problem and 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, test-statistics and model selection procedures.

The second essay, "Asymptotic properties of the maximum likelihood estimator of random effects models with serial correlation", considers the large sample behavior of the maximum likelihood estimator of random effects models with serial correlation in the form of AR(1) for the idiosyncratic or time-specific error component. Consistent estimation and asymptotic normality as N and/or T grows large is established for a comprehensive specification which nests these models as well as all commonly used random effects models. When N or $T \to \infty$ only a subset of the parameters are consistent and asymptotic normality is established for the consistent subsets. In addition we consider the properties of the maximum likelihood estimator under error component misspecification.

The third essay, "Specification and estimation of random effects models with serial correlation of general form", is also 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 or individual effects. The methods are applied to the estimation of a production function using a sample of 72 Japanese chemical firms observed during 1968-1987.

The fourth essay in this thesis is concerned with GARCH. GARCH models are designed to capture the dynamics of the conditional second moment of a time series. It is assumed that the data generating process is

$$y_t = \varepsilon_t$$

where $\{\varepsilon_t\}$ is a discrete-time stochastic process, $\varepsilon_t = z_t h_t$ and $z_t \sim iid(0,1)$. It is the dynamics of h_t^2, the conditional variance at time t, that the GARCH models wish to capture. The GARCH model is commonly used in its most simple form, the GARCH(1,1) model, in which the conditional variance is given by

$$h_t^2 = a + a_1 \varepsilon_{t-1}^2 + b h_{t-1}^2$$

This fourth essay, entitled "A simple efficient GMM estimator of GARCH models", considers efficient GMM based estimation of GARCH models. Sufficient conditions for the estimator to be consistent and asymptotically normal are established for the GARCH(1,1) conditional variance process. In addition efficiency results are obtained for a GARCH(1,1) model where the conditional variance is allowed to enter the mean as well. That is, the GARCH(1,1)-M model. An application to the returns to the SP500 index illustrates.

Part I

Random Effects Models

Essay 1

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

1.1 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. 5) for a review of the literature.

In this essay 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

*Co-authored with Sune Karlsson.

of interest is

$$\begin{aligned} y_{it} &= \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it} \\ \varepsilon_{it} &= \mu_i + \lambda_t + \nu_{it} \end{aligned} \tag{1.1}$$

with λ_t an AR(1),

$$\lambda_t = \rho\lambda_{t-1} + u_t, \tag{1.2}$$

or MA(1),

$$\lambda_t = u_t + \theta u_{t-1}, \tag{1.3}$$

process. This model introduces the empirically plausible phenomena of some factors affecting all individuals observed in the same time period and that these factors tend to persist over time. Examples of such factors are numerous and include business cycles, oil price shocks and economic policies that tend to persist during several of the periods for which the data are collected.

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 essay is as follows. Section 1.2 presents the maximum likelihood estimator of the model. Section 1.3 derives the specification tests. Section 1.4 contains results from a Monte-Carlo experiment and section 1.5 concludes.

1.2 The maximum likelihood estimator

In matrix form the two way model (1.1) 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 $\mathbf{Z}_\mu = (I_N \otimes \iota_T)$, $\mathbf{Z}_\lambda = (\iota_N \otimes \mathbf{I}_T)$, $\mathbf{Z} = [\iota_{NT}, \mathbf{X}]$, $\boldsymbol{\delta} = [\alpha, \boldsymbol{\beta}']'$, $\boldsymbol{\mu}' = (\mu_1, \dots, \mu_N)$, $\boldsymbol{\lambda}' = (\lambda_1, \dots, \lambda_T)$ and ι_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 (1.2) is stationary and the MA process (1.3) is strictly invertible.

The covariance matrix of the combined error term is given by

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

where $\mathbf{J}_T = \boldsymbol{\iota}_T\boldsymbol{\iota}_T'$ a $T \times T$ matrix of ones and $\boldsymbol{\Psi}$ is the covariance matrix of (1.2) or (1.3) 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 $\boldsymbol{\Psi}_\rho$ and the covariance matrix of the MA(1) process as Ψ_θ. More generally $\boldsymbol{\Psi}$ can be the covariance matrix of an arbitrary stationary and strictly invertible ARMA(p, q) process.

Direct inversion of $\boldsymbol{\Sigma}$ 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. For maximum likelihood estimation to be practical convenient expressions for $\boldsymbol{\Sigma}^{-1}$ and $|\boldsymbol{\Sigma}|$ 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

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

Using a well known result from matrix algebra

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\boldsymbol{\iota}_N \otimes \mathbf{I}_T)[\sigma_u^{-2}\boldsymbol{\Psi}^{-1} + N\mathbf{A}^*]^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{I}_T)\mathbf{A}^{-1} \qquad (1.4) \\ &= \mathbf{I}_N \otimes \mathbf{A}^* - (\boldsymbol{\iota}_N \otimes \mathbf{A}^*)[\sigma_u^{-2}\boldsymbol{\Psi}^{-1} + N\mathbf{A}^*]^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \\ &= \mathbf{I}_N \otimes \mathbf{A}^* - \sigma_u^2(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\left[\mathbf{I}_T + N\sigma_u^2\boldsymbol{\Psi}\mathbf{A}^*\right]^{-1}\boldsymbol{\Psi}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)\end{aligned}$$

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 $\boldsymbol{\Sigma}$ in a similar fashion as

$$\begin{aligned}|\boldsymbol{\Sigma}| &= |\mathbf{A}^*|^{-N}|\sigma_u^2\boldsymbol{\Psi}||\sigma_u^{-2}\boldsymbol{\Psi}^{-1} + N\mathbf{A}^*| \qquad (1.5) \\ &= \sigma_\nu^{2N(T-1)}\sigma_1^{2N}\left|\mathbf{I}_T + N\sigma_u^2\boldsymbol{\Psi}\mathbf{A}^*\right|\end{aligned}$$

Using these results we have the log-likelihood function as

$$
\begin{aligned}
l(\boldsymbol{\delta},\gamma) \; = \; & -\frac{TN}{2}\ln 2\pi - \frac{N(T-1)}{2}\ln\sigma_v^2 - \frac{N}{2}\ln\left(T\sigma_\mu^2+\sigma_v^2\right) \\
& -\frac{1}{2}\boldsymbol{\varepsilon}'\left(\mathbf{I}_N\otimes\mathbf{A}^*\right)\boldsymbol{\varepsilon} - \frac{1}{2}\ln\left|\mathbf{I}_T+N\sigma_u^2\boldsymbol{\Psi}\mathbf{A}^*\right| \\
& +\frac{\sigma_u^2}{2}\boldsymbol{\varepsilon}'(\boldsymbol{\iota}_N\otimes\mathbf{A}^*)\left[\mathbf{I}_T+N\sigma_u^2\boldsymbol{\Psi}\mathbf{A}^*\right]^{-1}\boldsymbol{\Psi}(\boldsymbol{\iota}_N'\otimes\mathbf{A}^*)\boldsymbol{\varepsilon}
\end{aligned} \tag{1.6}
$$

where γ is the vector of covariance parameters, $\left(\sigma_\mu^2,\sigma_\nu^2,\sigma_u^2,\rho\right)$ for (1.2) and $\left(\sigma_\mu^2,\sigma_\nu^2,\sigma_u^2,\theta\right)$ for (1.3). Evaluating the likelihood requires numerical calculation of the determinant and inverse of the $T\times T$ matrix $\mathbf{I}_T+N\sigma_u^2\boldsymbol{\Psi}\mathbf{A}^*$ 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 (1.1, 1.2) and (1.1, 1.3) are locally identified in the sense of Rothenberg (1971). The proof is given in appendix B.

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 (1.1, 1.2) and (1.1, 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 in the Monte Carlo experiment. Variance estimates can be based on either a numerical approximation to the Hessian matrix or the information matrix given in the appendix.

1.3 Specification tests

1.3.1 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

[1] If an analytic inverse and determinant is available for $\boldsymbol{\Psi}$ it is more convenient to work with $\sigma_u^{-2}\boldsymbol{\Psi}^{-1}+N\mathbf{A}^*$ (line 2 of (1.4), 1 of (1.5)) since the computations are much more efficient for symmetric positive definite matrices than for general matrices.

hypothesis are obtained from appendix A.2 and A.1 respectively by setting $\mathbf{\Psi} = \mathbf{I}_T$ 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

$$\xi_1 = \left(\frac{\partial l}{\partial \rho}\bigg|_{\rho=0} \right)' \mathcal{I}^{4,4} \left(\frac{\partial l}{\partial \rho}\bigg|_{\rho=0} \right) \tag{1.7}$$

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 (1.7) 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.

1.3.2 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

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

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 $\zeta_j = \operatorname{cov}(\lambda_t, \lambda_{t-j})$. 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

$$\xi_3 = \sqrt{T} \frac{\widehat{\zeta}_2}{\sqrt{\widehat{\zeta}_0^2 + 2\widehat{\zeta}_1^2}} \tag{1.8}$$

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 (1.8) as the BGT-MA test.

Let $\eta_j = \operatorname{corr}(\lambda_t \lambda_{t-j})$, 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

$$\xi_4 = \sqrt{T}(\widehat{\eta}_2 - (\widehat{\eta}_1)^2)/(1 - \widehat{\eta}_2) \tag{1.9}$$

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 (1.9) as the BGT-AR test. To get a test for which size approaches zero asymptotically when $|\eta_1| > 1/2$ and the true process can not be an MA(1) we may also accept the null hypothesis if $\widehat{\eta}_1 > \frac{1}{2} + \frac{1}{\sqrt{T}}$, see Baltagi and Li (1995).

1.4 Monte-Carlo study

1.4.1 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) (1.2) or MA(1) (1.3) process. The regressors, x_{it} are generated as

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

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 $\sigma_u^2 = \sigma_\lambda^2 \left(1 - \rho^2\right)$ 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.[2]

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 upon request.

1.4.2 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.

[2] Gauss code implementing the estimators and test statistics is available from http://swopec.hhs.se/hastef/abs/hastef0383.htm.

We experienced some convergence problems with the MA(1) model when the true model was AR(1) with $|\rho| = 0.8$[3]. This is not too surprising since the MA(1) model cannot match the moments of the AR(1) process for high values of $|\rho|$.

Figure 1.1a 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 1.1b 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.

1.4.3 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.

Tests of the null hypothesis of no serial correlation

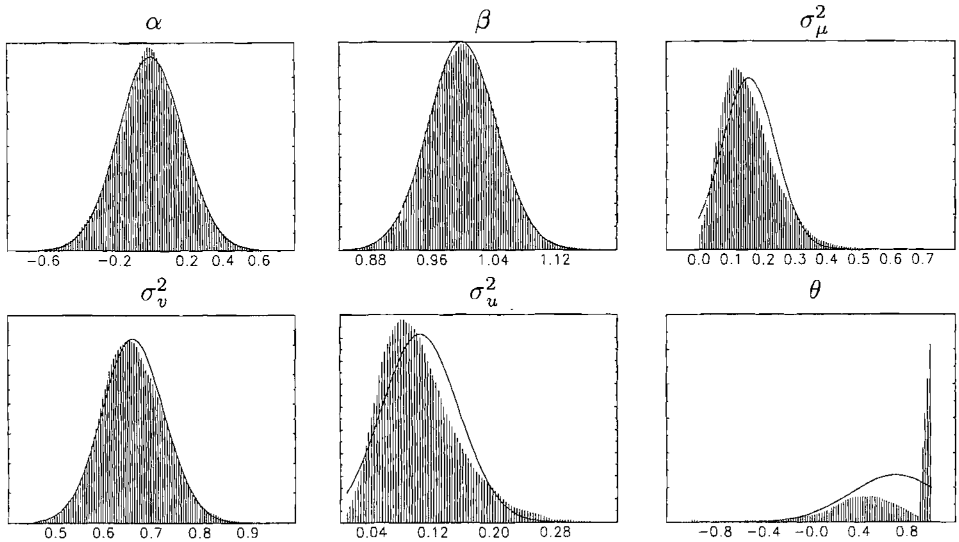
Size Figure 1.2 shows the nominal size (x–axis) and size discrepancy (y-axis) with 95% Kolmogornov-Smirnov "confidence bands" for the LR, Wald and LM-tests[4]. For $N = 20, T = 50$ the size properties are very good for the LM-test and the Wald and LR-tests against an AR(1) alternative (Figure 1.2a-1.2c). When testing against an MA(1) alternative the LR and, especially

[3] 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.

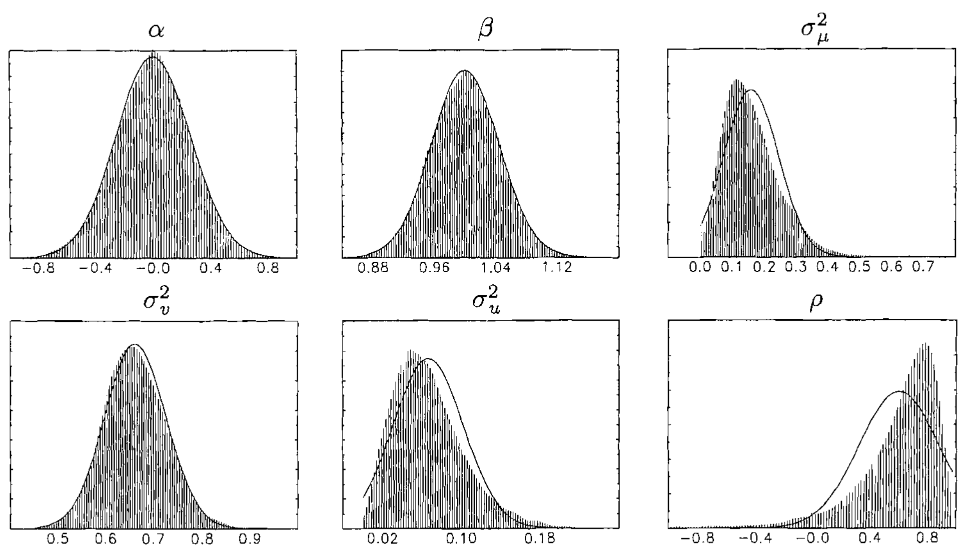
[4] In the graphs we refer to the parameter values of σ_μ^2, σ_v^2 as mi, $i = 1,..,4$ and vj, $j = 1,..,4$ respectively. For example m1v4 refer to $\sigma_\mu^2 = 1/6, \sigma_v^2 = 4/6$ and m2v1 refer to $\sigma_\mu^2 = 2/6, \sigma_v^2 = 1/6$.

Figure 1.1 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), $\rho = 0.8$



the Wald test suffer from size distortion and are sensitive to the choice of variance parameters (Figure 1.2d-1.2e).

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 1.2f-1.2h).

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 1.3a-1.3c 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 1.3d-1.3e 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 1.3c and 1.3d 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 1.3f shows the size and power of the LM-test in the AR(1) model with $\rho = 0.4$.

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

Size of BGT-AR and LM-AR Figure 1.4a-1.4d shows the size discrepancy of the BGT-AR and LM-AR tests for $N = 20, T = 50$. The BGT-AR test is undersized at usual significance levels and the size is also sensitive to the

Figure 1.2 Size discrepancy of tests of no serial correlation

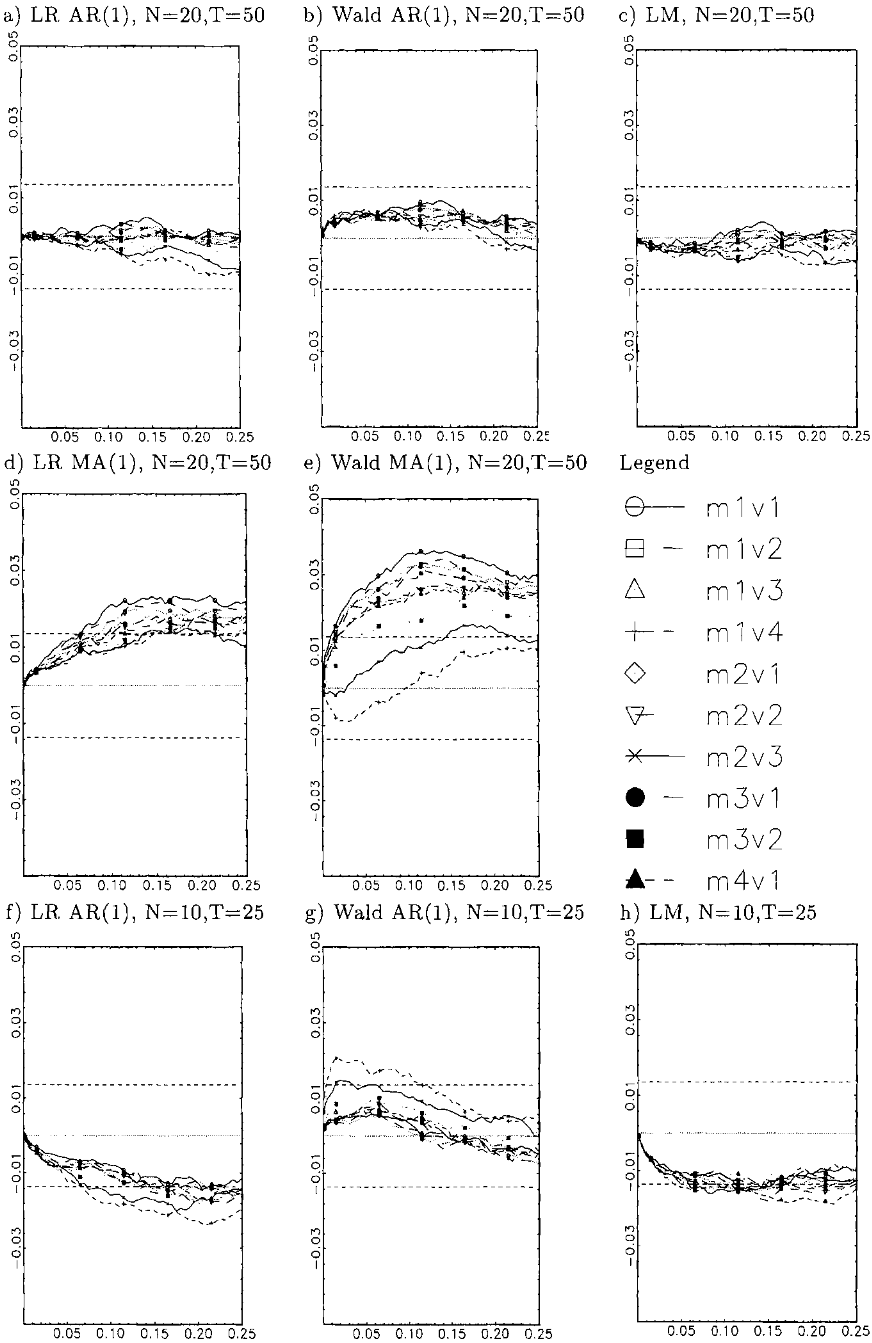
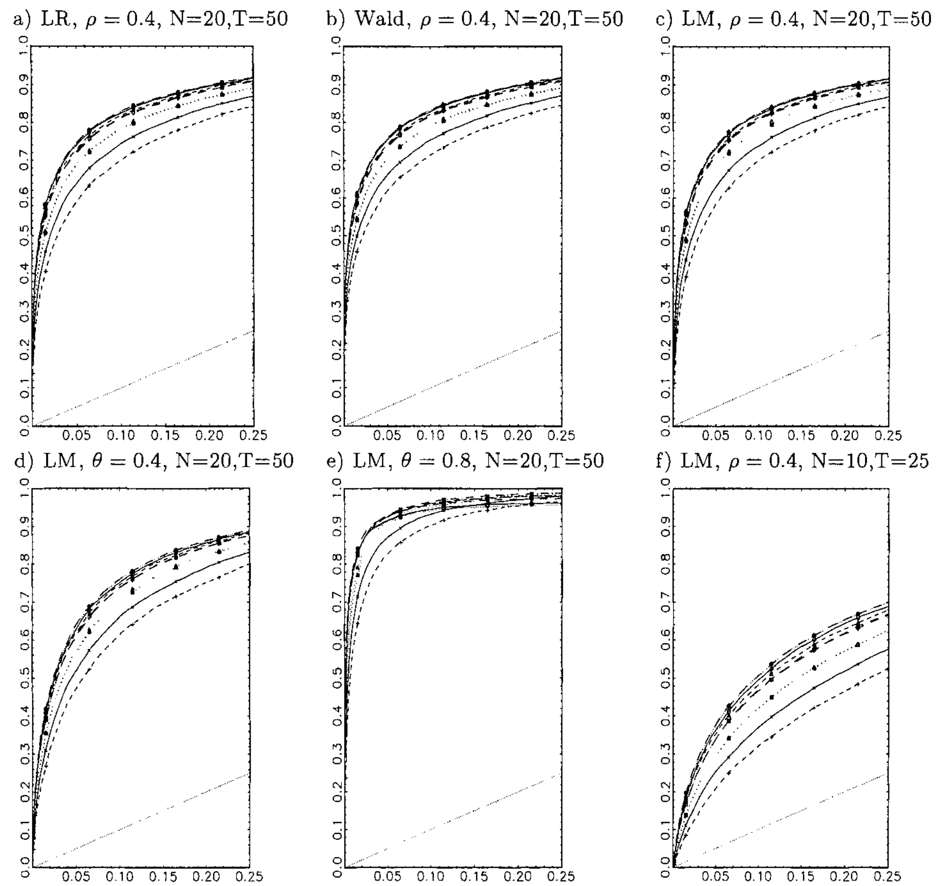


Figure 1.3 Power of tests of no serial correlation



choice of variance parameters[5]. A low σ_λ^2 with a relatively large σ_v^2 makes the BGT-AR test more undersized. For $|\rho| = 0.4$ 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 1.4e-1.4f 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.

Power of BGT-AR and LM-AR Figure 1.5a-1.5d compares the power functions for $N = 20, T = 50$. 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 1.5e compares the power functions for $N = 10, T = 25$. 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 1.6a-1.6d shows the size discrepancy of the BGT-MA and LM-MA tests for $N = 20, T = 50$. The BGT-MA test is insensitive to the choice of variance parameters, but severely undersized with the more severe cases occurring for positive θ. Given the sign of θ, the size of the BGT-MA test appears to be unaffected by the magnitude of θ. 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 1.6e-1.6f 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.

---

[5]All references in the text and in the graphs refer to the unadjusted BGT-AR test i.e. the statistic (1.9) without size adjustment.

Figure 1.4 Size discrepancy of BGT-AR and LM-AR

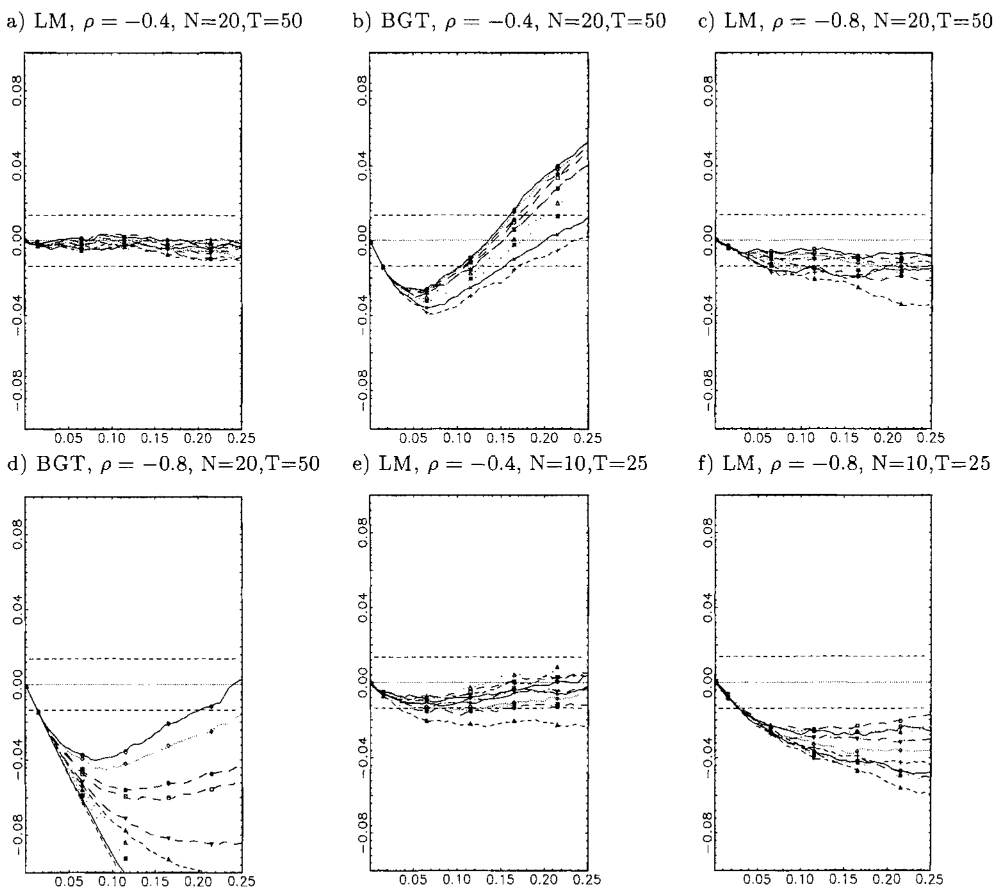


Figure 1.5 Power of BGT-AR and LM-AR

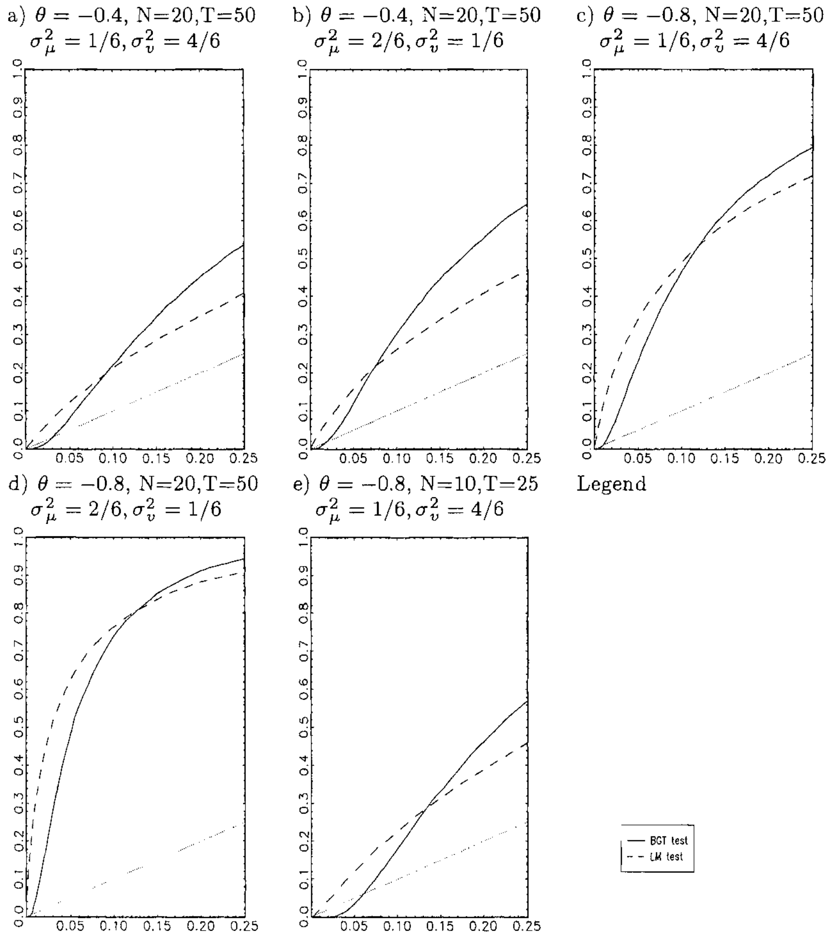
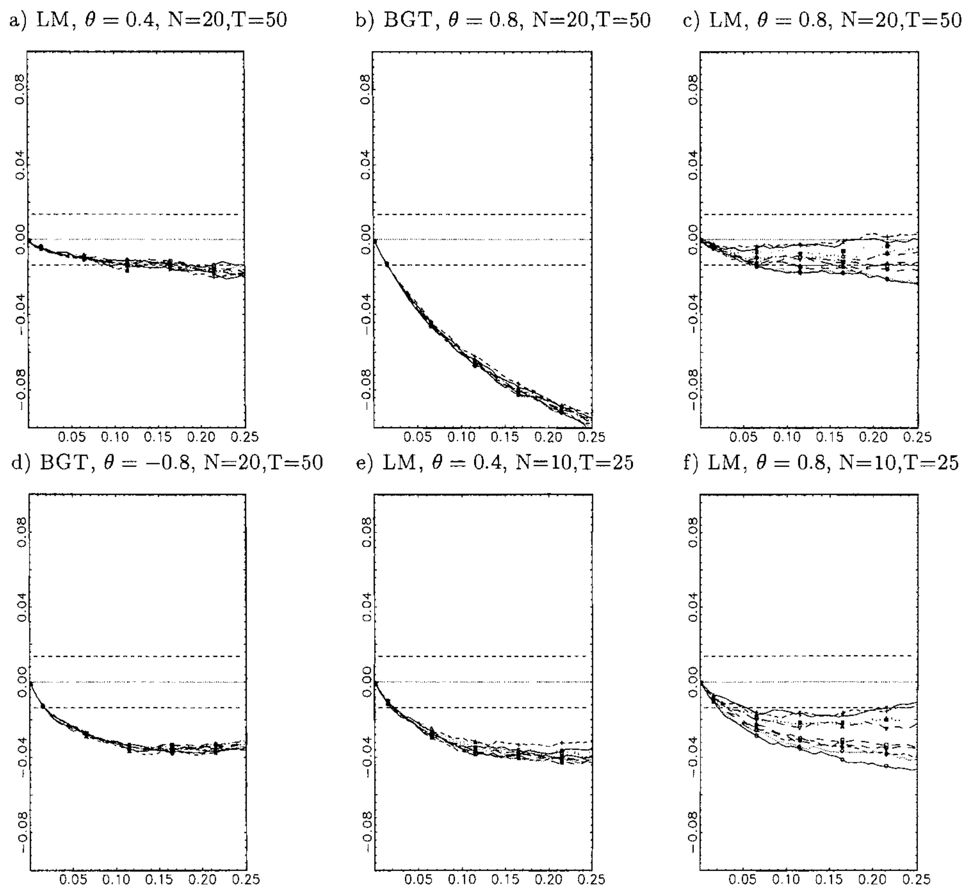


Figure 1.6 Size discrepancy of BGT-MA and LM-MA



Power of BGT-MA and LM-MA Figure 1.7a-1.7d shows the size-power curves for $N = 20, T = 50$. 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. Figure 1.7e illustrates the crossing point for $\rho = -0.8$. The LM-MA and BGT-MA tests have power equal to size at usual significance levels for $|\rho| = 0.4$.

1.4.4 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 discriminating between the AR(1) and MA(1) specification may very well be inconclusive. A decision can then be based on information criteria or comparison of the 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 (1.1, 1.2) and (1.1, 1.3). In the first case 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.

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

Discrimination is based on comparing the log-likelihoods (LL criteria) and p-values of the LM-tests (LM-p strategy) conditional on the LM-test of the null of no autocorrelation rejecting the null at the 5% level. We do not consider dicrimination based on the p-values of the BGT-tests due to their disappointing size properties.

Figure 1.7 Power of BGT-MA and LM-MA

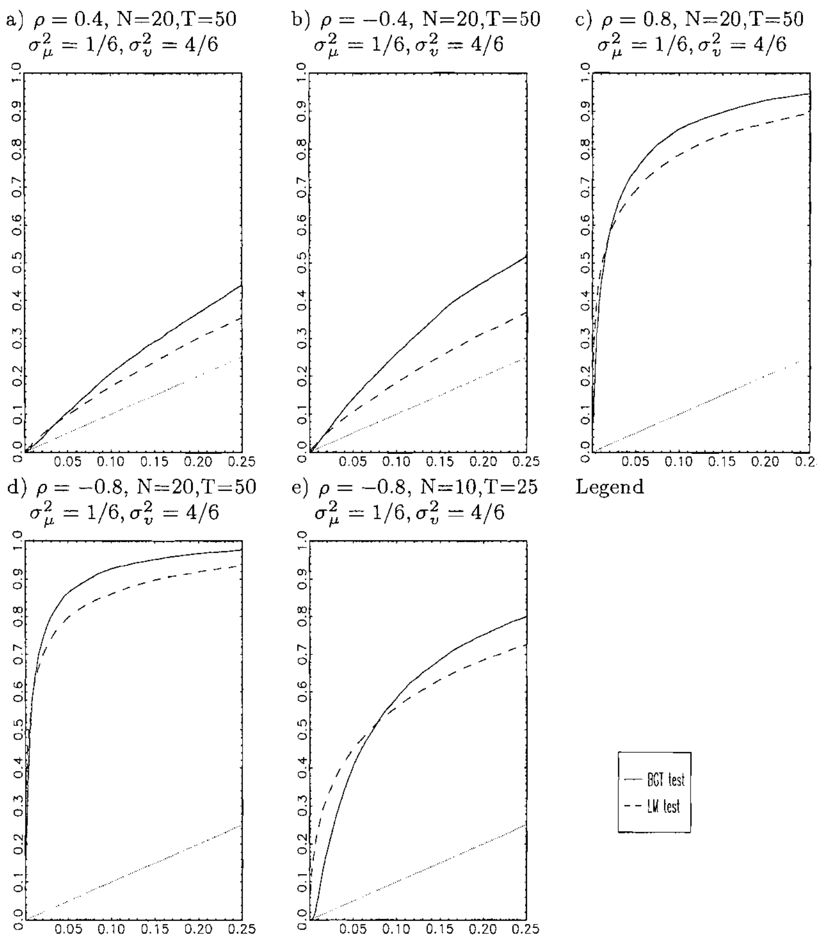


Table 1.1 Frequencies of correct classification of the AR(1) or MA(1) model, $\sigma_\mu^2 = 2/6$, $\sigma_v^2 = 2/6$, $N = 20, T = 50$

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

Table 1.1 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 insensitive 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_λ^2 and high σ_v^2) with a small $|\rho|$ or $|\theta|$.

Overall model selection

The AIC and BIC criteria are compared to a hypothesis testing/p-value approach denoted the LM/LM-p strategy. The first step in the LM/LM-p strategy strategy is to 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, the AR(1) ot MA(1) process is choosen based on the p-values of the discriminating LM-tests.

As for the LL criteria and LM-p strategy considered above the AIC and

Table 1.2 Classification frequencies for the standard two-way model (2-way), AR(1) and MA(1) models, $\sigma_\mu^2 = 2/6$, $\sigma_v^2 = 2/6$, $N = 20, T = 50$

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

BIC criteria and the LM/LM-p strategy are more or less sensitive to variance parametrization. Generally the performance deteriorate with decreasing σ_λ^2 and increasing σ_v^2.

Table 1.2 shows the classification frequencies for the standard two-way model (2-way), AR(1) and MA(1) models for $N = 20, T = 50$ with $\sigma_\mu^2 = 2/6, \sigma_v^2 = 2/6$.

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.

1.5 Conclusions

In this essay 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 an AR(1) alternative since they have the best size properties. Furthermore the power loss compared 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 BGT-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.

Information criteria 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. The ranking is reversed when the standard two-way model is the true model.

Appendix A

Score and Information

A.1 The score vector

This appendix derives the elements of the score vector for the models (1.1, 1.2) and (1.1, 1.3). For the regression parameters we have the standard result

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

and for the variance parameters the score is given by

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

where $\gamma = \left(\sigma_\mu^2, \sigma_\nu^2, \sigma_u^2, \rho\right)'$ for (1.2) and $\left(\sigma_\mu^2, \sigma_\nu^2, \sigma_u^2, \theta\right)'$ for (1.3).

For σ_μ^2 we have

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

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

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

Hence

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

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

For σ_ν^2 we have

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_v^2}) &= \operatorname{tr}(\boldsymbol{\Sigma}^{-1}) = N\operatorname{tr}(\mathbf{A}^*) - N\operatorname{tr}\left[(\mathbf{A}^*)^2\mathbf{B}^{-1}\right] \\
&= \frac{N}{\sigma_1^2} + \frac{N(T-1)}{\sigma_v^2} - \left(\frac{N}{T\sigma_1^4} - \frac{N}{T\sigma_v^4}\right)\boldsymbol{\iota}_T'\mathbf{B}^{-1}\boldsymbol{\iota}_T - \frac{N}{\sigma_v^4}\operatorname{tr}\mathbf{B}^{-1}, \\
\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\sigma_v^2}\boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Sigma}^{-2} = \mathbf{I}_N\otimes\mathbf{A}^{*2} - \mathbf{J}_N\otimes\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^{*2} \\
&\quad -\mathbf{J}_N\otimes\mathbf{A}^{*2}\mathbf{B}^{-1}\mathbf{A}^* + N\mathbf{J}_N\otimes\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^{*2}\mathbf{B}^{-1}\mathbf{A}^*
\end{aligned}$$

with

$$\begin{aligned}
\frac{\partial l}{\partial\sigma_\nu^2} &= -\frac{N}{2\sigma_1^2} - \frac{N(T-1)}{2\sigma_v^2} + \frac{1}{2}\left(\frac{N}{T\sigma_1^4} - \frac{N}{T\sigma_v^4}\right)\boldsymbol{\iota}_T'\mathbf{B}^{-1}\boldsymbol{\iota}_T + \frac{N}{2\sigma_v^4}\operatorname{tr}\mathbf{B}^{-1} \\
&\quad +\frac{1}{2}\boldsymbol{\varepsilon}'\left[\mathbf{I}_N\otimes\mathbf{A}^{*2}\right]\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{*\prime}\widetilde{\boldsymbol{\varepsilon}} + \frac{N}{2}\boldsymbol{\varepsilon}^{*\prime}\boldsymbol{\varepsilon}^*
\end{aligned}$$

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

For σ_u^2 we have

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

and

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

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

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

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

with $\mathbf{L}_\rho = \frac{\partial \boldsymbol{\Psi}_\rho}{\partial \rho} = \frac{2\rho}{1-\rho^2}\boldsymbol{\Psi}_\rho + \frac{1}{1-\rho^2}\mathbf{D}$ where $\mathbf{D}$ 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 (1.2) and $\mathbf{L}_\theta = \frac{\partial \boldsymbol{\Psi}_\theta}{\partial \theta}$ a bidiagonal matrix with 2θ on the main diagonal and ones on the subdiagonals for the MA(1) specification (1.3). This gives

$$\frac{\partial l}{\partial \gamma_4} = \sigma_u^2\left[\begin{array}{c} -\frac{N}{2}\,\mathrm{tr}(\mathbf{A}^*\mathbf{L}) + \frac{N^2}{2}\,\mathrm{tr}(\mathbf{A}^*\mathbf{L}\mathbf{A}^*\mathbf{B}^{-1}) \\ +\frac{1}{2}\widetilde{\varepsilon}'\varepsilon^{L} - N\varepsilon^{*\prime}\varepsilon^{L} + \frac{N^2}{2}\varepsilon^{*\prime}\left(\mathbf{I}_N \otimes \mathbf{L}\right)\varepsilon^* \end{array}\right]$$

with the appropriate $\mathbf{L}$ 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 (1.1, 1.2) and (1.1, 1.3). For the first element we have the result

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

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

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

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

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

We have for the I_{γ_1,γ_j} elements

$$\begin{aligned}\mathcal{I}_{\gamma_1,\gamma_1} &= \frac{1}{2}\operatorname{tr}\left[\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\gamma_1}\right)\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\gamma_1}\right)\right]\\ &= \frac{1}{2\sigma_1^4}\left[NT^2-\frac{2NT}{\sigma_1^2}\left(\iota_T'\mathbf{B}^{-1}\iota_T\right)+\frac{N^2}{\sigma_1^4}\left(\iota_T'\mathbf{B}^{-1}\iota_T\right)^2\right]\end{aligned}$$

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

$$\begin{aligned}\mathcal{I}_{\gamma_1,\gamma_2} &= \frac{1}{2}\operatorname{tr}\left[\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\gamma_1}\right)\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\gamma_3}\right)\right]\\ &= \frac{1}{2\sigma_1^4}\left[NT-\frac{2N}{\sigma_1^2}\left(\iota_T'\mathbf{B}^{-1}\iota_T\right)+N^2\left(\iota_T'\mathbf{B}^{-1}\mathbf{A}^{*2}\mathbf{B}^{-1}\iota_T\right)\right]\end{aligned}$$

$$\begin{aligned}\mathcal{I}_{\gamma_1,\gamma_3} &= \frac{1}{2}\operatorname{tr}\left[\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\gamma_1}\right)\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\gamma_3}\right)\right]\\ &= \frac{1}{2\sigma_1^4}\left[\begin{array}{c}N\left(\iota_T'\mathbf{\Psi}\iota_T\right)-2N^2\left(\iota_T'\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}\iota_T\right)\\ +N^3\left(\iota_T'\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}\mathbf{A}^*\mathbf{B}^{-1}\iota_T\right)\end{array}\right]\end{aligned}$$

$$\begin{aligned}\mathcal{I}_{\gamma_1,\gamma_4} &= \frac{1}{2}\operatorname{tr}\left[\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\gamma_1}\right)\mathbf{\Sigma}^{-1}\left(\frac{\partial\mathbf{\Sigma}}{\partial\gamma_4}\right)\right]\\ &= \frac{\sigma_u^2}{2\sigma_1^4}\left[\begin{array}{c}N\left(\iota_T'\mathbf{L}\iota_T\right)-2N^2\left(\iota_T'\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\iota_T\right)\\ +N^3\left(\iota_T'\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*\mathbf{B}^{-1}\iota_T\right)\end{array}\right]\end{aligned}$$

where $\mathbf{L}$ is $\mathbf{L}_\rho$ or $\mathbf{L}_\theta$ defined in appendix A.1. For the relevant I_{γ_2,γ_j} elements

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

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

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

Finally for the elements involving γ_3, γ_4 we have

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

$$\begin{aligned}
\mathcal{I}_{\gamma_3,\gamma_4} &= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\gamma_3}\right)\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\gamma_4}\right)\right] \\
&= \frac{\sigma_u^2}{2}\left[\begin{array}{c} N^2\operatorname{tr}(\mathbf{A}^*\boldsymbol{\Psi}\mathbf{A}^*\mathbf{L}) - 2N^3\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}\mathbf{A}^*\mathbf{L}\mathbf{A}^*\right) \\ +N^4\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*\right)\end{array}\right]
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}_{\gamma_4,\gamma_4} &= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\gamma_4}\right)\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\gamma_4}\right)\right] \\
&= \frac{\sigma_u^4}{2}\left[\begin{array}{c} N^2\operatorname{tr}\left((\mathbf{A}^*\mathbf{L})^2\right) - 2N^3\operatorname{tr}\left(\mathbf{B}^{-1}(\mathbf{A}^*\mathbf{L})^2\mathbf{A}^*\right) \\ +N^4\operatorname{tr}\left((\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}\mathbf{A}^*)^2\right)\end{array}\right]
\end{aligned}$$

Appendix 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

$$\boldsymbol{\Sigma}_1 = \mathbf{A} + \boldsymbol{\sigma}_u^2(\boldsymbol{\iota}_N \otimes \mathbf{I}_T)\boldsymbol{\Gamma}(\boldsymbol{\iota}_N' \otimes \mathbf{I}_T)$$

with inverse

$$\boldsymbol{\Sigma}_1^{-1} = \mathbf{I}_N \otimes \mathbf{A}^* - (\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\mathbf{B}^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)$$

where $\mathbf{B}_1^{-1} = \sigma_u^2 \left(\mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Gamma}\mathbf{A}^*\right)^{-1} \boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}$ is the covariance matrix of an ARMA(1,1) process with elements $\boldsymbol{\Gamma}_{tt} = \frac{1+\theta^2+2\theta\rho}{1-\rho^2}$ and $\boldsymbol{\Gamma}_{ts} = \frac{(\rho+\theta)(1+\rho\theta)\rho^{|t-s|-1}}{1-\rho^2}$ 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

$$\left.\frac{\partial l}{\partial \tau}\right|_{\tau=0} = -\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}_{1,\tau=0}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \tau}\right|_{\tau=0}\right)\right] + \frac{1}{2}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}_{1,\tau=0}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \tau}\right|_{\tau=0}\right)\boldsymbol{\Sigma}_{1,\tau=0}^{-1}\boldsymbol{\varepsilon}$$

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

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

where $\boldsymbol{\Psi}$ is given by $\boldsymbol{\Psi}_\rho$ if $\tau = \theta$ and $\boldsymbol{\Psi}_\theta$ if $\tau = \rho$, that is

$$\begin{aligned}
\boldsymbol{\Sigma}_{1,\rho=0}^{-1} &= \mathbf{I}_N \otimes \mathbf{A}^* - \sigma_u^2(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\left[\mathbf{I}_T + N\sigma_u^2\boldsymbol{\Psi}_\theta\mathbf{A}^*\right]^{-1}\boldsymbol{\Psi}_\theta(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \\
&= \mathbf{I}_N \otimes \mathbf{A}^* - (\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\mathbf{B}_\theta^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*).
\end{aligned}$$

Also $\frac{\partial \mathbf{\Sigma}_1}{\partial \tau}|_{\tau=0} = \sigma_u^2\left(\mathbf{J}_N \otimes \frac{\partial \mathbf{\Gamma}}{\partial \tau}|_{\tau=0}\right) = \sigma_u^2\left(\mathbf{J}_N \otimes \mathbf{K}_\tau\right)$ where $\mathbf{K}_\theta$ 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.

For testing $H_0 : \rho = 0$ we get

$$\begin{aligned}
\operatorname{tr}\left(\mathbf{\Sigma}^{-1} \left.\frac{\partial \mathbf{\Sigma}_1}{\partial \rho}\right|_{\rho=0}\right) &= \operatorname{tr}(\mathbf{\Sigma}^{-1}\sigma_u^2\left(\mathbf{J}_N \otimes \mathbf{K}_\rho\right)) \\
&= N\sigma_u^2 \operatorname{tr}(\mathbf{A}^*\mathbf{K}_\rho) - N^2\sigma_u^2 \operatorname{tr}(\mathbf{A}^*\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}) \\
\mathbf{\Sigma}^{-1}\left(\left.\frac{\partial \mathbf{\Sigma}_1}{\partial \rho}\right|_{\rho=0}\right)\mathbf{\Sigma}^{-1} &= \mathbf{\Sigma}^{-1}\sigma_u^2\left(\mathbf{J}_N \otimes \mathbf{K}_\rho\right)\mathbf{\Sigma}^{-1} \\
&= \sigma_u^2 \begin{bmatrix} \left(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{K}_\rho\mathbf{A}^*\right) \\ -N\left(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\right) \\ -N\left(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\mathbf{K}_\rho\mathbf{A}^*\right) \\ +N^2\left(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\right) \end{bmatrix}
\end{aligned}$$

and

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

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

$$\mathcal{I}_{\gamma_i,\gamma_j} = \frac{1}{2}\operatorname{tr}\left[\mathbf{\Sigma}^{-1}\left(\frac{\partial \mathbf{\Sigma}_1}{\partial \gamma_i}|_{\rho=0}\right)\mathbf{\Sigma}^{-1}\left(\frac{\partial \mathbf{\Sigma}_1}{\partial \gamma_j}|_{\rho=0}\right)\right]$$

where γ now is defined as $\gamma = (\sigma_\mu^2, \sigma_\nu^2, \sigma_u^2, \rho, \theta)'$. By noting that $(\frac{\partial \mathbf{\Sigma}_1}{\partial \gamma_i}|_{\rho=0}) = \frac{\partial \mathbf{\Sigma}_\theta}{\partial \gamma_i}$ for elements not involving ρ the only elements needed apart from those derived in appendix A.2 for the MA(1) specification are those containing ρ. We have

$$\begin{aligned}
\mathcal{I}_{\gamma_1,\rho} &= \frac{1}{2}\operatorname{tr}\left[\mathbf{\Sigma}^{-1}\left(\left.\frac{\partial \mathbf{\Sigma}_1}{\partial \gamma_1}\right|_{\rho=0}\right)\mathbf{\Sigma}^{-1}\left(\left.\frac{\partial \mathbf{\Sigma}_1}{\partial \rho}\right|_{\rho=0}\right)\right] \\
&= \frac{\sigma_u^2}{2\sigma_1^4}\begin{bmatrix} N\left(\boldsymbol{\iota}_T'\mathbf{K}_\rho\boldsymbol{\iota}_T\right) - 2N^2\left(\boldsymbol{\iota}_T'\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\boldsymbol{\iota}_T\right) \\ +N^3\left(\boldsymbol{\iota}_T'\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\boldsymbol{\iota}_T\right) \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}_{\gamma_2,\rho} &= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \gamma_2}\right|_{\rho=0}\right)\boldsymbol{\Sigma}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \rho}\right|_{\rho=0}\right)\right] \\
&= \frac{\sigma_u^2}{2}\left[\begin{array}{c} N\operatorname{tr}\left(\mathbf{A}^{*2}\mathbf{K}_\rho\right) - 2N^2\operatorname{tr}\left(\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^{*2}\right) \\ +N^3\operatorname{tr}\left(\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^{*2}\mathbf{B}_\theta^{-1}\mathbf{A}^*\right) \end{array}\right]
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}_{\gamma_3,\rho} &= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \gamma_3}\right|_{\rho=0}\right)\boldsymbol{\Sigma}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \rho}\right|_{\rho=0}\right)\right] \\
&= \frac{\sigma_u^2}{2}\left[\begin{array}{c} N^2\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}\mathbf{A}^*\mathbf{K}_\rho\right) - 2N^3\operatorname{tr}\left(\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\boldsymbol{\Psi}\mathbf{A}^*\right) \\ +N^4\operatorname{tr}\left(\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\boldsymbol{\Psi}\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\right) \end{array}\right]
\end{aligned}$$

$$\begin{aligned}
\mathcal{I}_{\gamma_4,\rho} &= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \gamma_4}\right|_{\rho=0}\right)\boldsymbol{\Sigma}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \rho}\right|_{\rho=0}\right)\right] \\
&= \frac{\sigma_u^4}{2}\left[\begin{array}{c} N^2\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_\theta\mathbf{A}^*\mathbf{K}_\rho\right) - 2N^3\operatorname{tr}\left(\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\mathbf{L}_\theta\mathbf{A}^*\right) \\ +N^4\operatorname{tr}\left(\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\mathbf{L}_\theta\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\right) \end{array}\right]
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{I}_{\rho,\rho} &= \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \rho}\right|_{\rho=0}\right)\boldsymbol{\Sigma}^{-1}\left(\left.\frac{\partial \boldsymbol{\Sigma}_1}{\partial \rho}\right|_{\rho=0}\right)\right] \\
&= \frac{\sigma_u^4}{2}\left[\begin{array}{c} N^2\operatorname{tr}\left(\left(\mathbf{A}^*\mathbf{K}_\rho\right)^2\right) - 2N^3\operatorname{tr}\left(\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\mathbf{K}_\rho\mathbf{A}^*\right) \\ +N^4\operatorname{tr}\left(\left(\mathbf{K}_\rho\mathbf{A}^*\mathbf{B}_\theta^{-1}\mathbf{A}^*\right)^2\right) \end{array}\right].
\end{aligned}$$

The corresponding expressions for testing $H_0 : \theta = 0$ are obtained by replacing $\mathbf{K}_\rho$, $\mathbf{B}_\theta^{-1}$ and $\mathbf{L}_\theta$ with $\mathbf{K}_\theta$, $\mathbf{B}_\rho^{-1}$ and $\mathbf{L}_\rho$ respectively.

Appendix 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

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

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

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

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 $a_3 = -2\sigma_u^2\rho / \left(1 - \rho^2\right)^2$. Substituting into row 4 yields

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

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

Bibliography

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.

Essay 2

Asymptotic properties of the maximum likelihood estimator of random effects models with serial correlation

2.1 Introduction

Ever since the seminal work of Balestra and Nerlove (1966) there has been a large interest in and use of random effects models. An important further development was the generalization of the one-way model with individual effects to allow for serial correlation by Lillard and Willis (1978). This model captures correlation in the data at the individual level and has been elaborated by, among others, Anderson and Hsiao (1982), MaCurdy (1982) and Baltagi and Li (1991, 1994). This is, however, not the only conceivable source of correlation. It is quite reasonable to expect random time effects to be correlated as well — reflecting serial correlation in the variables driving unobserved time specific heterogeneity. There are, consequently, a number of variations on random effects models allowing for correlation in the time effect. King (1986) studies a one-way model with serially correlated time effects, Magnus and Woodland (1988) consider a one-way model with both serially correlated time effects and idiosyncratic errors in a multivariate setting and Revankar (1979) proposed a two-way model with serially correlated time effects. Essay 1 derived a straightforward maximum likelihood estimator as well as hypothesis

*Co-authored with Sune Karlsson.

tests for the latter model.

While random effects models with serial correlation in the error components are being used extensively in empirical work the theoretical aspects are less well developed. Anderson and Hsiao (1982) consider the consistency properties of the one-way model with individual effects and serially correlated idiosyncratic effects. Amemiya (1971) proves the consistency and asymptotic normality of the maximum likelihood estimator of the standard two-way model as both N and T grows large. This essay extends the work of Anderson and Hsiao and Amemiya by establishing the asymptotic properties of a comprehensive random effects specification which nests the one-way models with serial correlation as well as the two-way model with serial correlation. More specifically, the model of interest is

$$\begin{aligned} y_{it} &= \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{d}_t'\boldsymbol{\pi} + \mathbf{h}_i'\boldsymbol{\tau} + \varepsilon_{it} \\ \varepsilon_{it} &= \mu_i + \lambda_t + v_{it} \end{aligned} \tag{2.1}$$

with λ_t an AR(1),

$$\lambda_t = \rho_\lambda \lambda_{t-1} + u_t, \tag{2.2}$$

and v_{it} an AR(1),

$$v_{it} = \rho_v v_{it-1} + e_{it}, \tag{2.3}$$

where $\mathbf{x}_{it}$ varies over both individuals and time, $\mathbf{d}_t$ is individual-invariant and $\mathbf{h}_i$ is time-invariant. If there are no time effects we obtain the one-way model with individual effects and serially correlated idiosyncratic errors and if there are no individual effects we obtain the one-way model with both serially correlated time effects and serially correlated idiosyncratic errors. Setting $\rho_v = 0$ obtains the two-way model with serially correlated time effects and setting $\rho_\lambda = 0$ obtains a model not discussed previously in the literature. That is, the two-way model with serially correlated idiosyncratic errors and independent time effects. The standard one-way models and the standard two-way model are, of course, nested in this specification as well.

In contrast to the earlier literature we consider both consistency and asymptotic normality with traditional large N and fixed T as well as with large T fixed N and both N and T large. We also pay special attention to the effects of including time or individual-invariant explanatory variables in the model.

The organization of the essay is as follows. Section 2.2 presents the comprehensive specification and the corresponding maximum likelihood estimator. Section 2.3 derives the asymptotic properties and section 2.4 concludes with some final remarks. All the proofs are in appendix B.

2.2 The comprehensive specification

In matrix form the comprehensive model is written

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

with $\mathbf{Z}_\mu = (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)$, $\mathbf{Z}_\lambda = (\boldsymbol{\iota}_N \otimes \mathbf{I}_T)$, $\mathbf{Z} = (\boldsymbol{\iota}_{NT}, \mathbf{X}, \mathbf{D}, \mathbf{H})$, where $\mathbf{X}$ is k_1-dimensional, $\mathbf{D} = (\boldsymbol{\iota}_N \otimes \mathbf{d})$, $\mathbf{d} = (\mathbf{d}_1, \ldots, \mathbf{d}_T)'$ is k_2-dimensional and $\mathbf{H} = (\mathbf{h} \otimes \boldsymbol{\iota}_T)$, $\mathbf{h} = (\mathbf{h}_1, \ldots, \mathbf{h}_N)'$ is k_3-dimensional, $k = \sum_{i=1}^{3} k_i$, $\boldsymbol{\delta} = \left(\alpha, \boldsymbol{\beta}', \boldsymbol{\pi}', \boldsymbol{\tau}'\right)'$, $\boldsymbol{\mu}' = (\mu_{1,}....., \mu_N)$, $\boldsymbol{\lambda}' = (\lambda_1, ..., \lambda_T)$ and $\boldsymbol{\iota}_N$ is a vector of ones of dimension N. Throughout we will maintain the assumption that $e_{it} \sim N\left(0, \sigma_e^2\right)$, $\mu_i \sim N\left(0, \sigma_\mu^2\right)$, $u_t \sim N\left(0, \sigma_u^2\right)$ independent of each other and $\mathbf{X}$, $\mathbf{d}$ and $\mathbf{h}$. In addition we assume that $\rho_\lambda, \rho_v \in (-1, 1)$.

The covariance matrix of the combined error term is given by

$$\begin{aligned}\boldsymbol{\Sigma} &= E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{Z}_\mu E(\boldsymbol{\mu}\boldsymbol{\mu}')\mathbf{Z}_\mu' + \mathbf{Z}_\lambda E(\boldsymbol{\lambda}\boldsymbol{\lambda}')\mathbf{Z}_\lambda' + E(\boldsymbol{\nu}\boldsymbol{\nu}') \\ &= \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_u^2(\mathbf{J}_N \otimes \boldsymbol{\Psi}_\lambda) + \sigma_e^2(\mathbf{I}_N \otimes \boldsymbol{\Psi}_v)\end{aligned} \tag{2.4}$$

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.

Let $\mathbf{A}$ be the covariance matrix of the one-way model with individual specific effects and serially correlated v_{it}. We can then write

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

where

$$\mathbf{A} = \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_e^2(\mathbf{I}_N \otimes \boldsymbol{\Psi}_v) = \mathbf{I}_N \otimes \left(\sigma_\mu^2 \mathbf{J}_T + \sigma_e^2 \boldsymbol{\Psi}_v\right)$$

Following Baltagi and Li (1991) let $\mathbf{C}$ be the Prais-Winsten transformation matrix for $\boldsymbol{\Psi}_v$ and write

$$\begin{aligned}\mathbf{C}^{-1}\mathbf{C}\left(\sigma_\mu^2 \mathbf{J}_T + \sigma_e^2 \boldsymbol{\Psi}_v\right)\mathbf{C}'\mathbf{C}^{-T} &= \mathbf{C}^{-1}\left(\sigma_\mu^2 \left(\mathbf{C}\boldsymbol{\iota}_T\right)\left(\mathbf{C}\boldsymbol{\iota}_T\right)' + \sigma_e^2 \mathbf{I}_T\right)\mathbf{C}^{-T} \\ &= \mathbf{C}^{-1}\left(\sigma_\alpha^2 \overline{\mathbf{J}}_T^\alpha + \sigma_e^2 \overline{\mathbf{E}}_T^\alpha\right)\mathbf{C}^{-T}\end{aligned}$$

where $\sigma_\alpha^2 = d^2\sigma_\mu^2 \left(1 - \rho_v\right)^2 + \sigma_e^2$, $\overline{\mathbf{J}}_T^\alpha = \boldsymbol{\iota}_T^\alpha \boldsymbol{\iota}_T^{\alpha\prime}/d^2$, $\boldsymbol{\iota}_T^{\alpha\prime} = \left(\alpha, \boldsymbol{\iota}_{T-1}'\right) = \left(\mathbf{C}\boldsymbol{\iota}_T\right)'$ and $\overline{\mathbf{E}}_T^\alpha = \mathbf{I}_T - \overline{\mathbf{J}}_T^\alpha$ with $d^2 = \boldsymbol{\iota}_T^{\alpha\prime}\boldsymbol{\iota}_T^\alpha = \alpha^2 + (T-1)$, $\alpha = \sqrt{\left(1 + \rho_v\right) / \left(1 - \rho_v\right)}$. We then have

$$\mathbf{A}^{-1} = \mathbf{I}_N \otimes \mathbf{C}'\left(\sigma_\alpha^{-2}\overline{\mathbf{J}}_T^\alpha + \sigma_e^{-2}\overline{\mathbf{E}}_T^\alpha\right)\mathbf{C} = \mathbf{I}_N \otimes \mathbf{A}^*$$

As in the first essay we can then write

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

and

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

Which gives the log-likelihood as

$$\begin{aligned}
l(\boldsymbol{\delta},\boldsymbol{\gamma}) &= -\frac{TN}{2}\ln 2\pi - \frac{N(T-1)}{2}\ln\sigma_e^2 + \frac{N}{2}\ln|\mathbf{C}|^2 - \frac{N}{2}\ln\sigma_\alpha^2 \qquad (2.5) \\
&\quad -\frac{1}{2}\boldsymbol{\varepsilon}'\left(\mathbf{I}_N \otimes \mathbf{A}^*\right)\boldsymbol{\varepsilon} - \frac{1}{2}\ln\left|\mathbf{I}_T + N\sigma_u^2\mathbf{\Psi}_\lambda\mathbf{A}^*\right| \\
&\quad +\frac{\sigma_u^2}{2}\boldsymbol{\varepsilon}'(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\left[\mathbf{I}_T + N\sigma_u^2\mathbf{\Psi}_\lambda\mathbf{A}^*\right]^{-1}\mathbf{\Psi}_\lambda(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)\boldsymbol{\varepsilon}
\end{aligned}$$

where $\boldsymbol{\delta} = \left(\alpha, \boldsymbol{\beta}', \boldsymbol{\pi}', \boldsymbol{\tau}'\right)'$ and $\boldsymbol{\gamma}$ is the vector of covariance parameters, $(\sigma_\mu^2, \sigma_e^2, \rho_v, \sigma_u^2, \rho_\lambda)$.

Evaluation of the likelihood requires numerical computation of the determinant and inverse of the $T \times T$ matrix $\mathbf{I}_T + N\sigma_u^2\mathbf{\Psi}_\lambda\mathbf{A}^*$. The elements of the score for the comprehensive log likelihood (2.5) are given in appendix A.1 and the information matrix in appendix A.2.

2.3 Asymptotic properties

Establishing consistency and asymptotic normality is complicated due to the fact that the likelihood contains terms of different orders. Furthermore the likelihood cannot be evaluated analytically which complicates matter further.

2.3.1 Assumptions

The following assumptions are sufficient for the results

(a) $\mu_i \sim N\left(0, \sigma_\mu^2\right)$, $u_t \sim N\left(0, \sigma_u^2\right)$, $e_{it} \sim N\left(0, \sigma_e^2\right)$ *independent of each other and* $\mathbf{X}$, $\mathbf{d}$ *and* $\mathbf{h}$. *In addition* $\mathbf{X}$, $\mathbf{d}$ *and* $\mathbf{h}$ *have full column ranks* k_1, k_2 *and* k_3 *respectively where* $(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT}, \mathbf{h}_i)$ *is iid across* i, $i = 1, \ldots, N$ *and* $(\mathbf{X}_{1t}, \ldots, \mathbf{X}_{Nt}, \mathbf{d}_t)$ *is strictly stationary and ergodic across* t, $t = 1, \ldots, T$ *with* $E\left|X_{jit}\right|^2 < \infty$, $j = 1, \ldots, k_1$, $E\left|d_{lt}\right|^2 < \infty$, $l = 1, \ldots, k_2$ *and* $E\left|h_{si}\right|^2 < \infty$, $s = 1, \ldots, k_3$

(b) $\Theta \equiv \{\theta : \boldsymbol{\delta}'\boldsymbol{\delta} \leq c < \infty, 0 < \sigma^2_{j,lb} \leq \sigma^2_j \leq \sigma^2_{j,ub}, -1 < \rho_{i,lb} \leq \rho_i \leq \rho_{i,ub} < 1\}$, *where ub, lb denote upper and lower bound respectively and $j = \mu, u, e$, $i = \lambda, v$ with θ_0 the true parameter vector belonging to the interior of Θ*

(c) *The normalized moment matrix, $\frac{1}{NT}\mathbf{Z}'\mathbf{Z}$, converge in probability to a finite positive-definite matrix as $N \to \infty, T \to \infty$ or $N, T \to \infty$. In addition there exists a diagonal matrix, say $\boldsymbol{\Upsilon}$, such that the normalized quadratic form*

$$\boldsymbol{\Upsilon}^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\boldsymbol{\Upsilon}^{-1}$$

converge in probability uniformly on Θ to a finite positive-definite matrix as $N \to \infty, T \to \infty$ or $N, T \to \infty$

The normality assumption on μ_i, u_t and e_{it} in (a) is certainly not necessary for consistency arguments. It is well-known that maximizing a normal log-likelihood even though the errors are non-normal will in general give consistent estimates given some moment conditions on μ_i, u_t, e_{it}. Inference is however more complicated so it is convenient to stay in the Gaussian framework.

Assumptions (b) is standard whereas assumption (c) may require some clarification. The first part of assumption (c) is the usual moment condition on the explanatory variables encountered in the asymptotic analysis of least squares models. The second part is concerned with the quadratic form, $\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}$. It amounts to assuming that the normalized quadratic form, $\boldsymbol{\Upsilon}^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}\boldsymbol{\Upsilon}^{-1}$, have the required limit properties. Lemma 6 in appendix B derives the scalings necessary for the block diagonal elements to converge to positive definite matrices. It follows from this that the scaling matrix must be given by

$$\operatorname{diag} \boldsymbol{\Upsilon} = \left(\min\left(\sqrt{N}, \sqrt{T}\right), \mathbf{F}_\beta, \mathbf{F}_\pi, \mathbf{F}_\tau\right) \tag{2.6}$$

where $\mathbf{F}_\beta$ is a vector containing k_1 $\sqrt{NT}$, and $\mathbf{F}_\pi, \mathbf{F}_\tau$ are vectors containing k_2 $\sqrt{T}$ and k_3 $\sqrt{N}$ respectively. Contrary to Amemiya (1971) we do not assume that $\operatorname{plim}_{N,T\to\infty} \frac{1}{NT}\left(\boldsymbol{\iota}_{NT}, \mathbf{X}\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\iota}_{NT}, \mathbf{X}\right)$ is non-singular. This is not true as can be seen from the form of the scaling matrix in (2.6). The constant needs different normalization and to complicate matters further the appropriate normalization depends on the relative rate of increase of N and T. This indicates a general problem with time-invariant and/or individual-invariant explanatory variables and in this sense we can interpret assumption (c) as that the $\mathbf{H}$ and $\mathbf{D}$ matrices contain variables with "sufficient variation" in the N and T dimension respectively. In fact, $\operatorname{plim}_{T\to\infty} \frac{1}{T}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{H}$ and $\operatorname{plim}_{N\to\infty} \frac{1}{N}\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}$

are null matrices whereas $\text{plim}_{T\to\infty} \mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H}$ and $\text{plim}_{N\to\infty} \mathbf{D}'\mathbf{\Sigma}^{-1}\mathbf{D}$ are random matrices. The appropriate normalizations of these information elements as both N and T grows large are $\frac{1}{N}$ and $\frac{1}{T}$ respectively and in contrast to the constant term these normalizations do not depend on the relative rate of increase of N and T. This illustrates that the behavior of the quadratic form, $\mathbf{Z}'\mathbf{\Sigma}^{-1}\mathbf{Z}$, may differ sharply from that of the "ordinary form", $\mathbf{Z}'\mathbf{Z}$.

We might remark here that the normalization matrix given in (2.6) and of course assumption (b) as well is only appropriate for the two-way model. For the one-way model with individual effects $\mathbf{D}'\mathbf{\Sigma}^{-1}\mathbf{D} = N\mathbf{d}'\mathbf{A}^*\mathbf{d}$ and hence $\text{plim}_{N\to\infty} \frac{1}{N}\mathbf{D}'\mathbf{\Sigma}^{-1}\mathbf{D}$ is a random matrix. Similarly in the one-way model with time effects $\text{plim}_{T\to\infty} \frac{1}{T}\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H}$ is a random matrix. The appropriate normalizations of the information elements $\mathbf{D}'\mathbf{\Sigma}^{-1}\mathbf{D}$, $\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H}$ are $\frac{1}{NT}$ and $\frac{1}{N}$ respectively in the one-way model with individual effects and $\frac{1}{T}$ and $\frac{1}{NT}$ respectively in the one-way model with time effects. The unique scaling matrix for the one-way model with individual effects is obtained by letting the first element of diag $\mathbf{\Upsilon}$ be replaced with $\sqrt{N}$ and $\mathbf{F}_\pi$ a vector containing k_2 $\sqrt{NT}$. For the time effects case this matrix is obtained by replacing the first element of diag $\mathbf{\Upsilon}$ with $\sqrt{T}$ and letting $\mathbf{F}_\tau$ be a vector containing k_3 $\sqrt{NT}$.

For the purpose of giving results for the one-way models we define $\Theta^{(i)}$ as the compact parameter space for the parameters of the individual effects model, $\theta^{(i)} = \left(\boldsymbol{\delta}, \gamma^{(i)}\right), \gamma^{(i)} = \left(\sigma_\mu^2, \sigma_e^2, \rho_v\right)$. Correspondingly we define $\Theta^{(t)}$ as the compact parameter space for the parameters of the time effects model, $\theta^{(t)} = \left(\boldsymbol{\delta}, \gamma^{(t)}\right), \gamma^{(t)} = \left(\sigma_e^2, \rho_v, \sigma_u^2, \rho_\lambda\right)$ and make the following additional assumptions

($\text{b}_{(i)}$) $\Theta^{(i)} \equiv \{\theta^{(i)} : \boldsymbol{\delta}'\boldsymbol{\delta} \leq c < \infty, 0 < \sigma_{j,lb}^2 \leq \sigma_j^2 \leq \sigma_{j,ub}^2, -1 < \rho_{v,lb} \leq \rho_v \leq \rho_{v,ub} < 1\}$, *where* ub, lb *denote upper and lower bound respectively and* $j = \mu, e$ *with* $\theta_0^{(i)}$ *the true parameter vector belonging to the interior of* $\Theta^{(i)}$

($\text{b}_{(t)}$) $\Theta^{(t)} \equiv \{\theta : \boldsymbol{\delta}'\boldsymbol{\delta} \leq c < \infty, 0 < \sigma_{j,lb}^2 \leq \sigma_j^2 \leq \sigma_{j,ub}^2, -1 < \rho_{i,lb} \leq \rho_i \leq \rho_{i,ub} < 1\}$, *where* ub, lb *denote upper and lower bound respectively and* $j = u, e$, $i = \lambda, v$ *with* $\theta_0^{(t)}$ *the true parameter vector belonging to the interior of* $\Theta^{(t)}$

Unless otherwise indicated in the following results for the comprehensive model use assumptions (a)-(c) and results for the one-way model with individual effects use assumptions (a), ($\text{b}_{(i)}$), (c). Accordingly, results for the one-way model with time effects use assumptions (a), ($\text{b}_{(t)}$) and (c).

2.3.2 Consistency

Our first result is for the comprehensive model specified by the log-likelihood (2.5). Define $\theta = (\boldsymbol{\delta}, \gamma)$ and let $\widehat{\theta} = \left(\widehat{\boldsymbol{\delta}}, \widehat{\gamma}\right)$, $\theta_0 = (\boldsymbol{\delta}_0, \gamma_0)$ denote the estimator and true parameters respectively

Theorem 1 *(Comprehensive model)*

(i) $\widehat{\theta} \xrightarrow{p} \theta_0$ *on* Θ *as* $N, T \to \infty$ *(it does not matter how)*

(ii) $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0, \widehat{\boldsymbol{\tau}} \xrightarrow{p} \boldsymbol{\tau}_0$ *on* Θ *as* $N \to \infty$ *and if in addition* $T \geq 2$, $\left(\widehat{\sigma}_\mu^2, \widehat{\sigma}_e^2, \widehat{\rho}_v\right) \xrightarrow{p} \left(\sigma_{\mu 0}^2, \sigma_{e0}^2, \rho_{v0}\right)$ *on* Θ *as* $N \to \infty$

(iii) $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0, \widehat{\boldsymbol{\pi}} \xrightarrow{p} \boldsymbol{\pi}_0$ *on* Θ *as* $T \to \infty$ *and if in addition* $N \geq 2$, $\left(\widehat{\sigma}_e^2, \widehat{\rho}_v, \widehat{\sigma}_u^2, \widehat{\rho}_\lambda\right) \xrightarrow{p} \left(\sigma_{e0}^2, \rho_{v0}, \sigma_{u0}^2, \rho_{\lambda 0}\right)$ *in an open neighborhood of* $\left(\sigma_{e0}^2, \rho_{v0}, \sigma_{u0}^2, \rho_{\lambda 0}\right)$

The proof proceeds by examining the probability limit of the log-likelihood standardized by $\frac{1}{NT}$. This method is not useful for dealing with the constant term but it allows us to prove some global consistency results for the variance parameters which are not easily obtained otherwise. The asymptotic properties of the constant term are essentially established in two lemmas, lemma 6 and lemma 7 given in appendix B. Lemma 6 shows that $\widehat{\alpha} = \min\left(\sqrt{N}, \sqrt{T}\right)$ consistent and hence the constant is not consistently estimated if only N or $T \to \infty$. Note that the inconsistency of the constant does not affect consistency of the $\sqrt{N}$ consistent parameters as $N \to \infty$. Nor does it affect consistency of the $\sqrt{T}$ consistent parameters as $T \to \infty$. The intuition for this is that these estimators do not (at least not asymptotically) use information about the constant. Analogously, inconsistency of for example $\boldsymbol{\pi}$ (the parameters of individual-invariant explanatory variables) as $N \to \infty$ does not affect consistency of the $\sqrt{N}$ consistent parameters[1].

Note that we assumed $T \geq 2$ as $N \to \infty$ to achieve identification of the variance parameters $\left(\sigma_\mu^2, \sigma_e^2, \rho_v\right)$ and $N \geq 2$ as $T \to \infty$ to achieve identification of the variance parameters $\left(\sigma_e^2, \rho_v, \sigma_u^2, \rho_\lambda\right)$. A similar requirement appears in assumption (a) and these conditions are frequently redundant when there are time or individual-invariant variables in the model.

A number of special cases emerges from the above theorem. For example, consistency results for the two-way model with serially correlated time effects and the two-way model with serially correlated idiosyncratic errors follow as

[1] The phrase "inconsistent parameters" is used here to refer to parameters whose estimator converge to non-degenerate random variables.

direct corollaries from theorem 1. In addition if $\rho_v = \rho_\lambda = 0$ and we have no time or individual-invariant explanatory variables theorem 1 (i) gives the consistency result of Amemiya (1971) for the standard two-way model. Theorem 1 (ii) and (iii) then gives consistency results as $N \to \infty$ and $T \to \infty$ respectively not covered in Amemiya (1971)[2].

Theorem 1 does not apply to the one-way model with both serially correlated time effects and serially correlated idiosyncratic errors since we have not allowed for $\sigma_\mu^2 = 0$. Consistency results for this model are however straightforward to obtain

Corollary 1 *(One-way model with time effects)*

(i) $\widehat{\theta}^{(t)} \xrightarrow{p} \theta_0^{(t)}$ *on* $\Theta^{(t)}$ *as* $N, T \to \infty$ *(it does not matter how)*

(ii) $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\tau}} \xrightarrow{p} \boldsymbol{\tau}_0$ *and* $\left(\widehat{\sigma}_e^2, \widehat{\rho}_v\right) \xrightarrow{p} \left(\sigma_{e0}^2, \rho_{v0}\right)$ *on* $\Theta^{(t)}$ *as* $N \to \infty$

(iii) $\widehat{\alpha} \xrightarrow{p} \alpha_{0,}\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\pi}} \xrightarrow{p} \boldsymbol{\pi}_0$ *and* $\widehat{\boldsymbol{\tau}} \xrightarrow{p} \boldsymbol{\tau}_0$ *on* $\Theta^{(t)}$ *as* $T \to \infty$ *and if in addition* $N \geq 2$, $\left(\widehat{\sigma}_e^2, \widehat{\rho}_v, \widehat{\sigma}_u^2, \widehat{\rho}_\lambda\right) \xrightarrow{p} \left(\sigma_{e0}^2, \rho_{v0,}\sigma_{u0}^2, \rho_{\lambda 0}\right)$ *in an open neighborhood of* $\left(\sigma_{e0}^2, \rho_{v0,}\sigma_{u0}^2, \rho_{\lambda 0}\right)$

In contrast to the comprehensive model considered in theorem 1 it is in this case possible to estimate all the parameters consistently as only $T \to \infty$. This follows since there is no individual effect which confounds with the constant term or the time-invariant explanatory variables. The constant is accordingly $\sqrt{T}$ consistent no matter what the relative rate of increase of N and T and $\boldsymbol{\tau}$ is accordingly $\sqrt{NT}$ consistent. The non-presence of individual effects also implies that there is a somewhat weaker identification condition on the variance parameters $\left(\sigma_e^2, \rho_v\right)$ as $N \to \infty$.

Finally we give consistency results for the one-way model with individual effects and serially correlated idiosyncratic errors

Corollary 2 *(One-way model with individual effects)*

(i) $\widehat{\theta}^{(i)} \to \theta_0^{(i)}$ *on* $\Theta^{(i)}$ *as* $N, T \to \infty$ *(it does not matter how)*

(ii) $\widehat{\alpha} \xrightarrow{p} \alpha_{0,}\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\tau}} \xrightarrow{p} \boldsymbol{\tau}_0$ *and* $\widehat{\boldsymbol{\pi}} \xrightarrow{p} \boldsymbol{\pi}_0$ *on* $\Theta^{(i)}$ *as* $N \to \infty$ *and if in addition* $T \geq 2$, $\left(\widehat{\sigma}_\mu^2, \widehat{\sigma}_e^2, \widehat{\rho}_v\right) \xrightarrow{p} \left(\sigma_{\mu 0}^2, \sigma_{e0}^2, \rho_{v0}\right)$ *on* $\Theta^{(i)}$ *as* $N \to \infty$

[2] For the standard two-way model it is straightforward to prove global consistency of $\widehat{\sigma}_u^2, \widehat{\sigma}_e^2$ as $T \to \infty$ (assuming $N \geq 2$).

Table 2.1 Consistency properties of random effects models

Model	Case		Case	
2-way $(\mu_i, \lambda_t, v_{it})$	$N \to \infty$		$T \to \infty$	
	C	NC	C	NC
λ_t, v_{it} *iid*	$\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\mu^2, \sigma_e^2$	$\alpha, \boldsymbol{\pi}, \sigma_u^2$	$\boldsymbol{\beta}, \boldsymbol{\pi}, \sigma_u^2, \sigma_e^2$	$\alpha, \boldsymbol{\tau}, \sigma_\mu^2$
λ_t $AR(1), v_{it}$ *iid*	$\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\mu^2, \sigma_e^2$	$\alpha, \boldsymbol{\pi}, \sigma_u^2, \rho_\lambda$	$\boldsymbol{\beta}, \boldsymbol{\pi}, \sigma_u^2, \rho_\lambda, \sigma_e^2$	$\alpha, \boldsymbol{\tau}, \sigma_\mu^2$
λ_t *iid*$, v_{it}$ $AR(1)$	$\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\mu^2, \sigma_e^2, \rho_v$	$\alpha, \boldsymbol{\pi}, \sigma_u^2$	$\boldsymbol{\beta}, \boldsymbol{\pi}, \sigma_u^2, \sigma_e^2, \rho_v$	$\alpha, \boldsymbol{\tau}, \sigma_\mu^2$
λ_t $AR(1), v_{it}$ $AR(1)$	$\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_\mu^2, \sigma_e^2, \rho_v$	$\alpha, \boldsymbol{\pi}, \sigma_u^2, \rho_\lambda$	$\boldsymbol{\beta}, \boldsymbol{\pi}, \sigma_u^2, \rho_\lambda, \sigma_e^2, \rho_v$	$\alpha, \boldsymbol{\tau}, \sigma_\mu^2$
1-way (λ_t, v_{it})				
	C	NC	C	NC
λ_t, v_{it} *iid*	$\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_e^2$	$\alpha, \boldsymbol{\pi}, \sigma_u^2$	$\alpha, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\tau}, \sigma_u^2, \sigma_e^2$	
λ_t $AR(1), v_{it}$ *iid*	$\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_e^2$	$\alpha, \boldsymbol{\pi}, \sigma_u^2, \rho_\lambda$	$\alpha, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\tau}, \sigma_u^2, \rho_\lambda, \sigma_e^2$	
λ_t *iid*$, v_{it}$ $AR(1)$	$\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_e^2, \rho_v$	$\alpha, \boldsymbol{\pi}, \sigma_u^2$	$\alpha, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\tau}, \sigma_u^2, \sigma_e^2, \rho_v$	
λ_t $AR(1), v_{it}$ $AR(1)$	$\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_e^2, \rho_v$	$\alpha, \boldsymbol{\pi}, \sigma_u^2, \rho_\lambda$	$\alpha, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\tau}, \sigma_u^2, \rho_\lambda, \sigma_e^2, \rho_v$	
1-way (μ_i, v_{it})				
	C	NC	C	NC
v_{it} *iid*	$\alpha, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\pi}, \sigma_\mu^2, \sigma_e^2$		$\boldsymbol{\beta}, \boldsymbol{\pi}, \sigma_e^2$	$\alpha, \boldsymbol{\tau}, \sigma_\mu^2$
v_{it} $AR(1)$	$\alpha, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\pi}, \sigma_\mu^2, \sigma_e^2, \rho_v$		$\boldsymbol{\beta}, \boldsymbol{\pi}, \sigma_e^2, \rho_v$	$\alpha, \boldsymbol{\tau}, \sigma_\mu^2$

Abbreviations: C=Consistent; NC=Not Consistent

(iii) $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\pi}} \xrightarrow{p} \boldsymbol{\pi}_0$ *and* $\left(\widehat{\sigma}_e^2, \widehat{\rho}_v\right) \xrightarrow{p} \left(\sigma_{e0}^2, \rho_{v0}\right)$ *on* $\Theta^{(i)}$ *as* $T \to \infty$

Since no time effect confounds with the constant or the individual-invariant explanatory variables α and $\boldsymbol{\pi}$ are $\sqrt{N}$ and $\sqrt{NT}$ consistent respectively implying that all parameters are consistently estimated as only $N \to \infty$. We also note that we do not need $N \geq 2$ as $T \to \infty$ to identify the variance parameters $\left(\sigma_e^2, \rho_v\right)$.

The results in theorem 1 and corollaries 1 and 2 covers a number of interesting models commonly used in practice and it is useful to summarize the consistency properties obtained. This is done in Table 2.1.

2.3.3 Asymptotic normality

Comprehensive model

In this section our interest centers on the asymptotic distribution of the appropriately scaled maximum likelihood estimator $\widehat{\theta} = \left(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\gamma}}\right)$. Before the statement of the main theorem it is useful to collect some preliminary results which appear in lemma 6 and 7 in appendix B.

Recall that assumption (c) ensures that the part of the limiting information matrix which belongs to the explanatory variables is a positive-definite matrix as either or both of the indices grow large. In case both N and $T \to \infty$ this limiting matrix, denoted $\mathbf{R}$, is obviously non-stochastic. A moments consideration also reveals that this matrix depends on the behavior of the ratio $\frac{N}{T}$

Lemma 1 *If* $\frac{N}{T} \to \infty$

$$\mathbf{R} = \begin{bmatrix} \frac{(1-\rho_\lambda)^2}{\sigma_u^2} & 0 & \frac{(1-\rho_\lambda)^2}{\sigma_u^2} E\mathbf{d}_t' & 0 \\ & \mathbf{R_X} & 0 & 0 \\ & & \frac{1}{\sigma_u^2} E\mathbf{d}_t^\lambda \mathbf{d}_t^{\lambda\prime} & 0 \\ & & & \frac{1}{\sigma_\mu^2}\left(E\mathbf{h}_i\mathbf{h}_i' - E\mathbf{h}_i E\mathbf{h}_i'\right) \end{bmatrix}$$

where $\mathbf{R_X} = \text{plim}_{N,T\to\infty} \frac{1}{NT}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}$ *and* $\mathbf{d}_t^\lambda = (\mathbf{d}_t - \rho_\lambda \mathbf{d}_{t-1})$.
If $\frac{T}{N} \to \infty$

$$\mathbf{R} = \begin{bmatrix} \frac{1}{\sigma_\mu^2} & 0 & 0 & \frac{1}{\sigma_\mu^2} E\mathbf{h}_i' \\ & \mathbf{R_X} & 0 & 0 \\ & & \frac{1}{\sigma_u^2} E\mathbf{d}_t^\lambda \mathbf{d}_t^{\lambda\prime} - \frac{(1-\rho_\lambda)^2}{\sigma_u^2} E\mathbf{d}_t E\mathbf{d}_t' & 0 \\ & & & \frac{1}{\sigma_\mu^2} E\mathbf{h}_i\mathbf{h}_i' \end{bmatrix}$$

Finally, if $N, T \to \infty$ *simultaneously*

$$\mathbf{R} = \begin{bmatrix} \omega & 0 & \omega E\mathbf{d}_t' & \omega E\mathbf{h}_i' \\ & \mathbf{R_X} & 0 & 0 \\ & & \frac{1}{\sigma_u^2} E\mathbf{d}_t^\lambda \mathbf{d}_t^{\lambda\prime} + \upsilon_1 E\mathbf{d}_t E\mathbf{d}_t' & \omega E\mathbf{d}_t E\mathbf{h}_i' \\ & & & \frac{1}{\sigma_\mu^2} E\mathbf{h}_i\mathbf{h}_i' + \upsilon_2 E\mathbf{h}_i E\mathbf{h}_i' \end{bmatrix}$$

where $\omega = \frac{1}{\sigma_\mu^2 + (1-\rho_\lambda)^{-2}\sigma_u^2}$, $\upsilon_1 = \omega - \frac{(1-\rho_\lambda)^2}{\sigma_u^2}$ *and* $\upsilon_2 = \omega - \sigma_\mu^{-2}$.

The lemma shows that when both N and $T \to \infty$ the variance formula, and hence the amount of information in the sample, depends on the behavior of the ratio $\frac{N}{T}$. It is important to notice that this result does not relate to assumptions about the sampling behavior of the time-invariant and the individual-invariant explanatory variables. In fact it holds regardless of wether the time-invariant and the individual-invariant explanatory variables are regarded as fixed or stochastic. Note however that if these variables are centered then $\mathbf{R}$ reduces to a block-diagonal matrix that only depends on the behavior of the ratio $\frac{N}{T}$ through the constant term.

If only N or $T \to \infty$ as in theorem 1 (ii) and (iii) part of the parameter vector is not consistently estimated. In fact, only the subvectors $\theta^i = \left(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\gamma}^{(i)}\right), \boldsymbol{\gamma}^{(i)} = (\sigma_\mu^2, \sigma_e^2, \rho_v)$ and $\theta^t = \left(\boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\gamma}^{(t)}\right), \boldsymbol{\gamma}^{(t)} = (\sigma_e^2, \rho_v, \sigma_u^2, \rho_\lambda)$ are consistently estimated as N and $T \to \infty$ respectively. The limiting distributions of the subsets of consistently estimated parameters are of course only interesting if they are information block-diagonal to the inconsistent parameters. The following lemma, which is a direct consequence of lemma 7 in appendix B, is useful in this respect

Lemma 2 *As $N \to \infty$ the information matrix is block-diagonal between θ^i and $\left(\alpha, \boldsymbol{\pi}', \sigma_u^2, \rho_\lambda\right)$ and as $T \to \infty$ the information matrix is block-diagonal between θ^t and $\left(\alpha, \boldsymbol{\tau}', \sigma_\mu^2\right)$*

Motivated by this lemma the theorem below applies a mean-value expansion to the part of the score vector which belongs to the consistent subvectors. In addition the elements of the limiting information matrix relating to the consistently estimated subvectors does not depend on the nuisance parameters $\left(\alpha, \boldsymbol{\pi}', \sigma_u^2, \rho_\lambda\right)$ as $N \to \infty$ nor on the nuisance parameters $\left(\alpha, \boldsymbol{\tau}', \sigma_\mu^2\right)$ as $T \to \infty$. This fact is important since it implies that we can obtain useful approximate variance formulas for the subsets of consistently estimated parameters.

We now obtain the main result of this section. For this purpose define $\mathbf{F}_{NT}$, $\mathbf{F}_N$ and $\mathbf{F}_T$ as diagonal matrices with

$$
\begin{aligned}
\operatorname{diag} \mathbf{F}_{NT} &= \left\{\min\left(\sqrt{N}, \sqrt{T}\right), \mathbf{F}_\beta, \mathbf{F}_\pi, \mathbf{F}_\tau, \sqrt{N}, \sqrt{NT}, \sqrt{NT}, \sqrt{T}, \sqrt{T}\right\} \\
\operatorname{diag} \mathbf{F}_N &= \left\{\mathbf{F}_\beta, \mathbf{F}_\tau, \sqrt{N}, \sqrt{NT}, \sqrt{NT}\right\} \\
\operatorname{diag} \mathbf{F}_T &= \left\{\mathbf{F}_\beta, \mathbf{F}_\pi, \sqrt{NT}, \sqrt{NT}, \sqrt{T}, \sqrt{T}\right\}
\end{aligned}
$$

We shall also need notation for limits of submatrices of the quadratic form in assumption (c). Let $\mathbf{Z}_N = (\mathbf{X}, \mathbf{H})$, $\mathbf{Z}_T = (\mathbf{X}, \mathbf{D})$, $\boldsymbol{\Upsilon}_N$ and $\boldsymbol{\Upsilon}_T$ diagonal such that

$$
\begin{aligned}
\operatorname{diag} \boldsymbol{\Upsilon}_N &= (\mathbf{F}_\beta, \mathbf{F}_\tau) \\
\operatorname{diag} \boldsymbol{\Upsilon}_T &= (\mathbf{F}_\beta, \mathbf{F}_\pi)
\end{aligned}
$$

with

$$
\begin{aligned}
\operatorname*{plim}_{N \to \infty} \boldsymbol{\Upsilon}_N^{-1} \mathbf{Z}_N' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_N \boldsymbol{\Upsilon}_N^{-1} &= \mathbf{R}_N \\
\operatorname*{plim}_{T \to \infty} \boldsymbol{\Upsilon}_T^{-1} \mathbf{Z}_T' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_T \boldsymbol{\Upsilon}_T^{-1} &= \mathbf{R}_T
\end{aligned}
$$

and we further let $\overline{\theta}$ denote a sequence such that $\operatorname{plim} \overline{\theta} = \theta_0$

Theorem 2 *(Comprehensive model)*

(i) $\mathbf{F}_{NT}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{V}\left(\theta_0\right)\right)$ *as* $N, T \to \infty$, *where*

$$\mathbf{V}^{-1}\left(\theta_0\right) = - \operatorname*{plim}_{N,T\to\infty} \left[\mathbf{F}_{NT}^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\boldsymbol{\gamma})}{\partial\theta\partial\theta'}|_{\overline{\theta}}\right)\mathbf{F}_{NT}^{-1}\right] = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{-1}\left(\theta_0\right)_\gamma \end{bmatrix}$$

a finite non-singular matrix, with $\mathbf{R} = \mathbf{R}\left(\theta_0\right)$ *a* $\sum_{i=1}^{3} k_i + 1$ *dimensional matrix given in lemma 1 and* $\mathbf{V}^{-1}\left(\theta_0\right)_\gamma$ *is a diagonal matrix with*

$$\operatorname{diag} \mathbf{V}^{-1}\left(\theta_0\right)_\gamma = \left\{\frac{1}{2\sigma_{\mu0}^4}, \frac{1}{2\sigma_{e0}^4}, \frac{1}{\left(1-\rho_{v0}^2\right)}, \frac{1}{2\sigma_{u0}^4}, \frac{1}{\left(1-\rho_{\lambda0}^2\right)}\right\}$$

(ii) $\mathbf{F}_N\left(\widehat{\theta}^i - \theta_0^i\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{V}_N\left(\theta_0^i\right)\right)$ *as* $N \to \infty$ *(assuming* $T \geq 2$*), where*

$$\mathbf{V}_N^{-1}\left(\theta_0^i\right) = - \operatorname*{plim}_{N\to\infty} \left[\mathbf{F}_N^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\boldsymbol{\gamma})}{\partial\theta^i\partial\theta^{i\prime}}|_{\overline{\theta}^i}\right)\mathbf{F}_N^{-1}\right] = \begin{bmatrix} \mathbf{R}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_N^{-1}\left(\theta_0^i\right)_{\gamma^{(i)}} \end{bmatrix}$$

a finite non-singular matrix, with $\mathbf{R}_N = \mathbf{R}_N\left(\theta_0^i\right)$ *a* $k_1 + k_3$ *dimensional matrix and* $\mathbf{V}_N^{-1}\left(\theta_0^i\right)_{\gamma^{(i)}}$ *given by*

$$\mathbf{V}_N^{-1}\left(\theta_0^i\right)_{\gamma^{(i)}} = \frac{1}{2}\begin{bmatrix} \left(\frac{\sigma_{\varpi0}^2}{\sigma_{\mu0}^2\sigma_{\alpha0}^2}\right)^2 & \frac{(1-\rho_{v0})^2}{\sigma_{\alpha0}^2\sqrt{T}}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\boldsymbol{\Psi}_v\boldsymbol{\iota}_T^{\alpha} & \frac{\sigma_{e0}^2(1-\rho_{v0})^2}{\sigma_{\alpha0}^2\sqrt{T}}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{L}_v\boldsymbol{\iota}_T^{\alpha} \\ & \frac{1}{T\sigma_{\alpha0}^4} + \frac{(T-1)}{T\sigma_{e0}^4} & \frac{\sigma_{e0}^2}{T}\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{L}_v\right) \\ & & \frac{\sigma_{e0}^4}{T}\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_v\right)^2 \end{bmatrix}$$

where $\sigma_{\varpi0}^2 = \sigma_{\alpha0}^2 - \sigma_{e0}^2$

(iii) $\mathbf{F}_T\left(\widehat{\theta}^t - \theta_0^t\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{V}_T\left(\theta_0^t\right)\right)$ *as* $T \to \infty$ *(assuming* $N \geq 2$*), where*

$$\mathbf{V}_T^{-1}\left(\theta_0^t\right) = - \operatorname*{plim}_{T\to\infty} \left[\mathbf{F}_T^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\boldsymbol{\gamma})}{\partial\theta^t\partial\theta^{t\prime}}|_{\overline{\theta}^t}\right)\mathbf{F}_T^{-1}\right] = \begin{bmatrix} \mathbf{R}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_T^{-1}\left(\theta_0^t\right)_{\gamma^{(t)}} \end{bmatrix}$$

a finite non-singular matrix, with $\mathbf{R}_T = \mathbf{R}_T\left(\theta_0^t\right)$ *a* $k_1 + k_2$ *dimensional matrix and* $\mathbf{V}_T^{-1}\left(\theta_0^t\right)_{\gamma(t)}$ *given by*

$$\begin{aligned}
&\mathbf{V}_T^{-1}\left(\theta_0^t\right)_{\gamma(t)} \\
&= \lim_{T\to\infty}\frac{1}{2T}\begin{bmatrix}
\frac{1}{\sigma_{e0}^4}\mathbf{V}^{+}_{\mathbf{\Psi}_v,\mathbf{\Psi}_v} & \frac{1}{\sigma_{e0}^2}\mathbf{V}^{+}_{\mathbf{L}_v,\mathbf{\Psi}_v} & \frac{\sqrt{N}}{\sigma_{e0}^4}\mathbf{V}_{\mathbf{\Psi}_v,\mathbf{\Psi}_\lambda} & \frac{\sigma_u^2\sqrt{N}}{\sigma_{e0}^4}\mathbf{V}_{\mathbf{\Psi}_v,\mathbf{L}_\lambda} \\
 & \mathbf{V}^{+}_{\mathbf{L}_v,\mathbf{L}_v} & \frac{\sqrt{N}}{\sigma_{e0}^2}\mathbf{V}_{\mathbf{L}_v,\mathbf{\Psi}_\lambda} & \frac{\sigma_{u0}^2\sqrt{N}}{\sigma_{e0}^2}\mathbf{V}_{\mathbf{L}_v,\mathbf{L}_\lambda} \\
 & & \frac{N^2}{\sigma_{e0}^4}\mathbf{V}_{\mathbf{\Psi}_\lambda,\mathbf{\Psi}_\lambda} & \frac{\sigma_{u0}^2 N^2}{\sigma_{e0}^4}\mathbf{V}_{\mathbf{L}_\lambda,\mathbf{\Psi}_\lambda} \\
 & & & \frac{\sigma_{u0}^4 N^2}{\sigma_{e0}^4}\mathbf{V}_{\mathbf{L}_\lambda,\mathbf{L}_\lambda}
\end{bmatrix}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{V}_{\mathbf{F},\mathbf{P}} &= \operatorname{tr}\left(\left(\mathbf{\Psi}_v^{-1}\mathbf{F}\mathbf{\Psi}_v^{-1}\mathbf{P}\right)\left(\mathbf{I}_T - 2\mathbf{M}\right)\right) + \operatorname{tr}\left(\mathbf{\Psi}_v^{-1}\mathbf{F}\mathbf{M}\mathbf{\Psi}_v^{-1}\mathbf{P}\mathbf{M}\right) \\
\mathbf{V}^{+}_{\mathbf{F},\mathbf{P}} &= \operatorname{tr}\left(\left(\mathbf{\Psi}_v^{-1}\mathbf{F}\mathbf{\Psi}_v^{-1}\mathbf{P}\right)\left(\mathbf{I}_T - \frac{2}{N}\mathbf{M}\right)\right) + \frac{1}{N}\operatorname{tr}\left(\mathbf{\Psi}_v^{-1}\mathbf{F}\mathbf{M}\mathbf{\Psi}_v^{-1}\mathbf{P}\mathbf{M}\right)
\end{aligned}$$

and $\mathbf{M} = \left(\mathbf{I}_T + \frac{\sigma_{e0}^2}{N\sigma_{u0}^2}\mathbf{\Psi}_\lambda\mathbf{\Psi}_v\mathbf{\Psi}_\lambda^{-2}\right)^{-1}$

Corresponding asymptotic normality results for the standard two-way model and the two-way model with serially correlated time effects or serially correlated idiosyncratic errors follow directly from theorem 2.

One-way models

Asymptotic normality results for the one-way models considered in corollary 1 and 2 can be derived quite easily given theorem 2. We concentrate on the one-way model with individual effects in this section, corresponding qualitative results for the one-way model with time effects follow similarly.

The limiting information matrix is, as in the two-way model, block-diagonal between consistent and inconsistent parameters. This allows us to obtain the marginal limiting distribution of the consistently estimated parameters when $T \to \infty$ in the same manner as for the two-way model. Also, the limiting information matrix for the consistently estimated parameters does not depend on the inconsistent nuisance parameters, ensuring that we can estimate the limiting variance consistently in the $T \to \infty$ case. In contrast to the two-way model all parameters are consistent as $N \to \infty$ and we obtain joint asymptotic normality for the full parameter vector under $N \to \infty$ as well as $N, T \to \infty$.

Make the following definitions

$$\begin{aligned}
\operatorname{diag} \mathbf{F}_{NT}^{(i)} &= \left\{\sqrt{N}, \mathbf{F}_\beta, \mathbf{F}_\pi, \mathbf{F}_\tau, \sqrt{N}, \sqrt{NT}, \sqrt{NT}\right\} \\
\operatorname{diag} \mathbf{F}_{T}^{(i)} &= \left\{\mathbf{F}_\beta, \mathbf{F}_\pi, \sqrt{NT}, \sqrt{NT}\right\}
\end{aligned}$$

where $\mathbf{F}_{NT}^{(i)}, \mathbf{F}_{T}^{(i)}$ are diagonal matrices and $\mathbf{F}_\pi$ is as in assumption $(c_{(i)})$. We also define $\theta^{(i)t} = \left(\boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\gamma}^{(i)t}\right)$, $\boldsymbol{\gamma}^{(i)t} = \left(\sigma_e^2, \rho_v\right)$ and with some further obvious notation we have

Corollary 3 *(One-way model with individual effects)*

(i) $\mathbf{F}_{NT}^{(i)}\left(\widehat{\theta}^{(i)} - \theta_0^{(i)}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{V}^{(i)}\left(\theta_0^{(i)}\right)\right)$ *as $N, T \to \infty$, where*

$$\begin{aligned}
\left[\mathbf{V}^{(i)}\left(\theta_0^{(i)}\right)\right]^{-1} &= -\operatorname*{plim}_{N,T\to\infty}\left[\left(\mathbf{F}_{NT}^{(i)}\right)^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta}, \boldsymbol{\gamma}^{(i)})}{\partial\theta^{(i)}\partial\theta^{(i)\prime}}|_{\overline{\theta}^{(i)}}\right)\left(\mathbf{F}_{NT}^{(i)}\right)^{-1}\right] \\
&= \begin{bmatrix} \mathbf{R}^{(i)} & \mathbf{0} \\ \mathbf{0} & \left[\mathbf{V}^{(i)}\left(\theta_0^{(i)}\right)_{\gamma^{(i)}}\right]^{-1} \end{bmatrix}
\end{aligned}$$

a finite non-singular matrix, with $\mathbf{R}^{(i)} = \mathbf{R}^{(i)}\left(\theta_0^{(i)}\right)$ a $\sum_{i=1}^{3} k_i + 1$ dimensional matrix given by

$$\mathbf{R}^{(i)} = \begin{bmatrix}
\frac{1}{\sigma_\mu^2} & \mathbf{0} & \mathbf{0} & \frac{1}{\sigma_\mu^2} E\mathbf{h}_i \\
 & \mathbf{R}_{\mathbf{X}}^{(i)} & \mathbf{R}_{\mathbf{X},\mathbf{D}}^{(i)} & \mathbf{0} \\
 & & \frac{1}{\sigma_e^2} E\mathbf{d}_t^{v\prime}\mathbf{d}_t^v - \frac{(1-\rho_v)^2}{\sigma_e^2} E\mathbf{d}_t' E\mathbf{d}_t & \mathbf{0} \\
 & & & \frac{1}{\sigma_\mu^2} E\mathbf{h}_i'\mathbf{h}_i
\end{bmatrix}$$

where $\mathbf{R}_{\mathbf{X}}^{(i)} = \operatorname{plim}_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{A}^*\mathbf{X}_i$, $\mathbf{R}_{\mathbf{X},\mathbf{D}}^{(i)} = \operatorname{plim}_{N,T\to\infty} \frac{1}{NT}$ $\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{A}^*\mathbf{d}$, $\mathbf{d}_t^v = (\mathbf{d}_t - \rho_v \mathbf{d}_{t-1})$ *and* $\left[\mathbf{V}^{(i)}\left(\theta_0^{(i)}\right)_{\gamma^{(i)}}\right]^{-1}$ *is a diagonal matrix with*

$$\operatorname{diag}\left[\mathbf{V}^{(i)}\left(\theta_0^{(i)}\right)_{\gamma^{(i)}}\right]^{-1} = \left\{\frac{1}{2\sigma_{\mu 0}^4}, \frac{1}{2\sigma_{e0}^4}, \frac{1}{\left(1-\rho_{v0}^2\right)}\right\}$$

(ii) $\mathbf{F}_{NT}^{(i)}\left(\widehat{\theta}^{(i)} - \theta_0^{(i)}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{V}_N^{(i)}\left(\theta_0^{(i)}\right)\right)$ *as* $N \to \infty$ *(assuming* $T \geq 2$*), where*

$$\begin{aligned}\left[\mathbf{V}_N^{(i)}\left(\theta_0^{(i)}\right)\right]^{-1} &= -\operatorname*{plim}_{N\to\infty}\left[\left(\mathbf{F}_{NT}^{(i)}\right)^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\boldsymbol{\gamma}^{(i)})}{\partial\theta^{(i)}\partial\theta^{(i)\prime}}|_{\overline{\theta}^{(i)}}\right)\left(\mathbf{F}_{NT}^{(i)}\right)^{-1}\right] \\ &= \begin{bmatrix}\mathbf{R}_N^{(i)} & \mathbf{0} \\ \mathbf{0} & \left[\mathbf{V}_N^{(i)}\left(\theta_0^{(i)}\right)_{\gamma^{(i)}}\right]^{-1}\end{bmatrix}\end{aligned}$$

a finite non-singular matrix, with $\mathbf{R}_N^{(i)} = \mathbf{R}_N^{(i)}\left(\theta_0^{(i)}\right)$ *a* $\sum_{i=1}^3 k_i + 1$ *dimensional matrix and* $\left[\mathbf{V}_N^{(i)}\left(\theta_0^{(i)}\right)_{\gamma^{(i)}}\right]^{-1}$ *given by*

$$\left[\mathbf{V}_N^{(i)}\left(\theta_0^{(i)}\right)_{\gamma^{(i)}}\right]^{-1} = \frac{1}{2}\begin{bmatrix}\left(\frac{\sigma_{\varpi 0}^2}{\sigma_{\mu 0}^2\sigma_{\alpha 0}^2}\right)^2 & \frac{(1-\rho_{v0})^2}{\sigma_{\alpha 0}^2\sqrt{T}}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\boldsymbol{\Psi}_v\boldsymbol{\iota}_T^{\alpha} & \frac{\sigma_{e0}^2(1-\rho_{v0})^2}{\sigma_{\alpha 0}^2\sqrt{T}}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{L}_v\boldsymbol{\iota}_T^{\alpha} \\ & \frac{1}{T\sigma_{\alpha 0}^4} + \frac{(T-1)}{T\sigma_{e0}^4} & \frac{\sigma_{e0}^2}{T}\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{L}_v\right) \\ & & \frac{\sigma_{e0}^4}{T}\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_v\right)^2\end{bmatrix}$$

where $\sigma_{\varpi 0}^2 = \sigma_{\alpha 0}^2 - \sigma_{e0}^2$

(iii) $\mathbf{F}_T^{(i)}\left(\widehat{\theta}^{(i)t} - \theta_0^{(i)t}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{V}_T^{(i)}\left(\theta_0^{(i)t}\right)\right)$ *as* $T \to \infty$*, where*

$$\begin{aligned}\left[\mathbf{V}_T^{(i)}\left(\theta_0^{(i)t}\right)\right]^{-1} &= -\operatorname*{plim}_{T\to\infty}\left[\left(\mathbf{F}_T^{(i)}\right)^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\boldsymbol{\gamma}^{(i)t})}{\partial\theta^{(i)t}\partial\theta^{(i)t\prime}}|_{\overline{\theta}^{(i)t}}\right)\left(\mathbf{F}_T^{(i)}\right)^{-1}\right] \\ &= \begin{bmatrix}\mathbf{R}_T^{(i)} & \mathbf{0} \\ \mathbf{0} & \left[\mathbf{V}_T^{(i)}\left(\theta_0^{(i)t}\right)_{\gamma^{(i)t}}\right]^{-1}\end{bmatrix}\end{aligned}$$

a finite non-singular matrix, with $\mathbf{R}_T^{(i)} = \mathbf{R}_T^{(i)}\left(\theta_0^{(i)t}\right)$ *a* $k_1 + k_2$ *dimensional matrix and* $\left[\mathbf{V}_T^{(i)}\left(\theta_0^{(i)t}\right)_{\gamma^{(i)t}}\right]^{-1}$ *is a diagonal matrix with*

$$\operatorname{diag}\left[\mathbf{V}_T^{(i)}\left(\theta_0^{(i)t}\right)_{\gamma^{(i)t}}\right]^{-1} = \left\{\frac{1}{2\sigma_{e0}^4}, \frac{1}{\left(1-\rho_{v0}^2\right)}\right\}$$

Comparing the results in the corollary above to the results in theorem 2 we note that

Property 1 *In contrast to the comprehensive model the information elements of $\alpha, \boldsymbol{\pi}$ and $\boldsymbol{\tau}$ does not depend on the behavior of the ratio $\frac{N}{T}$ as both N and T grows large*

Property 2 *As $N \to \infty$ (or $N, T \to \infty$) the variance matrix of the variance parameters $\gamma^{(i)}$ is the same in both models*

That is we have the same large N asymptotics for the variance parameters, $\gamma^{(i)}$ in the one-way model with individual effects and the two-way model. Noting that the one-way model with individual specific random effects is typically used in situations where large N asymptotics are appropriate this indicates that it is asymptotically costless to variance robustify by including time specific random effects as well. If in addition $\mathbf{h}$ is centered and $\mathbf{X}$ is centered in the N dimension we have the same large N limiting variance in these models for the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ as well.

2.3.4 Misspecification

It is well-known that in the framework of the classical linear model misspecification of the variance does in general not affect consistency of the regression parameters, only efficiency. Unfortunately, in the present situation this need not be true. As indicated by the results in theorem 1 and corollaries 1 and 2 problems arise since the true and the perceived error component structure need not agree on the appropriate probabilistic orders[3]. The theorem below illustrates what can happen

Theorem 3 (Misspecification of error components) *Suppose that the assumptions (a), (b) and (c) holds and the true model is the comprehensive model considered in theorem 1 but the estimated model is the one-way model with individual effects considered in corollary 2. Then, for $k_i = 1, i = 1, 2, 3$*

(i) $\widehat{\gamma}^{(i)t}$ is inconsistent as $N \to \infty, T \to \infty$ or $N, T \to \infty$ and $\widehat{\sigma}^2_\mu$ is inconsistent as $N \to \infty$

[3] Misspecification of the error components imply that the variance of the score and the negative expected hessian need not be equal. In addition they need not have the same probabilistic orders.

(ii) *As both $N, T \to \infty$ (it does not matter how) $\widehat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$ on Θ and as $T \to \infty$ $\left(\widehat{\beta}, \widehat{\pi}\right) \xrightarrow{p} (\beta_0, \pi_0)$ on Θ whereas $\widehat{\alpha}$ and $\widehat{\tau}$ are inconsistent. In case of $N \to \infty$ both $(\widehat{\alpha}, \widehat{\pi})$ are inconsistent, the situation for $\left(\widehat{\beta}, \widehat{\tau}\right)$ requires us to distinguish between if $\mathrm{plim}_{N\to\infty} \frac{1}{N} \sum X_{it} h_i$ is zero or not. If non-zero, $\widehat{\beta} \xrightarrow{p} \beta_0$ or $\widehat{\tau} \xrightarrow{p} \tau_0$ (or both) on Θ iff $\mathbf{h}$ is centered and $\mathbf{X}$ is centered in the N dimension. If zero, $\widehat{\beta} \xrightarrow{p} \beta_0$ on Θ iff $\mathbf{X}$ is centered in the N dimension and $\widehat{\tau} \xrightarrow{p} \tau_0$ on Θ iff $\mathbf{h}$ is centered*

Part two of the theorem might seem counterintuitive in the light of standard theory for linear regression. The key to understanding the result is to note that it is the case $N, T \to \infty$, where all regression parameters are estimated consistently, that corresponds to the standard theory. In the $N \to \infty$ case we may think of the time effects as dummy variables erroneously excluded from the model. Consistent estimation of the remaining regression parameters then requires that the corresponding explanatory variables are orthogonal to the excluded variables, hence the need for centering. Although centering of the data recovers some of the consistency properties for the mean parameters of a correctly specified one-way model it does not, and in contrast to the robustification result in property 2, lead to the same asymptotic distribution[4]. There is a loss of efficiency and a sandwich-type variance-covariance estimator should be used since the information matrix equality fails to hold. Also note that the driving force for the result is the presence of the time specific effects per se. Theorem 3 holds wether λ_t is serially correlated or not.

2.4 Final remarks

Panel data models which allow for serial correlation are extensively used in applied econometrics. This essay has explored the large sample theory for a comprehensive specification which nests most of the models used in practice[5]. In contrast to the previous literature we have treated the constant term appropriately as well as allowed for both time or individual-invariant random variables.

[4]The reader may notice that although we are only able to recover some of the consistency properties of the mean parameters in a correctly specified one-way model we obtain exactly the same consistency properties of the mean parameters as for the two-way model.

[5]Of course, none of the results in this essay are special to models with serial correlation. In addition the results for the variance components and the ordinary explanatory variables do not depend on the presence or non-presence of individual or time-invariant explanatory variables.

In terms of the consistency properties obtained our results reveal an interesting and, perhaps, unexpected difference between ordinary explanatory variables and explanatory variables that are time or individual-invariant. Whereas the parameters of ordinary explanatory variables are always estimated consistently whenever N or $T \to \infty$ the consistency properties of the parameters of time or individual-invariant explanatory variables depend crucially on the model. The source of this difference was attributed to confounding with time effects and/or individual effects and, of course, if there are neither individual nor time effects these parameters have the desirable properties of the parameters of ordinary explanatory variables.

Our results on asymptotic normality revealed a useful characterization of the limiting information matrix. The set of consistent parameters (as N or $T \to \infty$) are information block-diagonal to the set of inconsistent parameters and the set of consistent mean parameters are always information block-diagonal to the set of consistent variance parameters. In addition the elements of information of the consistent parameters do not depend on the inconsistent parameters, ensuring that the variance matrix of consistently estimated parameters can be consistently estimated.

As an application of the results obtained we considered the consequences of error component misspecification. In this situation it is useful to work with deviations from means to guard (incompletely) against possible inconsistency of the mean parameters and indeed the idea of centering is also useful in the context of robustification.

Possible extensions of our results include introducing dynamics in form of a lagged dependent variable as well as allowing for time trends commonly employed in practice. Given the present results one would suspect that a linear time trend is $T^{3/2}$ consistent in the two-way model and the one-way model with time effects but $\sqrt{N}T^{3/2}$ consistent in the one-way model with individual effects. However these and other issues are left for future work.

Appendix A

Score and Information

A.1 The score vector

This appendix derives the elements of the score vector. For the regression parameters we have the standard result

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

and for the variance parameters the score is given by

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

where $\gamma = (\sigma_\mu^2, \sigma_e^2, \rho_v, \sigma_u^2, \rho_\lambda)$

For σ_μ^2 we have

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

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

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

Hence

$$\begin{aligned}\frac{\partial l}{\partial \sigma_\mu^2} &= -\frac{N(1-\rho_v)^2 d^2}{2\sigma_\alpha^2} + \frac{N(1-\rho_v)^2}{2\sigma_\alpha^2}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha} \\ &\quad + \frac{1}{2}\boldsymbol{\varepsilon}'(\mathbf{I}_N \otimes \mathbf{A}^*\mathbf{J}_T\mathbf{A}^*)\boldsymbol{\varepsilon} - \widetilde{\boldsymbol{\varepsilon}}'\mathbf{J}_T\overline{\boldsymbol{\varepsilon}} + \frac{N}{2}\overline{\boldsymbol{\varepsilon}}'\mathbf{J}_T\overline{\boldsymbol{\varepsilon}}\end{aligned}$$

where $\widetilde{\boldsymbol{\varepsilon}} = (\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)\boldsymbol{\varepsilon}$ and $\overline{\boldsymbol{\varepsilon}} = \left(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\right)\boldsymbol{\varepsilon}$. For σ_e^2 we have

$$\begin{aligned}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_e^2}\right) &= \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{I}_N \otimes \boldsymbol{\Psi}_v)\right) \\ &= \operatorname{tr}(\mathbf{I}_N \otimes \mathbf{A}^*\boldsymbol{\Psi}_v) - \operatorname{tr}\left[(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\mathbf{B}^{-1}\left(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*\boldsymbol{\Psi}_v\right)\right] \\ &= N\operatorname{tr}(\mathbf{A}^*\boldsymbol{\Psi}_v) - N\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\right)\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_e^2}\boldsymbol{\Sigma}^{-1} &= \boldsymbol{\Sigma}^{-1}(\mathbf{I}_N \otimes \boldsymbol{\Psi}_v)\boldsymbol{\Sigma}^{-1} \\ &= \mathbf{I}_N \otimes \mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^* - (\mathbf{J}_N \otimes \mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*) \\ &\quad -(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*) \\ &\quad +N(\mathbf{J}_N \otimes \mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*)\end{aligned}$$

with

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

For σ_u^2 we have

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

and

$$\frac{\partial l}{\partial \sigma_u^2} = -\frac{N}{2}\operatorname{tr}(\mathbf{A}^*\mathbf{\Psi}_\lambda) + \frac{N^2}{2}\operatorname{tr}(\mathbf{A}^*\mathbf{\Psi}_\lambda\mathbf{A}^*\mathbf{B}^{-1}) + \frac{1}{2}\widetilde{\varepsilon}'\mathbf{\Psi}_\lambda\widetilde{\varepsilon} - N\widetilde{\varepsilon}'\mathbf{\Psi}_\lambda\overline{\varepsilon} + \frac{N^2}{2}\overline{\varepsilon}'\mathbf{\Psi}_\lambda\overline{\varepsilon}$$

Finally for the parameters in $\mathbf{\Psi}_\lambda$ and $\mathbf{\Psi}_v$ let $\mathbf{L}_\lambda = \frac{\partial \mathbf{\Psi}_\lambda}{\partial \rho_\lambda} = \frac{2\rho}{1-\rho^2}\mathbf{\Psi}_\lambda + \frac{1}{1-\rho^2}\mathbf{D}$ where $\mathbf{D}$ is a band matrix with zeros on the main diagonal and $i\rho_\lambda^{i-1}$ on the i$^{\text{th}}$ subdiagonal and define $\mathbf{L}_v$ similarly, we then have

$$\frac{\partial l}{\partial \rho_\lambda} = -\frac{\sigma_u^2 N}{2}\operatorname{tr}(\mathbf{A}^*\mathbf{L}_\lambda) + \frac{\sigma_u^2 N^2}{2}\operatorname{tr}(\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\mathbf{B}^{-1}) + \frac{\sigma_u^2}{2}\widetilde{\varepsilon}'\mathbf{L}_\lambda\widetilde{\varepsilon} - \sigma_u^2 N\widetilde{\varepsilon}'\mathbf{L}_\lambda\overline{\varepsilon} + \frac{N^2\sigma_u^2}{2}\overline{\varepsilon}'\mathbf{L}_\lambda\overline{\varepsilon}$$

and

$$\frac{\partial l}{\partial \rho_v} = -\frac{\sigma_e^2 N}{2}\operatorname{tr}(\mathbf{A}^*\mathbf{L}_v) + \frac{\sigma_e^2 N}{2}\operatorname{tr}(\mathbf{A}^*\mathbf{L}_v\mathbf{A}^*\mathbf{B}^{-1}) + \frac{\sigma_e^2}{2}\varepsilon'\left(\mathbf{I}_N \otimes \mathbf{A}^*\mathbf{L}_v\mathbf{A}^*\right)\varepsilon - \sigma_e^2\widetilde{\varepsilon}'\mathbf{L}_v\overline{\varepsilon} + \frac{N\sigma_e^2}{2}\overline{\varepsilon}'\mathbf{L}_v\overline{\varepsilon}$$

A.2 The information matrix

This appendix derives the elements of the information matrix. For the first element we have the result

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

and the elements $\mathcal{I}_{\delta,\gamma_i}$ are simply computed as

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

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

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

For the elements involving σ_μ^2

$$\mathcal{I}_{\sigma_\mu^2,\sigma_\mu^2} = \frac{1}{2}\left[\begin{array}{c} N\sigma_\alpha^{-4}d^4(1-\rho_v)^4 - 2N\sigma_\alpha^{-4}d^2(1-\rho_v)^4\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha} \\ +N^2\sigma_\alpha^{-4}(1-\rho_v)^4\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha}\end{array}\right]$$

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

$$\mathcal{I}_{\sigma_\mu^2,\sigma_e^2} = \frac{1}{2}\left[\begin{array}{c} N\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\boldsymbol{\Psi}_v\boldsymbol{\iota}_T^{\alpha} \\ -2N\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\boldsymbol{\iota}_T^{\alpha} \\ +N^2\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha} \end{array}\right]$$

$$\mathcal{I}_{\sigma_\mu^2,\sigma_u^2} = \frac{1}{2}\left[\begin{array}{c} N\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\boldsymbol{\Psi}_\lambda\boldsymbol{\iota}_T^{\alpha} \\ -2N^2\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_\lambda\boldsymbol{\iota}_T^{\alpha} \\ +N^3\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha} \end{array}\right]$$

$$\mathcal{I}_{\sigma_\mu^2,\rho_\lambda} = \frac{\sigma_u^2}{2}\left[\begin{array}{c} N\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{L}_\lambda\boldsymbol{\iota}_T^{\alpha} \\ -2N^2\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\boldsymbol{\iota}_T^{\alpha} \\ +N^3\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha} \end{array}\right]$$

$$\mathcal{I}_{\sigma_\mu^2,\rho_v} = \frac{\sigma_e^2}{2}\left[\begin{array}{c} N\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{L}_v\boldsymbol{\iota}_T^{\alpha} \\ -2N\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\boldsymbol{\iota}_T^{\alpha} \\ +N^2\sigma_\alpha^{-2}(1-\rho_v)^2\,\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha} \end{array}\right]$$

with $\mathbf{L}_\lambda$ and $\mathbf{L}_v$ defined in appendix A.1. Next for the relevant $\mathcal{I}_{\sigma_e^2,\gamma_j}$ elements

$$\mathcal{I}_{\sigma_e^2,\sigma_e^2} = \frac{1}{2}\left[\begin{array}{c} N\operatorname{tr}(\mathbf{A}^*\boldsymbol{\Psi}_v)^2 - 2N\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\right) \\ +N^2\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\right)^2 \end{array}\right]$$

$$\mathcal{I}_{\sigma_e^2,\sigma_u^2} = \frac{1}{2}\left[\begin{array}{c} N\operatorname{tr}(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\boldsymbol{\Psi}_\lambda) - 2N^2\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_\lambda\right) \\ +N^3\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_\lambda\mathbf{A}^*\right) \end{array}\right]$$

$$\mathcal{I}_{\sigma_e^2,\rho_\lambda} = \frac{\sigma_u^2}{2}\left[\begin{array}{c} N\operatorname{tr}(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{L}_\lambda) - 2N^2\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\right) \\ +N^3\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\right) \end{array}\right]$$

$$\mathcal{I}_{\sigma_e^2,\rho_v} = \frac{\sigma_e^2}{2}\left[\begin{array}{c} N\operatorname{tr}(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{L}_v) - 2N\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\right) \\ +N^2\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\mathbf{A}^*\right) \end{array}\right]$$

Finally for the elements involving σ_u^2, ρ_λ and ρ_v we have

$$\mathcal{I}_{\sigma_u^2,\sigma_u^2} = \frac{1}{2}\left[\begin{array}{c} N^2\operatorname{tr}(\mathbf{A}^*\boldsymbol{\Psi}_\lambda)^2 - 2N^3\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_\lambda\right) \\ +N^4\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_\lambda\mathbf{A}^*\right)^2 \end{array}\right]$$

$$\mathcal{I}_{\sigma_u^2,\rho_\lambda} = \frac{\sigma_u^2}{2}\left[\begin{array}{c} N^2\operatorname{tr}(\mathbf{A}^*\boldsymbol{\Psi}_\lambda\mathbf{A}^*\mathbf{L}_\lambda) - 2N^3\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\right) \\ +N^4\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\boldsymbol{\Psi}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\right) \end{array}\right]$$

$$\mathcal{I}_{\sigma_u^2,\rho_v} = \frac{\sigma_e^2}{2}\left[\begin{array}{c} N\operatorname{tr}\left(\mathbf{A}^*\mathbf{\Psi}_\lambda\mathbf{A}^*\mathbf{L}_v\right) - 2N^2\operatorname{tr}\left(\mathbf{A}^*\mathbf{\Psi}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\right) \\ +N^3\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\mathbf{A}^*\right) \end{array}\right]$$

$$\mathcal{I}_{\rho_\lambda,\rho_\lambda} = \frac{\sigma_u^4}{2}\left[\begin{array}{c} N^2\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_\lambda\right)^2 - 2N^3\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\right) \\ +N^4\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\right)^2 \end{array}\right]$$

$$\mathcal{I}_{\rho_\lambda,\rho_v} = \frac{\sigma_u^2\sigma_e^2}{2}\left[\begin{array}{c} N\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\mathbf{L}_v\right) - 2N^2\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\right) \\ +N^3\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_\lambda\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\mathbf{A}^*\right) \end{array}\right]$$

$$\mathcal{I}_{\rho_v,\rho_v} = \frac{\sigma_e^4}{2}\left[\begin{array}{c} N\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_v\right)^2 - 2N\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_v\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\right) \\ +N^2\operatorname{tr}\left(\mathbf{B}^{-1}\mathbf{A}^*\mathbf{L}_v\mathbf{A}^*\right)^2 \end{array}\right]$$

Appendix B

Proofs

A number of expressions involving the components of the variance matrix $\mathbf{\Sigma}$ appear frequently in the proofs. A series of lemmas below summarizes some basic results for these expressions. Unless otherwise indicated in case of joint convergence $(N, T \to \infty)$ no restriction on the indices are needed and joint limits can also be computed as sequential limits by letting $T \to \infty$ followed by $N \to \infty$, see Phillips and Moon (1999, corollary 1).

Lemma 3 *Let* $\mathbf{C}$ *be the Prais-Winsten transformation matrix for an* $AR(1)$ *process with parameter* ρ, $\mathbf{\Psi}$ *the variance covariance matrix of an* $AR(1)$ *process with parameter* r *and unit variance and let* $\boldsymbol{\iota}_T^\alpha$ *be a vector with first element* $\sqrt{(1+\rho)/(1-\rho)}$ *and remaining* $T-1$ *elements unity.*

$$\begin{aligned}
\operatorname{tr}\left(\mathbf{C\Psi C'}\right) &= \frac{2}{1-r^2} - \frac{2(T-1)}{1-r^2} r\rho + \frac{(T-2)}{1-r^2}\left(\rho^2+1\right) \\
\lim_{T\to\infty}\frac{1}{T}\operatorname{tr}\left(\mathbf{C\Psi C'}\right) &= \frac{\left(\rho^2+1\right)-2r\rho}{1-r^2}
\end{aligned}$$

$$\begin{aligned}
\operatorname{tr}\left(\mathbf{C\Psi C'}\boldsymbol{\iota}_T^\alpha\boldsymbol{\iota}_T^{\alpha\prime}\right) &= \boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C\Psi C'}\boldsymbol{\iota}_T^\alpha \\
&= \frac{1}{c_1}\left(\frac{1+\rho}{1-\rho}\right)\left(1-\rho^2\right) + \frac{2c_2}{c_1}\sqrt{\left(\frac{1+\rho}{1-\rho}\right)}\sum_{j=0}^{T-2} r^j \\
&\quad + \frac{c_3(T-1)}{c_1} + \frac{2}{c_1}\begin{pmatrix}\rho\left(r\rho-r^2\right) \\ +(r-\rho)\end{pmatrix}\sum_{j=2}^{T-1}(T-j)\,r^{j-2} \\
\lim_{T\to\infty}\frac{1}{T}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C\Psi C'}\boldsymbol{\iota}_T^\alpha &= \frac{1}{c_1}\left(1-2r\rho+p^2\right) - \frac{2}{c_1(r-1)}\left(\rho\left(r\rho-r^2\right)+(r-\rho)\right)
\end{aligned}$$

where $c_1 = \left(1 - r^2\right)$, $c_2 = (r - \rho)\left(1 - \rho^2\right)^{1/2}$ *and* $c_3 = \left(1 - 2r\rho + p^2\right)$. *Note that theses matrices are independent of* N *and that the limits hold when* $N, T \to \infty$ *as well.*

Lemma 4 *Let* $\mathbf{A}^* = \mathbf{C}'\left(\sigma_\alpha^{-2}\overline{\mathbf{J}}_T^\alpha + \sigma_e^{-2}\overline{\mathbf{E}}_T^\alpha\right)\mathbf{C}$ *and consider* $\operatorname{vech}\mathbf{A}^*$ *we then have elementwise convergence of* $\operatorname{vech}\mathbf{A}^*$ *to the infinite sequence* $\operatorname{vech}\left(\frac{1}{\sigma_e^2}\mathbf{\Psi}_{v,\infty}^{-1}\right)$ *at the rate* T^{-1} *as* $T \to \infty$.

Let $\mathbf{B} = \mathbf{\Psi}_\lambda^{-1} + N\sigma_u^2\mathbf{A}^*$ *we then have*

$$\lim_{T\to\infty} \operatorname{vech}\mathbf{B}^{-1} = \operatorname{vech}\left(\mathbf{\Psi}_{\lambda,\infty}^{-1} + N\sigma_u^2\sigma_e^{-2}\mathbf{\Psi}_{v,\infty}^{-1}\right)^{-1}$$

$$\begin{aligned}
\lim_{N\to\infty} \operatorname{vech}\mathbf{B}^{-1} &= \mathbf{0} \\
\lim_{N\to\infty} \boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^\alpha &= 0 \\
\lim_{T\to\infty} \frac{1}{T^p}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^\alpha &= 0, \forall p > 1 \\
\lim_{N\to\infty} \operatorname{tr}\left(\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}\right) &= 0
\end{aligned}$$

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\left|\mathbf{B}^{-1}\right| &= 0 \\
\lim_{T\to\infty} \frac{1}{T}\left|\mathbf{B}^{-1}\right| &= 0 \\
\lim_{N,T\to\infty} \frac{1}{NT}\left|\mathbf{B}^{-1}\right| &= 0
\end{aligned}$$

and

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\ln\left|\mathbf{B}^{-1}\right| &= 0 \\
\lim_{N,T\to\infty} \frac{1}{NT}\ln\left|\mathbf{B}^{-1}\right| &= 0.
\end{aligned}$$

Proof. To obtain the elementwise convergence of vech $\mathbf{A}^*$ we write

$$\mathbf{A}^* = \mathbf{C}'\left(\frac{1}{d^2\sigma_\mu^2\left(1-\rho_v\right)^2 + \sigma_e^2}\frac{\boldsymbol{\iota}_T^\alpha\boldsymbol{\iota}_T^{\alpha\prime}}{d^2} + \sigma_e^{-2}\left(\mathbf{I}_T - \frac{\boldsymbol{\iota}_T^\alpha\boldsymbol{\iota}_T^{\alpha\prime}}{d^2}\right)\right)\mathbf{C}$$

and note that $\iota_T^{\alpha\prime}\mathbf{C} = \left(\sqrt{\frac{(1+\rho_v)(1-\rho_v^2)}{1-\rho_v}} - \rho_v, 1-\rho_v, \ldots, 1-\rho_v, 1\right)$. The elements of $\frac{1}{d^2}\mathbf{C}'\iota_T^{\alpha}\iota_T^{\alpha\prime}\mathbf{C} \to 0$ as $T \to \infty$ since $d^2 = \alpha^2 + (T-1)$. Next the established limit for $\operatorname{vech}\mathbf{B}^{-1}$ as $T \to \infty$ follows from the elementwise convergence of $\operatorname{vech}\mathbf{A}^*$ as $T \to \infty$ and $\lim_{N\to\infty} \operatorname{vech}\mathbf{B}^{-1} = \mathbf{0}$ follows since $\lim_{N\to\infty} \frac{1}{N}\mathbf{B} = \sigma_u^2\mathbf{A}^*$ elementwise. Then $\lim_{N\to\infty} \iota_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\iota_T^{\alpha} = 0$ and $\lim_{N\to\infty} \operatorname{tr}\left(\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}\right) = 0$ follows immediately from $\lim_{N\to\infty} \operatorname{vech}\mathbf{B}^{-1} = \mathbf{0}$. To establish the $T \to \infty$ limit of $\frac{1}{T^p}\iota_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\iota_T^{\alpha} = 0$, $p > 1$ we note that for N fix and $T \to \infty$ $\operatorname{vech}\mathbf{B}^{-1}$ converges elementwise to the infinite sequence $\operatorname{vech}\left(\mathbf{\Psi}_{\lambda,\infty}^{-1} + N\sigma_u^2\sigma_e^{-2}\mathbf{\Psi}_{v,\infty}^{-1}\right)^{-1}$ which has the form of the inverse of a MA(1) covariance matrix, that is the off-diagonal elements decay exponentially. Since $\mathbf{A}^*$ converges elementwise to a band-diagonal matrix it follows that $\mathbf{A}^*\mathbf{B}^{-1}$ converges elementwise to a matrix with exponentially decaying off-diagonals. Hence $\frac{1}{T}\iota_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\iota_T^{\alpha}$ converges to a constant since this is the sum of the exponentially decaying elements in $\mathbf{A}^*\mathbf{B}^{-1}$ and

$$\lim_{T\to\infty} \frac{1}{T^p}\iota_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\iota_T^{\alpha} = 0, p > 1$$

follows.

To establish the limits for $\left|\mathbf{B}^{-1}\right|$ we note that $\mathbf{B} - \mathbf{\Psi}_\lambda^{-1} = N\sigma_u^2\mathbf{A}^*$ is positive definite which implies that $|\mathbf{B}| > \left|\mathbf{\Psi}_\lambda^{-1}\right| = 1 - \rho_\lambda^2$ and $\left|\mathbf{B}^{-1}\right| < |\mathbf{\Psi}_\lambda| = \frac{1}{1-\rho_\lambda^2}$. In addition $\left|\mathbf{B}^{-1}\right| > 0$ since $\mathbf{B}$ is positive definite and the results follow.

For $\ln\left|\mathbf{B}^{-1}\right|$ we have $\ln\left|\mathbf{B}^{-1}\right| < -\ln\left(1-\rho_\lambda^2\right)$, a lower bound is obtained from the Hadamard determinant theorem,

$$|\mathbf{B}| \leq \prod_{j=1}^{T} b_{jj} = \prod_{j=1}^{T}\left[\psi^{jj} + N\sigma_u^2 a_{jj}^*\right]$$

implying $\ln\left|\mathbf{B}^{-1}\right| \geq -\sum_{j=1}^{T}\ln\left[\psi^{jj} + N\sigma_u^2 a_{jj}^*\right] \geq -\sum_{j=1}^{T}\ln\left(1+\rho_\lambda^2+Nk\right) = -T\ln\left(1+\rho_\lambda^2+Nk\right)$ where $k = \max a_{jj}^*$. Note that k depends on T and approaches $\sigma_u^2\left(1+\rho_\lambda^2\right)/\sigma_e^2$ as $T \to \infty$ ■

Lemma 5 *Let $\mathbf{\Sigma}_0$ be the variance matrix $\mathbf{\Sigma}$ evaluated at θ_0. Then*

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{NT}\operatorname{tr}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0 &= \left(1-\frac{1}{N}\right)\frac{\sigma_{e0}^2}{\sigma_e^2}\left[\frac{\left(\rho_v^2+1\right)-2\rho_{v0}\rho_v}{1-\rho_{v0}^2}\right] \\ &\quad + \frac{1}{NT}\lim_{T\to\infty}\operatorname{tr}\left(\mathbf{P}_0\mathbf{P}^{-1}\right)\end{aligned}$$

where $\mathbf{P}_0 = \left(N\sigma_{u0}^2\mathbf{\Psi}_{\lambda 0} + \sigma_{e0}^2\mathbf{\Psi}_{v0}\right)$, $\mathbf{P} = \left(N\sigma_u^2\mathbf{\Psi}_\lambda + \sigma_e^2\mathbf{\Psi}_v\right)$.

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{NT}\operatorname{tr}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0 &= \frac{\sigma_{\mu 0}^2}{T\sigma_\alpha^2}d^2\left(1-\rho_v\right)^2 + \frac{\sigma_{e0}^2}{T\sigma_e^2}\operatorname{tr}\left(\mathbf{C\Psi}_{v0}\mathbf{C}'\right) \\ &\quad + \frac{\sigma_{e0}^2\left(\sigma_\alpha^{-2}-\sigma_e^{-2}\right)}{Td^2}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C\Psi}_{v0}\mathbf{C}'\boldsymbol{\iota}_T^\alpha\end{aligned}$$

with $\operatorname{tr}\left(\mathbf{C\Psi}_{v0}\mathbf{C}'\right)$ *and* $\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C\Psi}_{v0}\mathbf{C}'\boldsymbol{\iota}_T^\alpha$ *evaluated in lemma 3, and*

$$\lim_{N,T\to\infty}\frac{1}{NT}\operatorname{tr}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0 = \frac{\sigma_{e0}^2}{\sigma_e^2}\left[\frac{\left(\rho_v^2+1\right)-2\rho_{v0}\rho_v}{1-\rho_{v0}^2}\right]$$

Proof. Standard matrix algebra yields

$$\begin{aligned}\frac{1}{NT}\operatorname{tr}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_0 &= \frac{1}{T}\sigma_{\mu 0}^2\sigma_\alpha^{-2}d^2\left(1-\rho_v\right)^2 \\ &\quad + \frac{\sigma_{e0}^2\left(\sigma_\alpha^{-2}-\sigma_e^{-2}\right)}{Td^2}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C\Psi}_{v0}\mathbf{C}'\boldsymbol{\iota}_T^\alpha \\ &\quad + \frac{\sigma_{u0}^2\left(\sigma_\alpha^{-2}-\sigma_e^{-2}\right)}{Td^2}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C\Psi}_{\lambda 0}\mathbf{C}'\boldsymbol{\iota}_T^\alpha \\ &\quad + \frac{1}{T}\sigma_{e0}^2\sigma_e^{-2}\operatorname{tr}\left(\mathbf{C\Psi}_{v0}\mathbf{C}'\right) + \frac{1}{T}\sigma_{u0}^2\sigma_e^{-2}\operatorname{tr}\left(\mathbf{C\Psi}_{\lambda 0}\mathbf{C}'\right) \\ &\quad - \frac{1}{T}\sigma_{\mu 0}^2\sigma_u^2\sigma_\alpha^{-2}\left(1-\rho_v\right)^2\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^\alpha \\ &\quad - \frac{1}{T}\sigma_{e0}^2\sigma_u^2\operatorname{tr}(\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}_{v0}) \\ &\quad - \frac{1}{T}N\sigma_{u0}^2\sigma_u^2\operatorname{tr}(\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}_{\lambda 0})\end{aligned}$$

To establish the limit as $T\to\infty$ note that for the first term

$$\lim_{T\to\infty}\frac{1}{T}\sigma_{\mu 0}^2\sigma_\alpha^{-2}d^2\left(1-\rho_v\right)^2 = 0$$

since $\sigma_{\mu 0}^2\sigma_\alpha^{-2}d^2\left(1-\rho_v\right)^2 = O(1)$. For the next two terms

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sigma_\alpha^{-2}d^{-2}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C\Psi}_{j0}\mathbf{C}'\boldsymbol{\iota}_T^\alpha &= 0\ , j=\lambda,v \\ \lim_{T\to\infty}\frac{1}{T}d^{-2}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C\Psi}_{j0}\mathbf{C}'\boldsymbol{\iota}_T^\alpha &= 0\ , j=\lambda,v\end{aligned}$$

follows from lemma 3 since $\sigma_\alpha^2 = O(T)$ and $d^2 = O(T)$. The limits as $T\to\infty$ of the fourth and fifth term follow from lemma 3 and lemma 4 established that the sixth term converges to zero. For the last two terms we have by the elementwise convergence of vech $\mathbf{A}^*$ and vech $\mathbf{B}^{-1}$ established in lemma 4

$$\begin{aligned} & \lim_{T\to\infty} \frac{1}{T} \operatorname{tr}\left(\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}_{j0}\right) \\ = & \lim_{T\to\infty} \frac{1}{T} \operatorname{tr}\left(\sigma_e^{-2}\mathbf{\Psi}_v^{-1}\left(\mathbf{\Psi}_\lambda^{-1} + N\sigma_u^2\sigma_e^{-2}\mathbf{\Psi}_v^{-1}\right)^{-1}\sigma_e^{-2}\mathbf{\Psi}_v^{-1}\mathbf{\Psi}_{j0}\right) \end{aligned}$$

which is well defined (and non-zero) since the diagonal elements of the matrix are $O(1)$. Repeatedly applying elementary results on inverses of sums (Dhrymes (1984, p. 39)) to these last two terms and then collecting terms obtains the expression given in the theorem. This completes the proof of the $T \to \infty$ case.

Now consider the case when $N \to \infty$. Since all but the three last terms are independent of N we need only consider these. Then

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{T}\sigma_{\mu 0}^2\sigma_u^2\sigma_\alpha^{-2}(1-\rho_v)^2 \boldsymbol{\iota}_T^{\alpha\prime}\mathbf{A}^*\mathbf{B}^{-1}\boldsymbol{\iota}_T^{\alpha} &= 0 \\ \lim_{N\to\infty} \frac{1}{T}\sigma_{u0}^2\sigma_u^2 \operatorname{tr}(\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}_{\lambda 0}) &= 0 \end{aligned}$$

follows from lemma 3, and

$$\lim_{N\to\infty} \frac{1}{T} N\sigma_{u0}^2\sigma_u^2 \operatorname{tr}(\mathbf{A}^*\mathbf{B}^{-1}\mathbf{A}^*\mathbf{\Psi}_{\lambda 0}) = \frac{1}{T}\sigma_{u0}^2 \operatorname{tr}(\mathbf{\Psi}_{\lambda 0}\mathbf{A}^*)$$

follows since $\lim_{N\to\infty} N\mathbf{B}^{-1} = \left(\sigma_u^2\mathbf{A}^*\right)^{-1}$ elementwise. Collecting terms as in the $T \to \infty$ case then gives the result.

Finally the result for $N, T \to \infty$ follows by taking sequential limits and using lemma 3 ∎

The following lemma gives some basic limit results for the expressions $\mathbf{Z}'\mathbf{\Sigma}^{-1}\varepsilon$, $\boldsymbol{\iota}_{NT}'\mathbf{\Sigma}^{-1}\boldsymbol{\iota}_{NT}$, $\mathbf{D}'\mathbf{\Sigma}^{-1}\mathbf{D}$ and $\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H}$. In the proof of the results in this lemma we make extensive use of elementary results on inverses involving sums (Dhrymes (1984, p. 39)), applying them repeatedly to obtain manageable expressions.

Lemma 6 *As $N \to \infty, T \to \infty$ or $N, T \to \infty$*

$$\operatorname{plim} \frac{1}{NT}\mathbf{X}'\mathbf{\Sigma}^{-1}\varepsilon = \mathbf{0}$$

If both $N, T \to \infty$ and if $\frac{N}{T} \to \infty$

$$\begin{aligned} \operatorname*{plim}_{N,T\to\infty} \frac{1}{T}\boldsymbol{\iota}_{NT}'\mathbf{\Sigma}^{-1}\varepsilon &= 0 \\ \lim_{N,T\to\infty} \frac{1}{T}\boldsymbol{\iota}_{NT}'\mathbf{\Sigma}^{-1}\boldsymbol{\iota}_{NT} &= \frac{1}{\sigma_u^2}(1-\rho_\lambda)^2 \end{aligned}$$

If both $N,T \to \infty$ and if $\frac{T}{N} \to \infty$

$$\begin{aligned}
\underset{N,T\to\infty}{\text{plim}} \frac{1}{N}\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} &= 0 \\
\lim_{N,T\to\infty} \frac{1}{N}\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} &= \frac{1}{\sigma_\mu^2}
\end{aligned}$$

If $N,T \to \infty$ simultaneously

$$\begin{aligned}
\underset{N,T\to\infty}{\text{plim}} \frac{1}{N}\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} &= \underset{N,T\to\infty}{\text{plim}} \frac{1}{T}\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} = 0 \\
\lim_{N,T\to\infty} \frac{1}{N}\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} &= \lim_{N,T\to\infty} \frac{1}{T}\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} = \frac{1}{\sigma_\mu^2 + \sigma_u^2\left(1-\rho_\lambda\right)^{-2}}
\end{aligned}$$

As $T \to \infty$ or $N,T \to \infty$

$$\text{plim}\, \frac{1}{T}\mathbf{D}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} = \mathbf{0}$$

As $N \to \infty$

$$\text{plim}\, \frac{1}{NT}\mathbf{D}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} = \mathbf{0}$$

As $N \to \infty$ or $N,T \to \infty$

$$\text{plim}\, \frac{1}{N}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} = \mathbf{0}$$

As $T \to \infty$

$$\text{plim}\, \frac{1}{NT}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} = \mathbf{0}$$

If both $N,T \to \infty$ and if $\frac{N}{T} \to \infty$

$$\begin{aligned}
\underset{N,T\to\infty}{\text{plim}} \frac{1}{T}\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D} &= \frac{1}{T\sigma_{u0}^2} \underset{T\to\infty}{\text{plim}} \sum_{t=2}^{T} \left(\mathbf{d}_t - \rho_\lambda \mathbf{d}_{t-1}\right)\left(\mathbf{d}_t - \rho_\lambda \mathbf{d}_{t-1}\right)' \\
\underset{N,T\to\infty}{\text{plim}} \frac{1}{N}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{H} &= \frac{1}{\sigma_\mu^2} \underset{N\to\infty}{\text{plim}} \frac{1}{N}\mathbf{h}'\overline{\mathbf{E}}_N\mathbf{h}
\end{aligned}$$

where $\overline{\mathbf{E}}_N = \mathbf{I}_N - \overline{\mathbf{J}}_N, \overline{\mathbf{J}}_N = \frac{1}{N}\boldsymbol{\iota}_N\boldsymbol{\iota}'_N$. *If both $N,T \to \infty$ and if $\frac{T}{N} \to \infty$*

$$\begin{aligned}
\underset{N,T\to\infty}{\text{plim}} \frac{1}{T}\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D} &= \underset{T\to\infty}{\text{plim}} \frac{1}{T\sigma_u^2}\mathbf{S}_\lambda \\
\underset{N,T\to\infty}{\text{plim}} \frac{1}{N}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{H} &= \frac{1}{\sigma_\mu^2} \underset{N\to\infty}{\text{plim}} \frac{1}{N}\mathbf{h}'\mathbf{h}
\end{aligned}$$

where $\mathbf{S}_\lambda = \left(\sum_{t=2}^{T}(\mathbf{d}_t - \rho_\lambda \mathbf{d}_{t-1})(\mathbf{d}_t - \rho_\lambda \mathbf{d}_{t-1})' - \frac{(1-\rho_{\lambda 0})^2}{T}\sum_{t=2}^{T-1}\sum_{r=2}^{T-1}\mathbf{d}_t\mathbf{d}_r'\right)$.
Finally, if $N, T \to \infty$ *simultaneously*

$$\begin{aligned}
\operatorname*{plim}_{N,T\to\infty} \frac{1}{T}\mathbf{D}'\mathbf{\Sigma}^{-1}\mathbf{D} &= \operatorname*{plim}_{T\to\infty} \frac{1}{T\sigma_u^2}\mathbf{S}_\lambda + \operatorname*{plim}_{T\to\infty} \frac{1}{T^2\left(\sigma_\mu^2 + \frac{\sigma_u^2}{(1-\rho_\lambda)^2}\right)} \sum_{t=2}^{T-1}\sum_{r=2}^{T-1}\mathbf{d}_t\mathbf{d}_r' \\
\operatorname*{plim}_{N,T\to\infty} \frac{1}{N}\mathbf{H}'\mathbf{\Sigma}^{-1}\mathbf{H} &= \operatorname*{plim}_{N\to\infty} \frac{1}{N\sigma_\mu^2}\mathbf{h}'\overline{\mathbf{E}}_N\mathbf{h} + \operatorname*{plim}_{N\to\infty} \frac{1}{N\left(\sigma_\mu^2 + \frac{\sigma_u^2}{(1-\rho_\lambda)^2}\right)}\mathbf{h}'\overline{\mathbf{J}}_N\mathbf{h}
\end{aligned}$$

Proof. To obtain the limit results for $\frac{1}{NT}\mathbf{X}'\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}$ we write

$$\begin{aligned}
\frac{1}{NT}\mathbf{X}'\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon} &= \frac{(1-\rho_v)}{NT\sigma_\alpha^2}\sum_{t=1}^{T}\sum_{i=1}^{N}X_{i,t}r_t\mu_i \qquad \text{(B.1)} \\
&\quad -\frac{(1-\rho_v)}{N^2T\sigma_\alpha^2}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{i=1}^{N}X_{i,t}r_t\mu_j + \frac{1}{N^2T}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{i=1}^{N}X_{i,t}c_t\mu_j \\
&\quad +\frac{1}{NT}\sum_{r=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{N}X_{i,t}A_{t,r}^*v_{i,r} \\
&\quad -\frac{1}{N^2T}\sum_{r=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{i=1}^{N}X_{i,t}A_{t,r}^*v_{j,r} \\
&\quad +\frac{1}{N^2T}\sum_{r=1}^{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{i=1}^{N}X_{i,t}L^{t,r}v_{j,r} \\
&\quad +\frac{1}{NT}\sum_{r=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{N}X_{i,t}L^{t,r}\lambda_r
\end{aligned}$$

where r_t denotes the $t:th$ element of $\mathbf{C}'\boldsymbol{\iota}_T^\alpha$ and c_t denotes an element of the $T\times 1$ vector $\left(N\sigma_u^2\mathbf{\Psi}_\lambda + (\mathbf{A}^*)^{-1}\right)^{-1}\boldsymbol{\iota}_T$. Further $L^{t,r}$ denotes the $tr:th$ element of the $T\times T$ matrix $\mathbf{L}^{-1} = \left(N\sigma_u^2\mathbf{\Psi}_\lambda + (\mathbf{A}^*)^{-1}\right)^{-1}$ and $A_{t,r}^*$ denotes the $tr:th$ element of the $T\times T$ matrix $\mathbf{A}^*$.

First we consider the probability limits of the terms involving μ_i

$$\begin{aligned}
\operatorname{plim}\frac{(1-\rho_v)}{NT\sigma_\alpha^2}\sum_{t=1}^{T}\sum_{i=1}^{N}X_{i,t}r_t\mu_i &= 0 \\
\operatorname{plim}\frac{(1-\rho_v)}{N^2T\sigma_\alpha^2}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{i=1}^{N}X_{i,t}r_t\mu_j &= 0
\end{aligned}$$

as $N \to \infty, T \to \infty$ or $N, T \to \infty$ are straightforward to show since $\mathbf{C}'\boldsymbol{\iota}_T^\alpha$ is a constant vector. To establish corresponding results for

$$\operatorname{plim} \frac{1}{N^2T} \sum_{t=1}^{T} \sum_{j=1}^{N} \sum_{i=1}^{N} X_{i,t} c_t \mu_j \tag{B.2}$$

we need to consider the properties of c_t. First, since c_t is the $t : th$ element of $\mathbf{L}\boldsymbol{\iota}_T = \sigma_\alpha^{-2}(1-\rho_v)\mathbf{B}^{-1}\mathbf{C}'\boldsymbol{\iota}_T^\alpha$ and $\mathbf{B}^{-1}$ converges elementwise to a matrix with exponentially decaying off-diagonals, every element of $\mathbf{L}\boldsymbol{\iota}_T$ is an exponentially decaying sum. Secondly, by the properties of $\mathbf{B}^{-1}$ every element of $\mathbf{L}\boldsymbol{\iota}_T$ is $O\left((NT)^{-1}\right)$. This shows that (B.2) is zero as either or both of the indices grows large.

For the elements involving the idiosyncratic errors, v_{it}

$$\begin{aligned}
\operatorname{plim} \frac{1}{NT} \sum_{r=1}^{T} \sum_{t=1}^{T} \sum_{i=1}^{N} X_{i,t} A^*_{t,r} v_{i,r} &= 0 \\
\operatorname{plim} \frac{1}{N^2T} \sum_{r=1}^{T} \sum_{t=1}^{T} \sum_{j=1}^{N} \sum_{i=1}^{N} X_{i,t} A^*_{t,r} v_{j,r} &= 0
\end{aligned}$$

as $N \to \infty, T \to \infty$ or $N, T \to \infty$ holds since $\mathbf{A}^* = \mathbf{C}'\left(\sigma_\alpha^{-2}\overline{\mathbf{J}}_T^\alpha + \sigma_e^{-2}\overline{\mathbf{E}}_T^\alpha\right)\mathbf{C}$, $\mathbf{C}'\mathbf{C}$ is band-diagonal and $\mathbf{C}'\boldsymbol{\iota}_T^\alpha$ is a constant vector. To be able to write

$$\operatorname{plim} \frac{1}{N^2T} \sum_{r=1}^{T} \sum_{t=1}^{T} \sum_{j=1}^{N} \sum_{i=1}^{N} X_{i,t} L^{t,r} v_{j,r} = 0 \tag{B.3}$$

as $N \to \infty, T \to \infty$ or $N, T \to \infty$ we need to establish some properties of $\mathbf{L}^{-1}$. For this purpose we let $\mathbf{C}\boldsymbol{\Psi}_v\mathbf{C}' = \mathbf{I}_T$, $\mathbf{Q}$ be the eigenvectors of $\boldsymbol{\Psi}_\lambda$ in the metric of $\boldsymbol{\Psi}_v$. That is, $\mathbf{C}\boldsymbol{\Psi}_v\mathbf{C}' = \mathbf{Q}\boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda}$ is diagonal and $\mathbf{Q}$ is orthogonal. Further let $\mathbf{W} = \mathbf{C}'\mathbf{Q}$ we can then write

$$\begin{aligned}
\boldsymbol{\Psi}_v &= \mathbf{W}^{-1\prime}\mathbf{W}^{-1} = \mathbf{C}^{-1}\mathbf{Q}\mathbf{Q}'\mathbf{C}^{-1\prime} \\
\boldsymbol{\Psi}_\lambda &= \mathbf{C}^{-1}\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}'\mathbf{C}^{-1\prime} = \mathbf{W}^{-1\prime}\boldsymbol{\Lambda}\mathbf{W}^{-1}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{L} &= \mathbf{W}^{-1\prime}\mathbf{W}'\left(N\sigma_u^2\boldsymbol{\Psi}_\lambda + \sigma_\mu^2\mathbf{J}_T + \sigma_e^2\boldsymbol{\Psi}_v\right)\mathbf{W}\mathbf{W}^{-1} \\
&= \mathbf{W}^{-1\prime}\left(\mathbf{D} + \left(\sigma_\alpha^2 - \sigma_e^2\right)\overline{\mathbf{J}}_T^w\right)\mathbf{W}^{-1}
\end{aligned}$$

where $\overline{\mathbf{J}}_T^w = \mathbf{Q}'\overline{\mathbf{J}}_T^\alpha\mathbf{Q}$ is idempotent, $\mathbf{D} = N\sigma_u^2\mathbf{\Lambda} + \sigma_e^2\mathbf{I}_T$ is diagonal. Since $\mathbf{\Lambda}$ is diagonal with bounded constant elements setting $\mathbf{\Lambda} = \varphi\mathbf{I}_T$ will not change the order properties of $\mathbf{L}^{-1}$. Hence, defining $\overline{\mathbf{E}}_T^w = \mathbf{I}_T - \overline{\mathbf{J}}_T^w$ we obtain

$$\mathbf{L}^{-1} \approx \mathbf{W}'\left(\frac{1}{(N\sigma_u^2\varphi + \sigma_e^2)}\overline{\mathbf{E}}_T^w + \frac{1}{(N\sigma_u^2\varphi + \sigma_\alpha^2)}\overline{\mathbf{J}}_T^w\right)\mathbf{W}$$

which shows that $\mathbf{L}^{-1}$ is similar to $\mathbf{A}^*$ except that the elements of $\mathbf{L}^{-1}$ are $O\left(N^{-1}\right)$. This shows that (B.3) holds.

Finally for the term involving λ_t

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{r=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{N}X_{i,t}L^{t,r}\lambda_r = 0$$

follows since the elements of $\mathbf{L}^{-1}$ are $O\left(N^{-1}\right)$. Next

$$\operatorname*{plim}_{T\to\infty}\frac{1}{NT}\sum_{r=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{N}X_{i,t}L^{t,r}\lambda_r = 0$$

since λ_t have zero mean and by the properties of $\mathbf{L}^{-1}$. It follows that the probability limit is zero as both N and $T \to \infty$ as well. This completes the proof of the first result in the lemma. We consider next the limits of the terms involving the constant.

To obtain results for $\boldsymbol{\iota}_{NT}'\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon}$ we let $X_{it} = 1\ \forall i, t$ in (B.1). This gives

$$\boldsymbol{\iota}_{NT}'\mathbf{\Sigma}^{-1}\boldsymbol{\varepsilon} = (1-\rho_v)\sum_{t=1}^{T}\sum_{i=1}^{N}c_t\mu_i + \sum_{t=1}^{T}\sum_{i=1}^{N}c_t v_{i,t} + N\sum_{t=1}^{T}c_t\lambda_t$$

If $\frac{N}{T} \to \infty$ we normalize by $\frac{1}{T}$ to obtain

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty}\frac{(1-\rho_v)}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}c_t\mu_i &= 0\\
\operatorname*{plim}_{N\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}c_t v_{i,t} &= 0\\
\operatorname*{plim}_{T\to\infty}\operatorname*{plim}_{N\to\infty}\frac{N}{T}\sum_{t=1}^{T}c_t\lambda_t &= 0
\end{aligned}$$

as a consequence of the properties of c_t. To obtain corresponding results for $\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT}$ write

$$\begin{aligned}
\frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{A}^*\mathbf{X}_i - \frac{1}{N^2T}\sum_{j=1}^{N}\sum_{i=1}^{N}\mathbf{X}_j'\mathbf{A}^*\mathbf{X}_i \\
&+\frac{1}{N^2T}\sum_{j=1}^{N}\sum_{i=1}^{N}\mathbf{X}_j'\mathbf{L}^{-1}\mathbf{X}_i
\end{aligned}$$

and note that if $\mathbf{X}_i = \boldsymbol{\iota}_T\ \forall i$ we arrive at

$$\begin{aligned}
\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} &= N\boldsymbol{\iota}'_T\mathbf{L}^{-1}\boldsymbol{\iota}_T = N\boldsymbol{\iota}'_T\left(N\sigma_u^2\boldsymbol{\Psi}_\lambda + (\mathbf{A}^*)^{-1}\right)^{-1}\boldsymbol{\iota}_T \\
&= \sigma_\alpha^{-2}(1-\rho_v)N\boldsymbol{\iota}'_T\left(N\sigma_u^2\boldsymbol{\Psi}_\lambda\mathbf{A}^* + \mathbf{I}_T\right)^{-1}\mathbf{C}'\boldsymbol{\iota}_T^\alpha \\
&= \sigma_\alpha^{-2}\sigma_u^{-2}(1-\rho_v)\boldsymbol{\iota}'_T(\boldsymbol{\Psi}_\lambda\mathbf{A}^*)^{-1}\mathbf{C}'\boldsymbol{\iota}_T^\alpha + O\left(N^{-1}\right) \\
&= \sigma_\alpha^{-2}\sigma_u^{-2}(1-\rho_v)\boldsymbol{\iota}'_T(\mathbf{A}^*)^{-1}\boldsymbol{\Psi}_\lambda^{-1}\mathbf{C}'\boldsymbol{\iota}_T^\alpha + O\left(N^{-1}\right)
\end{aligned}$$

since $\eta_t = \boldsymbol{\iota}'_T\mathbf{C}^{-1}\boldsymbol{\iota}_T^\alpha = \frac{\alpha}{\sqrt{1-\rho_v^2}}\sum_{j=0}^{T}\rho_v^j + \sum_{j=0}^{T}(T-j)\rho_v^j = O(T)$

$$\begin{aligned}
&\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} \\
&= \frac{(1-\rho_v)}{\sigma_\alpha^2\sigma_u^2}\boldsymbol{\iota}'_T\mathbf{C}^{-1}\left(\frac{\sigma_\alpha^2-\sigma_e^2}{d^2}\boldsymbol{\iota}_T^\alpha\boldsymbol{\iota}_T^{\alpha\prime} + \sigma_e^2\mathbf{I}_T\right)\mathbf{C}^{-1\prime}\boldsymbol{\Psi}_\lambda^{-1}\mathbf{C}'\boldsymbol{\iota}_T^\alpha \\
&\quad +O\left(N^{-1}\right) \\
&= \frac{(1-\rho_v)}{\sigma_\alpha^2\sigma_u^2}\left(\sigma_\mu^2(1-\rho_v)^2\eta_t\boldsymbol{\iota}_T^{\alpha\prime} + \sigma_e^2\boldsymbol{\iota}'_T\mathbf{C}^{-1}\right)\mathbf{C}^{-1\prime}\boldsymbol{\Psi}_\lambda^{-1}\mathbf{C}'\boldsymbol{\iota}_T^\alpha \\
&\quad +O\left(N^{-1}\right)
\end{aligned}$$

hence if both $N, T \to \infty$ and $\frac{N}{T} \to \infty$

$$\begin{aligned}
\lim_{T\to\infty}\lim_{N\to\infty}\frac{1}{T}\boldsymbol{\iota}'_{NT}\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} &= \lim_{T\to\infty}\frac{\eta_t\sigma_\mu^2}{T\sigma_\alpha^2\sigma_u^2}(1-\rho_v)^3\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C}^{-1\prime}\boldsymbol{\Psi}_\lambda^{-1}\mathbf{C}'\boldsymbol{\iota}_T^\alpha \\
&= \lim_{T\to\infty}\frac{\eta_t\sigma_\mu^2}{T\sigma_\alpha^2\sigma_u^2}(1-\rho_v)^2\boldsymbol{\iota}'_T\mathbf{C}'\mathbf{C}^{-1\prime}\boldsymbol{\Psi}_\lambda^{-1}\mathbf{C}'\boldsymbol{\iota}_T^\alpha \\
&= \lim_{T\to\infty}\frac{\eta_t}{Td^2\sigma_u^2}(1-\rho_\lambda)\boldsymbol{\iota}_T^{\beta\prime}\mathbf{C}_\lambda\mathbf{C}'\boldsymbol{\iota}_T^\alpha \\
&= \frac{1}{\sigma_u^2}(1-\rho_\lambda)^2
\end{aligned}$$

where $\boldsymbol{\iota}_T^{\beta\prime} = \boldsymbol{\iota}'_T\mathbf{C}_\lambda$ and $\mathbf{C}_\lambda$ is the Prais-Winsten transformation matrix for $\boldsymbol{\Psi}_\lambda$. Alternatively this can be derived by noting that $\lim_{N\to\infty}\operatorname{vech}\left(\frac{1}{N}\mathbf{L}\right) =$

vech $\left(\sigma_u^2 \boldsymbol{\Psi}_\lambda\right)$ and hence $\lim_{N\to\infty} \frac{N}{T}\boldsymbol{\iota}_T' \mathbf{L}^{-1}\boldsymbol{\iota}_T = \frac{(1-\rho_\lambda)^2}{T\sigma_u^2}\boldsymbol{\iota}_T^{\beta\prime}\boldsymbol{\iota}_T^{\beta} \to \frac{(1-\rho_\lambda)^2}{\sigma_u^2}$ as $T \to \infty$. If $\frac{T}{N} \to \infty$ we have

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}}\ \underset{T\to\infty}{\text{plim}} \frac{(1-\rho_v)}{N}\sum_{t=1}^{T}\sum_{i=1}^{N} c_t\mu_i &= 0 \\
\underset{T\to\infty}{\text{plim}} \frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N} c_t v_{it} &= 0 \\
\underset{T\to\infty}{\text{plim}} \sum_{t=1}^{T} c_t\lambda_t &= 0
\end{aligned}$$

For $\boldsymbol{\iota}_{NT}' \boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT}$ we arrive at

$$\lim_{T\to\infty} \boldsymbol{\iota}_{NT}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} = \lim_{T\to\infty} N\boldsymbol{\iota}_T'\left(N\sigma_u^2\boldsymbol{\Psi}_\lambda + \sigma_\mu^2\mathbf{J}_T + \sigma_e^2\boldsymbol{\Psi}_v\right)^{-1}\boldsymbol{\iota}_T$$

and hence proceeding by induction

$$\lim_{N\to\infty}\lim_{T\to\infty}\frac{1}{N}\boldsymbol{\iota}_{NT}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} = \frac{1}{\sigma_\mu^2}$$

Finally if $N, T \to \infty$ simultaneously we obtain

$$\begin{aligned}
\underset{N,T\to\infty}{\text{plim}} \frac{1}{N}\boldsymbol{\iota}_{NT}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} &= \underset{N,T\to\infty}{\text{plim}} \frac{1}{T}\boldsymbol{\iota}_{NT}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} = 0 \\
\lim_{N,T\to\infty} \frac{1}{N}\boldsymbol{\iota}_{NT}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} &= \lim_{N,T\to\infty} \frac{1}{T}\boldsymbol{\iota}_{NT}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}_{NT} = \frac{1}{\sigma_\mu^2 + \sigma_u^2\left(1-\rho_\lambda\right)^{-2}}
\end{aligned}$$

This completes the proof for the terms involving the constant and we proceed to consider the limit results for the terms involving time-invariant explanatory variables or individual-invariant explanatory variables.

To prove that

$$\text{plim}\,\frac{1}{T}\mathbf{D}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} = \text{plim}\,\frac{1}{T}\sum_{i=1}^{N}\mathbf{d}'\mathbf{L}^{-1}\boldsymbol{\iota}_T\mu_i + \text{plim}\,\frac{1}{T}\sum_{i=1}^{N}\mathbf{d}'\mathbf{L}^{-1}\mathbf{v}_i + \text{plim}\,\frac{N}{T}\mathbf{d}'\mathbf{L}^{-1}\boldsymbol{\lambda}$$

is a null vector as either $T \to \infty$ or $N, T \to \infty$ it suffices to note the properties of $\mathbf{L}^{-1}$. By the properties of $\mathbf{L}^{-1}$ we similarly have $\text{plim}_{N\to\infty}\frac{N}{T}\mathbf{d}'\mathbf{L}^{-1}\boldsymbol{\lambda} \neq \mathbf{0}$ and $\text{plim}_{N\to\infty}\frac{1}{T}\mathbf{d}'\mathbf{L}^{-1}\boldsymbol{\lambda} = \mathbf{0}$. Results for $\text{plim}\,\frac{1}{N}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}$ and $\text{plim}\,\frac{1}{NT}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}$ can be shown analogously. Remaining results can be derived by noting that $\frac{1}{T}\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D} = \frac{N}{T}\mathbf{d}'\mathbf{L}^{-1}\mathbf{d}$, and $\frac{1}{N}\mathbf{H}'\boldsymbol{\Sigma}^{-1}\mathbf{H} = \frac{d^2(1-\rho_v)^2}{N\sigma_\alpha^2}\mathbf{h}'\overline{\mathbf{E}}_N\mathbf{h} + \frac{\boldsymbol{\iota}_\mathbf{T}'\mathbf{L}^{-1}\boldsymbol{\iota}_\mathbf{T}}{N}\mathbf{h}'\overline{\mathbf{J}}_N\mathbf{h}$ ■

The next and final lemma gives some important results about the limit behavior of the information cross-elements for the mean parameters, $\boldsymbol{\delta}$ and the elements $\mathcal{I}_{\delta,\gamma}$. Limit results for the elements $\mathcal{I}_{\gamma,\gamma}$ appear in theorem 2. To summarize some of the content in this lemma we can say that the set of consistent parameters (as $N \to \infty$ or $T \to \infty$) are information block-diagonal to the set of inconsistent parameters and that the set of consistent mean parameters and the set of consistent variance parameters are always information block-diagonal.

Lemma 7 *As either or both of N and $T \to \infty$ the cross-elements (properly normalized of course) $\mathcal{I}_{\beta,(\gamma,\alpha)}$, $\mathcal{I}_{\pi,(\gamma_1,\gamma_2)}$, $\mathcal{I}_{\tau,(\gamma_2,\gamma_3)}$ and $\mathcal{I}_{\alpha,\gamma_2}$, converge elementwise to zero in probability (or in expectation), where $\gamma_1 = \sigma_\mu^2$, $\gamma_2 = (\sigma_e^2, \rho_v)$ and $\gamma_3 = (\sigma_u^2, \rho_\lambda)$. As $N \to \infty$ (no matter what T is) this holds for $\mathcal{I}_{\gamma_1,(\alpha,\tau,\pi,\gamma_2)}$, $\mathcal{I}_{\gamma_2,\gamma_3}$, $\mathcal{I}_{\beta,\pi}$ and as $T \to \infty$ (no matter what N is) for $\mathcal{I}_{\gamma_2,(\alpha,\pi,\tau,\gamma_1)}$, $\mathcal{I}_{\gamma_2,\gamma_3}$, and $\mathcal{I}_{\beta,\tau}$. We now concentrate on mainly the non-zero cross elements of interest. If only $N \to \infty$*

$$\operatorname{plim} \frac{1}{\sqrt{T}} \mathcal{I}_{\alpha,\pi} = \frac{(1-\rho_\lambda)}{\sigma_u^2 \sqrt{T}} \mathbf{d}' \mathbf{C}_\lambda' \boldsymbol{\iota}_T^\beta$$

where $\mathbf{C}_\lambda' \mathbf{C}_\lambda = \boldsymbol{\Psi}_\lambda^{-1}, (1-\rho_\lambda) \boldsymbol{\iota}_T^\beta = \mathbf{C}_\lambda \boldsymbol{\iota}_T$, hence if $N, T \to \infty$ such that $\frac{N}{T} \to \infty$

$$\operatorname{plim} \frac{1}{T} \mathcal{I}_{\alpha,\pi} = \frac{(1-\rho_\lambda)^2}{\sigma_u^2} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T-1} \mathbf{d}_t'$$

and if $T \to \infty$ or $N, T \to \infty$ such that $\frac{T}{N} \to \infty$

$$\operatorname{plim} \frac{1}{\sqrt{T}} \mathcal{I}_{\alpha,\pi} = \mathbf{0}$$

If $N, T \to \infty$ simultaneously

$$\operatorname{plim} \frac{1}{T} \mathcal{I}_{\alpha,\pi} = \frac{1}{\sigma_u^2 (1-\rho_\lambda)^{-2} + \sigma_\mu^2} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T-1} \mathbf{d}_t'$$

If only $T \to \infty$

$$\operatorname{plim} \frac{1}{\sqrt{N}} \mathcal{I}_{\alpha,\tau} = \frac{1}{\sqrt{N} \sigma_\mu^2} \sum_{i=1}^{N} \mathbf{h}_i'$$

hence if $N, T \to \infty$ such that $\frac{T}{N} \to \infty$

$$\operatorname{plim} \frac{1}{N} \mathcal{I}_{\alpha,\tau} = \frac{1}{\sigma_\mu^2} \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{h}_i'$$

and if $N \to \infty$ *or* $N, T \to \infty$ *such that* $\frac{N}{T} \to \infty$

$$\operatorname{plim} \frac{1}{\sqrt{N}} \mathcal{I}_{\alpha,\tau} = \mathbf{0}$$

If $N, T \to \infty$ *simultaneously*

$$\begin{aligned}
\operatorname{plim} \frac{1}{N} \mathcal{I}_{\alpha,\tau} &= \frac{1}{\sigma_u^2 (1-\rho_\lambda)^{-2} + \sigma_\mu^2} \operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{h}_i' \\
\operatorname{plim} \frac{1}{\sqrt{NT}} \mathcal{I}_{\pi,\tau} &= \frac{1}{\sigma_u^2 (1-\rho_\lambda)^{-2} + \sigma_\mu^2} \operatorname{plim} \frac{1}{NT} \sum_{t=2}^{T-1} \mathbf{d}_t \sum_{i=1}^{N} \mathbf{h}_i'
\end{aligned}$$

and otherwise for the last term

$$\operatorname{plim} \frac{1}{\sqrt{NT}} \mathcal{I}_{\pi,\tau} = \mathbf{0}$$

Proof. These results can be proved with exactly the same methods as in lemma 6. In fact the same matrices are involved in the expressions and the proof is therefore omitted ■

Next we give the proofs of the theorems in the text.

Proof theorem 1. The method of proof is to examine the probability limit of the standardized log-likelihood. It is however not useful for dealing with the constant. In fact, the constant drops out of the analysis. The reason for adopting this method is that we can (in most cases) prove global consistency results for the other parameters which are not easily obtained otherwise. Asymptotic properties of the constant term are established separately at the end of the proof.

The negative of the log likelihood is up to an irrelevant term given by

$$\begin{aligned}
\phi(\boldsymbol{\delta}, \gamma) &= \frac{N}{2} \ln |\mathbf{A}^*| + \frac{1}{2} \ln \left| \mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Psi}_\lambda \mathbf{A}^* \right| \\
&\quad + \frac{1}{2} (\mathbf{y} - \mathbf{Z}\boldsymbol{\delta})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{Z}\boldsymbol{\delta}) \\
&= \frac{N}{2} \ln |\mathbf{A}^*| + \frac{1}{2} \ln \left| \mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Psi}_\lambda \mathbf{A}^* \right| \\
&\quad + \frac{1}{2} (\boldsymbol{\delta}_0 - \boldsymbol{\delta})' \mathbf{Z}' \boldsymbol{\Sigma}^{-1} \mathbf{Z} (\boldsymbol{\delta}_0 - \boldsymbol{\delta}) \\
&\quad + \frac{1}{2} \boldsymbol{\varepsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} + (\boldsymbol{\delta}_0 - \boldsymbol{\delta})' \mathbf{Z}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}
\end{aligned} \tag{B.4}$$

By theorem 4.1.1 of Amemiya (1985) we need to verify that (i) the parameter space Θ is a compact subset of the Euclidean K-space, (ii) $\phi(\boldsymbol{\delta}, \gamma)$ is continuous in $\theta \in \Theta$ for all $(\mathbf{y}, \mathbf{X})$ and is a measurable function of $(\mathbf{y}, \mathbf{X})$ for all $\theta \in \Theta$, (iii) $W^{-1}\phi(\boldsymbol{\delta}, \gamma)$ converges to a nonstochastic function, say ϕ_0, in probability uniformly in $\theta \in \Theta$ as $W \to \infty$ and ϕ_0 is uniquely minimized at θ_0. Since (i) follows from assumption (b) and (ii) is trivial it remains to show (iii). This involves finding the limit of $W^{-1}\phi(\boldsymbol{\delta}, \gamma)$ as $W \to \infty$ with $W = N$, $W = T$ and $W = NT$ respectively.

First we consider the uniform probability limit of (B.4) as $N, T \to \infty$. Note that

$$E\varepsilon'\boldsymbol{\Sigma}^{-1}\varepsilon = \operatorname{tr} \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0$$

where $\boldsymbol{\Sigma}_0$ denotes $\boldsymbol{\Sigma}$ evaluated at θ_0. Hence using lemma 5, lemma 6 and 7 and assumption (c)

$$\begin{aligned}
\operatorname*{plim}_{N,T\to\infty} \frac{1}{NT}\phi(\boldsymbol{\delta}, \gamma) &= -\lim_{T\to\infty} \frac{1}{2T} \ln |\mathbf{C}|^2 + \lim_{T\to\infty} \frac{1}{2T} \ln \left(d^2\sigma_\mu^2 (1-\rho_v)^2 + \sigma_e^2\right) \\
&\quad + \lim_{N,T\to\infty} \frac{N(T-1)}{2NT} \ln \sigma_e^2 \\
&\quad + \lim_{N,T\to\infty} \frac{1}{2NT} \ln \left|\mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Psi}_\lambda \mathbf{A}^*\right| \\
&\quad + \frac{1}{2}(\boldsymbol{\beta}_0 - \boldsymbol{\beta})' \mathbf{R}_{\mathbf{x}} (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \frac{\sigma_{e0}^2}{2\sigma_e^2}\left[\frac{(\rho_v^2+1) - 2\rho_{v0}\rho_v}{1-\rho_{v0}^2}\right]
\end{aligned}$$

with $\mathbf{R}_{\mathbf{x}} = \operatorname{plim}_{N,T\to\infty} \frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$, since $|\mathbf{C}| = O(1)$,

$$\lim_{N,T\to\infty} \frac{1}{2NT} \ln \left|\mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Psi}_\lambda \mathbf{A}^*\right| = 0$$

and using lemma 4 we arrive at

$$\frac{1}{2}\ln\sigma_e^2 + \frac{1}{2}(\boldsymbol{\beta}_0 - \boldsymbol{\beta})' \mathbf{R}_{\mathbf{x}} (\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \frac{\sigma_{e0}^2}{2\sigma_e^2}\left[\frac{(\rho_v^2+1) - 2\rho_{v0}\rho_v}{1-\rho_{v0}^2}\right] \qquad \text{(B.5)}$$

and it is straightforward to verify that (B.5) is uniquely minimized at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\sigma_e^2 = \sigma_{e0}^2$ and $\mathbf{C} = \mathbf{C}_0$. Having established the consistency of maximum likelihood estimators $\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_e^2, \widehat{\rho}_v$ as $N, T \to \infty$ we obtain the uniform probability limit of (B.4) as $N \to \infty$ with $T \geq 2$ a fix constant. For this purpose let

$\boldsymbol{\zeta} = \left(\boldsymbol{\beta}', \boldsymbol{\tau}'\right)'$

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}} \frac{1}{NT}\phi(\boldsymbol{\delta},\boldsymbol{\gamma}) &= \lim_{N\to\infty} \frac{1}{2T} \ln\left|\mathbf{A}^*\right| + \lim_{N\to\infty} \frac{1}{2NT} \ln\left|\mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Psi}_\lambda \mathbf{A}^*\right| \\
&\quad + \underset{N\to\infty}{\text{plim}} \frac{1}{2NT} \left(\boldsymbol{\zeta}_0 - \boldsymbol{\zeta}\right)' \mathbf{Z}_N' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_N \left(\boldsymbol{\zeta}_0 - \boldsymbol{\zeta}\right) \\
&\quad + \lim_{N\to\infty} \frac{1}{2NT} E\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} \\
&\quad + \underset{N\to\infty}{\text{plim}} \frac{1}{NT} \left(\boldsymbol{\zeta}_0 - \boldsymbol{\zeta}\right)' \mathbf{Z}_N' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \\
&= -\frac{1}{2T} \ln\left(1-\rho_v^2\right) + \frac{1}{2T} \ln\left(d^2\sigma_\mu^2 \left(1-\rho_v\right)^2 + \sigma_e^2\right) \\
&\quad + \left(\frac{1}{2} - \frac{1}{2T}\right) \ln \sigma_e^2 \\
&\quad + \frac{\sigma_{\mu 0}^2 d^2 \left(1-\rho_v\right)^2}{2T\left(d^2\sigma_\mu^2\left(1-\rho_v\right)^2 + \sigma_e^2\right)} + \frac{1}{2T}\sigma_{e0}^2 \operatorname{tr}\left(\mathbf{A}^* \boldsymbol{\Psi}_{v0}\right)
\end{aligned}$$

where we have used lemma 5 to evaluate $\lim_{N\to\infty} \frac{1}{2NT} E\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}$, lemma 4, 6 and lemma 7 and that $\boldsymbol{\zeta}$ is uniquely identified with $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0$. We then have

$$\begin{aligned}
\underset{N\to\infty}{\text{plim}} \frac{1}{NT}\phi(\boldsymbol{\delta},\boldsymbol{\gamma}) &= -\frac{1}{2T} \ln\left(1-\rho_v^2\right) + \frac{1}{2}\ln\sigma_e^2 \\
&\quad + \frac{1}{2}\sigma_{e0}^2\sigma_e^{-2}\frac{1}{T}\operatorname{tr}\left(\mathbf{C}\boldsymbol{\Psi}_{v0}\mathbf{C}'\right) - \frac{1}{2T}\ln\sigma_e^2 \\
&\quad - \frac{1}{2}\sigma_{e0}^2\sigma_e^{-2}\frac{d^{-2}}{T}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C}\boldsymbol{\Psi}_{v0}\mathbf{C}'\boldsymbol{\iota}_T^{\alpha} \\
&\quad + \frac{1}{2T}\ln\left(d^2\sigma_\mu^2\left(1-\rho_v\right)^2 + \sigma_e^2\right) \\
&\quad + \frac{1}{2T}\frac{\sigma_{\mu 0}^2 d^2\left(1-\rho_v\right)^2 + \sigma_{e0}^2 d^{-2}\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C}\boldsymbol{\Psi}_{v0}\mathbf{C}'\boldsymbol{\iota}_T^{\alpha}}{\left(d^2\sigma_\mu^2\left(1-\rho_v\right)^2 + \sigma_e^2\right)}
\end{aligned} \tag{B.6}$$

Evaluating $\operatorname{tr}\left(\mathbf{C}\boldsymbol{\Psi}_{v0}\mathbf{C}'\right)$ and $\boldsymbol{\iota}_T^{\alpha\prime}\mathbf{C}\boldsymbol{\Psi}_{v0}\mathbf{C}'\boldsymbol{\iota}_T^{\alpha}$ as in lemma 3 we can show that (B.6) is uniquely minimized at $\rho_v = \rho_{v0}, \sigma_e^2 = \sigma_{e0}^2$ and $\sigma_\mu^2 = \sigma_{\mu 0}^2$. This establishes the consistency of $\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_e^2, \widehat{\rho}_v$ as $N \to \infty$ as well as the consistency of $\widehat{\sigma}_\mu^2, \widehat{\boldsymbol{\tau}}$ as $N \to \infty$ or $N, T \to \infty$.

Consider next the uniform probability limit of (B.4) as $N \geq 2$ is fix and $T \to \infty$. Noting that lemma 6 and 7 and assumption (c) ensures that $\boldsymbol{\psi} =$

$(\beta', \pi')'$ is uniquely identified we have

$$\begin{aligned}\operatorname*{plim}_{T\to\infty} \frac{1}{NT}\phi(\boldsymbol{\delta},\boldsymbol{\gamma}) &= \lim_{T\to\infty}\frac{1}{2T}\ln|\mathbf{A}^*| + \lim_{T\to\infty}\frac{1}{2NT}\ln\left|\mathbf{I}_T + N\sigma_u^2\boldsymbol{\Psi}_\lambda\mathbf{A}^*\right| \\ &\quad + \lim_{T\to\infty}\frac{1}{2NT}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\end{aligned}$$

using lemma 5 and after some matrix manipulation we arrive at

$$\begin{aligned}\operatorname*{plim}_{T\to\infty} \frac{1}{NT}\phi(\boldsymbol{\delta},\boldsymbol{\gamma}) &= \left(\frac{1}{2}-\frac{1}{2N}\right)\ln\sigma_e^2 \\ &\quad + \left(\frac{1}{2}-\frac{1}{2N}\right)\frac{\sigma_{e0}^2}{\sigma_e^2}\left[\frac{\left(\rho_v^2+1\right)-2\rho_{v0}\rho_v}{1-\rho_{v0}^2}\right] \\ &\quad + \frac{1}{2N}\lim_{T\to\infty}\frac{1}{T}\ln\left|N\sigma_u^2\boldsymbol{\Psi}_\lambda+\sigma_e^2\boldsymbol{\Psi}_v\right| \\ &\quad + \frac{1}{2N}\lim_{T\to\infty}\frac{1}{T}\operatorname{tr}\left(\mathbf{P}_0\mathbf{P}^{-1}\right)\end{aligned} \tag{B.7}$$

where $\mathbf{P}_0$ and $\mathbf{P}$ are given in lemma 5. The first and second row of (B.7) are uniquely minimized at $\sigma_e^2 = \sigma_{e0}^2$, $\rho_v = \rho_{v0}$. However we cannot evaluate the last two rows analytically which complicates showing uniqueness globally[1]. We can prove the existence of a consistent root though (cf. Amemiya (1985, theorem 4.1.2)). Applying matrix differentiation to (B.7) using standard results for interchanging the limit and the derivative e.g. Rudin (1976, p 152) it is straightforward to show that the true parameters are a solution to the first order condition. Of course then we also need to verify that the second derivative matrix is positive-definite when evaluated at the true parameters. But this is straightforward to do as well. This proves the global consistency of $\widehat{\beta}, \widehat{\pi}$ as $T \to \infty$ (and also the global consistency of $\widehat{\pi}$ as $N, T \to \infty$) and the existence of a local consistent root for $\widehat{\sigma}_e^2, \widehat{\rho}_v, \widehat{\sigma}_u^2, \widehat{\rho}_\lambda$ as $T \to \infty$. Since the information matrix is positive definite over the full parameter space when $N, T \to \infty$ (as shown in theorem 2) this also proves the global consistency of $\widehat{\sigma}_u^2, \widehat{\rho}_\lambda$ as $N, T \to \infty$.

[1] In case of $\rho_v = \rho_\lambda = 0$ (B.7) reduces to

$$\begin{aligned}&\left(\frac{1}{2}-\frac{1}{2N}\right)\ln\sigma_e^2 + \left(\frac{1}{2}-\frac{1}{2N}\right)\frac{\sigma_{e0}^2}{\sigma_e^2} \\ &+\frac{1}{2N}\ln\left(N\sigma_u^2+\sigma_e^2\right) + \frac{1}{2N}\frac{N\sigma_{u0}^2+\sigma_{e0}^2}{N\sigma_u^2+\sigma_e^2}\end{aligned}$$

which is globally minimized at the true parameters if $N \geq 2$.

Finally we obtain results for the constant term. To obtain a local consistency result for $\widehat{\alpha}$ as $N, T \rightarrow \infty$ it suffices to consider lemma 6. In fact $\widehat{\alpha}$ can be shown to be globally consistent as $N, T \rightarrow \infty$ by the results in lemma 6 and lemma 7 and the fact that the information matrix is positive definite over the full parameter space for the remaining parameters. As a special case of lemma 6 we obtain the inconsistency of $\widehat{\alpha}$ as only $N \rightarrow \infty$ or $T \rightarrow \infty$ ∎

Proof theorem 2. We first derive the results when $N, T \rightarrow \infty$ and hence the full parameter vector is consistently estimated. For the purpose of establishing asymptotic normality of $\widehat{\theta}$ it is useful to structure $\boldsymbol{\delta}$ as $\boldsymbol{\delta} = \left(\alpha, \boldsymbol{\pi}', \boldsymbol{\tau}', \boldsymbol{\beta}'\right)'$ and we will do so below. By the mean value theorem for random functions Gourieroux and Monfort (1995, p 400)

$$\frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial \theta}|_{\widehat{\theta}} = \frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial \theta}|_{\theta_0} + \frac{\partial^2 l(\boldsymbol{\delta},\gamma)}{\partial \theta \partial \theta'}|_{\overline{\theta}} \left(\widehat{\theta} - \theta_0\right) \tag{B.8}$$

where $\overline{\theta}$ belongs to the segment $\left(\widehat{\theta}, \theta_0\right)$ with probability 1. Define $\mathbf{F}_{NT}$ as a diagonal matrix with

$$\operatorname{diag} \mathbf{F}_{NT} = \left\{\min\left(\sqrt{N}, \sqrt{T}\right), \mathbf{F}_{\pi}, \mathbf{F}_{\tau}, \mathbf{F}_{\beta}, \sqrt{N}, \sqrt{NT}, \sqrt{NT}, \sqrt{T}, \sqrt{T}\right\}$$

where $\mathbf{F}_{\beta}$ is a vector containing k_1 $\sqrt{NT}$ and $\mathbf{F}_{\pi}, \mathbf{F}_{\tau}$ are vectors containing k_1 $\sqrt{T}$ and k_2 $\sqrt{N}$ respectively. We can then write

$$\mathbf{F}_{NT}\left(\widehat{\theta} - \theta_0\right) = -\left[\mathbf{F}_{NT}^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\gamma)}{\partial\theta\partial\theta'}|_{\overline{\theta}}\right)\mathbf{F}_{NT}^{-1}\right]^{-1}\left[\mathbf{F}_{NT}^{-1}\left(\frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial\theta}|_{\theta_0}\right)\right] \tag{B.9}$$

From theorem 4.1.3 of Amemiya (1985) we need to show that (in addition to local consistency Amemiya (1985, theorem 4.1.2)) (i) $l(\boldsymbol{\delta},\gamma) \in C^2$ in a convex neighborhood of θ_0, (ii) $\left[\mathbf{F}_{NT}^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\gamma)}{\partial\theta\partial\theta'}|_{\overline{\theta}}\right)\mathbf{F}_{NT}^{-1}\right]$ converges to a finite non-singular matrix

$$\mathbf{V}^{-1}(\theta_0) = -\lim_{N,T\rightarrow\infty} E\left[\mathbf{F}_{NT}^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\gamma)}{\partial\theta\partial\theta'}|_{\theta_0}\right)\mathbf{F}_{NT}^{-1}\right]$$

in probability for any sequence $\overline{\theta}$ such that $\operatorname{plim}\overline{\theta} = \theta_0$ and (iii)

$$\mathbf{F}_{NT}^{-1}\left(\frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial\theta}|_{\theta_0}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{V}_1^{-1}(\theta_0)\right)$$

where

$$\mathbf{V}_1^{-1}(\theta_0) = \lim_{N,T\rightarrow\infty} E\left[\mathbf{F}_{NT}^{-1}\left(\frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial\theta}|_{\theta_0}\right)\left(\frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial\theta}|_{\theta_0}\right)'\mathbf{F}_{NT}^{-1}\right]$$

a finite non-singular matrix. Note that (i) is trivially satisfied and by assumption (a) (ii) follows if the convergence is uniform. Further note that

$$E\left[\left(\frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial\theta}|_{\theta_0}\right)\right]=0$$

is straightforward to verify from appendix A.1, and

$$\mathbf{V}_1(\theta_0)=\mathbf{V}(\theta_0)$$

follows from the information matrix equality. To show (ii) we take uniform limits of the appropriately scaled elements of the information matrix obtained from appendix A.2. The limits for the variance parameters are straightforward to derive using lemma 4 and repeatedly using elementary results on inverses involving sums. For the elements $\mathcal{I}_{\sigma_\mu^2,\gamma_j}$ we have

$$\begin{aligned}
\lim_{N,T\to\infty}\frac{1}{N}\mathcal{I}_{\sigma_\mu^2,\sigma_\mu^2} &= \lim_{N,T\to\infty}\frac{1}{2}\left[\frac{d^4}{\sigma_\alpha^4}(1-\rho_v)^4\right]\\
&= \frac{1}{2}\lim_{N,T\to\infty}\frac{\left(\alpha^2+(T-1)\right)^2(1-\rho_v)^4}{\left(\left(\alpha^2+(T-1)\right)\sigma_\mu^2(1-\rho_v)^2+\sigma_e^2\right)^2}\\
&= \frac{1}{2\sigma_\mu^4}
\end{aligned}$$

$$\lim_{N,T\to\infty}\frac{1}{N\sqrt{T}}\mathcal{I}_{\sigma_\mu^2,\sigma_e^2}=0$$

$$\lim_{N,T\to\infty}\frac{1}{\sqrt{NT}}\mathcal{I}_{\sigma_\mu^2,\sigma_u^2}=0$$

$$\lim_{N,T\to\infty}\frac{1}{\sqrt{NT}}\mathcal{I}_{\sigma_\mu^2,\rho_\lambda}=0$$

$$\lim_{N,T\to\infty}\frac{1}{N\sqrt{T}}\mathcal{I}_{\sigma_\mu^2,\rho_v}=0$$

and for the elements $\mathcal{I}_{\sigma_e^2,\gamma_j}$

$$\begin{aligned}
\lim_{N,T\to\infty}\frac{1}{NT}\mathcal{I}_{\sigma_e^2,\sigma_e^2} &= \lim_{N,T\to\infty}\frac{1}{2T}\operatorname{tr}\left(\mathbf{A}^*\boldsymbol{\Psi}_v\right)^2\\
&= \frac{1}{2\sigma_e^4}
\end{aligned}$$

$$\lim_{N,T\to\infty}\frac{1}{\sqrt{NT}}\mathcal{I}_{\sigma_e^2,\sigma_u^2}=0$$

$$\lim_{N,T\to\infty}\frac{1}{\sqrt{NT}}\mathcal{I}_{\sigma_e^2,\rho_\lambda}=0$$

$$\lim_{N,T\to\infty}\frac{1}{NT}\mathcal{I}_{\sigma_e^2,\rho_v}=\lim_{N,T\to\infty}\frac{1}{2\sigma_e^2 T}\operatorname{tr}\left(\boldsymbol{\Psi}_v^{-1}\mathbf{L}_v\right)=0$$

Finally, for the elements involving $\sigma_u^2, \rho_\lambda, \rho_v$

$$\lim_{N,T\to\infty}\frac{1}{T}\mathcal{I}_{\sigma_u^2,\sigma_u^2}=\frac{1}{2\sigma_u^4}$$

$$\begin{aligned}
\lim_{N,T\to\infty}\frac{1}{T}\mathcal{I}_{\sigma_u^2,\rho_\lambda} &= \lim_{N,T\to\infty}\frac{1}{2\sigma_u^2 T}\operatorname{tr}\left(\boldsymbol{\Psi}_\lambda^{-1}\mathbf{L}_\lambda\right)\\
&= \frac{\rho_\lambda}{\sigma_u^2\left(1-\rho_\lambda^2\right)}+\lim_{N,T\to\infty}\frac{1}{2\sigma_u^2 T\left(1-\rho_\lambda^2\right)}\operatorname{tr}\left(\boldsymbol{\Psi}_\lambda^{-1}\mathbf{D}\right)\\
&= \frac{\rho_\lambda}{\sigma_u^2\left(1-\rho_\lambda^2\right)}+\lim_{N,T\to\infty}\frac{\rho_\lambda}{\sigma_u^2 T\left(1-\rho_\lambda^2\right)}(1-T)=0
\end{aligned}$$

$$\lim_{N,T\to\infty}\frac{1}{\sqrt{NT}}\mathcal{I}_{\sigma_u^2,\rho_v}=0$$

and with

$$\begin{aligned}
\lim_{N,T\to\infty}\frac{1}{T}\mathcal{I}_{\rho_\lambda,\rho_\lambda} &= \lim_{N,T\to\infty}\frac{1}{2T}\operatorname{tr}\left(\boldsymbol{\Psi}_\lambda^{-1}\mathbf{L}_\lambda\right)^2\\
&= \lim_{N,T\to\infty}\frac{1}{2T}\operatorname{tr}\left(\boldsymbol{\Psi}_\lambda^{-1}\left(\frac{2\rho_\lambda}{\left(1-\rho_\lambda^2\right)}\boldsymbol{\Psi}_\lambda+\frac{1}{\left(1-\rho_\lambda^2\right)}\mathbf{D}\right)\right)^2\\
&= -\frac{2\rho_\lambda^2}{\left(1-\rho_\lambda^2\right)^2}+\lim_{N,T\to\infty}\frac{1}{2T\left(1-\rho_\lambda^2\right)^2}\operatorname{tr}\left(\boldsymbol{\Psi}_\lambda^{-1}\mathbf{D}\right)^2\\
&= \frac{1}{\left(1-\rho_\lambda^2\right)}
\end{aligned}$$

$$\lim_{N,T\to\infty}\frac{1}{\sqrt{NT}}\mathcal{I}_{\rho_\lambda,\rho_v}=0$$

$$\begin{aligned}
\lim_{N,T\to\infty}\frac{1}{NT}\mathcal{I}_{\rho_v,\rho_v} &= \lim_{N,T\to\infty}\frac{\sigma_e^4}{2T}\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_v\right)^2=\lim_{N,T\to\infty}\frac{1}{2T}\operatorname{tr}\left(\mathbf{C}'\mathbf{C}\mathbf{L}_v\right)^2\\
&= \lim_{N,T\to\infty}\frac{1}{2T}\operatorname{tr}\left(\boldsymbol{\Psi}_v^{-1}\mathbf{L}_v\right)^2=\frac{1}{(1-\rho_v^2)}
\end{aligned}$$

where we have used that $\mathbf{L}_\lambda=\frac{\partial\boldsymbol{\Psi}_\lambda}{\partial\rho_\lambda}=\frac{2\rho_\lambda}{\left(1-\rho_\lambda^2\right)}\boldsymbol{\Psi}_\lambda+\frac{1}{\left(1-\rho_\lambda^2\right)}\mathbf{D}$ with $\mathbf{D}$ a band matrix with zeros on the main diagonal and $i\rho_\lambda^{i-1}$ on the i:th subdiagonal and

$\operatorname{tr}\left(\mathbf{\Psi}_\lambda^{-1}\mathbf{D}\right) = 2\rho_\lambda (1-T)$, $\operatorname{tr}\left(\mathbf{\Psi}_\lambda^{-1}\mathbf{D}\right)^2 = 2\rho_\lambda^2 (T-1) + 2(T-1)$. Hence we arrive at (using assumption (c), lemma 6 and lemma 7)

$$\mathbf{V}^{-1}(\theta_0) = \begin{bmatrix} \phi & 0 & 0 \\ 0 & \mathbf{R_X} & 0 \\ 0 & 0 & \mathbf{V}^{-1}(\theta_0)_\gamma \end{bmatrix}$$

where $\mathbf{R_X} = \operatorname{plim}_{N,T\to\infty} \frac{1}{NT}\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}$, and with

$$\phi = \operatorname*{plim}_{T\to\infty} \begin{bmatrix} \frac{(1-\rho_{\lambda 0})^2}{\sigma_{u0}^2} & \frac{(1-\rho_\lambda)^2}{T\sigma_{u0}^2}\sum_{t=2}^{T-1}\mathbf{d}_t' & \mathbf{0} \\ & \frac{1}{T\sigma_{u0}^2}\sum_{t=2}^{T}\mathbf{d}_t^\lambda\mathbf{d}_t^{\lambda\prime} & \mathbf{0} \\ & & \frac{1}{\sigma_{\mu 0}^2}\operatorname{plim}_{N\to\infty}\frac{1}{N}\mathbf{h}'\overline{\mathbf{E}}_N\mathbf{h} \end{bmatrix}$$

if $\frac{N}{T}\to\infty$ where $\mathbf{d}_t^\lambda = (\mathbf{d}_t - \rho_\lambda\mathbf{d}_{t-1})$, $\overline{\mathbf{E}}_N = \mathbf{I}_N - \overline{\mathbf{J}}_N, \overline{\mathbf{J}}_N = \frac{1}{N}\boldsymbol{\iota}_N\boldsymbol{\iota}_N'$

$$\phi = \operatorname*{plim}_{N\to\infty} \begin{bmatrix} \frac{1}{\sigma_{\mu 0}^2} & 0 & \frac{1}{N\sigma_{\mu 0}^2}\sum_{i=1}^{N}\mathbf{h}_i' \\ & \operatorname{plim}_{T\to\infty}\frac{1}{T\sigma_{u0}^2}\mathbf{S}_\lambda & \mathbf{0} \\ & & \frac{1}{N\sigma_{\mu 0}^2}\mathbf{h}'\mathbf{h} \end{bmatrix}$$

if $\frac{T}{N}\to\infty$ where $\mathbf{S}_\lambda = \left(\sum_{t=2}^{T}\mathbf{d}_t^\lambda\mathbf{d}_t^{\lambda\prime} - \frac{(1-\rho_{\lambda 0})^2}{T}\sum_{t=2}^{T-1}\sum_{r=2}^{T-1}\mathbf{d}_t\mathbf{d}_r'\right)$, and

$$\phi = \operatorname*{plim}_{z\to\infty} \begin{bmatrix} \omega & \omega\frac{1}{T}\sum_{t=2}^{T-1}\mathbf{d}_t' & \omega\frac{1}{N}\sum_{i=1}^{N}\mathbf{h}_i' \\ & \frac{1}{T\sigma_{u0}^2}\mathbf{S}_\lambda + \omega\frac{1}{T^2}\sum_{t=2}^{T-1}\sum_{r=2}^{T-1}\mathbf{d}_t\mathbf{d}_r' & \frac{\omega}{NT}\sum_{t=2}^{T-1}\mathbf{d}_t\sum_{i=1}^{N}\mathbf{h}_i' \\ & & \frac{1}{N\sigma_{\mu 0}^2}\mathbf{h}'\overline{\mathbf{E}}_N\mathbf{h} + \omega\frac{1}{N}\mathbf{h}'\overline{\mathbf{J}}_N\mathbf{h} \end{bmatrix}$$

if $N,T\to\infty$ simultaneously, where $\omega = \frac{1}{\sigma_{\mu 0}^2 + (1-\rho_{\lambda 0})^{-2}\sigma_{u0}^2}$, $z = N$ or T and $\mathbf{V}^{-1}(\boldsymbol{\theta}_0)_\gamma$ is a diagonal matrix with

$$\operatorname{diag}\mathbf{V}^{-1}(\theta_0)_\gamma = \left\{\frac{1}{2\sigma_{\mu 0}^4}, \frac{1}{2\sigma_{e0}^4}, \frac{1}{\left(1-\rho_{v0}^2\right)}, \frac{1}{2\sigma_{u0}^4}, \frac{1}{\left(1-\rho_{\lambda 0}^2\right)}\right\}$$

To show (iii) note that the elements of the score for $\boldsymbol{\delta}$ is a linear combination of the normal ε and the score for the variance parameters, γ are linear combinations of quadratic forms in normal variates i.e. they can be written as

$$b + \varepsilon'\mathbf{P}\varepsilon$$

for suitable choice of b and symmetric matrix $\mathbf{P}$. We then apply the following lemma adapted from Amemiya (1971) to the quadratic forms in appendix A.1

Lemma. *Let an n-component vector random variable* $u \sim N(\mathbf{0}, \mathbf{\Lambda})$, $\mathbf{G}$ *be a non-negative definite symmetric matrix with rank* $r \leq n$. *Then* $\mathbf{u}'\mathbf{G}\mathbf{u}$ *is distributed as* $\sum_{i=1}^{r} \varphi_i \chi_i^2(1)$, *where the* φ*'s are* r *non-zero characteristic roots of* $\mathbf{\Lambda G}$ *and each* $\chi_i^2(1)$ *is an independent chi-square. If* $\mathbf{H}$ *is another non-negative definite symmetric matrix,* $\operatorname{cov}(\mathbf{u}'\mathbf{G}\mathbf{u}, \mathbf{u}'\mathbf{H}\mathbf{u}) = 2\operatorname{tr}(\mathbf{G\Lambda H\Lambda})$.

Asymptotic normality of the appropriately normalized score vector can then be shown by establishing sequential weak convergence results in case $\frac{N}{T} \to \infty$ or $\frac{T}{N} \to \infty$ (see Phillips and Moon (1999, section 3.3)) and in case $N, T \to \infty$ simultaneously a multivariate CLT for triangular arrays may be applied.

To establish the results as only N or $T \to \infty$ we apply the expansion (B.9) to the consistent subvectors $\theta^i = \left(\boldsymbol{\beta}, \boldsymbol{\tau}, \gamma^{(i)}\right), \gamma^{(i)} = \left(\sigma_\mu^2, \sigma_e^2, \rho_v\right)$ and $\theta^t = \left(\boldsymbol{\beta}, \boldsymbol{\pi}, \gamma^{(t)}\right), \gamma^{(t)} = \left(\sigma_e^2, \rho_v, \sigma_u^2, \rho_\lambda\right)$ as N and $T \to \infty$ respectively. This gives

$$\mathbf{F}_N\left(\widehat{\theta}^i - \theta_0^i\right) = -\left[\mathbf{F}_N^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\gamma)}{\partial\theta^i\partial\theta^{i\prime}}|_{\overline{\theta}^i}\right)\mathbf{F}_N^{-1}\right]^{-1}\left[\mathbf{F}_N^{-1}\left(\frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial\theta^i}|_{\theta_0^i}\right)\right]$$

and

$$\mathbf{F}_T\left(\widehat{\theta}^t - \theta_0^t\right) = -\left[\mathbf{F}_T^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\gamma)}{\partial\theta^t\partial\theta^{t\prime}}|_{\overline{\theta}^t}\right)\mathbf{F}_T^{-1}\right]^{-1}\left[\mathbf{F}_T^{-1}\left(\frac{\partial l(\boldsymbol{\delta},\gamma)}{\partial\theta^t}|_{\theta_0^t}\right)\right]$$

where $\mathbf{F}_N, \mathbf{F}_T$ are diagonal matrices with

$$\begin{aligned} \operatorname{diag}\mathbf{F}_N &= \left\{\mathbf{F}_\beta, \mathbf{F}_\tau, \sqrt{N}, \sqrt{NT}, \sqrt{NT}\right\} \\ \operatorname{diag}\mathbf{F}_T &= \left\{\mathbf{F}_\beta, \mathbf{F}_\pi, \sqrt{NT}, \sqrt{NT}, \sqrt{T}, \sqrt{T}\right\} \end{aligned}$$

To show (ii) for these cases we need to examine the convergence of the information matrices as N and $T \to \infty$ respectively. As $N \to \infty$ we find (using assumption (c), lemma 6, lemma 7 and straightforward computations)

$$\lim_{N\to\infty} E\left[\mathbf{F}_N^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\gamma)}{\partial\theta^i\partial\theta^{i\prime}}|_{\theta_0^i}\right)\mathbf{F}_N^{-1}\right] = \begin{bmatrix} \mathbf{R}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_N^{-1}\left(\theta_0^i\right)_{\gamma^{(i)}} \end{bmatrix}$$

where

$$\begin{aligned} &\mathbf{V}_N^{-1}\left(\theta_0^i\right)_{\gamma^{(i)}} \\ &= \frac{1}{2}\begin{bmatrix} \left(\frac{\left(\sigma_\alpha^2-\sigma_e^2\right)}{\sigma_\mu^2\sigma_\alpha^2}\right)^2 & \frac{(1-\rho_v)^2}{\sigma_\alpha^2\sqrt{T}}\iota_T^{\alpha\prime}\mathbf{A}^*\mathbf{\Psi}_v\iota_T^\alpha & \frac{\sigma_e^2(1-\rho_v)^2}{\sigma_\alpha^2\sqrt{T}}\iota_T^{\alpha\prime}\mathbf{A}^*\mathbf{L}_v\iota_T^\alpha \\ & \frac{1}{T}\left(\sigma_\alpha^{-4} + (T-1)\sigma_e^{-4}\right) & \frac{\sigma_e^2}{T}\operatorname{tr}\left(\mathbf{A}^*\mathbf{\Psi}_v\mathbf{A}^*\mathbf{L}_v\right) \\ & & \frac{\sigma_e^4}{T}\operatorname{tr}\left(\mathbf{A}^*\mathbf{L}_v\right)^2 \end{bmatrix} \end{aligned}$$

with $\mathbf{V}_N^{-1}\left(\theta_0^i\right)_{\gamma^{(i)}}$ positive-definite by theorem 1 and standard results of multivariate calculus.

As $T \to \infty$ we have (using assumption (c), lemma 6, lemma 7 and straightforward computations again)

$$\lim_{T\to\infty} E\left[\mathbf{F}_T^{-1}\left(\frac{\partial^2 l(\boldsymbol{\delta},\boldsymbol{\gamma})}{\partial\theta^t\partial\theta^{t\prime}}|_{\theta_0^t}\right)\mathbf{F}_T^{-1}\right] = \begin{bmatrix} \mathbf{R}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_T^{-1}\left(\theta_0^t\right)_{\gamma^{(t)}} \end{bmatrix}$$

with

$$\begin{aligned} &\mathbf{V}_T^{-1}\left(\theta_0^t\right)_{\gamma^{(t)}} \\ &= \lim_{T\to\infty}\frac{1}{2}\begin{bmatrix} \frac{1}{\sigma_e^4 T}\mathbf{V}^N_{\boldsymbol{\Psi}_v,\boldsymbol{\Psi}_v} & \frac{1}{\sigma_e^2 T}\mathbf{V}^N_{\mathbf{L}_v,\boldsymbol{\Psi}_v} & \frac{\sqrt{N}}{\sigma_e^4 T}\mathbf{V}_{\boldsymbol{\Psi}_v,\boldsymbol{\Psi}_\lambda} & \frac{\sigma_u^2\sqrt{N}}{\sigma_e^4 T}\mathbf{V}_{\boldsymbol{\Psi}_v,\mathbf{L}_\lambda} \\ & \frac{1}{T}\mathbf{V}^N_{\mathbf{L}_v,\mathbf{L}_v} & \frac{\sqrt{N}}{\sigma_e^2 T}\mathbf{V}_{\mathbf{L}_v,\boldsymbol{\Psi}_\lambda} & \frac{\sigma_u^2\sqrt{N}}{\sigma_e^2 T}\mathbf{V}_{\mathbf{L}_v,\mathbf{L}_\lambda} \\ & & \frac{N^2}{T\sigma_e^4}\mathbf{V}_{\boldsymbol{\Psi}_\lambda,\boldsymbol{\Psi}_\lambda} & \frac{\sigma_u^2 N^2}{\sigma_e^4 T}\mathbf{V}_{\mathbf{L}_\lambda,\boldsymbol{\Psi}_\lambda} \\ & & & \frac{\sigma_u^4 N^2}{T\sigma_e^4}\mathbf{V}_{\mathbf{L}_\lambda,\mathbf{L}_\lambda} \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} \mathbf{V}_{\mathbf{F},\mathbf{P}} &= \operatorname{tr}\left(\left(\boldsymbol{\Psi}_v^{-1}\mathbf{F}\boldsymbol{\Psi}_v^{-1}\mathbf{P}\right)(\mathbf{I}_T - 2\mathbf{M})\right) + \operatorname{tr}\left(\boldsymbol{\Psi}_v^{-1}\mathbf{F}\mathbf{M}\boldsymbol{\Psi}_v^{-1}\mathbf{P}\mathbf{M}\right) \\ \mathbf{V}^N_{\mathbf{F},\mathbf{P}} &= \operatorname{tr}\left(\left(\boldsymbol{\Psi}_v^{-1}\mathbf{F}\boldsymbol{\Psi}_v^{-1}\mathbf{P}\right)\left(\mathbf{I}_T - \frac{2}{N}\mathbf{M}\right)\right) + \frac{1}{N}\operatorname{tr}\left(\boldsymbol{\Psi}_v^{-1}\mathbf{F}\mathbf{M}\boldsymbol{\Psi}_v^{-1}\mathbf{P}\mathbf{M}\right) \end{aligned}$$

and $\mathbf{M} = \left(\mathbf{I}_T + \frac{\sigma_e^2}{N\sigma_u^2}\boldsymbol{\Psi}_\lambda\boldsymbol{\Psi}_v\boldsymbol{\Psi}_\lambda^{-2}\right)^{-1}$. The positive-definiteness of $\mathbf{V}_T^{-1}\left(\theta_0^t\right)_{\gamma^{(t)}}$ then follows from theorem 1.

These results show that the information elements of the subsets of consistent variance parameters do not depend on the inconsistent nuisance parameters as $N \to \infty$ and $T \to \infty$ respectively. To show this for the information elements of the subsets of consistent regression parameters as well we write, as in lemma 6,

$$\begin{aligned} \frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} &= \frac{1}{NT}\sum_{i=1}^N \mathbf{X}_i'\mathbf{A}^*\mathbf{X}_i - \frac{1}{N^2T}\sum_{j=1}^N\sum_{i=1}^N \mathbf{X}_j'\mathbf{A}^*\mathbf{X}_i \\ &\quad + \frac{1}{N^2T}\sum_{j=1}^N\sum_{i=1}^N \mathbf{X}_j'\mathbf{L}^{-1}\mathbf{X}_i \end{aligned}$$

where $\mathbf{L} = \left(N\sigma_u^2\boldsymbol{\Psi}_\lambda + (\mathbf{A}^*)^{-1}\right)^{-1} = O\left(N^{-1}\right)$. Hence

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} &= \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{A}^*\mathbf{X}_i - \operatorname*{plim}_{N\to\infty} \frac{1}{N^2T}\sum_{j=1}^{N}\sum_{i=1}^{N}\mathbf{X}_j'\mathbf{A}^*\mathbf{X}_i \\
&= \mathbf{R}_1\left(\theta_0^i\right)
\end{aligned}$$

and as $T \to \infty$ we find

$$\operatorname*{plim}_{T\to\infty} \frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} = \mathbf{R}_2\left(\theta_0^t\right)$$

since $\mathbf{A}^* = \mathbf{C}'\left(\sigma_\alpha^{-2}\overline{\mathbf{J}}_T^\alpha + \sigma_e^{-2}\mathbf{E}_T^\alpha\right)\mathbf{C}$ where $\sigma_\alpha^2 = O(T)$. Similarly one can show that cross-elements as well as information elements of time-invariant explanatory variables and individual-invariant random variables do not depend on nuisance parameters as N and $T \to \infty$ respectively. Finally, asymptotic normality of the limiting score vectors (suitably normalized of course) follows from applying a suitable multivariate CLT ■

Proof theorem 3. The negative of the log-likelihood is (apart from a constant term) given by

$$\begin{aligned}
\phi(\boldsymbol{\delta},\gamma) &= \frac{N}{2}\ln|\mathbf{A}^*| + \frac{1}{2}(\boldsymbol{\delta}_0-\boldsymbol{\delta})'\mathbf{Z}'(\mathbf{I}_N\otimes\mathbf{A}^*)\mathbf{Z}(\boldsymbol{\delta}_0-\boldsymbol{\delta}) \\
&\quad + \frac{1}{2}\boldsymbol{\varepsilon}'(\mathbf{I}_N\otimes\mathbf{A}^*)\boldsymbol{\varepsilon} + (\boldsymbol{\delta}_0-\boldsymbol{\delta})'\mathbf{Z}'(\mathbf{I}_N\otimes\mathbf{A}^*)\boldsymbol{\varepsilon}
\end{aligned}$$

Since the parameters ρ_λ, ρ_v play no role in what follows we assume $\rho_\lambda = \rho_v = 0$. To prove (i) note that

$$\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\right) = \frac{N\left(T\sigma_{\mu0}^2 + \sigma_{e0}^2\right)}{\left(T\sigma_\mu^2 + \sigma_e^2\right)} + \frac{N(T-1)\sigma_{e0}^2}{\sigma_e^2} + \frac{N\sigma_{u0}^2}{\left(T\sigma_\mu^2 + \sigma_e^2\right)} + \frac{N(T-1)\sigma_{u0}^2}{\sigma_e^2}$$

which contradicts a consistent root of σ_μ^2, σ_e^2 as $N \to \infty$ and a consistent root of σ_e^2 as $T \to \infty$ or $N, T \to \infty$ (and hence also of ρ_v as $N \to \infty, T \to \infty$ or $N, T \to \infty$). To show (ii) we need to investigate the behavior of $\mathbf{Z}'(\mathbf{I}_N\otimes\mathbf{A}^*)\mathbf{Z}$, $\mathbf{Z}'(\mathbf{I}_N\otimes\mathbf{A}^*)\boldsymbol{\varepsilon}$ which are explicitly written as

$$\begin{aligned}
&\mathbf{Z}'(\mathbf{I}_N\otimes\mathbf{A}^*)\mathbf{Z} \\
&= \begin{bmatrix}
\frac{NT}{\sigma_1^2} & \frac{1}{\sigma_1^2}\sum_{i=1}^N \boldsymbol{\iota}_T'\mathbf{X}_i & \frac{N}{\sigma_1^2}\boldsymbol{\iota}_T'\mathbf{d} & \frac{T}{\sigma_1^2}\sum_{i=1}^N h_i \\
 & \sum_{i=1}^N \mathbf{X}_i'\mathbf{A}^*\mathbf{X}_i & \sum_{i=1}^N \mathbf{X}_i'\mathbf{A}^*\mathbf{d} & \frac{1}{\sigma_1^2}\sum_{i=1}^N \mathbf{X}_i'\boldsymbol{\iota}_T h_i \\
 & & N\mathbf{d}'\mathbf{A}^*\mathbf{d} & \frac{1}{\sigma_1^2}\sum_{i=1}^N \mathbf{d}'\boldsymbol{\iota}_T h_i \\
 & & & \frac{T}{\sigma_1^2}\sum_{i=1}^N h_i'h_i
\end{bmatrix}
\end{aligned}$$

and

$$\mathbf{Z}'\left(\mathbf{I}_N \otimes \mathbf{A}^*\right)\boldsymbol{\varepsilon}$$
$$= \left[\begin{array}{c} \frac{T}{\sigma_1^2}\sum_{i=1}^N \mu_i + \frac{N}{\sigma_1^2}\sum_{t=1}^T \lambda_t + \frac{1}{\sigma_1^2}\sum_{t=1}^T\sum_{i=1}^N v_{it} \\ \frac{1}{\sigma_1^2}\sum_{t=1}^T\sum_{i=1}^N X_{it}\mu_i + \sum_{i=1}^N \mathbf{X}_i'\mathbf{A}^*\boldsymbol{\lambda} + \sum_{i=1}^N \mathbf{X}_i'\mathbf{A}^*\mathbf{v}_i \\ \frac{1}{\sigma_1^2}\sum_{t=1}^T\sum_{i=1}^N d_t\mu_i + N\mathbf{d}'\mathbf{A}^*\boldsymbol{\lambda} + \sum_{i=1}^N \mathbf{d}'\mathbf{A}^*\mathbf{v}_i \\ \frac{T}{\sigma_1^2}\sum_{i=1}^N h_i\mu_i + \frac{1}{\sigma_1^2}\sum_{i=1}^N h_i\sum_{t=1}^T \lambda_t + \frac{1}{\sigma_1^2}\sum_{t=1}^T\sum_{i=1}^N v_{it}h_i \end{array}\right]$$

with $\sigma_1^2 = T\sigma_\mu^2 + \sigma_e^2$. Proceeding as in the proof of theorem 1 then obtains the results in (ii) ■

Bibliography

Amemiya, T. (1971), 'The estimation of the variances in a variance-components model', *International Economic Review* **12**, 1–13.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

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

Balestra, P. and Nerlove, M. (1966), 'Pooling cross-section and time-series data in the estimation of a dynamic model: The demand for natural gas', *Econometrica* **34**, 585–612.

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.

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

Gourieroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volume Two*, Cambridge University Press.

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.

Phillips, P. C. and Moon, H. R. (1999), Linear regression limit theory for nonstationary panel data, Cowles Foundation Discussion Papers 1222, Yale University.

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

Rudin, W. (1976), *Principles of Mathematical Analysis*, McGraw-Hill.

Essay 3

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

3.1 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

*Co-authored with Sune Karlsson.

error terms or allows them both to be heteroskedastic, see Randolph (1988).

This essay 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), essay 1 and essay 2, among many others, contain further discussion of serially correlated period effects and/or serially correlated idiosyncratic errors.

The essay is organized as follows. In section 3.2 we present the model and discuss its relation to models previously suggested in the literature. Section 3.3 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 3.4 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 3.5 concludes.

3.2 The model

Consider the panel data regression model

$$y_{it} = \mathbf{z}'_{it}\boldsymbol{\delta}+\varepsilon_{it} \quad i = 1,\ldots,N;\ t = 1,\ldots,T \tag{3.1}$$

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),

$$\varepsilon_{it} = \mu_i + \lambda_t + v_{it} \tag{3.2}$$

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), essay 1 introduce serial correlation in the time effects via an

AR(1),

$$\lambda_t = \rho_\lambda \lambda_{t-1} + u_t, \tag{3.3}$$

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 essay 2 introduce serial correlation in the time effects and idiosyncratic errors via an AR(1) process for λ_t (3.3), and v_{it}

$$v_{it} = \rho_v v_{it-1} + e_{it} \tag{3.4}$$

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 (3.1,3.2). 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,

$$\varepsilon_{it} = \mu_i + v_{it} \tag{3.5}$$

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

$$\varepsilon_{it} = \lambda_t + v_{it} \tag{3.6}$$

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 (3.5) 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 (3.6) 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 (3.5) nor the auxiliary assumption of no individual effects in (3.6). Indeed the existence of such effects should be part of a hypothesis to be tested and not an assumption.

3.3 Estimation and model specification

3.3.1 Likelihood

In matrix form we can write the model (3.1,3.2) 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

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

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 $\boldsymbol{\Sigma}$. Direct inversion of $\boldsymbol{\Sigma}$ 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 essay 1 and 2 this is accomplished by using elementary results on inverses and determinants involving sums (Dhrymes (1984, p. 39-40)). This yields,

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\boldsymbol{\iota}_N \otimes \mathbf{I}_T)[\sigma_u^{-2}\boldsymbol{\Psi}_\lambda^{-1} + N\mathbf{A}^*]^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{I}_T)\mathbf{A}^{-1} \qquad (3.7) \\ &= \mathbf{I}_N \otimes \mathbf{A}^* - (\boldsymbol{\iota}_N \otimes \mathbf{A}^*)[\sigma_u^{-2}\boldsymbol{\Psi}_\lambda^{-1} + N\mathbf{A}^*]^{-1}(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \\ &= \mathbf{I}_N \otimes \mathbf{A}^* - \sigma_u^2(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\left[\mathbf{I}_T + N\sigma_u^2\boldsymbol{\Psi}_\lambda\mathbf{A}^*\right]^{-1}\boldsymbol{\Psi}_\lambda(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*) \end{aligned}$$

and

$$|\boldsymbol{\Sigma}| = |\mathbf{A}^*|^{-N}\left|\mathbf{I}_T + N\sigma_u^2\boldsymbol{\Psi}_\lambda\mathbf{A}^*\right| \qquad (3.8)$$

where $\mathbf{E}_N = \mathbf{I}_N - \overline{\mathbf{J}}_N$, $\overline{\mathbf{J}}_N = \mathbf{J}_N/N$, $\mathbf{A}^* = \left(\sigma_\mu^2 \mathbf{J}_T + \sigma_e^2 \boldsymbol{\Psi}_v\right)^{-1}$ 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 $\boldsymbol{\Sigma}$ has reduced the amount of numerical computation to the $T \times T$ matrices $\mathbf{A}^*$ and $\mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Psi}_\lambda \mathbf{A}^*$. 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} (3.4) 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, $\sigma_\alpha^2 = d^2 \sigma_\mu^2 (1-\rho_v)^2 + \sigma_e^2$, $\overline{\mathbf{J}}_T^\alpha = \boldsymbol{\iota}_T^\alpha \boldsymbol{\iota}_T^{\alpha\prime}/d^2$, $\boldsymbol{\iota}_T^{\alpha\prime} = \left(\alpha, \boldsymbol{\iota}_{T-1}'\right)$ and $\overline{\mathbf{E}}_T^\alpha = \mathbf{I}_T - \overline{\mathbf{J}}_T^\alpha$ with $d^2 = \boldsymbol{\iota}_T^{\alpha\prime} \boldsymbol{\iota}_T^\alpha = \alpha^2 + (T-1)$, $\alpha = \sqrt{(1+\rho_v)/(1-\rho_v)}$. 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 $\mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Psi}_\lambda \mathbf{A}^*$ is however more difficult. In the special case of $\boldsymbol{\Psi}_\lambda = \boldsymbol{\Psi}_v$ we can use the method of Baltagi and Li (1991, 1994) since

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

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

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

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 $\boldsymbol{\Psi}_\lambda \neq \boldsymbol{\Psi}_v$ the inverse of $\mathbf{I}_T + N\sigma_u^2 \boldsymbol{\Psi}_\lambda \mathbf{A}^*$ 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 (3.7) and (3.8) we can write the log-likelihood as

$$\begin{aligned} l(\theta) &= -\frac{TN}{2}\ln 2\pi + \frac{(N-1)}{2}\ln|\mathbf{A}^*| \\ &\quad -\frac{1}{2}\ln\left|\mathbf{I}_T + N\sigma_u^2\boldsymbol{\Psi}_\lambda\mathbf{A}^*\right| - \frac{1}{2}\boldsymbol{\varepsilon}'\left(\mathbf{I}_N \otimes \mathbf{A}^*\right)\boldsymbol{\varepsilon} \\ &\quad +\frac{\sigma_u^2}{2}\boldsymbol{\varepsilon}'(\boldsymbol{\iota}_N \otimes \mathbf{A}^*)\left[\mathbf{I}_T + N\sigma_u^2\boldsymbol{\Psi}_\lambda\mathbf{A}^*\right]^{-1}\boldsymbol{\Psi}_\lambda(\boldsymbol{\iota}_N' \otimes \mathbf{A}^*)\boldsymbol{\varepsilon} \end{aligned} \tag{3.9}$$

where $\theta = \left(\boldsymbol{\delta}', \boldsymbol{\gamma}\right)$ 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

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

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))

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

Note that if $\sigma_u^2 = 0$ in (3.9) 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 (3.9) 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.

3.3.2 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, γ. More specifically, in the framework of the model (3.1,3.2) 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.

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. Essay 1 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.

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 "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

[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.

effects to obtain an estimated Data Generating Process (DGP) with parameters $\widehat{\boldsymbol{\delta}}_0, \widehat{\gamma}_0$ 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

$$\xi_{\sup LM} = \sup_{\Xi} \left(\left. \frac{\partial l}{\partial \sigma_u^2} \right|_{\sigma_u^2 = 0} \right)' \mathcal{I}^{\sigma_u^2 \sigma_u^2} \left(\left. \frac{\partial l}{\partial \sigma_u^2} \right|_{\sigma_u^2 = 0} \right) \tag{3.10}$$

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.

Table 3.1 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

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.

3.4 Application

3.4.1 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 3.1.

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

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

$$\ln y_{it} = \alpha + \sum_{j=1}^{k} \beta_j \ln X_{jit} + \frac{1}{2} \sum_{s=1}^{k} \sum_{j=1}^{k} \phi_{sj} \ln X_{sit} \ln X_{jit} + \varepsilon_{it} \tag{3.11}$$

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).

3.4.2 Specification of the error components

Table 3.2 gives results from the LM-tests for serial correlation. For comparison Table 3.4 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 essay 1 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 3.2 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)

[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.

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 3.3 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 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 3.1 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 Figure 3.1 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 3.4 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.

3.4.3 Elasticities and returns to scale

Table 3.5 gives the overall mean of the input elasticities and returns to scale and Figure 3.2a-3.2d 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

[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 3.2 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 3.3 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, $B = 399$.

Figure 3.1 Size discrepancy of the supremum LM test

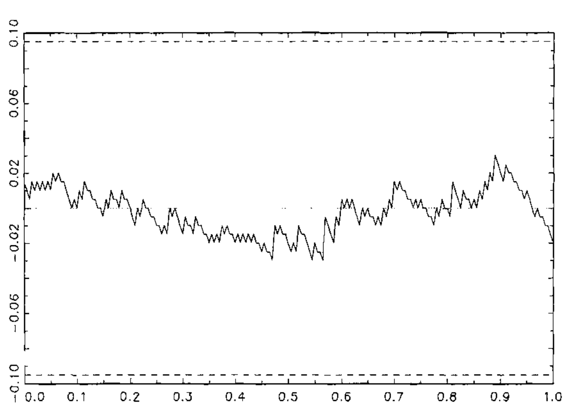


Table 3.4 Model selection criteria

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

Table 3.5 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

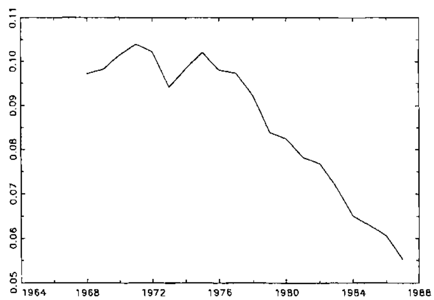
change in capital will result in 0.086% change in value added for given, labor, material and technology. The corresponding labor and material elasticities are 0.155 and 0.754 respectively. All three elasticities are statistically significant at the 1% level and are of expected (positive) sign. The sum of input elasticities is 0.995, indicating on average constant returns to scale. Looking at the temporal patterns of input elasticities (Figure 3.2) 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 3.2d 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.

3.4.4 Technical change

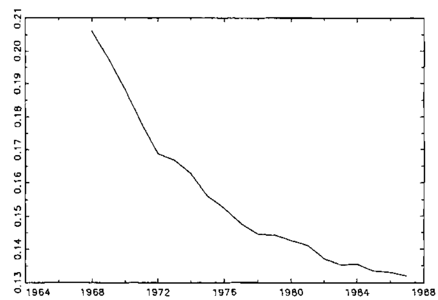
The last row of Table 3.5 gives the overall mean of total technical change and Figure 3.2e and 3.2f plots non-neutral and total technical change over time. The average rate of technical change is 0.9% per annum with contributions from the neutral and non-neutral components being −0.01% and 0.1% respectively. The time patterns of non-neutral and total technical change reflect changes in technology due to the oil crises, indicating substitution among in-

Figure 3.2 Elasticities, returns to scale and technical change

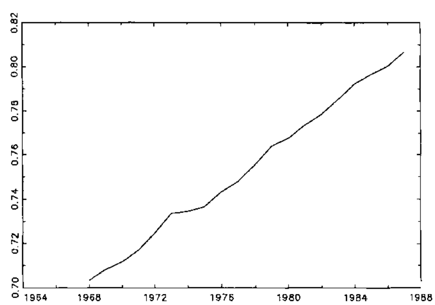
a) Elasticity of capital (K)



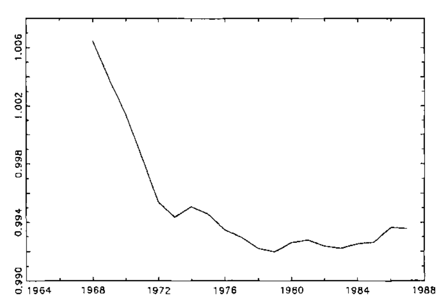
b) Elasticity of labor (L)



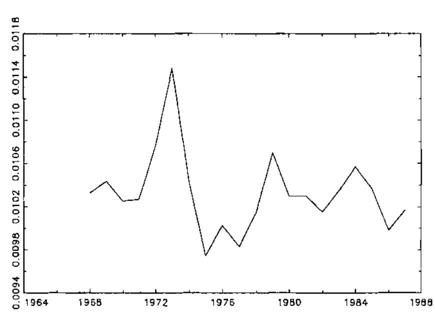
c) Elasticity of material (M)



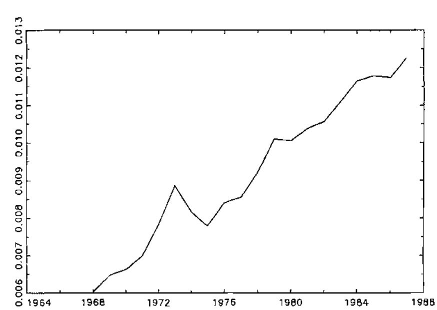
d) Returns to scale (RTS)



e) Non-neutral technical change



f) Total technical change



puts. As a consequence of the affine neutral component we do not observe major changes or any technical regress during the post oil crises period.

3.5 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.*

Appendix 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 essay 2 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 $(-1,1)$[1]. 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

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

where $\gamma = (\boldsymbol{\tau}, \rho, r)$, $\widetilde{\gamma} = (\widetilde{\boldsymbol{\tau}}, 0, 0)$, $\gamma_0 = (\boldsymbol{\tau}, 0, r)$, $\overline{\gamma}$ a mean value and $\mathbf{F}$ is an appropriate scaling matrix (see essay 2 for details). To introduce some notation for the information matrix we write $E\left[-\mathbf{F}^{-1}\left(\frac{\partial^2 l(\gamma)}{\partial \gamma \partial \gamma'}|_{\overline{\gamma}}\right)\mathbf{F}^{-1}\right] = \mathcal{I}(\gamma)$

[1]By theorem 2 in essay 2 we may ignore the mean parameters $\boldsymbol{\delta}$ for simplicity.

in partitioned form as,

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

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

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

where c is also an index obtained from the scaling matrix, $\mathbf{F}_{\tau}$ is the sub-scaling 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 $\mathbf{F}$ and (A.1) becomes

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

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

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

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

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

To investigate if the score equations (A.1) and (A.2) are in some sense equivalent it therefore suffices to consider the behavior of the term, $\mathcal{I}_{\rho_\lambda\rho_v,\tau}\rho_v\sqrt{NT}$ and the term, $\mathcal{I}_{\rho_v\rho_\lambda,\tau}\rho_\lambda\sqrt{T}$. By using the limit results of essay 2 it is straightforward to show that $\mathcal{I}_{\rho_\lambda\rho_v,\tau}$ and $\mathcal{I}_{\rho_v\rho_\lambda,\tau}$ shrink towards zero at the rate $N^{-\frac{3}{2}}$. Hence, $\mathcal{I}_{\rho_\lambda\rho_v,\tau}\rho_v\sqrt{NT}$ shrinks to zero if $\left(\frac{\sqrt{T}}{N}\right) \to 0$ and $\mathcal{I}_{\rho_v\rho_\lambda,\tau}\rho_\lambda\sqrt{T}$ 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 (A.2) 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 $(\sigma_\mu^2, \sigma_e^2, \rho_v)$ and the parameters $(\sigma_u^2, \rho_\lambda)$ as $N \to \infty$. This is accomplished by theorem 2 in essay 2. In addition theorem 2 shows that for large N the information of the parameters $(\sigma_\mu^2, \sigma_e^2, \rho_v)$ does not depend on the parameters $(\sigma_u^2, \rho_\lambda)$ 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 essay 2. Combining what we have obtained so far gives the results in property 1.

Bibliography

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.

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.

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.

Part II

GARCH

Essay 4

A simple efficient GMM estimator of GARCH models

4.1 Introduction

Unconditional return distributions are characterized by the 'stylized facts' of excess kurtosis, high peakedness and are often skewed, see Mills (1999, ch 5). But the conditional distribution is also characterized by excess kurtosis and skewness when the ARCH model of Engle (1982) or its generalization to GARCH by Bollerslev (1986) are fitted to return series, see e.g. French, Schwert and Staumbaugh (1987), Engle and Gonzales-Rivera (1991).

GARCH models are typically estimated by the method of maximum likelihood applied to the normal density, regardless of whether the conditional distribution is assumed normal or not. This may of course result in a large loss of efficiency relative to the true but unknown maximum likelihood estimator. In response Engle and Gonzales-Rivera (1991) introduced the semi-parametric maximum likelihood estimator of GARCH models. The semi-parametric estimator is a two-step estimator. In the first step consistent estimates of the parameters are obtained and are used to estimate a non-parametric conditional density. The second step consists of using this non-parametric density to adapt the initial estimator. The method has not been applied much in the literature though. This may be because the estimator is not so simple to compute, furthermore a choice of density estimator is required and specification testing is not straightforward.

This essay is concerned with efficient GMM estimation of GARCH models. In particular we show that efficient GMM is a feasible alternative to the quasi-maximum likelihood and semi-parametric estimators. Compared to the semi-parametric estimator efficient GMM has the advantage of being simple to compute and specification testing is straightforward. As a competitor to the common quasi-maximum likelihood estimator efficient GMM is asymptotically efficient with the coefficient of skewness and excess kurtosis of the conditional density being important in explaining the differences.

The organization of the essay is as follows. In section 4.2 we define the estimator in case of a GARCH(1,1) conditional variance model and give sufficient conditions for the estimator to be consistent and asymptotically normal. That is, to have the CAN property. Asymptotic relative efficiency comparison to the quasi-maximum likelihood estimator shows that efficient GMM is asymptotically more efficient under asymmetry of the conditional density and a small Monte-Carlo experiment confirms that the finite-sample gain can be substantial. Section 4.3 is concerned with efficient GMM estimation of the GARCH(1,1)-M regression model of Engle, Lilien and Robins (1987). It is shown that the introduction of a conditional mean makes the coefficient of excess kurtosis as well as of skewness important for explaining the relative efficiency gains of efficient GMM. In this section we also consider efficient GMM based specification tests. These tests are locally more powerful than the corresponding Bollerslev and Wooldridge (1992) robust classical tests whenever the efficient GMM estimator is asymptotically more efficient than quasi-maximum likelihood. Section 4.4 illustrates efficient GMM estimation and hypothesis testing with an application to the daily returns to the SP500 index, (1928-1991) and section 4.5 concludes. Proofs can be found in the appendix.

4.2 A GARCH(1,1) conditional variance process

4.2.1 The efficient GMM estimator

Consider the data generating process

$$\begin{aligned} \varepsilon_t &= z_t h_{0t}, z_t \sim iid(0,1) \\ h_{0t}^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{0t-1}^2 \end{aligned} \tag{4.1}$$

where the $l:th$ conditional moment of the ε_t process is $E[\varepsilon_t^l|^{t-1}] = v_l h_{0t}^l$, $^{t-1}$ is the information generated up to $t-1$ and $v_l = E\left(z_t^l\right)$. In practice we observe a finite segment of the process (4.1) and the objective is to estimate the parameters $\theta_0 \in \Theta$ where $\theta_0 = (\alpha_0, \alpha_1, \beta)'$ and Θ is a compact parameter space. To this end, define the (raw) vector

$$\mathbf{r}_t = [\varepsilon_t, \left(\varepsilon_t^2 - h_t^2\right)]' \tag{4.2}$$

and the generalized vector,

$$\mathbf{g}_t = \mathbf{F}_t'\mathbf{r}_t$$

where $\mathbf{F}_t$ is an instrumental variable function.

The GMM estimator of a parameter vector θ is the solution to (cf. Hansen (1982))

$$\text{argmin}_{\theta\in\Theta}\left[T^{-1}\sum_{t=1}^{T}\mathbf{g}_t\right]'\mathbf{W}_T\left[T^{-1}\sum_{t=1}^{T}\mathbf{g}_t\right]$$

with $\mathbf{W}_T = T^{-1}\sum_{t=1}^{T}\mathbf{W}_t$ an appropriate weighting matrix.

Efficient GMM corresponds to choosing $\mathbf{F}_t = \mathbf{\Sigma}_t^{-1}(\frac{\partial \mathbf{r}_t}{\partial\theta'})$ and $\mathbf{W}_t = (\frac{\partial \mathbf{r}_t'}{\partial\theta})\mathbf{\Sigma}_t^{-1}$ $\times(\frac{\partial \mathbf{r}_t}{\partial\theta'})$ where $\mathbf{\Sigma}_t = \text{var}(\mathbf{r}_t|^{t-1})$ and $(\frac{\partial \mathbf{r}_t}{\partial\theta'})$ is the Jacobian matrix, see Newey and McFadden (1994). The objective function for an operational efficient GMM estimator is then given by

$$Q_T = T^{-1}\left[\sum_{t=1}^{T}\mathbf{g}_t\right]'\left(\sum_{t=1}^{T}\mathbf{\Lambda}_t\right)^{-1}\left[\sum_{t=1}^{T}\mathbf{g}_t\right] \tag{4.3}$$

where $\mathbf{\Lambda}_t = \mathbf{g}_t\mathbf{g}_t'$ is a parameter dependent weighting matrix. The elements of the generalized moment and the weighting matrix are given by

$$\begin{aligned} g_{it} &= \frac{1}{\Delta}\left(\frac{\partial h_t^2}{\partial\theta_i}\right)h_t^{-2}\left[\left(\frac{\varepsilon_t}{h_t}\right)v_3 - \left(\left(\frac{\varepsilon_t^2}{h_t^2}\right)-1\right)\right] \\ \Lambda_{ijt} &= \frac{1}{\Delta^2}\left(\frac{\partial h_t^2}{\partial\theta_i}\right)h_t^{-2}\left[\left(\frac{\varepsilon_t}{h_t}\right)v_3 - \left(\left(\frac{\varepsilon_t^2}{h_t^2}\right)-1\right)\right]^2 h_t^{-2}\left(\frac{\partial h_t^2}{\partial\theta_j}\right) \end{aligned}$$

with $\Delta = (v_4 - 1) - v_3^2$[1].

By construction the objective function (4.3) is exactly identified and it is well-known from the literature that the choice of weighting matrix above is sufficient but not necessary for asymptotic efficiency. In fact asymptotic theory does not discriminate between a parameter dependent weighting matrix and a weighting matrix based on an initial consistent estimator of θ_0, or for that matter the identity matrix. Simplicity suggests that the identity matrix might be a good choice in applications. However finite-sample evidence indicates that allowing the weighting matrix to be parameter dependent is preferred and since the weighting matrix in (4.3) is typically constructed for inference little additional effort is needed for this choice.

To put quasi-maximum likelihood in the GMM framework choose $\mathbf{W}_t = \mathbf{s}_t\mathbf{s}_t'$, where $\mathbf{s}_t = \frac{\partial l_t}{\partial\theta}$ with l_t the normal quasi-loglikelihood for observation t

[1]By definition $\mathbf{\Sigma}_t$ is positive definite which implies $(v_4 - 1) - v_3^2 > 0$.

and $\mathbf{F}_t'\mathbf{r}_t = \mathbf{s}_t$. By noting that $\mathbf{g}_t = \frac{-2}{v_4-1}\mathbf{s}_t$ for $v_3 = 0$ the GMM objective function can be written

$$\left[T^{-1}\sum_{t=1}^{T}\mathbf{s}_t\right]' \mathbf{A}_T^{-1}\left[T^{-1}\sum_{t=1}^{T}\mathbf{s}_t\right]$$

where $\mathbf{A}_T = T^{-1}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{s}_t'$ and the asymptotic equivalence of efficient GMM based on (4.2) and quasi-maximum likelihood follows in case of symmetric z_t. Note however that in general efficient GMM requires an initial guess on the kurtosis and skewness of the rescaled variable. A guess on kurtosis can be based on the mean of $\frac{\varepsilon_t^4}{h_t^4(\widetilde{\theta}_T)}$, similarly the guess on skewness can be based on the mean of $\frac{\varepsilon_t^3}{h_t^3(\widetilde{\theta}_T)}$, where $\widetilde{\theta}_T$ is some initial estimator.

4.2.2 Asymptotic properties

All the asymptotic results below are derived with a parameter dependent weighting matrix. Compared to basing the weighting matrix on an initial consistent estimator of θ or simply the identity matrix no additional restrictions are needed. The latter estimators are of course special cases of the results given below. Furthermore allowing for a parameter dependent weighting matrix is unimportant for the asymptotic distribution.

The following assumptions are sufficient for the results

(a) $z_t \sim iid(0,1)$ and $\left(z_t, h_t^2\right)$ stochastically independent

(b) $E\left(z_t^4\right) = v_4 < \infty$

(c) $\beta + \alpha_1 < 1$

(d) $\Theta \equiv \{\theta : 0 < a_{0l} \leq a_0 \leq a_{0u}, 0 < a_{1l} \leq a_1 \leq a_{1u}, 0 < b_l \leq b \leq b_u < 1\}$

(e) $\theta_0 \in int(\Theta)$

In contrast to Lee and Hansen (1994) and Lumsdaine (1996) we do not allow for IGARCH. Besides giving a simpler asymptotic theory this means that we do not have to consider further restrictions on the parameter space as is necessary in the case of IGARCH. In particular the present framework can be used to establish consistency and asymptotic normality of the quasi-maximum likelihood estimator of the GARCH(1,1) process without restrictions on the parameter space.

Define $\widehat{\theta}_T$ as the sequence of minimizers of the objective function (4.3) and suppose some initial estimators $\widehat{v}_3, \widehat{v}_4$ are available and that these initial estimators only requires assumptions (a)-(e). We then have

Theorem 1 *Suppose assumptions (a)-(e) hold and that* $\widehat{v}_3 \xrightarrow{p} v_3^*, \widehat{v}_4 \xrightarrow{p} v_4^*$. *Then* $\widehat{\theta}_T \xrightarrow{p} \theta_0$ *on* Θ *regardless of* $v_3^* = v_3$ *or* $v_4^* = v_4$ *as long as both* v_3^*, v_4^* *are finite*

That is, $\widehat{\theta}_T$ is consistent for finite arbitrary guess on v_3, v_4. In practice we are of course interested in obtaining asymptotically valid inference about $\widehat{\theta}_T$ and for this purpose we need consistent initial estimators of v_3 and v_4. But this result has a useful consequence in terms of the asymptotic distribution of $\widehat{\theta}_T$. In particular we will be able to conclude that the asymptotic distribution of $\widehat{\theta}_T$ is the same regardless of wether v_3 and v_4 are known or estimated.

Let $\mathbf{\Lambda}_0 = E\mathbf{\Lambda}_t(\theta_0)$ and $\mathbf{G}_0 = E\mathbf{G}_t(\theta_0)$ where $\mathbf{G}_t(\theta) = \frac{\partial \mathbf{g}_t}{\partial \theta'}$

Theorem 2 *Suppose assumptions (a)-(e) holds and that* $\widehat{v}_3 \xrightarrow{p} v_3, \widehat{v}_4 \xrightarrow{p} v_4$ *or* v_3, v_4 *are known. Then* $T^{1/2}(\widehat{\theta}_T - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \left(\mathbf{G}_0\mathbf{\Lambda}_0^{-1}\mathbf{G}_0\right)^{-1}\right) \cong N\left(\mathbf{0}, \mathbf{\Lambda}_0^{-1}\right)$ *where " $\cong$ " denotes equality in distribution*

The above result allows us to compare asymptotic variances of efficient GMM and quasi-maximum likelihood. Taking expectations of $\mathbf{\Lambda}_t$, $\left(\mathbf{B}_t\mathbf{A}_t^{-1}\mathbf{B}_t\right)$ and evaluating them at θ_0, where $\mathbf{B}_t = \frac{\partial \mathbf{s}_t}{\partial \theta}$ and $\mathbf{A}_t = \mathbf{s}_t\mathbf{s}_t'$ we obtain

$$\mathbf{V}_{GMM} = \left[E\mathbf{\Lambda}_t(\theta_0)\right]^{-1} = \left[(v_4 - 1) - v_3^2\right]\left[E\left(\frac{\partial h_{0t}^2}{\partial \theta}\frac{\partial h_{0t}^2}{\partial \theta'}\right)\right]^{-1} h_{0t}^4$$

and

$$\begin{aligned}\mathbf{V}_{QMLE} &= \left[\left(E\mathbf{B}_t(\theta_0)\left[E\mathbf{A}_t(\theta_0)\right]^{-1}E\mathbf{B}_t(\theta_0)\right)\right]^{-1} \\ &= \left[(v_4 - 1)\right]\left[E\left(\frac{\partial h_{0t}^2}{\partial \theta}\frac{\partial h_{0t}^2}{\partial \theta'}\right)\right]^{-1} h_{0t}^4\end{aligned}$$

The relative efficiency ratio is seen to depend only on $\frac{(v_4-1)-v_3^2}{(v_4-1)}$ and is strictly decreasing in v_3. That is, efficient GMM is strictly more efficient than the quasi-maximum likelihood estimator when the conditional innovations have a skewed distribution.

4.2.3 Finite-sample properties

A small Monte-Carlo experiment is conducted to evaluate the finite-sample properties of the estimators[2]. We generate data from the GARCH(1,1) process (4.1) with two sets of parameter combinations close to what one commonly encounters in applying GARCH(1,1) to real data, $\theta_{01} = (0.1, 0.2, 0.7)'$ and $\theta_{02} = (0.05, 0.05, 0.9)'$ where $\theta_0 = (\alpha_0, \alpha_1, \beta)'$. The sample sizes considered are $T = 1000$ and 5000. Initial values were chosen arbitrarily as $(0.1, 0.25, 0.67)'$ and $(0.05, 0.07, 0.87)'$ for each set of parameters.

All the efficient GMM estimations are performed with the parameter dependent weighting matrix. Efficient GMM estimators with the weighting matrix provided by the initial consistent estimator, $\widetilde{\theta}_T$ (where $\widetilde{\theta}_T$ is the quasi-maximum likelihood estimate) and the identity matrix respectively performed less well. The performance was especially poor with the identity matrix where we experienced serious convergence problems.

Symmetric densities

For many financial return series a t-distribution with a few degrees of freedom fits the empirical density of z_t quite well. The question is if we can improve on quasi-maximum likelihood with the efficient GMM estimator when the rescaled variable, z_t has a fat-tailed density. The rescaled variable is assumed to follow a t-distribution with 5 degrees of freedom which gives true kurtosis of 9. The $t(5)$ distributed random variable is generated as the ratio of a standard normal and $\sqrt{\chi^2_5/5}$ variate. To obtain a $(0, 1)$ variable we divide by the standard deviation. Results are given in Table 4.1.

For the parameter combinations considered efficient GMM typically has a variance that is smaller than or equal to that of quasi-maximum likelihood, the gain is substantial for the case of θ_{02} and $T = 1000$. Furthermore the efficient GMM estimator has less bias than the quasi-maximum likelihood estimator in this case. The bias of the quasi-maximum likelihood estimator is small for θ_{01} but increases for the parameter vector θ_{02}. It appears that this tendency is not so strong for the efficient GMM estimator. For $T = 5000$ efficient GMM and quasi-maximum likelihood are equivalent for both θ_{01} and θ_{02}.

[2]Both estimators use the Newton algorithm. The efficient GMM estimator use Constrained Optimization (CO) module in GAUSS and quasi-maximum likelihood use the Constrained Maximum Likelihood (CML) module. In both cases analytical derivatives are used and constraints are imposed as $10^{-10} \leq a_0, a_1$ and $0 < b < 1$. All the results are based on 5000 replications and 100 initial values of the conditional variance process were discarded to avoid initialization effects. Quasi-maximum likelihood estimates are used to initialize a guess on v_3 and v_4 for the efficient GMM estimator.

Table 4.1 Finite sample comparison of efficient GMM (GMM) and quasi-maximum likelihood (QMLE), $z_t \sim t(5)$

	GMM		QMLE		
$T = 1000$	bias	std(1)	bias	std(2)	ratio (1)/(2)
θ_{01}					
a_0	0.010	0.045	0.011	0.046	0.978
a_1	0.005	0.066	0.005	0.068	0.970
b	−0.018	0.083	−0.018	0.085	0.976
θ_{02}					
a_0	0.011	0.051	0.027	0.081	0.629
a_1	0	0.030	0.005	0.032	0.938
b	−0.013	0.073	−0.034	0.104	0.702
$T = 5000$					
θ_{01}					
a_0	0.002	0.020	0.002	0.020	1
a_1	0.001	0.030	0.001	0.030	1
b	−0.005	0.037	−0.005	0.037	1
θ_{02}					
a_0	0.003	0.022	0.004	0.021	1.04
a_1	0.001	0.013	0.001	0.013	1
b	−0.005	0.030	−0.006	0.029	1.03

Table 4.2 Finite sample comparison of efficient GMM (GMM) and quasi-maximum likelihood (QMLE), $z_t \sim$ Gamma(2)

	GMM		QMLE		
$T = 1000$	bias	std(1)	bias	std(2)	ratio (1)/(2)
θ_{01}					
a_0	0.008	0.050	0.012	0.045	1.11
a_1	0.002	0.049	0.004	0.061	0.803
b	−0.013	0.074	−0.02	0.085	0.870
θ_{02}					
a_0	0.011	0.046	0.029	0.090	0.511
a_1	0.002	0.021	0.004	0.029	0.723
b	−0.015	0.059	−0.035	0.111	0.531
$T = 5000$					
θ_{01}					
a_0	0.001	0.013	0.002	0.018	0.722
a_1	0	0.020	0	0.026	0.789
b	−0.003	0.026	−0.005	0.035	0.743
θ_{02}					
a_0	0.002	0.013	0.004	0.018	0.722
a_1	0	0.008	0.001	0.011	0.727
b	−0.004	0.019	−0.005	0.026	0.730

Asymmetric densities

Unconditional return distributions are characterized by the 'stylized facts' of excess kurtosis, high peakedness and are often skewed. But the distribution of the rescaled variable as well is characterized by excess kurtosis and skewness when GARCH models are fitted to data see e.g. Engle and Gonzales-Rivera (1991). To consider asymmetric densities we generate z_t as Gamma-distributed with mean and variance parameter equal to 2. The Gamma(2) distributed random variable is obtained from *rndgam* in GAUSS and standardized by subtracting 2 from it and dividing by $2^{1/2}$. It has true kurtosis and skewness given by 6 and $2/\sqrt{2}$ respectively. Results are given in Table 4.2.

As for the case of symmetric z_t with $T = 1000$ the bias of the efficient GMM estimator seems to be smaller and the gain is larger for θ_{02}. For $T = 5000$ the efficient GMM estimator has smaller variance than quasi-maximum likelihood for both θ_{01} and θ_{02} which is what we expect since it is asymptotically more efficient. In fact the estimated standard deviation ratios are quite close to the theoretical ratio of approximately 0.775.

4.3 Extension to models with a conditional mean

4.3.1 The GARCH(1,1)-M regression model

The GARCH(1,1) conditional variance process considered in section 4.2 may be somewhat restrictive in practice. Here we consider some practical details of estimating more general models with the efficient GMM estimator introduced in section 4.2.1. The model of interest is the GARCH(1,1)-M regression model introduced by Engle et al. (1987)

$$\begin{aligned} y_t &= \mathbf{X}_t'\boldsymbol{\mu} + \delta f(h_t^2) + \varepsilon_t \\ \varepsilon_t &= z_t h_t \\ h_t^2 &= a_0 + a_1\varepsilon_{t-1}^2 + b h_{t-1}^2 \end{aligned} \tag{4.4}$$

where $f(h_t^2)$ is a function of the conditional variance and in addition to section 4.2.1 we assume that z_t is independent of $\mathbf{X}_t$. Sufficient conditions for consistency and asymptotic normality are, to the authors knowledge, not known even for the quasi-maximum likelihood estimator. In what follows we simply assume that such conditions are satisfied. More specifically, we assume that the CAN property holds for both quasi-maximum likelihood and efficient GMM.

By using the (raw) vector (4.2) we can in analogy with section 2.1 define the efficient GMM estimator of the model (4.4) as a solution to

$$\operatorname{argmin}_{\theta\in\Theta} T^{-2}\left[\sum_{t=1}^{T}\mathbf{g}_t\right]'\boldsymbol{\Lambda}_T^{-1}\left[\sum_{t=1}^{T}\mathbf{g}_t\right] \tag{4.5}$$

where the parameter vector θ is given by $\theta = (\gamma_1, \gamma_2)', \gamma_1 = (a_0, a_1, b)'$ and $\gamma_2 = (\delta, \mu')'$. The generalized moment, $\mathbf{g}_t$ is given by

$$\mathbf{g}_t = \frac{1}{\Delta}\left[\begin{array}{c} \left(\frac{\partial h_t^2}{\partial\theta}\right) h_t^{-2}\left[v_3\frac{\varepsilon_t}{h_t} - \left(\frac{\varepsilon_t^2}{h_t^2} - 1\right)\right] \\ + \left(\frac{\partial\varepsilon_t}{\partial\theta}\right) h_t^{-1}\left[\frac{\varepsilon_t}{h_t}(v_4 - 1) - v_3\left(\frac{\varepsilon_t^2}{h_t^2} - 1\right)\right] \end{array}\right] \tag{4.6}$$

with $\Delta = \left[(v_4 - 1) - v_3^2\right]$ and derivatives in (4.6) are computed recursively as

$$\frac{\partial h_t^2}{\partial\theta} = \mathbf{c}_{t-1} + \left(b - 2\delta a_1\varepsilon_{t-1}\frac{\partial f(h_{t-1}^2)}{\partial h_{t-1}^2}\right)\frac{\partial h_{t-1}^2}{\partial\theta}$$

where $\mathbf{c}_t = \left(1, \varepsilon_t^2, h_t^2, -2a_1\varepsilon_t f(h_t^2), -2a_1\varepsilon_t \mathbf{X}_t'\right)$, and

$$\frac{\partial \varepsilon_t}{\partial \theta} = \pi_t - \delta \frac{\partial f(h_t^2)}{\partial h_t^2}\frac{\partial h_t^2}{\partial \theta}$$

with $\pi_t = \left(0, 0, 0, -f(h_t^2), -\mathbf{X}_t'\right)$.

The equality of $E\mathbf{\Lambda}_t(\theta_0)$ and $E\mathbf{G}_t(\theta_0)$ is straightforward to verify from (4.6) using the iid property of z_t and hence the variance matrix of $T^{1/2}\widehat{\theta}_T$ can be consistently estimated by

$$\widehat{\mathbf{\Lambda}}_T^{-1} = \left(T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{\Lambda}}_t\right)^{-1}$$

or

$$\widehat{\mathbf{G}}_T^{-1} = \left(T^{-1}\sum_{t=1}^{T}\frac{\partial \widehat{\mathbf{g}}_t}{\partial \theta}\right)^{-1}$$

where $\widehat{\mathbf{\Lambda}}_t = \mathbf{\Lambda}_t\left(\widehat{\theta}_T\right)$ and $\frac{\partial \widehat{\mathbf{g}}_t}{\partial \theta} = \frac{\partial \mathbf{g}_t\left(\widehat{\theta}_T\right)}{\partial \theta}$. Inference based on $\widehat{\mathbf{\Lambda}}_T$ only involves first derivatives of the conditional mean and the conditional variance function. This is useful since estimation of GARCH models frequently rely on numerical approximations to the analytical derivatives.

4.3.2 Asymptotic efficiency comparison

For the purpose of comparing the asymptotic variance matrices of efficient GMM and quasi-maximum likelihood we let

$$\mathbf{S} = \left[\begin{array}{cc} h_{0t}^{-1}\frac{\partial m_{0t}}{\partial \theta} & \frac{1}{2}h_{0t}^{-2}\frac{\partial h_{0t}^2}{\partial \theta} \end{array}\right]$$

where $h_{0t}^2 = h_t^2(\theta_0)$ and $m_{0t} = m_t(\theta_0)$ is the regression function and define the matrices

$$\begin{aligned}
\mathbf{K} &= \left[\begin{array}{cc} 1 & 0 \\ 0 & 2 \end{array}\right] \\
\mathbf{L} &= \left[\begin{array}{cc} 1 & v_3 \\ v_3 & (v_4 - 1) \end{array}\right] \\
\mathbf{M} &= \frac{1}{\Delta}\left[\begin{array}{cc} (v_4 - 1) & -2v_3 \\ -2v_3 & 4 \end{array}\right]
\end{aligned}$$

with $\mathbf{L}, \mathbf{K}$ and $\mathbf{M}$ the 'weight matrices' for the $\mathbf{A}_t, \mathbf{B}_t$ and $\mathbf{\Lambda}_t$ matrices respectively. Define next

$$\mathbf{\Pi}_A = E\left(\mathbf{SDS}'\right)$$

for arbitrary positive semi-definite matrix $\mathbf{D}$.

We can then write the difference between the inverses of the asymptotic variance matrices as

$$\mathbf{V}_{GMM}^{-1} - \mathbf{V}_{QMLE}^{-1} = \mathbf{\Pi}_M - \mathbf{\Pi}_K \mathbf{\Pi}_L^{-1} \mathbf{\Pi}_K \tag{4.7}$$

which is positive semi-definite for a general conditional density and the model (4.4)[3].

Note that $v_3 = 0$, $v_4 = 3$ implies $\mathbf{K} = \mathbf{L} = \mathbf{M}$ and hence $\mathbf{V}_{QMLE} = \mathbf{V}_{GMM}$ if the conditional density is normal. In the case of excess kurtosis and/or skewness of the conditional density, $\mathbf{K} \neq \mathbf{L} \neq \mathbf{M}$. However a formal proof that this is sufficient for efficiency gains is too difficult and we consider some simple cases where positive results can be obtained.

Consider first the GARCH(1,1) regression model ($\delta = 0$ in (4.4)). Imposing symmetry of the conditional density is not sufficient for $\mathbf{V}_{QMLE} = \mathbf{V}_{GMM}$ but for the conditional variance parameters, γ_1 we have $\mathbf{V}(\gamma_1)_{QMLE} = \mathbf{V}(\gamma_1)_{GMM}$. Some tedious but straightforward algebra shows that

$$\begin{aligned}
\mathbf{V}^{-1}(\boldsymbol{\mu})_{GMM} - \mathbf{V}^{-1}(\boldsymbol{\mu})_{QMLE} &= \mathbf{P}_{11} + \frac{4}{(v_4 - 1)}\mathbf{P}_{12} - [\mathbf{P}_{11} + 2\mathbf{P}_{12}] \quad (4.8)\\
&\quad \times [\mathbf{P}_{11} + (v_4 - 1)\,\mathbf{P}_{12}]^{-1} [\mathbf{P}_{11} + 2\mathbf{P}_{12}]\\
&= \frac{(v_4 - 3)^2}{(v_4 - 1)^2} \left[\mathbf{P}_{11}^{-1} + \frac{1}{(v_4 - 1)}\mathbf{P}_{12}^{-1}\right]^{-1}
\end{aligned}$$

where $\mathbf{P}_{11} = E\mathbf{a}_{11,t}\mathbf{a}_{11,t}'$, $\mathbf{P}_{12} = E\mathbf{a}_{12,t}\mathbf{a}_{12,t}'$ with $\mathbf{a}_{11,t} = h_{0t}^{-1}\frac{\partial m_{0t}}{\partial \mu}$ and $\mathbf{a}_{12,t} = \frac{1}{2}h_{0t}^{-2}\frac{\partial h_{0t}^2}{\partial \mu}$. Note that (4.8) is positive definite for $v_4 > 3$ and increasing in v_4 implying there is efficiency gains for the conditional mean parameters in case of excess kurtosis.

A corresponding result for $\boldsymbol{\mu}$ and γ_1 under asymmetry of the conditional density is more difficult since the block-diagonal structure of $\mathbf{V}_{QMLE}$ and $\mathbf{V}_{GMM}$ is lost. We can however allow for $v_3 \neq 0$ if we set $\boldsymbol{\mu} = \mathbf{0}$. This gives the asymptotic relative efficiency result for the GARCH(1,1) conditional variance model, that is, $\frac{(v_4-1)-v_3^2}{v_4-1}$ obtained in section 2.2. In the GARCH(1,1)-M model (4.4) the block-diagonal structure of $\mathbf{V}_{QMLE}$ and $\mathbf{V}_{GMM}$ is lost even under symmetry of the conditional density. Thus in this case the algebra is too tedious to derive results such as (4.8) regardless of $v_3 = 0$ or not. However

[3] The matrix difference on the right-hand side of (4.7) appears as part of the difference between the inverse asymptotic variance matrices of the semi-parametric and maximum likelihood estimators in Gonzalez-Rivera and Drost (1999). They prove that this difference is positive semi-definite for the model (4.4).

we conjecture that due to the absence of block-diagonal structure there is efficiency gains for *both* conditional mean and variance parameters under excess kurtosis. Similar considerations apply for an asymmetric conditional density.

4.3.3 Specification tests

As in the familiar maximum likelihood setting the classical LM (score), Wald and LR-tests are available and derived with expansions involving the first and second derivatives of the objective function (cf. Newey and West (1987)). Under the null hypothesis, say $H_0 : \mathbf{a}(\theta_0) = 0$, the LM (score) statistic is given by

$$\xi_{LM} = \sqrt{T}\widetilde{\mathbf{g}}_T'\widetilde{\mathbf{\Lambda}}_T^{-1}\widetilde{\mathbf{G}}_T\left[\widetilde{\mathbf{G}}_T'\widetilde{\mathbf{\Lambda}}_T^{-1}\widetilde{\mathbf{G}}_T\right]^{-1}\widetilde{\mathbf{G}}_T'\widetilde{\mathbf{\Lambda}}_T^{-1}\sqrt{T}\widetilde{\mathbf{g}}_T \tag{4.9}$$

where "$\sim$" denotes evaluated under the null hypothesis. Under our assumptions on the rescaled variable, z_t, we obtain the asymptotically equivalent form

$$\xi_{LM} = \sum_{t=1}^{T}\widetilde{\mathbf{g}}_t'\left(\sum_{t=1}^{T}\widetilde{\mathbf{g}}_t\widetilde{\mathbf{g}}_t'\right)^{-1}\sum_{t=1}^{T}\widetilde{\mathbf{g}}_t \tag{4.10}$$

which is simply TR_u^2 from the linear regression of 1 on $\widetilde{\mathbf{g}}_t'$. As an example of the TR^2 form of the LM test consider the efficient GMM counterpart of the Engle (1982) ARCH(m) classical LM test with no parameters in the conditional mean, we have

$$\xi_{LM} = \frac{1}{\Delta}\mathbf{W}_1'\mathbf{W}_2\left(\mathbf{W}_2\mathbf{W}_2'\right)^{-1}\mathbf{W}_2'\mathbf{W}_1$$

where $\mathbf{W}_1' = (w_{11}, ..., w_{1T})$, $\mathbf{W}_2' = (w_{21}, ..., w_{2T})$ with $w_{1t} = v_3\frac{\varepsilon_t}{\widetilde{h}_t} - \left(\frac{\varepsilon_t^2}{\widetilde{h}_t^2} - 1\right)$ and $w_{2t} = \widetilde{h}_t^{-2}\widetilde{\frac{\partial h_t^2}{\partial\theta}}$. Since

$$\text{plim}_{T\to\infty}T^{-1}\sum_{t=1}^{T}\left[v_3\frac{\varepsilon_t}{\widetilde{h}_t} - \left(\frac{\varepsilon_t^2}{\widetilde{h}_t^2} - 1\right)\right]^2 = (v_4 - 1) - v_3^2$$

we can construct the asymptotically equivalent TR^2 form

$$\xi_{LM} = T\frac{\mathbf{W}_1'\mathbf{W}_2\left(\mathbf{W}_2\mathbf{W}_2'\right)^{-1}\mathbf{W}_2'\mathbf{W}_1}{\mathbf{W}_1'\mathbf{W}_1} = TR_u^2$$

where R_u^2 is the unadjusted squared multiple correlation coefficient from a regression of $\mathbf{W}_1$ on $\mathbf{W}_2$. Under the null hypothesis of no ARCH(m) $\widetilde{h}_t^2 =$

σ_ε^2 and $\mathbf{W}_2 = \left(1, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, ..., \varepsilon_{t-m}^2\right)$ so that the LM test of the null of no ARCH(m) is computed by TR_u^2 from the regression of $\left[\widetilde{v}_3 \frac{\varepsilon_t}{\sigma_\varepsilon} - \left(\frac{\varepsilon_t^2}{\sigma_\varepsilon^2} - 1\right)\right]$ on $\mathbf{W}_2$, where $\widetilde{v}_3$ is the skewness of y_t. That is, in contrast to the usual TR^2 form of the ARCH(m) LM test (which is computed by TR_u^2 from the regression of $\left(\frac{\varepsilon_t^2}{\sigma_\varepsilon^2} - 1\right)$ on $\mathbf{W}_2$) we have not imposed the auxiliary assumption of normality.

Similarly the Wald and quasi-LR statistics are available. The Wald statistic has its usual form and the quasi-LR test is given by

$$\xi_{QLR} = T\left[Q_T(\widetilde{\theta}_T) - Q_T(\widehat{\theta}_T)\right] \tag{4.11}$$

In the exactly identified case $Q_T(\widehat{\theta}_T) = 0$ and hence

$$\xi_{QLR} = T\left[Q_T(\widetilde{\theta}_T)\right] \tag{4.12}$$

which is computed holding the parameters in the null hypothesis fixed at their respective null hypothesis during the iterations. Note that $\xi_{QLR} = \xi_{LM}$ in this context.

As shown by Newey and West (1987) the LM, Wald and quasi-LR tests are asymptotically equivalent under the null and local alternatives. Under the alternative hypothesis $H_A : \mathbf{a}(\theta_0) = T^{-1/2}\boldsymbol{\lambda}$ these tests have the $\chi_q^2\left(\zeta_{GMM}\right)$ distribution, where $\zeta_{GMM} = \boldsymbol{\lambda}'\left[\mathbf{A}\left(\theta_0\right)\mathbf{V}_{GMM}^{-1}\mathbf{A}\left(\theta_0\right)\right]^{-1}\boldsymbol{\lambda}$ is the non-centrality parameter and $\mathbf{A}\left(\theta_0\right)$ is the Jacobian matrix of the $q\times 1$ vector function $\mathbf{a}(\theta)$ of the $p\times 1$ vector θ with $q \leq p$. It is useful to compare this non-centrality parameter to the non-centrality parameter of the Bollerslev and Wooldridge (1992) robust classical tests which is given by $\zeta_{QMLE} = \boldsymbol{\lambda}'\left[\mathbf{A}\left(\theta_0\right)\mathbf{V}_{QMLE}^{-1}\mathbf{A}\left(\theta_0\right)\right]^{-1}\boldsymbol{\lambda}$. From the results of the previous section it is clear that the GMM based tests are more powerful in terms of local asymptotic power. In particular, whenever efficient GMM is asymptotically more efficient than quasi-maximum likelihood.

4.4 Empirical illustration

In this section we apply the efficient GMM estimator to model the returns to the daily SP500 index, (1928-1991) as a GARCH-M model. Inclusion of a measure of volatility in the conditional mean of returns is an attempt to introduce a measure of risk. It is an implication of the 'mean-variance hypothesis' of the returns and under this hypothesis large returns are expected to be associated with high volatility.

Table 4.3 Unconditional moments of returns to the SP500 index

Mean	0.01821
Variance	1.32377
Skewness	−0.48725
Kurtosis	25.4164

The GARCH-M model has been applied by several researchers to model the relation between risk and return e.g. French et al. (1987) applied the GARCH-M model to subsets of the excess returns to the SP500 index and concluded that the data was consistent with a positive relation between conditional expected excess return and conditional variance. However Glosten, Jagannathan and Runkle (1993) obtained a significant negative relation between the conditional mean and the conditional variance of excess returns on stocks when the model was modified to allow positive and negative anticipated returns to have different impact on the conditional variance.

The distributional properties of the returns to the daily SP500 index has been analyzed in Mills (1999, ch 5) and Granger and Ding (1995) investigated the properties of absolute returns. Table 4.3 gives the estimates of the first four unconditional moments of the distribution of returns to the SP500 index.

The modelling strategy is to first specify the conditional mean and once satisfactory, tests for the conditional variance specification are performed. Only autocorrelation in the conditional mean is tested for and any possible nonlinearity of the conditional mean is disregarded. In tems of the conditional variance specification we restrict ourselves to the GARCH(1,1) case.

Fitting an AR(5) model to the returns and computing the LM (4.10) (quasi-LR (4.12)) test of the null hypothesis of no ARCH(4) gives a test-statistic of 55.67 with a corresponding p-value of 7.2×10^{-14} and hence we reject the null at any reasonable significance level. This leads us to estimate the following GARCH(1,1)-M model for the returns

$$\begin{aligned} y_t &= \mu_0 + \mu_1 y_{t-1} + \mu_2 y_{t-2} + \mu_3 y_{t-3} + \mu_4 y_{t-4} + \mu_5 y_{t-5} + \delta h_t + \varepsilon_t \\ h_t^2 &= a_0 + a_1 \varepsilon_{t-1}^2 + b h_{t-1}^2 \end{aligned}$$

where the choice of the conditional standard deviation specification of the GARCH(1,1)-M term is of course arbitrary.

Table 4.4 gives the estimation results together with Box-Pierce statistics of the levels and squares, $Q(\cdot)$ and $Q^2(\cdot)$ respectively (Box and Pierce (1970)) and Table 4.5 gives the first four moments of the rescaled residuals. Quasi-maximum likelihood estimates are given for comparison[4].

Inpsection of Table 4.4 shows that there is evidence for strong persistence in the conditional variance. The sum $a_1 + b$ is only slightly below unity for both efficient GMM and quasi-maximum likelihood. The estimates of the risk-premia term, δ are positive for both efficient GMM and quasi-maximum likelihood. Note however that the efficient GMM estimate is less than half the estimate of quasi-maximum likelihood and in contrast it is not significant at usual levels. Box-Pierce statistics of the levels and squares suggests that there might be some dynamics left in the data. If we follow the recommendation of Box and Pierce and compare with the χ^2_7 and χ^2_{19} distribution for the $Q(12)$ and $Q(24)$ statistics, respectively, they are significant at 5% level. The $Q^2(12)$ and $Q^2(24)$ statistics indicate that there might be some dynamics left in the conditional variance as well.

To entertain the possibility that we need to differentiate between bad and good news we consider a GJR-GARCH(1,1) specification for the conditional variance (Glosten et al. (1993)). The GJR-GARCH(1,1) specification is

$$h_t^2 = a_0 + \left(a_1 + a_2 S_{t-1}^-\right) \varepsilon_{t-1}^2 + b h_{t-1}^2 \tag{4.13}$$

where S_{t-1}^- is an indicator function which takes the value 1 when $\varepsilon_t < 0$ and value 0 when $\varepsilon_t > 0$.

An LM test of the null of no asymmetry against the alternative of the GJR-GARCH(1,1) specification gives a test-statistic of 31.55 with a corresponding p-value of 1.94×10^{-8} suggesting that there is a need to differentiate beween good and bad news.

Results from re-estimation with the GJR-GARCH(1,1) specification (4.13) for the conditional variance yield only marginally different results from Table 4.4 and are not reproduced here. In summary the Box-Pierce statistics have been reduced considerably and the GJR parameter is positive and strongly significant but δ is now close to zero and insignificant at usual levels for both efficient GMM and quasi-maximum likelihood. These results offer no evidence for a GARCH(1,1)-M formulation of the returns to the SP500 index but suggests that it is important to allow for asymmetric GARCH.

[4]Both estimators use the Newton algorithm with analytic first derivatives of the conditional mean and variance functions. The efficient GMM estimator use Constrained Optimization (CO) module in GAUSS and quasi-maximum likelihood use the Constrained Maximum Likelihood (CML) module. In both cases constraints are imposed which restricts a_0, a_1 from zero and $0 < b < 1$. Quasi-maximum likelihood estimates are used to initialize a guess on v_3 and v_4 for the efficient GMM estimator.

Table 4.4 GARCH(1,1)-M estimates of daily returns to the SP500, standard errors in parentheses based on weighting matrix for GMM and Bollerslev-Wooldridge robust standard errors for QMLE

	GMM	QMLE
μ_0	0.003194	−0.00172
	(0.01586)	(0.01778)
μ_1	0.140646	0.140532
	(0.00807)	(0.01015)
μ_2	−0.05300	−0.04540
	(0.00851)	(0.00982)
μ_3	0.01657	0.01715
	(0.00846)	(0.01053)
μ_4	0.01200	0.00854
	(0.00840)	(0.01012)
μ_5	0.01410	0.01850
	(0.00808)	(0.01034)
δ	0.02773	0.05700
	(0.02196)	(0.02404)
a_0	0.00754	0.00781
	(0.00190)	(0.00103)
a_1	0.08912	0.09082
	(0.00546)	(0.00889)
b	0.90782	0.90652
	(0.00627)	(0.00730)
$Q(12)$	18.53	16.91
$Q(24)$	35.15	33.87
$Q^2(12)$	21.93	20.65
$Q^2(24)$	32.10	30.88

Table 4.5 Conditional moments of rescaled residuals

	GMM	QMLE
Mean	−0.01129	−0.03466
Variance	1.00531	0.99850
Skewness	−0.51937	−0.51623
Kurtosis	7.40892	7.41753

4.5 Final remarks

This essay has introduced a feasible alternative to the common quasi-maximum likelihood estimator and the semi-parametric estimator of GARCH models. It has been shown that the efficient GMM estimator is simple to compute and asymptotically efficient relative to quasi-maximum likelihood. Hence in practice there is little reason not to prefer it over the common quasi-maximum likelihood estimator. Compared to the semi-parametric estimator efficient GMM has the advantage of being simple to compute and specification testing is also straightforward. One expects that the efficient GMM estimator will find its use in applications.

Appendix A

Proofs

We first give a series of lemmas that will be useful in the proofs of the theorems. For a random variable X_t let X_T denote $T^{-1}\sum_{t=1}^{T} X_t$, $||X_t||_p$ the L^p-norm of X_t and by $||X_t||$ the ordinary Euclidean norm.

Define the unobserved variance process, which is obtained by extending the observed process into the infinite past history

$$h_t^{2u} = \frac{a_0}{1-b} + a_1 \sum_{k=0}^{\infty} b^k \varepsilon_{t-1-k}^2$$

Lumsdaine (1996) lemma 1 show that $\left|\frac{\partial h_t^{2u}}{\partial a_0} h_t^{-2u}\right|$, $\left|\frac{\partial h_t^{2u}}{\partial a_1} h_t^{-2u}\right|$ are naturally bounded. The lemma below deals with the term $\left|\frac{\partial h_t^{2u}}{\partial b} h_t^{-2u}\right|$ which is more difficult

Lemma A.1 $E\left|\frac{\partial h_t^{2u}}{\partial b} h_t^{-2u}\right|^q < \infty$ *uniformly in* $\theta \in \Theta$ for all $1 \leq q < \infty$

Proof. By Minkowskis inequality

$$\left|\left|\frac{\partial h_t^{2u}}{\partial b} h_t^{-2u}\right|\right|_q^q = \left|\left|\sum_{k=0}^{\infty} b^k h_{t-k-1}^{2u} h_t^{-2u}\right|\right|_q^q \leq \left(\sum_{k=0}^{\infty} b^k \left|\left|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right|\right|_q\right)^q$$

Write

$$h_t^{2u} = \frac{a_0}{1-b}\left(1-b^{k+1}\right) + a_1 \sum_{j=0}^{k} b^j \varepsilon_{t-j-1}^2 + b^{k+1} h_{t-k-1}^{2u}$$

and

$$\begin{aligned}
\frac{h_{t-k-1}^{2u}}{h_t^{2u}} &= \frac{h_{t-k-1}^{2u}}{\frac{a_0}{1-b}\left(1-b^{k+1}\right)+a_1\sum_{j=0}^{k} b^j \varepsilon_{t-j-1}^2 + b^{k+1}h_{t-k-1}^{2u}} \\
&= \left(\frac{1}{b^{k+1}}\right)\frac{b^{k+1}h_{t-k-1}^{2u}}{\frac{a_0}{1-b}\left(1-b^{k+1}\right)+a_1\sum_{j=0}^{k} b^j \varepsilon_{t-j-1}^2 + b^{k+1}h_{t-k-1}^{2u}} \\
&\leq \left(\frac{1}{b^{k+1}}\right)\frac{b^{k+1}h_{t-k-1}^{2u}}{\frac{a_0}{1-b}\left(1-b^{k+1}\right)+b^{k+1}h_{t-k-1}^{2u}}
\end{aligned}$$

We then have, denoting $v = \frac{a_0}{1-b}\left(b^{-k}-b\right)$ and for c_1 a strictly positive constant

$$\begin{aligned}
\left(\sum_{k=0}^{\infty} b^k \left\|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right\|_q\right)^q &\leq \left(\sum_{k=0}^{\infty}\left(E\left|\frac{1}{\frac{v}{h_{t-k-1}^{2u}}+b}\right|^q\right)^{1/q}\right)^q \\
&\leq \left(\frac{1}{b}\sum_{k=0}^{\infty}\left(E\left|\frac{h_{t-k-1}^{2u}}{v+h_{t-k-1}^{2u}}\right|\right)^{1/q}\right)^q \\
&\leq \left(\frac{1}{b}\sum_{k=0}^{\infty}\left(\frac{1}{c_1\left(b^{-k}-b\right)+1}\right)^{1/q}\right)^q
\end{aligned}$$

where the first inequality follows from above, the second since $0<b<1$ and $q \geq 1$ and the third by Jensens inequality. Taking the limit, as $q \to \infty$, of the last term raised to the power of $\frac{1}{q}$ shows that it is not uniformly bounded. But for $q < \infty$ we have

$$\left(\frac{1}{b}\sum_{k=0}^{\infty}\left(\frac{1}{c_1\left(b^{-k}-b\right)+1}\right)^{1/q}\right)^q < \infty$$

uniformly in $\theta \in \Theta$ and hence all finite-order moments exist ∎

The following lemma bounds the expectation of the ratio $\left|\frac{h_t^{2u}(\theta_0)}{h_t^{2u}(\theta)}\right|$ uniformly in θ

Lemma A.2 $E\left|\frac{h_{0t}^2}{h_t^{2u}}\right|^q < \infty$ *uniformly in* $\theta \in \Theta$ for all $1 \leq q < \infty$

Proof. We have for $b \geq \beta$

$$\left|\frac{h_{0t}^2}{h_t^{2u}}\right|^q \leq \left|\frac{\frac{\alpha_0}{1-\beta}}{\frac{a_0}{1-b}}+\frac{\alpha_1}{a_1}\right|^q$$

and for $b \leq \beta$

$$\begin{aligned}
\left|\left|\frac{h_{0t}^2}{h_t^{2u}}\right|\right|_q^q &\leq \left|\left|\frac{\frac{\alpha_0}{1-\beta}}{\frac{a_0}{1-b}} + \frac{\alpha_1 \sum_{k=0}^{\infty} \beta^k \varepsilon_{t-k-1}^2}{h_t^{2u}}\right|\right|_q^q \\
&\leq \left|\frac{\frac{\alpha_0}{1-\beta}}{\frac{a_0}{1-b}}\right|^q + \left|\frac{\alpha_1}{a_1}\right|^q \left|\left|\sum_{k=0}^{\infty}\left(\frac{\beta}{b}\right)^k \frac{b^k h_{t-k}^{2u}}{\frac{a_0}{1-b}\left(1-b^k\right)+b^k h_{t-k}^{2u}}\right|\right|_q^q
\end{aligned}$$

By a similar argument to that in lemma A.1

$$\begin{aligned}
&\left|\left|\sum_{k=0}^{\infty}\left(\frac{\beta}{b}\right)^k \frac{h_{t-k}^{2u}}{\frac{a_0}{1-b}\left(b^{-k}-1\right)+h_{t-k}^{2u}}\right|\right|_q^q \\
\leq &\left(\sum_{k=0}^{\infty}\left(E\left|\left(\frac{\beta}{b}\right)^k \frac{h_{t-k}^{2u}}{\frac{a_0}{1-b}\left(b^{-k}-1\right)+h_{t-k}^{2u}}\right|^q\right)^{1/q}\right)^q \\
\leq &\left(\sum_{k=0}^{\infty}\left(\frac{\beta}{b}\right)^k \left(E\left|\frac{1}{\frac{\frac{a_0}{1-b}\left(b^{-k}-1\right)}{h_{t-k}^{2u}}+1}\right|\right)^{1/q}\right)^q
\end{aligned}$$

and using Jensens inequality

$$\begin{aligned}
&\left(\sum_{k=0}^{\infty}\left(\frac{\beta}{b}\right)^k \left(E\left|\frac{x}{\frac{a_0}{1-b}\left(b^{-k}-1\right)+x}\right|\right)^{1/q}\right)^q \qquad \text{(A.1)} \\
\leq &\left(\sum_{k=0}^{\infty}\left(\left(\frac{\beta}{b}\right)^{kq} \frac{1}{c_2\left(b^{-k}-1\right)+1}\right)^{1/q}\right)^q \\
= &\left(\frac{\beta}{b}\sum_{k=0}^{\infty}\left(\frac{\beta^k}{c_2\left(1-b^k\right)+b^k}\right)^{1/q}\right)^q
\end{aligned}$$

where $c_2 > 0$ and the last term is convergent uniformly in $\theta \in \Theta$ for all $q < \infty$ ■

The lemma below concerns the convergence of the unobserved objective function, Q_T^u based on h_t^{2u} to the limiting objective function, Q

Lemma A.3 $\sup_{\theta \in \Theta} |Q_T^u - Q| \xrightarrow{p} 0$

Proof. Applying the triangle and Cauchy-Schwarz inequalities to $|Q_T^u - Q|$ as in Newey and McFadden (1994) theorem 2.1 we need to show uniform convergence of $||\mathbf{g}_T^u - \mathbf{g}||$ and $||\mathbf{\Lambda}_T^u - \mathbf{\Lambda}||$ to zero. The method chosen here is to first establish a law of large numbers for all $\theta \in \Theta$. Uniform convergence (and continuity of the limiting function) will follow if we can establish stochastic equicontinuity, see Andrews (1992, theorem 3 (a)). Under assumption (c) Nelson (1990) show that h_{0t}^2 is strictly stationary (and covariance-stationary) and ergodic, hence $\mathbf{g}_T^u$ and $\mathbf{\Lambda}_T^u$ are strictly stationary and ergodic since they are measurable functions of h_{0t}.

Note that (ignoring some constants)

$$|g_{it}^u| \leq \left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right| + \left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\left(\frac{\varepsilon_t}{h_t^u}\right)\right| + \left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)\right|$$

From lemma A.1 we have for the first term

$$E\left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right| < \infty$$

Applying lemma A.1 and A.2 to the second and third terms respectively, using Hölders inequality and the independence of z_t we have $E\left|g_{it}^u\right| < \infty$.

Consider next the cross-product

$$\begin{aligned}
\left|g_{it}^u g_{jt}^{u\prime}\right| \leq\; & \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)' h_t^{-4u}\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)^2\right| \\
& +\left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)' h_t^{-4u}\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)\right| \\
& +\left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)'\right| + 2\left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)' h_t^{-4u}\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)^{3/2}\right| \\
& +2\left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j}\right)' h_t^{-4u}\left(\frac{\varepsilon_t}{h_t^{u}}\right)\right|
\end{aligned}$$

and again using lemma A.1, A.2, Hölders inequality and the independence of z_t gives $E\left|g_{it}^u g_{jt}^{u\prime}\right| < \infty$. This establishes a pointwise law of large numbers. To show convergence is uniform and continuity of the limiting function we establish that (a) $\sup_{\theta\in\Theta} E\left|\frac{\partial g_{it}^u}{\partial \theta_j}\right| < \infty$ and (b) $\sup_{\theta\in\Theta} E\left|\frac{\partial \mathbf{\Lambda}_{ijt}^u}{\partial \theta_k}\right| < \infty$ for all

i, j and k. Consider (a)

$$\begin{aligned}
\left|\frac{\partial g_{it}^u}{\partial \theta_j}\right| &\leq \left|\frac{\partial^2 h_t^{2u}}{\partial \theta_i \partial \theta_j} h_t^{-2u} \left(\frac{\varepsilon_t}{h_t^u}\right)\right| + \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)\left(\frac{\varepsilon_t}{h_t^u}\right)\right| \\
&+ \left|\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u} \frac{\partial h_t^u}{\partial \theta_j} h_t^{-u}\left(\frac{\varepsilon_t}{h_t^u}\right)\right| + \left|\frac{\partial^2 h_t^{2u}}{\partial \theta_i \partial \theta_j} h_t^{-2u}\left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right| \\
&+ \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)\left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right| \\
&+ \left|\left(\frac{\partial h_t^{2u}}{\partial \theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial \theta_j} h_t^{-2u}\right)\left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)\right|
\end{aligned}$$

and hence we need to find convenient expressions for the second derivative $\left|\frac{\partial^2 h_t^{2u}}{\partial \theta_i \partial \theta_j} h_t^{-2u}\right|$. Differentiating h_t^{2u} twice

$$\begin{aligned}
\frac{\partial h_t^{2u}}{\partial a_0 \partial b} h_t^{-2u} &= \sum_{k=0}^{\infty} b^k \frac{\partial h_{t-k-1}^{2u}}{\partial a_0} h_t^{-2u} \leq \frac{1}{1-b_u} \frac{1-b_l}{a_{0l}} \\
\frac{\partial h_t^{2u}}{\partial a_1 \partial b} h_t^{-2u} &= \sum_{k=0}^{\infty} b^k \frac{\partial h_{t-k-1}^{2u}}{\partial a_1} h_t^{-2u} \leq \sum_{k=0}^{\infty} b^k \varepsilon_{t-k-2}^2 h_t^{-2u} \\
&\leq \frac{1}{a_{1l}} \sum_{k=0}^{\infty} b^k h_{t-k-1}^{2u} h_t^{-2u}
\end{aligned}$$

where we used that $\varepsilon_{t-k-2}^2 \leq \frac{1}{a_{1l}} h_{t-k-1}^{2u}$. For the most demanding derivative with respect to b

$$\begin{aligned}
\frac{\partial^2 h_t^{2u}}{\partial b \partial b} h_t^{-2u} &= 2\sum_{k=0}^{\infty} b^k \frac{\partial h_{t-k-1}^{2u}}{\partial b} h_t^{-2u} = 2\sum_{k=0}^{\infty} b^k \left(\sum_{k=0}^{\infty} b^k \frac{h_{t-k-2}^{2u}}{h_{t-k-1}^{2u}}\right) \frac{h_{t-k-1}^{2u}}{h_t^{2u}} \\
&\leq 2\left(\sum_{k=0}^{\infty} b^k \left|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right|\right)^2
\end{aligned}$$

which shows that $\sup_{\theta \in \Theta} E\left|\frac{\partial g_{it}^u}{\partial \theta_j}\right| < \infty$ for all i, j by application of lemma A.1

and A.2. Next for (b) we have

$$\begin{aligned}
\frac{\partial\Lambda^u_{ijt}}{\partial\theta_k} &\leq \left|\frac{\partial h^{2u}_t}{\partial\theta_i}h_t^{-2u}\frac{\partial^2 h^{2u}_t}{\partial\theta_j\partial\theta_k}h_t^{-2u}\left(\frac{\varepsilon_t}{h^u_t}-(\frac{\varepsilon^2_t}{h^{2u}_t}-1)\right)^2\right| \\
&+\left|\left(\frac{\partial h^{2u}_t}{\partial\theta_i}h_t^{-2u}\right)\left(\frac{\partial h^{2u}_t}{\partial\theta_j}h_t^{-2u}\right)\left(\frac{\partial h^{2u}_t}{\partial\theta_k}h_t^{-2u}\right)\left(\frac{\varepsilon_t}{h^u_t}-(\frac{\varepsilon^2_t}{h^{2u}_t}-1)\right)^2\right| \\
&+\left|\begin{array}{c}\left(\frac{\partial h^{2u}_t}{\partial\theta_i}h_t^{-2u}\right)\left(\frac{\partial h^{2u}_t}{\partial\theta_j}h_t^{-2u}\right)\frac{\partial h^{u}_t}{\partial\theta_k}h_t^{-u} \\ \left(\frac{\varepsilon_t}{h^u_t}\right)\left(\left(\frac{\varepsilon_t}{h^u_t}\right)-(\frac{\varepsilon^2_t}{h^{2u}_t}-1)\right)\end{array}\right| \\
&+\left|\begin{array}{c}\left(\frac{\partial h^{2u}_t}{\partial\theta_i}h_t^{-2u}\right)\left(\frac{\partial h^{2u}_t}{\partial\theta_j}h_t^{-2u}\right)\left(\frac{\partial h^{2u}_t}{\partial\theta_k}h_t^{-2u}\right) \\ \left(\frac{\varepsilon^2_t}{h^{2u}_t}\right)\left(\left(\frac{\varepsilon_t}{h^u_t}\right)-(\frac{\varepsilon^2_t}{h^{2u}_t}-1)\right)\end{array}\right|
\end{aligned}$$

and we need to consider $\left|\frac{\partial h^{2u}_t}{\partial\theta_i}h_t^{-2u}\frac{\partial^2 h^{2u}_t}{\partial\theta_j\partial\theta_k}h_t^{-2u}\right|$. For the most demanding derivative

$$\left|\frac{\partial h^{2u}_t}{\partial b}h_t^{-2u}\frac{\partial^2 h^{2u}_t}{\partial b^2}h_t^{-2u}\right| \leq 2\left(\sum_{k=0}^{\infty} b^k\left|\frac{h^{2u}_{t-k-1}}{h^{2u}_t}\right|\right)^3$$

such that $\sup_{\theta\in\Theta} E\left|\frac{\partial\mathbf{\Lambda}^u_{ijt}}{\partial\theta_k}\right| < \infty$ for all i, j and k. Hence the sequences $||\mathbf{g}^u_T - \mathbf{g}||$, $||\mathbf{\Lambda}^u_T - \mathbf{\Lambda}||$ are stochastically equicontinuous ■

The next and final lemma is concerned with the convergence of the unobserved objective function, Q^u_T to the corresponding observed objective function, Q_T

Lemma A.4 $\sup_{\theta\in\Theta}|Q^u_T - Q_T| \xrightarrow{p} 0$

Proof. By the same argument as in lemma A.3 we need to show that $\sup_{\theta\in\Theta}||\mathbf{g}^u_T - \mathbf{g}_T|| \xrightarrow{p} 0$ and $\sup_{\theta\in\Theta}||\mathbf{\Lambda}^u_T - \mathbf{\Lambda}_T|| \xrightarrow{p} 0$. First we observe some properties of the conditional variance process

$$\begin{aligned}
h^{2u}_t &= h^2_t + b^{t-1}\left(a_1\sum_{k=0}^{\infty} b^k\varepsilon^2_{-k} + a_0\sum_{k=0}^{\infty} b^k\right) \\
&= h^2_t + b^{t-1}h_1
\end{aligned}$$

where $h_1 = h_1(\theta)$ is the initial condition. Next

$$E\sup_{\theta\in\Theta} h^{2u}_t \leq \frac{a_{0u}}{1-b_u} + a_{1u}\sum_{k=0}^{\infty} b^k_u E\sup_{\theta\in\Theta}\varepsilon^2_{t-1-k} < \infty$$

since $E \sup_{\theta\in\Theta} \varepsilon_t^2 < \infty$ under assumption (c) and does not depend on t, which in turn imply that $\sup_{\theta\in\Theta} \varepsilon_t^2 < \infty$. We also have

$$E \sup_{\theta\in\Theta} \frac{\varepsilon_t^2}{h_t^{2u}} \leq \frac{1-b_l}{a_{0l}} E \sup_{\theta\in\Theta} \varepsilon_t^2 < \infty$$

For some arbitrary i, j

$$\begin{aligned}\left| T^{-1} \sum_{t=1}^{T} \left(\Lambda_{ijt}^u - \Lambda_{ijt}\right) \right| \quad \leq \quad & T^{-1} \sum_{t=1}^{T} \left[|g_{it}^u| + |g_{it}| \right] \\ & \left\{ \begin{array}{c} \left| \begin{array}{c} \frac{\partial h_t^2}{\partial \theta_j} \frac{\varepsilon_t}{h_t} - \frac{\partial h_t^{2u}}{\partial \theta_j} \frac{\varepsilon_t}{h_t^u} - \frac{\partial h_t^2}{\partial \theta_j} \frac{\varepsilon_t^2}{h_t^2} - \\ \frac{\partial h_t^{2u}}{\partial \theta_j} \frac{\varepsilon_t^2}{h_t^{2u}} + \frac{\partial h_t^2}{\partial \theta_j} - \frac{\partial h_t^{2u}}{\partial \theta_j} \end{array} \right| h_t^{-2} - \\ \left| \frac{\partial h_t^{2u}}{\partial \theta_j} \left(\frac{\varepsilon_t}{h_t^u} - \left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1 \right) \right) \right| \left| h_t^{-2u} - h_t^{-2} \right| \end{array} \right\}\end{aligned}$$

since $\sup_{\theta\in\Theta} |g_{it}^u|$, $\sup_{\theta\in\Theta} |g_{it}|$ are finite and $h_t^{-2} \leq \frac{1-b_l}{a_{0l}}$, we must show that

$$(i)\ T^{-1} \sum_{t=1}^{T} \left| \frac{\partial h_t^2}{\partial \theta_j} \left(\frac{\varepsilon_t^r}{h_t^r} \right) - \frac{\partial h_t^{2u}}{\partial \theta_j} \left(\frac{\varepsilon_t^r}{h_t^{ru}} \right) \right| \xrightarrow{p} 0$$

$$(ii)\ T^{-1} \sum_{t=1}^{T} \left| \frac{\partial h_t^{2u}}{\partial \theta_j} \left(\frac{\varepsilon_t}{h_t^u} - \left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1 \right) \right) \right| \left| h_t^{-2u} - h_t^{-2} \right| \xrightarrow{p} 0$$

uniformly in θ for $r = 0, 1$ or 2. For (i) we have

$$\left| \frac{\partial h_t^2}{\partial \theta_j} \left(\frac{\varepsilon_t^r}{h_t^r} \right) - \frac{\partial h_t^{2u}}{\partial \theta_j} \left(\frac{\varepsilon_t^r}{h_t^{ru}} \right) \right| \leq \left| \begin{array}{c} \frac{\partial h_t^2}{\partial \theta_j} \left[\left(\frac{\varepsilon_t^r}{h_t^{ru}} \right) - \left(\frac{\varepsilon_t^r}{h_t^r} \right) \right] \\ + \frac{\partial}{\partial \theta_j} b^{t-1} h_1 \left(\frac{\varepsilon_t^r}{h_t^{ru}} \right) \end{array} \right|$$

By lemma 3 of Lee and Hansen (1994)

$$T^{-1} \sum_{t=1}^{T} \left[\left(\frac{\varepsilon_t^2}{h_t^{2u}} \right) - \left(\frac{\varepsilon_t^2}{h_t^2} \right) \right] \xrightarrow{p} 0$$

uniformly in θ, hence for $r = 1$ and trivially for $r = 0$. Note that

$$\sup_{\theta\in\Theta} \left\| \frac{\partial h_t^2}{\partial \theta} \right\| < \infty$$

since

$$
\begin{aligned}
\frac{\partial h_t^2}{\partial a_0} &= 1 + b\frac{\partial h_{t-1}^2}{\partial a_0} \leq \sum_{k=0}^{\infty} b^k \leq \frac{1}{1-b_u} \\
\frac{\partial h_t^2}{\partial a_1} &= \varepsilon_{t-1}^2 + b\frac{\partial h_{t-1}^2}{\partial a_0} \leq \sum_{i=0}^{\infty} b^k \varepsilon_{t-k-1}^2 \\
\frac{\partial h_t^2}{\partial b} &= h_{t-1}^2 + b\frac{\partial h_{t-1}^2}{\partial b} \leq \sum_{k=0}^{\infty} b^k h_{t-k-1}^2
\end{aligned}
$$

Furthermore $\sup_{\theta\in\Theta}\left(\frac{\varepsilon_t^r}{h_t^{ru}}\right) < \infty$ holds from above, and

$$
\begin{aligned}
\sum_{t=1}^{T}\left|\frac{\partial}{\partial\theta_i} b^{t-1} h_1\right| &\leq \sum_{t=1}^{T} b^{t-1}\left|\frac{\partial h_1}{\partial\theta_i}\right| + \sum_{t=1}^{T}(t-1)b^{t-2}h_1 \\
&\leq \frac{1}{1-b_u}\left|\frac{\partial h_1}{\partial\theta_i}\right| + \frac{b_u}{(1-b_u)^2}h_1
\end{aligned}
$$

hence

$$
T^{-1}\left[\frac{1}{1-b_u}\left|\frac{\partial h_1}{\partial\theta_i}\right| + \frac{b_u}{(1-b_u)^2}h_1\right] \xrightarrow{p} 0
$$

which shows (i). Next for (ii) $T^{-1}\sum_{t=1}^{T}\left|h_t^{-2u} - h_t^{-2}\right| \xrightarrow{p} 0$ uniformly in θ by lemma 6 (c) of Lumsdaine (1996), and

$$
\left|\frac{\partial h_t^{2u}}{\partial\theta_j}\left(\frac{\varepsilon_t}{h_t^u} - \left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right)\right|
$$

is bounded in probability uniformly in θ which shows (ii). Hence $\sup_{\theta\in\Theta}$ $||\mathbf{\Lambda}_T^u - \mathbf{\Lambda}_T|| \xrightarrow{p} 0$ holds. Finally $||\mathbf{g}_T^u - \mathbf{g}_T|| \xrightarrow{p} 0$ uniformly in θ follows from (i) above ■

We are now ready to give the proofs of the theorems in the text.

Proof theorem 1. By theorem 2.1 (Consistency theorem for extremum estimators) in Newey and McFadden (1994) we need to establish that (i) $\sup_{\theta\in\Theta}|Q_T - Q| \xrightarrow{p} 0$, (ii) Q is continuous, (iii) Θ is compact (which holds by assumption) and (iv) Q is uniquely minimized at θ_0. By the triangle inequality

$$
\sup_{\theta\in\Theta}|Q_T - Q| \leq \sup_{\theta\in\Theta}|Q_T^u - Q| + \sup_{\theta\in\Theta}|Q_T - Q_T^u|
$$

where $\sup_{\theta\in\Theta}|Q_T^u - Q| \to 0$ was established in lemma A.3 and $\sup_{\theta\in\Theta}|Q_T - Q_T^u| \to 0$ was established in lemma A.4. This proves (i) and (ii) follows from lemma A.3 and continuity of multiplication. To finally show (iv) we note that Lumsdaine (1996) lemma 5 prove that $E\left[\frac{\partial h_t^{2u}}{\partial\theta}\frac{\partial h_t^{2u}}{\partial\theta'}h_t^{-4u}\right]$ is a positive definite matrix for all $\theta \in \Theta$. It follows that $E\mathbf{\Lambda}_t^u$ is a positive definite matrix for $v_4 - v_3^2 > 1$, since

$$E\mathbf{\Lambda}_t^u = \frac{1}{(v_4-1)-v_3^2}E\frac{\partial h_t^{2u}}{\partial\theta}\frac{\partial h_t^{2u}}{\partial\theta'}h_t^{-4u}$$

Identification holds if θ_0 is unique solution to $\lim_{T\to\infty} E\mathbf{g}_T^u = 0$. By application of dominated convergence theorem $\lim_{T\to\infty} E\mathbf{g}_T^u = E\mathbf{g}_0$, where $\mathbf{g}_0 = \mathbf{g}(\theta_0)$ and since $E\left(\frac{\varepsilon_t}{h_{0t}}v_3 - \left(\frac{\varepsilon_t^2}{h_{0t}^2} - 1\right)\right) = 0$ a consistent root exists at θ_0. In addition this root exists for finite arbitrary initial guess on v_3, v_4 ■

Proof theorem 2. First we note that the fact that a root exists at θ_0 for finite arbitrary guess on v_3, v_4 ensures we can apply theorem 6.2 of Newey and McFadden (1994) to obtain that the asymptotic distribution of the estimator is independent of the guess on v_3, v_4.

Consider the gradient of the efficient GMM objective function

$$\begin{aligned}\frac{\partial}{\partial\theta}Q_T &= \frac{\partial \mathbf{g}_T'}{\partial\theta}\left(\mathbf{\Lambda}_T^{-1}+\mathbf{\Lambda}_T^{-1}\right)\mathbf{g}_T + \left[\left(\mathbf{g}_T'\otimes\mathbf{g}_T'\right)\frac{\partial \operatorname{vec}\mathbf{\Lambda}_T^{-1}}{\partial\theta'}\right]' \\ &= 2\mathbf{G}_T'\mathbf{\Lambda}_T^{-1}\mathbf{g}_T - \left[\left(\mathbf{g}_T'\otimes\mathbf{g}_T'\right)\left(\mathbf{\Lambda}_T^{-1}\otimes\mathbf{\Lambda}_T^{-1}\right)\frac{\partial \operatorname{vec}\mathbf{\Lambda}_T}{\partial\theta'}\right]'\end{aligned} \tag{A.2}$$

with $\mathbf{G}_T = T^{-1}\sum_{t=1}^T \frac{\partial \mathbf{g}_t}{\partial\theta'}$, and the second derivative

$$\begin{aligned}\frac{\partial Q_T^2(\theta)}{\partial\theta\partial\theta'} &= \frac{\partial \mathbf{g}_T'}{\partial\theta}\left(\mathbf{\Lambda}_T^{-1}+\mathbf{\Lambda}_T^{-1}\right)\frac{\partial \mathbf{g}_T}{\partial\theta'} \\ &\quad + \left[\mathbf{g}_T'\left(\mathbf{\Lambda}_T^{-1}+\mathbf{\Lambda}_T^{-1}\right)\otimes\mathbf{I}_p\right]\frac{\partial}{\partial\theta'}\left[\operatorname{vec}\left(\frac{\partial \mathbf{g}_T'}{\partial\theta}\right)\right] + o_p(1) \\ &= 2\mathbf{G}_T'\mathbf{\Lambda}_T^{-1}\mathbf{G}_T \\ &\quad + \left[\mathbf{g}_T'\left(\mathbf{\Lambda}_T^{-1}+\mathbf{\Lambda}_T^{-1}\right)\otimes\mathbf{I}_p\right]\frac{\partial \operatorname{vec}\mathbf{G}_T'}{\partial\theta'} + o_p(1)\end{aligned} \tag{A.3}$$

where the $o_p(1)$ term in (A.3)comes from the derivative of the second term in the gradient. It assumes that $\frac{\partial}{\partial\theta}\left[\frac{\partial \operatorname{vec}\mathbf{\Lambda}_T}{\partial\theta'}\right]' = o_p(T)$ uniformly in $\theta \in \Theta$ which can be shown. The proof is complete if we can show that the last terms of (A.2, A.3) are negligible asymptotically and the conditions of theorem 3.2 of Newey and McFadden (1994) holds (Asymptotic normality of minimum distance estimators). Asymptotic normality holds (given consistency) if (i)

$\sup_{\theta\in\Theta}||\mathbf{\Lambda}_T - \mathbf{\Lambda}|| \xrightarrow{p} 0$ and $\mathbf{\Lambda}$ is non-singular (which holds from above) (ii) $\sup_{\theta\in\Theta}\left|\left|\frac{\partial \mathbf{g}_T^u}{\partial\theta'} - \mathbf{G}\right|\right| \xrightarrow{p} 0$ and $\mathbf{G} = E\frac{\partial \mathbf{g}_T^u}{\partial\theta'}$ is non-singular (iii) $\theta_0 \in int(\Theta)$ (which holds by assumption) and (iv) asymptotic normality of $\sqrt{T}\mathbf{g}_T$.

To prove (ii) note that we have already shown that $\left|\left|\frac{\partial \mathbf{g}_T^u}{\partial\theta'} - \mathbf{G}\right|\right| \xrightarrow{p} 0$ in the proof of theorem 1. As in lemma A.3 uniform convergence and continuity of the limiting function follows if $\sup_{\theta\in\Theta} E\left|\frac{\partial \mathbf{G}_{ijt}^u}{\partial\theta_k}\right| < \infty$ for all i, j and k. Differentiating $\frac{\partial g_{it}^u}{\partial\theta_j}$ once more

$$
\begin{aligned}
\left|\frac{\partial^2 g_{it}^u}{\partial\theta_j\partial\theta_k}\right| \leq\ & \left|\frac{\partial^3 h_t^{2u}}{\partial\theta_i\partial\theta_j\partial\theta_k} h_t^{-2u}\left(\left(\frac{\varepsilon_t}{h_t^u}\right) - \left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right)\right| \\
& + \left| \begin{array}{c} \frac{\partial^2 h_t^{2u}}{\partial\theta_i\partial\theta_j} h_t^{-2u} \frac{\partial h_t^{2u}}{\partial\theta_k} h_t^{-2u} \\ \left[3\left(\left(\frac{\varepsilon_t}{h_t^u}\right) - \left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right)\right) - \left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)^{3/2}\right] \end{array} \right| \\
& + \left| \begin{array}{c} \left[\frac{\partial^2 h_t^{2u}}{\partial\theta_i\partial\theta_j} h_t^{-2u} \frac{\partial h_t^u}{\partial\theta_k} h_t^{-u} - \left(\frac{\partial h_t^{2u}}{\partial\theta_j} h_t^{-2u}\right)^2 \frac{\partial h_t^u}{\partial\theta_k} h_t^{-u}\right] \\ \left(\frac{\varepsilon_t}{h_t^u}\right)\left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right) \end{array} \right| \\
& + \left|2\left(\frac{\partial h_t^{2u}}{\partial\theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_j} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_k} h_t^{-2u}\right)\left(\frac{\varepsilon_t}{h_t^u}\right)\right| \\
& + \left| \begin{array}{c} 2\left(\frac{\partial h_t^{2u}}{\partial\theta_i} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_j} h_t^{-2u}\right)\left(\frac{\partial h_t^{2u}}{\partial\theta_k} h_t^{-2u}\right) \\ \left(-\left(\frac{\varepsilon_t^2}{h_t^{2u}} - 1\right) - \left(\frac{\varepsilon_t^2}{h_t^{2u}}\right)^{3/2}\right) \end{array} \right|
\end{aligned}
$$

and in view of lemma A.1 and A.2 we need to consider the third derivative of h_t^{2u}. For the derivative with respect to b we have

$$
\frac{\partial^3 h_t^{2u}}{\partial b^3} h_t^{-2u} \leq 3\sum_{k=0}^{\infty} b^k \frac{\partial^2 h_{t-k-1}^{2u}}{\partial b^2} h_t^{-2u} \leq 6\left(\sum_{k=0}^{\infty} b^k \left|\frac{h_{t-k-1}^{2u}}{h_t^{2u}}\right|\right)^3
$$

and hence applying lemma A.1 and A.2 using Hölders inequality and the independence of z_t we conclude that convergence is uniform and $\mathbf{G}$ is continuous on Θ. Then $\sup_{\theta\in\Theta}||\mathbf{G}_T - \mathbf{G}|| \xrightarrow{p} 0$ holds if $\sup_{\theta\in\Theta}||\mathbf{G}_T^u - \mathbf{G}_T|| \to 0$ which can be shown. It is straightforward to verify that $E\mathbf{G}_{ijt}(\theta_0) = E\mathbf{\Lambda}_{ijt}(\theta_0)$ for all i, j which shows the nonsingularity of $\mathbf{G}$ as well. Next (iv) follows since $\mathbf{g}_t$ is a stationary ergodic martingale difference sequence with finite variance and hence $T^{\frac{1}{2}}\mathbf{g}_T \xrightarrow{d} N(\mathbf{0}, \mathbf{\Lambda})$. It remains to show that the last terms in (A.2, A.3) are unimportant for the asymptotic distribution. We have from (A.2)

$$\frac{\partial}{\partial\theta}Q_T = 2\mathbf{G}_T'\mathbf{\Lambda}_T^{-1}\mathbf{g}_T - \left[\left(\mathbf{g}_T' \otimes \mathbf{g}_T'\right)\left(\mathbf{\Lambda}_T^{-1} \otimes \mathbf{\Lambda}_T^{-1}\right)\frac{\partial \operatorname{vec} \mathbf{\Lambda}_T}{\partial\theta'}\right]'$$

and scaling by $T^{1/2}$, since $T^{1/2}\mathbf{g}_T$ is bounded in probability, $\sup_{\theta\in\Theta}||\mathbf{\Lambda}_T - \mathbf{\Lambda}|| \xrightarrow{p} 0$ and $\frac{\partial \operatorname{vec} \mathbf{\Lambda}_T}{\partial\theta'} = O_p(1)$ uniformly in $\theta \in \Theta$ we have the result. A similar argument applied to the second term of (A.3) establishes the result here as well since $\frac{\partial \operatorname{vec} \mathbf{G}_T}{\partial\theta'} = O_p(1)$ uniformly in $\theta \in \Theta$. Applying a standard mean-value expansion of the gradient vector as in Newey and McFadden (1994) theorem 3.2 then obtains the distributional result given in theorem 2 ■

Bibliography

Andrews, D. W. K. (1992), 'Generic uniform convergence', *Econometric Theory* **8**, 241–257.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. and Wooldridge, J. M. (1992), 'Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometric Reviews* **11**, 143–172.

Box, G. E. P. and Pierce, D. A. (1970), 'Distribution of residual autocorrelation in autoregressive-integrated moving average time series models', *Journal of The American Statistical Association* **65**, 1509–1526.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation', *Econometrica* **50**, 987–1007.

Engle, R. F. and Gonzales-Rivera, G. (1991), 'Semiparametric ARCH models', *Journal of Business and Economic Statistics* **9/4**, 345–359.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987), 'Estimating time-varying risk premia in the term structure', *Econometrica* **55/2**, 391–407.

French, K. R., Schwert, W. G. and Staumbaugh, R. F. (1987), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**, 391–407.

Glosten, L., Jagannathan, R. and Runkle, D. (1993), 'On the relation between expected values and the volatility of the nominal excess return on stocks', *Journal of Finance* **48**, 1779–1801.

Gonzalez-Rivera, G. and Drost, F. C. (1999), 'Efficiency comparisons of maximum-likelihood based estimators in GARCH models', *Journal of Econometrics* **93**, 93–111.

Granger, C. W. J. and Ding, Z. (1995), 'Some properties of the absolute returns: An alternative measurement of risk', *Annales d'Economie et de Statistique* **40**, 67–91.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moment estimators', *Econometrica* **50**, 1029–1054.

Lee, S. W. and Hansen, B. (1994), 'Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator', *Econometric Theory* **10**, 29–52.

Lumsdaine, R. L. (1996), 'Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models', *Econometrica* **64**, 575–596.

Mills, T. C. (1999), *The Econometric Modelling of Financial Time Series*, Cambridge University Press.

Nelson, D. D. (1990), 'Stationarity and persistence in the GARCH(1,1) models', *Econometric Theory* **6**, 318–334.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle and D. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Amsterdam North Holland, chapter 36, pp. 2111–2245.

Newey, W. K. and West, K. (1987), 'Hypothesis testing with efficient method of moments estimation', *International Economic Review* **28**, 777–787.

EFI
The Economic Research Institute

Reports since 1995
A complete publication list can be found at www.hhs.se/efi

Published in the language indicated by the title

2001

Hill, M., Essays on Environmental Policy Analysis: Computable General Equilibrium Approaches Applied to Sweden.
Lange, F. och Wahlund, R., Category Management – När konsumenten är manager.

2000

Berg-Suurwee, U., Styrning före och efter stadsdelsnämndsreform inom kultur och fritid – Resultat från intervjuer och enkät.
Bergkvist, L., Advertising Effectiveness Measurement: Intermediate Constructs and Measures.
Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L., Koncernchefen och ägarna.
Bornefalk, A., Essays on Social Conflict and Reform.
Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform.
Edman, J., Information Use and Decision Making in Groups.
Emling, E., Svenskt familjeföretagande.
Ericson, M., Strategi, kalkyl, känsla.
Gunnarsson, J., Wahlund, R., Flink, H., Finansiella strategier i förändring: segment och beteenden bland svenska hushåll.
Hellman, N., Investor Behaviour – An Empirical Study of How Large Swedish Institutional Investors Make Equity Investment Decisions.
Hyll, M., Essays on the Term Structure of Interest Rates.
Håkansson, P., Beyond Private Label – The Strategic View on Distributor Own Brands.
I huvudet på kunden. Söderlund, M., (red).
Karlsson Stider, A., Familjen och firman.
Ljunggren, U., Styrning av grundskolan i Stockholms stad före och efter stadsdelsnämndsreformen – Resultat från intervjuer och enkät.
Ludvigsen, J., The International Networking between European Logistical Operators.
Nittmar, H., Produktutveckling i samarbete – Strukturförändring vid införande av nya Informationssystem.
Robertsson, G., International Portfolio Choice and Trading Behavior.
Schwarz, B., Weinberg, S., Serviceproduktion och kostnader – att söka orsaker till kommunala skillnader.
Stenström, E., Konstiga företag.
Styrning av team och processer – Teoretiska perspektiv och fallstudier. Bengtsson, L., Lind, J., Samuelson, L.A., (red).
Sweet, S., Industrial Change Towards Environmental Sustainability - The Case of Replacing Chloroflouorocarbons.
Tamm Hallström, K., Kampen för auktoritet – Standardiseringsorganisationer i arbete.

1999

Adler, N., Managing Complex Product Development.
Allgulin, M., Supervision and Monetary Incentives.
Andersson, P., Experto Credite: Three Papers on Experienced Decision Makers.
Ekman, G., Från text till batong – Om poliser, busar och svennar.
Eliasson, A-C., Smooth Transitions in Macroeconomic Relationships.
Flink, H., Gunnarsson, J., Wahlund, R., Svenska hushållens sparande och skuldsättning– ett konsumentbeteende-perspektiv.
Gunnarsson, J., Portfolio-Based Segmentation and Consumer Behavior: Empirical Evidence and Methodological Issues.
Hamrefors, S., Spontaneous Environmental Scanning.
Helgesson, C-F., Making a Natural Monopoly: The Configuration of a Techno-Economic Order in Swedish Telecommunications.
Japanese Production Management in Sunrise or Sunset. Karlsson, C., (red).
Jönsson, B., Jönsson, L., Kobelt, G., Modelling Disease Progression And the Effect of Treatment in Secondary Progressive MS. Research Report.
Lindé, J., Essays on the Effects of Fiscal and Monetary Policy.
Ljunggren, U., Indikatorer i grundskolan i Stockholms stad före stadsdels-nämndsreformen – en kartläggning.
Ljunggren, U., En utvärdering av metoder för att mäta produktivitet och effektivitet i skolan – Med tillämpning i Stockholms stads grundskolor.
Lundbergh, S., Modelling Economic High-Frequency Time Series.
Mägi, A., Store Loyalty? An Empirical Study of Grocery Shopping.
Mölleryd, B.G., Entrepreneurship in Technological Systems – the Development of Mobile Telephony in Sweden.
Nilsson, K., Ledtider för ledningsinformation.
Osynlig Företagsledning. Sjöstrand, S-E., Sandberg, J., Tyrstrup, M., (red).
Rognes, J., Telecommuting – Organisational Impact of Home Based – Telecommuting.
Sandström, M., Evaluating the Benefits and Effectiveness of Public Policy.
Skalin, J., Modelling Macroeconomic Time Series with Smooth Transition Autoregressions.
Spagnolo, G., Essays on Managerial Incentives and Product-Market Competition.
Strauss, T., Governance and Structural Adjustment Programs: Effects on Investment, Growth and Income Distribution.
Svedberg Nilsson, K., Effektiva företag? En studie av hur privatiserade organisationer konstrueras.
Söderström, U., Monetary Policy under Uncertainty.
Werr, A., The Language of Change The Roles of Methods in the Work of Management Consultants.
Wijkström, F., Svenskt organisationsliv – Framväxten av en ideell sektor.

1998

Andersson, M., On Testing and Forecasting in Fractionally Integrated Time Series Models.
Berg-Suurwee, U., Styrning av kultur- och fritidsförvaltning innan stadsdelsnämnds-reformen.
Berg-Suurwee, U., Nyckeltal avseende kultur- och fritidsförvaltning innan stadsdels-nämndsreformen.
Bergström, F., Essays on the Political Economy of Industrial Policy.
Bild, M., Valuation of Takeovers.
Charpentier, C., Samuelson, L.A., Effekter av en sjukvårdsreform – en analys av Stockholmsmodellen.

Eriksson-Skoog, G., The Soft Budget Constraint: The Emergence, Persistence and Logic of an Institution. The Case of Tanzania 1967-1992.
Gredenhoff, M., Bootstrap Inference in Time Series Econometrics.
Ioannidis, D., I nationens tjänst? Strategisk handling i politisk miljö – en nationell teleoperatörs interorganisatoriska, strategiska utveckling.
Johansson, S., Savings Investment, and Economic Reforms in Developing Countries.
Levin, J., Essays in Company Valuation.
Ljunggren, U., Styrning av grundskolan i Stockholms stad innan stadsdelsnämnds-reformen.
Mattsson, S., Från stat till marknad – effekter på nätverksrelationer vid en bolagise-ringsreform.
Nyberg, A., Innovation in Distribution Channels – An Evolutionary Approach.
Olsson, P., Studies in Company Valuation.
Reneby, J., Pricing Corporate Securities.
Roszbach, K., Essays on Banking Credit and Interest Rates.
Runsten, M., The Association Between Accounting Information and Stock Prices. Model development and empirical tests based on Swedish Data.
Segendorff, B., Essays on Bargaining and Delegation.
Sjöberg, L., Bagozzi, R., Ingvar, D.H., Will and Economic Behavior.
Sjögren, A., Perspectives on Human Capital: Economic Growth, Occupational Choice and Intergenerational Mobility.
Studier i kostnadsintäktsanalys. Jennergren, P., (red)
Söderholm, J., Målstyrning av decentraliserade organisationer. Styrning mot finansiella och icke-finansiella mål.
Thorburn, K., Cash Auction Bankruptcy and Corporate Restructuring
Wijkström, F., Different Faces of Civil Society.
Zethraeus, N., Essays on Economic Evaluation in Health Care. Evaluation of Hormone Replacement Therapy and Uncertainty in Economic Evaluations.

1997

Alexius, A., Essays on Exchange Rates, Prices and Interest Rates.
Andersson, B., Essays on the Swedish Electricity Market.
Berggren, N., Essays in Constitutional Economics.
Ericsson, J., Credit Risk in Corporate Securities and Derivatives. Valuation and Optimal Capital Structure Choice.
Charpentier, C., Budgeteringens roller, aktörer och effekter. En studie av budget-processerna i en offentlig organisation.
De Geer, H., Silfverberg, G., Citizens´Trust and Authorities´ Choices, A Report from The Fourth International Conference on Ethics in the Public Service, Stockholm June 15-18, 1994.
Friberg, R., Prices, Profits and Exchange Rates.
Från optionsprissättning till konkurslagstiftning. Bergström, C., Björk, T., (red)
Hagerud, G.E., A New Non-Linear GARCH Model.
Haksar, A., Environmental Effects of Economywide Policies: Case Studies of Costa Rica and Sri Lanka.
He, C., Statistical Properties of Garch Processes.
Holmgren, M., Datorbaserat kontrollrum inom processindustrin; erfarenheter i ett tidsperspektiv.
Jennergren, P., Tutorial on the McKinsey Model for Valuation of Companies.
Lagerlöf, J., Essays on Political Economy, Information, and Welfare.
Lange, F., Wahlund, R., Planerade och oplanerade köp - Konsumenternas planering och köp av dagligvaror.

Löthgren, M., Essays on Efficiency and Productivity; Contributions on Bootstrap, DEA and Stochastic Frontier Models.
Nilsson, H.E., Using Conceptual Data Modelling in Management Accounting: A Case Study.
Sjöberg, L., Ramsberg, J., En analys av en samhällsekonomisk bedömning av ändrade säkerhetsföreskrifter rörande heta arbeten.
Säfvenblad, P., Price Formation in Multi-Asset Securities Markets.
Sällström, S., On the Dynamics of Price Quality.
Södergren, B., På väg mot en horisontell organisation? Erfarenheter från näringslivet av decentralisering och därefter.
Tambour, M., Essays on Performance Measurement in Health Care.
Thorén, B., Berg-Surwee, U., Områdesarbete i Östra Hökarängen - ett försök att studera effekter av decentralisering.
Zhang Gang., Chinese Rural Enterprises Between Plan and Market.
Åhlström, P., Sequences in the Profess of Adopting Lean Production.
Åkesson, G., Företagsledning i strategiskt vakuum. Om aktörer och förändrings-processer.
Åsbrink, S., Nonlinearities and Regime Shifts in Financial Time Series.

1996

Advancing your Business. People and Information Systems in Concert. Lundeberg, M., Sundgren, B (red).
Att föra verksamheten framåt. Människor och informationssystem i samverkan. red. Lundeberg, M., Sundgren, B.
Andersson, P., Concurrence, Transition and Evolution - Perspectives of Industrial Marketing Change Processes.
Andersson, P., The Emergence and Change of Pharmacia Biotech 1959-1995. The Power of the Slow Flow and the Drama of Great Events.
Asplund, M., Essays in Industrial Economics.
Delmar, F., Entrepreneurial Behavior & Business Performance.
Edlund, L., The Marriage Market: How Do You Compare?
Gunnarsson, J., Three Studies of Financial Behavior. Research Report.
Hedborg, A., Studies of Framing, Judgment and Choice.
Holgersson, C., Höök, P., Ledarutveckling för kvinnor - Uppföljning av en satsning på Volvo - Research Report.
Holmberg, C., Stores and Consumers - Two Perspectives on Food Purchasing.
Håkansson, P., Wahlund, R., Varumärken. Från teori till praktik.
Karlsson, A., The Family Business as an Heirloom. Research Report.
Linghag, S., Man är handelsstudent. Research Report.
Molin, J., Essays on Corporate Finance and Governance.
Mägi, A., The French Food Retailing Industry - A Descriptive Study.
Mölleryd, B., Så byggdes en världsindustri - Entreprenörskapets betydelse för svensk mobiltelefoni. Research Report.
Nielsen, S., Omkostningskalkulation for avancerede produktions-omgivelser - en sammenligning af stokastiske og deterministiske omkost-ningskalkulationsmodeller.
Normark, P., Danielsson, L., Larsson, A., Lundblad, P., Kooperativa nyckeltal. Research Report.
Sandin, R., Heterogeneity in Oligopoly: Theories and Tests.
Sandén-Håkansson, U., Från kampanjmål till mediemix - en studie av samarbete mellan annonsörer, reklambyråer och mediebyråer. Research Report.

Stein, J., Söderlund, M., Framgång i arbetet, strategier för att utföra arbetsuppgifterna, arbetsuppgifternas karaktär och utbildningskvalitet. En empirisk studie av civil-ekonomer. Research Report.
Strömberg, P., Thorburn, K., An Empirical Investigation of Swedish Corporations in Liquidation Bankruptcy. Research Report.
Söderlund, M., Och ge oss den nöjda kunden. En studie av kundtillfredsställelse och dess orsaker och effekter. Research Report.
Thodenius, B., Användningen av ledningsinformationssystem i Sverige: Lägesbild 1995. Research Report.
Ulfsdotter, U., Internationalisering för expansion eller hemmamarknadsförsvar? De nordiska marknaderna för fruktyoghurt 1982-1994.
Westelius, A., A Study of Patterns of Communication in Management Accounting and Control Projects.
Wijkström, F., Den svenska ideella sektorn och pengarna. Research Report.
Örtendahl, M., Health and Time - A Problem of Discounting.

1995

Becker, T., Essays on Stochastic Fiscal Policy, Public Debt and Private Consumption.
Blomberg, J., Ordning och kaos i projektsamarbete - en socialfenomenologisk upplösning av en organisationsteoretisk paradox.
Brodin, B., Lundkvist, L., Sjöstrand, S-E., Östman, L., Styrelsearbete i koncerner
Brännström, T., Bias Approximation and Reduction in Vector Autoregressive Models. Research Report.
Ekonomisk politik i omvandling. Jonung, L (red).
Gunnarsson, J., Wahlund, R., Hushållens finansiella strategier. En explorativ studie.
Höök, P., Chefsutveckling ur könsperspektiv - Mentorskap och nätverk på Vattenfall - Research Report.
Levin, J., Olsson, P., Looking Beyond the Horizon and Other Issues in Company Valuation. Research Report.
Mägi, A., Customer Statisfaction in a Store Performance Framework. Research Report.
Nittmar, H., Produktutveckling i samarbete.
Persson, P-G., Modeling the Impact of Sales Promotion on Store Profits.
Roman, L., Institutions in Transition. A Study of Vietnamese Banking.
Sandberg, J., Statistisk metod - dess vetenskapliga hemvist, grundläggande principer och möjligheter inom samhällsvetenskapen. Research Report.
Sandberg, J., How Do We Justify Knowledge Produced by Interpretative Approaches? Research Report.
Schuster, W., Redovisning av konvertibla skuldebrev och konvertibla vinstandelsbevis - klassificering och värdering.
Söderberg, K., Farmartjänsten - Ny kooperation inom lantbruket. Research Report.
Söderqvist, T., Benefit Estimation in the Case of Nonmarket Goods. Four Essays on Reductions of Health Risks Due to Residential Radon Radiation.
Tamm Hallström, K., Kampen för auktoritet – standardiseringsorganisationer i arbete.
Thorén, B., Användning av information vid ekonomisk styrning - månadsrapporter och andra informationskällor.