



HAL
open science

Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model

Hugo Kruiniger

► **To cite this version:**

Hugo Kruiniger. Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model. *Econometrics*, 2008, 144 (2), pp.447. 10.1016/j.jeconom.2008.03.001 . hal-00495780

HAL Id: hal-00495780

<https://hal.science/hal-00495780>

Submitted on 29 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model

Hugo Kruiniger

PII: S0304-4076(08)00039-0
DOI: [10.1016/j.jeconom.2008.03.001](https://doi.org/10.1016/j.jeconom.2008.03.001)
Reference: ECONOM 3019

To appear in: *Journal of Econometrics*

Received date: 5 July 2006
Revised date: 5 March 2008
Accepted date: 22 March 2008

Please cite this article as: Kruiniger, H., Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model. *Journal of Econometrics* (2008), doi:[10.1016/j.jeconom.2008.03.001](https://doi.org/10.1016/j.jeconom.2008.03.001)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model

Hugo Kruiniger*
Queen Mary, University of London

This version March 2008.

Abstract

This paper considers Maximum Likelihood (ML) based estimation and inference procedures for linear dynamic panel data models with fixed effects.

The paper first studies the asymptotic properties of MaCurdy's (Journal of Econometrics, 1982) First Difference Maximum Likelihood (FDML) estimator for the covariance stationary panel AR(1)/unit root model with fixed effects, viz. $y_{i,t} = \rho y_{i,t-1} + (1-\rho)\mu_i + \varepsilon_{i,t}$, under a variety of asymptotic plans. Subsequently, the paper shows through Monte Carlo simulations for panels of various dimensions the favourable finite sample properties of the FDMLE for ρ as compared to those of a number of alternative fixed effects ML estimators for ρ under covariance stationarity and normality of the data. The paper also discusses panel unit root test procedures that are based on the FDMLE. A Monte Carlo study conducted for one version of these tests reveals that it has very good size and power properties in comparison with alternative panel unit root tests.

JEL classification: C12, C13, C23.

Keywords: dynamic panel data models, Maximum Likelihood, multi-index asymptotics, efficiency bounds, unit root test.

*Address: H.KRUIGIER@QMUL.AC.UK; Dept. of Economics, Mile End Road, London E1 4NS, United Kingdom.
Tel.: +44-20-7882-7808; fax: +44-20-8983-3580.

1 Introduction

In this paper we discuss various Maximum Likelihood (ML) based estimation and inference procedures for the covariance stationary panel AR(1)/unit root model with fixed effects (FE). We study and compare the properties of several estimators for the autoregressive parameter, ρ , under various asymptotic plans and/or for panels of various dimensions. We also propose a new ML based panel unit root (UR) test and compare it with various existing panel UR tests in a Monte Carlo study.

In the dynamic panel data literature, broadly speaking, two classes of estimators are considered: GMM (IV) estimators and ML estimators. There is now a sizeable literature on GMM estimation of the panel AR(1) model, see e.g. Ahn and Schmidt (1995, 1997) and Arellano (2003). The Generalized Method of Moments owes much of its popularity to its flexibility: one can add or drop moment conditions depending on whether or not specific assumptions about the model are likely to be satisfied by the data. In particular, GMM can be used in the presence of heterogeneous data. For instance, the GMM estimator due to Arellano and Bond (1991) allows for both time-series and cross-sectional heteroskedasticity. However, Monte Carlo studies have revealed that GMM estimators have poor finite sample properties in some cases. For instance, when the value of ρ is close to unity, the Arellano and Bond estimator suffers from a weak instruments problem, see e.g. Blundell and Bond (1998). Moreover, when the number of moment conditions is large relative to the number of observations, e.g. when the number of lags used to form instruments is large, the bias of the Arellano and Bond estimator becomes quite severe, see e.g. Bun and Kiviet (2006).

The other major estimation method, i.e., the ML method, is generally not regarded as a viable alternative to GMM in the case of dynamic panel data models with fixed effects, because it is widely believed that fixed effects ML estimators are inconsistent due to the incidental parameters problem (cf. Neyman and Scott, 1948). The latter belief is probably based on the papers by Kiefer (1980) and Nickell (1981). Nickell has shown that the standard FEML estimator for the panel AR(1) model with arbitrary initial conditions, i.e., the Within Groups (WG) or Least Squares Dummy Variables estimator, is inconsistent when the cross-sectional dimension of the panel, N , tends to infinity whereas the time dimension of the panel, T , is fixed, while Kiefer (1980) has argued that the standard FEML estimator for the covariance matrix of the possibly autocorrelated errors of an otherwise static panel regression model is inconsistent when T is fixed.¹

The assessment of the usefulness of the ML method for estimating dynamic panel data models is not so bleak if one looks further. MaCurdy (1981, 1982) argued that in a situation where T can be treated as fixed,

¹Nickell derived a formula for the asymptotic bias of the WG estimator assuming covariance stationarity.

the ML method yields consistent estimators for covariance stationary panel ARMA models with fixed effects when it is applied to first-differences of the data. The resulting First Difference ML (FDML) estimators for the AR and MA parameters are still consistent under cross-sectional heteroskedasticity of the errors. More recently, Hsiao et al. (2002) and Kruiniger (2001) have independently shown that the panel AR(1) model with fixed effects and arbitrary initial conditions can be consistently estimated by the ML method, viz. the Restricted FEML (RFEML) estimator, if the differences between the initial observations and the individual effects or, equivalently, the differenced data satisfy a very mild condition, i.e., finite second (or $2+\delta$ -th) moments.^{2 3}

In this paper we aim to extend the asymptotic results of MaCurdy (1981, 1982) in two ways. First we investigate further the large N , fixed T asymptotic properties of the FDMLE for the covariance stationary panel AR(1)/UR model giving special attention to the unit root case and the question of efficiency. Among other things we show that the FDMLE does not attain the generalized Cramér-Rao lowerbound for estimators for the covariance stationary panel AR(1) model with fixed effects when $N \rightarrow \infty$ and T is fixed. Next, we examine the asymptotic properties of the FDMLE when T cannot be considered fixed. The results of the study provide further insight into the usefulness of the FDMLE as compared to other fixed effects ML estimators for the panel AR(1) model and permit the formulation of new powerful unit root test procedures.

Traditionally, the large N , fixed T asymptotic properties were considered the most relevant asymptotic properties of panel data estimators since the panel data sets used in econometric studies typically had a short time dimension. Because of the increasing availability of panel data sets that have a relatively long time dimension, e.g. the Penn World tables, some attention has recently been given to the properties of estimators under various alternative asymptotic plans in which T grows large, see also Phillips and Moon (1999). For instance, Alvarez and Arellano (2003) and Hahn and Kuersteiner (2002) have derived the asymptotic distribution of the WG estimator under large T , arbitrary N asymptotics and diagonal path asymptotics, respectively. These papers found that if the data are covariance stationary and $0 < \lim(N/T) < \infty$, then the WG estimator has a bias term in its asymptotic distribution. Hahn and Kuersteiner have also developed a bias-corrected version of the WG estimator using the formula for its diagonal path asymptotic bias under

²Hsiao et al. called the estimator the Transformed ML estimator rather than the Restricted FEML estimator.

³The standard FEML estimators treat the individual effects as N different parameters. On the other hand, the RFEMLE and the related FDMLE assume that the differences between the initial conditions and the individual effects are random variables with a common mean and a common variance parameter. This assumption allows one to formulate a likelihood function for the first-differences of the data that is free of incidental parameters. The RFEMLE is equal to the FDMLE when stationarity is imposed on (the common variance parameter of) the initial conditions.

covariance stationarity of the data.

In this paper we derive the large T , arbitrary N asymptotic properties of the FDMLE and the standard FEMLE for the covariance stationary panel AR(1) model. We find that under normality and covariance stationarity of the data, the FDMLE for ρ is asymptotically equivalent to the WG estimator when $T \rightarrow \infty$. But in contrast to the WG estimator, the FDMLE for ρ does not exhibit a bias term in its large T , arbitrary N asymptotic distribution unless the assumption of covariance stationarity is not satisfied by the data. Under large T , arbitrary N asymptotics the FDMLE is also asymptotically equivalent to the standard FEMLE for the covariance stationary panel AR(1) model. However, the latter estimator is inconsistent for fixed T and, like the WG estimator, asymptotically biased under large T , arbitrary N asymptotics.

We also show that in the unit root case both under large N , fixed T asymptotics and under joint N, T asymptotics the FDMLE for ρ has a normal limiting distribution. These findings immediately suggest a simple Wald-type panel UR test. The results of Monte Carlo experiments for panels of various dimensions indicate that our FDMLE based UR test has higher power against stationary ($|\rho| < 1$) alternatives than a number of well-known panel UR test from the literature including the test of Harris and Tzavalis (1999) which is based on the WG estimator, and a test of Levin et al. (2002). We also discuss whether and how our panel UR testing procedure should be modified to allow for possibly heterogeneous AR(p) and/or MA(q) dependence, drift or trend parameters, and for non-Gaussian or heterogeneously distributed errors. However, our panel UR test cannot (easily) be modified to allow for unknown structural breaks or general cross-sectional dependence. Recent surveys of the literature on panel UR tests are provided by Bond et al. (2005) and Breitung and Pesaran (2005).

The paper is organised as follows. Section 2 examines the asymptotic properties of the FDMLE for the panel AR(1)/UR model under a variety of asymptotic plans and discusses how the FDML framework can be used to conduct panel UR tests. Section 2 also investigates the asymptotic properties of the standard FEMLE for the covariance stationary panel AR(1) model. Section 3 contains two Monte Carlo studies. First it compares the power and size properties of various panel UR tests including FDMLE based UR tests. Subsequently, it compares the finite sample properties of the FDMLE for ρ with various other fixed effects ML estimators for the covariance stationary panel AR(1) model including two bias-corrected WG estimators. Section 4 concludes. The proofs are collected in the appendix.

2 ML estimation of the panel AR(1)/UR model with fixed effects

Consider the following panel AR(1) model with individual effects:

$$\begin{aligned} y_{i,t} &= \rho y_{i,t-1} + (1 - \rho)\mu_i + \varepsilon_{i,t}, \\ \eta_i &= (1 - \rho)\mu_i, \\ \varepsilon_{i,t} | \eta_i, y_{i,1} &\sim N(0, \sigma^2) \quad i.i.d., \quad i = 1, \dots, N, \quad t = 2, \dots, T, \end{aligned} \quad (1)$$

where i indicates the individual unit and t indicates the time period. Note that when $\rho = 1$ the individual effects, the μ_i , drop out from the model.

Below we consider various fixed effects (FE) estimators for the covariance stationary panel AR(1) model with possibly a unit root (UR). In the FE version of this model the individual effects are left completely unrestricted. Furthermore, the FE estimators for this model only exploit first-differences of the data.

The following assumptions imply that $\{y_{i,t}\}$ is covariance stationary:

$$\begin{aligned} |\rho| &< 1, \quad \text{and} \\ (y_{i,1} - \mu_i) | \mu_i &\sim N\left(0, \frac{\sigma^2}{1 - \rho^2}\right) \quad (i.i.d.), \quad i = 1, \dots, N. \end{aligned}$$

Let y_i denote the T -vector of all the observations for individual i including the first observation and let $V(\rho)$ or V for short denote a $(T \times T)$ matrix that is defined as follows: ⁴

$$V(\rho) = V = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \dots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \dots & \rho & 1 \end{pmatrix}. \quad (2)$$

Then we can write the covariance stationary panel AR(1) model with individual effects as

$$(y_i - \mu_i \iota) | \mu_i \sim N(0, \sigma^2 V(\rho)) \quad (i.i.d.), \quad i = 1, \dots, N, \quad (3)$$

where ι is a T -vector of ones.

We now define two ML estimators for the covariance stationary panel AR(1) model with fixed effects. Let μ be an N -vector that contains all the individual effects. Then the FEML estimator for ρ , μ and σ^2 in

⁴We usually omit the argument ρ of V .

the covariance stationary panel AR(1) model globally maximizes the following log-likelihood function:

$$\begin{aligned} \sum_{i=1}^N l_{CS,FE}(y_i; \rho, \mu, \sigma^2) &= -\frac{NT}{2} \log(2\pi) - \frac{NT}{2} \log \sigma^2 \\ &\quad - \frac{N}{2} \log |V(\rho)| - \frac{1}{2\sigma^2} \sum_{i=1}^N ((y_i - \mu_i \iota)' V^{-1}(\rho) (y_i - \mu_i \iota)). \end{aligned} \quad (4)$$

Under large N , fixed T asymptotics, this FEML estimator is an inconsistent estimator for ρ , σ^2 and μ due to the incidental parameters problem (cf Neyman and Scott, 1948, and Kiefer, 1980), see appendix A.5.

MaCurdy (1981, 1982) has suggested to take first differences of the data, so that the individual effects drop out from the model, before applying the ML estimation method. Let D denote the $(T-1 \times T)$ first-difference matrix with $D_{k,k} = -1$ and $D_{k,k+1} = 1$, $k = 1, \dots, T-1$, and $D_{k,l} = 0$ elsewhere. Then after taking first-differences of the observations in (3), we obtain:

$$D(y_i - \mu_i \iota) = Dy_i \sim N(0, \sigma^2 DV(\rho)D'), \quad (i.i.d.), \quad i = 1, \dots, N. \quad (5)$$

Note that the covariance matrix of Dy_i , $\sigma^2 DV(\rho)D'$, has an ARMA(1,1) structure with the MA parameter equal to -1.

The First Difference ML estimator (FDMLE) for ρ and σ^2 globally maximizes ⁵

$$\begin{aligned} \sum_{i=1}^N l(Dy_i; \rho, \sigma^2) &= -\frac{N(T-1)}{2} \log(2\pi) - \frac{N(T-1)}{2} \log \sigma^2 \\ &\quad - \frac{N}{2} \log |DV(\rho)D'| - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i' D' (DV(\rho)D')^{-1} Dy_i). \end{aligned} \quad (6)$$

Computation of the value of this log-likelihood function can be simplified by making use of the following equalities $|D'VD'| = (\iota'V^{-1}\iota)|V| = (T(1-\rho) + 2\rho)/(1+\rho)$ and $D'(D'VD')^{-1}D = V^{-1} - V^{-1}\iota\iota'V^{-1}/\iota'V^{-1}\iota$, which were established by Lancaster and Lindenhovius (1996). Note that V^{-1} has a very simple band-diagonal structure, see appendix A.3.2. Furthermore, $D'VD'$ is Positive Definite Symmetric (PDS) as long as $-1 < \rho < T/(T-2)$.

At this point it is useful to introduce some additional notation that is used in the appendices. $\theta = (r \ s^2)'$ denotes the vector of the common parameters. θ_0 denotes the vector of the true values of the common parameters, i.e. $\theta_0 = (\rho \ \sigma^2)'$. We also use $R(\rho) = DV(\rho)D'$ and $\Omega(\theta_0) = \sigma^2 R(\rho)$. Finally, $u_{i,t}$ is the homogeneous autoregressive process $u_{i,t} = y_{i,t} - \mu_i$.

When $|\rho| < 1$, the log-likelihood function given in (4) can be written as the sum of the log-likelihood function of the first-differences of the data, which is given in (6), and the N marginal log-likelihood functions

⁵It is easy to show that when σ^2 is unknown the FDMLE only exists for $T \geq 3$.

of the sufficient statistics for the fixed effects, $\tilde{\mu}_i = [l'V^{-1}l]^{-1}l'V^{-1}y_i$, $i = 1, \dots, N$, see appendix A.3.2. It follows that the FDML estimator for ρ and σ^2 is equal to the Conditional ML estimator for ρ and σ^2 .

When $\rho = 1$ the model specified in (1) can no longer be rewritten as (3). Nevertheless, as is shown in appendix A.1, the log-likelihood function of the first-differences given in (6) is still well-defined for $\rho = 1$, cf lemma 6. It is also shown in appendix A.1 that the likelihood function of the first-differences of the data can be differentiated with respect to the parameters as many times as is desired at and in the neighbourhood of the unit root, cf lemma 7.

2.1 Asymptotic properties of the First Difference ML estimator

We first present some results on the large N , fixed T asymptotic properties of the FDML estimator.

Theorem 1 *The FDML estimator for ρ and σ^2 in the covariance stationary panel AR(1)/UR model with fixed effects is consistent when $N \rightarrow \infty$ and T is fixed.*

PROOF

See appendix A.2.

Theorem 2 *When $-1 < \rho \leq 1$, the large N , fixed T limiting distribution of $(\hat{\rho}_{FDML}, \hat{\sigma}_{FDML}^2)$ is given by*

$$\sqrt{N} \begin{pmatrix} \hat{\rho}_{FDML} - \rho \\ \hat{\sigma}_{FDML}^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N(0, COV_{FDML,T}), \quad (7)$$

where

$$COV_{FDML,T} = \begin{pmatrix} I_{\rho\rho,T} & -\frac{(T-1)}{(1+\rho)(T(1-\rho)+2\rho)\sigma^2} \\ -\frac{(T-1)}{(1+\rho)(T(1-\rho)+2\rho)\sigma^2} & \frac{T-1}{2\sigma^4} \end{pmatrix}^{-1}, \quad (8)$$

with

$$I_{\rho\rho,T} = \frac{-2(1+\rho)(\rho^{T-1} - 2(T-2)\rho^2 + 2T-3)(T(1-\rho)+2\rho)}{(1-\rho^2)^2(T(1-\rho)+2\rho)^2} + \frac{2(1+\rho)^2((T-2)(\rho-1)-1)^2 + ((T-2)(1-\rho^2)+1+\rho^2)(T(1-\rho)+2\rho)^2}{(1-\rho^2)^2(T(1-\rho)+2\rho)^2}. \quad (9)$$

PROOF

See appendix A.3.

It can easily be verified that the numerator of $I_{\rho\rho}$ contains the factor $(1-\rho)^2$ so that the same factor in the denominator cancels out. Furthermore, COV_{FDML} is PDS for $-1 < \rho \leq 1$. Below we discuss the unit root case in more detail.

The limiting covariance matrix in (8) has been obtained under normality of the data. In general it will depend on the kurtosis of the data.

Lemma 1 in Neyman and Scott (1948) generalizes the large N , fixed T Cramér-Rao efficiency lower bound (henceforth CR bound) to models with incidental parameters.⁶ The generalized CR bound for (asymptotically) unbiased estimators for ρ and σ^2 in the covariance stationary panel AR(1) model with fixed effects is equal to $-(EHF_{\rho,\sigma^2})^{-1}$, where EHF_{ρ,σ^2} is the Expected Hessian with respect to ρ and σ^2 of the log-likelihood function $l_{CS,FE}(y_i; \rho, \mu, \sigma^2)$ given in (4). EHF_{ρ,σ^2} is derived in appendix A.3.2. It can easily be verified that under large N , fixed T asymptotics the FDMLE for ρ and σ^2 , which is equal to the Conditional MLE for ρ and σ^2 , does not attain this generalized CR bound; when T is fixed, the conditioning statistics, $\tilde{\mu}_i$, $i = 1, \dots, N$, are not ancillary for the common parameters, ρ and σ^2 .^{7 8}

We have the following results on the large T , arbitrary N asymptotic properties of the FDMLE:

Theorem 3 *The FDMLE estimator for ρ and σ^2 in the covariance stationary panel AR(1) model with fixed effects is consistent irrespective of whether $T \rightarrow \infty$, or $N \rightarrow \infty$, or $N, T \rightarrow \infty$ jointly. When $|\rho| < 1$, the large T and arbitrary N limiting distribution of $(\hat{\rho}_{FDMLE}, \hat{\sigma}_{FDMLE}^2)$ is given by*

$$\sqrt{NT} \begin{pmatrix} \hat{\rho}_{FDMLE} - \rho \\ \hat{\sigma}_{FDMLE}^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 - \rho^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \right). \quad (10)$$

PROOF

See appendix A.4 and in particular appendix A.4.1.

Note that the limiting variance of $\hat{\theta}_{FDMLE}$ is equal to $\lim_{T \rightarrow \infty} T \times COV_{FDMLE, T}$, where $COV_{FDMLE, T}$ is given in (8).

The large T , fixed N asymptotic variance of the FDMLE for ρ is equal to the large T , fixed N CR bound. Furthermore, the joint N, T asymptotic variance of the FDMLE for ρ is equal to both the (N, T) and the (T, N) sequential limit CR efficiency lower bound for the covariance stationary panel AR(1) model with fixed effects, and also to the diagonal path asymptotic (i.e. with $N/T \rightarrow \kappa$, where κ is a constant) efficiency

⁶Andersen (1970b) derives the Fisher efficiency lower bound for so-called regular estimators for models with incidental parameters. This bound is equal to the generalized CR bound of Neyman and Scott. Note that regular estimators are characterized by normality rather than (asymptotic) unbiasedness.

⁷See appendix A.3.2. For instance, the term $\frac{2}{1-\rho^2}(\rho^{T-1})$ from $\iota' \frac{dV^{-1}}{d\rho} V \frac{dV^{-1}}{d\rho} \iota$ in EHM does not cancel out. Since the panel AR(1) model with fixed effects is a special case of the panel ARMA(p,q) model with fixed effects, the FDMLE's for more general panel ARMA models are not asymptotically efficient either when T is fixed.

⁸Andersen (1970a) presents a necessary and sufficient condition for the equality of the variance of a Conditional MLE to the Cramér-Rao bound, see his 5th theorem.

lower bound for regular estimators for the panel AR(1) model with fixed initial observations which has been derived by Hahn and Kuersteiner (2002), i.e. $1 - \rho^2$.^{9 10}

The large T , arbitrary N asymptotic properties of the FDMLE for ρ can be obtained under assumptions that are considerably weaker than those we have made above. For instance, large T , arbitrary N consistency of the FDMLE does not depend on whether or not the initial observations satisfy covariance stationarity. Nor does consistency of the FDMLE require normality of the data. Moreover, under covariance stationarity of the data the large T , arbitrary N asymptotic distribution of the FDMLE for ρ is equal to that of the estimator that maximizes the function $l_{CS,FE}(y_i; \rho, \mu, \sigma^2)$ given in (4), i.e. the FEMLE for ρ , when the fixed effects are known. Therefore the limiting distribution of the FDMLE for ρ given in theorem 3 remains valid as long as the FDMLE for ρ is consistent, $(NT)^{-1/2} \sum_{i=1}^N \sum_{t=2}^T u_{i,t-1} \varepsilon_{i,t} \xrightarrow{d} N(0, \sigma^4(1 - \rho^2)^{-1})$, and $\text{plim}(NT)^{-1} \sum_{i=1}^N \sum_{t=2}^{T-1} u_{i,t}^2 = \sigma^2(1 - \rho^2)^{-1}$ as $T \rightarrow \infty$ regardless of N . In particular, its validity does not depend on normality of the data.¹¹

Under normality and covariance stationarity of the data, the FDMLE for ρ is large T , arbitrary N asymptotically equivalent to the WG estimator. However, unlike the WG estimator, the FDMLE for ρ does not exhibit a bias term in its large T , arbitrary N asymptotic distribution unless the assumption of covariance stationarity is violated by the initial observations. In that case the asymptotic distribution of the FDMLE for ρ contains a bias term of order $1/T$, namely $T^{-1} \rho \sigma^{-2} (1 - \rho^2) (N^{-1} \sum_{i=1}^N E(u_{i,1}^2) - \sigma^2(1 - \rho^2)^{-1})$. When $N/T \rightarrow 0$ this bias term can be omitted, but when $N/T \rightarrow \kappa > 0$ and $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E(u_{i,1}^2) \neq \sigma^2(1 - \rho^2)^{-1}$ this bias term must be kept.

⁹When $|\rho| < 1$, it is easily seen that both the scaled score and the scaled Hessian of the joint marginal likelihood function of the initial observations vanish as $T \rightarrow \infty$, regardless of whether N is fixed or tends to infinity (where the scaling is by appropriate powers of NT). Thus the diagonal path asymptotic efficiency lower bound for the panel AR(1) model with fixed initial observations is also valid for the covariance stationary panel AR(1) model, as the information contained in the initial observations becomes negligible when $T \rightarrow \infty$ and $|\rho| < 1$.

¹⁰Hahn and Kuersteiner derived their diagonal path asymptotic efficiency lower bound for the panel AR(1) model with fixed initial observations under the assumption that both $N^{-1} \sum_{i=1}^N y_{i,1}^2 = O(1)$ and $N^{-1} \sum_{i=1}^N \mu_i^2 = O(1)$. However, the fixed effects version of this model would only impose $N^{-1} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 = O(1)$, see Kruiniger (2001). Therefore, strictly speaking, their bound applies to random effects estimators. Nevertheless, for reasons similar to those in the previous footnote, when $T \rightarrow \infty$, the information contained in the level observations of one period becomes negligible so that the bound is also valid for fixed effects estimators which only exploit data in first-differences.

¹¹On the other hand, the large T , arbitrary N limiting distribution of the FDMLE for σ^2 does depend on the assumption of normality.

Under large T , arbitrary N asymptotics the FDMLE is also asymptotically equivalent to the FEMLE for the covariance stationary panel AR(1) model with unknown μ , which is consistent in this case, see appendix 5. However, when the data are covariance stationary the FEMLE for ρ has the same large T , arbitrary N asymptotic bias as the WG estimator, i.e. $-T^{-1}(1 + \rho)$.

2.2 A new panel unit root test

One can obtain the large N , fixed T limiting distribution of the FDMLE for ρ and σ^2 when $\rho = 1$ from theorem 2 by evaluating $\lim_{\rho \uparrow 1} COV_{FDMLE}^{-1}$ using de l'Hôpital's rule. This leads to the following result:

Corollary 4 *When $\rho = 1$, the large N , fixed T limiting distribution of $(\hat{\rho}_{FDMLE}, \hat{\sigma}_{FDMLE}^2)$ is given by*

$$\sqrt{N} \begin{pmatrix} \hat{\rho}_{FDMLE} - 1 \\ \hat{\sigma}_{FDMLE}^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N \left(0, \begin{bmatrix} \frac{8}{(T-1)(T-2)} & \frac{4\sigma^2}{(T-1)(T-2)} \\ \frac{4\sigma^2}{(T-1)(T-2)} & \frac{2\sigma^4}{T-2} \end{bmatrix} \right). \quad (11)$$

Corollary 4 can be generalized to results for the FDMLE of the panel ARIMA(p,1,q) model.

When $\rho = 1$ and $T \rightarrow \infty$ the FDMLE remains consistent irrespective of the asymptotic behaviour or the value of N . Moreover, when $\rho = 1$ the joint N, T limiting distribution of $\hat{\rho}_{FDMLE}$ is very similar to its large N , fixed T limiting distribution:

Theorem 5 *When $\rho = 1$, the FDMLE estimator for ρ and σ^2 in the panel AR(1)/UR model with fixed effects is consistent irrespective of whether $T \rightarrow \infty$, or $N \rightarrow \infty$, or $N, T \rightarrow \infty$ jointly. The joint N, T limiting distribution of $(\hat{\rho}_{FDMLE}, \hat{\sigma}_{FDMLE}^2)$ is given by*

$$\begin{pmatrix} T\sqrt{N}(\hat{\rho}_{FDMLE} - 1) \\ \sqrt{NT}(\hat{\sigma}_{FDMLE}^2 - \sigma^2) \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 8 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \right). \quad (12)$$

PROOF

See appendix A.4 and in particular appendix A.4.2.

When $\rho = 1$ the large T , fixed N limiting distribution of the FDMLE for ρ is non-normal. However, the joint N, T limiting distribution of the FDMLE for ρ is normal when $\rho = 1$ owing to averaging across individuals. Note that the joint N, T limiting variance of $\hat{\rho}_{FDMLE}$ is equal to the limit of T^2 times its large N , fixed T limiting variance as $T \rightarrow \infty$, i.e. $\lim_{T \rightarrow \infty} T^2 \times 8/[(T-1)(T-2)] = 8$.

Just as in the stationary case given in theorem 3, the validity of the joint N, T limiting distribution of the FDMLE for ρ given in theorem 5 does not depend on normality of the data. The same limiting distribution will be obtained when the normality assumption is replaced by $E|\varepsilon_{i,t}|^{4+\delta} < \infty$. In particular, in contrast to its large N , fixed T limiting variance, the joint N, T limiting variance of the FDMLE for ρ does not depend on the kurtosis of the idiosyncratic errors.

Theorem 5 can be extended to the FDMLE for the panel AR(p) model with a unit root.

The results in corollary 4 and theorem 5 suggest a new unit root test for panel data models with fixed T or large T which is a Wald-type t-test that is based on the FDMLE for ρ and its asymptotic standard error.

The FDMLE based panel UR test can be modified so as to allow for heteroskedasticity, heterogeneous drift parameters, AR(p) dependence and/or MA(q) dependence. For instance, to allow for individual drift parameters, the data should be differenced twice rather than only once. In order to test the unit root hypothesis in a panel AR(1) model with possibly heteroskedastic idiosyncratic errors, e.g. with $E(\varepsilon_{i,t}^2) = \sigma_{i,t}^2$, $i = 1, \dots, N$, $t = 2, \dots, T$, one can use a t-test based on a modified version of the log-likelihood function in (6) with $SDV(\rho)D'S$ replacing $\sigma^2 DV(\rho)D'$ where $S = \text{diag}(\varsigma_2, \dots, \varsigma_T)$ and $\varsigma_2, \varsigma_3, \dots, \varsigma_T$ are auxiliary parameters. If $\rho = 1$, then $N^{-1} \sum_{i=1}^N E(Dy_i y_i' D')$ = $N^{-1} \sum_{i=1}^N E(\varepsilon_i \varepsilon_i')$ = $\text{diag}(\bar{\sigma}_2^2, \dots, \bar{\sigma}_T^2)$, where $\varepsilon_i = (\varepsilon_{i,2}, \dots, \varepsilon_{i,T})'$, $i = 1, \dots, N$, and $\bar{\sigma}_t^2 = N^{-1} \sum_{i=1}^N \sigma_{i,t}^2$, $t = 2, \dots, T$, and in addition $SDV(\rho)D'S = S^2 = \text{diag}(\varsigma_2^2, \dots, \varsigma_T^2)$ (cf. the proof of lemma 6 in appendix A.1). Therefore we can expect the Quasi ML estimator for ρ that is based on the modified version of the log-likelihood function to be consistent and asymptotically normal when $\rho = 1$, and the corresponding t-test for $\rho = 1$ to have correct size. Before a specific version of the FDMLE based panel UR test is used, it is important to ascertain that its underlying assumptions, e.g. homoskedasticity or no structural breaks, are satisfied by the data.

2.2.1 A comparison of the local power of various panel unit root tests

Various unit root tests for panels with fixed T or large T have already been suggested in the literature. In a fixed T situation Breitung and Meyer (1994) have proposed a Modified Dickey-Fuller test-statistic which is based on the OLS estimator for ρ in a model for deviations of the data from the initial observations, viz. $y_{i,t} - y_{i,1} = \rho(y_{i,t-1} - y_{i,1}) + \nu_{i,t}$, where $\nu_{i,t} = \varepsilon_{i,t} - (1 - \rho)(y_{i,1} - \mu_i)$ for $t = 3, \dots, T$. This estimator is unbiased for $\rho = 1$. Straightforward calculations show that the large N , fixed T local-to-unity limiting variance of this estimator is equal to $\text{limvar}(\hat{\rho}_{BM,T}) = 2/((T - 1)(T - 2))$. Harris and Tzavalis (1999) have presented an LM-type UR test-statistic for the panel AR(1) model with individual effects which is based on an adjusted version of the WG estimator. The adjustment is equal to minus the large N , fixed T asymptotic bias of the WG estimator when $\rho = 1$, namely $3/T$. The large N , fixed T local-to-unity limiting variance of this estimator is equal to $\text{limvar}(\hat{\rho}_{WG,T}) = 3(17(T - 1)^2 - 20(T - 1) + 17)/(5T^3(T - 2))$. Hahn and Kuersteiner (2002) have derived the joint N, T asymptotic distribution of the WG estimator when $\rho = 1$, which is a normal distribution. The joint N, T asymptotic bias of the WG estimator is equal to the large N , fixed T asymptotic bias of the WG estimator, $-3/T$, while the joint N, T limiting variance of

$\hat{\rho}_{WG}$, $\limvar(\hat{\rho}_{WG})$, is equal to the limit of T^2 times its large N fixed T limiting variance as $T \rightarrow \infty$, i.e. $\limvar(\hat{\rho}_{WG}) = \lim_{T \rightarrow \infty} T^2 \times \limvar(\hat{\rho}_{WG,T}) = 10.2$.

Under covariance stationarity of the data both the OLS estimator of Breitung and Meyer, $\hat{\rho}_{BM}$, and the bias-corrected WG estimator, i.e. $\hat{\rho}_{WG} + 3/T$, are biased upwards. When T is fixed and the data are covariance stationary the asymptotic bias of the former estimator is $(1 - \rho)/2$ (see Breitung and Meyer) while tedious calculations show that local-to-unity the asymptotic bias of the latter estimator is equal to $(1 - \rho)(1 + 3/T)/4$.¹² These biases reduce the power of the UR tests that are based on the corresponding estimators. On the other hand the FDMLE for ρ is asymptotically unbiased. The local power of the aforementioned UR tests is given by the following large N , fixed T limiting distributions of the corresponding test statistics under local alternatives of the form $\rho = 1 - cN^{-1/2}$: $N^{1/2}(\hat{\rho}_{BM} - 1)/(\limvar(\hat{\rho}_{BM,T}))^{1/2} \xrightarrow{d} N(-\frac{c}{2}\sqrt{\frac{(T-1)(T-2)}{2}}, 1)$, $N^{1/2}(\hat{\rho}_{WG} + 3/T - 1)/(\limvar(\hat{\rho}_{WG,T}))^{1/2} \xrightarrow{d} N(-\frac{3c}{4}(1 - T^{-1})(\limvar(\hat{\rho}_{WG,T}))^{-1/2}, 1)$, and $N^{1/2}(\hat{\rho}_{FDMLE} - 1)/(\limvar(\hat{\rho}_{FDMLE,T}))^{1/2} \xrightarrow{d} N(-c\sqrt{\frac{(T-1)(T-2)}{8}}, 1)$.¹³ Thus the panel UR tests based on $\hat{\rho}_{BM}$ and $\hat{\rho}_{FDMLE}$, respectively, have the same local power, while the test based on $\hat{\rho}_{WG} + 3/T$ has less power against local covariance stationary alternatives than the two tests just mentioned.

Finally, we consider how the power of the UR test that is based on the $\hat{\rho}_{FDMLE}$ is affected when the initial observations do not satisfy covariance stationarity. Assume that $E((y_{i,1} - \mu_i)^2) = \lambda\sigma^2(1 - \rho^2)^{-1}$ for all $i \in \{1, \dots, N\}$, where $\lambda \geq 0$. Note that in this case $(NT)^{1/2}(\hat{\rho}_{FDMLE} - \rho - T^{-1}\rho(\lambda - 1)) \xrightarrow{d} N(0, 1 - \rho^2)$. In addition let $\rho = 1 - c(NT)^{-1/2}$ and $N/T \rightarrow \kappa$. Then $\limvar(\hat{\rho}_{FDMLE}) = 8$ and $(NT)^{1/2}(\hat{\rho}_{FDMLE} - 1)/(\limvar(\hat{\rho}_{FDMLE}))^{1/2} \xrightarrow{d} N((-c + \kappa^{1/2}(\lambda - 1))/\sqrt{8}, 1)$. Thus the local power of the $\hat{\rho}_{FDMLE}$ based UR test against alternatives with $\lambda < 1$ is larger than its local power against covariance stationary alternatives.

3 Simulation experiments

3.1 The finite sample performance of panel unit root tests

In this section we compare the finite sample performance of our panel unit root test, which is based on the FDMLE for ρ , with four other panel UR tests, namely (1) the test of Breitung and Meyer (henceforth BM) discussed in section 2, (2) the test due to Harris and Tzavalis which is based on the bias-corrected WG estimator, (3) a test which is based on the Restricted FEMLE (hereafter RFEMLE) for ρ , i.e. the

¹²In deriving the latter asymptotic bias we used a well-known expression for the bias of $\hat{\rho}_{WG}$ due to Nickell (1981).

¹³The limiting distributions of the UR test statistics based on $\hat{\rho}_{BM,T}$ and $\hat{\rho}_{FDMLE,T}$ under local alternatives have already been derived by Bond et al. (2005).

FEMLE for ρ in the panel AR(1) model with fixed effects and arbitrary initial conditions which is given by (1) and the additional assumption that $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 < \infty$ (see Kruiniger 2001, and Hsiao et al., 2002), and (4) the test due to Levin et al. (2002) (henceforth LL) for their model 2. Under the null hypothesis of a unit root, i.e. $\rho = 1$, the first three tests are similar with respect to the initial conditions because the test statistics only exploit data in first differences. Furthermore, the power of the first three tests only depends on the distributional properties of the $\varepsilon_{i,t}$ and the differences between the initial conditions and the individual effects, $y_{i,1} - \mu_i$, $i = 1, \dots, N$. The UR test due to LL is only valid for panels where T is large compared to N , i.e., where $N/T \rightarrow 0$.

In most simulation experiments we have conducted the errors have been drawn from normal distributions: $\varepsilon_{i,t} \sim N(0, \sigma^2)$ with $\sigma^2 = 1$. We note that the value chosen for σ^2 is of no consequence. To assess how the assumptions with respect to $y_{i,1} - \mu_i$, $i = 1, \dots, N$, affect the power of the tests, we have conducted three different kinds of experiments: in one set the initial observations have been drawn from covariance stationary distributions, i.e. $(y_{i,1} - \mu_i) | \mu_i \sim N(0, \sigma^2 / (1 - \rho^2))$, while in the other two sets the initial observations are non-stationary. The two non-stationary cases considered are: (1) $y_{i,1} - \mu_i = 0$; and (2) $(y_{i,1} - \mu_i) | \mu_i \sim N(0, 2\sigma^2 / (1 - \rho^2))$. Note that in all situations $E(y_{i,t} - y_{i,t-1}) = 0$ as is the case under the null hypothesis. In both case (1) and case (2) the variance of $y_{i,1} - \mu_i$ is different from the variance under covariance stationarity. We have also conducted experiments in which the errors and $y_{i,1} - \mu_i$, $i = 1, \dots, N$, are drawn from standardized/centred $\chi^2(1)$ distributions and experiments in which the data are drawn from heterogenous distributions, i.e., with $\rho_i \sim \bar{\rho} + 0.9(1 - \bar{\rho}) \times \text{Uniform}[-1, 1]$, $i = 1, \dots, N$.

In the simulation experiments panel data sets of various dimensions have been considered. We generated both traditional ‘micro-economic’ panels, where T is small as compared to N , and panels with dimensions that can be found in the recent empirical macro-economic literature, where T is not small as compared to N . The simulation evidence for these ‘macro’ panels allows us to investigate the relevancy and accuracy of the joint asymptotics results. All simulation results are based on 10,000 replications and the (nominal) level of the tests is always 5%.

In an i.i.d. situation, one can envisage two versions of the FDMLE based UR tests: the Wald version which uses an unrestricted estimate of the standard error of the estimator, and an LM version which uses a restricted estimate of the standard error that is obtained under the null hypothesis. The latter estimate of the standard error is an extremely simple function of N and T only, see corollary 4 and theorem 5.

The implementation of the RFEMLE based unit root test raises a number of issues. First, one could in principle restrict the estimate of the variance of the ‘constant error component,’ i.e. $\sigma_v^2 \equiv (1 - \rho)^2 \text{Var}(y_{i,1} -$

μ_i), to be positive as is often done in ML estimation of static random effects models. However, when ρ is close or equal to one, this constraint is binding in about 50% of the cases and often causes computational problems. Therefore during the iterations we only restricted the values of $\hat{\sigma}_{RFEMLE}^2$ and $\hat{\sigma}_{RFEMLE}^2 + (T-1)\hat{\sigma}_{v,RFEMLE}^2$ to be positive. These restrictions are satisfied by the estimators by construction. Secondly, under homoskedasticity of the $\varepsilon_{i,t}$ the expected Hessian of the log-likelihood function is singular when $\rho = 1$ (although $\hat{\rho}_{RFEMLE}$ is still consistent in this case). As a result when $\rho = 1$ the Wald test-statistic $(\hat{\rho}_{RFEMLE} - 1)/SD(\hat{\rho}_{RFEMLE})$ has a distribution that is different from a standard normal distribution.¹⁴ Therefore we had to obtain this distribution through simulations. Moreover, when ρ is close or equal to one, the Hessian of the log-likelihood function evaluated at the estimates was sometimes ill-conditioned and its inverse could not be computed. In all cases considered this happened in less than 0.2% of the replications. When it occurred, we added $(0.0001 - \lambda_1)I$ to the Hessian where λ_1 is its smallest eigenvalue.

The empirical quantiles of the RFEMLE based test statistic for $\rho = 1$ as well as its empirical size when using the conventional (i.e. standard normal) critical value of 1.645 (at the 5% level of significance) are reported for i.i.n.d. data in table 1 for different combinations of N and T . The empirical sizes of this test differ substantially from the nominal level of 5%. Therefore, we also investigated the performance of the size-adjusted RFEMLE based test, which uses the critical values from table 1. Because the empirical sizes of the FDMLE based tests are always close to the nominal level whenever N is large (i.e. ≥ 100), we did not consider size-adjusted versions of these tests in the case of ‘micro’ panels. Finally, we only considered a size-adjusted version of the test due to LL.

The test due to BM has already been compared with other panel UR tests for the case of ‘micro’ panels in a Monte Carlo study conducted by Bond et al. (2005). That study found that the BM test performed very similar to our FDMLE based test. Furthermore, as mentioned above, the LL test is not valid for ‘micro’ panels. Therefore, for both the BM test and the LL test we only report simulation results for the ‘macro’ panels. On the other hand, we only report results on the RFEMLE based test for a subset of the ‘macro’ panels that we consider, because using this test makes more sense for ‘micro’ panels with possibly non-stationary initial observations under the alternative hypothesis than for ‘macro’ panels where the distributions of the initial observations under the alternative hypothesis are of limited consequence.

Tables 2-9 report the simulation results on the empirical size and power of the panel unit root tests that were mentioned above. Tables 2, 4, 6, 7, 8 and 9 report results on power against covariance stationary

¹⁴Binder et al. (2005) have also considered this test-statistic. However their critical values seem to have been taken from the table of the standard normal distribution.

alternatives, whereas tables 3 and 5 report results on power against non-stationary alternatives. In the tables ‘W’ denotes the Wald version of a test, ‘LM’ stands for the LM version of a test, while ‘SA’ indicates that the test has been size-adjusted. Finally, ‘SW’ indicates that the test has been based on a ‘sandwich’ estimator for the variance of the estimator for ρ . Inspection of the results in tables 2-9 leads to the following conclusions:

1. The Wald version of the FDMLE based test which uses a simple standard error based on the empirical Hessian does not suffer from size distortions unless N is small.
2. When the data are asymmetrically distributed the FDMLE-W-SW test, which uses a ‘sandwich’ formula for the standard error, and the WG based LM test over-reject.
3. The FDMLE based test has uniformly and substantially higher power against covariance stationary alternatives than the size-adjusted WG and RFEMLE based tests. This conclusion still holds when the ρ_i are heterogenous or when the data are asymmetrically distributed.
4. The FDMLE based test has uniformly and substantially higher power against ‘non-stationary’ alternatives with $\rho < 1$ and $y_{i,1} = \mu_i$ than the size-adjusted WG and RFEMLE based tests, with the WG based test performing uniformly better than the RFEMLE based test. When considering ‘non-stationary’ alternatives with $\rho < 1$ and $(y_{i,1} - \mu_i)|\mu_i \sim N(0, 2/(1 - \rho^2))$, the ordering of the tests on the basis of their power properties is reversed although the differences in power are small. The power of the FDMLE based test decreases as $Var(y_{i,1} - \mu_i)$ increases as predicted by the theory in section 2.
5. In the scenarios considered the size-adjustment of the RFEMLE based test reduces its power by up to 0.21.
6. When T is large and $N \geq 10$, the size-adjusted FDMLE based test has slightly higher power against covariance stationary alternatives than the size-adjusted BM test and substantially higher power against covariance stationary alternatives than the LL test.
7. The power of each test increases when N or T increases or when the value of ρ decreases.
8. The power properties of the Wald and LM versions of the FDMLE based tests are similar when T is small.

T	N	1%	5%	10%	50%	90%	95%	99%	size
4	100	-3.93	-2.56	-1.75	-0.0074	1.93	2.89	4.85	0.110
4	500	-3.86	-2.40	-1.58	0.0018	1.94	2.88	4.43	0.095
7	100	-3.80	-2.49	-1.78	0.0000	2.19	3.27	5.28	0.111
7	500	-3.82	-2.44	-1.74	-0.0005	2.01	2.94	4.83	0.107
6	100	-3.74	-2.47	-1.75	-0.0095	2.06	3.10	5.44	0.110
25	10	-3.35	-2.24	-1.58	0.0027	3.56	5.41	9.42	0.095
25	25	-3.32	-2.19	-1.42	0.0127	2.81	4.21	6.71	0.088
$N(0, 1)$		-2.33	-1.65	-1.28	0.0000	1.28	1.65	2.33	0.050

Table 1: Empirical quantiles of $(\hat{\rho}_{RFEML} - 1)/SD(\hat{\rho}_{RFEML})$ when $\rho = 1$; $\varepsilon_{i,t} \sim N(0, 1)$.

T	N	test	size	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
4	100	FDML-W	0.056	0.13	0.25	0.53
		RFEML-W-SA	0.049	0.10	0.13	0.19
		RFEML-W	0.110	0.20	0.26	0.33
4	500	FDML-W	0.050	0.26	0.59	0.98
		RFEML-W-SA	0.053	0.14	0.18	0.32
		RFEML-W	0.095	0.23	0.29	0.45
7	100	FDML-W	0.057	0.26	0.59	0.98
		RFEML-W-SA	0.046	0.12	0.19	0.42
		RFEML-W	0.111	0.24	0.33	0.59
7	500	FDML-W	0.054	0.68	0.99	1.00
		RFEML-W-SA	0.047	0.16	0.31	0.82
		RFEML-W	0.107	0.28	0.46	0.90

Table 2: power against covariance stationary alternatives.

T	N	test	size	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	FDML-W	0.056	0.457	0.914	1.000
		RFEML-W-SA	0.049	0.138	0.182	0.290
		RFEML-W	0.110	0.246	0.310	0.506
		WG-LM-SA	0.050	0.308	0.732	0.994
		FDML-LM	0.063	0.473	0.924	1.000

Table 3: power against “non-stationary” alternatives with $y_{i,1} = \mu_i$.

T	N	test	size	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	FDML-W	0.056	0.200	0.466	0.925
		RFEML-W-SA	0.049	0.118	0.162	0.319
		RFEML-W	0.110	0.230	0.287	0.483
		WG-LM-SA	0.050	0.153	0.327	0.812
		FDML-LM	0.063	0.217	0.486	0.935

Table 4: power against covariance stationary alternatives.

T	N	test	size	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	FDML-W	0.056	0.068	0.116	0.351
		RFEML-W-SA	0.049	0.098	0.170	0.396
		RFEML-W	0.110	0.192	0.304	0.569
		WG-LM-SA	0.050	0.060	0.101	0.378
		FDML-LM	0.063	0.080	0.127	0.388

Table 5: power against “non-stationary” alternatives with $(y_{i,1} - \mu_i) | \mu_i \sim N(0, 2/(1 - \rho^2))$.

T	N	test	size	$\bar{\rho} = 0.95$	$\bar{\rho} = 0.9$	$\bar{\rho} = 0.8$
6	100	FDML-W	0.056	0.201	0.470	0.911
		RFEML-W-SA	0.049	0.119	0.173	0.353
		RFEML-W	0.110	0.231	0.308	0.523
		WG-LM-SA	0.050	0.145	0.334	0.805
		FDML-LM	0.063	0.217	0.491	0.921

Table 6: power against heterogeneous covariance stationary alternatives with $\rho_i \sim \bar{\rho} + 0.9(1 - \bar{\rho}) \times \text{Uniform}[-1, 1]$, $i = 1, \dots, N$.

T	N	test	size	cr. value	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
25	5	FDML-W	0.080		0.247	0.510	0.937
		FDML-W-SA		-2.10	0.126	0.314	0.829
		WG-LM-SA		-2.36	0.127	0.279	0.763
		BM-LM-SA		-2.85	0.152	0.328	0.750
		FDML-LM	0.125		0.331	0.620	0.972
25	10	FDML-W	0.086		0.359	0.773	0.998
		FDML-W-SA		-1.94	0.249	0.660	0.995
		RFEML-W-SA		-2.24	0.174	0.365	0.871
		RFEML-W	0.095		0.310	0.545	0.940
		LL-SA				0.229	
		WG-LM-SA		-2.14	0.201	0.519	0.978
		BM-LM-SA		-2.40	0.252	0.615	0.978
		FDML-LM	0.118		0.431	0.840	0.999
25	25	FDML-W	0.064		0.623	0.985	1.000
		FDML-W-SA		-1.80	0.562	0.978	1.000
		RFEML-W-SA		-2.19	0.258	0.620	0.994
		RFEML-W	0.088		0.380	0.743	0.997
		LL-SA				0.448	
		WG-LM-SA		-1.97	0.389	0.888	1.000
		BM-LM-SA		-2.07	0.544	0.963	1.000
		FDML-LM	0.081		0.686	0.991	1.000
25	50	FDML-W	0.061		0.872	1.000	1.000
		FDML-W-SA		-1.78	0.836	1.000	1.000
		LL-SA				0.739	
		WG-LM-SA		-1.96	0.620	0.994	1.000
		BM-LM-SA		-1.95	0.821	1.000	1.000
		FDML-LM	0.075		0.896	1.000	1.000

Table 7: power against covariance stationary alternatives. For the LL test $T = 26$.

T	N	test	size	cr. value	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$	
50	5	FDML-W	0.085		0.527	0.942	1.000	
		FDML-W-SA		-2.17	0.293	0.808	1.000	
		WG-LM-SA		-2.41	0.273	0.736	0.999	
			BM-LM-SA		-2.81	0.361	0.778	0.994
			FDML-LM	0.126		0.648	0.974	1.000
		10	FDML-W	0.085		0.783	0.999	1.000
		FDML-W-SA		-1.96	0.659	0.996	1.000	
		LL-SA				0.574		
		WG-LM-SA		-2.09	0.528	0.980	1.000	
		BM-LM-SA		-2.39	0.645	0.986	1.000	
		FDML-LM	0.120		0.854	1.000	1.000	
50	25	FDML-W	0.073		0.990	1.000	1.000	
		FDML-W-SA		-1.83	0.981	1.000	1.000	
		LL-SA				0.925		
			WG-LM-SA		-1.99	0.889	1.000	1.000
			BM-LM-SA		-2.05	0.975	1.000	1.000
			FDML-LM	0.094		0.994	1.000	1.000
50	50	FDML-W	0.064		1.000	1.000	1.000	
		FDML-W-SA		-1.77	1.000	1.000	1.000	
		LL-SA				0.998		
			WG-LM-SA		-1.86	0.995	1.000	1.000
			BM-LM-SA		-1.88	1.000	1.000	1.000
			FDML-LM	0.079		1.000	1.000	1.000
100	5	FDML-W	0.086		0.948	1.000	1.000	
		FDML-LM	0.141		0.983	1.000	1.000	
100	10	FDML-W	0.078		0.999	1.000	1.000	
		FDML-LM	0.117		1.000	1.000	1.000	

Table 8: power against covariance stationary alternatives. For the LL test $T = 51$.

T	N	test	size	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	FDML-W	0.046	0.208	0.481	0.858
		FDML-W-SW	0.087			
		RFEML-W-SA	0.047	0.102	0.108	0.173
		RFEML-W	0.176	0.307	0.350	0.526
		WG-LM	0.089	0.205	0.361	0.749
		FDML-LM	0.053	0.223	0.502	0.872

Table 9: power against covariance stationary alternatives; errors and $y_{i,1} - \mu_i$, $i = 1, \dots, N$, drawn from standardized/centred $\chi^2(1)$ distributions.

3.2 A comparison of ML estimators

We have also conducted some Monte Carlo experiments to compare the finite sample performance of the FDMLE for ρ , $\hat{\rho}_{FDML}$, the RFEMLE for ρ , $\hat{\rho}_{RFEML}$, and two Bias-Corrected Within Groups (BCWG) estimators for ρ . The first BCWG estimator, $\hat{\rho}_{BCWG1}$, is given by equation (43) in Bun and Carree (2005) and is based on the asymptotic bias formula for $\hat{\rho}_{WG}$ due to Nickell (1981): $\hat{\rho}_{BCWG1} = h^{-1}(\hat{\rho}_{WG})$, where $h(\rho) = \rho + \text{plim}_{N \rightarrow \infty}(\hat{\rho}_{WG} - \rho)$. The second BCWG estimator, $\hat{\rho}_{BCWG2}$, is due to Hahn and Kuersteiner (2002) and is based on a formula for the diagonal path asymptotic bias of $\hat{\rho}_{WG}$: $\hat{\rho}_{BCWG2} = \frac{T}{T-1}\hat{\rho}_{WG} + \frac{1}{T-1}$. The design of our experiments is similar to that of Hahn and Kuersteiner. We have generated artificial data from the covariance stationary panel AR(1) model with $\sigma^2 = 1$ for several values of ρ : $\rho \in \{0, 0.3, 0.6, 0.9\}$. The value chosen for σ^2 is again of no consequence. Nor does it matter how the individual effects have been generated as they are differenced out by the estimators. Finally, the dimensions of the panels are $N \in \{100, 200\}$ and $T \in \{6, 11, 21, 51\}$. We also compared the results for panels with $N = 50$ and $T = 51$. All simulation results are based on 10,000 replications.

The results are reported in table 9. It is easily seen that in all cases the bias of $\hat{\rho}_{FDML}$ is smaller than the bias of $\hat{\rho}_{BCWG2}$, often by a large factor, e.g. 100, and also smaller than the bias of $\hat{\rho}_{RFEML}$. For every panel size considered the bias of each of these three estimators is the largest for $\rho = 0.9$. The bias-corrected estimator due to Bun and Carree, $\hat{\rho}_{BCWG1}$, is unbiased by definition. Furthermore the rmse of $\hat{\rho}_{FDML}$ is always smaller than the rmse's of the other estimators but the differences between the rmse's of the estimators vanish when T grows large. Moreover, the rmse's of $\hat{\rho}_{BCWG1}$ and $\hat{\rho}_{RFEML}$ are always smaller than the rmse of $\hat{\rho}_{BCWG2}$. The latter finding is interesting because unlike $\hat{\rho}_{BCWG2}$ the consistency of $\hat{\rho}_{RFEML}$ is robust to violations of covariance stationarity of the data. The rmse of $\hat{\rho}_{FDML}$ tends to decrease when the value of ρ increases except when T is small. On the other hand the rmse's of $\hat{\rho}_{RFEML}$, $\hat{\rho}_{BCWG2}$ and to a lesser extent $\hat{\rho}_{BCWG1}$ tend to increase when the value of ρ increases, except when T is large as compared to N . We conclude that it is better to use $\hat{\rho}_{FDML}$ rather than $\hat{\rho}_{BCWG1}$ and $\hat{\rho}_{BCWG2}$, especially when ρ is close to unity and T is relatively small, unless the assumption of covariance stationarity is (seriously) violated in which case $\hat{\rho}_{RFEML}$ could be used.

Case			Bias $\times 100$			RMSE			
T	N	ρ	$\hat{\rho}_{RFEML}$	$\hat{\rho}_{FDML}$	$\hat{\rho}_{BCWG2}$	$\hat{\rho}_{RFEML}$	$\hat{\rho}_{FDML}$	$\hat{\rho}_{BCWG1}$	$\hat{\rho}_{BCWG2}$
6	100	0.0	0.105	0.038	-4.016	0.057	0.057	0.058	0.067
		0.3	0.003	-0.084	-7.057	0.067	0.062	0.064	0.090
		0.6	1.326	-0.122	-11.54	0.112	0.066	0.071	0.129
		0.9	-1.881	-0.262	-17.77	0.109	0.065	0.077	0.187
6	200	0.0	0.025	0.022	-3.991	0.040	0.040	0.041	0.055
		0.3	0.031	-0.006	-6.959	0.048	0.044	0.046	0.080
		0.6	0.515	0.005	-11.42	0.074	0.046	0.050	0.121
		0.9	-0.932	-0.159	-17.71	0.091	0.047	0.055	0.182
11	100	0.0	-0.014	-0.048	-1.051	0.036	0.035	0.035	0.036
		0.3	-0.033	-0.104	-1.929	0.037	0.036	0.036	0.040
		0.6	-0.047	-0.125	-3.739	0.039	0.035	0.037	0.050
		0.9	0.547	-0.193	-7.937	0.063	0.033	0.040	0.086
11	200	0.0	0.011	-0.018	-1.021	0.025	0.025	0.025	0.027
		0.3	-0.008	-0.031	-1.868	0.026	0.026	0.026	0.031
		0.6	-0.039	-0.080	-3.689	0.028	0.025	0.026	0.044
		0.9	0.743	-0.080	-7.816	0.053	0.023	0.028	0.081
21	100	0.0	-0.010	0.001	-0.250	0.023	0.024	0.024	0.024
		0.3	-0.025	-0.070	-0.534	0.024	0.023	0.023	0.024
		0.6	-0.081	-0.042	-1.024	0.022	0.022	0.022	0.024
		0.9	0.247	-0.103	-3.159	0.031	0.017	0.020	0.036
21	200	0.0	0.027	0.016	-0.234	0.017	0.017	0.017	0.017
		0.3	-0.025	-0.012	-0.476	0.017	0.017	0.017	0.017
		0.6	-0.056	-0.040	-1.015	0.015	0.015	0.016	0.018
		0.9	0.085	-0.048	-3.108	0.020	0.012	0.014	0.033
51	50	0.0	0.020	0.021	-0.020	0.021	0.020	0.020	0.020
		0.3	-0.033	-0.024	-0.098	0.020	0.020	0.020	0.020
		0.6	-0.046	-0.046	-0.206	0.017	0.017	0.017	0.017
		0.9	-0.075	-0.070	-0.762	0.013	0.012	0.013	0.014

Table 10: bias and rmse of $\hat{\rho}_{RFEML}$, $\hat{\rho}_{FDML}$, $\hat{\rho}_{BCWG1}$ and $\hat{\rho}_{BCWG2}$.

4 Concluding remarks

In this paper we have studied Maximum Likelihood based estimation and inference procedures for the covariance stationary panel AR(1)/Unit Root model with fixed effects.

We have shown that the First Difference ML estimator for this model is consistent and has a limiting normal distribution whether one employs large N , fixed T asymptotics, joint N, T asymptotics or in case $|\rho| < 1$ large T , fixed N asymptotics. Under large N , fixed T asymptotics the FDMLE does not attain the (generalized) Cramér-Rao efficiency lower bound for estimators of ρ in the covariance stationary panel AR(1) model with fixed effects. Furthermore, under covariance stationarity and normality of the data, the FDMLE

and the Fixed Effects MLE for the covariance stationary panel AR(1) model are asymptotically equivalent to the WG estimator for ρ when $T \rightarrow \infty$ regardless of whether N tends to infinity or remains fixed but, unlike the last two estimators, the FDMLE for ρ does not exhibit a bias term in its large T , arbitrary N asymptotic distribution. Finally, the results of Monte Carlo simulations for panels of various dimensions show that under covariance stationarity and normality of the data the FDMLE for ρ has better finite sample properties than the bias-corrected WG estimators due to Hahn and Kuersteiner (2002) and Bun and Carree (2005).

We have also presented a Wald-type panel UR test that is based on the FDMLE for the panel AR(1)/UR model. A Monte Carlo study revealed that for panels of very different dimensions our FDMLE based UR test has considerably higher power against covariance stationary alternatives as well as various important non-stationary alternatives (with $|\rho| < 1$) than the WG estimator based UR test, the Restricted FE(Q)MLE based UR test and a UR test due to Levin et al. but only slightly higher power against covariance stationary alternatives than the UR test due to Breitung and Meyer.

Most of the results that we have obtained for the estimators of the panel AR(1)/UR model can be generalized to estimators for more general models. Under covariance stationarity and normality of the data the FDMLE for the panel ARMA model has the same large T , arbitrary N asymptotic properties as the MLE for the covariance stationary panel ARMA model with known individual effects. The large T , arbitrary N asymptotic properties of the latter estimator are similar to the asymptotic properties of the MLE for the covariance stationary ARMA model for a single time-series, which are discussed in Pierce (1971). When T is fixed the results for the FDMLE of the panel AR(1)/UR model can be extended to FDMLE's for more general panel ARMA models with possibly an autoregressive unit root by using some results on the covariance matrices of ARMA errors that can be found in, for instance, Van der Leeuw (1994). Furthermore, when T grows large the results for the FDMLE of the panel AR(1)/UR model can be extended to FDMLE's for more general panel AR(p) models with possibly a unit root.

Finally, the Wald-type panel UR test that is based on the FDMLE for the panel AR(1)/UR model can be modified so as to allow for individual drift parameters, heteroskedasticity, and general possibly heterogeneous AR(p) and/or MA(q) dependence. When the data exhibit heterogeneous AR(p) dependence under the null hypothesis, a limit distribution of the appropriate FDMLE based UR test-statistic can be derived under sequential asymptotics where first $T \rightarrow \infty$ and subsequently $N \rightarrow \infty$. The derivation of this limit distribution as well as the limit distributions of the test-statistic under other asymptotic plans, and the potential effect of lag order selection on these limit distributions are topics for further research.

A PROOFS

A.1 Two lemmata

We first show that $E(Dy_i y_i' D) = \sigma^2 DV(\rho)D' = \sigma^2 R(\rho)$ still holds when $\rho = 1$, even though $V(\rho)$ does not exist when $\rho = 1$.

Lemma 6 *Let $T \geq 3$ and let $\rho = 1$. Then $E(Dy_i y_i' D) = \sigma^2 R(1)$, $i = 1, \dots, N$.*

PROOF OF LEMMA 6

Let $R = R(r)$. It is easily verified that $R_{i,i} = 2/(1+r) \forall i \in \{1, \dots, T-1\}$, $R_{i,j} = r^{i-j-1}(r-1)/(1+r) \forall i, j$ that satisfy $i > j \geq 1$, and $R_{i,j} = R_{j,i} \forall i, j \in \{1, \dots, T-1\}$. Hence $R(1) = I$.

Note that when $\rho = 1$, $\Delta y_{i,t} = \varepsilon_t$ and $E(\Delta y_{i,s} \Delta y_{i,t}) = E(\varepsilon_{i,s} \varepsilon_{i,t}) \forall s, t$. It follows that $E(Dy_i y_i' D | \rho = 1) = \sigma^2 I = \sigma^2 R(1)$. \square

Lemma 6 implies that the formula of the likelihood function of the differenced data, $\prod_{i=1}^N L(Dy_i; r, s^2)$, which is given in (13) below, is still valid when $\rho = 1$. The likelihood function $\prod_{i=1}^N L(Dy_i; r, s^2)$ is also differentiable with respect to r and s^2 at and in the neighbourhood of $\rho = 1$:

Lemma 7 *Let $T \geq 3$. Furthermore let $\Theta_r^e = (-1, 1 + \frac{2}{T-2})$ and let $\Theta^e = \Theta_r^e \times (0, \infty)$. Then (i) $|R(r)| > 0$ on Θ_r^e and (ii) $|R(r)|$, $R(r)^{-1}$ and the likelihood function of the first differences Dy_i , $i = 1, \dots, N$,*

$$\prod_{i=1}^N L(Dy_i; r, s^2) = (2\pi s^2)^{-\frac{N(T-1)}{2}} (|Dy_i' D|)^{-\frac{N}{2}} \exp\left(-\frac{1}{2s^2} \sum_{i=1}^N (y_i' D' (Dy_i D')^{-1} Dy_i)\right) \quad (13)$$

are infinitely many times differentiable with respect to r (θ) on Θ_r^e (Θ^e).

PROOF OF LEMMA 7

Recall that $|R(r)| = (T(1-r) + 2r)/(1+r)$ (cf Lancaster and Lindenhovius, 1996). It is easily verified that $|R(r)| > 0$ for any $r \in \Theta_r^e$. Moreover $|R(r)|$ is infinitely many times differentiable with respect to r as long as $r \neq -1$.

Note that the elements of $R(r)^{-1}$ are equal to ratios of the cofactors of $R(r)$ and $|R(r)|$. The elements of $R(r)$ are given in the proof of lemma 6. It is easily seen that the cofactors of $R(r)$ are infinitely many times differentiable with respect to r as long as $r \neq -1$. It follows that the elements of $R(r)^{-1}$ are infinitely many times differentiable with respect to r as long as $|R(r)| \neq 0$ and $r \neq -1$. The latter conditions are satisfied for any $r \in \Theta_r^e$.

The last claim in the lemma now follows straightforwardly from the results above. \square

A.2 Proof of theorem 1

To prove large N , fixed T consistency of the FDMLE for the covariance stationary panel AR(1)/UR model with fixed effects we verify the conditions of theorem 2.5. in Newey and McFadden (1994). The proof proceeds as follows:

We assume that $\theta_0 \in \bar{\Theta}$ where $\bar{\Theta}$ can be any compact subset of $(-1, 1] \times (0, \infty)$.

Lemma 7 implies that $l(Dy_i; r, s^2)$ is differentiable with respect to θ on $\Theta_r^e \times (0, \infty)$, where $\Theta_r^e = (-1, 1 + \frac{2}{T-2})$. It follows that $l(Dy_i; r, s^2)$ is continuous on $\bar{\Theta}$.

The last two steps of the proof are as follows:

For identification it suffices to show that $\Omega(\theta) \neq \Omega(\theta_0) \forall \theta \in \bar{\Theta}$ with $\theta \neq \theta_0$, because $(\Omega(\theta) \neq \Omega(\theta_0))$ if and only if $\theta \neq \theta_0 \Leftrightarrow (\Omega(\theta)^{-1} \neq \Omega(\theta_0)^{-1} \text{ iff } \theta \neq \theta_0)$. The latter equivalence follows trivially from the fact that $\Omega(\theta)^{-1}$ is the unique inverse of $\Omega(\theta) \forall \theta \in \bar{\Theta}$. Now, the $T - 1$ elements of the first column of the covariance matrix $\Omega(\theta)$ are all different functions of the 2 elements of θ . It follows that if $T - 1 \geq 2$, then $\Omega(\theta) \neq \Omega(\theta_0) \forall \theta \in \bar{\Theta}$ with $\theta \neq \theta_0$.

To show that $E(\sup_{(r,s^2) \in \bar{\Theta}} l(Dy_i; r, s^2)) < \infty$, we note that $R(r)^{-1}$ and $|R(r)|$ are differentiable and hence continuous on Θ_r^e by lemma 7. It follows that $\Omega(\theta)^{-1}$ and $|\Omega(\theta)|$ are bounded on the compact set $\bar{\Theta}$. Moreover, $E(Dy_i y_i' D') = \Omega(\theta_0)$ is bounded. We conclude that $E(\sup_{(r,s^2) \in \bar{\Theta}} l(Dy_i; r, s^2)) < \infty$. \square

A.3 Proof of theorem 2

We first establish asymptotic normality of the FDMLE and then derive its limiting covariance matrix:

A.3.1 Large N , fixed T asymptotic normality of the FDMLE

To establish large N , fixed T asymptotic normality of the FDMLE for the covariance stationary panel AR(1)/UR model with fixed effects we verify the conditions of theorem 3.3. in Newey and McFadden (1994), henceforth NMcf. The proof proceeds as follows:

First we note that the FDMLE is consistent under large N , fixed T asymptotics, see appendix A.2.

Next we assume that there exists a number $\delta_2 > 0$ such that $\sigma^2 \in [\delta_2, \frac{1}{\delta_2}]$. Furthermore, for any $\rho \in (-1, 1]$ we can define a compact set $\bar{\mathcal{N}}^\delta(\rho) = \{\tilde{r} \mid |\rho - \tilde{r}| \leq \delta(\rho)\}$ such that $\delta(\rho) > 0$ and $\bar{\mathcal{N}}^\delta(\rho) \subset (-1, 1 + \frac{2}{T-2})$. Note that the value of $\delta(\rho)$ may depend on ρ . Let $\bar{\mathcal{N}}^{\frac{1}{2}\delta}(\rho) = \{\tilde{r} \mid |\rho - \tilde{r}| \leq \frac{1}{2}\delta(\rho)\}$. Then condition 3.3(i) of NMcf can be shown to hold since $\theta_0 \in \text{interior}(\bar{\mathcal{N}}^{\frac{1}{2}\delta}(\rho) \times [\frac{1}{2}\delta_2, \frac{2}{\delta_2}])$.

Condition 3.3(ii) of NMcf holds because $L(Dy_i; r, s^2)$ is infinitely many times differentiable with respect to θ on Θ^e , where $\Theta^e = \Theta_r^e \times (0, \infty)$ with $\Theta_r^e = (-1, 1 + \frac{2}{T-2})$ (cf lemma 7). Furthermore $L(Dy_i; r, s^2) > 0$ on Θ^e .

It also follows from lemma 7 that $|R(r)| > 0$ on Θ_r^e and that $R(r)^{-1}$ and $|R(r)|$ and their derivatives with respect to r are bounded on any compact subset of Θ_r^e . These results together with the fact that Dy_i has finite moments by virtue of being Gaussian, and the fact that the standard normal density $\phi(v)$ is uniformly bounded imply that conditions (iii) and (v) of theorem 3.3 in NMcf are satisfied for a neighbourhood $\mathcal{N}(\theta_0)$ of θ_0 such that $\bar{\mathcal{N}}^{\frac{1}{2}\delta}(\rho) \times [\frac{1}{2}\delta_2, \frac{2}{\delta_2}] \subset \mathcal{N}(\theta_0) \subset \Theta^e$.

Lemma 7 and the fact that the Dy_i are Gaussian also imply existence of the information matrix $J = E[\{\partial l(Dy_i; \rho, \sigma^2)/\partial \theta_0\} \{\partial l(Dy_i; \rho, \sigma^2)/\partial \theta_0\}']$ for any $\theta_0 \in \Theta$, where $\Theta = (-1, 1] \times (0, \infty)$.

Let $\Omega = \Omega(\theta)$. Then the Hessian of $l(Dy_i; r, s^2)$ is given by $H(\theta, y_i) = \frac{1}{2} \sum_{1 \leq s, t < T} (\Omega_{s,t} - \Delta y_{i,s+1} \Delta y_{i,t+1}) \frac{\partial^2 [\Omega^{-1}]_{s,t}}{\partial \theta \partial \theta'} - \frac{1}{2} ([\frac{d \text{vec} \Omega}{d \theta'} |_\theta]' (\Omega^{-1} \otimes \Omega^{-1}) [\frac{d \text{vec} \Omega}{d \theta'} |_\theta])$, see Magnus (1978).

The second Bartlett identity holds because condition (iii) of theorem 3.3 and lemma 3.6 in NMcf allow us to interchange the order of differentiation and integration. Thus $J = -E[H(\theta_0, y_i)] = \frac{1}{2} ([\frac{d \text{vec} \Omega}{d \theta'} |_{\theta_0}]' (\Omega(\theta_0)^{-1} \otimes \Omega(\theta_0)^{-1}) [\frac{d \text{vec} \Omega}{d \theta'} |_{\theta_0}])$. Finally, $J = -E[H(\theta_0, y_i)]$ is nonsingular for any $\theta_0 \in \Theta$ iff $\frac{d \text{vec} \Omega}{d \theta'} |_{\theta_0}$ has full rank. Since Ω is a Toeplitz matrix, we have that $\frac{d \text{vec} \Omega}{d \theta'} |_{\theta_0}$ has full rank iff $\frac{d \Omega e_1}{d \theta'} |_{\theta_0}$ has full rank, where Ωe_1 is equal to the first column of Ω . In appendix A.2 it was argued that θ can be identified from the elements of the first column of Ω whenever $T - 1 \geq 2$. It follows that J is nonsingular iff $T \geq 3$.

We conclude that $\sqrt{N}(\hat{\theta}_{FDMLE} - \theta_0) \xrightarrow{d} N(0, J^{-1})$.

A.3.2 Large N, fixed T limiting distribution of the FDMLE

Above we have already established large N , fixed T asymptotic normality of the FDMLE. Here we derive the large N , fixed T limiting covariance matrix of the FDMLE, which is equal to minus the inverse of the expectation of the Hessian of $N^{-1} \sum_{i=1}^N l(Dy_i; r, s^2)$ evaluated at θ_0 .

Let $f(r) = \iota' V^{-1} \iota$, and $b(r)' = b' = \frac{\iota' V^{-1} \iota}{\iota' V^{-1} \iota}$. The log-likelihood function $\sum_{i=1}^N l_{CS,FE}(y_i; r, m, s^2)$ can be decomposed as $\sum_{i=1}^N l(Dy_i; r, s^2) + \sum_{i=1}^N l(\tilde{\mu}_i; r, m_i, s^2)$, where $l(\tilde{\mu}_i; r, m_i, s^2)$ is the log-likelihood function of $\tilde{\mu}_i = b(r)' y_i$:

$$\begin{aligned} l(\tilde{\mu}_i; r, m_i, s^2) = l(\tilde{\mu}_i) &= -\frac{1}{2} \log 2\pi - \frac{1}{2} \log s^2 + \frac{1}{2} \log f(r) - \frac{f(r)}{2s^2} (\tilde{\mu}_i - m_i)^2 \\ &= -\frac{1}{2} \log 2\pi - \frac{1}{2} \log s^2 + \frac{1}{2} \log f(r) - \frac{f(r)}{2s^2} (b(r)'(y_i - m_i \iota))^2. \end{aligned} \quad (14)$$

It follows that the (Expected) Hessian of $N^{-1} \sum_{i=1}^N l(Dy_i; \rho, \sigma^2)$ can be obtained from the (Expected) Hessians of $N^{-1} \sum_{i=1}^N l_{CS,FE}(y_i; \rho, \mu, \sigma^2)$ and $N^{-1} \sum_{i=1}^N l(\tilde{\mu}_i; \rho, \mu_i, \sigma^2)$, $i = 1, \dots, N$:

$$N^{-1} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta_0} = N^{-1} \sum_{i=1}^N \frac{\partial^2 l_{CS,FE}(y_i; r, \mu, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta_0} - N^{-1} \sum_{i=1}^N \frac{\partial^2 l(\tilde{\mu}_i; r, \mu_i, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta_0}.$$

The elements of the Expected Hessian of the (Marginal) log-likelihood function $l(\tilde{\mu}_i)$, *EHM*, are obtained as follows: ¹⁵

$$\begin{aligned} \frac{dl(\tilde{\mu}_i)}{dr} &= \frac{1}{2} \frac{f'(r)}{f(r)} - \frac{f'(r)}{2s^2} (b(r)'(y_i - m_i \iota))^2 - \frac{f(r)}{s^2} b(r)'(y_i - m_i \iota) \frac{db(r)'}{dr} (y_i - m_i \iota), \text{ and} \\ \frac{d^2 l(\tilde{\mu}_i)}{dr^2} &= \frac{1}{2} \left[\frac{f''(r)}{f(r)} - \left(\frac{f'(r)}{f(r)} \right)^2 \right] - \frac{f''(r)}{2s^2} (\tilde{\mu}_i - m_i)^2 - 2 \frac{f'(r)}{s^2} b(r)'(y_i - m_i \iota) \frac{db(r)'}{dr} (y_i - m_i \iota) \\ &\quad - \frac{f(r)}{s^2} \left(\frac{db(r)'}{dr} (y_i - m_i \iota) \right)^2 - \frac{f(r)}{s^2} b(r)'(y_i - m_i \iota) \frac{d^2 b(r)'}{dr^2} (y_i - m_i \iota). \end{aligned}$$

Now, we have $E \left(\frac{f''(\rho)}{2\sigma^2} (\tilde{\mu}_i - \mu_i)^2 \right) = \frac{1}{2} \frac{f''(\rho)}{f(\rho)}$, and $E \left(b(\rho)'(y_i - \mu_i \iota) \frac{db(\rho)'}{d\rho} (y_i - \mu_i \iota) \right) =$

$$E \left(b(\rho)' u_i u_i' \frac{db(\rho)'}{d\rho} \right) = \sigma^2 b(\rho)' V \frac{db(\rho)'}{d\rho} = \frac{\sigma^2}{f(\rho)} \iota' \frac{db(\rho)'}{d\rho} = 0, \text{ where } u_i = y_i - \mu_i \iota.$$

Likewise, $E \left(b(\rho)'(y_i - \mu_i \iota) \frac{d^2 b(\rho)'}{d\rho^2} (y_i - \mu_i \iota) \right) = 0$, and

$$\begin{aligned} E \left(\frac{f(\rho)}{\sigma^2} \left(\frac{db(\rho)'}{d\rho} (y_i - \mu_i \iota) \right)^2 \right) &= E \left(\frac{f(\rho)}{\sigma^2} \left(\frac{db(\rho)'}{d\rho} (u_i u_i') \frac{db(\rho)'}{d\rho} \right) \right) = E \left(f(\rho) \frac{db(\rho)'}{d\rho} V \frac{db(\rho)'}{d\rho} \right) = \\ &= \frac{1}{f(\rho)} \frac{d(\iota' V^{-1} \iota)}{d\rho} V \frac{d(V^{-1} \iota)}{d\rho} - 2 \frac{1}{f(\rho)} \frac{d(\iota' V^{-1} \iota)}{d\rho} V b(\rho) f'(\rho) + \frac{1}{f(\rho)} b(\rho)' V b(\rho) (f'(\rho))^2 = \\ &= \frac{1}{f(\rho)} \frac{d(\iota' V^{-1} \iota)}{d\rho} V \frac{d(V^{-1} \iota)}{d\rho} - 2 \left(\frac{1}{f(\rho)} \right)^2 \frac{d(\iota' V^{-1} \iota)}{d\rho} \iota f'(\rho) + \left(\frac{f'(\rho)}{f(\rho)} \right)^2 = \\ &= \frac{1}{f(\rho)} \frac{d(\iota' V^{-1} \iota)}{d\rho} V \frac{d(V^{-1} \iota)}{d\rho} - 2 \left(\frac{f'(\rho)}{f(\rho)} \right)^2 + \left(\frac{f'(\rho)}{f(\rho)} \right)^2 = \frac{1}{f(\rho)} \frac{d(\iota' V^{-1} \iota)}{d\rho} V \frac{d(V^{-1} \iota)}{d\rho} - \left(\frac{f'(\rho)}{f(\rho)} \right)^2. \end{aligned}$$

It follows that $E \left(\frac{d^2 l(\tilde{\mu}_i)}{d\rho^2} \right) = -\frac{1}{f(\rho)} \frac{d(\iota' V^{-1} \iota)}{d\rho} V \frac{d(V^{-1} \iota)}{d\rho} + \frac{1}{2} \left(\frac{f'(\rho)}{f(\rho)} \right)^2$.

Explicit expressions for the terms on the RHS of the latter equality can easily be obtained.

¹⁵The prime ' denotes a derivative in the case of a scalar function and transpose in case of a vector.

Note that V^{-1} has a very simple structure:

$$V^{-1}(\rho) = \begin{pmatrix} 1 & -\rho & 0 & \cdot & \cdot & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdot & \cdot & 0 \\ 0 & -\rho & 1+\rho^2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 1+\rho^2 & -\rho \\ 0 & 0 & \cdot & 0 & -\rho & 1 \end{pmatrix}, \quad \text{with } |\rho| < 1.$$

Hence $f(\rho) = \iota'V^{-1}\iota = (1-\rho)(T(1-\rho) + 2\rho)$ and $\frac{d(V^{-1}\iota)}{d\rho} = \zeta_1 + \zeta_2$, where $\zeta_1 = -2(1-\rho)(0 \ 1 \ 1 \ \dots \ 1 \ 1 \ 0)$ and $\zeta_2 = -(1 \ 0 \ 0 \ \dots \ 0 \ 0 \ 1)$. From $\zeta_1'V\zeta_1 = 4(1-\rho)(1+\rho)^{-1}(T-2 + 2\sum_{t=1}^{T-3}(T-2-t)\rho^t)$, $\zeta_1'V\zeta_2 = 4\rho(1-\rho^2)^{-1}(1-\rho^{T-2})$ and $\zeta_2'V\zeta_2 = 2(1-\rho^2)^{-1}(1+\rho^{T-1})$, it then follows that $\frac{d(\iota'V^{-1}\iota)}{d\rho}V\frac{d(V^{-1}\iota)}{d\rho} = \frac{2}{1-\rho^2}(\rho^{T-1} - 2(T-2)\rho^2 + 2T - 3)$. Finally $f'(\rho) = -2(T-1)(1-\rho) - 2\rho$.

The other elements of EHM are $E\left(\frac{d^2l(\tilde{\mu}_i)}{d\rho d\sigma^2}\right) = \frac{f'(\rho)}{2\sigma^2 f(\rho)}$, $E\left(\frac{d^2l(\tilde{\mu}_i)}{d(\sigma^2)^2}\right) = -\frac{1}{2\sigma^4}$, $E\left(\frac{d^2l(\tilde{\mu}_i)}{d\rho d\mu_i}\right) = 0$, $E\left(\frac{d^2l(\tilde{\mu}_i)}{d\sigma^2 d\mu_i}\right) = 0$, and $E\left(\frac{d^2l(\tilde{\mu}_i)}{d\mu_i^2}\right) = -\frac{f(\rho)}{\sigma^2}$.

We now derive the elements of the Expected Hessian corresponding to $l_{CS,FE}(y_i; \rho, \mu, \sigma^2)$, EHF .

Note that

$$l_{CS,FE}(y_i; \rho, \mu, \sigma^2) = l(y_i) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log \sigma^2 - \frac{1}{2} \log |V| - \frac{1}{2\sigma^2} u_i' V^{-1} u_i. \quad (15)$$

It is easily seen that $E\left(\frac{d(u_i' V^{-1} u_i)}{d\rho}\right) = -\sigma^2 \frac{2\rho}{1-\rho^2}$ and $E\left(\frac{d^2(u_i' V^{-1} u_i)}{d\rho^2}\right) = \sigma^2 \frac{2(T-2)}{1-\rho^2}$.

Moreover, since $|V| = (1-\rho^2)^{-1}$, we have $\frac{d|V|}{d\rho} = 2\rho(|V|)^2$ and $\frac{d^2|V|}{d\rho^2} = 2(|V|)^2 + 8\rho^2(|V|)^3$.

It follows that $E\left(\frac{d^2l(y_i)}{d\rho^2}\right) = -\left[\frac{(T-2)(1-\rho^2)+1+\rho^2}{(1-\rho^2)^2}\right]$ and $E\left(\frac{d^2l(y_i)}{d\rho d\sigma^2}\right) = -\frac{1}{\sigma^2} \frac{\rho}{1-\rho^2}$.

The other elements of EHF are $E\left(\frac{d^2l(y_i)}{d(\sigma^2)^2}\right) = -\frac{T}{2\sigma^4}$, $E\left(\frac{d^2l(y_i)}{d\rho d\mu_i}\right) = 0$, $E\left(\frac{d^2l(y_i)}{d\sigma^2 d\mu_i}\right) = 0$, and $E\left(\frac{d^2l(y_i)}{d\mu_i^2}\right) = -\frac{f(\rho)}{\sigma^2}$. Notice that the last three elements of EHF are the same as those of EHM because $\tilde{\mu}_i$, $i = 1, \dots, N$, are sufficient statistics. Notice also that both EHF and EHM are block-diagonal.

Thus the large N , fixed T limiting distribution of $\{\hat{\rho}_{FDML}, \hat{\sigma}_{FDML}^2\}$ is given by

$$\sqrt{N} \begin{pmatrix} \hat{\rho}_{FDML} - \rho \\ \hat{\sigma}_{FDML}^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N\left(0, (EHM_{\rho, \sigma^2} - EHF_{\rho, \sigma^2})^{-1}\right). \quad (16)$$

An explicit expression for $(EHM_{\rho, \sigma^2} - EHF_{\rho, \sigma^2})$ is given in (8)-(9). \square

A.4 Large T asymptotic properties of the FDMLE

Some general results:

We can write the first-difference log-likelihood function (6) as

$$\sum_{i=1}^N l(Dy_i; r, s^2) = -\frac{N(T-1)}{2} \ln s^2 - \frac{N}{2} \ln |DVD'| - \frac{1}{2s^2} \sum_{i=1}^N q(Dy_i; r),$$

where $q(Dy_i; r) = y_i' D' (DVD')^{-1} Dy_i$. (17)

Recall that $D'(DVD')^{-1}D = V^{-1} - V^{-1}\iota'\iota^{-1}/(\iota'V^{-1}\iota)$. Note also that $Dy_i = Du_i$ and $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$.

Thence we have:

$$q(Du; r) = -\frac{(1-r)}{(T(1-r)+2r)}(u_1^2 + 2u_1u_T + u_T^2 + 2(1-r)(u_1 + u_T) \sum_{t=2}^{T-1} u_t$$

$$+ (1-r)^2(\sum_{t=2}^{T-1} u_t)^2) + (1+r^2) \sum_{t=2}^{T-1} u_t^2 - 2r \sum_{t=2}^T u_t u_{t-1} + u_1^2 + u_T^2. \quad (18)$$

Differentiating $q(Du; r)$ and $\ln |DVD'| = \ln \left(\frac{T(1-r)+2r}{1+r} \right)$ with respect to r yields:

$$\frac{\partial q(Du; r)}{\partial r} = \frac{2}{(T(1-r)+2r)^2} (u_1^2 + 2u_1u_T + u_T^2 + 2(1-r)(u_1 + u_T) \sum_{t=2}^{T-1} u_t + (1-r)^2(\sum_{t=2}^{T-1} u_t)^2)$$

$$+ \frac{2(1-r)}{(T(1-r)+2r)} ((u_1 + u_T) \sum_{t=2}^{T-1} u_t + (1-r)(\sum_{t=2}^{T-1} u_t)^2) + 2r \sum_{t=2}^{T-1} u_t^2 - 2 \sum_{t=2}^T u_t u_{t-1}, \quad (19)$$

$$\frac{\partial^2 q(Dy_i; r)}{\partial r^2} = \frac{4(T-2)}{(T(1-r)+2r)^3} (u_1^2 + 2u_1u_T + u_T^2 + 2(1-r)(u_1 + u_T) \sum_{t=2}^{T-1} u_t + (1-r)^2(\sum_{t=2}^{T-1} u_t)^2)$$

$$- \frac{8}{(T(1-r)+2r)^2} ((u_1 + u_T) \sum_{t=2}^{T-1} u_t + (1-r)(\sum_{t=2}^{T-1} u_t)^2) - \frac{2(1-r)}{(T(1-r)+2r)} (\sum_{t=2}^{T-1} u_t)^2 + 2 \sum_{t=2}^{T-1} u_t^2, \quad (20)$$

$$\frac{\partial \ln |DVD'|}{\partial r} = \frac{-2(T-1)}{(1+r)(T(1-r)+2r)}, \quad (21)$$

and

$$\frac{\partial^2 \ln |DVD'|}{\partial r^2} = -\frac{4(T-1)((T-2)r-1)}{(1+r)^2(T(1-r)+2r)^2}. \quad (22)$$

A.4.1 Some results specific to the case $|\rho| < 1$:

Lemma 8 *Let $u_t = y_t - \mu$ and let $\{u_t\}$ be covariance stationary. Then*

$$\begin{aligned} E(u_t^2) &= \sigma^2(1 - \rho^2)^{-1}, \quad E(u_t^4) = 3\sigma^4(1 - \rho^2)^{-2}, \quad E(u_1 u_T) = \sigma^2(1 - \rho^2)^{-1} \rho^{T-1}, \\ E((u_1 + u_T) \sum_{t=2}^{T-1} u_t) &= 2\sigma^2(1 - \rho^2)^{-1} \rho(1 - \rho)^{-1}(1 - \rho^{T-2}), \\ E((\sum_{t=2}^{T-1} u_t)^2) &= T \frac{\sigma^2}{(1-\rho)^2} + o(T), \quad E((u_1 \sum_{t=2}^{T-1} u_t)^2) = O(T) \quad \text{and} \quad E((u_T \sum_{t=2}^{T-1} u_t)^2) = O(T). \\ \text{Furthermore, when } T \rightarrow \infty, & \quad (E((\sum_{t=2}^{T-1} u_t)^2))^{-1/2} \sum_{t=2}^{T-1} u_t \xrightarrow{d} N(0, 1) \quad \text{and} \\ T^{-1/2} \sum_{t=2}^T u_{t-1} \varepsilon_t & \xrightarrow{d} N(0, \sigma^4(1 - \rho^2)^{-1}). \end{aligned}$$

PROOF

We only prove the last five claims of lemma 8:

$$\begin{aligned} E((\sum_{t=2}^{T-1} u_t)^2) &= E(\sum_{s=2}^{T-1} u_s \sum_{t=2}^{T-1} u_t) = \\ E(\sum_{s=2}^{T-1} \sum_{t=2}^s u_s u_t) &+ E(\sum_{s=2}^{T-1} \sum_{t=s}^{T-1} u_s u_t) - E(\sum_{t=2}^{T-1} u_t^2) = \\ \sigma^2(1 - \rho^2)^{-1} (\sum_{s=2}^{T-1} \sum_{t=2}^s \rho^{s-t} &+ \sum_{s=2}^{T-1} \sum_{t=s}^{T-1} \rho^{t-s} - (T-2)) = \\ \sigma^2(1 - \rho^2)^{-1} (\sum_{s=2}^{T-1} (1 - \rho)^{-1} (1 - \rho^{s-1}) &+ \sum_{s=2}^{T-1} (1 - \rho)^{-1} (1 - \rho^{T-s}) - (T-2)) = \\ \sigma^2(1 - \rho^2)^{-1} (2(1 - \rho)^{-1} (T-2) - \rho(1 - \rho)^{-1} (1 - \rho^{T-2})) &- T + 2 = T \frac{\sigma^2}{(1-\rho)^2} + o(T). \end{aligned}$$

$$\begin{aligned} \text{Note that } E((u_1 \sum_{t=2}^{T-1} u_t)^2) &= E(u_1^2 \sum_{s=2}^{T-1} \sum_{t=2}^{T-1} u_s u_t) = \\ E(u_1^2 \sum_{s=2}^{T-1} \sum_{t=2}^s u_s u_t) &+ E(u_1^2 \sum_{s=2}^{T-1} \sum_{t=s}^{T-1} u_s u_t) - E(u_1^2 \sum_{t=2}^{T-1} u_t^2). \end{aligned}$$

$$\text{Moreover, if } t > 1, \text{ then } E(u_1^2 u_t^2) = \frac{\sigma^4}{1-\rho^2} \left(\frac{1-\rho^{2(t-1)}}{1-\rho^2} \right) + \rho^{2(t-1)} E(u_1^4).$$

$$\text{Hence } E(u_1^2 \sum_{s=2}^{T-1} \sum_{t=2}^s u_s u_t) = \sum_{s=2}^{T-1} \sum_{t=2}^s \rho^{s-t} E(u_1^2 u_t^2) = T \frac{\sigma^4}{(1-\rho^2)^2(1-\rho)} + o(T),$$

$$E(u_1^2 \sum_{s=2}^{T-1} \sum_{t=s}^{T-1} u_s u_t) = \sum_{s=2}^{T-1} \sum_{t=s}^{T-1} \rho^{t-s} E(u_1^2 u_s^2) = T \frac{\sigma^4}{(1-\rho^2)^2(1-\rho)} + o(T), \text{ and}$$

$$E(u_1^2 \sum_{t=2}^{T-1} u_t^2) = \sum_{t=2}^{T-1} E(u_1^2 u_t^2) = T \frac{\sigma^4}{(1-\rho^2)^2} + o(T).$$

It follows that $E((u_1 \sum_{t=2}^{T-1} u_t)^2) = O(T)$.

$$\begin{aligned} \text{Note that } E((u_T \sum_{t=2}^{T-1} u_t)^2) &= E(u_T^2 \sum_{s=2}^{T-1} \sum_{t=2}^{T-1} u_s u_t) = \\ E(u_T^2 \sum_{s=2}^{T-1} \sum_{t=2}^s u_s u_t) &+ E(u_T^2 \sum_{s=2}^{T-1} \sum_{t=s}^{T-1} u_s u_t) - E(u_T^2 \sum_{t=2}^{T-1} u_t^2). \end{aligned}$$

$$\text{Moreover, if } t < T, \text{ then } E(u_t^2 u_T^2) = \frac{\sigma^4}{1-\rho^2} \left(\frac{1-\rho^{2(T-t)}}{1-\rho^2} \right) + \rho^{2(T-t)} E(u_t^4).$$

$$\text{Hence } E(u_T^2 \sum_{s=2}^{T-1} \sum_{t=2}^s u_s u_t) = \sum_{s=2}^{T-1} \sum_{t=2}^s \rho^{s-t} E(u_t^2 u_T^2) = T \frac{\sigma^4}{(1-\rho^2)^2(1-\rho)} + o(T),$$

$$E(u_T^2 \sum_{s=2}^{T-1} \sum_{t=s}^{T-1} u_s u_t) = \sum_{s=2}^{T-1} \sum_{t=s}^{T-1} \rho^{t-s} E(u_s^2 u_T^2) = T \frac{\sigma^4}{(1-\rho^2)^2(1-\rho)} + o(T), \text{ and}$$

$$E(u_T^2 \sum_{t=2}^{T-1} u_t^2) = E(u_1^2 \sum_{t=2}^{T-1} u_t^2) = O(T).$$

We conclude that also $E((u_T \sum_{t=2}^{T-1} u_t)^2) = O(T)$.

The last two claims of lemma 8 follow from well-known central limit theorems for autoregressive processes, see e.g. section 5.5 of Anderson (1971). \square

Proof of theorem 3:

We first establish large T , arbitrary N consistency of the FDMLE when $|\rho| < 1$:

A priori we assume that $(\rho, \sigma^2) \in (-1, 1] \times [0, \infty)$, that is, we do not rule out the possibility that $\rho = 1$, even though in reality $|\rho| < 1$.

The likelihood equations implied by (6) are given by:

$$\frac{1}{NT} \sum_{i=1}^N \frac{\partial l(Dy_i; r, s^2)}{\partial s^2} = -\frac{(T-1)}{2Ts^2} + \frac{1}{2s^4} \frac{1}{NT} \sum_{i=1}^N q(Dy_i; r) = 0 \quad (23)$$

and

$$\frac{2}{NT} \sum_{i=1}^N \frac{\partial l(Dy_i; r, s^2)}{\partial r} = -\frac{1}{T} \frac{\partial \ln |DVD'|}{\partial r} - \frac{1}{s^2} \frac{1}{NT} \sum_{i=1}^N \frac{\partial q(Dy_i; r)}{\partial r} = 0. \quad (24)$$

Solving (23) for s^2 yields $s^2 = \frac{1}{T-1} \frac{1}{N} \sum_{i=1}^N q(Dy_i; r)$. Replacing s^2 in (24) by this expression gives:

$$\frac{1}{T} \frac{\partial \ln |DVD'|}{\partial r} \frac{1}{T-1} \frac{1}{N} \sum_{i=1}^N q(Dy_i; r) + \frac{1}{NT} \sum_{i=1}^N \frac{\partial q(Dy_i; r)}{\partial r} = 0. \quad (25)$$

Note that $\lim_{T \rightarrow \infty} T^{-1} \frac{\partial \ln |DVD'|}{\partial r} = 0$ for $|r| < 1$, while $\lim_{T \rightarrow \infty} T^{-1} \frac{\partial \ln |DVD'|}{\partial r} = -\frac{1}{2}$ for $r = 1$.

Let $\Upsilon_1(Du; r) = \frac{1}{(T(1-r)+2r)}(u_1^2 + 2u_1u_T + u_T^2 + 2(1-r)(u_1 + u_T) \sum_{t=2}^{T-1} u_t + (1-r)^2(\sum_{t=2}^{T-1} u_t)^2) + u_1^2 + u_T^2$ and $\Upsilon_2(Du; r) = \frac{1}{(T(1-r)+2r)}((u_1 + u_T) \sum_{t=2}^{T-1} u_t + (1-r)(\sum_{t=2}^{T-1} u_t)^2)$. Then lemma 8 implies that $T^{-1}\Upsilon_1(Du; r) = o_p(1)$ and $T^{-1}\Upsilon_2(Du; r) = o_p(1)$ for $-1 < r \leq 1$ and also that $T^{-1} \sum_{t=2}^T u_{t-1}\varepsilon_t = o_p(1)$.

Furthermore, we have $\text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T u_t^2 = E(u_t^2) = \sigma^2(1 - \rho^2)^{-1}$ by the ergodicity theorem.

Therefore, regardless of whether N is fixed or tends to infinity,

$$\text{p lim}_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N q(Dy_i; r) = ((1+r^2) - 2r\rho)\sigma^2(1 - \rho^2)^{-1}. \quad (26)$$

It follows that if $|r| < 1$, $\text{plim}_{T \rightarrow \infty} \frac{1}{T} \frac{\partial \ln |DVD'|}{\partial r} \frac{1}{T-1} \frac{1}{N} \sum_{i=1}^N q(Dy_i; r) = 0$ regardless of N , while if $r = 1$ $\text{plim}_{T \rightarrow \infty} \frac{1}{T} \frac{\partial \ln |DVD'|}{\partial r} \frac{1}{T-1} \frac{1}{N} \sum_{i=1}^N q(Dy_i; r) = -\sigma^2(1 + \rho)^{-1}$ regardless of N .

Note also that

$$\text{p lim}_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \frac{\partial q(Dy_i; r)}{\partial r} = 2(r - \rho)\sigma^2(1 - \rho^2)^{-1} \quad (27)$$

regardless of N .

We need to consider two cases:

1) $|r| < 1$: Equation (25) holds true if and only if $r = \rho$.

2) $r = 1$: Equation (25) is not satisfied for $r = 1$ since $-\sigma^2(1 + \rho)^{-1} + 2(1 - \rho)\sigma^2(1 - \rho^2)^{-1} \neq 0$.

Moreover, it follows from (26) that $\text{plim}_{T \rightarrow \infty} \frac{1}{T-1} \frac{1}{N} \sum_{i=1}^N q(Dy_i; \rho) = \sigma^2$ regardless of N .

Thus when $|\rho| < 1$, $(r, s^2)' = (\rho, \sigma^2)'$ is the unique solution to the likelihood equations implied by (6).

The large T , arbitrary N limiting distribution of the FDMLE when $|\rho| < 1$:

First consider the stochastic limit of the Hessian evaluated at θ^* , where $\theta_k^* = \lambda_k \hat{\theta}_k + (1 - \lambda_k) \theta_{0,k}$ for some $\lambda_k \in [0, 1]$ and $k = 1, 2$. The elements of the Hessian are given by:

$$\frac{1}{NT} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial r^2} = -\frac{1}{2T} \frac{\partial^2 \ln |D'VD'|}{\partial r^2} - \frac{1}{2s^2} \frac{1}{NT} \sum_{i=1}^N \frac{\partial^2 q(Dy_i; r)}{\partial r^2}, \quad (28)$$

$$\frac{1}{NT} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial r \partial s^2} = \frac{1}{2s^4} \frac{1}{NT} \sum_{i=1}^N \frac{\partial q(Dy_i; r)}{\partial r}, \text{ and} \quad (29)$$

$$\frac{1}{NT} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial (s^2)^2} = \frac{(T-1)}{2Ts^4} - \frac{1}{s^6} \frac{1}{NT} \sum_{i=1}^N q(Dy_i; r). \quad (30)$$

Note that since $\text{plim}_{T \rightarrow \infty} \hat{\rho} = \rho$ and $\text{plim}_{T \rightarrow \infty} \hat{\sigma}^2 = \sigma^2$ regardless of N , we have $\text{plim}_{T \rightarrow \infty} \theta^* = \theta_0$ regardless of N . Then it follows from (26) that $\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N q(Dy_i; r^*) = \sigma^2$ and hence $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial (s^2)^2} \Big|_{\theta=\theta^*} = -\frac{1}{2\sigma^4}$ regardless of N .

Let $\Upsilon_3(Du; r) = \frac{1}{(T(1-r)+2r)} (\sum_{t=2}^{T-1} u_t)^2$. Then lemma 8 implies that $T^{-1} \Upsilon_3(Du; r) = o_p(1)$ for $|r| < 1$. We have already seen in the consistency proof above that $T^{-1} \Upsilon_1(Du; r) = o_p(1)$ and $T^{-1} \Upsilon_2(Du; r) = o_p(1)$ for $|r| < 1$. Therefore, we obtain that $\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \frac{\partial^2 q(Dy_i; r)}{\partial r^2} \Big|_{r=r^*} = 2\sigma^2(1 - \rho^2)^{-1}$ regardless of N . Note also that $\lim_{T \rightarrow \infty} \frac{1}{T} \frac{\partial^2 \ln |D'VD'|}{\partial r^2} = 0$. It follows that $\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial r^2} \Big|_{\theta=\theta^*} = -(1 - \rho^2)^{-1}$ regardless of N .

Finally, it follows from (27) that $\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial r \partial s^2} \Big|_{\theta=\theta^*} = 0$ regardless of N .

Consider now the scaled score vector:

$$\text{We have } (NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \rho} = -\frac{1}{2} (N/T)^{1/2} \frac{\partial \ln |D'VD'|}{\partial r} \Big|_{r=\rho} - \frac{1}{2\sigma^2} (NT)^{-1/2} \sum_{i=1}^N \frac{\partial q(Dy_i; \rho)}{\partial \rho}.$$

$$\text{Note that } \rho \sum_{t=2}^{T-1} u_t^2 - \sum_{t=2}^T u_t u_{t-1} = \rho \sum_{t=2}^{T-1} u_t^2 - \sum_{t=2}^T \rho u_{t-1}^2 - \sum_{t=2}^T u_{t-1} \varepsilon_t = -\rho u_1^2 - \sum_{t=2}^T u_{t-1} \varepsilon_t.$$

Then we can write $\frac{\partial q(Du; \rho)}{\partial \rho} = \sum_{k=1}^4 \Psi_k(Du; \rho)$, where

$$\Psi_1(Du; \rho) = \frac{2}{(T(1-\rho)+2\rho)^2} (u_1^2 + 2u_1 u_T + u_T^2 + 2(1-\rho)(u_1 + u_T) \sum_{t=2}^{T-1} u_t + (1-\rho)^2 (\sum_{t=2}^{T-1} u_t)^2),$$

$$\Psi_2(Du; \rho) = \frac{2(1-\rho)T}{(T(1-\rho)+2\rho)} (T^{-1}(u_1 + u_T) \sum_{t=2}^{T-1} u_t + (1-\rho)(T^{-1/2} \sum_{t=2}^{T-1} u_t)^2),$$

$$\Psi_3(Du; \rho) = -2\rho u_1^2, \text{ and } \Psi_4(Du; \rho) = -2 \sum_{t=2}^T u_{t-1} \varepsilon_t.$$

Let $\tilde{\Psi}_k(Du; \rho) = \Psi_k(Du; \rho) - E\Psi_k(Du; \rho)$ for $k = 1, 2, 3$. Note that $\text{Var}(T^{-1/2} \tilde{\Psi}_1(Du; \rho)) = O(T^{-3})$, $\text{Var}(T^{-1/2} \tilde{\Psi}_2(Du; \rho)) = O(T^{-1})$ and $\text{Var}(T^{-1/2} \tilde{\Psi}_3(Du; \rho)) = O(T^{-1})$. Therefore, as $T \rightarrow \infty$ regardless of whether $N \rightarrow \infty$ or is fixed, we have $(NT)^{-1/2} \sum_{i=1}^N \sum_{k=1}^3 \tilde{\Psi}_k(Du_i; \rho) = o_p(1)$.

Let $R_{N,T} = -\frac{1}{2} \left(\frac{N}{T} \right)^{1/2} \left(\frac{\partial \ln |D'VD'|}{\partial r} \Big|_{r=\rho} + \sigma^{-2} \sum_{k=1}^3 E\Psi_k(Du_i; \rho) \right)$. Note that $E\Psi_1(Du_i; \rho) = \frac{2\sigma^2}{(T(1-\rho)+2\rho)(1-\rho)}$. We also have $\left(\frac{\partial \ln |D'VD'|}{\partial r} \Big|_{r=\rho} + \sigma^{-2} \sum_{k=2}^3 E\Psi_k(Du_i; \rho) \right) = -\frac{2}{(1-\rho)(T(1-\rho)+2\rho)}$. It follows that $R_{N,T} = 0$.

Finally note that $E\Psi_4(Du_i; \rho) = 0$ and $T^{-1/2} \sum_{t=2}^T u_{t-1}\varepsilon_t \xrightarrow{d} N(0, \sigma^4(1-\rho^2)^{-1})$ (see lemma 8). We conclude that $(NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \rho} \xrightarrow{d} N(0, (1-\rho^2)^{-1})$ as $T \rightarrow \infty$ regardless of whether $N \rightarrow \infty$ or is fixed.

We can write $(NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \sigma^2} = -\left(\frac{N}{T}\right)^{1/2} \frac{(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4} (NT)^{-1/2} \sum_{i=1}^N q(Dy_i; \rho) = \frac{1}{2\sigma^2} (NT)^{-1/2} \sum_{i=1}^N \left(\frac{1}{\sigma^2} q(Dy_i; \rho) - (T-1)\right)$.

Note that $\sigma^{-2}q(Dy_i; \rho) = \sigma^{-2}u_i'D'(DVD')^{-1}Du_i$ is equal to the sum of $T-1$ independent $\chi^2(1)$ r.v.'s so that $\sigma^{-2}q(Dy_i; \rho) \sim \chi^2(T-1)$.

Therefore, as $T \rightarrow \infty$ regardless of whether $N \rightarrow \infty$ or is fixed, we have $(NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \sigma^2} \xrightarrow{d} N(0, \frac{1}{2\sigma^4})$ and $Cov((NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \rho}, (NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \sigma^2}) \rightarrow 0$. \square

A.4.2 Some results specific to the case $\rho = 1$:

When $\rho = 1$ it is useful to rewrite the expressions for $q(Du; r)$, $\partial q(Du; r)/\partial r$, and $\partial^2 q(Du; r)/\partial r^2$ given in (18)-(20). Note that when $\rho = 1$, $Du_i = \varepsilon_i$ and $u_{i,t} = u_{i,t-1} + \varepsilon_{i,t} = u_{i,1} + \sum_{s=2}^t \varepsilon_{i,s}$. Therefore if in $q(Du; r)$ we replace u_t by $u_1 + \sum_{s=2}^t \varepsilon_s$, for $t = 2, \dots, T$, then all terms involving u_1 would cancel out because $q(Du; r)$ only depends on ε . Thus when $\rho = 1$ we can assume w.l.o.g. that $u_1 = 0$. Then after replacing $u_t u_{t-1}$ by $u_{t-1}^2 + u_{t-1}\varepsilon_t$ in one of the terms in (18), we obtain:

$$q(Du; r) = -\frac{(1-r)}{(T(1-r)+2r)}(u_T^2 + 2(1-r)u_T \sum_{t=2}^{T-1} u_t + (1-r)^2(\sum_{t=2}^{T-1} u_t)^2) + (1-r)^2 \sum_{t=2}^{T-1} u_t^2 - 2r \sum_{t=2}^T u_{t-1}\varepsilon_t + u_T^2, \quad (31)$$

$$\begin{aligned} \frac{\partial q(Du; r)}{\partial r} &= \frac{2}{(T(1-r)+2r)^2}(u_T^2 + 2(1-r)u_T \sum_{t=2}^{T-1} u_t + (1-r)^2(\sum_{t=2}^{T-1} u_t)^2) + \\ &\frac{2(1-r)}{(T(1-r)+2r)}(u_T \sum_{t=2}^{T-1} u_t + (1-r)(\sum_{t=2}^{T-1} u_t)^2) - 2(1-r) \sum_{t=2}^{T-1} u_t^2 - 2 \sum_{t=2}^T u_{t-1}\varepsilon_t, \end{aligned} \quad (32)$$

and

$$\begin{aligned} \frac{\partial^2 q(Du; r)}{\partial r^2} &= -\frac{4(-T+2)}{(T(1-r)+2r)^3}(u_T^2 + 2(1-r)u_T \sum_{t=2}^{T-1} u_t + (1-r)^2(\sum_{t=2}^{T-1} u_t)^2) - \\ &\frac{8}{(T(1-r)+2r)^2}(u_T \sum_{t=2}^{T-1} u_t + (1-r)(\sum_{t=2}^{T-1} u_t)^2) - \frac{2(1-r)}{(T(1-r)+2r)}(\sum_{t=2}^{T-1} u_t)^2 + 2 \sum_{t=2}^{T-1} u_t^2 \end{aligned} \quad (33)$$

Lemma 9 Let $z_{1,i,T} = T^{-1}u_{i,T}^2$, $z_{2,i,T} = T^{-2}u_{i,T} \sum_{t=2}^{T-1} u_{i,t}$, $z_{3,i,T} = T^{-3/2} \sum_{t=2}^{T-1} u_{i,t}$, $z_{4,i,T} = T^{-1} \sum_{t=2}^T u_{i,t-1} \varepsilon_{i,t}$, and $z_{5,i,T} = T^{-2} \sum_{t=2}^{T-1} u_{i,t}^2$. Moreover, let $w_1 = \sigma^2 W^2(1)$, $w_2 = \sigma^2 W(1) \int_0^1 W(t) dt$, $w_3 = \sigma \int_0^1 W(t) dt$, $w_4 = (1/2)\sigma^2(W^2(1) - 1)$, and $w_5 = \sigma^2 \int_0^1 W^2(t) dt$. Then as $T \rightarrow \infty$, $z_{k,i,T} \xrightarrow{d} w_k$ for $k = 1, 2, 3, 4, 5$ and for all $i \in \{1, \dots, N\}$. Furthermore, $E(w_1) = \sigma^2$, $E(w_2) = \sigma^2/2$, $E(w_3^2) = \sigma^2/3$, $E(w_4) = 0$, and $E(w_5) = \sigma^2/2$.

PROOF

We only prove the claims with respect to $E(w_2)$ and $E(w_3)$:

$$E(w_2) = \sigma^2 E \int_0^1 W^2(t) dt = \sigma^2 \int_0^1 E(W^2(t)) dt = \sigma^2 \int_0^1 t dt = \sigma^2/2, \text{ while}$$

$$E(w_3^2) = \sigma^2 E((\int_0^1 W(t) dt)^2) = \sigma^2/3 \text{ in view of the fact that } \int_0^1 W(t) dt \sim N(0, 1/3). \quad \square$$

Proof of theorem 5:

We first establish large T , arbitrary N consistency of the FDMLE when $\rho = 1$:

A priori we assume that $(\rho \sigma^2) \in (-1, 1] \times [0, \infty)$.

The likelihood equations (23) and (24) imply that

$$\frac{1}{T} \frac{\partial \ln |DVD'|}{\partial r} - \frac{1}{T-1} \frac{1}{N} \sum_{i=1}^N q(Dy_i; r) + \frac{1}{NT} \sum_{i=1}^N \frac{\partial q(Dy_i; r)}{\partial r} = 0. \quad (34)$$

After multiplying (34) by $(T(1-r) + 2r)^2(1+r)$, setting $N = 1$, replacing Dy_i by Du , and substituting (31) and (32) for $q(Du; r)$ and $\partial q(Du; r)/\partial r$, respectively, we obtain:

$$\begin{aligned} & 2(1-r)(u_T^2 + 2(1-r)u_T \sum_{t=2}^{T-1} u_t + (1-r)^2 (\sum_{t=2}^{T-1} u_t)^2) - \\ & 2(T(1-r) + 2r)((1-r)^2 \sum_{t=2}^{T-1} u_t^2 - 2r \sum_{t=2}^T u_{t-1} \varepsilon_t + u_T^2) + \\ & 2(1+r)(u_T^2 + 2(1-r)u_T \sum_{t=2}^{T-1} u_t + (1-r)^2 (\sum_{t=2}^{T-1} u_t)^2) + \\ & 2(1-r)(1+r)(T(1-r) + 2r)(u_T \sum_{t=2}^{T-1} u_t + (1-r)(\sum_{t=2}^{T-1} u_t)^2) - \\ & 2(1-r)(1+r)(T(1-r) + 2r)^2 \sum_{t=2}^{T-1} u_t^2 - 2(1+r)(T(1-r) + 2r)^2 \sum_{t=2}^T u_{t-1} \varepsilon_t = 0. \end{aligned} \quad (35)$$

Let $x = 1 - r$. Then using (35) it can be shown that (34) is equivalent to:

$$N^{-1} \sum_{i=1}^N (g_{4,i,T} x^4 + g_{3,i,T} x^3 + g_{2,i,T} x^2 + g_{1,i,T} x + g_{0,i,T}) = 0, \quad (36)$$

where

$$\begin{aligned}
g_{0,i,T} &= -4z_{4,i,T}, \\
g_{1,i,T} &= -(Tz_{1,i,T} - 2z_{1,i,T} - 8Tz_{2,i,T} + 6Tz_{4,i,T} - 12z_{4,i,T} + 8Tz_{5,i,T}), \\
g_{2,i,T} &= -2(-T^2z_{2,i,T} + 3Tz_{2,i,T} - 3T^2z_{3,i,T}^2 + T^2z_{4,i,T} - 5Tz_{4,i,T} + 6z_{4,i,T} + 4T^2z_{5,i,T} - 9Tz_{5,i,T}), \\
g_{3,i,T} &= -(T^2z_{2,i,T} - 2Tz_{2,i,T} - 2T^3z_{3,i,T}^2 + 6T^2z_{3,i,T}^2 - T^2z_{4,i,T} + 4Tz_{4,i,T} \\
&\quad - 4z_{4,i,T} + 2T^3z_{5,i,T} - 11T^2z_{5,i,T} + 14Tz_{5,i,T}), \\
g_{4,i,T} &= (T - 2)(-T^2z_{3,i,T}^2 + T^2z_{5,i,T} - 2Tz_{5,i,T}). \tag{37}
\end{aligned}$$

It follows from lemma 9 that

$$\begin{aligned}
g_{1,i,T} &= -T(z_{1,i,T} - 8z_{2,i,T} + 6z_{4,i,T} + 8z_{5,i,T}) + O_p(1), \\
g_{2,i,T} &= -2T^2(-z_{2,i,T} - 3z_{3,i,T}^2 + z_{4,i,T} + 4z_{5,i,T}) + O_p(T), \\
g_{3,i,T} &= -2T^3(z_{5,i,T} - z_{3,i,T}^2) + O_p(T^2), \\
g_{4,i,T} &= T^3(z_{5,i,T} - z_{3,i,T}^2) + O_p(T^2). \tag{38}
\end{aligned}$$

Let $\tilde{g}_{k,i,T} = T^{-3}g_{k,i,T}$ for $k = 0, 1, 2, 3, 4$ and for all $i \in \{1, \dots, N\}$. Then (36) is equivalent to

$$N^{-1} \sum_{i=1}^N (\tilde{g}_{4,i,T}x^4 + \tilde{g}_{3,i,T}x^3 + \tilde{g}_{2,i,T}x^2 + \tilde{g}_{1,i,T}x + \tilde{g}_{0,i,T}) = 0. \tag{39}$$

Note that (a) $\tilde{g}_{k,i,T}$ are iid across i for all T and k ; (b) the $\tilde{g}_{k,i,T}$ are Gaussian and hence integrable for all T and k ; (c) $\text{plim}_{T \rightarrow \infty} \tilde{g}_{k,i,T} = 0$ for $k = 0, 1, 2$; as $T \rightarrow \infty$ $\tilde{g}_{3,i,T} \xrightarrow{d} -2(w_5 - w_3^2)$ and $\tilde{g}_{4,i,T} \xrightarrow{d} w_5 - w_3^2$ by lemma 9; and (d) the $\tilde{g}_{k,i,T}$ are Gaussian and hence uniformly L_2 -bounded and uniformly integrable in T for all i and k .¹⁶ Therefore $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N \tilde{g}_{k,i,T} = 0$ for $k = 0, 1, 2$; $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N \tilde{g}_{3,i,T} = -2E(w_5 - w_3^2)$; and $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N \tilde{g}_{4,i,T} = E(w_5 - w_3^2)$ by corollary 1 in Phillips and Moon (1999), which is henceforth referred to as PM. Finally lemma 9 also implies that $E(w_5 - w_3^2) = \sigma^2/6$.

Thus $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N (\tilde{g}_{4,i,T}x^4 + \tilde{g}_{3,i,T}x^3 + \tilde{g}_{2,i,T}x^2 + \tilde{g}_{1,i,T}x + \tilde{g}_{0,i,T}) = (x^4 - 2x^3)\sigma^2/6 = 0$.

Furthermore, let $\bar{u}_i = (T - 2)^{-1} \sum_{t=2}^{T-1} u_{i,t}$. Then it is easily seen that $z_{5,i,T} - z_{3,i,T}^2 = T^{-2} \sum_{t=2}^{T-1} (u_{i,t} - \bar{u}_i)^2 + 2T^{-3}(T - 2)\bar{u}_i^2 \geq 0$ and $(N^{-1} \sum_{i=1}^N (z_{5,i,T} - z_{3,i,T}^2))^{-1} = O_p(1)$.

Consequently, as $T \rightarrow \infty$, regardless of whether N is fixed or tends to infinity, $\hat{x} (= 1 - \hat{\rho})$ converges in probability to 0 or 2. Since we assumed that $-1 < \rho \leq 1$, it follows that $\text{plim}_{T \rightarrow \infty} \hat{\rho} = \rho = 1$ regardless of N .

¹⁶See e.g. Davidson (1994) for definitions of uniform L_p -boundedness and uniform integrability.

Let $\xi = Tx = T(1-r)$ and $\check{g}_{k,i,T} = T^{-k}g_{k,i,T}$ for $k = 0, 1, 2, 3, 4$ and for all $i \in \{1, \dots, N\}$. Then (36) is equivalent to

$$N^{-1} \sum_{i=1}^N (\check{g}_{4,i,T} \xi^4 + \check{g}_{3,i,T} \xi^3 + \check{g}_{2,i,T} \xi^2 + \check{g}_{1,i,T} \xi + \check{g}_{0,i,T}) = 0. \quad (40)$$

Now by corollary 1 of PM we have $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N \check{g}_{0,i,T} = E(-4w_4) = 0$, $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N \check{g}_{1,i,T} = E(-(w_1 - 8w_2 + 6w_4 + 8w_5)) = -\sigma^2$, $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N \check{g}_{2,i,T} = E(-2(-w_2 - 3w_3^2 + w_4 + 4w_5)) = -\sigma^2$, $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N \check{g}_{3,i,T} = E(-2(w_5 - w_3^2)) = -\sigma^2/3$, and $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N \check{g}_{4,i,T} = 0$.

Thus $\text{plim}_{N,T \rightarrow \infty} \hat{\xi}$ is a solution of $-(\xi^3 + 3\xi^2 + 3\xi)\sigma^2/3 = 0$. Note that $\xi^3 + 3\xi^2 + 3\xi = \xi(\xi^2 + 3\xi + 3) = 0 \Leftrightarrow \xi = 0$. Therefore $\text{plim}_{N,T \rightarrow \infty} \hat{\xi} = \text{plim}_{N,T \rightarrow \infty} T(1 - \hat{\rho}) = 0$.

Finally, let $0 < \delta < 1$ and impose $T(1 - \hat{\rho}) + 2\hat{\rho} > \delta$. This restriction on $\hat{\rho}$ implies that $\hat{\rho} < \frac{T-\delta}{T-2} = 1 + \frac{2-\delta}{T-2}$. It follows that $\text{plim}_{T \rightarrow \infty} \frac{(1-\hat{\rho})}{(T(1-\hat{\rho})+2\hat{\rho})} = 0$ regardless of N . Moreover $\text{plim}_{N,T \rightarrow \infty} T^{-1}((1-\hat{\rho})^2 \sum_{t=2}^{T-1} u_t^2 - 2\hat{\rho} \sum_{t=2}^T u_{t-1}\varepsilon_t + u_T^2) = \text{plim}_{N,T \rightarrow \infty} T^{-1}(T^2(1-\hat{\rho})^2 T^{-2} \sum_{t=2}^{T-1} u_t^2 + 2T(1-\hat{\rho})T^{-1} \sum_{t=2}^T u_{t-1}\varepsilon_t + \sum_{t=2}^T \varepsilon_t^2) = \sigma^2$, where we have used that $-2 \sum_{t=2}^T u_{t-1}\varepsilon_t + u_T^2 = \sum_{t=2}^T \varepsilon_t^2$.

We conclude that $\text{plim}_{N,T \rightarrow \infty} \hat{\sigma}^2 = \text{plim}_{N,T \rightarrow \infty} \frac{1}{T-1} \frac{1}{N} \sum_{i=1}^N q(Dy_i; \hat{\rho}) = \sigma^2$.

Joint N, T limiting distribution of the FDMLE when $\rho = 1$:

First we consider the stochastic limit of the Hessian evaluated at θ^* , where $\theta_k^* = \lambda_k \hat{\theta}_k + (1 - \lambda_k) \theta_{0,k}$ for some $\lambda_k \in [0, 1]$ and $k = 1, 2$:

Let $0 < \delta < 1$ and impose $T(1 - \hat{\rho}) + 2\hat{\rho} > \delta$. This restriction on $\hat{\rho}$ implies that $\hat{\rho} < \frac{T-\delta}{T-2} = 1 + \frac{2-\delta}{T-2}$.

Since $\text{plim}_{T \rightarrow \infty} \hat{\rho} = \rho = 1$ regardless of N , and $\text{plim}_{N,T \rightarrow \infty} \hat{\sigma}^2 = \sigma^2$, we have $\text{plim}_{T \rightarrow \infty} r^* = \rho = 1$ regardless of N , and $\text{plim}_{N,T \rightarrow \infty} \theta^* = \theta_0$. Furthermore, in view of $\text{plim}_{N,T \rightarrow \infty} T(1 - \hat{\rho}) = 0$, we have $\text{plim}_{N,T \rightarrow \infty} T(1 - r^*) = 0$. Finally, since $T(1 - \hat{\rho}) + 2\hat{\rho} > \delta$ and $T(1 - \rho) + 2\rho = 2 > \delta$, we also have $T(1 - r^*) + 2r^* > \delta$.

These results imply that $\text{plim}_{T \rightarrow \infty} \frac{(1-r^*)}{(T(1-r^*)+2r^*)} = 0$ regardless of N . Moreover, it is easily seen that $\text{plim}_{N,T \rightarrow \infty} T^{-1}((1-r^*)^2 \sum_{t=2}^{T-1} u_t^2 - 2r^* \sum_{t=2}^T u_{t-1}\varepsilon_t + u_T^2) = \text{plim}_{N,T \rightarrow \infty} T^{-1}(T^2(1-r^*)^2 T^{-2} \sum_{t=2}^{T-1} u_t^2 + 2T(1-r^*)T^{-1} \sum_{t=2}^T u_{t-1}\varepsilon_t + \sum_{t=2}^T \varepsilon_t^2) = \sigma^2$, where we have used that $-2 \sum_{t=2}^T u_{t-1}\varepsilon_t + u_T^2 = \sum_{t=2}^T \varepsilon_t^2$. It follows that $\text{plim}_{N,T \rightarrow \infty} (NT)^{-1} \sum_{i=1}^N q(Dy_i; r^*) = \sigma^2$. Noting that $-\frac{(T-1)}{2} \frac{\partial^2 \ln s^2}{\partial (s^2)^2} = \frac{(T-1)}{2} \frac{1}{s^4}$, we conclude that $\text{plim}_{N,T \rightarrow \infty} (NT)^{-1} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial (s^2)^2} \Big|_{\theta=\theta^*} = -\frac{1}{2\sigma^4}$.

Note that $\text{plim}_{N,T \rightarrow \infty} \frac{1}{T^2} \frac{\partial^2 \ln |DV D'|}{\partial r^2} \Big|_{r=r^*} = \lim_{T \rightarrow \infty} \frac{1}{T^2} \frac{-4(T-1)(T-3)}{16} = -\frac{1}{4}$. Furthermore, $\frac{\partial q(Dy_i; r)}{\partial r} = \frac{2}{(T(1-r)+2r)^2} (Tz_{1,i,T} + 2(1-r)T^2 z_{2,i,T} + (1-r)^2 T^3 z_{3,i,T}^2) - \frac{(1-r)}{(T(1-r)+2r)} (-2T^2 z_{2,i,T} - 2(1-r)T^3 z_{3,i,T}^2) - 2(1-r)T^2 z_{5,i,T} - 2Tz_{4,i,T}$, and $\frac{\partial^2 q(Dy_i; r)}{\partial r^2} = -\frac{4(-T+2)}{(T(1-r)+2r)^3} (Tz_{1,i,T} + 2(1-r)T^2 z_{2,i,T} + (1-r)^2 T^3 z_{3,i,T}^2) - \frac{8}{(T(1-r)+2r)^2} (T^2 z_{2,i,T} + (1-r)T^3 z_{3,i,T}^2) - \frac{2(1-r)}{(T(1-r)+2r)} T^3 z_{3,i,T}^2 + 2T^2 z_{5,i,T}$.

Corollary 1 of PM and lemma 9 imply that $\text{plim}_{N,T \rightarrow \infty} N^{-1} \sum_{i=1}^N z_{k,i,T} = E(w_k)$ for $k = 1, 2, 3, 4, 5$. It follows that $\text{plim}_{N,T \rightarrow \infty} T^{-2} N^{-1} \sum_{i=1}^N \frac{\partial^2 q(Dy_i; r)}{\partial r^2} \Big|_{r=r^*} = \sigma^2/2$ and hence $\text{plim}_{N,T \rightarrow \infty} T^{-2} N^{-1} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial r^2} \Big|_{\theta=\theta^*} = -\frac{1}{2} \times (-\frac{1}{4}) - \frac{1}{4} = -\frac{1}{8}$. It also follows that $\text{plim}_{N,T \rightarrow \infty} T^{-3/2} N^{-1} \sum_{i=1}^N \frac{\partial^2 l(Dy_i; r, s^2)}{\partial r \partial s^2} \Big|_{\theta=\theta^*} = \text{plim}_{N,T \rightarrow \infty} T^{-3/2} N^{-1} \sum_{i=1}^N \frac{1}{2(s^*)^4} \frac{\partial q(Dy_i; r)}{\partial r} \Big|_{r=r^*} = 0$.

Consider now the score vector:

We have $T^{-1} N^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \rho} \Big|_{\rho=1} = -\frac{1}{2} T^{-1} N^{1/2} \frac{\partial \ln |D V D'|}{\partial r} \Big|_{r=1} - \frac{1}{2\sigma^2} T^{-1} N^{-1/2} \sum_{i=1}^N \frac{\partial q(Dy_i; \rho)}{\partial \rho} \Big|_{\rho=1}$.

It is easily verified that $\frac{\partial q(Dy_i; \rho)}{\partial \rho} \Big|_{\rho=1} = \frac{\partial \tilde{q}(Dy_i; \rho)}{\partial \rho} \Big|_{\rho=1}$ where $\tilde{q}(Dy_i; r) = -\frac{(1-r)}{(T(1-r)+2r)} u_T^2 - 2r \sum_{t=2}^T u_{t-1} \varepsilon_t$.

Note that $\frac{\partial \tilde{q}(Dy_i; r)}{\partial r} = \frac{2}{(T(1-r)+2r)^2} u_T^2 - 2 \sum_{t=2}^T u_{t-1} \varepsilon_t$ and $2 \sum_{t=2}^T u_{t-1} \varepsilon_t = u_T^2 - \sum_{t=2}^T \varepsilon_t^2$. Therefore, we have $\frac{\partial q(Dy_i; \rho)}{\partial \rho} \Big|_{\rho=1} = \frac{1}{2} u_T^2 - 2 \sum_{t=2}^T u_{t-1} \varepsilon_t = -\frac{1}{2} u_T^2 + \sum_{t=2}^T \varepsilon_t^2$.

Furthermore note that $-\frac{1}{2} \frac{N^{1/2}}{T} \frac{\partial \ln |D V D'|}{\partial r} \Big|_{r=1} = \frac{N^{-1/2}(T-1)}{4T}$. It follows that $T^{-1} N^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \rho} = \Psi_5 + \Psi_6$, where $\Psi_5 = -\frac{1}{2\sigma^2} N^{-1/2} T^{-1} \sum_{i=1}^N \sum_{t=2}^T (\varepsilon_t^2 - \sigma^2)$ and $\Psi_6 = \frac{1}{4} N^{-1/2} T^{-1} (T-1) \sum_{i=1}^N (\frac{u_T^2}{(T-1)\sigma^2} - 1)$.

Since $u_T = \sum_{t=2}^T \varepsilon_t$ we have $(T-1)^{-1} \sigma^{-2} u_T^2 \sim \chi^2(1)$ and $E((T-1)^{-1} \sigma^{-2} u_T^2) = 1 \forall T$. Because of Gaussianity we also have that $((T-1)^{-1} \sigma^{-2} u_T^2 - 1)^2$ is uniformly integrable in T . Therefore if $N, T \rightarrow \infty$ jointly, then by theorem 3 in PM $\Psi_6 \xrightarrow{d} N(0, \frac{1}{8})$. Note also that $\Psi_5 = O_p(N^0 T^{-1/2})$. We conclude that if $N, T \rightarrow \infty$ jointly, then $T^{-1} N^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \rho} \xrightarrow{d} N(0, \frac{1}{8})$.

Note that $\frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \sigma^2} = -\frac{(T-1)}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} q(Dy_i; \rho) \Big|_{\rho=1} = -\frac{(T-1)}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} (-2 \sum_{t=2}^T u_{t-1} \varepsilon_t + u_T^2)$, and $(-2 \sum_{t=2}^T u_{t-1} \varepsilon_t + u_T^2) = \sum_{t=2}^T \varepsilon_t^2$. Hence $\frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \sigma^2} = \sum_{t=2}^T (\varepsilon_t^2 - \sigma^2) / 2\sigma^4$.

Clearly $T^{-1/2} \sum_{t=2}^T (\varepsilon_t^2 - \sigma^2) / 2\sigma^4 \xrightarrow{d} N(0, \frac{1}{2\sigma^4})$. Therefore if $N, T \rightarrow \infty$ jointly, then $T^{-1/2} N^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \sigma^2} \xrightarrow{d} N(0, \frac{1}{2\sigma^4})$.

Finally, since $u_T^2 = 2 \sum_{t=2}^T u_{t-1} \varepsilon_t + \sum_{t=2}^T \varepsilon_t^2$, we have $\lim_{T \rightarrow \infty} \text{Cov}((T-1)^{-1} \sigma^{-2} u_T^2, T^{-1/2} \sum_{t=2}^T (\varepsilon_t^2 - \sigma^2)) = 0$. It follows that $\lim_{N,T \rightarrow \infty} \text{Cov}(T^{-1} N^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \rho}, T^{-1/2} N^{-1/2} \sum_{i=1}^N \frac{\partial l(Dy_i; \rho, \sigma^2)}{\partial \sigma^2}) = 0$.

□

A.5 Various asymptotic results for the FEMLE for the covariance stationary panel AR(1) model

$l_{CS,FE}(y; r, m, s^2) = l(Dy; r, s^2) + \sum_{i=1}^N l(\tilde{\mu}_i; r, m_i, s^2)$, where $\tilde{\mu}_i = \frac{\iota' V^{-1} y_i}{\iota' V^{-1} \iota}$ and $l(\tilde{\mu}_i; r, m_i, s^2) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log s^2 + \frac{1}{2} \log \iota' V^{-1} \iota - \frac{1}{2\sigma^2} (\iota' V^{-1} \iota) (\tilde{\mu}_i - m_i)^2$. Furthermore, $\iota' V^{-1} \iota = (1-r)(T(1-r) + 2r)$.

Inconsistency of the FEMLE when \mathbf{T} is fixed:

We first concentrate the log-likelihood function $l_{CS,FE}(y; r, m, \sigma^2)$. For any value of T , $l_{CS,FE}(y; r, m, \sigma^2)$ is maximized at $m_i = \tilde{\mu}_i$, $i = 1, \dots, N$. Therefore the concentrated log-likelihood function $l_{CS,FE}(y; r, \tilde{\mu}, \sigma^2)$ divided by N equals $\frac{1}{N} l(Dy; r, s^2) - \frac{1}{2} \log(2\pi) - \frac{1}{2} \log s^2 + \frac{1}{2} \log \iota' V^{-1} \iota$.

The large N probability limits of the (scaled) concentrated likelihood equations corresponding to r and s^2 are not satisfied for the true values ρ and σ^2 , because the FDMLE for ρ and σ^2 is consistent and $\partial \log \iota' V^{-1} \iota / \partial \rho \neq 0$ and $\partial \log \sigma^2 / \partial \sigma^2 \neq 0$. It follows that the FEMLE for ρ and σ^2 is inconsistent.

Consistency of the FEMLE when T tends to infinity:

The concentrated log-likelihood function $l_{CS,FE}(y; r, \tilde{\mu}, s^2)$ divided by NT equals $\frac{1}{NT} l(Dy; r, s^2) - \frac{1}{2T} \log(2\pi) - \frac{1}{2T} \log s^2 + \frac{1}{2T} \log \iota' V^{-1} \iota$. As $T \rightarrow \infty$ regardless of whether $N \rightarrow \infty$ or remains fixed, the last three terms vanish. Thus when $T \rightarrow \infty$, the limiting criterion for the FEMLE is the same as the limiting criterion for the FDMLE. It follows that $\text{plim}_{T \rightarrow \infty} (\hat{\rho}_{FEMLE} - \hat{\rho}_{FDMLE}) = 0$ and $\text{plim}_{T \rightarrow \infty} (\hat{\sigma}_{FEMLE}^2 - \hat{\sigma}_{FDMLE}^2) = 0$. Since the FDMLE for ρ and σ^2 is consistent under large T , arbitrary N asymptotics, this must also be the case for the FEMLE for ρ and σ^2 .

Large T , arbitrary N asymptotic distribution of the FEMLE:

Theorem 10 *Provided that $|\rho| < 1$ and $N/T^3 \rightarrow 0$, the large T , arbitrary N asymptotic distribution of the*

$$FEMLE \text{ is given by } \sqrt{NT} \begin{pmatrix} \hat{\rho}_{FEMLE} - (\rho - \frac{1}{T}(1 + \rho)) \\ \hat{\sigma}_{FEMLE}^2 - (\sigma^2 - \frac{1}{T}\sigma^2) \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 - \rho^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \right).$$

PROOF

Note that $\lim_{T \rightarrow \infty} \frac{\iota' V^{-1} \iota}{T} = (1 - r)^2$. Therefore, also when $T \rightarrow \infty$, $(NT)^{-1} l_{CS,FE}(y; r, m, s^2)$ is still maximized at $m_i = \tilde{\mu}_i$, $i = 1, \dots, N$.

The scaled score vector of the concentrated log-likelihood function $l_{CS,FE}(y; r, \tilde{\mu}, s^2)$ can be written as

$$(NT)^{-1/2} \frac{\partial l_{CS,FE}(y; r, \tilde{\mu}, s^2)}{\partial \theta_0} = (NT)^{-1/2} \frac{\partial l(Dy; r, s^2)}{\partial \theta_0} + (NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(\tilde{\mu}_i; r, \tilde{\mu}_i, s^2)}{\partial \theta_0}.$$

$$\text{Let } \Xi_{NT} = E \left[(NT)^{-1/2} \frac{\partial l_{CS,FE}(y; r, \tilde{\mu}, s^2)}{\partial \theta_0} \right]. \text{ Then } \Xi_{NT} = (NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(\tilde{\mu}_i; r, \tilde{\mu}_i, s^2)}{\partial \theta_0}.$$

$$\text{Since } (NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(\tilde{\mu}_i; r, \tilde{\mu}_i, s^2)}{\partial \theta_0} = (N/T)^{1/2} \frac{\partial(-0.5 \log s^2 + 0.5 \log \iota' V^{-1} \iota)}{\partial \theta_0}, \quad \frac{\partial \log \iota' V^{-1} \iota}{\partial \rho} = \frac{-2(T-2)(1-\rho)-2}{(1-\rho)(T(1-\rho)+2\rho)},$$

and $\frac{\partial \log \sigma^2}{\partial \sigma^2} = \frac{1}{\sigma^2}$, we have $\Xi_{NT} = O(\sqrt{N/T})$.

The scaled Hessian of the concentrated log-likelihood function can be written as

$$(NT)^{-1} \frac{\partial^2 l_{CS,FE}(y; r, \tilde{\mu}, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^*} = (NT)^{-1} \frac{\partial^2 l(Dy; r, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^*} + (NT)^{-1} \sum_{i=1}^N \frac{\partial^2 l(\tilde{\mu}_i; r, \tilde{\mu}_i, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^*}, \text{ where}$$

$$\theta_k^* = \lambda_k \hat{\theta}_k + (1 - \lambda_k) \theta_{0,k} \text{ for some } \lambda_k \in [0, 1] \text{ and } k = 1, 2.$$

$$\text{From } (NT)^{-1} \sum_{i=1}^N \frac{\partial^2 l(\tilde{\mu}_i; r, \tilde{\mu}_i, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^*} = T^{-1} \frac{\partial^2(-0.5 \log s^2 + 0.5 \log \iota' V^{-1} \iota)}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^*} = O(T^{-1}),$$

we have $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \frac{\partial^2 l_{CS,FE}(y; r, \tilde{\mu}, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^*} = \text{plim}_{T \rightarrow \infty} (NT)^{-1} \frac{\partial^2 l(Dy; r, s^2)}{\partial \theta \partial \theta'} \Big|_{\theta=\theta^*}$, regardless of N , i.e.,

regardless of whether $N \rightarrow \infty$ or remains fixed.

It follows using results from appendix A.4 that $\sqrt{NT}(\hat{\theta} - \theta_0) + \left[E \left((NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta^*} \right) \right]^{-1} \Xi_{NT} + \tilde{R}_{NT} \xrightarrow{d} N(0, \Phi^{-1})$ as $T \rightarrow \infty$ regardless of N , where $\tilde{R}_{NT} = \left(\left[(NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta^*} \right]^{-1} - \left[E \left((NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta_0} \right) \right]^{-1} \right) \Xi_{NT}$ and $\Phi = -\text{plim}_{T \rightarrow \infty} (NT)^{-1} \frac{\partial^2 l(Dy;r,s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta^*} = \begin{pmatrix} 1 - \rho^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}^{-1}$. Moreover, $\theta^* - \theta_0 = O_p(1/\sqrt{NT})$ and hence $(NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta^*} - E \left((NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta_0} \right) = O_p(1/\sqrt{NT})$ as $T \rightarrow \infty$ regardless of N . Then we have by the delta method that $\left[(NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta^*} \right]^{-1} - \left[E \left((NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta_0} \right) \right]^{-1} = O_p(1/\sqrt{NT})$ and hence $\tilde{R}_{NT} = o_p(1)$ as $T \rightarrow \infty$ regardless of N .

We also have by (8) that $E \left((NT)^{-1} \frac{\partial^2 l(Dy;r,s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta_0} \right) = -\Phi + O(T^{-1})$ and hence $E \left((NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta_0} \right) = -\Phi + O(T^{-1})$. It follows that $\left[E \left((NT)^{-1} \frac{\partial^2 l_{CS,FE}(y;r,\tilde{\mu},s^2)}{\partial\theta\partial\theta'} \Big|_{\theta=\theta_0} \right) \right]^{-1} = -\Phi^{-1} + O(T^{-1})$. Finally, we have $\lim_{T \rightarrow \infty} \sqrt{T/N} \Xi_{NT} = \left(-\frac{1}{(1-\rho)} \quad -\frac{1}{2\sigma^2} \right)'$.

We conclude that the large T , arbitrary N asymptotic distribution of the FEMLE for ρ and σ^2 is identical to the large T , arbitrary N asymptotic distribution of the FDMLE for ρ and σ^2 , apart from the fact that the former has an asymptotic bias term. Assuming that $N/T^3 \rightarrow 0$ so that we only need to consider leading terms in the power expansions (in powers of T^{-1}) of the expected score vector and the inverse of the expected Hessian, this asymptotic bias term is equal to $-\frac{1}{T} \Phi^{-1} \left(\frac{1}{(1-\rho)} \quad \frac{1}{2\sigma^2} \right)' = -\frac{1}{T} (1 + \rho \quad \sigma^2)'$. \square

Acknowledgements: I thank two anonymous referees for helpful comments. All errors are solely mine. This research was supported by the ESRC through Grant R000239139.

References

- [1] Ahn, S.C. and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5-28.
- [2] Ahn, S.C. and P. Schmidt, 1997, Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation. *Journal of Econometrics* 76, 309-321.
- [3] Alvarez, J. and M. Arellano, 2003, The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71, 1121-1159.
- [4] Andersen, E.B., 1970a, Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society, Series B* 32, 283-301.

- [5] Andersen, E.B., 1970b, On Fisher's lower bound to asymptotic variances in case of infinitely many incidental parameters. *Skandinavisk ActuarTidskrift*, 78-85.
- [6] Anderson, T.W., 1971, *The statistical analysis of time series*. Wiley, New York.
- [7] Arellano, M. and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277-297.
- [8] Arellano, M., 2003, *Panel data econometrics*. Oxford University Press, Oxford, United Kingdom.
- [9] Binder, M., C. Hsiao and M.H. Pesaran, 2005, Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory* 21, 795-837.
- [10] Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.
- [11] Bond, S.R., C. Nauges and F. Windmeijer, 2005, Unit roots: identification and testing in micro panels. CEMMAP working paper CWP07/05, Institute for Fiscal Studies, London.
- [12] Breitung, J. and W. Meyer, 1994, Testing for unit roots using panel data: are wages on different bargaining levels cointegrated? *Applied Economics* 26, 353-361.
- [13] Breitung, J. and M.H. Pesaran, 2008, Unit roots and cointegration in panels, in: L. Mátyás and P. Sevestre (Eds.), *The econometrics of panel data: fundamentals and recent developments in theory and practice*. Springer, New York.
- [14] Bun, M.J.G. and M.A. Carree, 2005, Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics* 23, 200-210.
- [15] Bun, M.J.G. and J.F. Kiviet, 2006, The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132, 409-444.
- [16] Davidson, J., 1994, *Stochastic limit theory*. Oxford University Press, Oxford, United Kingdom.
- [17] Hahn, J. and G. Kuersteiner, 2002, Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639-1657.
- [18] Harris, R.D.F. and E. Tzavalis, 1999, Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91, 201-226.

- [19] Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel models covering short time periods. *Journal of Econometrics* 109,107-150.
- [20] Kiefer, N.M., 1980, Estimation of fixed effect models for time series of cross-sections with arbitrary intertemporal covariance. *Journal of Econometrics* 14, 195-202.
- [21] Kruiniger, H., 2001, On the estimation of panel regression models with fixed effects. Working paper no. 450, Queen Mary, University of London.
- [22] Lancaster, T. and B. Lindenhovius, 1996, Biases in dynamic panel data models: a reconsideration. Working paper no. 96-11, Brown University, Providence, Rhode Island.
- [23] Levin, A., C.-F. Lin and C.-S. J. Chu, 2002, Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics* 108, 1-24.
- [24] MaCurdy, T., 1981, Multiple time series models applied to panel data. Working paper no. W0646, NBER, Cambridge, Massachusetts.
- [25] MaCurdy, T., 1982, The use of time series processes to model the time structure of earnings in a longitudinal data analysis. *Journal of Econometrics* 18, 83-114.
- [26] Magnus, J.R., 1978, Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics* 7, 281-312.
- [27] Newey, W. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in: R.F. Engle and D. McFadden, (Eds.), *Handbook of econometrics*, Vol. 4. North-Holland, Amsterdam, pp. 2111-2245.
- [28] Neyman, J. and E.L. Scott, 1948, Consistent estimates based on partially consistent observations. *Econometrica* 16, 1-32.
- [29] Nickell, S., 1981, Biases in dynamic models with fixed effects. *Econometrica* 49, 1417-1426.
- [30] Phillips, P.C.B. and H. Moon, 1999, Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* 67, 1057-1111.
- [31] Pierce, D.A., 1971, Least squares estimation in the regression model with autoregressive-moving average errors. *Biometrika* 58, 299-312.

- [32] Van der Leeuw, J., 1994, The covariance matrix of ARMA errors in closed form. *Journal of Econometrics* 63, 397-405.

ACCEPTED MANUSCRIPT