Department of Economics

Maximum Likelihood Estimation of Dynamic Linear Panel Data Models with Fixed Effects

Hugo Kruiniger

Working Paper No. 458

June 2002

ISSN 1473-0278

Queen Mary
University of London

Maximum likelihood estimation of dynamic linear panel data models with fixed effects

Hugo Kruiniger*

Department of Economics,
Queen Mary, University of London

Previous version December 2000.

This version April 2002.

JEL classification: C11, C12, C13, C23.

Keywords: dynamic panel data models, fixed effects, Maximum Likelihood, GMM, Modified ML, Bayesian methods, redundancy, Cramér-Rao and Fisher efficiency bounds, unit root test.

*Address: H.KRUINIGER@QMUL.AC.UK. This paper is a revised and extended version of Queen Mary working paper no. 429. I should like to thank Richard Baillie, Richard Blundell, Grant Hillier, Hashem Pesaran, Jerzy Szroeter and Elias Tzavalis for helpful comments. All errors are solely mine. This research was supported by the European Commission through grant ERBFMBICT961594 and by the ESRC through grant R000239139.

Abstract

In this paper we consider inference procedures for two types of dynamic linear panel data models with fixed effects. First, we show that the closure of the stationary ARMA panel model with fixed effects can be consistently estimated by the First Difference Maximum Likelihood Estimator and we derive its large N , fixed T , as well as its large T , arbitrary N asymptotic distributions. These results allow us to formulate two simple likelihood based unit root tests which use the table of the standard normal distribution. We also establish the asymptotic properties of the Modified ML Estimator for the conditional AR(1) panel model with fixed effects under various asymptotic plans. Then we show that when $N \rightarrow \infty$ but T is fixed, the FDMLE does not attain the (generalised) Cramér-Rao lowerbound for the stationary AR(1) panel model with fixed effects and is asymptotically less efficient than the Random Effects MLE. However, when $T \rightarrow \infty$, the FDMLE, the Fixed Effects MLE, the REMLE are asymptotically equivalent to the Modified MLE and the Within ML (or LSDV) Estimator, i.e. under normality some of the moment conditions implicitly exploited by the first three estimators become asymptotically redundant.

1 Introduction

This paper studies Maximum Likelihood based estimation and testing procedures for two types of dynamic panel data models with fixed effects: the AR(1) panel model with fixed effects that conditions on the initial observations irrespective of the value of the autoregressive parameter, r , and the ARMA panel model with fixed effects that conditions on the initial observations when the individual time series contain a unit root but otherwise assumes stationarity. As the latter model is the closure of the stationary ARMA model, we will refer to it as the inclusive ARMA panel model with fixed effects. The former model will be referred to as the conditional AR(1) panel model with fixed effects.

Throughout the paper we will assume cross-sectional independence. Furthermore, we will assume that the errors are homoskedastic.

There is now a sizeable literature on GMM / IV estimation of dynamic panel data models, see e.g. Ahn and Schmidt (1995, 1997). 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. However, various Monte Carlo studies have revealed that the GMM estimators have poor finite sample properties in some cases. For instance, when r is close to one, they may suffer from weak instrument problems, see 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 GMM estimators becomes quite severe due to correlation between the estimated weight matrix and the sample moments, see Ziliak (1997).

The other major estimation principle, Maximum Likelihood, is generally not regarded as offering a real alternative to the Generalized Method of Moments in the case of dynamic panel data models, because ML estimators are subject to the incidental parameters problem, see Neyman and Scott (1948). In a well-known paper Nickell (1981) has proved that the (Within) ML estimator for the conditional AR(1) panel model with fixed effects is inconsistent when $N \rightarrow \infty$ and T is fixed, while Kiefer (1980) argues that the ML estimator for the covariance matrix in a static panel model with fixed effects is inconsistent when T remains fixed.

However, the situation for ML is not so bleak if one looks further. MaCurdy (1981a,

1982) suggested that the ML method does yield consistent estimators for stationary ARMA panel models with fixed effects when the data are first differenced. Kruiniger (2002) has argued that the conditional AR(1) panel model with fixed effects can only be consistently estimated if the initial observations and the individual effects, or equivalently the data, satisfy a particular condition. That paper subsequently derives the Fixed Effects Quasi ML estimator for the conditional AR(1) model under the assumption that the above mentioned specific condition is satisfied by the data, and it shows that this estimator is consistent.

Traditionally, the large N , fixed T asymptotic properties of panel data estimators were considered the most relevant 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 focused on the properties of estimators under various alternative asymptotic plans in which T grows large, see also Phillips and Moon (1999). Alvarez and Arellano (1998) and Hahn and Kuersteiner (2000) have shown that the Within ML (or LSDV) estimator is consistent under large T , arbitrary N asymptotics and diagonal path asymptotics, respectively.

In the first part of this paper we study the asymptotic properties of various Maximum Likelihood type estimators for the stationary/inclusive ARMA panel model with fixed effects. We first establish the large N , fixed T asymptotic properties of the First Difference ML estimator for the inclusive model in a rigorous manner.¹ To this end, we prove that the First Difference likelihood function corresponding to the inclusive ARMA model with fixed effects is differentiable by using results on covariance matrices of ARMA errors obtained by Van der Leeuw (1994). Thus this likelihood function is also differentiable at the unit root, despite the fact that the covariance matrix of the ARMA errors no longer exists at the unit root. Next, we derive the large T , arbitrary N asymptotic properties of the FDMLE.

One important result is that under any conceivable asymptotic plan, including the large T , fixed N plan, the FDMLE has a normal limiting distribution in the unit root case. This result leads to the formulation of two simple Wald-type (panel) unit root tests which use the table of the standard normal distribution. One of these unit root tests can also be used

¹MaCurdy (1981a, 1982) did not consider the unit root case.

when only a single time-series is available.

Under large T , arbitrary N asymptotics the FDMLE is asymptotically equivalent to the Fixed Effects MLE for the stationary AR(1) panel model with fixed effects. However, the latter estimator is inconsistent for fixed T and asymptotically biased under large T and arbitrary N asymptotics.

When T is fixed, the conditional AR(1) panel model with fixed effects can also be consistently estimated by the Modified Maximum Likelihood estimator, see Kruiniger (1998, 2002) and Lancaster (2001). In this paper we show that the Modified MLE is in fact consistent under any type of asymptotics. On the other hand, it is not clear whether the FEQMLE for the conditional model is also consistent under large T asymptotics. Therefore, one may prefer the Modified MLE over the FEQMLE, even though the FEQMLE is more efficient than the Modified MLE when T is fixed and the data are normally distributed (see Kruiniger (2002)). We also derive the large T , arbitrary N asymptotic distribution of the Modified MLE.

The FDMLE does not attain the generalized Cramér-Rao lowerbound for estimators for the stationary AR(1) panel model with fixed effects when $N \rightarrow \infty$ and T is fixed. Moreover, it is shown that under this kind of asymptotics the Random Effects MLE for r is more efficient than the FDMLE for r . Under large T and arbitrary N asymptotics, the FDMLE, the REMLE and the FEMLE for r in the stationary AR(1) model are asymptotically equivalent to the Modified MLE and the Within MLE for r in the conditional AR(1) model. This implies that some moment conditions that are implicitly exploited by the first three estimators, i.e. those involving levels of the data and those that are related to the initial conditions are asymptotically redundant relative to the moment conditions that are implicitly exploited by the Within ML estimator (see Breusch et al. (1999) for the notion of redundancy).

The paper is organised as follows. Section 2 examines the asymptotic properties of the First Difference ML estimators for the inclusive AR(1) panel model with fixed effects and for more general ARMA panel models with fixed effects. It also describes two very simple Wald-type (panel) unit root tests. Section 3 compares Fixed Effects and Random Effects ML and GMM estimators for the stationary AR(1) panel model with individual effects in terms of efficiency. Section 4 discusses the asymptotic properties of the Modified ML estimator for the conditional AR(1) panel model with fixed effects, and section 5 concludes.

2 ML estimation of the AR(1) panel model: the stationary case and the unit root case

In this section and section 4 we will consider the estimation of the autoregression parameter r in the AR(1) panel model with fixed effects

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

where $i \in \{1, 2, \dots, N\}$ indicates the individual unit and t indicates the time period. The i.i.d. assumption applies to both the time-series dimension and the cross-section dimension of the panel.

In this paper we will study the estimation of two versions of the AR(1) panel model with fixed effects. The versions differ with regard to the assumptions that are made about the initial conditions and the set of admissible values for r . In section 4 we will discuss the estimation of the AR(1) panel model with fixed effects when conditioning on the values of the initial observations. However, if the distribution of the initial observations is known, conditioning on these observations when estimating the parameters would entail a loss of information and as a consequence the estimator obtained would (in general) be less efficient than the most efficient estimator. An important case where the distribution of the initial observations is known is the situation where the data follow a stationary process. This assumption about the data is often satisfied. We will discuss ML type estimation methods for the stationary AR(1) panel model with fixed effects in this section. We will also study the unit root case even though that model belongs to the class of conditional AR(1) models, and we will discuss procedures for testing the unit root hypothesis against AR(1) models with $|r| < 1$.

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

STAT1: $|r| < 1$

STAT2: $y_{i,1} \sim N(\mu_i, \frac{\sigma^2}{1-r^2})$

Let us first consider unconditional ML estimation of the stationary model. Before we can define the unconditional (stationary, Fixed Effects) ML estimators, we have to introduce some additional notation. Let y_i denote the T -vector of all the observations for individual i including the first observation and let $\sigma^2 V_T(r)$ denote the covariance matrix of the y_i 's. Then we can restate the stationary version of model (1) as

$$(y_i - \mu_i \iota) \equiv u_i \sim N(0, \sigma^2 V_T(r)), \quad i = 1, \dots, N, \quad (i.i.d.), \quad (2)$$

where ι is a T -vector of ones and V_T is a covariance matrix of AR(1) errors:²

$$V_T(r) = \frac{1}{1-r^2} \begin{pmatrix} 1 & r & r^2 & \cdot & \cdot & \cdot & r^{T-1} \\ r & 1 & r & \cdot & \cdot & \cdot & r^{T-2} \\ r^2 & r & 1 & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & \cdot & r \\ r^{T-1} & r^{T-2} & \cdot & \cdot & \cdot & r & 1 \end{pmatrix} \quad \text{where } |r| < 1. \quad (3)$$

The unconditional ML estimators for the triple $\{r, \mu, \sigma^2\}$ in the stationary model are the statistics that maximize the following log-likelihood function:

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

Under large N , fixed T asymptotics, the unconditional ML estimators for r and σ^2 are inconsistent due to the incidental parameters problem (cf Neyman and Scott (1948) and Kiefer (1980)): see appendix A.5.

MaCurdy (1981a, 1982) has suggested to take first differences of the data, so that the individual effects drop out from the AR(1) panel model, before applying the ML method.

Let D denote the $(T-1 \times T)$ first-difference matrix with $D_{i,i} = -1$ and $D_{i,i+1} = 1$, $i = 1, \dots, T-1$, and $D_{i,j} = 0$ elsewhere. If we take first-differences of the observations in (2),

²We will drop the subscript T when knowledge of the time-dimension of the panel is unimportant.

We will also omit the argument r of V .

we obtain

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

The covariance matrix of Dy_i , $\sigma^2 DV D'$, has an ARMA(1,1) structure with the MA parameter equal to -1. Later on we will denote $\sigma^2 DV(r) D'$ by $\Omega(\theta)$, where $\theta = (r \ \sigma^2)'$.

The First Difference MLE (FDMLE) for the triple $\{r, \mu, \sigma^2\}$ globally maximizes^{3 4}

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

When $|r| < 1$, the unconditional likelihood function, (4), can be factorized as the product of the first-difference likelihood function, (6), and the N (marginal) likelihood functions of the sufficient statistics for the incidental parameters, $\tilde{\mu}_i = \frac{\iota' V^{-1} y_i}{\iota' V^{-1} \iota}$, $i = 1, \dots, N$.⁵ It follows that the FDML estimator for r and σ^2 is equal to the Conditional ML estimator for r and σ^2 .

Sofar, we have maintained that $|r| < 1$. We will now generalize the model by including the unit root case:

Definition 1 *The inclusive AR(1) panel model with fixed effects is characterized by (1) and the following assumptions:*

$$\text{INCL1:} \quad -1 < r \leq 1,$$

$$\text{INCL2:} \quad y_{i,1} \sim N(\mu_i, \frac{\sigma^2}{1-r^2}) \text{ if } -1 < r < 1; \text{ (p)lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 < \infty \text{ if } r = 1.$$

The unit root case is special because the model specified in (1) cannot be rewritten as (2) and also because the incidental parameters, i.e. the μ_i 's, drop out from the model. Nevertheless, the log-likelihood function of the first-differences (6) is still well-defined for $r = 1$. In particular, in appendix A.1 it will be shown that the matrix $DV D'$ and the likelihood of the first-differences can be differentiated as many times as is desired in (the neighbourhood of) $r = 1$.

³It is easy to show that when σ^2 is unknown the FDMLE for r in the AR(1) model with fixed effects only exists for $T \geq 3$.

⁴Hsiao et al. (1998) have also considered the FDMLE for the stationary model.

⁵See the earlier working paper version, Kruiniger (2000).

2.1 Asymptotic properties of the FDMLE for r

We will first present various results on the asymptotic properties of the FDMLE estimator of the traditional large N , fixed T variety but later on we will also discuss its large T , fixed N and its joint (N, T) asymptotic properties.⁶ Among other things, we will see that the FDMLE estimator is still consistent and asymptotically normal when $r = 1$, despite the fact that $r = 1$ is on the boundary of the parameter space and corresponds to nonstationary data.

Theorem 1 *The FDMLE estimator for r and σ^2 in the inclusive $AR(1)$ panel model with fixed effects, which maximizes (6), is consistent when $N \rightarrow \infty$ and T is fixed.*

PROOF

See appendix A.2. The proof is given for the more general $ARMA(p, q)$ panel model with fixed effects. The true values of the parameters are denoted by upperbars.

In what follows we will first consider the stationary case, $|r| < 1$, but at the end of this section we will return to the unit root case, $r = 1$.

Theorem 2 *The large N , fixed T asymptotic distribution of $(\hat{r}_{FDMLE}, \hat{\sigma}_{FDMLE}^2)$ is given by*

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

where

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

with

$$I_{rr} = \frac{-2(1+r)(r^{T-1} - 2(T-2)r^2 + 2T-3)(T(1-r)+2r)}{(1-r^2)^2(T(1-r)+2r)^2} + \frac{2(1+r)^2((T-2)(r-1)-1)^2 + ((T-2)(1-r^2)+1+r^2)(T(1-r)+2r)^2}{(1-r^2)^2(T(1-r)+2r)^2} \quad (9)$$

PROOF

See appendix A.3.

⁶See Phillips and Moon (1999) for a discussion of different types of convergence for panel data.

It can easily be verified that the asymptotic covariance matrix is positive definite symmetric for $-1 < r < 1$ (see also appendix A.3 and section 3).

Lemma 1 in Neyman and Scott (1948) generalizes the large N , fixed T Cramér-Rao lower bound (hereafter CR bound) to models with incidental parameters.⁷ The CR bound for (asymptotically) unbiased estimators for r and σ^2 is equal to $-(EHU_{r,\sigma^2})^{-1}$, where EHU_{r,σ^2} is the Expected Hessian (w.r.t. r and σ^2) of the Unconditional log-likelihood function. EHU_{r,σ^2} is given in appendix A.3.1. It is easy to verify that under large N , fixed T asymptotics the FDMLE for (r, σ^2) does not attain the CR bound for the inclusive AR(1) panel model with fixed effects; when T is fixed, the conditioning statistics, $\tilde{\mu}_i$, $i = 1, \dots, N$, are not ancillary for the common parameters, r and σ^2 .^{8 9} We will see in section 3 that more efficient root- N consistent (Random Effects) ML and GMM estimators exist for the inclusive AR(1) panel model with individual effects if $N^{-1} \sum_{i=1}^N \mu_i^2 = O(1)$ and $N^{-1} \sum_{i=1}^N y_{i,1}^2 = O(1)$.

Using the results of Phillips and Moon (1999), we can easily derive the large T and arbitrary N asymptotic properties of the FDMLE:

Theorem 3 *The FDML estimator for r and σ^2 in the stationary AR(1) panel model with fixed effects, which maximizes (6), is consistent regardless of whether $T \rightarrow \infty$, or $N \rightarrow \infty$, or $N, T \rightarrow \infty$ jointly. The large T and arbitrary N asymptotic distribution of $(\hat{r}_{FDML}, \hat{\sigma}_{FDML}^2)$ is given by*

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

PROOF

See appendix A.4.

⁷Andersen (1970b) derives the Fisher efficiency lower bound for 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.

⁸See appendix A.3.1: for instance, the term $\frac{2}{1-r^2}(r^{T-1})$ from $\iota' \frac{dV^{-1}}{dr} V \frac{dV^{-1}}{dr} \iota$ in EHM does not cancel out. Since the AR(1) model with fixed effects is just a special case of the ARMA(p,q) model with fixed effects, the FDMLE's for more general models are not asymptotically efficient either.

⁹Andersen (1970a) presents a necessary and sufficient condition for the equality of the variance of a Conditional MLE to the Cramér-Rao lower bound; see his 5th theorem.

The large T , fixed N asymptotic variance of $\hat{r}_{FDMLE}$ is equal to the large T , fixed N CR bound, $1 - r^2$. Furthermore, the joint (N, T) asymptotic variance of the FDMLE for r is equal to both $((N, T)$ and $(T, N))$ sequential limit CR efficiency bounds (and to the (N, T) sequential limit Fisher efficiency bound), as well as to the diagonal path asymptotic (i.e. with $N/T = \text{constant}$) efficiency bound for regular estimators for the *conditional* AR(1) panel model with fixed effects, which has been derived by Hahn and Kuersteiner (2000).¹⁰

When $T \rightarrow \infty$, the difference between the FDMLE (CMLE) for r and σ^2 and the unconditional (stationary, Fixed Effects) MLE for r and σ^2 converges in probability to zero; see appendix A.5. Thus although the unconditional MLE for r and σ^2 is inconsistent under large N , fixed T asymptotics, this estimator is consistent under large T and arbitrary N asymptotics, i.e. regardless of whether N is fixed or tends to infinity, just as in the single time-series case. The large T and arbitrary N asymptotic distribution of the FEMLE contains a bias term but is otherwise identical to the that of the FDMLE. The bias term for r is equal to the large T and arbitrary N asymptotic bias of the Least Squares Dummy Variables estimator (see Alvarez and Arellano (1998)).

The unconditional likelihood function for the stationary AR(1) panel model with *random* effects can be factorised in a way that is analogous to the factorisation that was used in the fixed effects case, i.e. as the product of the (marginal) likelihood function for the sufficient statistic for the common mean parameter (i.e. the constant term) and the first difference (or conditional) likelihood function. In appendix A.5. it is shown that when $T \rightarrow \infty$, the unconditional REMLE and FEMLE for the stationary AR(1) panel model are large- T asymptotically equivalent, regardless of whether N is fixed or tends to infinity. However, the former estimator is asymptotically unbiased.

We will now turn to the unit root case. We can obtain $\lim_{r \uparrow 1} V_{FDMLE}$ by applying de l'Hôpital's rule to (8). This leads to the following result:

¹⁰It is very easily seen that when $|r| < 1$, both the scaled score and Hessian of the joint likelihood function of the initial conditions vanish as $T \rightarrow \infty$, regardless of whether N is fixed or tends to infinity (where the scaling is by powers of NT). Thus the diagonal path asymptotic efficiency bound for the conditional model is also valid for the unconditional model, as the information contained in the initial observations becomes negligible when $T \rightarrow \infty$ (provided that $|r| < 1$).

Theorem 4 *When $r = 1$, the large N and fixed T asymptotic distribution of $(\hat{r}_{FDML}, \hat{\sigma}_{FDML}^2)$ is given by*

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

PROOF

See appendix A.3. and above. Notice that T is fixed and that the likelihood function is differentiable at $r = 1$.

The result in theorem 4 suggests a new unit root test for panel data models with fixed T which is simply a t-test that is based on the FDMLE for r and its large N , fixed T asymptotic variance, i.e. $8/[N(T-1)(T-2)]$.

Various unit root tests for panels with fixed T have already been suggested in the literature. Breitung and Meyer (1994) have proposed a t-test which is based on an estimator for r in the AR(1) panel model with fixed effects that, like the FDMLE, is consistent both when $r = 1$ and when $|r| < 1$. This estimator is based on deviations from the initial observations and has the same large N , fixed T asymptotic variance as the FDMLE. Harris and Tzavalis (1999) present a unit root test statistic for the AR(1) panel model with individual effects which is based on an adjusted version of the LSDV estimator. The adjustment corrects for the inconsistency of the LSDV estimator. The corrected estimator has a normal large N , fixed T asymptotic distribution with a variance that is approximately equal to $10.2/[N(T-1)(T-2)]$.

When $r = 1$, the large T , arbitrary N asymptotic distribution of $\hat{r}_{FDML}$ is very similar to its large N , fixed T asymptotic distribution:

Theorem 5 *When $r = 1$, the FDML estimator for r and σ^2 in the inclusive AR(1) panel model with fixed effects, which maximizes (6), is consistent whether $T \rightarrow \infty$, or $N \rightarrow \infty$, or $N, T \rightarrow \infty$ jointly. The large T and arbitrary N asymptotic distribution of $(\hat{r}_{FDML}, \hat{\sigma}_{FDML}^2)$ is given by*

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

PROOF

See appendix A.4. and above. The joint N, T limiting variance of $\hat{r}_{FDML}$ has been obtained by evaluating $\lim_{T \rightarrow \infty} 8T^2/[(T-1)(T-2)]$.

The large T and arbitrary N asymptotic distribution of the FDMLE for r is still normal when $r = 1$ even for $N = 1$, because the corresponding likelihood function only depends on Dy_i , which is still stationary in this case. Thus a very simple time-series unit root test, which uses the table of the standard normal distribution, is given by the t-test that is based on the FDMLE for r . To the best of our knowledge, this test has not been proposed in the literature so far.

Hahn and Kuersteiner (2000) have also derived the joint asymptotic distribution of the bias corrected LSDV estimator when $r = 1$. This asymptotic distribution is normal with the variance given by $10.2N^{-1}T^{-2}$. Thus when $r = 1$, the asymptotic variance of the $\hat{r}_{FDMLE}$ is about 80% of the asymptotic variance of the inconsistency or bias corrected LSDV estimator, both under large N , fixed T asymptotics and under joint asymptotics.

Since the FDMLE is consistent whether or not $r = 1$, the unit root test based on the FDMLE is consistent. Therefore, analytical asymptotic results on the power of the FDMLE based unit test against local stationary alternatives can easily be obtained. On the other hand, the large N , fixed T , inconsistency and the diagonal path (i.e. with $N/T = \text{constant}$) asymptotic bias of the LSDV estimator for r increase in r (see Nickell (1981), Alvarez and Arellano (1998), and Hahn and Kuersteiner (2000)). This reduces the power of the unit root test based on the LSDV estimator. Bond, Nauge, and Windmeijer (2002) present Monte Carlo results on the performance of the FDMLE and LSDV based unit root tests.

2.2 The First Difference ML estimator for the inclusive ARMA(p,q) panel model with fixed effects.

In this subsection we will consider estimation of more general inclusive ARMA(p,q) panel models with fixed effects. We will replace (1) by

$$\begin{aligned} A(L)(y_{i,t} - \mu_i) &= B(L)\varepsilon_{i,t}, \\ \check{\mu}_i &= A(1)\mu_i, \\ \varepsilon_{i,t} | \check{\mu}_i, y_{i,1}, \dots, y_{i,p} &\sim N(0, \sigma^2) \quad (i.i.d.), \quad t = 1 + p, \dots, T, \end{aligned} \tag{13}$$

where $A(L) = 1 - \sum_{k=1}^p r_k L^k$ and $B(L) = 1 + \sum_{k=1}^q r_{p+k} L^k$.

Let d denote the order of integration of the processes $\{y_{i,t}\}$, $i = 1, \dots, N$, then we can decompose $A(L)$ as $A(L) = (1 - L)^d \tilde{A}(L)$.

We will now define the inclusive ARMA(p,q) panel model with fixed effects:

Definition 2 *The inclusive ARMA(p,q) panel model with fixed effects is given by (13) and the following assumption:*

$$\begin{aligned} \text{INCL:} \quad & d \in \{0, 1\}; \\ & \{(1 - L)^d y_{i,t}\} \text{ is a stationary process;} \\ & \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 < \infty \text{ if } d = 1. \end{aligned}$$

The assumption about the maximal order of integration (d) can easily be relaxed.

Assumption INCL allows us to rewrite (13) as

$$\tilde{A}(L)D(y_{i,t} - \mu_i) = B(L)(1 - L)^{1-d}\varepsilon_{i,t}. \quad (14)$$

This in turn leads to

$$D(y_i - \mu_i) = Dy_i = Du_i \sim N(0, \sigma^2 DV D') \text{ i.i.d.}, \quad (15)$$

where $V (= V(r))$ is now a $(T \times T)$ covariance matrix of ARMA(p,q) errors; see, for instance, Van der Leeuw (1994) (henceforth VdL) for the structure of such a covariance matrix. The vector r contains the $p + q$ ARMA parameters. If $A(L)$ has a unit root (i.e. $d = 1$), then $\{Dy_{i,t}\}$ will be an ARMA(p-1,q) process instead of an ARMA(p,q+1) process. Moreover, when $d = 1$, the distribution of Dy_i is conditional on the value of $y_{i,1}$, as is the case for the AR(1) panel model with fixed effects (thus not conditional on $y_{i,1}, \dots, y_{i,-p+2}$!).

Before we consider various estimators for the inclusive ARMA(p,q) panel model with fixed effects, we will first introduce some additional notation. Let $Z_i = (Dy_i)(Dy_i)'$, $Z = \frac{1}{N} \sum_{i=1}^N Z_i$, and let $\bar{\Omega} = E(Z_i) = \Omega(\bar{\theta}) = \bar{\sigma}^2 DV(\bar{r})D'$, where $\Omega(\theta) = \Omega = \sigma^2 DV(r)D'$, $\theta = (r', \sigma^2)'$, and where upperbars denote the true values of the parameters. Moreover, let $\Theta = \Theta_r \times (0, \infty)$ denote the parameter space for θ . Finally, let $r^S = 1 - A(1)$.

It is shown in appendix A.1 that the first-difference log-likelihood function corresponding to the ARMA(p,q) panel model with fixed effects is infinitely many times differentiable in (the neighbourhood of) every $r \in \Theta_r$ with $r^S = 1$, when condition INCL is satisfied.

The FDMLE for $\theta, \hat{\theta}_{FDMLE}$, satisfies the following likelihood equations:

$$\begin{aligned} tr((\Omega(\theta) - Z)\frac{d\Omega(\theta)^{-1}}{dr_j}) &= 0, \quad j = 1, \dots, p+q, \\ tr((\Omega(\theta) - Z)\frac{d\Omega(\theta)^{-1}}{d\sigma^2}) &= 0, \end{aligned} \quad (16)$$

(see Magnus (1978)) or equivalently

$$[\frac{dvec\Omega}{d\theta'}]'(\Omega^{-1} \otimes \Omega^{-1})vec(Z - \Omega) = 0, \quad (17)$$

where we have used that $\frac{d\Omega^{-1}}{d\theta_j} = -\Omega^{-1}\frac{d\Omega}{d\theta_j}\Omega^{-1}$ and the general result that $trGWHW = (vecG)'(W \otimes W)vecH$. These likelihood equations are highly nonlinear in r .

We can now generalize theorems 2 and 3:

Theorem 6 *The FDMLE for θ in the inclusive ARMA(p, q) panel model with fixed effects is consistent when $N \rightarrow \infty$ while T remains fixed, and its large N and fixed T asymptotic distribution is given by*

$$\sqrt{N}(\hat{\theta}_{FDMLE} - \theta) \xrightarrow{d} N(0, 2([\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1})[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}])^{-1}) \quad (18)$$

PROOF

See appendices A.1-A.3 for details. We will assume that $\bar{\theta} \in \bar{\Theta}$, where $\bar{\Theta}$ is a compact subset of Θ . The set $\bar{\Theta}$ includes parameter vectors that correspond to the unit root case. MaCurdy (1981a) has observed that we need that $T > p+q+1$ for identification of all the parameters. Then DVD' has a unique ARMA($p, q+1$) structure (after deleting common divisors). It follows that $\Omega(\theta) \neq \Omega(\bar{\theta})$ if and only if $\theta \neq \bar{\theta}$ and $T > p+q+1$. Furthermore, $\Omega(\theta)$ is nonsingular and differentiable by INCL and results in VdL. It follows that $\log|\Omega(\theta)|$ and $\Omega(\theta)^{-1}$ are differentiable and $E(\sup_{\theta \in \bar{\Theta}} l(y_i|\tilde{\mu}_i; \theta)) < \infty$.

The asymptotic distribution of $\hat{\theta}_{FDMLE}$ can be obtained by Taylor-expanding (17) around $\bar{\theta}$ and using the second Bartlett identity (see als Magnus (1978)) $\square$.

The (Quasi) FDMLE for the more general ARMA(p, q) panel model with fixed effects can be used for testing the unit root hypothesis, $r^S = 1$.

In principle, we can also derive large T , arbitrary N asymptotic results for the FDMLE for the more general model analogous to those in theorem 3.

3 A comparison of the FDMLE and the REMLE for the stationary AR(1) panel model when T is fixed

In the previous section we have seen that the FDMLE does not attain the Cramér-Rao lowerbound for the stationary AR(1) panel model with fixed effects when T is finite. On the other hand, when $T \rightarrow \infty$, $N^{-1} \sum_{i=1}^N \mu_i^2 = O(1)$, and $N^{-1} \sum_{i=1}^N y_{i,1}^2 = O(1)$, the FDMLE for r in the stationary AR(1) panel model with individual effects is asymptotically equivalent to the REMLE for r . However, a direct comparison of the asymptotic variances of these ML estimators when T is fixed is rather cumbersome. Instead, we will compare two GMM estimators for the stationary AR(1) panel model that are asymptotically equivalent to the FDMLE and the REMLE respectively.

The Optimal Minimum Distance Estimator (OMDE) for the ARMA(p,q) panel model with fixed effects is defined as

$$\begin{aligned} \hat{\theta}_{OMD} = \arg \min_{\theta \in \Theta} [vech(Z - \Omega(\theta))] & \\ \times \left(\frac{1}{N} \sum_{i=1}^N ((vech Z_i)(vech Z_i)') - (vech Z)(vech Z)' \right)^{-1} & vech(Z - \Omega(\theta)), \end{aligned} \quad (19)$$

where $(\frac{1}{N} \sum_{i=1}^N ((vech Z_i)(vech Z_i)') - (vech Z)(vech Z)')$ is a consistent estimator of $E((vech(Z_i - \bar{\Omega}))(vech(Z_i - \bar{\Omega})))'$, which is the inverse of the (asymptotically) optimal weighting matrix (see Chamberlain (1982, 1984)).

The OMDE can be interpreted as a GMM estimator which is based on the following $0.5(T-1)T$ -vector of moment conditions:

$$E(vech(Z_i)) - vech(\Omega(\theta)) = 0. \quad (20)$$

The moment conditions in (20) summarize all the information in the second moments of the (first-)differenced data.

In appendix B.1 we examine the large N and fixed T asymptotic properties of the OMDE that is defined in (19). In particular, we show that under normality of the disturbances the OMDE and the FDMLE for the inclusive ARMA(p,q) panel model with fixed effects are asymptotically equivalent.

Ahn and Schmidt (1997) and Kruiniger (2000) have presented efficient GMM estimators for the stationary AR(1) panel model with random effects. These estimators are asymptotically equivalent but the Optimal Linear GMM estimator of Kruiniger has the advantage that it only exploits moment conditions that are linear in r . The OLGMM estimator optimally exploits the following $0.5T(T+1) - 2$ moment conditions which are implied by covariance stationarity of $\{y_{i,t}\}$, $i = 1, \dots, N$:

$$E(y_{i,t-s}\Delta\varepsilon_{i,t}) = 0 \text{ for } s = 2, \dots, t-1 \text{ and } t = 3, \dots, T, \quad (21)$$

where $\Delta\varepsilon_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$,

$$E((y_{i,t} - ry_{i,t-1})\Delta y_{i,t-1}) = 0 \text{ for } t = 3, \dots, T, \quad (22)$$

$$E(y_{i,t}\varepsilon_{i,t} - y_{i,t-1}\varepsilon_{i,t-1}) = 0 \text{ for } t = 3, \dots, T, \quad (23)$$

and

$$E((r-1)(\Delta y_{i,2})^2 - 2\Delta y_{i,2}\Delta y_{i,3}) = 0. \quad (24)$$

Since the OLGMM estimator exploits all the information about r that is contained in the second moments of the data, it is asymptotically equivalent to the REMLE for the stationary model under normality (cf Ahn and Schmidt (1995)). Moreover, the Optimal Linear GMM estimator is at least as efficient as the OMDE. It turns out that under normality the Optimal Linear GMM estimator is strictly more efficient than the OMDE when T is finite:

Theorem 7 *Under normality the Optimal Linear GMM estimator for r is asymptotically more efficient than the OMD estimator, i.e. the Optimal GMM estimator based on $E(\text{vech}(Z_i)) - \text{vech}(\Omega(\theta)) = 0$ when $|r| < 1$, $N \rightarrow \infty$ and T is fixed.*

PROOF

See appendix B.2. It is shown that under the assumption of normality the moment condition $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ is not redundant relative to $E(\text{vech}(Z_i)) - \text{vech}(\Omega(r, \sigma^2)) = 0$ when $|r| < 1$. The property of redundancy is defined by Breusch et al. (1999).

Under non-normality the OMDE is generally more efficient than the FDMLE, see MaCurdy (1981b). We conclude that when T is fixed, the Optimal Linear GMM estimator is generally more efficient than the FDMLE.

4 ML estimation of the conditional AR(1) panel model.

Sometimes we cannot make the stationarity assumption. For instance, when the initial observations $y_{i,1}$, $i = 1, 2, \dots, N$, are the start-up values of stochastic processes, they may have a distribution that is different from the distribution of the subsequent observations. In that case we may decide to condition on the initial observations, i.e. to treat the initial observations as given. The resulting *conditional* AR(1) panel model with fixed effects can be stated as

$$\begin{aligned} y_i &= r y_{i,-1} + \check{\mu}_i + \varepsilon_i \\ \check{\mu}_i &= (1 - r) \mu_i \\ \varepsilon_i | \check{\mu}_i, y_{i,1} &\sim N(0, \sigma^2 I_{T-1}) \text{ (i.i.d.)}, \end{aligned} \tag{25}$$

where $i \in \{1, 2, \dots, N\}$. For each individual unit, we again have T observations on y including the initial observation $y_{i,1}$.

4.1 Modified ML estimation of the conditional AR(1) panel model with fixed effects

It is well known that the Within MLE for the conditional AR(1) panel model with fixed effects equals the LSDV estimator and that this estimator is inconsistent under traditional large N , and fixed T asymptotics (see Nickell (1981)).

Lancaster (1997) has outlined a Bayesian approach which yields consistent estimators for r and σ^2 in the conditional AR(1) panel model with fixed effects. His approach involves two steps: first, a reparametrization of the model such that the ‘new’ fixed effects are information orthogonal to the common parameters r and σ^2 , and second, the new fixed effects in the likelihood are integrated out using a noninformative, uniform prior.

In the first step Lancaster obtains the following reparametrization

$$\begin{aligned} \check{\mu}_i &= (1 - r) y_{i,1} + \nu_i \exp(-\beta(r)), \quad r = r, \quad \text{and} \quad \sigma^2 = \sigma^2, \\ \text{where } \beta(r) &= \frac{1}{T-1} \sum_{t=1}^{T-2} \frac{T-1-t}{t} r^t, \end{aligned} \tag{26}$$

and where ν_i , $i = 1, 2, \dots, N$, are the new fixed effects. Then in the second step he derives the marginal posterior density of r and σ^2 , $p(r, \sigma^2) = \prod_{i=1}^N p_i(r, \sigma^2)$, as follows

$$\begin{aligned} p(r, \sigma^2) &\propto \pi(r, \sigma^2) \prod_{i=1}^N \int \sigma^{-T-1} \exp \left(-\frac{1}{2\sigma^2} \sum_{t=2}^T (y_{i,t} - ry_{i,t-1} - \check{\mu}_i)^2 \right) d\nu_i \\ &= \pi(r, \sigma^2) \exp(N\beta(r)) \prod_{i=1}^N \int \sigma^{-T-1} \exp \left(-\frac{1}{2\sigma^2} \sum_{t=2}^T (y_{i,t} - ry_{i,t-1} - \check{\mu}_i)^2 \right) d\check{\mu}_i \end{aligned} \quad (27)$$

$$\begin{aligned} &= \pi(r, \sigma^2) \exp(N\beta(r)) \times \\ &\quad \sigma^{-N(T-2)} \prod_{i=1}^N \exp \left(-\frac{1}{2\sigma^2} \sum_{t=2}^T (y_{i,t} - \bar{y}_i - r(y_{i,t-1} - \bar{y}_{i,-1}))^2 \right), \end{aligned} \quad (28)$$

where $\pi(r, \sigma^2)$ is a noninformative prior for r and σ^2 . Notice that in the prior density the new fixed effects and the common parameters are independent. Lancaster's estimator for r and σ^2 is given by a maximum of $p(r, \sigma^2)$. This estimator can also be viewed as a Modified Maximum likelihood Estimator.

Theorem 8 *If $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 < \infty$, then a maximum of (28) on $(-1, 1) \times (0, \infty)$ is a consistent estimator for r and σ^2 in the conditional AR(1) panel model, (25), with (r, σ^2) belonging to a compact subset of $(-1, 1) \times (0, \infty)$, when N tends to infinity and T is fixed.*

PROOF

See Kruiniger (1998, 2002). Lancaster (2001) proves consistency of this estimator assuming that $y_{i,1} - \mu_i$, $i = 1, \dots, N$ are i.i.d.

In appendix C.1, it is proven that Lancaster's two-step Bayesian procedure is equivalent to integrating the incidental parameters out of the model with the original parametrization with respect to Jeffrey's invariant noninformative prior.

The large N , fixed T asymptotic distribution of the Modified Maximum Likelihood Estimator for $(\hat{r}_{MML}, \hat{\sigma}_{MML}^2)$ can easily be obtained by employing a Taylor expansion of the estimator ¹¹.

¹¹The large N , fixed T asymptotic distribution of the Modified ML Estimator has been obtained independently by Lancaster (1997).

Both the expected Hessian and the expected outer product of the gradient of the modified log-likelihood function — which is given by (28) apart from the factor $\pi(r, \sigma^2)$ — are easily derived. Let us define

$$\check{P} = \begin{pmatrix} 0 & . & . & 0 & 0 & 0 \\ 1 & 0 & & & 0 & 0 \\ r & 1 & 0 & & & 0 \\ . & r & 1 & 0 & & . \\ . & & r & 1 & 0 & . \\ r^{T-3} & . & . & r & 1 & 0 \end{pmatrix}, \quad \check{\lambda}_1 = \begin{pmatrix} 1 \\ r \\ r^2 \\ \vdots \\ r^{T-3} \\ r^{T-2} \end{pmatrix} \quad (29)$$

then

$$y_{i,-1} - \mu_i \iota = \check{\lambda}_1 (y_{i,1} - \mu_i \iota) + \check{P} \varepsilon_i \quad (30)$$

The expected modified Hessian equals

$$EMH = \lim_{N \rightarrow \infty} \begin{pmatrix} \beta''(r) - \frac{1}{N} \sum_{i=1}^N \frac{E(y'_{i,-1} Q y_{i,-1})}{\sigma^2} & \frac{\beta'(r)}{\sigma^2} \\ \frac{\beta'(r)}{\sigma^2} & -\frac{T-2}{2\sigma^4} \end{pmatrix} \quad (31)$$

and the modified information matrix equals

$$MIM = \lim_{N \rightarrow \infty} \begin{pmatrix} tr(Q \check{P} Q \check{P}) + \frac{1}{N} \sum_{i=1}^N \frac{E(y'_{i,-1} Q y_{i,-1})}{\sigma^2} & -\frac{\beta'(r)}{\sigma^2} \\ -\frac{\beta'(r)}{\sigma^2} & \frac{T-2}{2\sigma^4} \end{pmatrix} \quad (32)$$

where we have used that $E(\varepsilon'_i Q y_{i,-1}) = \sigma^2 tr(Q \check{P}) = -\sigma^2 \beta'(r)$. It is easily seen that $E(y'_{i,-1} Q y_{i,-1}) = E((y_{i,-1} - \mu_i) Q (y_{i,-1} - \mu_i)) = \sigma^2 tr(\check{P}' Q \check{P}) + (y_{i,1} - \mu_i)^2 \check{\lambda}'_1 Q \check{\lambda}_1$. Moreover, $tr(Q \check{P} Q \check{P}) \neq -\beta''(r)$, and therefore $EMH \neq -MIM$. Thus the limiting covariance matrix of $(\hat{r}_{MML}, \hat{\sigma}_{MML}^2)$ will be in the form of a ‘sandwich’ formula.

Theorem 9 *The large N , fixed T asymptotic distribution of $(\hat{r}_{MML}, \hat{\sigma}_{MML}^2)$ is given by*

$$\sqrt{N} \begin{pmatrix} \hat{r}_{MML} - r \\ \hat{\sigma}_{MML}^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N(0, (EMH)^{-1} MIM (EMH)^{-1}) \quad (33)$$

PROOF

See above.

The large T , arbitrary N asymptotic distribution of the Modified MLE for r and σ^2 is identical to the large T and arbitrary N asymptotic distribution of the FDMLE for r and σ^2 :

Theorem 10 When $|r| < 1$, the Modified ML estimator for r and σ^2 in the conditional AR(1) panel model with fixed effects is consistent whether $T \rightarrow \infty$, or $N \rightarrow \infty$, or $N, T \rightarrow \infty$ jointly. The large T and arbitrary N asymptotic distribution of $(\hat{r}_{MML}, \hat{\sigma}_{MML}^2)$ is given by

$$\sqrt{NT} \begin{pmatrix} \hat{r}_{MML} - r \\ \hat{\sigma}_{MML}^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 - r^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} \right) \quad (34)$$

PROOF

See appendix C.2.

Under large T , arbitrary N asymptotics, the bias corrected LSDV estimator for r due to Hahn and Kuersteiner (2000), $\hat{r}_{HK} = \frac{T+1}{T} \hat{r}_{LSDV} + \frac{1}{T}$, has the same limiting distribution as the Modified ML estimator for r (see Alvarez and Arellano, 1998). However, $\hat{r}_{HK}$ is inconsistent under large N , fixed T asymptotics unlike $\hat{r}_{MML}$.

Kruiniger (2002) has shown that under large N , fixed T asymptotics and normality the (conditional) Fixed Effects ML estimator is an efficient estimator for r in the conditional AR(1) panel model with fixed effects and that the Modified MLE for r is less efficient than the (conditional) FEMLE for r . Moreover, it is shown in that paper that under large N , fixed T asymptotics both estimators are very imprecise when r is close to one. The latter result is related to the fact that the expected Hessians of the respective log-likelihood functions are singular when $r = 1$, which implies that r is asymptotically not identified by the likelihood functions when $r = 1$. Thus the large T , arbitrary N limiting distribution for $\hat{r}_{MML}$ provides a poor approximation to the true distribution for $\hat{r}_{MML}$ local to unity when T is small.¹² On the other hand, since $\hat{r}_{MML}$ attains both the large T , fixed N CR efficiency bound and the diagonal path limit efficiency bound for regular estimators, which has been derived by Hahn and Kuersteiner (2000), there seems little harm in employing $\hat{r}_{MML}$ instead of $\hat{r}_{cFEMLE}$ whenever both N and T are large and r is not very close to unity.¹³

¹²One can consider at least three dimensions for asymptotics: $N \rightarrow \infty$, $T \rightarrow \infty$, and $r \uparrow 1$. The asymptotic results depend crucially on whether first $T \rightarrow \infty$ or $r \uparrow 1$. Since $\lim_{T \rightarrow \infty} \frac{1}{T} \beta'(r) = 0$ when $|r| < 1$ and $\lim_{r \uparrow 1} \frac{1}{T} \beta'(r) = \frac{1}{2} \frac{T-2}{T}$, it is easily seen that $\lim_{r \uparrow 1} \lim_{T \rightarrow \infty} \frac{1}{T} \beta'(r) = 0$, whereas $\lim_{T \rightarrow \infty} \lim_{r \uparrow 1} \frac{1}{T} \beta'(r) = \frac{1}{2}$.

¹³Moreover, only the large N , fixed T asymptotic properties of the (conditional) FEMLE are known.

5 Summary

In this paper we have studied Maximum Likelihood estimators for two types of dynamic panel data models with fixed effects: the conditional AR(1) panel model with fixed effects and the inclusive ARMA panel model with fixed effects which includes both the stationary model and the unit root model.

We have proved that the First Difference ML estimator for the inclusive ARMA panel model with fixed effects is consistent and has a limiting normal distribution whether one employs large N and fixed T asymptotics, large T and fixed N asymptotics, or joint N, T asymptotics. In this connection we have shown that the likelihood function of the first-differences of the data is differentiable (also at the unit root). Next, we have presented two Wald-type panel unit root tests that are based on the FDMLE. One of these tests can also be used when only a single time-series is available.

We have also shown that the Modified ML approach yields an estimator for the conditional AR(1) panel model with fixed effects that is consistent and has a limiting normal distribution under any type of asymptotics that could possibly be employed.

Under large N and fixed T asymptotics the FDMLE does not attain the (generalized) Cramér-Rao lowerbound for estimators of the stationary AR(1) panel model with fixed effects. It has been proved that under this kind of asymptotics and normality the most efficient Random Effects GMM estimator for r in the stationary model, e.g. the Optimal Linear GMM estimator due to Kruiniger (2000), is more efficient than the Optimal Minimum Distance Estimator for r in the stationary model with fixed effects, which is due to Chamberlain (1982, 1984). Since under normality the OMD estimator for r is asymptotically equivalent to the FDMLE and the OLGMM estimator for r is asymptotically equivalent to the REMLE, it follows that the REMLE is asymptotically more efficient than the FDMLE for r when N tends to infinity but T is fixed.

Under large T and arbitrary N asymptotics and normality, the FDMLE, the REMLE and the FEMLE for r in the stationary AR(1) model, and the Modified MLE and the Within MLE for r in the conditional AR(1) model are asymptotically equivalent. This implies that when $T \rightarrow \infty$, some moment conditions, i.e. those involving levels of the data and the moment conditions that are related to the initial conditions, are asymptotically redundant relative to the moment conditions that are implicitly exploited by the Within ML estimator.

A ASYMPTOTIC PROPERTIES OF FDML ESTIMATORS FOR INCLUSIVE ARMA(p,q) PANEL MODELS WITH FIXED EFFECTS.

We will first establish some useful results and then prove consistency and asymptotic normality. Van der Leeuw (1994) is henceforth abbreviated as VdL. Let $\|A\| = \text{tr}(A'A)^{1/2}$ be the Euclidean norm of a matrix A .

A.1 Introduction and two lemmas

Recall that the $(p+q)$ -vector r comprises p AR parameters and q MA parameters and that $r^S = \sum_{k=1}^p r_k$. We will distinguish among the following parameter spaces for $\theta = (r' \sigma^2)'$: the parameter space corresponding to the inclusive model, $\Theta = \Theta_r \times (0, \infty)$, where $\Theta_r = \{r \in \mathbb{R}^{p+q} \mid 0 < r^S \leq 1\}$; the set $\Theta_r^{UR} = \{r \mid r \in \Theta_r, r^S = 1\}$; the set $\Theta_r^{URZ} = \{r \mid r \in \Theta_r^{UR}, r_{p+1} = 0, \dots, r_q = 0\}$; the set $\Theta_r^S = \{r \mid r \in \Theta_r, r^S < 1\}$; the compact parameter space $\bar{\Theta} = \bar{\Theta}_r \times [\delta_2, \frac{1}{\delta_2}]$, where $\delta_2 > 0$ is a small number, and which is such that $\bar{\Theta} \subset \Theta$ and $\Theta_r^{URZ} \subset \bar{\Theta}_r$; and the *extended* parameter space $\Theta^\delta = \Theta_r^\delta \times (0, \infty)$, where Θ_r^δ is an open set and $\Theta_r \subset \Theta_r^\delta$. A more precise definition of Θ_r^δ will be given below.

VdL has shown that an ARMA(p,q) matrix V can be factorized as $V = KW^{-1}K'$, where K is linear in the MA parameters, and W consists of second order polynomials of the AR parameters. W^{-1} is an AR(p) covariance matrix.

The matrix W may be singular for particular values of the vector of AR parameters. Nevertheless, expressions involving W^{-1} , e.g. $DW^{-1}D'$, may still exist for such values of the vector of AR parameters.

We can factorize the autoregressive lag-polynomial $A(L)$ as $\prod_{k=1}^p (1 - a_k L)$, where w.l.o.g. $|a_1| \leq 1$, $|a_i| < 1$, $i = 2, \dots, p$. The unit root case, $r^S = 1$, corresponds to $a_1 = 1$ and stationarity, $r^S < 1$, corresponds to $a_1 < 1$.

Let $a = (a_1 \ a_2 \ \dots \ a_p)'$, $\Theta_a = \{a \in \mathcal{C}^p \mid |a_1| \leq 1, |a_i| < 1, i = 2, \dots, p\}$ and let $\Theta_a^{UR} = \{a \mid |a_1| = 1, |a_i| < 1, i = 2, \dots, p\}$.

We will first show that $E(Dy_i y_i' D')$ is left continuous for every $r \in \Theta_r^{UR}$, i.e. $E(Dy_i y_i' D')$ is continuous between the stationary cases and the unit root cases. Therefore, $E(Dy_i y_i' D) =$

$\sigma^2 D V D'$ for every $r \in \Theta_r^{UR}$ as well. Let $R(r) = D V D'$, then we can state

Lemma 11 *For any $p, q, T < \infty$, $\lim_{n \rightarrow \infty} R(\check{r}_n) = R(r)$ for every $r \in \Theta_r^{UR}$ and for every sequence $\{\check{r}_n\}_{n=1}^\infty$ with $\lim_{n \rightarrow \infty} \check{r}_n = r$, and $\check{r}_n \in \Theta_r, \forall n \geq 1$.*

PROOF

We will first prove the lemma for $R = D W^{-1} D'$, where W^{-1} is an arbitrary $\text{AR}(p)$ covariance matrix. So initially we will assume that $B(L) = I$ or $K = I$.

In the lemma we consider any sequence $\{\check{r}_n\}$ in Θ_r that converges to $r \in \Theta_r^{UR}$. Equivalently, we could consider any sequence $\{\check{a}_n\}$ in Θ_a that converges to $a \in \Theta_a^{UR}$, since by definition $\forall r \in \Theta_r, \exists! a \in \Theta_a$ such that $1 - \sum_{k=1}^p r_k L^k = \prod_{k=1}^p (1 - a_k L)$, and since $\forall a \in \Theta_a, \exists! r \in \Theta_r$ such that $1 - \sum_{k=1}^p r_k L^k = \prod_{k=1}^p (1 - a_k L)$. We can restate the lemma using the alternative "a"-parametrisation. To prove the lemma we need to introduce some additional definitions.

We decompose $A(L)$ as $\check{A}(L)(1 - a_1 L)$. Let's define $\omega_{i,t} = \check{A}(L)^{-1} B(L) \varepsilon_{i,t}$. Then we can write $A(L)(y_{i,t} - \mu_i) = B(L) \varepsilon_{i,t}$ as $(1 - a_1 L)(y_{i,t} - \mu_i) = \omega_{i,t}$, and

$$\Delta y_{i,t} = \sum_{m=0}^\infty a_1^m \omega_{i,t-m} - \sum_{m=0}^\infty a_1^m \omega_{i,t-1-m} = \sum_{m=0}^\infty a_1^m (a_1 - 1) \omega_{i,t-1-m} + \omega_{i,t}.$$

Define for arbitrary s, t , and a_1 , with $|t-s| < \infty$ and $|a_1| \leq 1$, $E(s, t, a_1) \equiv E(\Delta y_{i,s} \Delta y_{i,t}) = E[(\sum_{m=0}^\infty a_1^m (a_1 - 1) \omega_{i,s-1-m} + \omega_{i,s})(\sum_{m=0}^\infty a_1^m (a_1 - 1) \omega_{i,t-1-m} + \omega_{i,t})]$. We assume w.l.o.g. that $s \leq t$. Let us also define

$$E(s, t, a_1, 1) \equiv E[(\sum_{m=0}^\infty a_1^m (a_1 - 1) \omega_{i,s-1-m})(\sum_{m=0}^\infty a_1^m (a_1 - 1) \omega_{i,t-1-m})],$$

$$E(s, t, a_1, 2) \equiv E[(\sum_{m=0}^\infty a_1^m (a_1 - 1) \omega_{i,s-1-m})(\omega_{i,t})], \text{ and}$$

$$E(s, t, a_1, 3) \equiv E[(\omega_{i,s})(\sum_{m=0}^\infty a_1^m (a_1 - 1) \omega_{i,t-1-m})].$$

Then we can write $E(s, t, a_1) = E(s, t, a_1, 1) + E(s, t, a_1, 2) + E(s, t, a_1, 3) + E(\omega_{i,s} \omega_{i,t})$.

W.l.o.g., we will only consider sequences $\{\check{a}_n\}$ in which only a_1 changes. We will prove the lemma by showing that $\lim_{a_1 \uparrow 1} E(s, t, a_1) = E(\omega_{i,s} \omega_{i,t}) \forall s, t$, i.e. by showing that $\lim_{a_1 \uparrow 1} E(s, t, a_1, 1) = 0$, $\lim_{a_1 \uparrow 1} E(s, t, a_1, 2) = 0$, and $\lim_{a_1 \uparrow 1} E(s, t, a_1, 3) = 0 \forall s, t$.

For this purpose, we will first examine $E[\sum_{m=0}^\infty a_1^m \omega_{i,t-m} \omega_{i,t}]$.

Since $\{w_{i,t}\}$ is a stationary $\text{AR}(p-1)$ process, $E[\sum_{s=0}^\infty a_1^s \omega_{i,t} \omega_{i,t-s}] = [\sum_{s=0}^\infty a_1^s \rho_s] E(\omega_{i,t}^2)$, where ρ_s is the s -th autocorrelation. Stationarity of $\{\omega_{i,t}\}$ also implies that $E(\omega_{i,t}^2) < \infty$. Moreover, the autocorrelations ρ_s satisfy the Yule-Walker equations: $\rho_s = \sum_{k=1}^{p-1} \Psi_k \lambda_k^s$, where Ψ_k and λ_k are constants and $|\lambda_k| < 1$. Upon substitution, we obtain $\sum_{s=0}^\infty a_1^s \rho_s =$

$\sum_{s=0}^{\infty} a_1^s \sum_{k=1}^{p-1} \Psi_k \lambda_k^s = \sum_{k=1}^{p-1} \Psi_k \sum_{s=0}^{\infty} a_1^s \lambda_k^s = \sum_{k=1}^{p-1} \frac{\Psi_k}{1-a_1 \lambda_k}$ since $|a_1 \lambda_k| < 1$, $k = 1, 2, \dots$, $p-1$.

Let us define $E(t, a_1) \equiv (1 - a_1)E[\sum_{s=0}^{\infty} a_1^s \omega_{i,s} \omega_{i,t-s}] \forall t$.

We find that $\lim_{a_1 \uparrow 1} E(t, a_1) = \lim_{a_1 \uparrow 1} E(\omega_{i,t}^2) \sum_{k=1}^{p-1} \frac{\Psi_k(1-a_1)}{1-a_1 \lambda_k} = 0 \forall t$.

We will now show that $\lim_{a_1 \uparrow 1} E(s, t, a_1, 1) = 0$.

$$E(s, t, a_1, 1) = (a_1 - 1)^2 E[(\sum_{m=0}^{\infty} a_1^m \omega_{i,s-1-m})(\sum_{m=0}^{\infty} a_1^m \omega_{i,t-1-m})].$$

$$\text{We have } E[(\sum_{m=0}^{\infty} a_1^m \omega_{i,s-m})(\sum_{m=0}^{\infty} a_1^m \omega_{i,t-m})] =$$

$$E[(\sum_{m=0}^{t-s-1} a_1^m \omega_{i,t-m})(\sum_{m=0}^{\infty} a_1^m \omega_{i,s-m})] + a_1^{t-s} E(\sum_{m=0}^{\infty} a_1^m \omega_{i,s-m})^2.$$

$$\text{Now } E(\sum_{m=0}^{\infty} a_1^m \omega_{i,s-m})^2 = E[-\sum_{m=0}^{\infty} a_1^{2m} \omega_{i,s-m}^2 + 2 \sum_{m=0}^{\infty} \sum_{l=m}^{\infty} a_1^{m+l} \omega_{i,s-m} \omega_{i,s-l}] =$$

$$E[-\sum_{m=0}^{\infty} a_1^{2m} \omega_{i,s-m}^2] + 2E[\omega_{i,s}^2] \sum_{m=0}^{\infty} a_1^{2m} \sum_{l=m}^{\infty} a_1^{l-m} \rho_{i,l-m} =$$

$$-\frac{1}{1-a_1^2} E[\omega_{i,s}^2] + \frac{2}{1-a_1^2} E[\omega_{i,s}^2] \sum_{k=1}^{p-1} \frac{\Psi_k}{1-a_1 \lambda_k}.$$

$$\text{Therefore } \lim_{a_1 \uparrow 1} (a_1 - 1)^2 a_1^{t-s} E(\sum_{m=0}^{\infty} a_1^m \omega_{i,s-m})^2 = 0.$$

$$\text{We also have } E(\sum_{m=0}^{t-s-1} a_1^m \omega_{i,t-m})(\sum_{m=0}^{\infty} a_1^m \omega_{i,s-m}) =$$

$$E(\sum_{k=0}^{t-s-1} a_1^k \omega_{i,t-k} a_1^{s+k-t} (\sum_{m=0}^{\infty} a_1^m \omega_{i,t-k-m} - \sum_{m=0}^{t-k-s-1} a_1^m \omega_{i,t-k-m})).$$

By stationarity of $\{\omega_{i,t}\}$ there is a constant C (which does not depend on a_1) such that

$$|E(\sum_{k=0}^{t-s-1} a_1^k \omega_{i,t-k} a_1^{s+k-t} \sum_{m=0}^{t-k-s-1} a_1^m \omega_{i,t-k-m})| < C \text{ and therefore}$$

$$\lim_{a_1 \uparrow 1} (a_1 - 1)^2 E(\sum_{k=0}^{t-s-1} a_1^k \omega_{i,t-k} a_1^{s+k-t} \sum_{m=0}^{t-k-s-1} a_1^m \omega_{i,t-k-m}) = 0.$$

$$\text{The second product } E(\sum_{k=0}^{t-s-1} a_1^k \omega_{i,t-k} a_1^{s+k-t} \sum_{m=0}^{\infty} a_1^m \omega_{i,t-k-m}) =$$

$$a_1^{s-t} E(\sum_{k=0}^{t-s-1} a_1^{2k} \omega_{i,t-k} \sum_{m=0}^{\infty} a_1^m \omega_{i,t-k-m}) = \frac{a_1^{s-t}}{1-a_1} \sum_{k=0}^{t-s-1} a_1^{2k} E(t-k, a_1).$$

Since $\lim_{a_1 \uparrow 1} E(t, a_1) = 0 \forall t$, we find that

$$\lim_{a_1 \uparrow 1} (a_1 - 1)^2 E(\sum_{k=0}^{t-s-1} a_1^k \omega_{i,t-k} a_1^{s+k-t} \sum_{m=0}^{\infty} a_1^m \omega_{i,t-k-m}) = 0.$$

$$\text{It follows that } \lim_{a_1 \uparrow 1} (a_1 - 1)^2 E(\sum_{m=0}^{t-s-1} a_1^m \omega_{i,t-m})(\sum_{m=0}^{\infty} a_1^m \omega_{i,s-m}) = 0.$$

From the previous results, we conclude that $\lim_{a_1 \uparrow 1} E(s, t, a_1, 1) = 0 \forall s, t$.

We will now show that $\lim_{a_1 \uparrow 1} E(s, t, a_1, 2) = 0 \forall s, t$.

$$\text{First, } E(s, t, a_1, 2) = E[(\sum_{m=0}^{\infty} a_1^m (a_1 - 1) \omega_{i,s-1-m})(\omega_{i,t})] =$$

$$E[(\sum_{m=s-t-1}^{\infty} a_1^m (a_1 - 1) \omega_{i,s-1-m})(\omega_{i,t})] - E[(\sum_{m=s-t-1}^{-1} a_1^m (a_1 - 1) \omega_{i,s-1-m})(\omega_{i,t})] =$$

$$a_1^{s-t-1} E(t, a_1) - E[(\sum_{m=s-t-1}^{-1} a_1^m (a_1 - 1) \omega_{i,s-1-m})(\omega_{i,t})] =$$

$$a_1^{s-t-1} E(t, a_1) - (a_1 - 1) E[(\sum_{m=s-t-1}^{-1} a_1^m \omega_{i,s-1-m})(\omega_{i,t})].$$

The sum consists of a finite number of terms. Stationarity of $\{\omega_{i,t}\}$ implies that

$$\lim_{a_1 \uparrow 1} (1 - a_1) E[\sum_{m=s-t-1}^{-1} a_1^m \omega_{i,s-1-m} \omega_{i,t}] = 0. \text{ Above, it was shown that } \lim_{a_1 \uparrow 1} E(t, a_1) = 0$$

$\forall t$. It follows that $\lim_{a_1 \uparrow 1} E(s, t, a_1, 2) = 0 \forall s, t$.

Likewise, we can show that $\lim_{a_1 \uparrow 1} E(s, t, a_1, 3) = 0 \forall s, t$.

Recall that $E(s, t, a_1, 3) = E[(\omega_{i,s})(\sum_{m=0}^{\infty} a_1^m (a_1 - 1) \omega_{i,t-1-m})]$.

If $s = t$, then $E(s, t, a_1, 3) = E(s, a_1)$. Next, assume that $s \leq t - 1$.

$$\begin{aligned} & \text{In that case } E[(\omega_{i,s})(\sum_{m=0}^{\infty} a_1^m (a_1 - 1) \omega_{i,t-1-m})] = \\ & a_1^{t-s-1} E(s, a_1) + E[(\sum_{m=0}^{t-s-2} a_1^m (a_1 - 1) \omega_{i,t-1-m})(\omega_{i,s})] = \\ & a_1^{t-s-1} E(s, a_1) + (a_1 - 1) E[(\sum_{m=0}^{t-s-2} a_1^m \omega_{i,t-1-m})(\omega_{i,s})]. \end{aligned}$$

We can show that $\lim_{a_1 \uparrow 1} (1 - a_1) E[(\sum_{m=0}^{t-s-2} a_1^m \omega_{i,t-1-m})(\omega_{i,s})] = 0$ using that the sum consists of a finite number of terms and $\{\omega_{i,t}\}$ is a stationary process. In addition, $\lim_{a_1 \uparrow 1} E(s, a_1) = 0 \forall s$. We conclude that $\lim_{a_1 \uparrow 1} E(s, t, a_1, 3) = 0 \forall s, t$.

Using the results from above, we obtain $\lim_{a_1 \uparrow 1} E(s, t, a_1) = E(\omega_{i,s} \omega_{i,t}) \forall s, t$ with $|t - s| < \infty$. When $a_1 = 1$, $\Delta y_{i,t} = \omega_{i,t} \forall t$, and therefore $E(\Delta y_{i,s} \Delta y_{i,t}) = E(\omega_{i,s} \omega_{i,t}) \forall s, t$. It follows that $\lim_{n \rightarrow \infty} R(\tilde{r}_n) = R(r)$ for every $r \in \Theta_r^{URZ}$.

We will now prove the lemma for $R = DVD'$, where V is an arbitrary ARMA(p,q) covariance matrix.

VdL shows that an $T \times T$ ARMA(p,q) matrix V_T can be factorized as $V_T = K_T W_{T+q}^{-1} K_T'$, where K_T is an $T \times (T+q)$ matrix which is linear in the MA parameters and W_{T+q}^{-1} is an $(T+q) \times (T+q)$ AR(p) covariance matrix. Let $\varphi_1, \dots, \varphi_q$ denote the MA parameters and let $O_{T,k}$ denote a $T \times k$ matrix of zeros, then we can write $K_T = [O_{T,q} \ I_T] + \sum_{k=1}^q \varphi_k [O_{T,q-k} \ I_T \ O_{T,k}]$. Noting that the $(T-1) \times T$ first difference matrix $D_T = [I_{T-1} \ 0] - [0 \ I_{T-1}]$, it can easily be verified that $D_T K_T = K_{T-1} D_{T+q}$. We obtain that $D_T V_T D_T' = K_{T-1} D_{T+q} W_{T+q}^{-1} D_{T+q}' K_{T-1}'$.

Above we have shown that $\lim_{n \rightarrow \infty} \tilde{R}(\tilde{r}_n) = \tilde{R}(r)$ where $\tilde{R}(r) = D_T W_T^{-1} D_T'$, for every $r \in \Theta_r^{URZ}$ and for arbitrary p , $T < \infty$. Since K_{T-1} is linear in (the elements of) r , K_{T-1} is continuous in r . Finally, we notice that the product of left-continuous functions is left-continuous itself. $\square$

When $a_1 = 1$, $E(Dy_i y_i' D')$ is a stationary ARMA(p-1,q) covariance matrix. Since all stationary ARMA matrices are nonsingular (see VdL) and bounded, $\|E(\omega_i \omega_i')\| < \infty$ and $E(\omega_i \omega_i')$ is nonsingular or more precisely $\sup_{r \in \Theta_r^{UR}} \|R(r)\| < \infty$, and $R(r)$ is nonsingular $\forall r \in \Theta_r^{UR}$.

We will now show that $l(Dy_i; r, \sigma^2)$ is differentiable with respect to θ .

Lemma 12 *The likelihood function of the first differences Dy_i , $i = 1, \dots, N$,*

$$\begin{aligned} l(Dy_i; r, \sigma^2) = & -\frac{N(T-1)}{2} \log(2\pi) - \frac{N(T-1)}{2} \log \sigma^2 \\ & -\frac{N}{2} \log |DVD'| - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i' D' (DVD')^{-1} Dy_i) \end{aligned} \quad (35)$$

is infinitely many times differentiable with respect to θ in every $\check{\theta} \in \Theta^\delta$, where $\Theta^\delta = \Theta_r^\delta \times (0, \infty)$, with $\Theta_r^\delta = \left(\bigcup_{r \in \Theta_r^{UR}} \mathcal{N}^\delta(r) \right) \cup \Theta_r$ and $\mathcal{N}^\delta(r) \equiv \{\check{r} \mid \|r - \check{r}\| < \delta(r)\}$.

PROOF

Lemma 11 implies that $E(Dy_i y_i' D) = \sigma^2 DVD'$ also when $r \in \Theta_r^{UR}$ even though V does not exist when $r \in \Theta_r^{UR}$. Thus the formula of the likelihood function of the differenced data, $l(Dy_i; r, \sigma^2)$, which is given in (35), is also valid when $r \in \Theta_r^{UR}$.

It follows from VdL that the covariance matrix of a stationary ARMA process, (a stationary ARMA matrix for short) is nonsingular and bounded, that its elements are ratios of polynomials in the AR and MA parameters, and that its elements are differentiable with respect to the AR and MA parameters (r) in every $\check{r} \in \Theta_r^S$ (the stationary parameter space). It also follows from VdL that the determinant and the inverse of a stationary ARMA matrix are differentiable with respect to r in every $\check{r} \in \Theta_r^S$. In addition, the inverse and the determinant of the stationary ARMA matrix are bounded because the original matrix is nonsingular and its elements are bounded. Since the stationary ARMA matrices and their inverses and determinants consist of ratios of polynomials in the AR and MA parameters, they are in fact infinitely many times differentiable with respect to r in every $\check{r} \in \Theta_r^S$.

At this point it should be noted that although DVD' is a stationary ARMA(p,q+1) matrix when $\{y_t\}$ is an ARMA(p,q) process and $r^S < 1$ or a stationary ARMA(p-1,q) matrix when $r^S = 1$, this matrix corresponds to the difference of an ARMA process and it is therefore parametrized in a different way than the “standard” ARMA(p,q+1) or ARMA(p-1,q) matrices found in VdL. In case that $r^S < 1$, one of the MA parameters is already known, so in that case the ARMA(p,q+1) matrix DVD' depends on only q MA parameters. In the unit root case, the ARMA(p-1,q) matrix DVD' contains p AR parameters that satisfy the constraint $r^S = 1$.

Now, $\forall r \in \Theta_r$ (including every $r \in \Theta_r^{UR}$) $R(r) = DVD'$ exists and is stationary and hence bounded and also nonsingular. Furthermore, $\forall r \in \Theta_r$ $R(r)$ consists of ratios of polynomials

in the AR and MA parameters just like V . Because $R(r)$ is bounded, we have $\forall r \in \Theta_r$ that the denominators of $R(r)$ do not equal zero. Since polynomials are continuous and since the function $1/x$ is continuous in every $x \in \mathfrak{R}$, $x \neq 0$, it follows that $\forall \varepsilon$ and $\forall r \in \Theta_r^{UR}$, $\exists \delta$ such that $\|R(r) - R(\check{r})\| < \varepsilon$ $\forall \check{r}$ that satisfy $\|r - \check{r}\| < \delta$. So $\forall r \in \Theta_r^{UR}$, $\exists \delta(r)$ such that $R(r)$ is bounded and positive definite on $\mathcal{N}^\delta(r) \equiv \{\check{r} \mid \|r - \check{r}\| < \delta(r)\}$. It follows that $R(r)$ is differentiable with respect to r , i.e. with respect to the AR and MA parameters, in every $r \in \Theta_r^\delta$, where $\Theta_r^\delta = \left(\bigcup_{r \in \Theta_r^{UR}} \mathcal{N}^\delta(r)\right) \cup \Theta_r$. Since the elements of the inverse of a matrix consist of ratios of minors and the determinant, the elements of $R(r)^{-1}$ also consist of ratios of polynomials in the AR and MA parameters. Because $R(r)$ is nonsingular and differentiable with respect to r in every $r \in \Theta_r^\delta$, $R(r)^{-1}$ and $|R(r)|$ are also differentiable with respect to r in every $r \in \Theta_r^\delta$. Moreover, $R(r)^{-1}$ and $|R(r)|$ are bounded on any compact subset of Θ^δ . Because the matrices and their inverses and the determinant consist of ratios of polynomials in the AR and MA parameters, they are in fact infinitely many times differentiable with respect to r in every $r \in \Theta_r^\delta$. $\square$

The function $L(Dy_i; r, \sigma^2)$ is a normal probability density function (p.d.f.) $\forall \theta \in \Theta^\delta$, because $R(r)$ is positive definite symmetric $\forall r \in \Theta_r^\delta$. However, $\forall \theta \in \Theta^\delta$ with $r^S > 1$, this density is not the p.d.f. of Dy_i .

A.2 Large N, fixed T consistency of the FDMLE for the inclusive ARMA(p,q) panel model with fixed effects

To prove consistency of the FDMLE we will verify the conditions of theorem 2.5. in Newey and McFadden (1994).

We assume that $\bar{\theta} \in \bar{\Theta} \subset \Theta$, where $\bar{\Theta}$ can be any compact subset of Θ . However, we will assume that $\bar{\Theta}$ is such that $\Theta_r^{URZ} \subset \bar{\Theta}_r$.

Lemma 12 implies that $l(Dy_i; r, \sigma^2)$ is differentiable with respect to θ in every $\theta \in \Theta^\delta$. It follows that $l(Dy_i; r, \sigma^2)$ is continuous in every $\theta \in \bar{\Theta}$.

The last two steps of the proof are as follows:

For identification it suffices to show that $\sigma^2 DV(r)D' \neq \bar{\sigma}^2 DV(\bar{r})D' \forall \theta \in \bar{\Theta}$ with $\theta \neq \bar{\theta}$, because $(\sigma^2 DV(r)D' \neq \bar{\sigma}^2 DV(\bar{r})D' \text{ if and only if } \theta \neq \bar{\theta}) \Leftrightarrow ((\sigma^2 DV(r)D')^{-1} \neq (\bar{\sigma}^2 DV(\bar{r})D')^{-1} \text{ iff } \theta \neq \bar{\theta})$. The latter follows trivially from the fact that $\sigma^2 DV(r)D' \times$

$(\sigma^2 DV(r)D')^{-1} = I \forall \theta \in \bar{\Theta}$. Now, the $T - 1$ elements of the first column of the (stationary) covariance matrix $\sigma^2 DV(r)D'$ are all different functions of the $p + q + 1$ parameters. It follows that if $p + q + 1 \leq T - 1$, $(\sigma^2 DV(r)D' \neq \sigma^2 DV(\bar{r})D') \Leftrightarrow \theta \neq \bar{\theta}$.

To show that $E(\sup_{(r, \sigma^2) \in \bar{\Theta}} l(Dy_i; r, \sigma^2)) < \infty$, we recall that $|\sigma^2 DV(r)D'| > 0 \forall \theta \in \Theta^\delta$ and that $(DV(r)D')^{-1}$ and $|DV(r)D'|$ are differentiable and hence continuous in every $r \in \Theta_r^\delta$. It follows that $\sigma^{-2}(DV(r)D')^{-1}$ and $|\sigma^2 DV(r)D'|$ are bounded on the compact set $\bar{\Theta}$. Furthermore, $E(Dy_i y_i' D')$ is bounded by stationarity of Dy_i . We conclude that $E(\sup_{(r, \sigma^2) \in \bar{\Theta}} l(Dy_i; r, \sigma^2)) < \infty$. $\square$

A.3 Large N, fixed T asymptotic distribution of the FDMLE for the inclusive ARMA(p,q) panel model with fixed effects

Asymptotic normality of the FDMLE for the inclusive ARMA(p,q) panel model with fixed effects can be established by verifying the conditions of theorem 3.3. in Newey and McFadden (1994), henceforth NMcf. The proof proceeds as follows:

We have seen that $\forall r \in \Theta_r^{UR}, \exists \delta(r)$ such that $R(r) = DV(r)D'$ is bounded and positive definite on $\mathcal{N}^\delta(r) = \{\tilde{r} \mid \|r - \tilde{r}\| < \delta(r)\}$. Also recall that $\Theta_r^\delta = \left(\bigcup_{r \in \Theta_r^{UR}} \mathcal{N}^\delta(r)\right) \cup \Theta_r$.

Let us now define the compact set $\bar{\mathcal{N}}^{\frac{1}{2}\delta}(\bar{r}) \equiv \{\tilde{r} \mid \|\bar{r} - \tilde{r}\| \leq \frac{1}{2}\delta(\bar{r})\}$, where $\bar{r}$ is the true value of r . Next, let $\bar{\Theta}_r^{\frac{1}{2}\delta} = \bar{\mathcal{N}}^{\frac{1}{2}\delta}(\bar{r}) \cup \bar{\Theta}_r$, where $\bar{\Theta}_r$ is a compact set with $\Theta_r^{URZ} \subset \bar{\Theta}_r \subset \Theta_r$. Note that $\bar{\Theta}_r^{\frac{1}{2}\delta} \subset \Theta_r^\delta$.

The true value $\bar{\theta} \in \bar{\Theta}_r \times [\delta_2, \frac{1}{\delta_2}]$ where $\delta_2 > 0$ and very small. It follows that $\bar{\theta} \in \text{interior}(\bar{\Theta}_r^{\frac{1}{2}\delta} \times [\frac{1}{2}\delta_2, \frac{2}{\delta_2}])$. We note that the conditions of theorem 2.5 in NMcf are also satisfied for the larger compact set $\bar{\Theta}_r^{\frac{1}{2}\delta} \times [\frac{1}{2}\delta_2, \frac{2}{\delta_2}]$ (of course, we still assume that the true value $\bar{\theta} \in \bar{\Theta}_r \times [\delta_2, \frac{1}{\delta_2}]$, e.g. when evaluating expectations). It can also be verified that lemma 2.2 and lemma 2.4 of NMcf are still applicable when we use the larger compact set $\bar{\Theta}_r^{\frac{1}{2}\delta} \times [\frac{1}{2}\delta_2, \frac{2}{\delta_2}]$.

Above, we have shown that $L(Dy_i; r, \sigma^2)$ is infinitely many times differentiable with respect to AR and MA parameters in every $\theta \in \Theta^\delta$ (see lemma 12). Furthermore we have observed that $L(Dy_i; r, \sigma^2)$ is a proper normal p.d.f. $\forall \theta \in \Theta^\delta$ and therefore $L(Dy_i; r, \sigma^2) > 0$ on Θ^δ .

Recall that $\Omega = \Omega(\theta) = \sigma^2 DV(r)D'$. The Hessian of $l(Dy_i; r, \sigma^2)$ is given by

$H(\theta, y_i) = \frac{1}{2} \sum_{1 \leq s, t < T} (\Omega_{s,t} - \Delta y_{i,s+1} \Delta y_{i,t+1}) \frac{\partial^2 [\Omega^{-1}]_{s,t}}{\partial \theta \partial \theta'} - \frac{1}{2} ([\frac{dvec\Omega}{d\theta'}]_{\bar{\theta}})' (\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1}) [\frac{dvec\Omega}{d\theta'}]_{\bar{\theta}})$, (cf Magnus (1978)).

Above we have shown that $R(r)$ is nonsingular and that $R(r)$, $R(r)^{-1}$, and $|R(r)|$ are ratios of polynomials in the elements of r and infinitely many times differentiable on Θ^δ . It follows that $R(r)$, $R(r)^{-1}$, and $|R(r)|$ and their derivatives with respect to r are bounded on any compact subset of Θ^δ . We also noticed that $L(Dy_i; r, \sigma^2)$ is a proper p.d.f. $\forall \theta \in \Theta^\delta$. These results together with stationarity of Dy_i , which gives that $E(Dy_i y_i' D')$ is bounded, and the fact that the standard normal density $\phi(v)$ and its first and second derivatives are uniformly bounded imply that conditions (iii) and (v) of theorem 3.3 in NMCF are satisfied (since we can choose the neighbourhood of $\bar{\theta}$, $\mathcal{N}(\bar{\theta})$, such that $\mathcal{N}(\bar{\theta}) \subset \bar{\Theta}_r^{\frac{1}{2}\delta} \times [\frac{1}{2}\delta_2, \frac{2}{\delta_2}]$).

The second Bartlett identity holds because condition (iii) of theorem 3.3 in NMCF allows us to interchange the order of differentiation and integration. So the information matrix equals minus the expected value of the Hessian of $l(Dy_i; r, \sigma^2)$, i.e. $\frac{1}{2} ([\frac{dvec\Omega}{d\theta'}]_{\bar{\theta}})' (\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1}) [\frac{dvec\Omega}{d\theta'}]_{\bar{\theta}})$. Using the results concerning $R(r)^{-1}$ from above we can easily establish existence of the information matrix for any $\bar{\theta} \in \Theta$. Finally, the information matrix is nonsingular if $\frac{dvec\Omega}{d\theta'}$ has full rank. Because Ω has the structure of a covariance matrix of a stationary process, it is a Toeplitz matrix and therefore we can restrict attention to the derivatives of the elements of the first column of Ω . The fact that the elements of θ can be identified from this column when $T - 1 \geq p + q + 1$ implies that $\frac{dvec\Omega}{d\theta'}$ has full rank when $T \geq p + q + 2$. It follows that $E[H(\bar{\theta}, y_i)]$ is negative definite symmetric for any $\bar{\theta} \in \Theta$. $\square$

A.3.1 Large N, fixed T asymptotic distribution of the FDMLE for the stationary AR(1) panel model with fixed effects

Let $f(r) = \iota' V^{-1} \iota$, and $b(r)' = b' = \frac{\iota' V^{-1}}{\iota' V^{-1} \iota}$. The unconditional log-likelihood function $l_{U,FE}(y_i; r, \mu, \sigma^2)$ can be decomposed as $l(Dy_i; r, \sigma^2) + \sum_{i=1}^N l(\tilde{\mu}_i; r, \mu_i, \sigma^2)$, where $\tilde{\mu}_i = b(r)' y_i$ is the sufficient statistic for μ_i (see Kruiniger (1998)). It follows that the Hessian of $l(Dy_i; r, \sigma^2)$ can be obtained from the Hessians of $l_{U,FE}(y_i; r, \mu, \sigma^2)$ and $l(\tilde{\mu}_i; r, \mu_i, \sigma^2)$, $i = 1, \dots, N$:

$$(NT)^{-1} \frac{\partial^2 l(Dy_i; r, \sigma^2)}{\partial \theta \partial \theta'} = (NT)^{-1} \frac{\partial^2 l_{U,FE}(y_i; r, \mu, \sigma^2)}{\partial \theta \partial \theta'} - (NT)^{-1} \sum_{i=1}^N \frac{\partial^2 l(\tilde{\mu}_i; r, \mu_i, \sigma^2)}{\partial \theta \partial \theta'}, \text{ where } \theta = (r' \sigma^2)'$$

The marginal log-likelihood function of $\tilde{\mu}_i = b' y_i$ is given by

$$l(\tilde{\mu}_i) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 + \frac{1}{2} \log f(r) - \frac{f(r)}{2\sigma^2} (\tilde{\mu}_i - \mu_i)^2 \quad (36)$$

$$= -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 + \frac{1}{2} \log f(r) - \frac{f(r)}{2\sigma^2} (b(r)'(y_i - \mu_i \iota))^2.$$

The elements of the Expected Hessian of this (Marginal) log-likelihood function, EHM , are obtained as follows: ¹⁴

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

Now, we have $E \left(\frac{f''(r)}{2\sigma^2} (\tilde{\mu}_i - \mu_i)^2 \right) = \frac{1}{2} \frac{f''(r)}{f(r)}$, and

$$E \left(b(r)'(y_i - \mu_i \iota) \frac{db(r)'}{dr} (y_i - \mu_i \iota) \right) = E \left(b(r)' u_i u_i' \frac{db(r)'}{dr} \right) = \sigma^2 b(r)' V \frac{db(r)'}{dr} = \frac{\sigma^2}{c} \iota' \frac{db(r)'}{dr} = 0.$$

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

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

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

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

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

It follows that $f(r) = \iota' V^{-1} \iota = (1-r)((T-2)(1-r)+2)$ and $\frac{d(\iota' V^{-1})}{dr} V \frac{d(V^{-1} \iota)}{dr} = \iota' \frac{dV^{-1}}{dr} V \frac{dV^{-1}}{dr} \iota = \frac{2}{1-r^2} (r^{T-1} - 2(T-2)r^2 + 2T - 3)$. The latter result is not obvious and has been obtained by using *Mathematica* and by exploiting the fact that $\frac{dV^{-1}}{dr} V \frac{dV^{-1}}{dr}$ has a particular structure for arbitrary values of T . The other elements of EHM are $E \left(\frac{d^2 l(\tilde{\mu}_i)}{dr d\sigma^2} \right) = \frac{f'(r)}{2\sigma^2 f(r)}$, $E \left(\frac{d^2 l(\tilde{\mu}_i)}{d(\sigma^2)^2} \right) = -\frac{1}{2\sigma^4}$, $E \left(\frac{d^2 l(\tilde{\mu}_i)}{dr d\mu_i} \right) = 0$, $E \left(\frac{d^2 l(\tilde{\mu}_i)}{d\sigma^2 d\mu_i} \right) = 0$, and $E \left(\frac{d^2 l(\tilde{\mu}_i)}{d\mu_i^2} \right) = -\frac{f(r)}{\sigma^2}$.

¹⁴The prime ' denotes a derivative in the case of a scalar function and transpose in case of a vector.

We will now derive the Expected Hessian corresponding to the Unconditional log-likelihood function, EHU . The unconditional log-likelihood of the y_i 's divided by N is given by

$$\frac{1}{N}l(y_i) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2}\log|V| - \frac{1}{2\sigma^2}u_i'V^{-1}u_i. \quad (37)$$

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

Since $|V| = (1-r^2)^{-1}$, we have $\frac{d|V|}{dr} = 2r(|V|)^2$ and $\frac{d^2|V|}{dr^2} = 2(|V|)^2 + 8r^2(|V|)^3$.

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

The other elements of EHU are $E\left(\frac{d^2l(y_i)}{d(\sigma^2)^2}\right) = -\frac{NT}{2\sigma^4}$, $E\left(\frac{d^2l(y_i)}{drd\mu_i}\right) = 0$, $E\left(\frac{d^2l(y_i)}{d\sigma^2d\mu_i}\right) = 0$, and $E\left(\frac{d^2l(y_i)}{d\mu_i^2}\right) = -\frac{Nf(r)}{\sigma^2}$.

The last three elements of EHU are the same as those of $(N \times)$ EHM because $\tilde{\mu}_i, i = 1, \dots, N$, are sufficient statistics. Notice that both EHU and EHM are block-diagonal.

Thus the large N , fixed T asymptotic distribution of $\{\hat{r}_{FDMLE}, \hat{\sigma}_{FDMLE}^2\}$ is given by

$$\sqrt{N}\begin{pmatrix} \hat{r}_{FDMLE} - r \\ \hat{\sigma}_{FDMLE}^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N(0, (EHM_{r,\sigma^2} - EHU_{r,\sigma^2}/N)^{-1}) \quad (38)$$

An explicit expression for $(EHM_{r,\sigma^2} - EHU_{r,\sigma^2}/N)$ is given in (8)-(9).

Below we have plotted graphs of the asymptotic variance of $\hat{r}_{FDMLE}$ (normalized by T) for various values of T .

Figure 1: the asymptotic variance of the FDMLE for r

$$T = 3 \quad w(r) = \frac{3}{4}(-3+r)^2(1+r)^2$$

$$T = 8 \quad x(r) = 8 \frac{(-4+3r)^2(1+r)^2}{3r^7+5r^6+3r^5+r^4-r^3-66r^2-5r+81}$$

$$T = 15$$

$$y(r) = \frac{15}{2} \frac{(-15+13r)^2(1+r)^2}{13r^{14}+24r^{13}+20r^{12}+16r^{11}+12r^{10}+8r^9+4r^8-4r^6-8r^5-12r^4-16r^3-1203r^2-24r+1352}$$

$$T = \infty \quad k(r) = 1 - r^2$$

A.4 Large T, arbitrary N asymptotic properties of the FDMLE for the inclusive AR(1) panel model with fixed effects

Recall that $\theta = (r \sigma^2)'$, $\Omega = \sigma^2 DV(r)D'$, and that upperbars denote the true values of the parameters. Let $d(\tilde{\theta}, \bar{\theta}) = \left\| \tilde{\theta} - \bar{\theta} \right\| = [(\tilde{\theta} - \bar{\theta})'(\tilde{\theta} - \bar{\theta})]^{1/2}$. Furthermore, let $\tilde{T} \equiv \begin{pmatrix} T^{-1} & 0 \\ 0 & T^{-0.5} \end{pmatrix}$.

We will first prove two lemmas:

Lemma 13 *Let $N = 1$. If $\bar{\theta} \in \Theta_r^S \times (0, \infty)$, then (A) $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2 l(Dy_i; r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}} = -\Phi(\bar{\theta}) = - \begin{pmatrix} 1 - \bar{r}^2 & 0 \\ 0 & 2\bar{\sigma}^4 \end{pmatrix}^{-1}$, and (B) $\exists R(\bar{\theta}) > 0$ such that $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^3 l(Dy_i; r, \sigma^2)}{\partial \theta_i \partial \theta_j \partial \theta_k} \Big|_{\bar{\theta}} = m_{ijk}(\tilde{\theta})$, with $|m_{ijk}(\tilde{\theta})| < \hat{M}_{ijk}(\bar{\theta})$, $i, j, k = 1, 2, \forall \tilde{\theta}$ such that $d(\tilde{\theta}, \bar{\theta}) < R(\bar{\theta})$.*

If $\bar{\theta} \in \Theta_r^{UR} \times (0, \infty)$, then (A') $\text{plim}_{T \rightarrow \infty} \tilde{T} \left(\frac{\partial^2 l(Dy_i; r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}} \right) \tilde{T} = - \begin{pmatrix} 8 & 0 \\ 0 & 2\bar{\sigma}^4 \end{pmatrix}^{-1}$, and (B') $\exists R(\bar{\theta}) > 0$ such that $\text{plim}_{T \rightarrow \infty} \tilde{T} \left(\frac{\partial^3 l(Dy_i; r, \sigma^2)}{\partial \theta \partial \theta' \partial \theta_k} \Big|_{\bar{\theta}} \right) \tilde{T} = M_k(\tilde{\theta}) = [m_{ijk}(\tilde{\theta})]$, with $|m_{ijk}(\tilde{\theta})| < \hat{M}_{ijk}(\bar{\theta})$, $i, j, k = 1, 2, \forall \tilde{\theta}$ such that $d(\tilde{\theta}, \bar{\theta}) < R(\bar{\theta})$.

PROOF

For $N = 1$, we have $l_{U,FE}(y; \mu, r, \sigma^2) = l(Dy; r, \sigma^2) + l(\tilde{\mu}; \mu, r, \sigma^2)$, where $\tilde{\mu} = \frac{\iota' V^{-1} y}{\iota' V^{-1} \iota}$ and $l(\tilde{\mu}; r, \mu, \sigma^2) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 + \frac{1}{2} \log \iota' V^{-1} \iota - \frac{1}{2\sigma^2} (\iota' V^{-1} \iota)(\tilde{\mu} - \mu)^2$ (see appendix A.3.1). Recall that $\iota' V^{-1} \iota = (1 - r)((T - 2)(1 - r) + 2)$. We will first show part (A).

It is easily seen that $\lim_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2 \log \iota' V^{-1} \iota}{\partial r^2} = 0$ and $\lim_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2 \log \sigma^2}{\partial (\sigma^2)^2} = 0$.

Thus $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2 l(\tilde{\mu}; \mu, r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}} = \text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2}{\partial \theta \partial \theta'} \left[-\frac{1}{2\sigma^2} (\iota' V^{-1} \iota)(\tilde{\mu} - \mu)^2 \right] \Big|_{\bar{\theta}}$. To calculate the latter limit we use the following results:

Firstly, $\lim_{T \rightarrow \infty} (T)^{-1} \iota' V^{-1} \iota = (1 - r)^2$ and $E(\iota' V^{-1} y) = \iota' V^{-1} \iota \bar{\mu}$. Let $\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$. Then $\text{plim}_{T \rightarrow \infty} \bar{y} = E(y_t) = \bar{\mu}$ by the ergodicity theorem.

Hence, $\text{plim}_{T \rightarrow \infty} (T)^{-1} \iota' V^{-1} y = (1 - r)^2 \text{plim}_{T \rightarrow \infty} \bar{y} = (1 - r)^2 \bar{\mu} = \lim_{T \rightarrow \infty} (T)^{-1} E(\iota' V^{-1} y)$.

Moreover, $\text{plim}_{T \rightarrow \infty} \tilde{\mu} = \bar{\mu} = E(\tilde{\mu})$ and $\text{plim}_{T \rightarrow \infty} (\tilde{\mu})^2 = (\bar{\mu})^2 = \lim_{T \rightarrow \infty} E[(\tilde{\mu})^2]$. The last result follows from $E[(\tilde{\mu})^2] = \text{Var}(\tilde{\mu}) + E[(\tilde{\mu})]^2$, and $\lim_{T \rightarrow \infty} \text{Var}(\tilde{\mu}) = \lim_{T \rightarrow \infty} (\iota' V^{-1} \iota)^{-1} = 0$.

We also have $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial \iota' V^{-1} y}{\partial r} = 2(r - 1) \text{plim}_{T \rightarrow \infty} \bar{y} = 2(r - 1) \bar{\mu} = \lim_{T \rightarrow \infty} (T)^{-1} E\left(\frac{\partial \iota' V^{-1} y}{\partial r}\right)$,

and $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2 l' V^{-1} y}{\partial r^2} = 2 \text{plim}_{T \rightarrow \infty} \bar{y} = 2\bar{\mu} = \lim_{T \rightarrow \infty} (T)^{-1} E(\frac{\partial^2 l' V^{-1} y}{\partial r^2})$.

It follows from all these results that

$$\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2 l(\tilde{\mu}; \mu, r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}} = \lim_{T \rightarrow \infty} (T)^{-1} E(\frac{\partial^2 l(\tilde{\mu}; \mu, r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}}). \quad (39)$$

It follows from results in Pierce (1971) that

$$\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2 l_{U,FE}(y; \mu, r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}} = \lim_{T \rightarrow \infty} (T)^{-1} E(\frac{\partial^2 l_{U,FE}(y; \mu, r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}}). \quad (40)$$

Together (39) and (40) imply that $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^2 l(Dy; r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}} = \lim_{T \rightarrow \infty} (T)^{-1} E(\frac{\partial^2 l(Dy; r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}})$.

Now, $-E(\frac{\partial^2 l(Dy; r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}}) \equiv T \times \Phi_T(\bar{\theta}) = \frac{1}{2} ([\frac{dvec \Omega}{d\theta'}] \Big|_{\bar{\theta}})' (\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1}) [\frac{dvec \Omega}{d\theta'} \Big|_{\bar{\theta}}]$, which is given in (8)-(9). Letting T tend to infinity, we obtain $\lim_{T \rightarrow \infty} (T)^{-1} E(\frac{\partial^2 l(Dy; r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}}) = -\lim_{T \rightarrow \infty} \Phi_T(\bar{\theta}) = -\Phi(\bar{\theta}) = -\begin{pmatrix} 1 - \bar{r}^2 & 0 \\ 0 & 2\bar{\sigma}^4 \end{pmatrix}^{-1}$. With regard to (A') we have $\lim_{\bar{r} \uparrow 1} E(\frac{\partial^2 l(Dy; r, \sigma^2)}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}}) = -\begin{pmatrix} \frac{8}{(T-1)(T-2)} & \frac{4\bar{\sigma}^2}{(T-1)(T-2)} \\ \frac{4\bar{\sigma}^2}{(T-1)(T-2)} & \frac{2\bar{\sigma}^4}{(T-2)} \end{pmatrix}^{-1}$ and $\lim_{T \rightarrow \infty} \tilde{T} \begin{pmatrix} \frac{8}{(T-1)(T-2)} & \frac{4\bar{\sigma}^2}{(T-1)(T-2)} \\ \frac{4\bar{\sigma}^2}{(T-1)(T-2)} & \frac{2\bar{\sigma}^4}{(T-2)} \end{pmatrix}^{-1} \tilde{T} = \begin{pmatrix} 8 & 0 \\ 0 & 2\bar{\sigma}^4 \end{pmatrix}^{-1}$.

We will now turn to part (B):

It also follows from Pierce (1971) that $\forall \bar{\theta} \in \Theta_r^S \times (0, \infty)$, $\exists R_1(\bar{\theta}) > 0$ such that $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^3 l_{U,FE}(y; \mu, r, \sigma^2)}{\partial \theta_i \partial \theta_j \partial \theta_k} \Big|_{\bar{\theta}} = n_{ijk}(\tilde{\theta})$, with $|n_{ijk}(\tilde{\theta})| < \hat{N}_{ijk}(\bar{\theta})$, $i, j, k = 1, 2$, $\forall \tilde{\theta}$ such that $d(\tilde{\theta}, \bar{\theta}) < R_1(\bar{\theta})$.

Using the results presented above (39), we can show that $\exists R_2(\bar{\theta}) > 0$ such that $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^3 l(\tilde{\mu}; \mu, r, \sigma^2)}{\partial \theta_i \partial \theta_j \partial \theta_k} \Big|_{\bar{\theta}} = p_{ijk}(\tilde{\theta})$, with $|p_{ijk}(\tilde{\theta})| < \hat{P}_{ijk}(\bar{\theta})$, $i, j, k = 1, 2$, $\forall \tilde{\theta}$ such that $d(\tilde{\theta}, \bar{\theta}) < R_2(\bar{\theta})$.

It follows that $\exists R(\bar{\theta}) > 0$ such that $\text{plim}_{T \rightarrow \infty} (T)^{-1} \frac{\partial^3 l(Dy_i; r, \sigma^2)}{\partial \theta_i \partial \theta_j \partial \theta_k} \Big|_{\bar{\theta}} = m_{ijk}(\tilde{\theta})$, with $|m_{ijk}(\tilde{\theta})| < \hat{M}_{ijk}(\bar{\theta})$, $i, j, k = 1, 2$, $\forall \tilde{\theta}$ such that $d(\tilde{\theta}, \bar{\theta}) < R(\bar{\theta})$.

Mutatis mutandis for $\bar{\theta} \in \Theta_r^{UR} \times (0, \infty)$. $\square$

Lemma 14 Let $N = 1$, $\tilde{\theta} \in \Theta_r \times (0, \infty)$, and let $K_k(\tilde{\theta}, y_i) = (T)^{-1} \frac{\partial^3 l(Dy_i; r, \sigma^2)}{\partial \theta \partial \theta' \partial \theta_k} \Big|_{\tilde{\theta}}$, $k = 1, 2$ (pre- and postmultiply $l(Dy_i; r, \sigma^2)$ by $\tilde{T}$ instead of $T^{-0.5} I_2$ when $\tilde{\theta} \in \Theta_r^{UR} \times (0, \infty)$). Then $\|K_k(\tilde{\theta}, y_i)\|$, $k = 1, 2$, are uniformly integrable in T .

PROOF

Since $y_{i,t}$ is Gaussian, all moments of $K_k(\tilde{\theta}, y_i)$ exist, and therefore we can easily prove uniform L_2 -boundedness of $K_k(\tilde{\theta}, y_i)$. This in turn implies uniform integrability of $\|K_k(\tilde{\theta}, y_i)\|$ by theorem 12.10 in Davidson (1994). $\square$

A.4.1 Proof of Theorem 3

The FDMLE for θ is based on a regular likelihood function (without incidental parameters), which is differentiable at $r^S = 1$ (see lemma 12). Therefore, it follows by standard asymptotic results that the FDMLE for θ is consistent whether $N \rightarrow \infty$ or $T \rightarrow \infty$.

For joint N, T consistency of $\hat{\theta}_{FDMLE}$, i.e. $\hat{\theta}_{FDMLE} \xrightarrow{p} \theta$ as $(N, T \rightarrow \infty)$ (jointly), it is sufficient to show that $\frac{1}{NT} \sum_{i=1}^N \text{tr}[(DV(\bar{r})D')^{-1}Dy_i y_i' D'] \xrightarrow{p} \bar{\sigma}^2$ as $(N, T \rightarrow \infty)$. Let $X_{i,T} = \frac{1}{T-1} \text{tr}[(DV(\bar{r})D')^{-1}Dy_i y_i' D']$. It is easily seen that the random variables $X_{i,T}/\bar{\sigma}^2$ are i.i.d. χ^2 across $i \forall T$, with $E(X_{i,T}) = \bar{\sigma}^2 \forall T$. Since $y_{i,t}$ is Gaussian with finite moments, all moments of $X_{i,T}$ exist, and therefore $X_{i,T}$ is uniform L_2 -bounded. Application of theorem 12.10 in Davidson (1994) yields that $\|X_{i,T}\|$ is uniformly integrable in T . Moreover, $X_{i,T} \xrightarrow{d} X_i$ with $X_i = \bar{\sigma}^2$. Since all the conditions of corollary 1 in Phillips and Moon (1999), henceforth PM, are satisfied, we conclude that $\frac{1}{N} \sum_{i=1}^N X_{i,T} \xrightarrow{p} \bar{\sigma}^2$ as $(N, T \rightarrow \infty)$.

It is straightforward to obtain the large T , fixed N asymptotic distribution of $\hat{\theta}_{FDMLE}$. We will therefore focus in this appendix on the joint N, T asymptotic distribution of $\hat{\theta} = \hat{\theta}_{FDMLE}$. Using results from Magnus (1978), we obtain the following Taylor expansion of $\hat{\theta}$ around $\bar{\theta}$:

$$0 = N^{-0.5} \sum_{i=1}^N G(\bar{\theta}, y_i) - \left[N^{-1} \sum_{i=1}^N H(\bar{\theta}, y_i) \right] (NT)^{0.5}(\hat{\theta} - \bar{\theta}) + \frac{1}{2} \left[N^{-1} \sum_{i=1}^N F(\bar{\theta}, \tilde{\theta}, y_i) \right] (NT)^{0.5}(\hat{\theta} - \bar{\theta}),$$

where $\tilde{\theta}$ is a convex combination of $\hat{\theta}$ and $\bar{\theta}$,
 $G(\bar{\theta}, y_i) = \frac{1}{2} T^{-0.5} \sum_{1 \leq s, t < T} (\bar{\Omega}_{s,t} - \Delta y_{i,s+1} \Delta y_{i,t+1}) \left(\frac{\partial[\Omega^{-1}]_{s,t}}{\partial \theta} \Big|_{\bar{\theta}} \right)$ (premultiply by $\tilde{T}$ instead of $T^{-0.5} I_2$ when $\bar{r} = 1$),

$$H(\bar{\theta}, y_i) = \frac{1}{2T} \sum_{1 \leq s, t < T} (\bar{\Omega}_{s,t} - \Delta y_{i,s+1} \Delta y_{i,t+1}) \left(\frac{\partial^2[\Omega^{-1}]_{s,t}}{\partial \theta \partial \theta'} \Big|_{\bar{\theta}} \right) - \frac{1}{2T} ([\frac{dvec \Omega}{d\theta'}] \Big|_{\bar{\theta}})' (\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1}) [\frac{dvec \Omega}{d\theta'} \Big|_{\bar{\theta}}]$$

(pre- and postmultiply by $\tilde{T}$ instead of $T^{-0.5} I_2$ when $\bar{r} = 1$), and

$$F_{k,\cdot}(\bar{\theta}, \tilde{\theta}, y_i) = (\hat{\theta} - \bar{\theta})' K_k(\tilde{\theta}, y_i), \quad k = 1, 2, \text{ where } F_{k,\cdot}(\bar{\theta}, \tilde{\theta}, y_i) \text{ is the } k\text{-th row of } F(\bar{\theta}, \tilde{\theta}, y_i) \text{ and}$$

$$K_k(\tilde{\theta}, y_i) = (T)^{-1} \frac{\partial^3 l(Dy_i; r, \sigma^2)}{\partial \theta \partial \theta' \partial \theta_k} \Big|_{\tilde{\theta}} \text{ (pre- and postmultiply by } \tilde{T} \text{ instead of } T^{-0.5} I_2 \text{ when } \bar{r} = 1).$$

Let $\tilde{\theta}$ be such that $d(\tilde{\theta}, \bar{\theta}) < R(\bar{\theta})$. It is easily seen that $K_k(\tilde{\theta}, y_i)$ are i.i.d. across $i \forall T$. It follows from Pierce (1971) that $E(K_k(\tilde{\theta}, y_i))$ exists $\forall T$. Moreover, it follows from lemma 14 that $\|K_k(\tilde{\theta}, y_i)\|$ is uniformly integrable in T . Lemma 13 then implies that $\left\{ (N)^{-1} \sum_{i=1}^N K_k(\tilde{\theta}, y_i) \right\}$ is uniformly bounded in probability as $(N, T \rightarrow \infty)$. This result and consistency of $\hat{\theta}$ imply that $N^{-1} \sum_{i=1}^N F(\bar{\theta}, \tilde{\theta}, y_i) \xrightarrow{p} 0$ as $(N, T \rightarrow \infty)$. It follows that $(NT)^{0.5}(\hat{\theta} - \bar{\theta}) - \left[N^{-1} \sum_{i=1}^N H(\bar{\theta}, y_i) \right]^{-1} N^{-0.5} \sum_{i=1}^N G(\bar{\theta}, y_i) \xrightarrow{d} 0$ as $(N, T \rightarrow \infty)$, that is,

$(NT)^{0.5}(\hat{\theta} - \bar{\theta})$ and $\left[N^{-1} \sum_{i=1}^N H(\bar{\theta}, y_i)\right]^{-1} N^{-0.5} \sum_{i=1}^N G(\bar{\theta}, y_i)$ have the same joint N, T asymptotic distributions.

We will derive the joint N, T asymptotic distribution of $\hat{\theta}$ by applying some results of PM. Let $Y_{i,T} = G(\bar{\theta}, y_i) = T^{-0.5} \left\{ \frac{1}{2} \sum_{1 \leq s, t < T} (\bar{\Omega}_{s,t} - \Delta y_{i,s+1} \Delta y_{i,t+1}) \frac{\partial [\Omega^{-1}]_{s,t}}{\partial \theta} \right\}$. Then the following results can easily be verified: the $Y_{i,T}$ are i.i.d. across $i \forall T$, with $E(Y_{i,T}) = 0$, and $E(Y_{i,T} Y'_{i,T}) = \Phi_T(\bar{\theta}) = \frac{1}{2T} ([\frac{dvec\Omega}{d\theta'} |_{\bar{\theta}}]' (\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1}) [\frac{dvec\Omega}{d\theta'} |_{\bar{\theta}}])$; $\liminf_T \lambda_{\min}(\Phi_T) > 0$, where $\lambda_{\min}(\Phi_T)$ is the smallest eigenvalue of Φ_T ; and $\lim_{T \rightarrow \infty} \Phi_T(\bar{\theta}) = \Phi(\bar{\theta})$. Furthermore, using similar arguments as above, it can be shown that $\|Y_{i,T}\|^2$ is uniformly integrable in T .

Then $N^{-0.5} \sum_{i=1}^N Y_{i,T} \xrightarrow{d} N(0, \Phi(\bar{\theta}))$ by theorem 3 of PM.

Next, we will show that $N^{-1} \sum_{i=1}^N H(\bar{\theta}, y_i) \xrightarrow{p} -\Phi$ as $(N, T \rightarrow \infty)$. Let $Z_{i,T}(\bar{\theta}) = H(\bar{\theta}, y_i) + \Phi_T(\bar{\theta})$. It is easily seen that $Z_{i,T}(\bar{\theta})$ are i.i.d. across $i \forall T$, with $E(Z_{i,T}(\bar{\theta})) = 0 \forall T$. Moreover, $\|Z_{i,T}\|$ is uniformly integrable in T , and $Z_{i,T}(\bar{\theta}) \xrightarrow{d} 0$ by lemma 13.

Then it follows by corollary 1 in PM that $\frac{1}{N} \sum_{i=1}^N Z_{i,T}(\bar{\theta}) \xrightarrow{p} 0$ as $(N, T \rightarrow \infty)$, that is, $\frac{1}{N} \sum_{i=1}^N H(\bar{\theta}, y_i) \xrightarrow{p} \lim_{T \rightarrow \infty} -\frac{1}{2T} ([\frac{dvec\Omega}{d\theta'} |_{\bar{\theta}}]' (\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1}) [\frac{dvec\Omega}{d\theta'} |_{\bar{\theta}}]) = -\Phi(\bar{\theta})$ as $(N, T \rightarrow \infty)$.

It follows by Cramér's theorem that $(NT)^{0.5}(\hat{\theta} - \bar{\theta}) \xrightarrow{d} N(0, [\Phi(\bar{\theta})]^{-1})$ as $(N, T \rightarrow \infty)$, where $[\Phi(\bar{\theta})]^{-1} = \begin{pmatrix} 1 - \bar{r}^2 & 0 \\ 0 & 2\bar{\sigma}^4 \end{pmatrix}$ if $\bar{\theta} \in \Theta_r^S \times (0, \infty)$ and $[\Phi(\bar{\theta})]^{-1} = \begin{pmatrix} 8 & 0 \\ 0 & 2\bar{\sigma}^4 \end{pmatrix}$ if $\bar{\theta} \in \Theta_r^{UR} \times (0, \infty)$. $\square$

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

A.5.1 The unconditional FEML estimator

The stationary AR(1) panel model with fixed effects is given by $(y_i - \mu_i \iota) \equiv u_i \sim N(0, \sigma^2 V_T(r))$, $i = 1, \dots, N$, (*i.i.d.*). The unconditional log-likelihood function $l_{U,FE}(y; r, \mu, \sigma^2)$ can be decomposed as $l(Dy; r, \sigma^2) + \sum_{i=1}^N l(\tilde{\mu}_i; r, \mu_i, \sigma^2)$, where $\tilde{\mu}_i = \frac{\iota' V^{-1} y_i}{\iota' V^{-1} \iota}$ and $l(\tilde{\mu}_i; r, \mu_i, \sigma^2) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 + \frac{1}{2} \log \iota' V^{-1} \iota - \frac{1}{2\sigma^2} (\iota' V^{-1} \iota) (\tilde{\mu}_i - \mu_i)^2$ (see Kruiniger (1998)). Furthermore, $\iota' V^{-1} \iota = (1-r)((T-2)(1-r) + 2)$.

Inconsistency of the unconditional FEMLE when T is fixed:

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

The large- N probability limits of the (scaled) concentrated likelihood equations corresponding to r and σ^2 are not satisfied for the true values of r and σ^2 , because the FDMLE for r and σ^2 is consistent and the derivatives of $-\frac{1}{2} \log \sigma^2 + \frac{1}{2} \log \iota' V^{-1} \iota$ with respect to r and σ^2 are different from zero at the true values of r and σ^2 . It follows that the unconditional (stationary) FEMLE for r and σ^2 is inconsistent.

Consistency of the unconditional FEMLE when T tends to infinity:

The concentrated log-likelihood function $l_{U,FE}(y; r, \tilde{\mu}, \sigma^2)$ divided by NT equals $\frac{1}{NT} l(Dy; r, \sigma^2) - \frac{1}{2T} \log(2\pi) - \frac{1}{2T} \log \sigma^2 + \frac{1}{2T} \log \iota' V^{-1} \iota$. As $T \rightarrow \infty$ regardless of the asymptotic behaviour of N , 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 $\text{plim}_{T \rightarrow \infty} (\hat{r}_{FEMLE} - \hat{r}_{FDMLE}) = 0$ and $\text{plim}_{T \rightarrow \infty} (\hat{\sigma}_{FEMLE}^2 - \hat{\sigma}_{FDMLE}^2) = 0$. Since the FDMLE for r and σ^2 is consistent under large T , arbitrary N asymptotics, this must also be the case for the FEMLE for r and σ^2 .

Large T, arbitrary N asymptotic distribution of the unconditional FEMLE:

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

The scaled Hessian of the concentrated log-likelihood function can be written as $(NT)^{-1} \frac{\partial^2 l_{U,FE}(y; r, \tilde{\mu}, \sigma^2)}{\partial \theta \partial \theta'} = (NT)^{-1} \frac{\partial^2 l(Dy; r, \sigma^2)}{\partial \theta \partial \theta'} + (NT)^{-1} \sum_{i=1}^N \frac{\partial^2 l(\tilde{\mu}_i; r, \tilde{\mu}_i, \sigma^2)}{\partial \theta \partial \theta'}$, where $\theta = (r \ \sigma^2)'$.

From $\lim_{T \rightarrow \infty} (NT)^{-1} \sum_{i=1}^N \frac{\partial^2 l(\tilde{\mu}_i; r, \tilde{\mu}_i, \sigma^2)}{\partial \theta \partial \theta'} = \lim_{T \rightarrow \infty} T^{-1} \frac{\partial^2 (-0.5 \log \sigma^2 + 0.5 \log \iota' V^{-1} \iota)}{\partial \theta \partial \theta'} = 0$,

we have $\text{plim}_{N, T \rightarrow \infty} (NT)^{-1} \frac{\partial^2 l_{U,FE}(y; r, \tilde{\mu}, \sigma^2)}{\partial \theta \partial \theta'} = \text{plim}_{N, T \rightarrow \infty} (NT)^{-1} \frac{\partial^2 l(Dy; r, \sigma^2)}{\partial \theta \partial \theta'}$, regardless of the relative rate of increase of N and T . The latter probability limit is given in appendix A.4:

$$\text{plim}_{N, T \rightarrow \infty} (NT)^{-1} \frac{\partial^2 l(Dy; r, \sigma^2)}{\partial \theta \partial \theta'} = -\Phi = - \begin{pmatrix} 1 - r^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix}^{-1}.$$

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

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

$$\text{Now, } (NT)^{-1/2} \sum_{i=1}^N \frac{\partial l(\tilde{\mu}_i; r, \tilde{\mu}_i, \sigma^2)}{\partial \theta} = (N/T)^{1/2} \frac{\partial (-0.5 \log \sigma^2 + 0.5 \log \iota' V^{-1} \iota)}{\partial \theta}.$$

Since $\frac{\partial \log \iota' V^{-1} \iota}{\partial r} = \frac{-2(T-2)(1-r)-2}{(1-r)((T-2)(1-r)+2)}$, we have $\lim_{T \rightarrow \infty} \frac{\partial \log \iota' V^{-1} \iota}{\partial r} = \frac{-2}{(1-r)}$. Moreover, $\frac{\partial \log \sigma^2}{\partial \sigma^2} = \frac{1}{\sigma^2}$.

We conclude that the large T , arbitrary N asymptotic distribution of the unconditional MLE for r and σ^2 is identical to the large T , arbitrary N asymptotic distribution of the FDMLE for r and σ^2 , apart from the fact that the former has an asymptotic bias term that is equal to $-\frac{1}{T} \Phi^{-1} \left(\frac{1}{(1-r)} \quad \frac{1}{2\sigma^2} \right)' = -\frac{1}{T} (1 + r \quad 1)'$. Thus as $T \rightarrow \infty$, regardless of whether N is fixed or tends to infinity,

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

A.5.2 The unconditional REML estimator

The stationary AR(1) panel model with random effects is given by $(y_i - c) \equiv v_i \sim N(0, \sigma_\mu^2 \iota \iota' + \sigma^2 V(r))$, $i = 1, \dots, N$, (*i.i.d.*). Let $W = \sigma_\mu^2 \iota \iota' + \sigma^2 V(r)$. The unconditional log-likelihood function $l_{U,RE}(y; r, c, \sigma_\mu^2, \sigma^2)$ can be decomposed as $l_{C,RE}(y|y_1; r, c, \sigma_\mu^2, \sigma^2) + l_{M,RE}(y_1; r, c, \sigma_\mu^2, \sigma^2)$, where $l_{M,RE}(y_1; r, c, \sigma_\mu^2, \sigma^2) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma_\mu^2 + \frac{\sigma^2}{1-r^2}) - \frac{1}{2}(\sigma_\mu^2 + \frac{\sigma^2}{1-r^2})^{-1} \sum_{i=1}^N (y_{i,1} - c)^2$.

The conditional REML estimator, which is based on the conditional likelihood function $l_{C,RE}(y|y_1; r, c, \sigma_\mu^2, \sigma^2)$, has been considered by Alvarez and Arellano (albeit with $c = 0$).

Consistency of the unconditional REMLE:

The REMLE for $(r, c, \sigma_\mu^2, \sigma^2)$ is based on a regular likelihood function which does not depend on incidental parameters and which is differentiable. Therefore, it follows by standard

asymptotic results that the REMLE is consistent under large N and arbitrary T asymptotics (provided that $T \geq 2$). When $T \rightarrow \infty$ and N is fixed, the REMLE for r and σ^2 is still consistent, although the REMLE for σ_μ^2 is inconsistent. This can be seen as follows. When $T \rightarrow \infty$, $\frac{1}{NT}l_{M,RE}(y_1; r, c, \sigma_\mu^2, \sigma^2) \rightarrow 0$. Thus the large T limiting criterion for the unconditional REMLE is equal to the large T limiting criterion for the conditional REMLE. Consistency of the unconditional REMLE for r and σ^2 then follows from consistency of the conditional REMLE for r and σ^2 , which was shown in Alvarez and Arellano (consistency of $\hat{\sigma}_{CREML}^2$ follows from the limiting likelihood equation (A111) in their paper).

Large N,T asymptotic distribution of the unconditional REMLE:

The scaled score vector of the log-likelihood function can be written as $(NT)^{-1/2} \frac{\partial l_{U,RE}(y; r, c, \sigma_\mu^2, \sigma^2)}{\partial \theta} = (NT)^{-1/2} \frac{\partial l_{C,RE}(y; r, c, \sigma_\mu^2, \sigma^2)}{\partial \theta} + (NT)^{-1/2} \frac{\partial l_{M,RE}(y; r, c, \sigma_\mu^2, \sigma^2)}{\partial \theta}$, where now $\theta = (r \ c \ \sigma_\mu^2 \ \sigma^2)'$. When $N, T \rightarrow \infty$, we have $(NT)^{-1/2} \frac{\partial l_{M,RE}(y; r, c, \sigma_\mu^2, \sigma^2)}{\partial \theta} = o_p(1)$.

The scaled Hessian matrix of the log-likelihood function can be written as $(NT)^{-1} \frac{\partial^2 l_{U,RE}(y; r, c, \sigma_\mu^2, \sigma^2)}{\partial \theta \partial \theta'} = (NT)^{-1} \frac{\partial^2 l_{C,RE}(y; r, c, \sigma_\mu^2, \sigma^2)}{\partial \theta \partial \theta'} + (NT)^{-1} \frac{\partial^2 l_{M,RE}(y; r, c, \sigma_\mu^2, \sigma^2)}{\partial \theta \partial \theta'}$. When $N, T \rightarrow \infty$, we also have $(NT)^{-1} \frac{\partial^2 l_{M,RE}(y; r, c, \sigma_\mu^2, \sigma^2)}{\partial \theta \partial \theta'} = o_p(1)$.

It follows that the large N, T asymptotic distribution of the unconditional REMLE is equal to the large N, T asymptotic distribution of the conditional REMLE. Alvarez and Arellano have shown for $c = 0$ that $(NT)^{1/2}(\hat{r}_{CREML}^2 - r) \xrightarrow{d} N(0, 1 - r^2)$ when $N, T \rightarrow \infty$, provided $0 \leq \lim(N/T) < \infty$. However, when $c \neq 0$, the same result can be obtained: the proof follows Alvarez and Arellano verbatim. We conclude that $(NT)^{1/2}(\hat{r}_{REML}^2 - r) \xrightarrow{d} N(0, 1 - r^2)$ when $N, T \rightarrow \infty$, provided $0 \leq \lim(N/T) < \infty$.

For fixed T the unconditional REMLE has a different asymptotic variance.

B PROOFS OF RESULTS IN SECTION 3

B.1 Asymptotic properties of the OMD estimator

In this appendix we discuss the large N and fixed T asymptotic properties of the OMDE.

Theorem 15 *The OMDE for θ in the inclusive ARMA(p, q) panel model with fixed effects is consistent and its large N and fixed T asymptotic distribution is given by*

$$\sqrt{N}(\hat{\theta}_{OMD} - \theta) \xrightarrow{d} N(0, COV_{OMD}) \quad (41)$$

where

$$COV_{OMD} = ([\frac{d \text{vech} \Omega}{d\theta'}|_{\bar{\theta}}]'(E((\text{vech} Z_i)(\text{vech} Z_i)') - (\text{vech} \bar{\Omega})(\text{vech} \bar{\Omega})')^{-1}[\frac{d \text{vech} \Omega}{d\theta'}|_{\bar{\theta}}])^{-1}$$

PROOF

For details see also appendix A. Identification of $\bar{\theta}$ follows from $\text{plim}_{N \rightarrow \infty} Z = E(Z_i) = \bar{\Omega}$, $\bar{\Omega} = \Omega(\theta)$ iff $\theta = \bar{\theta}$, and the fact that the limit of the weighting matrix is nonsingular; moreover, we can find a compact subset of Θ , say $\bar{\Theta}$; it follows from INCL and results in VdL that $\Omega(\theta)$ is differentiable (and hence continuous) in every $\theta \in \bar{\Theta}$ and therefore that $\sup_{\theta \in \bar{\Theta}} \|\Omega(\theta)\| < \infty$, where $\|\cdot\|$ is the Euclidean matrix norm. (For example: in the AR(1) panel model with fixed effects, $r = -1$ is ruled out by assumption INCL, and $\Omega(r = 1, \sigma^2) = \sigma^2 I$). This together implies consistency of $\hat{\theta}_{OMD}$ (see Newey and McFadden (1994)).

The asymptotic distribution can easily be obtained by performing a Taylor expansion around the first order condition corresponding to the minimization problem in (19) $\square$.

We now compare the OMDE with the FDMLE:

Theorem 16 *Under normality the FDMLE for the inclusive ARMA(p, q) panel model with fixed effects, $\hat{\theta}_{FDMLE}$, is asymptotically equivalent to the OMDE, $\hat{\theta}_{OMD}$, when $N \rightarrow \infty$ and T is fixed.*

PROOF

Let $\check{D}_{T-1}(= \check{D})$ denote the duplication matrix corresponding to a square matrix of dimension $T - 1$, let $\check{D}_{T-1}^+ = (\check{D}'\check{D})^{-1}\check{D}'$ (the Moore-Penrose inverse of $\check{D}$), and let P_{T-1} be the

commutation matrix (see Magnus and Neudecker (1988)). Then we have under normality $E((vecZ_i)(vecZ_i)' - (vec\bar{\Omega})(vec\bar{\Omega})' = (\bar{\Omega} \otimes \bar{\Omega})(I_{(T-1)^2} + P_{T-1})$. Using that $P_{T-1}\check{D}_{T-1} = \check{D}_{T-1}$, we obtain $COV_{OMD}^{-1} = [\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(E((vechZ_i)(vechZ_i)' - (vech\bar{\Omega})(vech\bar{\Omega})')^{-1}[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}] = [\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(\check{D}_{T-1}^+(\bar{\Omega} \otimes \bar{\Omega})(I_{(T-1)^2} + P_{T-1})(\check{D}_{T-1}^+)')^{-1}[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}] = \frac{1}{2}[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(\check{D}^+(\bar{\Omega} \otimes \bar{\Omega})(\check{D}^+)')^{-1} \times [\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}] = \frac{1}{2}[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(\check{D}'\check{D})(\check{D}'(\bar{\Omega} \otimes \bar{\Omega})\check{D})^{-1}(\check{D}'\check{D})\check{D}^+[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}] = -\frac{1}{2}[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(\check{D}'\check{D})(\check{D}'(\bar{\Omega} \otimes \bar{\Omega})\check{D})^{-1}\check{D}'(\bar{\Omega} \otimes \bar{\Omega})[\frac{dvec\Omega^{-1}}{d\theta'}|_{\bar{\theta}}] = -\frac{1}{2}[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(\check{D}'\check{D})(\check{D}'(\bar{\Omega} \otimes \bar{\Omega})\check{D})^{-1}\check{D}'(\bar{\Omega} \otimes \bar{\Omega})\check{D}[\frac{dvec\Omega^{-1}}{d\theta'}|_{\bar{\theta}}] = -\frac{1}{2}[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(\check{D}'\check{D})[\frac{dvec\Omega^{-1}}{d\theta'}|_{\bar{\theta}}] = \frac{1}{2}[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}]'(\bar{\Omega}^{-1} \otimes \bar{\Omega}^{-1})[\frac{dvec\Omega}{d\theta'}|_{\bar{\theta}}] = COV_{FDM L}^{-1} \square.$

B.2 Proof of theorem 7:

Note that the data are stationary and normal and that T is fixed (i.e. finite).

The (Optimal) Linear GMM estimator exploits all the second order moment conditions that are implied by covariance stationarity, while the OMD estimator exploits a subset of these moment conditions, namely the moment conditions in $E(\text{vech}(Z_i)) - \text{vech}(\Omega(\theta)) = 0$.

We will prove the theorem by showing for arbitrary fixed $T \geq 3$ that there exists a moment condition that is i) exploited by the OLGMM estimator but not by the OMD estimator, and ii) not redundant relative to the moment conditions that are optimally exploited by the OMD estimator.

We will show that $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ satisfies conditions i) and ii) for any $T \geq 3$.

First, the OLGMM estimator exploits $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ for any $T \geq 3$ (see (21)). On the other hand, it is easily seen that the OMD estimator does not exploit $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$, because this moment condition involves data in levels.

To prove non-redundancy of $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$, we consider a GMM estimator that exploits both $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ and the moment conditions in $E(\text{vech}(Z_i)) - \text{vech}(\Omega(\theta)) = 0$. We will denote this estimator by GMM^+ . Furthermore, let $f(\theta) = (y_{i,1}\Delta\varepsilon_{i,3} \ (\text{vech}(Z_i) - \text{vech}(\Omega(\theta))))'$.

Note that under stationarity and normality $N^{-1} \sum_{i=1}^N \text{Var}(y_{i,1}\Delta\varepsilon_{i,3}) = 2\sigma^4(1-r^2)^{-1} + 2N^{-1} \sum_{i=1}^N \mu_i^2 \sigma^2$, while $N^{-1} \sum_{i=1}^N E[(\text{vech}(Z_i) - \text{vech}(\Omega(\theta)))(\text{vech}(Z_i) - \text{vech}(\Omega(\theta)))'] = \check{D}_{T-1}^+(\bar{\Omega} \otimes \bar{\Omega})(I_{(T-1)^2} + P_{T-1})(\check{D}_{T-1}^+)'$, where $\check{D}_{T-1}^+$ and P_{T-1} have been defined in appendix B.1. The latter does not involve $N^{-1} \sum_{i=1}^N \mu_i^2$. The covariance of $y_{i,1}\Delta\varepsilon_{i,3}$ and $\text{vech}(Z_i) - \text{vech}(\Omega(\theta))$ does not involve $N^{-1} \sum_{i=1}^N \mu_i^2$ either. It follows that the optimal weighting matrix of the GMM^+ estimator, $[E(f(\theta)f(\theta)')]^{-1}$, will depend on $N^{-1} \sum_{i=1}^N \mu_i^2$.

It is also easily verified that $E(\frac{df}{d\theta'})'$ does not depend on $N^{-1} \sum_{i=1}^N \mu_i^2$. Note that $y_{i,1}\Delta\varepsilon_{i,3} = y_{i,1}(\Delta y_{i,3} - r\Delta y_{i,2})$. Then we have under stationarity $E(\frac{d(y_{i,1}\Delta\varepsilon_{i,3})}{dr}) = -E(y_{i,1}\Delta y_{i,2}) = \frac{1}{1+r}\sigma^2$.

We conclude that the variance Optimal GMM^+ estimator for r depends on $N^{-1} \sum_{i=1}^N \mu_i^2$, whereas the variance of the OMD estimator for r does not depend on $N^{-1} \sum_{i=1}^N \mu_i^2$. It follows that $\text{Var}(\hat{r}_{\text{OMD}}) > \text{Var}(\hat{r}_{\text{GMM}^+})$, and therefore that $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ is not redundant relative to the moment conditions that are optimally exploited by the OMD estimator. $\square$

C PROOF OF RESULTS IN SECTION 4

C.1 Parameter orthogonalization and using a uniform prior is equivalent to using Jeffrey's prior:

We will prove that Lancaster's two-step Bayesian procedure is equivalent to integrating the incidental parameters out of the model with the original parametrization with respect to Jeffrey's invariant noninformative prior $\det(IM)^{0.5}$, where IM stands for the Information Matrix. Let $L(y, \phi)$ denote the likelihood function, let ϕ be the old parameter vector (including the $\check{\mu}'_i$ s) and let η be the new parameter vector (including the ν'_i s) chosen such that IM_η is diagonal (the parameters in η are information orthogonal). Moreover let $J = \frac{d\eta}{d\phi'}$. Now, if the prior $\pi(\eta)$ is chosen equal to $\det(IM_\eta)^{0.5}$, which is a noninformative, 'uniform' prior, then

$$\begin{aligned}
 & \int \cdots \int L(y|\phi) \pi(\eta) \left(\prod_{i=1}^N \frac{d\nu_i}{d\check{\mu}_i} \right) d\check{\mu}_1 \dots d\check{\mu}_N = \\
 & \int \cdots \int L(y|\phi) \pi(\eta) d\nu_1 \dots d\nu_N = \\
 & \int \cdots \int L(y|\phi) \det(IM_\eta)^{0.5} d\nu_1 \dots d\nu_N = \\
 & \int \cdots \int L(y|\phi) \det(IM_\eta)^{0.5} \det(J) \left(\prod_{i=1}^N \left(\frac{d\nu_i}{d\check{\mu}_i} \right)^{-1} \right) d\nu_1 \dots d\nu_N = \\
 & \int \cdots \int L(y|\phi) \det(IM_\eta)^{0.5} \det(J) d\check{\mu}_1 \dots d\check{\mu}_N = \\
 & \int \cdots \int L(y|\phi) \det(IM_\phi)^{0.5} d\check{\mu}_1 \dots d\check{\mu}_N
 \end{aligned} \tag{42}$$

where the first line in (42) corresponds to (27). The fourth line makes use of $\det(J) = \prod_{i=1}^N \frac{d\nu_i}{d\check{\mu}_i} = \exp(N\beta(r))$ and the last line follows from the invariance property of Jeffrey's prior (see also Zellner (1971) p. 47). Although Lancaster's reparametrization approach yields exactly the same marginal posterior distribution and estimator as the approach which uses the original parametrization and Jeffrey's prior $\det(IM_\phi)^{0.5}$, in the case at hand Lancaster's approach is more convenient, since $\det(IM_\phi)^{0.5}$ is quite a complicated function.

C.2 Large T, arbitrary N asymptotic results for the Modified MLE for the conditional AR(1) panel model with fixed effects.

In the sequel we will need the results in the following lemma:

Lemma 17 *Let the vector $\check{\lambda}_1 = \check{\lambda}_1(r)$, the matrix $\check{P} = \check{P}(r)$, and the function $\beta(r)$ be defined as in section 4. Then $\lim_{T \rightarrow \infty} \frac{1}{T} \text{tr}(\check{P}'Q\check{P}) = \frac{1}{1-r^2}$, $\lim_{T \rightarrow \infty} \frac{1}{T} \text{tr}(\check{P}'Q\check{P}\check{P}'Q\check{P}) = \frac{1+r^2}{(1-r^2)^3}$, $\lim_{T \rightarrow \infty} T^{-1}\check{\lambda}_1'Q\check{\lambda}_1 = 0$, $\text{tr}(Q\check{P}) = -\beta'(r)$, $\lim_{T \rightarrow \infty} \frac{1}{T}\beta'(r) = 0$, $\lim_{T \rightarrow \infty} \frac{1}{T}\beta''(r) = 0$, and $\lim_{T \rightarrow \infty} \frac{1}{T} \text{tr}(Q\check{P}Q\check{P}) = 0$.*

PROOF

$$\lim_{T \rightarrow \infty} \frac{1}{T} \text{tr}(\check{P}'Q\check{P}) = \lim_{T \rightarrow \infty} \frac{1}{T} \text{tr}(\check{P}'\check{P}) = \lim_{T \rightarrow \infty} \frac{1}{T} \frac{1}{1-r^2} (T - 1 - r^2 \left(\frac{1-r^{2(T-1)}}{1-r^2} \right)) = \frac{1}{1-r^2};$$

$$\text{likewise, } \lim_{T \rightarrow \infty} \frac{1}{T} \text{tr}(\check{P}'Q\check{P}\check{P}'Q\check{P}) = \lim_{T \rightarrow \infty} \frac{1}{T} \text{tr}(\check{P}'\check{P}\check{P}'\check{P}) = \frac{1+r^2}{(1-r^2)^3};$$

it is straightforward to verify that $\lim_{T \rightarrow \infty} T^{-1}\check{\lambda}_1'Q\check{\lambda}_1 = 0$, and $\text{tr}(Q\check{P}) = -\beta'(r)$;

$$\lim_{T \rightarrow \infty} \frac{1}{T}\beta'(r) = 0, \text{ because } 0 \leq \lim_{T \rightarrow \infty} \frac{1}{T}\beta'(r) = \lim_{T \rightarrow \infty} \frac{1}{T} \frac{1}{T-1} \sum_{t=1}^{T-2} (T-1-t)r^{t-1} \leq$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-3} r^t = \lim_{T \rightarrow \infty} \frac{1}{T} \frac{1-r^{T-2}}{1-r} = 0;$$

$$\lim_{T \rightarrow \infty} \frac{1}{T}\beta''(r) = 0, \text{ since } 0 \leq \lim_{T \rightarrow \infty} \frac{1}{T}\beta''(r) = \lim_{T \rightarrow \infty} \frac{1}{T} \frac{1}{T-1} \sum_{t=1}^{T-2} (T-1-t)(t-1)r^{t-2} \leq$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-4} (t+1)r^t = \lim_{T \rightarrow \infty} \frac{1}{T} \frac{1}{1-r} \left(\sum_{t=0}^{T-4} r^t - (T-3)r^{T-3} \right) = 0; \text{ and}$$

$\lim_{T \rightarrow \infty} \frac{1}{T} \text{tr}(Q\check{P}Q\check{P}) = 0$ is again easily verified noting that $\text{tr}(Q\check{P}Q\check{P}) = \text{tr}(\check{P}\check{P}) - \frac{2}{T}\iota'\check{P}\check{P}\iota + \frac{1}{T^2}(\iota'\check{P}\iota)^2$. $\square$

Let $\theta = (r \ \sigma^2)'$, and let upperbars denote true values of the parameters. The modified MLE for θ , $\hat{\theta}_{MLL}$, is defined as a solution to the modified likelihood equations, i.e. the first order conditions for a maximum of the modified log-likelihood function. The modified log-likelihood function for the conditional AR(1) panel model with fixed effects, divided by NT , is given by

$$\log \tilde{p}(r, \sigma^2) = \frac{1}{NT} \sum_{i=1}^N \log \tilde{p}_i(r, \sigma^2), \quad (43)$$

$$\text{where } \log \tilde{p}_i(r, \sigma^2) = \beta(r) - (T-2) \log \sigma - \frac{1}{2\sigma^2} (y_i - ry_{i,-1})' Q (y_i - ry_{i,-1}),$$

with $\beta(r) = \frac{1}{T-1} \sum_{t=1}^{T-2} \frac{T-1-t}{t} r^t$. We have assumed that $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 < \infty$.

Since (43) is not a regular log-likelihood function, we cannot use standard asymptotic results for ML estimators to prove that $\hat{\theta}_{MML}$ is consistent.

Kruiniger (1998, 2002) has shown large N , fixed T consistency of the $\hat{\theta}_{MML}$. When N is fixed and $T \rightarrow \infty$, $\log \tilde{p}(r, \sigma^2)$ converges uniformly in θ to the same probability limit as the within log-likelihood function. It follows that $\hat{\theta}_{MML} - \hat{\theta}_{LSDV} \xrightarrow{p} 0$ and therefore that $\hat{\theta}_{MML}$ is consistent when N is fixed and $T \rightarrow \infty$, because the LSDV estimator is consistent under large T , fixed N asymptotics.

We will now prove joint N, T consistency of $\hat{\theta}_{MML}$, i.e. $\hat{\theta}_{MML} \xrightarrow{p} \theta$ as $(N, T \rightarrow \infty)$, by showing that $\log \tilde{p}(r, \sigma^2)$ converges in probability uniformly in θ when $(N, T) \rightarrow \infty$.

We can rewrite $(y_i - ry_{i,-1})'Q(y_i - ry_{i,-1})$ as $\varepsilon_i'Q\varepsilon_i + 2(\bar{r} - r)y_{i,-1}'Q\varepsilon_i + (\bar{r} - r)^2 y_{i,-1}'Qy_{i,-1}$. Moreover, $\varepsilon_i'Qy_{i,-1} = \varepsilon_i'Q(y_{i,-1} - \mu_i) = \varepsilon_i'Q(\check{P}\varepsilon_i + \check{\lambda}_1(y_{i,1} - \mu_i))$, and $y_{i,-1}'Qy_{i,-1} = (y_{i,-1} - \mu_i)'Q(y_{i,-1} - \mu_i) = \varepsilon_i'\check{P}'Q\check{P}\varepsilon_i + 2\varepsilon_i'\check{P}'Q\check{\lambda}_1(y_{i,1} - \mu_i) + \check{\lambda}_1'Q\check{\lambda}_1(y_{i,1} - \mu_i)^2$.

Since ε_i is i.i.d. Gaussian with finite moments and $\frac{1}{T}\varepsilon_i'Q\varepsilon_i \xrightarrow{q.m.} \bar{\sigma}^2$ as $T \rightarrow \infty$, $\frac{1}{NT} \sum_{i=1}^N \varepsilon_i'Q\varepsilon_i \xrightarrow{p} \bar{\sigma}^2$ as $(N, T \rightarrow \infty)$ by corollary 1 of Phillips and Moon (1999), (henceforth PM).

Likewise, $\frac{1}{NT} \sum_{i=1}^N \varepsilon_i'\check{P}'Q\check{P}\varepsilon_i \xrightarrow{p} \bar{\sigma}^2/(1 - \bar{r}^2)$ as $(N, T \rightarrow \infty)$, and $\frac{1}{NT} \sum_{i=1}^N \varepsilon_i'Q\check{P}\varepsilon_i \xrightarrow{p} 0$ as $(N, T \rightarrow \infty)$ by corollary 1 of PM, lemma 15 and $E((\varepsilon_i'M_1\varepsilon_i)(\varepsilon_i'M_2\varepsilon_i)) = \bar{\sigma}^4(tr(M_1)tr(M_2) + tr(M_1M_2 + M_1'M_2))$ for arbitrary constant square matrices M_1 and M_2 .

We can show that $\frac{1}{NT} \sum_{i=1}^N \varepsilon_i'Q\check{\lambda}_1(y_{i,1} - \mu_i) \xrightarrow{p} 0$ as $(N, T \rightarrow \infty)$ by applying theorem 1 of PM. Let $X_{i,T} = T^{-1}\varepsilon_i'Q\check{\lambda}_1(y_{i,1} - \mu_i)$. The random variables $X_{i,T}$ are independent across i , $\forall T$, $E(X_{i,T}) = 0$, and $X_{i,T} \xrightarrow{d} X_i = 0$, as $T \rightarrow \infty$, $\forall i$. It follows immediately that conditions (ii) and (iv) of theorem 1 of PM are satisfied. Condition (iii) of theorem 1 holds because $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 < \infty$, and $\lim_{T \rightarrow \infty} T^{-1}\check{\lambda}_1'Q\check{\lambda}_1 = 0 < \infty$, and therefore $\limsup_{N,T} N^{-1} \sum_{i=1}^N E(X_{i,T}^2) = \limsup_{N,T} T^{-2}\check{\lambda}_1'Q\check{\lambda}_1 N^{-1} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 = 0 < \infty$. Finally, noting that $N^{-1} \sum_{i=1}^N E|X_{i,T}| = (E|T^{-1}\varepsilon_i'Q\check{\lambda}_1|) N^{-1} \sum_{i=1}^N |y_{i,1} - \mu_i|$, condition (i) follows by applying the Cauchy-Schwarz inequality, which yields $E|T^{-1}\varepsilon_i'Q\check{\lambda}_1| \leq (T^{-2}\bar{\sigma}^2\check{\lambda}_1'Q\check{\lambda}_1)^{0.5}$, and $N^{-1} \sum_{i=1}^N |y_{i,1} - \mu_i| \leq \left(N^{-1} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 \right)^{0.5}$, and by lemma 15.

Likewise, $\frac{1}{NT} \sum_{i=1}^N \varepsilon_i'\check{P}'Q\check{\lambda}_1(y_{i,1} - \mu_i) \xrightarrow{p} 0$ as $(N, T \rightarrow \infty)$ and $T^{-1}\check{\lambda}_1'Q\check{\lambda}_1 N^{-1} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 \xrightarrow{p} 0$ as $(N, T \rightarrow \infty)$.

Hence $\frac{1}{NT} \sum_{i=1}^N \varepsilon'_i Q y_{i,-1} \xrightarrow{p} 0$ as $(N, T \rightarrow \infty)$, and $\frac{1}{NT} \sum_{i=1}^N y'_{i,-1} Q y_{i,-1} \xrightarrow{p} \bar{\sigma}^2 \check{P}(\bar{r})' Q \check{P}(\bar{r})$ as $(N, T \rightarrow \infty)$.

Joint N, T consistency of $\hat{\theta}_{MML}$, i.e. $\hat{\theta}_{MML} \xrightarrow{p} \theta$ as $(N, T \rightarrow \infty)$, now follows immediately.

Noting that $\hat{\theta}_{MML}$ and the Within MLE for θ are asymptotically equivalent when N is fixed and $T \rightarrow \infty$, it is straightforward to obtain the large T , fixed N asymptotic distribution of $\hat{\theta}_{MML}$.

We will now derive the joint N, T asymptotic distribution of $\hat{\theta} = \hat{\theta}_{MML}$. We obtain the following Taylor expansion of $\hat{\theta}$ around $\bar{\theta}$:

$(NT)^{0.5}(\hat{\theta} - \bar{\theta}) = \left[N^{-1} \sum_{i=1}^N H(\tilde{\theta}, y_i) \right]^{-1} N^{-0.5} \sum_{i=1}^N G(\bar{\theta}, y_i)$, where $\tilde{\theta}$ is a convex combination of $\hat{\theta}$ and $\bar{\theta}$, $G(\theta, y_i) = T^{-0.5} \frac{\partial}{\partial \theta} \log \tilde{p}_i(\theta)$, and $H(\theta, y_i) = \frac{1}{T} \frac{\partial^2}{\partial \theta \partial \theta'} \log \tilde{p}_i(\theta)$.

We obtain the joint N, T asymptotic distribution of $N^{-0.5} \sum_{i=1}^N G(\bar{\theta}, y_i)$ by applying theorem 2 of PM:

Let $Y_{i,T} = G(\bar{\theta}, y_i)$, and $MIM_{i,T} = \frac{1}{T} \begin{pmatrix} tr(Q \check{P} Q \check{P}) + \frac{E(y'_{i,-1} Q y_{i,-1})}{\bar{\sigma}^2} & -\frac{\beta'(\bar{r})}{\bar{\sigma}^2} \\ -\frac{\beta'(\bar{r})}{\bar{\sigma}^2} & \frac{T-1}{2\bar{\sigma}^4} \end{pmatrix}$, where

$E(y'_{i,-1} Q y_{i,-1}) = E((y_{i,-1} - \mu_i) Q (y_{i,-1} - \mu_i)) = \bar{\sigma}^2 tr(\check{P}' Q \check{P}) + (y_{i,1} - \mu_i)^2 \check{\lambda}'_1 Q \check{\lambda}_1$. Then, $E(Y_{i,T}) = 0$, and $E(Y_{i,T} Y'_{i,T}) = \Phi_{i,T}(\bar{\theta}) = (MIM_{i,T})$.

Using the results in lemma 15, it is easily verified that $\liminf_T \lambda_{\min}(\Phi_{i,T}) > 0$, where $\lambda_{\min}(\Phi_{i,T})$ is the smallest eigenvalue of $\Phi_{i,T}$, and $\lim_{T \rightarrow \infty} \Phi_{i,T}(\bar{\theta}) = \Phi(\bar{\theta})$, where $[\Phi(\bar{\theta})]^{-1} =$

$$\begin{pmatrix} 1 - \bar{r}^2 & 0 \\ 0 & 2\bar{\sigma}^4 \end{pmatrix}.$$

Because $Y_{i,T}$ is a vector, we will use the Cramér-Wold device to derive the asymptotic distribution of $N^{-0.5} \sum_{i=1}^N Y_{i,T}$. Let $\xi_{i,n,T} = \lambda' Y_{i,T} / s_{n,T}$, where $s_{n,T}^2 = \sum_{i=1}^N \lambda' \Phi_{i,T} \lambda$. Since $y_{i,t}$ and $\varepsilon_{i,t}$ are Gaussian, all moments of $Y_{i,T}$ exist, and therefore $\xi_{i,n,T}$ is uniform $L_{4+\delta}$ bounded. This implies that $\xi_{i,n,T}^4$ is uniform integrable in T by theorem 12.10 in Davidson (1994). It follows by theorem 2 of PM that $\sum_{i=1}^N \xi_{i,n,T} \xrightarrow{d} N(0, 1)$ as $(N, T \rightarrow \infty)$ for arbitrary λ . We conclude that $N^{-0.5} \sum_{i=1}^N G(\bar{\theta}, y_i) \xrightarrow{d} N(0, [\Phi(\bar{\theta})]^{-1})$ as $(N, T \rightarrow \infty)$.

Next, we will show that $N^{-1} \sum_{i=1}^N H(\tilde{\theta}, y_i) \xrightarrow{p} -\Phi(\bar{\theta})$ as $(N, T \rightarrow \infty)$.

Above we have argued that $\hat{\theta}_{MML} \xrightarrow{p} \bar{\theta}$ as $(N, T \rightarrow \infty)$. Let $Z_{i,T} = H(\theta, y_i)$ and

$$EMH_{i,T}(\theta) = \frac{1}{T} \begin{pmatrix} \beta''(r) - \frac{E(y'_{i,-1} Q y_{i,-1})}{\sigma^2} & -\frac{1}{2\sigma^4} E(y_{i,-1}' Q (y_i - r y_{i,-1})) \\ -\frac{1}{2\sigma^4} E(y_{i,-1}' Q (y_i - r y_{i,-1})) & -\frac{1}{2\sigma^6} E(y_i - r y_{i,-1})' Q (y_i - r y_{i,-1}) \end{pmatrix}.$$

We will now show that $\frac{1}{N} \sum_{i=1}^N Z_{i,T} \xrightarrow{p} -\Phi(\theta)$ uniformly in θ as $(N, T \rightarrow \infty)$ by verifying the conditions of theorem 1 of PM. The random matrices $Z_{i,T}$ are independent across i , $\forall T$, $E(Z_{i,T}) = EMH_{i,T}(\theta)$, and $\lim_{T \rightarrow \infty} EMH_{i,T}(\theta) = -\Phi(\theta)$, and $Z_{i,T} \xrightarrow{d} Z_i = -\Phi(\theta)$, as $T \rightarrow \infty \forall i$. Above, we have shown that $\frac{1}{NT} \sum_{i=1}^N y'_{i,-1} Q y_{i,-1} \xrightarrow{p} \bar{\sigma}^2 \check{P}(\bar{r})' Q \check{P}(\bar{r})$ as $(N, T \rightarrow \infty)$, $\frac{1}{NT} \sum_{i=1}^N \varepsilon'_i Q y_{i,-1} \xrightarrow{p} 0$ as $(N, T \rightarrow \infty)$, and $\frac{1}{NT} \sum_{i=1}^N \varepsilon'_i Q \varepsilon_i \xrightarrow{p} \bar{\sigma}^2$ as $(N, T \rightarrow \infty)$ by theorem 1 and corollary 1 of PM and by lemma 15.

Then it follows immediately that $N^{-1} \sum_{i=1}^N H(\tilde{\theta}, y_i) \xrightarrow{p} -\Phi(\bar{\theta})$ as $(N, T \rightarrow \infty)$.

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, 1998, The time series and cross-section asymptotics of dynamic panel data estimators, working paper no. 9808, CEMFI, Madrid.
- [4] Andersen, E.B., 1970a, Asymptotic properties of conditional maximum-likelihood estimators, *JRSS 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] Andersen, E.B., 1973, Conditional inference and models for measuring, *Mentalhygiejnisk Forlag*, Copenhagen.
- [7] Blundell, R., and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of econometrics* 87, 115-143.
- [8] Bond, S., Nauge, C., and F. Windmeijer, 2002, Unit roots and identification in autoregressive panel data models: a comparison of alternative tests, mimeo., IFS, London.
- [9] 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.
- [10] Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski, 1999, Redundancy of moment conditions, *Journal of econometrics* 91, 89-111.
- [11] Chamberlain, G., 1982, Multivariate regression models for panel data, *Journal of econometrics* 18, 5-46.
- [12] Chamberlain, G., 1984, Panel data, in *Handbook of econometrics*, edited by Z. Griliches and M. Intriligator, North Holland, Amsterdam.

- [13] Davidson. J., 1994, *Stochastic limit theory*, Oxford University Press.
- [14] Hahn, J., and G. Kuersteiner, 2000, Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, unpublished manuscript, M.I.T., Cambridge, Mass.
- [15] 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.
- [16] Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu, 1998, Maximum likelihood estimation of fixed effects dynamic panel models covering short time periods, unpublished manuscript, revised 2001, University of Southern California.
- [17] 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.
- [18] Kruiniger, H., 1998, Conditional maximum likelihood estimation of dynamic panel data models, University College London Economics Working Paper 98-04, also available at <http://www.qmw.ac.uk/~ugte185/>
- [19] Kruiniger, H., 2000, GMM estimation of dynamic panel data models with persistent data, working paper no. 428, Queen Mary, University of London, also available at <http://www.qmw.ac.uk/~ugte185/>
- [20] Kruiniger, H., 2002, On the estimation of panel regression models with fixed effects, working paper no. 450, Queen Mary, University of London, also available at <http://www.qmw.ac.uk/~ugte185/>
- [21] Lancaster, T., 1997, Orthogonal parameters and panel data, working paper no. 97-32, revised in 1999 and 2001, Brown University, Providence, Rhode Island.
- [22] MaCurdy, T., 1981a, Multiple time series models applied to panel data, NBER working paper no. W0646, Cambridge, Massachusetts.
- [23] MaCurdy, T., 1981b, Asymptotic properties of quasi-maximum likelihood estimators and test statistics, NBER technical working paper no. T0014, Cambridge, Massachusetts.

- [24] 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.
- [25] 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.
- [26] Magnus, J.R., and H. Neudecker, 1988, *Matrix differential calculus with applications in statistics and econometrics*, Wiley, Chicester, United Kingdom.
- [27] Newey, W., and D. McFadden, 1994, Large sample estimation and hypothesis testing, Ch. 36 in *Handbook of econometrics*, vol. 4, R.F. Engle and D. McFadden eds., Elsevier, Amsterdam.
- [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.
- [33] Zellner, A., 1971, *An introduction to Bayesian inference in econometrics*, Wiley, New York.
- [34] Zilliak, J.P., 1997, Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators, *Journal of Business & Economic Statistics* 15, 419-431.

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2002 Hugo Kruiniger
All rights reserved.**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmul.ac.uk
Website: www.econ.qmul.ac.uk/papers/wp.htm**