

On the estimation of panel regression models with fixed effects

Hugo Kruiniger*
Department of Economics
Queen Mary, University of London

This version February 2002.

Abstract

This paper considers estimation of panel data models with fixed effects. First, we show that a consistent “unrestricted fixed effects” estimator does not exist for autoregressive panel data models with initial conditions. We derive necessary and sufficient conditions for the consistency of estimators for these models. In particular, we show that various widely used GMM estimators for the conditional AR(1) panel model are inconsistent under trending fixed effects sequences. Next, we derive, justify, and compare restricted Fixed Effects GMM and (Q)ML estimators for this model. We find that the FEML estimator is asymptotically efficient, whereas the Modified ML estimator is not. We also compare the fixed effects approach for estimating the conditional AR(1) panel model and covariance parameters in static panel data models with the correlated random effects approach.

Keywords: fixed effects, correlated effects, (essentially) random effects, conditional likelihood, modified likelihood, GMM, quasi likelihood, cross-sectional dependence.

JEL Classification: C11, C14, C23.

*Email: h.kruiniger@qmul.ac.uk. I should like to thank Gary Chamberlain and Rod McCrorie for helpful discussions. This research was funded by the ESRC under grant R000239139.

1. Introduction.

One of the advantages of panel data over cross-sectional data when estimating regression models is the possibility to allow for so-called ‘individual’ effects that can capture unobservable, time-invariant effects which may be correlated with the observable variables. Whilst in most econometric applications these individual effects can be modeled as (correlated) Random Effects (RE), in some cases a Fixed Effects (FE) approach should be adopted in order to obtain consistent estimators. In this paper we will consider GMM and ML estimation of two panel regression models with fixed effects: a panel AR(1) model with initial conditions and a static panel model with covariance parameters.

Various widely used GMM estimators for dynamic panel data models, e.g. those due to Arellano and Bond (1991), Ahn and Schmidt (1995), and Arellano and Bover (1995), exploit moment conditions that involve differences of the data. However, since these moment conditions also involve levels of the data, the corresponding GMM estimators are not consistent for all fixed effects sequences, unlike e.g. the First Difference Maximum Likelihood Estimators for stationary panel models (see MaCurdy (1981a)). For instance, the Arellano-Bond GMM estimator (1991) which uses instruments in levels is inconsistent if the sequence of squared individual mean effects follows a linear trend. Such a situation could arise, for instance, when modeling the behaviour of firms of different sizes.

In this paper we will prove an impossibility result and derive necessary and sufficient conditions for the consistency of fixed effects estimators for *conditional* dynamic panel data models, that is, for dynamic panel models with initial conditions. In particular, we will show that a consistent “unrestricted fixed effects” estimator does not exist for these models, and we will derive a necessary condition (that is sometimes also sufficient) for the consistency of any fixed effects estimator for the conditional AR(1) panel model. This condition is related to the deviations of the initial conditions from the individual means. Subsequently, we will propose various restricted FE GMM estimators and a FE (Quasi) ML estimator that are consistent for arbitrary sequences of initial conditions and individual means that satisfy this necessary condition for consistency. The aforementioned necessary condition is also sufficient for the consistency of the Bayesian/Modified Maximum Likelihood estimator due to Lancaster (1997), but not for the consistency of GMM estimators which involve levels of the data.

The FEML estimator for the conditional AR(1) panel model has also been proposed by Hsiao et al. (1998). Their derivation and justification of the FEML estimator are different from ours, which are based on the necessary condition for consistency mentioned above.¹ The FEML estimator can also be viewed as a FE

¹Hsiao et al.’s derivation and justification of the FEML estimator are based on unnecessarily

Quasi ML estimator. We will show that the latter is consistent under much weaker and more reasonable assumptions than the “pure” FEML estimator. In particular, we can relax the assumption that the first differences of the data have a common mean and a common variance, and the assumption that they are independent along the cross-section dimension of the panel.

Next, we will establish the connection under normality and (cross-sectional) homoskedasticity between the FE and (Correlated) RE (Q)ML estimators for the conditional AR(1) panel model, and their GMM counterparts. As a by-product, we find that the FEML estimator is asymptotically efficient under these conditions.

We will also show that the FEML estimator for the conditional AR(1) panel model is asymptotically more efficient than the Bayesian/Modified ML estimator of Lancaster (1997). This implies that the Modified ML approach of Neyman and Scott (1948) and the Bayesian approach (see Lindley and El-Sayyed (1968), Lancaster (1997), and Kruiniger (1998)) do not always yield the most efficient FE estimator if the ordinary ML estimator is inconsistent. The Conditional ML approach yields an inconsistent estimator in this case.

Finally, we will compare various fixed effects estimators with the corresponding correlated random effects estimators. It is well known that in a static regression model a correlated random effects estimator for the slope parameters is (asymptotically) equivalent to a fixed effects estimator when the individual effects are correlated with all regressors (see e.g. Hsiao (1986)). This result holds true under any intertemporal covariance matrix of the disturbances. Below we will show that this result does not carry over to estimators for the parameters of the covariance matrix itself: in that case a correlated random effects approach leads to a more efficient estimator than a fixed effects approach. We will also compare the fixed effects approach to estimation with the random effects approach for the conditional AR(1) panel data model within a framework due to Chamberlain (1980). In particular, we will show that the RE (Quasi) ML estimator is more efficient than the FE (Quasi) ML estimator for this model.

The outline of the paper is as follows. Section 2 derives a necessary condition for consistency of any estimator for the conditional AR(1) model. It also discusses sufficient conditions for the consistency of various widely used GMM estimators for this model, proposes new FE GMM estimators, derives their asymptotic distributions under various assumptions about the individual mean effects sequences, and examines the properties of the Modified ML estimator and the Conditional ML estimator. Section 3 derives the “pure” FEML estimator and the FEQML estimator for the conditional AR(1) panel model, gives an alternative justification

strong assumptions. On the other hand their paper does not mention that the fixed effects have to satisfy a restriction for the FEML estimator to be consistent.

for the FEML estimator, and compares the FE(Q)ML estimator with the Modified ML estimator, the Conditional ML estimator, and the (correlated) RE(Q)ML estimator. Section 4 presents and compares various ML and GMM estimators for covariance parameters in static panel regression models. Section 5 concludes.

2. Fixed effects estimators for dynamic panel models with fixed initial conditions.

One possible way of dealing with individual effects in regression models is to apply a ‘within’ transformation to the model such that the individual effects drop out. In this section we will examine the asymptotic properties of two types of estimators for the AR(1) panel model with initial conditions that make use of such a transformation: GMM estimators, e.g. those due to Arellano and Bond (1991) and Ahn and Schmidt (1995), and a Maximum Likelihood (ML) type estimator, i.e. the Modified ML Estimator due to Lancaster (1997). Notwithstanding the application of the transformation, the consistency of the GMM and ML-type estimators for the conditional AR(1) panel model does still depend on the asymptotic behaviour of the sequence of the individual effects. As a matter of fact, the distribution of the transformed data still depends on the individual effects.

We will present necessary and sufficient conditions on the sequences of individual effects and initial conditions for consistency of estimators for the conditional AR(1) panel model. It turns out that a necessary condition for the consistency of *any* estimator for the conditional AR(1) panel model is that the average of the squared differences between the initial conditions and individual effects converges in probability. This condition will also be sufficient for consistency in the case of the Modified ML estimator, and for what we will call Fixed Effects GMM estimators. We stress that the necessary condition for consistency does not require that $(p)\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \mu_i^2 = \sigma_\mu^2 < \infty$.

We will also examine the Conditional ML estimator for the conditional AR(1) panel model.² This estimator will be shown to be inconsistent.

The AR(1) panel model with fixed initial conditions and fixed individual effects, μ_i , is given by

$$y_i = \rho y_{i,-1} + (1 - \rho)\mu_i \iota + \varepsilon_i, \quad -1 < \rho \leq 1, \quad (2.1)$$

where $y_i = (y_{i,2} \dots y_{i,T})'$, $y_{i,-1} = (y_{i,1} \dots y_{i,T-1})'$, ι is a vector of ones, and $i \in \mathcal{I} = \{1, 2, \dots, N\}$. So for each individual unit we have T observations on y , including the initial observation $y_{i,1}$. The designation ‘fixed’ means that we

²In the likelihood context a consistent estimator for the parameters of interest can sometimes be obtained by conditioning on sufficient statistics for the incidental parameters, and maximizing the resulting conditional joint density of the data. This is the Conditional Maximum Likelihood approach (see e.g. Andersen (1970, 1973)).

make minimal assumptions about the initial conditions and individual effects respectively. In the unit root case the individual effects disappear.

Let $\mathcal{T} = \{2, \dots, T\}$. We make the following assumptions (cf Ahn and Schmidt (1997)):

Standard assumptions (SA):

$\varepsilon_i \sim (0, \sigma_i^2 I_{T-1})$, $i = 1, \dots, N$, (*ind. dist.*); $E(|\varepsilon_{i,t}|^{2+\delta}) < \Delta_1 < \infty$, for some $\delta > 0$, and $\Delta_1 > 0$, $\forall i \in \mathcal{I}$, and $\forall t \in \mathcal{T}$; $|E(y_{i,t})| < \infty$, $E(y_{i,1}\varepsilon_i) = 0$, $\forall i \in \mathcal{I}$; and $|\mu_i| < \infty$, $E(\mu_i \varepsilon_i) = 0$, $\forall i \in \mathcal{I}$, if $-1 < \rho < 1$.

Below we also assume that $\sigma^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^2 < \infty$, where $\sigma_i^2 = E(\varepsilon_{i,t}^2)$.

The version of the conditional AR(1) panel model which assumes SA implies the following restrictions on the second moments of y_i (cf Ahn and Schmidt):

$$\begin{aligned} E(y_{i,1} \Delta \varepsilon_{i,t}) &= 0, \quad t = 3, \dots, T, \\ E((y_{i,t} - \rho y_{i,t-1})^2 - (y_{i,2} - \rho y_{i,1})^2) &= 0, \quad t = 3, \dots, T, \quad \text{and} \\ E((y_{i,t} - \rho y_{i,t-1})(y_{i,s} - \rho y_{i,s-1}) - (y_{i,3} - \rho y_{i,2})(y_{i,2} - \rho y_{i,1})) &= 0, \\ &\quad 2 \leq s < t, \quad t = 4, \dots, T. \end{aligned} \tag{2.2}$$

The *standard* Arellano-Bond (AB) GMM estimator, $\hat{\rho}_{AB}$, exploits only the following moment conditions, which do not require homoskedasticity of the $\varepsilon_{i,t}$'s :

$$E(y_{i,s} \Delta \varepsilon_{i,t}) = 0, \quad 1 \leq s \leq t-2, \quad t = 3, \dots, T. \tag{2.3}$$

It is often useful to rewrite the conditional AR(1) panel model (2.1) as

$$y_i - \mu_i \iota = \rho(y_{i,-1} - \mu_i \iota) + \varepsilon_i, \quad -1 < \rho \leq 1. \tag{2.1'}$$

We will now introduce the notion of essentially random effects:

Definition 2.1. We will say that $\{\eta_i\}_{i=1}^\infty$ is a sequence of *essentially random effects* when $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \eta_i^2 = \sigma_\eta^2 < \infty$.

Notice that the definition of *essentially* random effects allows for individual effects η_i that are not random.

Let $v_{i,1} \equiv y_{i,1} - \mu_i$. In the sequel we will need the following assumption

Restricted Fixed Effects and Initial Conditions assumption (RFE):

- 1) $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}^2 = \sigma_v^2 < \infty$, and $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1} \varepsilon_{i,t} = 0$, $\forall t \in \mathcal{T}$,
- 2) $(p)\lim_{N \rightarrow \infty} N^{-1-\kappa} \sum_{i=1}^N \mu_i^2 = \sigma_\mu^2 < \infty$ for some $\kappa \geq 0$,
- 3) $N^{-1-\lambda} \sum_{i=1}^N \mu_i v_{i,1} \xrightarrow{d} \sigma_{\mu,v}$ for some $\lambda \geq 0$, where $|\sigma_{\mu,v}| < \infty$ if $\sigma_{\mu,v}$ is constant, and $\sigma_{\mu,v} \neq 0$ if $\lambda > 0$.

Notice that $\lambda \leq 0.5\kappa$ by the Cauchy-Schwarz inequality.

Below we will show that a Fixed Effects GMM or ML type estimator for ρ can only be consistent if the $v_{i,1}$'s are essentially random effects.

An alternative more restrictive assumption is:

Restricted Fixed Effects and Initial Conditions assumption ' (RFE)':

- 1) $v_{i,1}$, $i = 1, \dots, N$, are i.h.d. random variables that satisfy

$$E(v_{i,1}) = \bar{v}, \text{ and } E(|v_{i,1}|^{2+\delta}) < \Delta_1 < \infty \text{ for some } \delta, \Delta_1 > 0, \forall i \in \mathcal{I},$$

- 2) $(p)\lim_{N \rightarrow \infty} N^{-1-\kappa} \sum_{i=1}^N \mu_i^2 = \sigma_\mu^2 < \infty$ for some $\kappa \geq 0$,

- 3) $N^{-1-\lambda} \sum_{i=1}^N \mu_i v_{i,1} \xrightarrow{d} \sigma_{\mu,v}$ for some $\lambda \geq 0$, where

$$|\sigma_{\mu,v}| < \infty \text{ if } \sigma_{\mu,v} \text{ is constant, and } \sigma_{\mu,v} \neq 0 \text{ if } \lambda > 0.$$

Note that SA implies that $E(v_{i,1}\varepsilon_{i,t}) = 0$, $\forall i \in \mathcal{I}$ and $\forall t \in \mathcal{T}$, and that SA and RFE1' together imply that RFE1 holds. Although RFE' is weaker than assuming random individual effects, it imposes restrictions on the individual effects and initial conditions that — as we will see — are not necessary for consistency of fixed effects estimators of the conditional AR(1) panel model. Assumption RFE relaxes RFE' substantially because it allows $E(v_{i,1})$ to be different across individuals and it does not require cross-sectional independence of the $v_{i,1}$'s.

The conditional AR(1) model can be rewritten as

$$\begin{aligned} \Delta y_{i,2} &= (\rho - 1)(y_{i,1} - \mu_i) + \varepsilon_{i,2} \\ \Delta y_{i,t} &= \rho \Delta y_{i,t-1} + \Delta \varepsilon_{i,t} \quad t = 3, \dots, T. \end{aligned} \tag{2.4}$$

In appendix A, we prove the following lemmas:

Lemma 2.2. *If RFE1 (or SA and RFE1') holds, and $E(|\varepsilon_{i,t}|^{2+\delta}) < \Delta_1 < \infty$, $\forall i \in \mathcal{I}$, and $\forall t \in \mathcal{T}$, then $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,t} - \mu_i)^2 < \infty$, $\forall t \in \mathcal{T}$.*

Lemma 2.3. *If RFE1 (or SA and RFE1') holds, and $E(|\varepsilon_{i,t}|^{2+\delta}) < \Delta_1 < \infty$, $\forall i \in \mathcal{I}$, and $\forall t \in \mathcal{T}$, then $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,t})^2 < \infty$, $\forall t \in \mathcal{T}$.*

We also prove the reverse:

Lemma 2.4. *If $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,2})^2 < \infty$, $\forall t \in \mathcal{T}$, $E(|\varepsilon_{i,t}|^{2+\delta}) < \Delta_1 < \infty$, $\forall i \in \mathcal{I}$, and $\forall t \in \mathcal{T}$, and if $p\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}\varepsilon_{i,t} = 0$, $\forall t \in \mathcal{T}$, then $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}^2 = \sigma_v^2 < \infty$.*

Maximum Likelihood estimators for ρ in (2.1) as well as any reasonable GMM estimator for ρ depend on second order sample moments of the data. These estimators will be consistent only if these sample moments converge in probability.³ ⁴ The second order sample moments can be rewritten as sums of cross-sectional averages of squared levels, squared differences, and/or crossproducts of levels and differences of the data. If the fixed effects and initial conditions satisfy $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}^2 = \sigma_v^2 < \infty$, then lemma 2.3 implies that the cross-sectional averages of the squared differences will converge in probability, whereas cross-sectional averages of the squared levels may well diverge. The implication of lemma 2.4 is that $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}^2 = \sigma_v^2 < \infty$ is in fact a necessary condition for the convergence in probability of the cross-sectional averages of squared differences and therefore for the consistency of GMM or ML estimators for (2.1) which depend on these averages.⁵

We conclude that a fixed effects GMM or ML estimator for (2.1) should only depend on first differences of the data. Moreover, a fixed effects GMM or ML estimator for (2.1) that is consistent for any sequences $\{y_{i,1}\}$ and $\{\mu_i\}$ does not exist. However, consistent *restricted* fixed effects GMM and ML estimators for models that include assumption RFE1 do exist.

We note that RFE1, or $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,2})^2 < \infty$, is a reasonable assumption that is met in most applications, possibly after rescaling the data.

2.1. GMM estimation under fixed effects.

In the appendix we prove the following result:

Theorem 2.5. *Assume that SA and RFE hold, $\sigma_{\mu,v}$ is constant, $T = 3$, and $|\rho| < 1$. Then the standard Arellano-Bond GMM estimator $\hat{\rho}_{AB}$ is consistent if $N^{-2-2\lambda} \sum_{i=1}^N \mu_i^2 = o(1)$.*

PROOF

See Appendix A. Notice that $N^{-2-2\lambda} \sum_{i=1}^N \mu_i^2 = o(1)$ is equivalent to $\kappa < 1 + 2\lambda$.

³We require convergence in probability of these sample moments in order to prove uniform convergence of the criterion function (see e.g. Newey and McFadden (1994)).

⁴Under non-normality of the data, one may want to consider GMM estimators for ρ which also exploit information contained in third and higher order sample moments. In that case, issues similar to those discussed in this paper will arise. In particular, such a GMM estimator is a consistent fixed effects estimator only if $\{v_{i,1}\}$ satisfies a generalized version of condition RFE1. Below we will confine our attention to estimators that only exploit second moments of the data.

⁵As can be seen from (2.1') and (2.4), likelihood functions for the conditional AR(1) model can always be rewritten such that they only depend on cross-sectional averages of squares and cross-products of $(r - \rho)v_{i,1}$, $v_{i,1}$, and the disturbances, where r is the variable in the likelihood function that corresponds to the true parameter ρ .

When $\sigma_{\mu,v}$ is not constant, $\kappa < 1 + 2\lambda$ is no longer sufficient for the consistency of the standard AB GMM estimator.

Theorem 2.5. implies that if the individual effects are trending, then $\hat{\rho}_{AB}$ may be a poor estimator.

A panel of firm data may fit the conditions of the following example:

Example 2.6. Let RFE' hold with $E(v_{i,1}) = 0$ and $E(v_{i,1}^2) > \delta_1 > 0 \forall i \in \mathcal{I}$, and $\mu_i = \beta\sqrt{i}$. Then $\frac{1}{N} \sum_{i=1}^N \mu_i^2 = 0.5\beta^2(N+1) = O(N)$, and $\overline{\mu v} = \frac{1}{N} \sum_{i=1}^N \mu_i v_{i,1} \xrightarrow{d} N(0, \text{Var}(\overline{\mu v}))$, where $\text{Var}(\overline{\mu v}) = \lim_{N \rightarrow \infty} N^{-2} \sum_{i=1}^N \mu_i^2 E(v_{i,1}^2)$. Thus $\kappa = 1$ and $\lambda = 0$. Moreover, $\overline{y_1 \Delta \varepsilon_3} = \frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3} \xrightarrow{d} N(0, \text{Var}(\overline{y_1 \Delta \varepsilon_3}))$, where $\text{Var}(\overline{y_1 \Delta \varepsilon_3}) = \lim_{N \rightarrow \infty} N^{-2} \sum_{i=1}^N 2\mu_i^2 \sigma_i^2$. It follows that $\kappa = 1 + 2\lambda$, and that the standard AB GMM estimator, $\hat{\rho}_{AB}$, is inconsistent.

The theorem below describes how the asymptotic distribution of $\hat{\rho}_{AB}$ depends on the values of κ and λ :

Theorem 2.7. Let assumptions SA and RFE hold, let $\sigma_{\mu,v}$ be constant, $T = 3$, and $|\rho| < 1$. Moreover, let $\lim_{N \rightarrow \infty} N^{-1-\kappa} \sum_{i=1}^N E(\mu_i \Delta \varepsilon_{i,3})^2 = \zeta_1$, $0 < \zeta_1 < \infty$, and let $\lim_{N \rightarrow \infty} N^{-1-\kappa} \sum_{i=1}^N E(\mu_i \varepsilon_{i,2})^2 = \zeta_2$, $0 < \zeta_2 < \infty$. If $\kappa = \lambda = 0$, and $\{y_{i,1} \Delta \varepsilon_{i,3}\}$ satisfies the Lindeberg condition, and if also $0 < \zeta_0 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E(y_{i,1} \Delta \varepsilon_{i,3})^2 < \infty$, then

$$N^{0.5} (\hat{\rho}_{AB} - \rho) \xrightarrow{d} N(0, [(\rho - 1)\sigma_v^2 + (\rho - 1)\sigma_{\mu,v}]^{-2} \zeta_0).$$

If $\kappa > 0$, and $\{\mu_i \Delta \varepsilon_{i,3}\}$ satisfies the Lindeberg condition, then:

$$\text{if } 0 < \kappa < 1, \text{ and } \lambda = 0, N^{0.5(1-\kappa)} (\hat{\rho}_{AB} - \rho) \xrightarrow{d} N(0, [(\rho - 1)\sigma_v^2 + (\rho - 1)\sigma_{\mu,v}]^{-2} \zeta_1),$$

$$\text{if } 0 < 2\lambda \leq \kappa < 1 + 2\lambda, N^{0.5(1-\kappa+2\lambda)} (\hat{\rho}_{AB} - \rho) \xrightarrow{d} N(0, [(\rho - 1)\sigma_{\mu,v}]^{-2} \zeta_1).$$

Finally, if $\kappa \geq 1 + 2\lambda$, and both $\{\mu_i \Delta \varepsilon_{i,3}\}$ and $\{\mu_i \varepsilon_{i,2}\}$ satisfy the Lindeberg condition, then:

$$\text{if } \kappa = 1 \text{ and } \lambda = 0, (\hat{\rho}_{AB} - \rho) \xrightarrow{d} \frac{N(0, \zeta_1)}{(\rho - 1)\sigma_v^2 + (\rho - 1)\sigma_{\mu,v} + N(0, \zeta_2)},$$

$$\text{if } \kappa = 1 + 2\lambda, \text{ and } \lambda > 0, (\hat{\rho}_{AB} - \rho) \xrightarrow{d} \frac{N(0, \zeta_1)}{(\rho - 1)\sigma_{\mu,v} + N(0, \zeta_2)},$$

$$\text{if } \kappa > 1 + 2\lambda, (\hat{\rho}_{AB} - \rho) \xrightarrow{d} \frac{N(0, \zeta_1)}{N(0, \zeta_2)}.$$

PROOF

See Appendix A.

When the individual effects are trending, random sequences like $\{\mu_i \Delta \varepsilon_{i,3}\}$ can still satisfy the Lindeberg condition:

Example 2.8. Let $\sigma_h^2 = \sigma^2$, $\forall h \in \mathcal{I}$, and let $\mu_h = \beta h^\alpha$ with $\alpha > -0.5 \forall h \in \mathcal{I}$. Then $\sum_{i=1}^N E(\mu_i \Delta \varepsilon_{i,3})^2 = 2\sigma^2 \beta^2 \sum_{h=1}^N h^{2\alpha} = O(N^{2\alpha+1})$. More generally, if $0 < \delta_1 < \sigma_i^2 < \Delta_1 < \infty$, $\forall i \in \mathcal{I}$, then we still have $\sum_{i=1}^N E(\mu_i \Delta \varepsilon_{i,3})^2 = O(N^{2\alpha+1})$. We can show that $\{\mu_i \Delta \varepsilon_{i,3}\}$ satisfies the Lindeberg condition by applying theorem 23.18 in Davidson (1994). Take $X_i = \mu_i \Delta \varepsilon_{i,3}$, and $c_i = \mu_i$. Then $X_i^2/c_i^2 = (\Delta \varepsilon_{i,3})^2$. Since $E(|\varepsilon_{i,t}|^{2+\delta}) < \Delta_1 < \infty$, $\forall i \in \mathcal{I}$, $\forall t \in \mathcal{T}$, we obtain that $\Delta \varepsilon_{i,3}$ is uniformly $L_{2+\delta}$ bounded by applying the Minkowski inequality. It then follows from theorem 12.10 in Davidson that $\{X_i^2/c_i^2\}$ is uniformly integrable. Let $M_N = c_N = \mu_N$ and $s_N^2 = \sum_{i=1}^N E(\mu_i \Delta \varepsilon_{i,3})^2$. Since $\sup_N [NM_N^2/s_N^2] = \sup_N O(N^{-2\alpha})\beta^2 N^{2\alpha} = C < \infty$, we can conclude that $\{\mu_i \Delta \varepsilon_{i,3}\}$ satisfies the Lindeberg condition.

We have the following alternative to theorem 2.5, which involves assumption RFE1' instead of assumption RFE1:

Theorem 2.9. Let assumptions SA, RFE1', and RFE2' hold, $T = 3$, and $|\rho| < 1$. Let λ_1 be the smallest positive number ($\lambda_1 \geq 0$) such that $\lim_{N \rightarrow \infty} N^{-1-\lambda_1} \sum_{i=1}^N E(\mu_i v_{i,1}) = c_{\mu,v}$ with $|c_{\mu,v}| < \infty$ and $c_{\mu,v} \neq 0$ if $\lambda_1 > 0$. Moreover, let $\lim_{N \rightarrow \infty} N^{-1-\kappa_1} \sum_{i=1}^N \text{Var}(\mu_i[(\rho-1)v_{i,1} + \varepsilon_{i,2}]) = \zeta_3$, $0 < \zeta_3 < \infty$, for some $\kappa_1 \geq 0$, and let $\{\mu_i[(\rho-1)v_{i,1} + \varepsilon_{i,2}] - E(\mu_i(\rho-1)v_{i,1})\}$ satisfy the Lindeberg condition. Then the standard Arellano-Bond GMM estimator, $\hat{\rho}_{AB}$, is consistent if $\kappa_1 < 1 + 2\lambda_1$.

PROOF

See Appendix A.

The Arellano-Bond *differences* GMM estimator, $\hat{\rho}_{ABdif}$, uses lagged differences of $y_{i,t-1}$ as instruments instead of levels, that is, it exploits the following moment conditions

$$E(\Delta y_{i,s} \Delta \varepsilon_{i,t}) = 0, \quad 2 \leq s \leq t-2, \quad t = 4, \dots, T. \quad (2.5)$$

This estimator only involves differences of the data and will be consistent for any sequence of fixed effects as long as assumption RFE1 (or RFE1') holds.

Theorem 2.10. Assume that SA, and RFE1 hold. Then the Arellano-Bond differences GMM estimator, $\hat{\rho}_{ABdif}$, is consistent.

PROOF

See Appendix A.

The asymptotic distribution of $\hat{\rho}_{ABdif}$ is easily derived. We make the following assumptions:

Moment assumption 1 (MA1): $0 < \bar{\sigma}^4 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sigma_i^4 < \infty$.

Moment assumption 2 (MA2):

$$\sigma_{v,\varepsilon}^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E(v_{i,1}^2 \varepsilon_{i,t}^2) < \infty, \forall t \in \mathcal{T}.$$

Then it is straightforward to prove the following result:

Theorem 2.11. *Let assumptions SA, RFE1, MA1, and MA2 hold, $T = 4$, and $|\rho| < 1$. Moreover, let $\{\Delta y_{i,2} \Delta \varepsilon_{i,4}\}$ satisfy the Lindeberg condition.*

Then $N^{0.5} (\hat{\rho}_{ABdif} - \rho) \xrightarrow{d} N(0, [(\rho - 1)^2 \rho \sigma_v^2 + (\rho - 1) \sigma^2]^{-2} [2(\rho - 1)^2 \sigma_{v,\varepsilon}^2 + 2\bar{\sigma}^4])$.

When a GMM estimator for (2.1) exploits all the AB moment conditions that involve only first-differences of the data, (2.5), in an optimal way, the AB moment conditions that involve levels of the data, (2.3), will be redundant if $\kappa > 2\lambda$.

Corollary 2.12. *Assume that SA and RFE hold, $\sigma_{\mu,v}$ is constant, and that $|\rho| < 1$. In addition, let assumptions MA1, and MA2 hold and let $\{\Delta y_{i,s} \Delta \varepsilon_{i,t}\}$ satisfy the Lindeberg condition $\forall s, t \in \mathcal{T}$ that satisfy $s + 2 \leq t$. Then if $\kappa > 2\lambda$ the optimal Arellano-Bond GMM estimator for ρ only exploits the moment conditions in (2.5): $\hat{\rho}_{OAB} = \hat{\rho}_{OABdif}$.*

PROOF

Trivial. As can be seen from Theorems 2.7 and Theorem 2.11, if $\kappa > 2\lambda$ the rate of convergence of $\hat{\rho}_{AB}$ (when $T = 3$) is lower than that of $\hat{\rho}_{ABdif}$ (when $T = 4$).

Just like the standard AB GMM estimator for ρ does not exhaust the set of all the second moment conditions, the AB differences GMM estimator for ρ does not exhaust the set of all the second moment conditions that follow from assumption SA and involve only differences of the data.

The complete set of second moment conditions implied by the conditional AR(1) panel model corresponds to

$$E \left[\begin{pmatrix} y_{i,1} \\ \Delta y_{i,2} \\ D\varepsilon_i \end{pmatrix} \begin{pmatrix} y_{i,1} & \Delta y_{i,2} & (D\varepsilon_i)' \end{pmatrix} \right],$$

where D is the $(T-2 \times T-1)$ first difference matrix with $D_{k,k} = -1$ and $D_{k,k+1} = 1$, $k = 1, \dots, T-2$, and zeros elsewhere. The Random Effects Conditional GMM estimator for ρ exploits all these second moment conditions, or equivalently, the moment conditions in (2.2).

As we have seen above, GMM estimators for ρ which exploit moment conditions involving levels of the data, e.g. $y_{i,1}$, are not consistent for all sequences of fixed effects that satisfy RFE. On the other hand, Fixed Effects GMM estimators — which are consistent for any sequence of fixed effects that satisfies RFE1 —

only exploit moment conditions that only involve differences of the observations. Noting that $\Delta y_{i,2} = \varepsilon_{i,2} - (1 - \rho)v_{i,1}$, and $D\varepsilon_i = Dy_i - \rho Dy_{i,-1}$, SA, and RFE1 imply

$$\text{p} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[\begin{pmatrix} \Delta y_{i,2} \\ Dy_i - \rho Dy_{i,-1} \end{pmatrix} \begin{pmatrix} \Delta y_{i,2} & (Dy_i - \rho Dy_{i,-1})' \end{pmatrix} \right] = \quad (2.6)$$

$$\begin{bmatrix} \sigma^2 + (1 - \rho)^2 \sigma_v^2 & -\sigma^2 & 0' \\ -\sigma^2 & \sigma^2 H & \\ 0 & & \end{bmatrix},$$

where $H = DD'$.⁶

We will call the GMM estimator for ρ in the Conditional AR(1) panel model which exploits all the moment conditions in (2.6) the Fixed Effects Conditional GMM estimator. This GMM estimator is the solution of a third-order polynomial in ρ just like the Random Effects Conditional GMM estimator for ρ .

The ‘moment condition’ $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,2})^2 = \sigma^2 + (1 - \rho)^2 \sigma_v^2$ is redundant for estimating ρ , because it is the only moment condition in (2.6) that involves σ_v^2 . Cross-sectional heteroskedasticity does not pose a problem for estimating ρ , since the $2 + \delta$ -th moments of the $\varepsilon_{i,t}$, $E(|\varepsilon_{i,t}|^{2+\delta})$, are bounded $\forall i \in \mathcal{I}$, and $\forall t \in \mathcal{T}$, and the cross-sectional average of the σ_i^2 converges. One could, of course, also allow for heteroskedasticity in the time-series dimension of the panel at the cost of introducing additional parameters.

Compared with the (Optimal) FECGMM estimator for ρ , the (Optimal) RECGMM estimator exploits some additional moment conditions which involve levels of the data. These extra moment conditions are not redundant. Therefore, we have the following result:

Theorem 2.13. *The Optimal Random Effects Conditional GMM estimator for ρ , which exploits (2.2), is asymptotically more efficient than the Optimal Fixed Effects Conditional GMM estimator for ρ , which exploits (2.6).*

PROOF

See Appendix A.

A special case arises when the $\{y_{i,t}\}$ are stationary. In that case $E(v_{i,1}^2) = \sigma_i^2 / (1 - \rho^2)$ and it follows that $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,2})^2 = 2\sigma^2 / (1 + \rho)$. After substituting this moment condition for $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,2})^2 = \sigma^2 + (1 - \rho)^2 \sigma_v^2$ in (2.6), we can obtain the Minimum Distance estimator for the stationary model

⁶We can extend these results to the conditional AR(1) panel model with exogenous regressors: $y_i = \rho y_{i,-1} + \iota \mu_i (1 - \rho) + X_i \beta + \varepsilon_i$. In this model we define $v_{i,1} \equiv y_{i,1} - \mu_i - \bar{x}_i' \beta / (1 - \rho)$. Then we obtain $\Delta y_{i,2} = (x_{i,2} - \bar{x}_i)' \beta + \varepsilon_{i,2} - (1 - \rho)v_{i,1}$.

with fixed effects (see also Chamberlain (1984)). Under homoskedasticity and normality the OMD estimator is asymptotically equivalent to the Conditional MLE for the stationary model with fixed effects (see Kruiniger (1998, 2000a)).

2.2. Modified Maximum Likelihood estimation.

The conditional AR(1) panel model with fixed effects can also be estimated by the maximum likelihood method. Nickell (1981) has shown that the ordinary and the within maximum likelihood estimator for ρ in this model, which are both equal to the LSDV estimator for ρ , are inconsistent. However, Lancaster (1997) has proposed a Modified Maximum Likelihood Estimator (MMLE) for this model.⁷ We will briefly discuss this estimator and then derive a necessary and sufficient condition for consistency of this estimator.

For the remainder of this section we will assume that the $\varepsilon_{i,t}$ are normally distributed and homoskedastic: $\varepsilon_i|y_{i,1}, \mu_i \sim N(0, \sigma^2 I) \forall i \in \mathcal{I}$.

The Modified MLE for (ρ, σ^2) is equal to the maximizer [with respect to (r, s^2)] of the Modified Likelihood function, which is proportional to

$$\begin{aligned} \tilde{p}(r, s^2) &\equiv \prod_{i=1}^N \tilde{p}_i(r, s^2) \equiv \\ &\exp(N\xi(r)) \times s^{-N(T-2)} \prod_{i=1}^N \exp\left(-\frac{1}{2s^2} \sum_{t=2}^T (y_{i,t} - \bar{y}_i - r(y_{i,t-1} - \bar{y}_{i,-1}))^2\right), \\ \text{where } \xi(r) &= \frac{1}{T-1} \sum_{t=1}^{T-2} \frac{T-1-t}{t} r^t. \end{aligned} \tag{2.7}$$

Taking logarithms and dividing by the sample size (N) yields

$$\begin{aligned} \frac{1}{N} \ln \tilde{p}(r, s^2) &= \xi(r) - (T-2) \log s - \frac{1}{2s^2} \frac{1}{N} \sum_{i=1}^N (y_i - r y_{i,-1})' Q (y_i - r y_{i,-1}) \\ &= \xi(r) - (T-2) \log s \\ &\quad - \frac{1}{2s^2} \frac{1}{N} \sum_{i=1}^N (\varepsilon_i + (\rho - r) y_{i,-1})' Q (\varepsilon_i + (\rho - r) y_{i,-1}) \\ &= \xi(r) - (T-2) \log s - \frac{1}{2s^2} \frac{1}{N} \sum_{i=1}^N (\varepsilon_i' Q \varepsilon_i) \\ &\quad + \frac{1}{s^2} \frac{1}{N} \sum_{i=1}^N (\rho - r) y_{i,-1}' Q \varepsilon_i - \frac{1}{2s^2} \frac{1}{N} \sum_{i=1}^N (\rho - r)^2 y_{i,-1}' Q y_{i,-1}, \end{aligned} \tag{2.8}$$

⁷See Neyman and Scott (1948) for the Modified Maximum Likelihood approach.

where Q is the so-called within matrix. Since $Q\iota = 0$, the last two terms are equal to

$$\frac{1}{s^2} \frac{1}{N} \sum_{i=1}^N (\rho - r)(y_{i,-1} - \mu_i \iota)' Q \varepsilon_i - \frac{1}{2s^2} \frac{1}{N} \sum_{i=1}^N (\rho - r)^2 (y_{i,-1} - \mu_i \iota)' Q (y_{i,-1} - \mu_i \iota).$$

Using that $(y_i - \mu_i \iota) = \rho(y_{i,-1} - \mu_i \iota) + \varepsilon_i$, we can rewrite the last two terms as a function of ε_i and $(y_{i,1} - \mu_i)$.

To prove consistency of the Modified Maximum Likelihood Estimator we need to verify that $\frac{1}{N} \log \tilde{p}(r, s^2)$ converges uniformly in probability, among other conditions. Uniform convergence of $\frac{1}{N} \log \tilde{p}(r, s^2)$ follows from assumptions SA and RFE1. We therefore have the following result:

Theorem 2.14. *Let assumption SA hold, let $-1 < \rho < 1$, and let $\sigma^2 > 0$. Then the Modified MLE for (2.1) is consistent if and only if assumption RFE1 holds.*

PROOF

See appendix A. The proof follows Kruiniger (1998) apart from the uniform convergence part.⁸

When the model contains exogenous regressors $(X_i \beta(1 - \rho))$, assumption RFE1 has to be replaced by a more general version in order to prove consistency of the Modified MLE. After redefining $v_{i,1}$ as $y_{i,1} - \mu_i - \bar{x}_i' \beta$, we not only need to assume that $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}^2 = \sigma_v^2 < \infty$, and $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1} \varepsilon_{i,t} = 0$, but also $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1} Q X_i = \Sigma_{vqx}$, where Σ_{vqx} is constant and finite. We require similar assumptions in order to prove consistency of the FECGMM estimator for the model with exogenous regressors.

Both the FECGMM estimator for the model without exogenous regressors and the Modified MLE for that model depend only on the elements of $\frac{1}{N} \sum_{i=1}^N (Dy_i y_i' D')$. When $T = 3$, these estimators are equal to each other.

⁸As a matter of fact, we prove consistency of the Modified MLE for the more general conditional ARX(1) panel model with fixed effects and exogenous regressors. Lancaster (2001) proves consistency of this Modified MLE under the assumption that the $v_{i,1}$ are i.i.d. .

2.3. Conditional Maximum Likelihood estimation under fixed effects.

We can rewrite the conditional AR(1) panel model as

$$y_i - y_{i,1}\iota = \underline{\varphi}_1(\rho)[y_{i,1} - \mu_i] + \underline{P}(\rho)\varepsilon_i, \quad i = 1, \dots, N, \quad (2.9)$$

where

$$\underline{P}(\rho) = \begin{bmatrix} 1 & 0 & \cdot & \cdot & 0 \\ \rho & 1 & 0 & & \cdot \\ \cdot & \rho & 1 & 0 & \cdot \\ \cdot & & \rho & 1 & 0 \\ \rho^{T-2} & \cdot & \cdot & \rho & 1 \end{bmatrix}, \text{ and } \underline{\varphi}_1(\rho) = (\rho - 1)\underline{P}(\rho)\iota.$$

It is clear from formulation (2.9) that the conditional AR(1) panel model contains only one sequence of incidental parameters, i.e. $\{y_{i,1} - \mu_i\}$.

The Conditional ML approach studied by Andersen (1970, 1973) attempts to derive consistent estimators for the parameters of interest in a model by conditioning on sufficient statistics for the incidental parameters in that model.

Let $\underline{V}(\rho) = \sigma^2 \underline{P}(\rho) \underline{P}(\rho)'$. Then a sufficient statistic for $(y_{i,1} - \mu_i)$ is given by

$$\hat{y}_{i,1} - \hat{\mu}_i = (\underline{\varphi}_1' \underline{V}^{-1} \underline{\varphi}_1)^{-1} \underline{\varphi}_1' \underline{V}^{-1} (y_i - y_{i,1}\iota) = \frac{1}{T-1} \iota' \underline{P}(\rho)^{-1} (y_i - y_{i,1}\iota). \quad (2.10)$$

We note that the sufficient statistic for $(y_{i,1} - \mu_i)$ is not a true statistic since it depends on ρ .

The Conditional ML estimators for ρ and σ^2 in the conditional AR(1) panel model are then defined as the maximizers of the conditional likelihood function $\prod_{i=1}^N L_c(y_i - y_{i,-1} | \hat{y}_{i,1} - \hat{\mu}_i, y_{i,1})$. These Conditional ML estimators for ρ and σ^2 can be shown to be inconsistent, even if assumption RFE1 holds.⁹

Theorem 2.15. *Let assumptions RFE1 and SA hold, and let $\sigma^2 > 0$. Then the Conditional MLE for (2.1) which is based on $\prod_{i=1}^N L_c(y_i - y_{i,-1} | \hat{y}_{i,1} - \hat{\mu}_i, y_{i,1})$ is inconsistent.*

PROOF

See appendix A. See also Kruiniger (1998).

⁹ Andersen (1970) has only proven consistency of CMLE's that are obtained by conditioning on *true* statistics.

3. Fixed and random effects ML estimators for dynamic panel data models with initial conditions.

In this section we will derive the Fixed Effects ML estimator for the autoregressive parameter in the AR(1) panel model with fixed initial conditions. We will show that the correlated Random Effects ML estimator and the Fixed Effects ML estimator for this parameter are asymptotically equivalent to the ORECMM and OFECMM estimators, respectively, which were discussed in section 2. Finally, we will compare the FE MLE with the Modified MLE and the Conditional MLE.

In the sequel we will refer to the AR(1) panel model with initial conditions

$$\begin{aligned} y_i &= \rho y_{i,-1} + \check{\mu}_i + \varepsilon_i, \\ \text{where } \check{\mu}_i &\equiv (1 - \rho)\mu_i, \end{aligned} \tag{2.1}$$

as “the conditional model.”

3.1. Correlated random effects ML and GMM estimators.

Chamberlain (1980) describes a maximum likelihood approach for estimating ρ in the conditional model when both the individual effects and the initial conditions are essentially random. He allows the distribution of the individual effects to depend upon the initial conditions. This leads to the following ‘correlated effects’ specification ¹⁰

$$\check{\mu}_i = \pi y_{i,1} + v_i,$$

where v_i is a new random effect and where

$$\pi = (p) \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N y_{i,1} \check{\mu}_i}{\sum_{i=1}^N y_{i,1}^2},$$

so that the new random effects are uncorrelated with the initial conditions

$$(p) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,1} v_i = 0,$$

and

$$(p) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \check{\mu}_i^2 = \pi^2 (p) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 + (p) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_i^2.$$

¹⁰W.l.o.g. we will assume that $E(v_i) = 0 \forall i \in \mathcal{I}$. A nonzero common mean of the v_i ’s can be handled by including an intercept.

The model then becomes

$$\begin{aligned} y_i &= \rho y_{i,-1} + \pi y_{i,1} \iota + u_i, \\ \text{where } u_i &\equiv v_i \iota + \varepsilon_i. \end{aligned} \tag{3.1}$$

After adding the assumption that the u_i are normally distributed, application of the Maximum Likelihood method to this model will yield Chamberlain's correlated Random Effects ML estimator for ρ in the conditional AR(1) panel model.¹¹ This estimator is consistent and asymptotically efficient (see also appendix B).

A special case is $\pi = 0$. When $\pi = 0$, imposing this restriction leads to a more efficient Random Effects ML estimator, i.e. the Uncorrelated Random Effects MLE of Balestra and Nerlove (1966).

The parameter ρ in the conditional model can also be estimated by the Generalized Method of Moments. The formulation of the conditional model given in (3.1) implies the same moment conditions for ρ as the original formulation given in (2.1), i.e. the moment conditions in (2.2). When using formulation (2.1) of the conditional model, one loses four of the “second moment” conditions for the purpose of estimating ρ by eliminating the nuisance parameters, i.e. σ^2 , and the cross-sectional averages of $y_{i,1}^2$, $\check{\mu}_i^2$ and $y_{i,1}\check{\mu}_i$. When using formulation (3.1) of the model, one has to eliminate σ^2 , π , and the averages of $y_{i,1}^2$ and v_i^2 from the moment conditions. It is easy to see that there is a one-to-one relationship between both sets of parameters. Therefore, the optimal GMM estimator for ρ which exploits all the second moment conditions is the same for both versions of the conditional model. This estimator for ρ is the ORECGMM estimator.

We will now show that under normality and homoskedasticity the (correlated) REML estimator for ρ is asymptotically equivalent to the ORECGMM estimator for ρ .¹² The joint density $f(y_{i,1}, y_i)$ can be factorized as $g(y_i|y_{i,1})h(y_{i,1})$. Notice that $f(y_{i,1}, y_i) \sim f(y_{i,1}, u_i)$ and $g(y_i|y_{i,1}) \sim g(u_i)$. The conditional model does not impose any restrictions on (the parameters of) the marginal density of $y_{i,1}$, $h(y_{i,1})$; in particular, $h(y_{i,1})$ does not depend on the parameters in $g(y_i|y_{i,1})$. Therefore, Chamberlain's REMLE, which is based on $\prod_{i=1}^N g(y_i|y_{i,1})$, is equal to the Quasi MLE for the same parameters which is based on the product of the joint densities, $\prod_{i=1}^N f(y_{i,1}, y_i)$. Under normality and homoskedasticity, this QMLE for ρ

¹¹Sims (2000) proposed an estimation approach for the AR(1) panel model where the initial conditions and the individual means follow a bivariate (normal) distribution, which allows for correlation between both variables. Thus Sims' random effects approach also specifies a (marginal) distribution for the initial condition, whereas Chamberlain's approach does not, i.e. his approach still conditions on the initial observations and is therefore more general.

¹²If the conditional model contains an intercept, then the GMM estimator that is asymptotically equivalent to the REMLE also exploits the information in the first moments of the data.

is asymptotically equivalent to an Optimal Minimum Distance estimator for ρ , which in turn is equivalent to the ORECGMM estimator for ρ .

We note that if the $u_i \equiv v_i\iota + \varepsilon_i$, $i = 1, \dots, N$, are normally distributed and if the model is linear in u_i only the first and second moments matter (see also Ahn and Schmidt (1995)). Moreover, the ORECGMM estimator attains the semiparametric efficiency bound, while the REMLE attains the Cramér-Rao lowerbound.

3.2. Fixed effects ML and GMM estimators.

When neither the initial conditions nor the individual effects are essentially random, we can still use a variant of Chamberlain's approach to obtain a consistent (Quasi) ML estimator for ρ , i.e. the Fixed Effects (Quasi) ML estimator for ρ . We obtain this estimator by replacing μ_i by $y_{i,1} - v_{i,1}$, by assuming that the $v_{i,1}$ are (essentially) random effects¹³, i.e. satisfy RFE1' and homoskedasticity in the case of the FEMLE, or satisfy RFE1 in the case of the FEQMLE, and by assuming (or imposing) normality. This is equivalent to imposing the restriction $\pi = (1 - \rho)$ on the conditional model in (3.1) and treating the $v_i = -(1 - \rho)v_{i,1}$ as (essentially) random effects, which are drawn from a normal distribution. This leads to the following formulation of the conditional model

$$y_i = \rho y_{i,-1} + (1 - \rho)y_{i,1}\iota + u_i, \quad (3.2)$$

where $u_i \equiv v_i\iota + \varepsilon_i$,

with the additional assumption $u_i \sim N(0, (1 - \rho)^2 \sigma_v^2 \iota\iota' + \sigma^2 I)$.¹⁴

We have the following result:

Theorem 3.1. *Let assumption SA hold and let $-1 < \rho < 1$. Then the FE Quasi ML estimator for the conditional AR(1) model, (2.1), is consistent if and only if assumption RFE1 holds.*

PROOF

See appendix B. Consistency of the “pure” FEMLE estimator follows as a special case.¹⁵

¹³In section 2 we have argued that the assumption that the $v_{i,1}$ are essentially random effects is necessary for the consistency of any fixed effects estimator for ρ in the conditional AR(1) panel model. See lemmas 2.3 and 2.4 and the text following lemma 2.4.

¹⁴We assume that the cross-sectional average of the $E(v_{i,1})$, $i = 1, \dots, N$, is zero. If this average is non-zero, our FE(Q)MLE is still consistent but an asymptotically more efficient fixed effects MLE can be obtained after adding an intercept to the model.

¹⁵The “pure” FEMLE requires that the $v_{i,1}$ are i.i.d. and normally distributed. However, this requirement is not in the spirit of the fixed effects approach. Therefore, strictly speaking, the estimator that we refer to in the text as a FE Quasi MLE could be considered as the true Fixed Effects MLE, provided that the disturbances $\varepsilon_{i,t}$ are i.i.d. and normal.

The FEQML estimator is “Quasi” in a double sense: it allows for heterogeneity in the moments of the $v_{i,1}$ ’s and for non-normality of the $\varepsilon_{i,t}$ ’s and the $v_{i,1}$ ’s (under certain conditions, e.g. RFE1).

The FEML estimator for ρ based on (3.2) is equal to the QMLE for ρ which is based on the density of $y_i - y_{i,-1}$, and is (under normality and homoskedasticity of the u_i ’s) asymptotically equivalent to the OFECGMM estimator from section 2.¹⁶ This can be seen by taking differences between the equations for adjacent observations in (3.2) and keeping the equation for the first ‘observation’ $\Delta y_{i,2}$. This nonsingular constant transformation of system (3.2) yields

$$\begin{aligned}\Delta y_{i,2} &= u_{i,2} = v_i + \varepsilon_{i,2} = \varepsilon_{i,2} - (1 - \rho)v_{i,1} \\ \Delta y_{i,t} &= \rho \Delta y_{i,t-1} + \Delta \varepsilon_{i,t} \quad t = 3, \dots, T.\end{aligned}\tag{3.3}$$

This system gives rise to the moment conditions in (2.6) which are exploited by the OFECGMM estimator. It follows that the FEML estimator for ρ is the most efficient Fixed Effects estimator under normality and (cross-sectional) homoskedasticity. However, in general the FEQML estimator for ρ is less efficient than the OFECGMM estimator. Moreover, since the ORECGMM estimator is more efficient than the OFECGMM estimator (see Theorem 2.13), the RE(Q)MLE is in general more efficient than the FE(Q)MLE. This is due to the fact that the RE estimators assume and exploit that $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 < \infty$ and $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mu_i^2 < \infty$, unlike the FE estimators.

We notice that Hsiao et al. (1998) have also proposed the “pure” FEML estimator for the conditional model. However, the assumptions that they have made in order to derive the FEML estimator and to justify its use are unnecessarily strong. In particular, they have made strong assumptions about the presample differences of the $y_{i,t}$ ’s. Moreover, they have assumed that all the individual processes started in the same period in order to achieve homoskedasticity of the $\Delta y_{i,2}$, $i = 1, \dots, N$.¹⁷ ¹⁸ In contrast, we assume that the deviations of the initial observations from the individual effects, i.e. the v_i ’s, are i.i.d. and normally distributed. More importantly, we have noted that the FEML estimator can also be viewed as a Fixed Effects *Quasi* ML estimator, which is valid under much weaker conditions than the pure FEML estimator. In particular, the FEQMLE

¹⁶If the model contains an intercept, then the FE GMM estimator that is asymptotically equivalent to the FEMLE also exploits the information in the first moments of the data.

¹⁷Hsiao et al. condition on the first presample realizations of the $\{y_{i,t}\}$ processes instead of the initial observations, which is what the conditional model does. Furthermore, in their footnote 3, they implicitly assume that $Var(\Delta y_{i,-m+1}) = 0 \forall i \in \mathcal{I}$, where $y_{i,-m}$ is the first realization of the $\{y_{i,t}\}$ process.

¹⁸On the other hand, Hsiao et al. do not mention that the individual effects have to satisfy a condition for the FEMLE to be consistent.

for ρ and σ^2 is still consistent under heteroskedasticity of the $v_{i,1}$'s (or $\Delta y_{i,2}$'s) (cf assumptions RFE1' and RFE1), when the means of the $v_{i,1}$'s ($\Delta y_{i,2}$'s) are different across individuals (cf assumption RFE1), and also under cross-sectional dependence of the $v_{i,1}$'s (cf assumption RFE1).¹⁹ Lemmas 2.3 and 2.4 allow us to assess the plausibility of assumption RFE1 directly.

The FEQML estimator may be an attractive alternative to the OFECGMM estimator for a number of reasons. First, these estimators are asymptotically equivalent under normality and homoskedasticity. Second, when T is large the number of moment conditions exploited by the OFECGMM estimator will be large, which may result in poor finite sample properties of this GMM estimator. Third, to implement the OFECGMM estimator, one has to estimate a possibly large-dimensional optimal weighting matrix. This may also result in poor finite sample properties of this GMM estimator. The covariance matrix in the (quasi) likelihood function on the other hand is parsimoniously parametrized.

When the u_i 's are not i.i.d. and normally distributed, the asymptotic distribution of the FEQMLE will in general be different from the asymptotic distribution of the FEMLE. If the $u_{i,1}$'s are i.h.d., one may obtain the asymptotic distribution of the FEQMLE by applying the Lindeberg-Feller Central Limit Theorem. The asymptotic variance of the FEQMLE can then be computed by using the sandwich formula $H^{-1}GH^{-1}$, where H is the asymptotic Hessian and G is the limit of the outerproduct of the gradient of the log-likelihood function (cf MaCurdy (1981b)).

The FEML approach (or the OFECGMM approach for that matter) can easily be extended to higher order conditional AR panel models. Consider the AR(2) version: $y_i = \rho_1 y_{i,-1} + \rho_2 y_{i,-2} + (1 - \rho_1 - \rho_2)\mu_i \iota + \varepsilon_i$. In this case we assume that $v_{i,2} = y_{i,2} - \mu_i$ and $v_{i,1} = y_{i,1} - \mu_i$ are essentially random. Then the first equation of the transformed model reads $\Delta y_{i,3} = v_i + \varepsilon_{i,3} = \varepsilon_{i,3} + \rho_2 v_{i,1} - (1 - \rho_1)v_{i,2}$.

Finally, we present a general framework for obtaining consistent (Q)ML and GMM estimators

$$y_i - y_{i,1}\iota = \rho(y_{i,-1} - y_{i,1}\iota) + \tilde{\pi}y_{i,1}\iota + u_i, \quad (3.4)$$

where $u_i \equiv v_i\iota + \varepsilon_i$.

If $\tilde{\pi} = 0$ we obtain the FE model, whereas if $\tilde{\pi} \neq 0$ we obtain the (correlated) RE model, where both y_1 and μ_i are essentially random. The RE model implies the moment conditions in (2.2) which are exploited by the ORECGMM estimator. If $\tilde{\pi} = \rho - 1$, we obtain the uncorrelated random effects model. This model yields one extra moment condition, $E[(y_{i,2} - \rho y_{i,1})y_{i,1}] = 0$, relative to those in (2.2).

¹⁹So in contrast to what Hsiao et al. suggest, within a Quasi ML framework one can in fact deal with $E(\Delta y_{i,2}) = b_i$, where the b_i are different across individuals, as long as $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,2})^2 < \infty$. The latter requires that $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N b_i^2 < \infty$.

3.3. A comparison of the FEML estimator, the Conditional ML estimator, and the Modified ML estimator.

In appendix B we show that if the u_i 's are i.i.d. normal, then the FEML estimator for ρ in the conditional AR(1) panel model is more efficient than the Modified ML estimator (or Bayesian estimator) that was proposed by Lancaster (1997). Thus, although both the Modified ML approach of Neyman and Scott (1948) and the Bayesian method suggested by Lancaster (1997) and Lindley and El-Sayyed (1968) may help to overcome the inconsistency problem of direct ML estimators, they do not necessarily solve the inefficiency problem of ML which is also related to the presence of incidental parameters in the model. On the other hand, if the u_i 's are not i.i.d. normal, in particular, if the $v_{i,1}$'s are not i.i.d. normal (but still satisfy condition RFE1), then the FEQMLE for ρ is no longer asymptotically efficient and the Modified MLE for ρ may asymptotically be more precise than the FEQMLE for ρ . In this case only the OFECGMM estimator for ρ is asymptotically efficient. Moreover, the limiting variances of the FEQML estimator and the OFECGMM estimator will in general depend on the fourth order moment(s) of the $v_{i,1}$'s, whereas the limiting variance of the Modified MLE only depends on the second moment(s) of the $v_{i,1}$'s, whatever the distribution(s) of the $v_{i,1}$'s is (are).

In the case of the conditional AR(1) panel model, the Conditional ML method fares worse than the Modified ML method: it does not solve the inconsistency problem of ML. In this case, the Conditional ML method makes use of a sufficient statistic that depends on ρ . Indeed, Andersen (1970) has only proven consistency of Conditional ML estimators that are based on sufficient statistics which are true statistics.

4. Inefficiency of fixed effects estimators for covariance parameters.

It is well known that in static regression models a fixed effects approach and correlated random effects approach yield the same estimators for the mean equation parameters. In this section we will show that this does not hold true for the covariance parameters. When the individual effects are essentially random (effects), only a correlated random effects approach yields efficient estimators for the covariance parameters, whereas generally only the fixed effects approach yields consistent estimators for the covariance parameters when the individual effects are not (essentially) random.

The static panel model with individual effects is given by

$$\begin{aligned} y_i &= X_i\beta + \mu_i\iota + \varepsilon_i, \\ \varepsilon_i &\sim i.i.d. (0, V), \quad i = 1, \dots, N, \end{aligned}$$

where $\dim(\varepsilon_i) = T$. When discussing ML estimators, we will assume normality.

4.1. Estimation under fixed effects.

Kiefer (1980) has examined ML and GLS estimation of the static panel model with fixed effects and arbitrary covariance matrix V . The standard fixed effects MLE is always inconsistent for the covariance matrix, V , and in general also for the slope parameters, β , unless $V \propto I$. In the latter case the MLE for β is equal to the LSDV estimator for β and does not depend on the inconsistent estimator for the variance of the $\varepsilon_{i,t}$, σ^2 .

The Feasible GLS estimator that Kiefer proposes for β when V is unknown, is calculated in two steps: first β is consistently estimated by the LSDV estimator and next a Within FGLS estimator of β is computed using a consistent estimator for QVQ : $N^{-1} \sum_{i=1}^N Q(y_i - X_i \hat{\beta}_{LSDV})(y_i - X_i \hat{\beta}_{LSDV})' Q$, where $Q = I_T - \frac{1}{T} \iota \iota'$.

We make two remarks. First, as shown in appendix C, any within transformation matrix K , such that $K\iota = 0$ and $\text{rank}(K) = T - 1$, will yield the same FGLS estimator. Secondly, it does not matter whether one first applies a GLS transformation related to V and next a transformation to remove the terms involving the μ_i 's or applies these transformations in reverse order: the fact that V is not consistently estimated by $\hat{V} = N^{-1} \sum_{i=1}^N (y_i - X_i \hat{\beta}_{LSDV} - \hat{\mu}_{i,OLS\iota})(y_i - X_i \hat{\beta}_{LSDV} - \hat{\mu}_{i,OLS\iota})'$ does not affect the consistency of the resulting GLS estimator. To see this, let $P'P = \hat{V}^{-1}$, and let $M_p = I_T - P\iota(\iota'P'P\iota)^{-1}\iota'P'$. Then $M_p M_p' M_p = M_p$ and $P'M_p P = \hat{V}^{-1} - \frac{\hat{V}^{-1}\iota\iota'\hat{V}^{-1}}{\iota'\hat{V}^{-1}\iota} = D'(D\hat{V}D')^{-1}D = Q(Q\hat{V}Q)^{-1}Q$, where the structure of the first-difference matrix D was described in section 2. The equality in the middle was established by Lancaster and Lindenhovius (1996).

When V is parameterized, say $V = V(\theta)$, and the μ_i are fixed effects, a “within” ML approach will yield consistent estimators of both β and θ , provided that θ is identified. All the “within” ML estimators for β and θ which make use of a “within” transformation matrix K such that $K\iota = 0$ and $\text{rank}(K) = T - 1$, are in fact equal to the Conditional MLE (see Kruiniger (1998) for the relationship between transformed MLE's and CMLE's). Now, the CMLE will be consistent under any sequence of fixed effects since the conditional likelihood function does not depend on them.

4.2. Estimation under correlated random effects.

We will now discuss the correlated random effects approach to estimating β and θ . In the sequel we will assume that the μ_i 's are essentially random effects, i.e. $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mu_i^2 = \sigma_\mu^2 < \infty$, and $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i' X_i < \infty$. In addition, we will assume that all the regressors are (essentially) correlated with the indi-

vidual effects, i.e. $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (x_{k,i,t} - \bar{x}_{k,t}) \mu_i \neq 0 \forall k \in \{1, \dots, \dim(\beta)\}$, and $\forall t \in \{1, \dots, T\}$. The correlated random effects approach amounts to decomposing the individual effects into a part that captures the correlation with the regressors, and a remainder, say ν_i , that is uncorrelated with the regressors: $\mu_i = \tilde{x}_i' \xi + \nu_i$, where the elements of \tilde{x}_i are linear functions of the regressors in $vec(X_i')$, and $\xi = \text{plim}_{N \rightarrow \infty} (\sum_{i=1}^N \tilde{x}_i \tilde{x}_i')^{-1} \sum_{i=1}^N \tilde{x}_i \mu_i$, so that $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \tilde{x}_i \nu_i = 0$. In addition, we have that $(p)\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \nu_i^2 = \sigma_v^2 < \infty$. If we take the \tilde{x}_i' equal to $\iota'(\sigma_v^2 \iota \iota' + V(\theta))^{-1} X_i$, then the Random Effects ML estimator of β and θ based on

$$\begin{aligned} y_i &= c\iota + X_i \beta + \iota \tilde{x}_i' \xi + \nu_i \iota + \varepsilon_i, \\ \nu_i \iota + \varepsilon_i &\sim i.i.d. (0, \sigma_v^2 \iota \iota' + V(\theta)), \quad i = 1, \dots, N, \end{aligned}$$

can be shown to be consistent and asymptotically efficient. Consistency follows from $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i' (\sigma_v^2 \iota \iota' + V)^{-1} \nu_i = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \tilde{x}_i \nu_i = 0$, and $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \tilde{x}_i \iota' (\sigma_v^2 \iota \iota' + V)^{-1} \nu_i = [\iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota] \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \tilde{x}_i \nu_i = 0$.

Let $u_i \equiv y_i - c\iota - X_i \beta - \iota \tilde{x}_i' \xi = \nu_i \iota + \varepsilon_i$. Noting that $E[Du_i u_i' (\sigma_v^2 \iota \iota' + V)^{-1} \iota] = 0$, we can factorize the joint normal density of the u_i 's as $f(u_i) = g(Du_i) \times h([\iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota' (\sigma_v^2 \iota \iota' + V)^{-1} u_i)$, where D is the first difference matrix defined in section 2.²⁰ The marginal density of $[\iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota' (\sigma_v^2 \iota \iota' + V)^{-1} u_i$ depends on β , ξ , σ_v^2 , and θ , whereas $g(Du_i)$ only involves β and θ . We note that $h([\iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota' (\sigma_v^2 \iota \iota' + V)^{-1} u_i)$ does not contain any information on β , because β and ξ correspond essentially to the same regressors, i.e. $[\iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota' (\sigma_v^2 \iota \iota' + V)^{-1} X_i$ and $\iota' (\sigma_v^2 \iota \iota' + V)^{-1} X_i$. We can reparametrize the model so that $[\iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota' (\sigma_v^2 \iota \iota' + V)^{-1} u_i = [\iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota' (\sigma_v^2 \iota \iota' + V)^{-1} (y_i - X_i \tilde{\xi})$ with $\tilde{\xi} = \beta + \iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota \xi$. Therefore, $\hat{\beta}_{ML, RE}$ can be obtained by maximizing $\prod_{i=1}^N g(Du_i)$ alone given $V(\hat{\theta}_{ML, RE})$. Thus $\hat{\beta}_{ML, RE}$ is a First Difference FGLS estimator which makes use of $DV(\hat{\theta}_{ML, RE})D'$. It follows that the correlated random effects MLE for β , the fixed effects MLE for β , and the Within FGLS estimator for β which uses a consistent estimate of QVQ' , are asymptotically equivalent. However, $h([\iota' (\sigma_v^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota' (\sigma_v^2 \iota \iota' + V)^{-1} u_i)$ does contain additional information about θ and therefore the random effects MLE of θ is asymptotically more efficient than the fixed effects MLE of θ which only maximizes $\prod_{i=1}^N g(Du_i)$:

Theorem 4.1. *When T is fixed and the individual effects are essentially random but correlated with all the regressors, the correlated random effects MLE of θ , $\hat{\theta}_{ML, RE}$, is asymptotically more efficient than the fixed effects MLE of θ , $\hat{\theta}_{ML, FE}$.*

²⁰It can easily be shown that $|DW D'| = \iota' W^{-1} \iota |W|$. See also Theorem 2 in Lancaster and Lindenhovius (1996).

PROOF

Straightforward: see the discussion above.

When $T \rightarrow \infty$, the contribution of the $h([\iota'(\sigma_\nu^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota'(\sigma_\nu^2 \iota \iota' + V)^{-1} u_i)$ to the joint likelihood function $\prod_{i=1}^N f(u_i)$ becomes negligible and the fixed effects MLE of θ is asymptotically equivalent to the random effects MLE of θ .

We notice that ML applied to a model with $\mu_i = [vec(X_i')]'\pi + w_i$ (cf Mundlak (1978), and Chamberlain (1980, 1984)), i.e. with $\tilde{x}_i = vec(X_i')$, will lead to a First Difference FGLS type estimator for β , which is asymptotically equivalent to $\hat{\beta}_{FE}$ (and $\hat{\beta}_{RE}$) as well. This approach also yields an estimator for θ , $\hat{\theta}_{ML,\pi}$, that is more efficient than the fixed effects estimator for θ . This can again be seen from the factorisation of $f(u_i)$ as $g(Du_i)h([\iota'(\sigma_\nu^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota'(\sigma_\nu^2 \iota \iota' + V)^{-1} u_i)$. In this case we can reparametrize the model so that $[\iota'(\sigma_w^2 \iota \iota' + V)^{-1} \iota]^{-1} \times \iota'(\sigma_w^2 \iota \iota' + V)^{-1} u_i = [\iota'(\sigma_w^2 \iota \iota' + V)^{-1} \iota]^{-1} \iota'(\sigma_w^2 \iota \iota' + V)^{-1} (y_i - \iota[vec(X_i')]'\tilde{\pi})$ with $\tilde{\pi} = [\iota'(\sigma_w^2 \iota \iota' + V)^{-1} \iota]^{-1} vec[\beta \iota'(\sigma_w^2 \iota \iota' + V)^{-1}] + \pi$. Since using the specification $\mu_i = \iota'(\sigma_w^2 \iota \iota' + V)^{-1} X_i \xi + \nu_i$ amounts to imposing restrictions on π which involve θ , i.e. $\pi = vec(\xi \iota'(\sigma_w^2 \iota \iota' + V(\theta))^{-1})$, $\hat{\theta}_{RE}$ is more efficient than $\hat{\theta}_{ML,\pi}$.

When π in $\mu_i = [vec(X_i')]'\pi + w_i$ is not restricted to depend on θ in some way, the information matrix is block-diagonal, i.e. the non-diagonal elements in the information matrix corresponding to the slope parameters β and π on the one hand and the covariance parameters θ on the other hand are zero. It follows that it does not matter for the efficiency of $\hat{\theta}_{ML,\pi}$ whether the elements of π that correspond to those regressors in X_i which are not correlated with μ_i , are restricted to equal zero or not. However, imposing zero restrictions on π (or ξ) does lead to more efficient ML estimators for β , provided that these restrictions are valid of course.

4.3. Generalized method of moments estimators for covariance parameters.

There also exist fixed and random effects versions of the Generalized Method of Moments for estimating θ . A Fixed Effects GMM estimator for θ minimizes $\{\sum_i vech[D(y_i - X_i \hat{\beta}_{FE})(y_i - X_i \hat{\beta}_{FE})' D' - D V D']\} W \{\sum_i vech[D(y_i - X_i \hat{\beta}_{FE}) \times (y_i - X_i \hat{\beta}_{FE})' D' - D V D']\}$ for some weighting matrix W . A Random Effects GMM estimator for θ would minimize $\{\sum_i vech[(y_i - X_i \hat{\beta}_{FE})(y_i - X_i \hat{\beta}_{FE})' - \sigma_\mu^2 \iota \iota' - V]\} W \{\sum_i vech[(y_i - X_i \hat{\beta}_{FE})(y_i - X_i \hat{\beta}_{FE})' - \sigma_\mu^2 \iota \iota' - V]\}$ for some weighting matrix W . Consistency of $\hat{\theta}_{GMM,RE}$ requires that the μ_i 's are essentially random effects. This Random Effects estimation procedure is equivalent to a GMM estimator that exploits the moment conditions in $E\{vech[D(y_i - X_i \hat{\beta}_{FE})(y_i - X_i \hat{\beta}_{FE})' D' - D V D']\} = 0$, and $E\{[D(y_i - X_i \hat{\beta}_{FE})(y_i - X_i \hat{\beta}_{FE})' e_1 - D V e_1]\} = 0$, where e_1 is the

first column of the identity matrix. Since the REGMM estimator for θ exploits more moment conditions than the FEGMM estimator, which are not redundant, the former is asymptotically more efficient than the latter.

The Optimal FEGMM estimator for θ is asymptotically equivalent to the FEMLE for θ under normality. The Optimal REGMM estimator for θ in turn is asymptotically equivalent to $\hat{\theta}_{ML,\pi}$, but in general asymptotically less efficient than $\hat{\theta}_{ML,RE}$, except when the regressors in X_i are not correlated with the μ'_i s.

5. Conclusions

In this paper we have studied estimation of various versions of the AR(1) panel model with initial conditions. We have derived a necessary condition for consistency of any estimator for the fixed effects version of this model. This condition involves the deviations of the initial conditions from the individual (mean) effects. The condition allows for cross-sectional correlation and heterogeneity of the data. By imposing this condition on Chamberlain's correlated effects formulation of the model, we obtained the Fixed Effects (Quasi) Maximum Likelihood estimator.

A related but perhaps not very surprising result is that any consistent fixed effects estimator for the AR(1) panel model with initial conditions involves only first differences of the data. We have seen that if the individual effects are trending, a GMM estimator for ρ which depends on levels of the data, e.g. the Arellano-Bond GMM estimator, may have poor properties.

We have established the connection under normality and (cross-sectional) homoskedasticity between the FE(Q)MLE for ρ and the Optimal Fixed Effects Conditional GMM estimator for ρ , which exploits all the second order moment conditions that only involve differences of the data. It follows that the FE(Q)MLE for ρ is asymptotically efficient under these conditions. Similarly, we have proven asymptotic equivalence of the REMLE for ρ and the ORECGMM estimator for ρ due to Ahn and Schmidt (1995). The latter is shown to be asymptotically more efficient than the OFECGMM estimator for ρ .

The FEMLE for ρ was also compared with alternative likelihood approaches. The Bayesian estimator using Jeffreys prior, or equivalently, the Modified MLE, is asymptotically inefficient. Moreover, both the FEMLE for ρ and the Modified MLE for ρ are very imprecise near the unit root. The Conditional ML approach yields an inconsistent estimator for ρ .

Finally, it is shown in this paper that the Fixed Effects approach but also Chamberlain's π -approach lead to less precise estimators for covariance parameters in static regression models than the Correlated Random Effects approach.

A. Proofs for section 2.

PROOF OF LEMMA 2.2

Let $\Delta_1 > 0$ be a generic constant. Recall that $v_{i,1} = y_{i,1} - \mu_i$.

The model reads as $(y_{i,t} - \mu_i) = \rho(y_{i,t-1} - \mu_i) + \varepsilon_{i,t}$, $\forall t \in \mathcal{T}$.

Therefore, $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,t} - \mu_i)^2 = \rho^2(p) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,t-1} - \mu_i)^2 + 2\rho \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N [(y_{i,t-1} - \mu_i)\varepsilon_{i,t}] + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{i,t}^2$, $\forall t \in \mathcal{T}$.

Now, $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,1} - \mu_i)\varepsilon_{i,t} = 0$, $\forall t \in \mathcal{T}$, because of RFE1 (or

because of SA and RFE1' which imply $E((y_{i,1} - \mu_i)\varepsilon_{i,t}) = 0$, and

$E(|(y_{i,1} - \mu_i)\varepsilon_{i,t}|^{1+\delta}) < \Delta_1$). It follows that $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,2} - \mu_i)\varepsilon_{i,t} = \rho \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,1} - \mu_i)\varepsilon_{i,t} + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{i,2}\varepsilon_{i,t} = 0$, $\forall t \in \{3, \dots, T\}$.

By induction over t , we then obtain $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,t-1} - \mu_i)\varepsilon_{i,t} = 0$, and

$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,t} - \mu_i)^2 < \infty$, $\forall t \in \mathcal{T}$. \square

PROOF OF LEMMA 2.3

We have $\Delta y_{i,t} = y_{i,t} - \mu_i + \mu_i - y_{i,t-1} = (\rho - 1)(y_{i,t-1} - \mu_i) + \varepsilon_{i,t}$, $\forall t \in \mathcal{T}$.

Therefore, $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,t})^2 = (\rho - 1)^2(p) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,t-1} - \mu_i)^2 + 2(\rho - 1) \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,t-1} - \mu_i)\varepsilon_{i,t} + \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{i,t}^2$, $\forall t \in \mathcal{T}$.

Using results from the proof of lemma 2.2, we obtain by induction over t that

$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,t-1} - \mu_i)\varepsilon_{i,t} = 0$, and $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,t})^2 < \infty$, $\forall t \in \mathcal{T}$. \square

PROOF OF LEMMA 2.4

We have $\Delta y_{i,2} = y_{i,2} - \mu_i + \mu_i - y_{i,1} = (\rho - 1)(y_{i,1} - \mu_i) + \varepsilon_{i,2}$.

Therefore, $(\rho - 1)^2(p) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta y_{i,2})^2 - 2(\rho - 1) \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,1} - \mu_i)\varepsilon_{i,2} - \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{i,2}^2 < \infty$. \square

PROOF OF THEOREM 2.5

Recall that $v_{i,1} = y_{i,1} - \mu_i$. W.l.o.g. we have assumed that $T = 3$. Then Arellano and Bond's GMM estimator exploits only one moment condition, $E[y_{i,1}\Delta\varepsilon_{i,3}] = 0$, where $\Delta\varepsilon_{i,3} = \Delta y_{i,3} - \rho\Delta y_{i,2}$:

$$\hat{\rho}_{AB} = [\sum y_{i,1}\Delta y_{i,2}]^{-1}[\sum y_{i,1}\Delta y_{i,3}] = \rho + [\sum y_{i,1}\Delta y_{i,2}]^{-1}[\sum y_{i,1}\Delta\varepsilon_{i,3}].$$

We can write the numerator as $\frac{1}{N}\sum y_{i,1}\Delta\varepsilon_{i,3} = \frac{1}{N}\sum \mu_i\Delta\varepsilon_{i,3} + \frac{1}{N}\sum v_{i,1}\Delta\varepsilon_{i,3}$, and the denominator as $\frac{1}{N}\sum y_{i,1}\Delta y_{i,2} = \frac{1}{N}\sum y_{i,1}[(\rho - 1)(y_{i,1} - \mu_i) + \varepsilon_{i,2}] = \frac{1}{N}\sum[(\rho - 1)v_{i,1}^2 + v_{i,1}\varepsilon_{i,2}] + \frac{1}{N}\sum \mu_i[(\rho - 1)v_{i,1} + \varepsilon_{i,2}]$.

Let Δ_1 be a generic positive constant.

By assumption SA we have $E(\mu_i\Delta\varepsilon_{i,3}) = 0$ and $E(\varepsilon_{i,t}^2) < \Delta_1 < \infty$, $\forall i \in \mathcal{I}$.

Hence $Var(N^{-1-\lambda}\sum \mu_i\Delta\varepsilon_{i,3}) = N^{-2-2\lambda}\sum E[(\mu_i\Delta\varepsilon_{i,3})^2] < N^{-2-2\lambda}\sum \mu_i^2\Delta_1$.

Therefore, $\lim_{N \rightarrow \infty} Var(N^{-1-\lambda}\sum \mu_i\Delta\varepsilon_{i,3}) = 0$ if $N^{-2-2\lambda}\sum \mu_i^2 = o(1)$.

It follows that $N^{-1-\lambda}\sum \mu_i\Delta\varepsilon_{i,3} \xrightarrow{q.m.} 0$ if $N^{-2-2\lambda}\sum \mu_i^2 = o(1)$.

Now, $N^{-1-\lambda}\sum \mu_i\Delta\varepsilon_{i,3} \xrightarrow{q.m.} 0$ implies that $\text{plim}_{N \rightarrow \infty} N^{-1-\lambda}\sum \mu_i\Delta\varepsilon_{i,3} = 0$.

Therefore, $\text{plim}_{N \rightarrow \infty} N^{-1-\lambda}\sum \mu_i\Delta\varepsilon_{i,3} = 0$ if $N^{-2-2\lambda}\sum \mu_i^2 = o(1)$.

Likewise, $\text{plim}_{N \rightarrow \infty} N^{-1-\lambda}\sum \mu_i\varepsilon_{i,2} = 0$ if $N^{-2-2\lambda}\sum \mu_i^2 = o(1)$.

Since $E(v_{i,1}\varepsilon_{i,t}) = 0$, we obtain that $\text{plim}_{N \rightarrow \infty} \frac{1}{N}\sum[(\rho - 1)v_{i,1}^2 + v_{i,1}\varepsilon_{i,2}] = (\rho - 1)\sigma_v^2$. In addition, we find that $\text{plim}_{N \rightarrow \infty} \frac{1}{N}\sum v_{i,1}\Delta\varepsilon_{i,3} = 0$.

Finally, the assumptions imply that $\text{plim}_{N \rightarrow \infty} N^{-1-\lambda}\sum(\rho - 1)\mu_i v_{i,1} = (\rho - 1)\sigma_{\mu,v}$.

Thus the probability limit of the denominator of $\hat{\rho}_{AB}$ will be different from zero.

□

PROOF OF THEOREM 2.7

Recall that $\sigma_v^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}^2 < \infty$.

In addition, $\lim_{N \rightarrow \infty} N^{-1-\kappa} \sum_{i=1}^N E(\mu_i \Delta \varepsilon_{i,3})^2 = \zeta_1$, with $0 < \zeta_1 < \infty$,

$\lim_{N \rightarrow \infty} N^{-1-\kappa} \sum_{i=1}^N E(\mu_i \varepsilon_{i,2})^2 = \zeta_2$, with $0 < \zeta_2 < \infty$, and

$\lim_{N \rightarrow \infty} N^{-1-\lambda} \sum_{i=1}^N \mu_i v_{i,1} = \sigma_{\mu,v}$, $0 \leq \lambda \leq 0.5\kappa$.

Then the following results can easily be verified:

If $\kappa > 0$, $\text{plim}_{N \rightarrow \infty} N^{-0.5(1+\kappa)} \sum v_{i,1} \Delta \varepsilon_{i,3} = 0$.

If $\kappa > 1$, $\text{plim}_{N \rightarrow \infty} N^{-\kappa} \sum [(\rho - 1)v_{i,1}^2 + v_{i,1} \varepsilon_{i,2}] = 0$.

If $0 \leq 2\lambda \leq \kappa < 1 + 2\lambda$, $\text{plim}_{N \rightarrow \infty} N^{-1-\lambda} \sum \mu_i [(\rho - 1)v_{i,1} + \varepsilon_{i,2}] = (\rho - 1)\sigma_{\mu,v}$.

If $0 < \zeta_0 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E(y_{i,1} \Delta \varepsilon_{i,3})^2 < \infty$, and $\{y_{i,1} \Delta \varepsilon_{i,3}\}$ satisfies

the Lindeberg condition, then $N^{-0.5} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3} \xrightarrow{d} N(0, \zeta_0)$.

If $0 \leq \kappa < 1$ and $\lambda = 0$, then $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N y_{i,1} \Delta y_{i,2} =$

$\text{plim}_{N \rightarrow \infty} N^{-1} \sum [(\rho - 1)v_{i,1}^2 + v_{i,1} \varepsilon_{i,2}] + \text{plim}_{N \rightarrow \infty} N^{-1} \sum \mu_i [(\rho - 1)v_{i,1} + \varepsilon_{i,2}] =$
 $(\rho - 1)\sigma_v^2 + (\rho - 1)\sigma_{\mu,v}$.

If $0 < 2\lambda \leq \kappa < 1 + 2\lambda$, then $\text{plim}_{N \rightarrow \infty} N^{-1-\lambda} \sum_{i=1}^N y_{i,1} \Delta y_{i,2} = (\rho - 1)\sigma_{\mu,v}$.

So if $\kappa = \lambda = 0$, $0 < \zeta_0 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E(y_{i,1} \Delta \varepsilon_{i,3})^2 < \infty$, and $\{y_{i,1} \Delta \varepsilon_{i,3}\}$ satisfies the Lindeberg condition, then

$N^{0.5} (\hat{\rho}_{AB} - \rho) \xrightarrow{d} N(0, [(\rho - 1)\sigma_v^2 + (\rho - 1)\sigma_{\mu,v}]^{-2} \zeta_0)$.

If $\kappa > 0$ and $\{\mu_i \Delta \varepsilon_{i,3}\}$ satisfies the Lindeberg condition, then

$N^{-0.5(1+\kappa)} \sum y_{i,1} \Delta \varepsilon_{i,3} = N^{-0.5(1+\kappa)} \sum v_{i,1} \Delta \varepsilon_{i,3} + N^{-0.5(1+\kappa)} \sum \mu_i \Delta \varepsilon_{i,3} \xrightarrow{d} N(0, \zeta_1)$;

if $0 < \kappa < 1$ and $\lambda = 0$, $N^{0.5(1-\kappa)} (\hat{\rho}_{AB} - \rho) \xrightarrow{d} N(0, [(\rho - 1)\sigma_v^2 + (\rho - 1)\sigma_{\mu,v}]^{-2} \zeta_1)$;

if $0 < 2\lambda \leq \kappa < 1 + 2\lambda$, $N^{0.5(1-\kappa+2\lambda)} (\hat{\rho}_{AB} - \rho) \xrightarrow{d} N(0, [(\rho - 1)\sigma_{\mu,v}]^{-2} \zeta_1)$.

Finally, if $\kappa \geq 1 + 2\lambda$ and $\{\mu_i \varepsilon_{i,2}\}$ satisfies the Lindeberg condition, then

if $\kappa = 1$ and $\lambda = 0$, $N^{-0.5(1+\kappa)} \sum y_{i,1} \Delta y_{i,2} \xrightarrow{d} (\rho - 1)\sigma_v^2 + (\rho - 1)\sigma_{\mu,v} + N(0, \zeta_2)$;

if $\kappa = 1 + 2\lambda$ and $\lambda > 0$, $N^{-0.5(1+\kappa)} \sum y_{i,1} \Delta y_{i,2} \xrightarrow{d} (\rho - 1)\sigma_{\mu,v} + N(0, \zeta_2)$;

if $\kappa > 1 + 2\lambda$, $N^{-0.5(1+\kappa)} \sum y_{i,1} \Delta y_{i,2} \xrightarrow{d} N(0, \zeta_2)$.

So if $\{\mu_i \Delta \varepsilon_{i,3}\}$ and $\{\mu_i \varepsilon_{i,2}\}$ satisfy the Lindeberg condition, then

if $\kappa = 1$ and $\lambda = 0$, $(\hat{\rho}_{AB} - \rho) \xrightarrow{d} \frac{N(0, \zeta_1)}{(\rho - 1)\sigma_v^2 + (\rho - 1)\sigma_{\mu,v} + N(0, \zeta_2)}$;

if $\kappa = 1 + 2\lambda$ and $\lambda > 0$, $(\hat{\rho}_{AB} - \rho) \xrightarrow{d} \frac{N(0, \zeta_1)}{(\rho - 1)\sigma_{\mu,v} + N(0, \zeta_2)}$;

if $\kappa > 1 + 2\lambda$, $(\hat{\rho}_{AB} - \rho) \xrightarrow{d} \frac{N(0, \zeta_1)}{N(0, \zeta_2)}$. \square

PROOF OF THEOREM 2.9

The proof is similar to the proof of theorem 2.5. with one modification.

It is assumed that λ_1 is the smallest positive number, $\lambda_1 \geq 0$, such that $\lim_{N \rightarrow \infty} N^{-1-\lambda_1} \sum_{i=1}^N E(\mu_i v_{i,1}) = c_{\mu,v}$ with $|c_{\mu,v}| < \infty$ and $c_{\mu,v} \neq 0$ if $\lambda_1 > 0$. Moreover, $\lim_{N \rightarrow \infty} N^{-1-\kappa_1} \sum_{i=1}^N \text{Var}(\mu_i[(\rho-1)v_{i,1} + \varepsilon_{i,2}]) = \zeta_3$, with $0 < \zeta_3 < \infty$, for some $\kappa_1 \geq 0$, and $\{\mu_i[(\rho-1)v_{i,1} + \varepsilon_{i,2}] - E(\mu_i(\rho-1)v_{i,1})\}$ satisfies the Lindeberg condition.

It follows from these assumptions that $\lim_{N \rightarrow \infty} N^{-1-\kappa_1} \sum_{i=1}^N \mu_i^2 \text{Var}(\varepsilon_{i,t}) < \infty$. Therefore, if $\kappa_1 < 1 + 2\lambda_1$, we obtain that $\text{plim}_{N \rightarrow \infty} N^{-1-\lambda_1} \sum \mu_i[(\rho-1)v_{i,1}] = (\rho-1)c_{\mu,v}$, and $\text{plim}_{N \rightarrow \infty} N^{-1-\lambda_1} \sum \mu_i \Delta \varepsilon_{i,3} = 0$. \square

PROOF OF THEOREM 2.10

Trivial. Notice that the first-differences of the data $\Delta y_{i,t}$, $t = 2, \dots, T$, only depend on $y_{i,1}$ and μ_i through $v_{i,1}$.

PROOF OF THEOREM 2.13

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

We will show that $E(y_{i,1} \Delta \varepsilon_{i,3}) = 0$ is such a moment condition.

First note that the FECGMM estimator does not exploit $E(y_{i,1} \Delta \varepsilon_{i,3}) = 0$, because this moment condition involves data in levels. On the other hand, the moment conditions exploited by the ORECGMM estimator can be formulated in such a way that $E(y_{i,1} \Delta \varepsilon_{i,3}) = 0$ is one of them (see (2.2)), that is, the ORECGMM estimator exploits $E(y_{i,1} \Delta \varepsilon_{i,3}) = 0$, either explicitly or implicitly.

Now, the moment conditions and the weighting matrix of the OFECGMM estimator only depend on $N^{-1} \sum_{i=1}^N y_{i,1}^2$ through $N^{-1} \sum_{i=1}^N v_{i,1}^2 = N^{-1} \sum_{i=1}^N (y_{i,1} - \mu_i)^2$ (i.e. σ_v^2). However, the optimal weighting matrix for the ORECGMM estimator — which exploits $E(y_{i,1} \Delta \varepsilon_{i,3}) = 0$ — also depends on $N^{-1} \sum_{i=1}^N y_{i,1}^2$ separately: $E[(N^{-1} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3})^2] = 2\sigma^2 N^{-1} \sum_{i=1}^N y_{i,1}^2$.

We conclude that $E(y_{i,1} \Delta \varepsilon_{i,3}) = 0$ is not redundant, and that $\text{Var}(\hat{\rho}_{ORECGMM}) < \text{Var}(\hat{\rho}_{OFECGMM})$. \square

PROOF OF THEOREM 2.14

The conditional ARX(1) panel model with fixed effects (and K exogenous regressors) can be stated as

$$\begin{aligned} y_i &= \rho y_{i,-1} + (1 - \rho)\mu_i + X_i\beta(1 - \rho) + \varepsilon_i, \quad -1 < \rho \leq 1 \\ \text{with } \varepsilon_i | X_i, y_{i,1}, \mu_i &\sim N(0, \sigma^2 I), \quad \forall i \in \mathcal{I}. \end{aligned} \quad (\text{A.1})$$

In addition, we will assume that assumption SA holds, and $\sigma^2 > 0$.

Let us define

$$\check{P} = \begin{pmatrix} 0 & . & . & 0 & 0 & 0 \\ 1 & 0 & & 0 & 0 & \\ \rho & 1 & 0 & & 0 & \\ . & \rho & 1 & 0 & & . \\ . & & \rho & 1 & 0 & . \\ \rho^{T-3} & . & . & \rho & 1 & 0 \end{pmatrix}, \quad \check{\varphi}_1 = \begin{pmatrix} 1 \\ \rho \\ \rho^2 \\ \vdots \\ \rho^{T-3} \\ \rho^{T-2} \end{pmatrix} \quad (\text{A.2})$$

and $\check{\beta} = \beta(1 - \rho)$, then

$$\begin{aligned} y_{i,-1} - \mu_i - \iota \bar{x}'_i \beta &= \check{\varphi}_1(y_{i,1} - \mu_i - \bar{x}'_i \beta) + \check{P}QX_i\check{\beta} + \check{P}\varepsilon_i \\ &\equiv \check{Z}_i + \check{P}\varepsilon_i, \end{aligned} \quad (\text{A.3})$$

where $\check{Z}_i = \check{\varphi}_1(y_{i,1} - \mu_i - \bar{x}'_i \beta) + \check{P}QX_i\check{\beta}$.

Let $v_{i,1} = y_{i,1} - \mu_i - \bar{x}'_i \beta$. We will add the following assumptions

$$\begin{aligned} QX_i \text{ are i.h.d. with } E(|\Delta x_{k,i,t}|^{2+\delta}) &< \Delta_1 < \infty, \\ \forall k \in \{1, \dots, K\}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \end{aligned} \quad (\text{A.4})$$

$$\text{and } \text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (\Delta x_{i,s})(\Delta x'_{i,t}) = \Sigma_{s,t}, \quad \forall s, t \in \mathcal{T},$$

and

$$\text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}^2 = \sigma_v^2 < \infty, \quad \text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1} \varepsilon_i = 0, \quad (\text{A.5})$$

$$\begin{aligned} \text{and } \text{p lim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1} \Delta x_{k,i,s} &= \Sigma_{k,t} \text{ with } |\Sigma_{k,t}| < \infty, \\ \forall k \in \{1, \dots, K\}, \forall t \in \mathcal{T}. \end{aligned}$$

Assumption (A.5) generalizes assumption RFE1.

The modified log-likelihood function for the conditional ARX(1) panel model (divided by N) is given by

$$\begin{aligned} \frac{1}{N} \log \tilde{p}(r, s^2, \check{b}) &= \xi(r) - \frac{T-2}{2} \log s^2 \\ &\quad - \frac{1}{2s^2} \frac{1}{N} \sum_{i=1}^N (y_i - ry_{i,-1} - X_i \check{b})' Q (y_i - ry_{i,-1} - X_i \check{b}), \\ \text{where } \xi(r) &= \frac{1}{T-1} \sum_{t=1}^{T-2} \frac{T-1-t}{t} r^t, \end{aligned} \quad (\text{A.6})$$

and the modified log-likelihood equations are given by the sums (over i) of

$$\begin{aligned} \Psi_{\rho,i}(r, s^2, \check{b}) &= \xi'(r) + \frac{1}{s^2} (y_i - ry_{i,-1} - X_i \check{b})' Q y_{i,-1} = 0, \\ \Psi_{\sigma^2,i}(r, s^2, \check{b}) &= -\frac{T-2}{2s^2} + \frac{1}{2s^4} (y_i - ry_{i,-1} - X_i \check{b})' Q (y_i - ry_{i,-1} - X_i \check{b}) = 0, \\ \Psi_{\beta,i}(r, s^2, \check{b}) &= \frac{1}{s^2} (y_i - ry_{i,-1} - X_i \check{b})' Q X_i = 0. \end{aligned} \quad (\text{A.7})$$

We can rewrite these functions using

$$y_i - ry_{i,-1} - X_i \check{b} = (\rho - r)y_{i,-1} + X_i(\check{\beta} - \check{b}) + \varepsilon_i. \quad (\text{A.8})$$

It follows from $Q\iota = 0$, (A.3) and (A.8) that

$$\begin{aligned} Q(y_i - ry_{i,-1} - X_i \check{b}) &= \\ Q((\rho - r)(y_{i,-1} - \mu_i \iota - \iota \bar{x}_i' \beta) + X_i(\check{\beta} - \check{b}) + \varepsilon_i) &= \\ Q((\rho - r)(\check{Z}_i + \check{P} \varepsilon_i) + X_i(\check{\beta} - \check{b}) + \varepsilon_i). \end{aligned} \quad (\text{A.9})$$

We will prove that *one* of the solutions of the modified log-likelihood equations is a consistent estimator for $(\rho, \sigma^2, \check{\beta})$ by verifying the conditions of theorem 4.1.2. of Amemiya (1986).

By using (A.9), it can easily be shown under the assumptions of the model (including (A.4) and (A.5)) that $\frac{1}{N} \log \tilde{p}(r, s^2, \check{b}) = \frac{1}{N} \sum_{i=1}^N \log \tilde{p}_i(r, s^2, \check{b})$ converges uniformly in probability to

$$\begin{aligned} \log \tilde{p}_A(r, s^2, \check{b}; \rho, \sigma^2, \check{\beta}) &= \xi(r) - \frac{T-2}{2} \log s^2 - \frac{\sigma^2}{2s^2} \text{tr}(Q) \\ &\quad - \frac{1}{2s^2} [(r - \rho)^2 [\Sigma_{zqz} + \sigma^2 \text{tr}(\check{P}' Q \check{P})] - 2(r - \rho) \sigma^2 \text{tr}(Q \check{P}(\rho)) \\ &\quad + (\check{b} - \check{\beta})' \Sigma_{xqx} (\check{b} - \check{\beta}) + 2(r - \rho) \Sigma_{zqx} (\check{b} - \check{\beta})], \end{aligned} \quad (\text{A.10})$$

where $\Sigma_{zqz} = p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \check{Z}'_i Q \check{Z}_i < \infty$, $\Sigma_{xqx} = p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X'_i Q X_i < \infty$, $\Sigma_{zqx} = p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \check{Z}'_i Q X_i$, and $(r, s^2, \check{b}) \in \Theta = (-1, 1) \times (0, \infty) \times \mathbb{R}^K$.

Notice that assumption (A.5) is a necessary condition for uniform convergence (in probability) of $\frac{1}{N} \log \tilde{p}(r, s^2, \check{b})$.

Next, we have to show that the probability limit of the modified log-likelihood function is maximized at the true values.

We will first show that

$$\begin{aligned} p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Psi_{\rho,i}(\rho, \sigma^2, \check{\beta}) &= 0, \\ p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Psi_{\sigma^2,i}(\rho, \sigma^2, \check{\beta}) &= 0, \\ p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Psi_{\beta,i}(\rho, \sigma^2, \check{\beta}) &= 0. \end{aligned} \quad (\text{A.11})$$

Evaluating the terms of the modified log-likelihood equations at the true values of the parameters, we obtain

$$\begin{aligned} \Psi_{\rho,i}(\rho, \sigma^2, \check{\beta}) &= \xi'(\rho) + \frac{\varepsilon'_i Q y_{i,-1}}{\sigma^2}, \\ \Psi_{\sigma^2,i}(\rho, \sigma^2, \check{\beta}) &= -\frac{T-2}{2\sigma^2} + \frac{\varepsilon'_i Q \varepsilon_i}{2\sigma^4}, \\ \Psi_{\beta,i}(\rho, \sigma^2, \check{\beta}) &= \frac{\varepsilon'_i Q X_i}{\sigma^2}. \end{aligned} \quad (\text{A.12})$$

Then

$$\begin{aligned} p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Psi_{\rho,i}(\rho, \sigma^2, \check{\beta}) &= \xi'(\rho) + tr(Q \check{P}(\rho)) \\ &= \xi'(\rho) - \xi'(\rho) = 0, \\ p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Psi_{\sigma^2,i}(\rho, \sigma^2, \check{\beta}) &= -\frac{T-2}{2\sigma^2} + \frac{\sigma^2}{2\sigma^4} tr(Q) = 0, \\ p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Psi_{\beta,i}(\rho, \sigma^2, \check{\beta}) &= \frac{1}{\sigma^2} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E(\varepsilon'_i Q X_i) = 0, \end{aligned} \quad (\text{A.13})$$

where we have used that $tr(Q \check{P}(\rho)) = tr(\check{P}(\rho)) - (T-1)^{-1} \iota' \check{P}(\rho) \iota = -\xi'(\rho)$.

Finally, we have to verify that the Hessian of the probability limit of the modified log-likelihood function is a negative definite symmetric matrix at the true values of the parameters. This Hessian is given by

$$MH = \begin{pmatrix} \xi''(\rho) - \frac{\Sigma_{zqz} + \sigma^2 \text{tr}(\check{P}'Q\check{P})}{\sigma^2} & \frac{\xi'(\rho)}{\sigma^2} & -\frac{\Sigma_{zqx}}{\sigma^2} \\ \frac{\xi'(\rho)}{\sigma^2} & -\frac{T-2}{2\sigma^4} & 0 \\ -\frac{\Sigma'_{zqx}}{\sigma^2} & 0 & -\frac{\Sigma_{xqx}}{\sigma^2} \end{pmatrix}.$$

We can rewrite MH as

$$\begin{pmatrix} \xi''(\rho) - \text{tr}(\check{P}(\rho)'Q\check{P}(\rho)) & \frac{\xi'(\rho)}{\sigma^2} & 0 \\ \frac{\xi'(\rho)}{\sigma^2} & -\frac{T-2}{2\sigma^4} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} -\frac{\Sigma_{zqz}}{\sigma^2} & 0 & -\frac{\Sigma_{zqx}}{\sigma^2} \\ 0 & 0 & 0 \\ -\frac{\Sigma'_{zqx}}{\sigma^2} & 0 & -\frac{\Sigma_{xqx}}{\sigma^2} \end{pmatrix}. \quad (\text{A.14})$$

It can be shown that $\begin{pmatrix} \xi''(\rho) - \text{tr}(\check{P}(\rho)'Q\check{P}(\rho)) & \frac{\xi'(\rho)}{\sigma^2} \\ \frac{\xi'(\rho)}{\sigma^2} & -\frac{T-2}{2\sigma^4} \end{pmatrix}$ is negative definite symmetric for $-1 < \rho < 1$. When $\rho = 1$, this matrix is singular: ignoring the powers of σ , its diagonal elements are equal to $-\frac{1}{2}(T-2)$ and its off-diagonal elements are equal to $\frac{1}{2}(T-2)$. Moreover, $\Sigma_{zqz} = 0$ and $\Sigma_{zqx} = 0$ when $\rho = 1$. Thus when $\rho = 1$ the parameters ρ and σ^2 are not uniquely identified.

The 2×2 matrix obtained after eliminating the middle row and column from the second matrix in (A.14) is also negative definite symmetric and hence MH is negative definite symmetric when $-1 < \rho < 1$. \square

The asymptotic distribution of the Modified MLE for the parameters in the ARX(1) panel model (A.1) is given by

$$\sqrt{N} \begin{pmatrix} \hat{\rho}_{MML} - \rho \\ \hat{\sigma}_{MML}^2 - \sigma^2 \\ \hat{\beta}_{MML} - \check{\beta} \end{pmatrix} \rightarrow N(0, (MH)^{-1} MIM (MH)^{-1}), \quad (\text{A.15})$$

where the Modified Information Matrix (MIM) is given by

$$MIM = \begin{pmatrix} \text{tr}(Q\check{P}Q\check{P}) + \frac{\Sigma_{zqz} + \sigma^2 \text{tr}(\check{P}'Q\check{P})}{\sigma^2} & -\frac{\xi'(\rho)}{\sigma^2} & \frac{\Sigma_{zqx}}{\sigma^2} \\ -\frac{\xi'(\rho)}{\sigma^2} & \frac{T-2}{2\sigma^4} & 0 \\ \frac{\Sigma'_{zqx}}{\sigma^2} & 0 & \frac{\Sigma_{xqx}}{\sigma^2} \end{pmatrix}. \quad (\text{A.16})$$

To derive MIM we have used the general result $E((\varepsilon'_i M_1 \varepsilon_i)(\varepsilon'_i M_2 \varepsilon_i)) = \sigma^4(\text{tr}(M_1)\text{tr}(M_2) + \text{tr}(M_1 M_2 + M'_1 M_2))$.

An explicit expression for $\text{tr}(Q\check{P}Q\check{P})$ can be found using *Maple*. It can easily be checked that $\text{tr}(Q\check{P}Q\check{P}) \neq -\xi''(\rho)$ and therefore $MH \neq -MIM$.

PROOF OF THEOREM 2.15

The conditional AR(1) panel model can be rewritten as

$$y_i - y_{i,1}\iota = \underline{\varphi}_1(\rho)[y_{i,1} - \mu_i] + \underline{P}(\rho)\varepsilon_i, \quad i = 1, \dots, N, \quad (\text{A.17})$$

where $\underline{\varphi}_1(\rho)$ and $\underline{P}(\rho)$ have been defined in the main text.

Recall that $\underline{V}(\rho) = \sigma^2 \underline{P}(\rho) \underline{P}(\rho)'$. The Conditional ML estimators for ρ and σ^2 in the conditional AR(1) panel model are defined as the maximizers of the conditional likelihood function $\prod_{i=1}^N L_c(y_i - y_{i,-1} | \hat{y}_{i,1} - \hat{\mu}_i, y_{i,1})$, where $\hat{y}_{i,1} - \hat{\mu}_i$ is a sufficient statistic for $y_{i,1} - \mu_i$. This sufficient statistic is given by

$$\hat{y}_{i,1} - \hat{\mu}_i = (\underline{\varphi}_1' \underline{V}^{-1} \underline{\varphi}_1)^{-1} \underline{\varphi}_1' \underline{V}^{-1} (y_i - y_{i,1}\iota) = \frac{1}{T-1} \iota' \underline{P}(\rho)^{-1} (y_i - y_{i,1}\iota). \quad (\text{A.18})$$

The following results are easily verified:

$$(y_i - y_{i,-1}) - E(y_i - y_{i,-1} | \hat{y}_{i,1} - \hat{\mu}_i, y_{i,1}) = \underline{P}(\rho) Q \varepsilon_i,$$

$$\text{Var}(\hat{y}_{i,1} - \hat{\mu}_i | y_{i,1}) = \frac{1}{T-1} \sigma^2,$$

$$\text{Cov}(y_i - y_{i,-1}, \hat{y}_{i,1} - \hat{\mu}_i | y_{i,1}) = \frac{1}{T-1} \sigma^2 \underline{P}(\rho) \iota,$$

$$\text{Var}(y_i - y_{i,-1} | \hat{y}_{i,1} - \hat{\mu}_i, y_{i,1}) = \sigma^2 [\underline{V}(\rho) - \frac{1}{T-1} \underline{P}(\rho) \iota \iota' \underline{P}(\rho)] = \sigma^2 \underline{P}(\rho) Q \underline{P}(\rho)',$$

$$\text{and } \varepsilon_i' Q \underline{P}(\rho)' [\underline{P}(\rho) Q \underline{P}(\rho)']^{-1} \underline{P}(\rho) Q \varepsilon_i = \varepsilon_i' Q \varepsilon_i.$$

It follows that the conditional likelihood function is equal to the joint normal density of the $\underline{P}(\rho) Q \varepsilon_i$, $i = 1, \dots, N$.

The Conditional ML estimator for ρ based on $\prod_{i=1}^N L_c(y_i - y_{i,-1} | \hat{y}_{i,1} - \hat{\mu}_i, y_{i,1})$ is inconsistent, because the probability limit of the conditional likelihood is not maximized at the true value ρ . This can be seen by comparing the probability limits of the conditional likelihood equation for ρ and the modified likelihood equation for ρ when both are evaluated at the true values of the parameters. The difference between them is $\xi'(\rho) + \frac{1}{2} \frac{d}{d\rho} \log |\underline{P}(\rho) Q \underline{P}(\rho)'|^+ = f(\rho) \neq 0$, where $|X|^+$ is the product of non-zero eigenvalues of the matrix X . Since the Modified ML estimators for ρ and σ^2 are consistent, it follows that the Conditional ML estimator for ρ must be inconsistent. \square

Notice that the Conditional MLE for ρ is different from the Within MLE/LSDV estimator for ρ , which however is also inconsistent.

B. Results for section 3.

CONSISTENCY OF THE REMLE FOR THE CONDITIONAL AR(1) PANEL MODEL:

This ML estimator is defined as the maximizer of a product of proper density functions. This likelihood function does not involve any incidental parameters. Consistency of the REMLE therefore follows by standard arguments. Nevertheless, it is interesting to see that the asymptotic first order conditions are satisfied at the true parameter values:

Recall that $u_i \equiv v_i\iota + \varepsilon_i$, where both v_i and ε_i are i.i.d., with $E(y_{i,1}v_i) = 0$, and $E(y_{i,1}\varepsilon_i) = 0$, [and $E(v_i\varepsilon_i) = 0$], $\forall i \in \mathcal{T}$. Let $\tilde{\Sigma} = E(u_i u_i') = \sigma^2 I + \sigma_v^2 \iota \iota'$. Then $E(y_{i,-1}' \tilde{\Sigma}^{-1} u_i) = 0$, since $E(y_{i,-1}' \tilde{\Sigma}^{-1} u_i) = E[\text{tr}(\tilde{\Sigma}^{-1} u_i y_{i,-1}')] =$

$$E \left[\text{tr} \left(\tilde{\Sigma}^{-1} \begin{bmatrix} 0 & u_1^2 & u_1 u_2 + \rho u_1^2 & \cdots \\ 0 & u_1 u_2 & u_2^2 + \rho u_1 u_2 & \cdots \\ 0 & u_1 u_3 & u_2 u_3 + \rho u_1 u_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \right) \right] = \text{tr} \left(\tilde{\Sigma}^{-1} \tilde{\Sigma} \begin{bmatrix} 0 & 1 & \rho & \rho^2 & \cdots \\ 0 & 0 & 1 & \rho & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \right)$$

$= 0$.

It is also easily verified that $E(y_{i,1} \iota' \tilde{\Sigma}^{-1} u_i) = 0$, $E[\text{tr}(u_i u_i' - \tilde{\Sigma}) \frac{\partial \tilde{\Sigma}^{-1}}{\partial \sigma^2}] = 0$, and $E[\text{tr}(u_i u_i' - \tilde{\Sigma}) \frac{\partial \tilde{\Sigma}^{-1}}{\partial \sigma_v^2}] = 0$.

PROOF OF THEOREM 3.1

The FE (Quasi) MLE for $\{\rho, \sigma^2\}$ in the conditional AR(1) panel model is based on the (quasi) likelihood function corresponding to the following (auxiliary) model

$$y_i - y_{i,1}\iota = \rho(y_{i,-1} - y_{i,1}\iota) + u_i, \quad (\text{B.1})$$

where $u_i = -(1 - \rho)v_{i,1}\iota + \varepsilon_i \sim N(0, (1 - \rho)^2 \sigma_v^2 \iota \iota' + \sigma^2 I)$.

Recall that $v_{i,1} = y_{i,1} - \mu_i$. We assume that $-1 < \rho < 1$, $\sigma^2 > 0$, and that assumptions SA and RFE1 hold. Notice that if $\rho = 1$, the term $v_i = -(1 - \rho)v_{i,1}$ would vanish and the parameter $\sigma_v^2 = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N v_{i,1}^2$ would also vanish.

It is convenient to reparametrize the auxiliary model, that is, to replace $(1 - \rho)^2 \sigma_v^2$ by $\tilde{\sigma}_v^2$. Furthermore, let $\tilde{\Sigma} = E(u_i u_i') = \tilde{\sigma}_v^2 \iota \iota' + \sigma^2 I_{T-1}$, $\tilde{\Delta} y_i = y_i - y_{i,1}\iota$ and $\tilde{\Delta} y_{i,-1} = y_{i,-1} - y_{i,1}\iota$. Then the (quasi) log-likelihood function is given by

$$\begin{aligned} \log L &= -\frac{1}{2} N(T-1) \log 2\pi - \frac{N}{2} \log |\tilde{\Sigma}| \\ &\quad - \frac{1}{2} \sum_{i=1}^N (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})' \tilde{\Sigma}^{-1} (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1}). \end{aligned} \quad (\text{B.2})$$

After noting that $\tilde{\Sigma}$ equals $\sigma^2 Q + (\sigma^2 + (T-1)\tilde{\sigma}_v^2) \frac{1}{T-1} \iota \iota'$, where $Q = I_{T-1} - \frac{1}{T-1} \iota \iota'$, it follows by applying some classical matrix algebra results from the panel data literature that $\tilde{\Sigma}^{-1} = \frac{1}{\sigma^2} Q + \frac{1}{\sigma^2 + (T-1)\tilde{\sigma}_v^2} \frac{1}{T-1} \iota \iota'$ and $|\tilde{\Sigma}| = \sigma^{2(T-2)} (\sigma^2 + (T-1)\tilde{\sigma}_v^2)$ (see e.g. Hsiao (1986)). These results allow us to rewrite the likelihood function as

$$\begin{aligned} \log L = & -\frac{1}{2} N(T-1) \log 2\pi - \frac{N(T-2)}{2} \log \sigma^2 - \frac{N}{2} \log(\sigma^2 + (T-1)\tilde{\sigma}_v^2) \\ & - \frac{1}{2\sigma^2} \sum_{i=1}^N (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})' Q (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1}) \\ & - \frac{1}{2(\sigma^2 + (T-1)\tilde{\sigma}_v^2)} \frac{1}{T-1} \sum_{i=1}^N [\iota' (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})]^2. \end{aligned} \quad (\text{B.3})$$

To prove consistency of the FE(Q)MLE, we will verify the conditions of theorem 4.1.1. in Amemiya (1986). The main difference with a proof for standard ML estimators is that the $v_{i,1}$ are not i.i.d. but satisfy RFE1.

We can express $\tilde{\Delta} y_{i,-1}$ in terms of $v_{i,1}$ and ε_i ,

$$\tilde{\Delta} y_{i,-1} = \check{P} \iota (\rho - 1) (y_{i,1} - \mu_i) + \check{P} \varepsilon_i, \quad (\text{B.4})$$

where the matrix \check{P} is defined in (A.2).

Next, we can rewrite the (quasi) log-likelihood function using that

$$\begin{aligned} \tilde{\Delta} y_i - r \tilde{\Delta} y_{i,-1} &= (\rho - r) \tilde{\Delta} y_{i,-1} + v_i \iota + \varepsilon_i = \\ &= [(\rho - r) \check{P} \iota + \iota] (\rho - 1) (y_{i,1} - \mu_i) + [(\rho - r) \check{P} + I] \varepsilon_i. \end{aligned} \quad (\text{B.5})$$

It can easily be seen from (B.5) that the (quasi) log-likelihood function, divided by N , converges uniformly in probability if and only if assumption RFE1 holds.

We note that one would obtain the same probability limit of the (quasi) log-likelihood function if the $v_{i,1}$ were i.i.d. normal with $E(v_{i,1}) = 0$, $E(v_{i,1}^2) = \sigma_v^2$, and $E(v_{i,1} \varepsilon_i) = 0$. In that case we would have a standard ML estimation problem. It immediately follows that the probability limit of the (quasi) log-likelihood function attains a uniquely maximum at the true values of the parameters. \square

It is interesting to see that the probability limit of the (quasi) log-likelihood function is indeed maximized at the true values of the parameters, despite correlation between the $y_{i,-1} - y_{i,1} \iota$ and the u_i :

First, $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \tilde{\Delta} y'_{i,-1} \tilde{\Sigma}^{-1} u_i =$

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \text{tr}[\tilde{\Sigma}^{-1} u_i \tilde{\Delta} y'_{i,-1}] = \text{tr} \left(\tilde{\Sigma}^{-1} \tilde{\Sigma} \begin{bmatrix} 0 & 1 & \rho & \rho^2 & \cdots \\ 0 & 0 & 1 & \rho & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \right) = 0,$$

of the likelihood equation for ρ for the REMLE given above.

Furthermore, it is obvious that $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N [\text{tr}(u_i u'_i - \tilde{\Sigma}) \frac{\partial \tilde{\Sigma}^{-1}}{\partial \sigma^2}] = 0$, and

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N [\text{tr}(u_i u'_i - \tilde{\Sigma}) \frac{\partial \tilde{\Sigma}^{-1}}{\partial \sigma_v^2}] = 0.$$

A COMPARISON OF THE MODIFIED MLE AND THE FEMLE FOR ρ :

We will assume that $u_i = v_i \iota + \varepsilon_i \sim i.i.d. N(0, \tilde{\Sigma})$, where $\tilde{\Sigma} = \tilde{\sigma}_v^2 \iota \iota' + \sigma^2 I_{T-1}$ with $\tilde{\sigma}_v^2 = (1 - \rho)^2 \sigma_v^2$.

The Modified ML estimator is a solution of the modified likelihood equations for ρ and σ^2 , which are given by:

$$\begin{aligned} \xi'(\rho) + \frac{1}{\sigma^2} \frac{1}{N} \sum_{i=1}^N (y_i - \rho y_{i,-1})' Q y_{i,-1} &= 0, \text{ and} \\ -\frac{T-2}{2\sigma^2} + \frac{1}{2\sigma^2} \frac{1}{N} \sum_{i=1}^N (y_i - \rho y_{i,-1})' Q (y_i - \rho y_{i,-1}) &= 0, \end{aligned}$$

where $\xi(r) = \frac{1}{T-1} \sum_{t=1}^{T-2} \frac{T-1-t}{t} r^t$.

The Fixed Effects (Quasi) ML estimator is defined as the global maximizer of the (quasi) log-likelihood function given in (B.3). The corresponding likelihood equations for ρ , σ^2 and $\tilde{\sigma}_v^2$ are given by:

$$\frac{\partial \log L}{\partial \rho} = \sum_{i=1}^N \tilde{\Delta} y'_{i,-1} \tilde{\Sigma}^{-1} (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1}) = 0, \quad (\text{B.6})$$

$$\begin{aligned} \frac{\partial \log L}{\partial \sigma^2} &= -\frac{N(T-2)}{2\sigma^2} - \frac{N}{2\tilde{\sigma}^2} + \frac{1}{2\sigma^4} \sum_{i=1}^N (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})' Q (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1}) \\ &\quad + \frac{1}{2\tilde{\sigma}^4} \frac{1}{T-1} \sum_{i=1}^N [\iota' (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})]^2 \\ &= 0 \end{aligned} \quad (\text{B.7})$$

and

$$\frac{\partial \log L}{\partial \tilde{\sigma}_v^2} = -\frac{N(T-1)}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4} \sum_{i=1}^N [\iota' (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})]^2 = 0, \quad (\text{B.8})$$

where $\tilde{\sigma}^2 = \sigma^2 + (T-1)\tilde{\sigma}_v^2$. We cannot solve this system of first-order conditions of the maximum likelihood problem analytically, but we can simplify it. This yields the following system of equations:

$$\begin{aligned} & [\frac{1}{N} \sum_{i=1}^N (\tilde{\Delta} y'_{i,-1} [\tilde{\sigma}_v^2 (T-1)Q + \sigma^2 I] \tilde{\Delta} y_{i,-1})] \rho = \\ & \quad \frac{1}{N} \sum_{i=1}^N (\tilde{\Delta} y'_{i,-1} [\tilde{\sigma}_v^2 (T-1)Q + \sigma^2 I] \tilde{\Delta} y_i), \\ \sigma^2 &= \frac{1}{(T-2)} \frac{1}{N} \sum_{i=1}^N (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})' Q (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1}) = \\ & \quad \frac{1}{(T-2)} \frac{1}{N} \sum_{i=1}^N (y_i - \rho y_{i,-1})' Q (y_i - \rho y_{i,-1}), \\ \text{and } \tilde{\sigma}_v^2 &= \frac{1}{(T-1)^2} \frac{1}{N} \sum_{i=1}^N [\iota' (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})]^2 - \frac{\sigma^2}{(T-1)}. \end{aligned}$$

When $T = 3$, it is easily verified that $\hat{\rho}_{MML} = \hat{\rho}_{FEMLE}$ and $\hat{\sigma}_{MML}^2 = \hat{\sigma}_{FEMLE}^2$. When $T \geq 4$, the Modified MLE for ρ and the FEMLE for ρ differ, because $\frac{1}{N} \sum_{i=1}^N \tilde{\Delta} y'_{i,-1} (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1}) \neq [\tilde{\sigma}_{v,FEMLE}^2(\rho)](T-1)\xi'(\rho)$.

We will now compare the limiting variances of the Modified MLE and the FEMLE. The asymptotic variance of the Modified MLE was derived in the proof of Theorem 2.14. in appendix A. To derive the limiting variance of the FEMLE, we will use that $\tilde{\Delta} y_{i,-1} = \check{P}\iota(\rho-1)(y_{i,1} - \mu_i) + \check{P}\varepsilon_i = \check{P}(v_i\iota + \varepsilon_i)$ (see (B.4)). Then the following results are easily obtained:

$$\begin{aligned} E[(\tilde{\Delta} y_{i,-1})' \tilde{\Sigma}^{-1} (\tilde{\Delta} y_{i,-1})] &= \sigma^2 \text{tr}(\check{P}' \tilde{\Sigma}^{-1} \check{P}) + \tilde{\sigma}_v^2 \iota' \check{P}' \tilde{\Sigma}^{-1} \check{P} \iota, \\ E[\iota' (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})]^2 &= (T-1)\tilde{\sigma}^2, \\ E[(\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})' Q (\tilde{\Delta} y_{i,-1})] &= E[\varepsilon_i' Q \check{P}(v_i\iota + \varepsilon_i)] = \sigma^2 \text{tr}(Q \check{P}) = -\sigma^2 \xi'(\rho), \\ \text{and } E[\iota' (\tilde{\Delta} y_{i,-1}) (\tilde{\Delta} y_i - \rho \tilde{\Delta} y_{i,-1})' \iota] &= E[\iota' \check{P}(v_i\iota + \varepsilon_i) (v_i\iota + \varepsilon_i)' \iota] = \\ &= [(T-1)\tilde{\sigma}_v^2 + \sigma^2] \iota' \check{P} \iota = \tilde{\sigma}^2 \iota' \check{P} \iota = \tilde{\sigma}^2 (T-1) \xi'(\rho). \end{aligned}$$

Let $\delta = (\rho \ \sigma^2 \ \tilde{\sigma}_v^2)'$. Then the limiting variance of the FEMLE for δ is given by

$$\text{Var}(N^{0.5} \hat{\delta}_{FEMLE}) = NE \left[-\frac{\partial^2 \log L}{\partial \delta \partial \delta'} \right]^{-1} = \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{12} & \frac{T-2}{2\tilde{\sigma}^4} + \frac{1}{2\tilde{\sigma}^4} & \frac{V_{13}}{2\tilde{\sigma}^4} \\ V_{13} & \frac{T-1}{2\tilde{\sigma}^4} & \frac{(T-1)^2}{2\tilde{\sigma}^4} \end{bmatrix},$$

where $V_{11} = \sigma^2 \text{tr}(\check{P}' \tilde{\Sigma}^{-1} \check{P}) + \tilde{\sigma}_v^2 \iota' \check{P}' \tilde{\Sigma}^{-1} \check{P} \iota = \sigma^2 \text{tr}(Q \check{P}' \tilde{\Sigma}^{-1} \check{P}) + \frac{\tilde{\sigma}^2}{T-1} \iota' \check{P}' \tilde{\Sigma}^{-1} \check{P} \iota$, $V_{12} = -\frac{1}{\sigma^2} \xi'(\rho) + \frac{1}{\tilde{\sigma}^2} \xi'(\rho)$, and $V_{13} = \frac{1}{\tilde{\sigma}^2} (T-1) \xi'(\rho)$.

If $u_i = v_i\iota + \varepsilon_i \sim i.i.d. N(0, \tilde{\Sigma})$, then the FEMLE attains the Cramér-Rao lowerbound for the conditional AR(1) panel model with (restricted) fixed effects and is therefore asymptotically efficient.

When $T = 3$ and $-1 < \rho < 1$ (and $u_i \sim i.i.d. N(0, \tilde{\Sigma})$), it is easily verified that $\text{Var}(\hat{\rho}_{MML}) = \text{Var}(\hat{\rho}_{FEMLE}) = \frac{2(1-\rho)^2 \sigma_v^2 \sigma^2 + \sigma^4}{(1-\rho)^4 \sigma_v^4}$. When $T \geq 4$ (and $-1 < \rho < 1$ and $u_i \sim i.i.d. N(0, \tilde{\Sigma})$), one can use *Maple* to verify that $\text{Var}(\hat{\rho}_{MML}) > \text{Var}(\hat{\rho}_{FEMLE})$.

When ρ tends to one, the variances of $\hat{\rho}_{MML}$ and $\hat{\rho}_{FEML}$ explode: $\lim_{\rho \uparrow 1} Var(\hat{\rho}_{MML}) = \infty$, and $\lim_{\rho \uparrow 1} Var(\hat{\rho}_{FEML}) = \infty$. Thus both the Modified ML estimator for ρ and the FEML estimator for ρ are very imprecise near the unit root. In the proof of Theorem 2.14 we have seen that when $\rho = 1$ the probability limit of the Hessian of the Modified log-likelihood function, MH , is singular, which means that the parameters ρ and σ^2 are not uniquely identified in this case. Similarly, we can easily show that $Var(N^{0.5}\hat{\delta}_{FEML})$ is a singular matrix when $\rho = 1$. In that case $\tilde{\sigma}_v^2 = 0$, $\tilde{\sigma}^2 = \sigma^2$ and $\tilde{\Sigma} = \sigma^2 I$. Moreover, $tr(\check{P}'\check{P}) = \frac{1}{2}(T-1)(T-2)$ and $\xi'(1) = \frac{1}{2}(T-2)$. These properties of $\hat{\rho}_{MML}$ and $\hat{\rho}_{FEML}$ near and at the unit root are consistent with results in Kruiniger (2000b), namely that all the first and second moment conditions for the conditional AR(1) panel model are weak near $\rho = 1$ and do not identify the unit root.

When the assumption that $u_i = v_i\iota + \varepsilon_i \sim i.i.d. N(0, \tilde{\Sigma})$ does not hold, the limiting variance of $\hat{\delta}_{FEQML}$ takes the form of a sandwich formula, $H^{-1}GH^{-1}$, where H is the probability limit of the Hessian and G is the probability limit of the outerproduct of the gradient of the quasi log-likelihood function (cf MaCurdy (1981b)). In this case only the OFECGMM estimator is still asymptotically efficient and the FEQMLE for ρ and σ^2 may asymptotically now be less efficient than the Modified MLE for these parameters. Furthermore, the middle part of the sandwich formula will in general depend on fourth order moments of v_i and $\varepsilon_{i,t}$. On the other hand, the limiting variance of the Modified MLE still only depends on the second moment of v_i , which one may consider as an advantage of the MMLE over the FEQMLE if one is only prepared to make minimal assumptions about the fixed effects (like assumption RFE).

C. Proof for section 4.

We will prove that the FGLS estimator for the static model with fixed effects is invariant under different within transformations of the data. It suffices to show that

$$K'(KVK')^{-1}K = Q'(QVQ)^{-}Q \quad (\text{C.1})$$

for each $(T - 1 \times T)$ matrix K such that $K\iota = 0$, and $\text{rank}(K) = T - 1$.

PROOF

Define the $(T \times T)$ matrix H as $VQ + [0 \ \iota]$, and let D be the $(T - 1 \times T)$ first-difference matrix.

Note that $\text{span}(VQ) = \text{span}(VD')$, because $\text{span}(Q) = \text{span}(D')$. Moreover, $V[D' \ 0] + [0 \ \iota]$ is a non-singular matrix, i.e. $VD'x_1 + \iota x_2 = 0 \Leftrightarrow x_1 = 0 \wedge x_2 = 0$, because $D(VD'x_1 + \iota x_2) = DVD'x_1 = 0 \Leftrightarrow x_1 = 0$, and $\iota x_2 = 0 \Leftrightarrow x_2 = 0$. It follows that the matrix H is nonsingular.

After pre- and postmultiplying (C.1) by H , we obtain $QVK'(KVK')^{-1}KVQ = QVQ$.

Define $G' = KV^{1/2}$. Then $QVK'(KVK')^{-1}KVQ = QV^{1/2}G(G'G)^{-1}G'V^{1/2}Q$.

Noting that $\text{span}(G) = \text{span}(V^{1/2}K') = \text{span}(V^{1/2}Q)$, we conclude that $QV^{1/2}G(G'G)^{-1}G'V^{1/2}Q = QVQ$. \square

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] Amemiya, T., 1986, *Advanced Econometrics*, Harvard University Press.
- [4] Andersen, E.B., 1970, Asymptotic properties of conditional maximum-likelihood estimators, *JRSS Series B* 32, 283-301.
- [5] Andersen, E.B., 1973, *Conditional inference and models for measuring*, Mentalhygiejnisk Forlag, Copenhagen.
- [6] Arellano, M., and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of economic studies* 58, 277-297.
- [7] Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of econometrics* 68, 29-51.
- [8] Balestra, P., and M. Nerlove, 1966, Pooling cross-section and time series data in the estimation of a dynamic model: the demand for natural gas, *Econometrica* 34, 585-612.
- [9] Chamberlain, G., 1980, Analysis of covariance with qualitative data, *Review of economic studies* XLVII, 225-238.
- [10] Chamberlain, G., 1984, Panel data, Ch. 22 in *Handbook of econometrics*, Vol. III, North Holland/Elsevier, Amsterdam.
- [11] Davidson, J., 1994, *Stochastic Limit Theory*, Oxford University Press.
- [12] Hsiao, C., 1986, *Analysis of Panel Data*, Cambridge University Press.
- [13] Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu, 1998, Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods, unpublished manuscript, revised April 2001, University of Cambridge.
- [14] 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.

- [15] Kruiniger, H., 1998, CML estimation of dynamic panel data models, Discussion Paper 98-04, Dept. of Economics, University College London.
- [16] Kruiniger, H., 2000a, ML and GMM estimation of dynamic panel data models with fixed effects, Working paper 429, Dept. of Economics, Queen Mary, University of London. Also available at <http://papers.ssrn.com/> .
- [17] Kruiniger, H., 2000b, GMM estimation of dynamic panel data models with persistent data, Working paper 428, Dept. of Economics, Queen Mary, University of London. Also available at <http://papers.ssrn.com/> .
- [18] Lancaster, T., and B. Lindenhovius, 1996, Biases in dynamic panel data models: a reconsideration. Working paper 96-11, Brown University, Providence, Rhode Island.
- [19] Lancaster, T., 1997, Orthogonal parameters and panel data, Working paper 97-32, revised May 1999 and March 2001, Brown University, Providence, Rhode Island.
- [20] Lindley, D.V., and G.M. El-Sayyed, 1968, The Bayesian estimation of a linear functional relationship, JRSS Series B 30, 190-202.
- [21] MaCurdy, T., 1981a, Multiple time series models applied to panel data, NBER working paper no. W0646, Cambridge, Massachusetts.
- [22] MaCurdy, T., 1981b, Asymptotic properties of quasi-maximum likelihood estimators and test statistics, NBER Technical working paper no. T0014, Cambridge, Massachusetts.
- [23] Mundlak, Y., 1978, On the pooling of time series and cross section data, *Econometrica* 46, 69-85.
- [24] Newey, N., and D. McFadden, 1994, Large sample estimation and hypothesis testing, Ch. 36 in *Handbook of econometrics*, Vol. IV, North Holland/Elsevier, Amsterdam.
- [25] Neyman, J., and E.L. Scott, 1948, Consistent estimates based on partially consistent observations, *Econometrica* 16, 1-32.
- [26] Nickell, S., 1981, Biases in dynamic models with fixed effects, *Econometrica* 49, 1417-1426.
- [27] Sims, C., 2000, Using the likelihood perspective to sharpen econometric discourse: three examples, *Journal of econometrics* 95, 443-462.