Department of Economics

Quasi ML Estimation of the Panel AR(1) Model with Arbitrary Initial Conditions

Hugo Kruiniger

Working Paper No. 582 December 2006 ISSN 1473-0278

Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions

Hugo Kruiniger*
Queen Mary, University of London

This version: December 2006

JEL classification: C12, C13, C23.

Keywords: dynamic panel data, initial conditions, Quasi ML, GMM, weak moment conditions, local-to-zero asymptotics.

*Address: h.kruiniger@qmw.ac.uk; Dept. of Economics, Mile End Road, London E1 4NS.
An earlier verion of this paper was presented at the UK Econometric Study Group meeting in Bristol in July 2004 and at the ESWC at UCL in August 2005. I would like to thank Steve Bond and Frank Windmeijer for helpful suggestions and for sharing some of their computer code with me. This research was funded by the ESRC under grant R000239139.

Abstract

In this paper we show that the Quasi ML estimation method yields consistent Random and Fixed Effects estimators for the autoregression parameter ρ in the panel AR(1) model with arbitrary initial conditions even when the errors are drawn from heterogenous distributions. We compare both analytically and by means of Monte Carlo simulations the QML estimators with the GMM estimator proposed by Arellano and Bond (1991) [AB], which ignores some of the moment conditions implied by the model. Unlike the AB GMM estimator, the QML estimators for ρ only suffer from a weak instruments problem when ρ is close to one if the cross-sectional average of the variances of the errors is constant over time, e.g. under time-series homoskedasticity. However, even in this case the QML estimators are still consistent when ρ is equal to one and they display only a relatively small bias when ρ is close to one. In contrast, the AB GMM estimator is inconsistent when ρ is equal to one, and is severly biased when ρ is close to one. Finally, we study the finite sample properties of two types of estimators for the standard errors of the QML estimators for ρ, and the bounds of QML based confidence intervals for ρ.

1 Introduction

In this paper we show that the Quasi ML estimation method yields consistent estimators for the autoregression parameter ρ in the *conditional* panel AR(1) model, i.e. the panel AR(1) model with arbitrary initial conditions, when the errors are drawn from heterogenous distributions. We consider both Random Effects and Fixed Effects QML estimators for ρ and compare them with various GMM estimators for ρ. In particular we analyze the distributional properties of various QML and GMM estimators for ρ when ρ is close to unity. We also compare the QML estimators with various prominent GMM estimators for ρ in a Monte Carlo study.

In the panel data literature, broadly speaking, two classes of estimators are considered: GMM (IV) estimators and Maximum Likelihood estimators. There is now a sizeable literature on GMM estimation of the panel AR(1) model, see e.g. Arellano and Bond (1991) and Ahn and Schmidt (1995, 1997). The Generalized Method of Moments owes much of its popularity to its flexibility: one can add or drop moment conditions depending on whether or not specific assumptions about the model are likely to be satisfied by the data. Furthermore GMM can be used in the presence of heterogenous data. For instance, the GMM estimator due to Arellano and Bond [AB], which only exploits orthogonality of lagged values of the dependent variable and the idiosyncratic errors, allows for both time-series and cross-sectional heteroskedasticity and arbitrary initial conditions. On the other hand, Monte Carlo studies have revealed that this GMM estimator is badly biased and very inprecise when the value of ρ is close to unity. This is due to a weak instruments problem, see Blundell and Bond (1998) and Kruiniger (2006a). However, Arellano and Bover (1995) and Blundell and Bond have shown that if the data also satisfy a mean-stationarity assumption, a GMM estimator with much better finite sample properties can be obtained, namely the System estimator.

The other major estimation principle, Maximum Likelihood, is generally not regarded as a viable alternative to the Generalized Method of Moments in the case of dynamic panel data models, because [one believes that] ML does not allow for heterogeneity of the idiosyncratic errors and fixed effects ML estimators are subject to the incidental parameters problem (cf Neyman and Scott, 1948). Nickell (1981) has shown that under stationarity of the data the naive fixed effects ML estimator for the panel AR(1) model with homoskedastic errors, i.e.

the Within Groups (WG) estimator, is inconsistent when the cross-sectional dimension of the panel, N, tends to infinity and the time dimension of the panel, T, is fixed, while Kiefer (1980) has shown that the fixed effects ML estimator for the covariance matrix of possibly autocorrelated errors of an otherwise static panel regression model is inconsistent when T is fixed.

The situation for the ML method is not so bleak if one looks further. MaCurdy (1981a, 1982) argued that the ML method does yield large-N consistent estimators for covariance stationary panel ARMA models with fixed effects when it is applied to first-differences of the data. The First Difference ML estimators for ρ are still consistent under cross-sectional heteroskedasticity. Both Chamberlain (1980) and Anderson and Hsiao (1982) proposed the same Random Effects MLE for the conditional panel AR(1) model. Recently, Hsiao et al. (2002) and Kruiniger (2001) have independently shown that the conditional panel AR(1) model with fixed effects (FE) and homogenously distributed errors can be consistently estimated by (Quasi) ML if the difference between the initial observations and the individual effects (or equivalently the differenced data) satisfy a particular condition.

The above survey suggests that it is difficult to obtain reliable estimates for the panel AR(1) model when the data are heterogenous and persistent but do not satisfy some form of stationarity, such as mean-stationarity. However, in this paper we show that the aforementioned RE MLE and FE (Quasi) MLE for the conditional panel AR(1) model still contain large-N consistent estimators for ρ when the errors display arbitrary heterogeneity of the type usually found in real panel data. [1] In this case these estimators for ρ should be reinterpreted as *Quasi* ML estimators in the most general sense.

When the cross-sectional average of the variances of the errors is constant over time and ρ is close to one, the aforementioned QML estimators for ρ also suffer from a 'weak instruments' problem, but — unlike the AB GMM estimator — they are still consistent when ρ is equal to unity. Furthermore, the REMLE attains the Cramér-Rao lowerbound for the conditional

[1] Earlier studies, e.g. Blundell and Bond (1998) and Alvarez and Arellano (2003), insisted that consistency of the REMLE (or the related FGLS estimator) requires homogeneity (or homoskedasticity) of the errors. Since real panel data are almost never homogenous, the REMLE would be of limited practical interest if this claim were true.

model if the errors are i.i.d. and Gaussian, whereas the AB GMM estimator ignores some of the moment conditions implied by the conditional panel AR(1) model. Therefore, these QML estimators may offer an attractive alternative to the AB GMM estimator.

To gain further insight into the performance of the Quasi ML estimators we have conducted various Monte Carlo experiments. In particular, we have compared the finite sample properties of the REQMLE, the FEQMLE, the Optimal AB GMM estimator and the Optimal System estimator for various values of ρ larger than 0.5. In the experiments we have studied how the properties of the estimators are affected if we change (1) the distribution of the differences between the initial conditions and the individual effects; (2) the distributions of the idiosyncratic errors; and/or (3) the ratio of the variances of the error components. For instance, we have carried out experiments where the idiosyncratic errors display arbitrary heteroskedasticity in both dimensions of the panel and where the errors follow MA(1) processes with heterogenous MA parameters. Among other things, we find that the QML estimators display only small biases and that they are (much) more precise than the AB GMM estimator in all scenarios that we have considered. Furthermore the differences in precision and RMSE between the QML estimators and the AB GMM estimator increase considerably as ρ gets closer to unity. The System estimator generally performs better than the QML estimators when all the moment conditions exploited by the former estimator are valid. However, when the assumption of mean-stationarity is violated, the System estimator can be substantially biased, whereas the QML estimators continue to perform well.

At the end of the paper we examine the finite sample properties of two types of estimators for the standard errors of the QML estimators for ρ, and the bounds of QML based confidence intervals for ρ. The first type of estimator that we consider is based on first-order fixed parameter asymptotics while the second type of estimator is the product of a simple bootstrap procedure.

The paper is organised as follows. Section 2 reviews various versions of the panel AR(1) model and their underlying assumptions. Section 3 discusses GMM and (Quasi) ML estimation of the panel AR(1) model. Section 4 analyzes the distributional properties of various QML and GMM estimators for ρ when ρ is close to unity. Section 5 discusses QML and GMM estimation of panel AR(1) models which include exogenous regressors. Section 6 conducts a Monte Carlo study into the finite sample properties of the QML estimators and various

prominent GMM estimators when the data are persistent and section 7 concludes. Proofs are collected in two appendices. Below P(S)DS denotes Positive (Semi-)Definite Symmetric.

2 The panel AR(1) model

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

$$\begin{aligned} y_{i,t} &= \rho y_{i,t-1} + w_{i,t}, \\ w_{i,t} &= \eta_i + \varepsilon_{i,t}, \text{ where } \eta_i = (1-\rho)\mu_i, \end{aligned} \tag{1}$$

for $i = 1, ..., N$ and $t = S+2, ..., T$. We assume that the number of 'individuals', N, is large, $S \leq 0$ and that the number of observations per 'individual', T, is fixed. Furthermore $-1 < \rho \leq 1$. Note that in the unit root case the individual effect, η_i, vanishes and that the model can be rewritten as

$$y_{i,t} - \mu_i = \rho(y_{i,t-1} - \mu_i) + \varepsilon_{i,t}.$$

The observations on $y = (y_{.,1} \ ... \ y_{.,T})'$ are independently distributed across the N individuals.

The idiosyncratic error term, $\varepsilon_{i,t}$, satisfies the following Standard Assumptions (SA):

$$E(\varepsilon_{i,t}) = 0 \text{ and } E\left|\varepsilon_{i,t}\right|^{2+\delta} < \infty \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T, \tag{2}$$

where $\delta > 0$ is arbitrarily small. The individual effects, the μ_i, are often treated as Random Effects. In this case we make the following assumptions (REA):

$$(y_{i,1} \ \eta_i)', \ i = 1, ..., N, \text{ are i.i.d. with } E(y_{i,1}^2) = \sigma_y^2 < \infty, \tag{3}$$

and [2]

$$E(\mu_i) = 0, \ E(\mu_i^2) = \sigma_\mu^2 < \infty, \text{ and } E(\mu_i y_{i,1}) = \sigma_{\mu y} \text{ when } |\rho| < 1. \tag{4}$$

We will assume that higher order moments of $(y_{i,1} \ \eta_i)'$ exist whenever this is required.

Unlike the RE estimators, the Fixed Effects estimators only exploit data in first differences. Kruiniger (2001) has shown that consistent estimation of FE versions of the model is

[2] We could allow $(y_{i,1} \ \eta_i)'$, $i = 1, ..., N$, to be i.h.d. at the cost of more complicated derivations.

only possible if $v_{i,1} \equiv y_{i,1} - \mu_i$, $i = 1, ..., N$, satisfy the Restricted Fixed Effects and Initial Conditions Assumption (RFEA):

$$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} v_{i,1}^2 = \sigma_v^2 < \infty \text{ when } |\rho| < 1. \tag{5}$$

Note that in the FE versions of the model only the differences between the initial conditions and the individual effects are restricted; the individual effects themselves remain unrestricted.

Apart from distinguishing between RE and FE versions of the model, we can distinguish among three versions of the panel AR(1) model which impose different restrictions on the initial conditions and the error components.

The least restricted RE version of the model only imposes Assumption B: [3]

$$E(\varepsilon_{i,t} y_{i,1}) = 0 \text{ and } E(\varepsilon_{i,t}\eta_i) = 0 \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T, \tag{6}$$

and

$$E(\varepsilon_{i,s}\varepsilon_{i,t}) = 0 \text{ for } i = 1, ..., N \text{ and } t \neq s. \tag{7}$$

In the FE case, assumption (6) is replaced by:

$$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} v_{i,1}\varepsilon_{i,t} = 0 \text{ for } t = 2, ..., T \text{ when } |\rho| < 1. \tag{8}$$

Note that REA plus (6) imply RFEA plus (8). We will refer to the model that only imposes Assumption B as the *conditional* model.

A stronger version of the RE version of model also imposes Assumption M:

$$E(y_{i,1} - \mu_i) = 0 \text{ and } E((y_{i,1} - \mu_i)\mu_i) = 0 \text{ for } i = 1, ..., N \text{ when } |\rho| < 1. \tag{9}$$

This assumption is also known as mean-stationarity (cf Arellano and Bover, 1995, and Blundell and Bond, 1998). Assumption M holds for instance when $S \to -\infty$.

Finally, the strongest version of the model imposes full covariance stationarity on the

[3] We can relax this assumption and allow for autocorrelation in $\{\varepsilon_{i,t}\}$ and correlation between ε_i and $y_{i,1}$, see e.g. Blundell ands Smith (1991). For instance, if $\varepsilon_{i,t} \sim MA(1)$, we would assume that $E(\varepsilon_{i,t}\eta_i) = 0$ for $t = 2, ..., T$, $E(\varepsilon_{i,t} y_{i,1}) = 0$ for $t = 3, ..., T$, and $E(\varepsilon_{i,s}\varepsilon_{i,t}) = 0$ for $|s - t| > 1$.

data, i.e. it also adds Assumption C which consists of assumptions TSH and STIV:

$$\sigma_{i,t}^2 = E(\varepsilon_{i,t}^2) = \sigma_i^2 < \infty, \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T, \quad \text{(TSH)}, \tag{10}$$

$$Var(y_{i,1} - \mu_i) = \frac{\sigma_i^2}{1 - \rho^2}, \text{ for } i = 1, ..., N \text{ when } |\rho| < 1 \quad \text{(STIV)}. \tag{11}$$

The first part of assumption C, TSH, means that the idiosyncratic errors, the $\varepsilon_{i,t}$, are homoskedastic over time; the second part, STIV, imposes stationarity on the variances of the initial conditions. Assume that TSH holds. Then STIV holds when $S \to -\infty$.

Below we will often use a weaker version of assumption TSH which we call TSH*. Let $\overline{\sigma}_t^2 = N^{-1} \sum_{i=1}^N \sigma_{i,t}^2$. Then assumption TSH* holds if and only if $\overline{\sigma}_t^2 = \overline{\sigma}_2^2$, for $t = 3, ..., T$. Finally, we will also consider estimators for conditional models that satisfy assumption TSH or assumption TSH*.

3 The estimators for the panel AR(1) model

3.1 GMM estimators

We will review the moment conditions that are available for GMM estimation of ρ in the covariance stationary panel AR(1) model with random effects.

Arellano and Bond have derived the following $(T-1)(T-2)/2$ linear moment conditions from assumption B, i.e. from (6)-(7):

$$E(y_{i,t-s} \Delta w_{i,t}) = 0 \text{ for } s = 2, ..., t-1 \text{ and } t = 3, ..., T, \tag{12}$$

where $\Delta w_{i,t} = w_{i,t} - w_{i,t-1} = \Delta \varepsilon_{i,t} = \Delta y_{i,t} - \rho \Delta y_{i,t-1}$. Assumption B also implies $T-3$ non-linear moment conditions (see Ahn and Schmidt, 1995):

$$E(w_{i,T} \Delta w_{i,t-1}) = 0 \text{ for } t = 4, ..., T. \tag{13}$$

The GMM estimator that exploits all the moment conditions in (12) and (13) will be referred to as the RE Conditional GMM (or RECGMM) estimator.

If mean-stationarity holds as well, we can add $T-2$ moment conditions to the moment conditions in (12) (see Arellano and Bover, 1995): [4]

$$E(w_{i,t} \Delta y_{i,t-1}) = 0 \text{ for } t = 3, ..., T. \tag{14}$$

[4] These moment conditions do not require that $E(y_{i,1} - \mu_i) = 0$.

The estimator that exploits the moment conditions in both (12) and (14) is known as the System (or SYS) estimator.

The assumption of time-series homoskedasticity, (10), also implies $T-2$ moment conditions (see Ahn and Schmidt, 1995):

$$E(w_{i,t}^2 - w_{i,t-1}^2) = 0 \text{ for } t = 3, ..., T. \tag{15}$$

The GMM estimator that exploits all the moment conditions in (12), (13) and (15) will be referred to as the RE HOmoskedastic Conditional GMM (or REHOCGMM) estimator.

If the mean-stationarity assumption is valid, we can replace the non-linear moment conditions in (15) by:

$$E(y_{i,t}w_{i,t} - y_{i,t-1}w_{i,t-1}) = 0 \text{ for } t = 3, ..., T. \tag{16}$$

The AB GMM estimator, which only exploits the moment conditions in (12), is inconsistent when $\rho = 1$. On the other hand the HOCGMM and SYS estimators are still consistent when $\rho = 1$. However, as the following result makes clear, consistency at $\rho = 1$ does not require that the data satisfy some form of stationarity:

Theorem 1 (Consistency of the RECGMM estimator for the conditional model) *Let assumptions SA, REA and B hold, and let* $-1 < \rho \leq 1$. *Then the RE Conditional GMM estimator for* ρ *is consistent if (and when* $\rho = 1$ *only if)* $T \geq 4$.

PROOF

See appendix A.1. There we show that if $\rho = 1$, then ρ is uniquely identified by the moment conditions in (13) unless $T < 4$ or $\overline{\sigma}_s^2/\overline{\sigma}_{s-1}^2 = \overline{\sigma}_3^2/\overline{\sigma}_2^2 \neq 1$ for all $s \in \{3, ..., T-1\}$. If $\rho = 1$ and the average variance $\overline{\sigma}_t^2$ changes at a constant rate between $t = 2$ and $t = T-1 \geq 3$, then ρ is only locally identified: $r = 1$ or $r = \overline{\sigma}_3^2/\overline{\sigma}_2^2$. However, when $\rho \neq 1$ and $T \geq 3$, then ρ is uniquely identified whatever the values of $\overline{\sigma}_t^2$ are. Therefore, when ρ is not uniquely identified, we know that $\rho = 1$. In appendix 1 we also show that if $\rho = 1$ and $T \geq 3$, then ρ is uniquely identified by the 'homoskedasticity' moment conditions in (15).

Thus although some of its underlying moment conditions are non-linear in ρ, an REC-GMM estimator can easily be implemented for any given weight matrix. Consistent estimation of the elements of the optimal weight matrix will be discussed at the end of section 3.

Sofar we have only discussed RE GMM estimators. The FE versions of these estimators

are obtained by considering the maximum subset of moment conditions that only involve first differences of the data. Note that the FE SYS estimator is equal to the FE AB GMM estimator. Note also that the variances of the FE estimators do not depend on σ_μ^2.

3.2 ML estimators for the conditional model

A Random Effects ML or GLS estimator for the conditional version of the panel AR(1) model stated in (1) will generally be inconsistent due to correlation between the individual effect η_i and the regressor, $y_{i,t-1}$. Given Assumption B, which maintains that $E(\varepsilon_{i,t}y_{i,1}) = 0$, $t = 2, ..., T$, the regressors and the η_i are correlated if and only if the initial conditions $y_{i,1}$ and the η_i are correlated. Following Chamberlain (1980), we can decompose the η_i into a term that depends on the $y_{i,1}$, and a term that does not. This leads to the following 'correlated effects' specification: [5]

$$\eta_i = \pi(1-\rho)y_{i,1} + (1-\rho)v_i,\ i = 1, ..., N, \tag{17}$$

where v_i is a new random effect and

$$\pi(1-\rho) = \text{plim}_{N\to\infty}\frac{\sum_{i=1}^N y_{i,1}\eta_i}{\sum_{i=1}^N y_{i,1}^2},$$

so that the v_i are uncorrelated with the initial conditions and subsequent observations on y:

$$\text{plim}_{N\to\infty}(1-\rho)N^{-1}\sum_{i=1}^N y_{i,t}v_i = 0,\ t = 1, ..., T.$$

Let $y_i = (y_{i,2}\ ...\ y_{i,T})'$, $y_{i,-1} = (y_{i,1}\ ...\ y_{i,T-1})'$ and let ι denote a vector of ones. Then using the decomposition of η_i given in (17) we can rewrite the conditional panel AR(1) model as

$$\begin{aligned} &y_i = \rho y_{i,-1} + \pi(1-\rho)y_{i,1}\iota + u_i, \\ &\text{where} \quad u_i = (1-\rho)v_i\iota + \varepsilon_i \quad \text{with} \quad E(\varepsilon_i\varepsilon_i') = diag(\sigma_t^2). \end{aligned} \tag{18}$$

After adding the assumption that the errors are i.i.d. and Gaussian, that is $u_i \sim i.i.d.\ N(0, (1-\rho)^2\sigma_v^2\iota\iota' + diag(\sigma_t^2))$, application of the Maximum Likelihood method to (18) yields

[5] W.l.o.g. we assume that $E(y_{i,1}) = 0$ so that $E(v_i) = 0$. $E(y_{i,1}) \neq 0$ can be handled by including an intercept in the model, see section 5.

Chamberlain's correlated RE ML Estimator for ρ, π, σ_v^2, $\sigma_2^2, ..., \sigma_{T-1}^2$ and σ_T^2. [6] When calculating the REMLE it is convenient to use the reparameterization $\widetilde{\sigma}_v^2 = (1-\rho)^2\sigma_v^2$.

A special case arises for $\pi = 0$. 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).

Blundell and Smith (1991) have discussed ML estimation of generalized versions of the model in (18) allowing for correlation between ε_i and $y_{i,1}$ but maintaining homogeneity of the distributions of the errors. For instance, if the $\varepsilon_{i,t}$ follow an MA(1) process and $E(\varepsilon_{i,2}y_{i,1}) \neq 0$, then one should apply the ML method to the following extended model:

$$\begin{aligned} &y_i = \rho y_{i,-1} + \pi(1-\rho)y_{i,1}\iota + \tau_1 y_{i,1}e_1 + \widetilde{u}_i, \\ &\text{where} \quad \widetilde{u}_i = (1-\rho)v_i\iota + \widetilde{\varepsilon}_i, \quad \text{with} \quad E(\widetilde{\varepsilon}_i\widetilde{\varepsilon}_i') = \Psi, \end{aligned} \tag{19}$$

where e_1 is the first column of an identity matrix and where $\Psi_{s,t} = 0$ for $|s-t| > 1$. By including the term $\tau_1 y_{i,1}e_1$ in the model, one ensures that the resulting error terms, the $\widetilde{\varepsilon}_{i,t}$, are uncorrelated with the regressors.

Hsiao et al. (2002) and Kruiniger (2001) have independently derived the FE ML estimator and the FE Quasi ML estimator, respectively, for the conditional model. We follow the exposition in Kruiniger (2001). One can obtain the FE (Quasi) MLE by replacing μ_i in the conditional model by $y_{i,1} - v_{i,1}$, and by assuming (or imposing) that the $v_{i,1}$ (and the ε_i) are i.i.d. and Gaussian. This is equivalent to imposing the restriction $\pi = 1$ on the conditional model in (18) and treating the $v_i = -v_{i,1}$ as random effects which are drawn from a normal distribution. This leads to the following formulation of the conditional model

$$\begin{aligned} &y_i = \rho y_{i,-1} + (1-\rho)y_{i,1}\iota + u_i, \\ &\text{where} \quad u_i = -(1-\rho)v_{i,1}\iota + \varepsilon_i \quad \text{with} \quad E(\varepsilon_i\varepsilon_i') = diag(\sigma_t^2), \end{aligned} \tag{20}$$

[6] Sims (2000) proposed an estimation approach for the panel AR(1) model where the initial conditions and the individual means follow a bivariate (normal) distribution that 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. The latter approach is therefore more general.

and where the $v_{i,1}$ satisfy assumptions RFEA and B. After imposing $u_i \sim i.i.d.\ N(0, \widetilde{\sigma}_v^2 \iota\iota' + diag(\sigma_t^2))$, application of the Maximum Likelihood method to (20) yields the FE (Quasi) ML Estimator for ρ, $\widetilde{\sigma}_v^2$, $\sigma_2^2, ..., \sigma_{T-1}^2$ and σ_T^2. [7] [8]

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.

Applying a nonsingular constant transformation to the model in (20), we obtain a model that only involves first differences of the data:

$$\begin{aligned} \Delta y_{i,2} &= u_{i,2} = -(1-\rho)v_{i,1} + \varepsilon_{i,2}, \\ \Delta y_{i,t} &= \rho \Delta y_{i,t-1} + \Delta \varepsilon_{i,t} \quad t = 3, ..., T. \end{aligned} \tag{21}$$

Thus the FE(Q)MLE only exploits data in first differences.

3.3 Quasi ML estimation of ρ when the data are heterogenously distributed

So far we have assumed homogeneity of the distributions of the $\varepsilon_{i,t}$ although we have allowed for possible heterogeneity in the distributions (i.e. the first two moments) of the $v_{i,1}$ in the case of the FEQMLE. However, such strong distibutional assumptions with respect to the error terms are almost never satisfied by panel data and therefore the REMLE and the FEQMLE discussed above seem almost useless for practical purposes.

It turns out that the above RE and FE ML estimation procedures still yield consistent estimators for ρ (and only for ρ) when the errors display arbitrary heterogeneity of the type usually found in real panel data provided that the relevant moments of the data exist. In that case one should reinterpret the above estimators as Quasi ML estimators for ρ where the label Quasi reflects the fact that the estimators allow for heterogenously distributed data.

[7] We assume that $N^{-1}\sum_{i=1}^{N} E(v_{i,1}) = 0$. RE and FE QML estimation of panel AR(1) models with a constant term and possibly exogenous regressors is discussed in section 5.

[8] The 'pure' FEMLE requires that the $v_{i,1}$ are i.i.d. and Gaussian. 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 Gaussian.

We have the following result:

Theorem 2 *Let assumptions SA, REA or RFEA, and B hold. Moreover let* $-1 < \rho \leq 1$. *Then the RE Quasi MLE and the FE Quasi MLE that are based on the likelihood fuctions corresponding to the models in (18) and (20), respectively, are consistent estimators for* ρ *if* ρ *is identified.*

PROOF

See appendix A.2. The conditions for identification of ρ in the case of the REQML estimator are the same as those for the RECGMM estimator which were discussed below theorem 1.

We have the following result on the fixed-parameter first-order asymptotic distribution of the FEMLE:

Theorem 3 *Let assumptions SA, RFEA and B hold. Moreover let the* ε_i *and* $v_{i,1}$ *be i.i.d. and Gaussian. Then the fixed-parameter first-order asymptotic distribution of the FEMLE based on (20) is normal when* $|\rho| < 1$ *but non-normal when the* $\varepsilon_{i,t}$ *are homoskedastic and* $\rho = 1$. *The limiting variance of the FEMLE for the homoskedastic case with* $|\rho| < 1$ *is given in appendix A.3.*

PROOF

See appendix A.3. One can easily prove a similar result with respect to the asymptotic distribution of the REMLE. Unlike the limiting variance of the FEMLE for ρ, the limiting variance of the REMLE for ρ depends on σ_y^2 and also on σ_μ^2 through $\widetilde{\sigma}_v^2 = (1-\rho)^2\sigma_v^2 = (1-\rho)^2(\sigma_\mu^2 - \pi^2\sigma_y^2)$. When $\rho = 1$, the Expected Hessian of the (quasi) log-likelihood function corresponding to the FE(Q)MLE is singular when it is evaluated at the true values of the parameters and when TSH* holds.

When the error components are not i.i.d. and Gaussian, the asymptotic distribution of the REQMLE (FEQMLE) will in general be different from the asymptotic distribution of the REMLE (FEMLE). When the error components are i.h.d., e.g. when they are cross-sectionally heteroskedastic, asymptotic normality of the REQMLE (FEQMLE) is implied by the Lindeberg-Feller Central Limit Theorem. When the error components are i.h.d. or i.i.d. and non-normal, the asymptotic variance of the REQMLE can 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). Provided that the $4+\delta-th$ moments of the data exist, the matrix G can be consistently estimated by a cross-sectional average of the outerproducts of the 'individual' contributions to the score vector where the parameters have been replaced by their REQML (FEQML) estimates.

In Kruiniger (2001) it has been shown that when the error components are i.i.d. and Gaussian the Optimal REHOCGMM estimator for ρ is asymptotically equivalent to the REMLE for ρ in the homoskedastic version of (18). A similar equivalence result holds for the Optimal RECGMM, FECGMM and FEHOCGMM estimators for ρ.

When the error components are i.i.d. and Gaussian, the QML estimators for ρ attain the relevant Cramér-Rao lowerbounds. Under non-normality or heterogeneity of the distributions of the error components an Optimal CGMM estimator for ρ is generally asymptotically more efficient than the corresponding QMLE for ρ.

The optimal versions of the CGMM estimators can only be computed in two-steps: an initial consistent estimator for ρ is required for consistent estimation of the elements of the optimal weight matrix. Clearly, the finite sample properties of these two-step CGMM estimators depend on the choice of the initial estimator that is used to compute the optimal weight matrix. One possibility is to use the corresponding QMLE as the initial estimator.

4 The distributional properties of QML and GMM estimators for ρ when ρ is close to unity

When ρ is close to unity, the (RE) AB GMM estimator suffers from a weak instruments problem, see appendix B.2 and also Blundell and Bond (1998) and Kruiniger (2006a). Stock and Wright (1997) have argued that doing local-to-zero asymptotics provides a better approximation to the finite sample distribution of a GMM estimator that exploits weak moment conditions than traditional first-order fixed parameter asymptotics. In this context local-to-zero means that the covariance between the (weak) instrument and the regressor becomes smaller when the sample size increases. To be specific let $Z_{3,i}^{\prime} = [y_{i,1}(\Delta y_{i,3} - \rho \Delta y_{i,2})$ $y_{i,1}(\Delta y_{i,4} - \rho \Delta y_{i,3})$ $y_{i,2}(\Delta y_{i,4} - \rho \Delta y_{i,3})$ ... $y_{i,T-2}(\Delta y_{i,T} - \rho \Delta y_{i,T-1})]$ and let $Z_{4,i} = -d(Z_{3,i})/d\rho$.

The local-to-zero approach recognizes that for $\rho = 1 - \lambda N^{-0.5}$ $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} Z_{4,i} = 0$ and that for this parameter sequence the mean of the vector $Z_4 = N^{-0.5} \sum_{i=1}^{N} Z_{4,i}$ remains of the same order of magnitude as the standard deviations of its elements when N grows large. Therefore, when ρ is close to one, the local-to-zero large sample distribution of $\widehat{\rho}_{AB} - \rho = Z_4' W_N Z_3 / Z_4' W_N Z_4$, where $Z_3 = N^{-0.5} \sum_{i=1}^{N} Z_{3,i}$ and $\{W_N\}$ is a sequence of weight matrices such that $\text{plim}_{N\to\infty} W_N = W$ exists, may provide a better approximation to the finite sample distribution of $\widehat{\rho}_{AB}$ than the fixed parameter large sample distribution of $N^{0.5}(\widehat{\rho}_{AB} - \rho) = \overline{Z}_4' W_N Z_3 / \overline{Z}_4' W_N \overline{Z}_4$, where $\overline{Z}_4 = N^{-1} \sum_{i=1}^{N} Z_{4,i}$. The following result is based on theorem 1 in Kruiniger (2006a).

Theorem 4 *Let assumptions SA, REA and B hold and let S be fixed so that $Var(y_{i,1} - \mu_i) \propto (1-\rho)^0$ $\forall i \in \{1, 2, ..., N\}$. In addition, let $\rho = 1 - \lambda N^{-0.5}$ with $\lambda \geq 0$, let $\sigma_{\mu y} = E(y_{i,1}\mu_i)$, let $\sigma_y^2 = E(y_{i,1}^2)$ and let W_N be an arbitrary sequence of PDS weight matrices with $plim_{N\to\infty} W_N = W$, where W is PDS and finite. Finally, let the $\varepsilon_{i,t}$ be i.i.d.*

(a) If $T = 3$ then $\widehat{\rho}_{AB} \xrightarrow{d} \rho + \frac{\check{Z}_1}{\check{Z}_2}$, with $\begin{bmatrix} \check{Z}_1 \\ \check{Z}_2 \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ \lambda\left(\sigma_{\mu y} - \sigma_y^2\right) \end{pmatrix}, \sigma^2 \sigma_y^2 \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right]$.

(b) If $T > 3$ then $\widehat{\rho}_{AB} \xrightarrow{d} \rho + \check{Z}_4' W \check{Z}_3 / \check{Z}_4' W \check{Z}_4$, where $\check{Z}_3$ and $\check{Z}_4$ are Gaussian random vectors with $E(\check{Z}_3) = 0$, $E(\check{Z}_{4,k}) \propto \lambda$ for $k = 1, 2, ..., (1/2)(T-1)(T-2)$, and $E(\check{Z}_3 \check{Z}_4') \neq 0$.

One obtains the asymptotic distribution of the AB GMM estimator for $\rho = 1$ by taking $\lambda = 0$. Kruiniger (2006a) also derives a local-to-unity asymptotic distribution of the AB GMM estimator when the data are covariance stationary, e.g. when assumption TSH holds and $S \to -\infty$.

The local asymptotic distributions of the AB GMM estimator also capture the fact that this estimator is biased when ρ is close to unity. The bias stems from the fact that the instruments are weak and the fact that $E(Z_3 Z_4') \neq 0$, again see Kruiniger (2006a) for details.

We now consider the distributional properties of the RECGMM and REHOCGMM estimators when ρ is close to unity. We have the following result:

Theorem 5 *Let assumptions SA, REA and B hold and let S be fixed so that $Var(y_{i,1} - \mu_i) \propto (1-\rho)^0$ $\forall i \in \{1, 2, ..., N\}$. Then the moment conditions in (12) and (15) are weak when ρ*

is close to unity. Moreover, when ρ is close to unity the moment conditions in (13) are all weak as well if and only if assumption TSH* (almost) holds.

PROOF

See appendix B.1.

Thus if the cross-sectional average of the variances of the errors is not constant over time, then one can find some moment conditions that are not weak when ρ is local to unity without having to invoke mean-stationarity of the data, namely some or all of the moment conditions in (13). We expect a GMM estimator which exploits these non-linear moment conditions, e.g. the RECGMM estimator, to have much better finite sample properties than the AB GMM estimator when ρ is close to one and assumption TSH* does not hold (approximately). On the other hand, if all the moment conditions in (12), (13), and (15) are weak, then again local-to-zero asymptotics yields a better approximation to the finite sample distributions of the RECGMM and REHOCGMM estimators than traditional fixed parameter asymptotics.

Since the RE and FE CGMM estimators are asymptotically equivalent to QML estimators when the error components are i.i.d and Gaussian, we expect to have a result related to theorem 5 for the latter estimators.

Theorem 6 *Let assumptions SA, REA (or RFEA) and B hold and let S be fixed so that $Var(y_{i,1} - \mu_i) \propto (1-\rho)^0 \ \forall i \in \{1, 2, ..., N\}$. Moreover, let ρ be local to unity. Then the probability limit of the Hessian (divided by N) of the (quasi) log-likelihood function of the REQMLE (FEQMLE) is almost singular if and only if assumption TSH* (almost) holds, irrespective of whether assumption TSH has been imposed or not.*

PROOF

See appendix B.2.

It is well-known that (Quasi) ML estimators can be reinterpreted as GMM estimators. The underlying moment conditions can be obtained by setting the expected score vector equal to zero. For instance, when assumption TSH holds and has been imposed, the RE-QMLE can be reinterpreted as a GMM estimator which exploits the following moment conditions: $E(y'_{i,-1}\Phi^{-1}u_i) = 0$, $E(y_{i,1}\iota'\Phi^{-1}u_i) = 0$, $E(tr((u_iu'_i - \Phi)\partial\Phi^{-1}/\partial\sigma^2)) = 0$, and $E(tr((u_iu'_i - \Phi)\partial\Phi^{-1}/\partial\widetilde{\sigma}_v^2)) = 0$, where $\Phi = E(u_iu'_i) = \sigma^2 I + \widetilde{\sigma}_v^2\iota\iota'$ and $\Phi^{-1} = \sigma^{-2}Q +$

$(\sigma^2 + (T-1)\widetilde{\sigma}_v^2)^{-1}\frac{1}{T-1}\iota\iota'$ with $Q = I_{T-1} - \frac{1}{T-1}\iota\iota'$. It follows that the expected Hessian of the (quasi) log-likelihood function equals the first derivative of the vector of moment conditions exploited by the QMLE with respect to the parameters. Therefore when the expected Hessian of the (quasi) log-likelihood function is almost singular, the (Quasi) MLE suffers from a 'weak moment conditions problem.' In that case one can also obtain the local-to-zero asymptotic distribution of the (Quasi) MLE.

For the FEMLE we have the following local-to-zero asymptotic result:

Theorem 7 *Let assumptions SA, RFEA, B and TSH hold, let* $T \geq 4$ *and let* S *be fixed so that* $Var(y_{i,1} - \mu_i) \propto (1-\rho)^0 \ \forall i \in \{1, 2, ..., N\}$. *In addition, let* $FH(d) = -\frac{1}{N}\frac{\partial^2 \log L_F(\delta)}{\partial\delta\partial\delta'}|_{\delta=d}$ *and* $FEH(\delta) = E(FH(\delta))$, *where* $L_F(\delta)$ *is the FE likelihood function corresponding to (20) and* $\delta = \delta(\rho) = (\rho \ \sigma^2 \ \widetilde{\sigma}_v^2(\rho))'$. *Let* $c_{\det}(\rho) = \det(FEH(\delta(\rho)))/(1-\rho)^2$, *and* $\overline{FEH}(\delta) = (FEH(\delta))^{-1}\det(FEH(\delta))$. *Let* $\check{\delta}$ *be such that* $(FH(\check{\delta}))(\widehat{\delta}_{FEML} - \delta) = N^{-1}\frac{\partial \log L_F(\delta)}{\partial\delta}$ *and* $\check{\delta}_k = \mu_k\widehat{\delta}_{FEML,k} + (1-\mu_k)\delta_k$ *for some* $\mu_k \in [0,1]$, $k = 1, 2, 3$. *Finally, let* $\rho = 1 - \lambda N^{-0.25}$ *with* $\lambda > 0$. *Then* $N^{0.25}(\widehat{\rho}_{FEML} - \rho) \xrightarrow{d} \widetilde{Z}_1/\widetilde{Z}_2$, *where* $\widetilde{Z}_1 \sim N\left(0, \lambda^2 \lim_{\rho\uparrow 1} c_{\det}(\rho)\overline{FEH}_{11}(\delta(1))\right)$, *and* $\widetilde{Z}_2 \equiv \lambda^2 \lim_{\rho\uparrow 1} c_{\det}(\rho) + Z_{\det}(\lambda)$ *is the limiting distribution of* $N^{0.5}\det(FH(\check{\delta}))$. *Moreover,* $\widetilde{Z}_1$ *and* $\widetilde{Z}_2$ *are correlated.*

PROOF

See appendix B.3. There we also show that $\overline{FEH}_{11}(\delta(1)) \neq 0$ and that $\det(FEH(\delta(\rho))) = (1-\rho)^2 c(\rho)$ where $c(\rho)$ is a polynomial in ρ with $c(1) \neq 0$. Hence $\lim_{\rho\uparrow 1} c_{\det}(\rho) = c_{\det}(1) \neq 0$. Furthermore $\det(FEH(\delta(1))) = 0$. The latter result is also shown in appendix A.3.

When ρ is close to unity and the REQMLE and the FEQMLE suffer from a 'weak moment conditions problem,' then the finite sample distributional properties of these QML estimators are different from those of the AB GMM estimator. Note also that the local-to-zero rate of convergence of the FEMLE is positive while that of the AB GMM estimator is zero. Indeed the FEMLE is consistent when ρ is (local to) one whereas the AB GMM estimator is not.

When ρ is local to unity but assumption TSH* does not (nearly) hold, the RE and FE QML estimators, just like the RE and FE CGMM estimators, do not suffer from a 'weak moment conditions problem' and their first-order fixed parameter asymptotic distributions are expected to give a reasonable approximation to their finite sample distributions.

5 Models with exogenous regressors

In this section we discuss QML and GMM estimation of dynamic panel data models with a constant term and one time-varying exogenous variable, i.e.

$$y_{i,t} = \alpha(1-\rho) + \rho y_{i,t-1} + \beta(1-\rho)x_{i,t} + (1-\rho)\mu_i + \varepsilon_{i,t}, \quad t = 2, ..., T, \tag{22}$$

To keep the discussion simple we assume that TSH holds, i.e. $E(\varepsilon_i\varepsilon_i') = \sigma^2 I$. We also assume that the $x_{i,1}$ are observed whenever this is required. We consider estimation of both FE and RE versions of this model and we also consider the possibility that the $x_{i,t}$ are correlated with the μ_i. Below the terms and moment conditions in curly brackets should be added whenever the $x_{i,t}$ are correlated with the μ_i. Furthermore we distinguish between the case where the $x_{i,t}$ are weakly exogenous, i.e. $E(x_i^t\varepsilon_{i,t}) = 0$, $t = 2, ..., T$, where $x_i^t = (x_{i,1}\ ...\ x_{i,t})'$, and the case where the $x_{i,t}$ are strictly exogenous, i.e. $E(x_i^T\varepsilon_{i,t}) = 0$, $t = 2, ..., T$.

We first assume that the $x_{i,t}$ are strictly exogenous. Then the RE and FE models can be estimated by the QML method. In addition, when the error components are i.i.d. and Gaussian, one can construct asymptotically equivalent GMM estimators.

To obtain the REQMLE for ρ and β, apply ML to the model in (22) with μ_i replaced by

$$\pi y_{i,1}\ \{+\gamma_1' x_i\} + v_i,$$

where $x_i = (x_{i,2}\ ...\ x_{i,T})'$.

Let $u_i = (1-\rho)v_i\iota + \varepsilon_i$. Then the REQMLE is asymptotically equivalent to a GMM estimator that exploits $E(y_{i,-1}'\Phi^{-1}u_i) = 0$, $E(y_{i,1}\iota'\Phi^{-1}u_i) = 0$, $E(\iota'\Phi^{-1}u_i) = 0$, $E(x_i'\Phi^{-1}u_i) = 0$, $\{E(x_i\iota'\Phi^{-1}u_i) = 0\}$, $E(tr((u_iu_i' - \Phi)\partial\Phi^{-1}/\partial\sigma^2)) = 0$, and $E(tr((u_iu_i' - \Phi)\partial\Phi^{-1}/\partial\widetilde{\sigma}_v^2)) = 0$, where $\Phi = E(u_iu_i') = \sigma^2 I + \widetilde{\sigma}_v^2\iota\iota'$.

To obtain the FEQMLE for ρ and β in the model given in (22), μ_i should be replaced by

$$y_{i,1} - \beta x_{i,1}\ \{+\gamma_2'(x_i - x_{i,-1})\} + v_i,$$

where $x_{i,-1} = (x_{i,1}\ ...\ x_{i,T-1})'$.

A GMM estimator that is asymptotically equivalent to the FEQMLE exploits $E((y_{i,-1} - y_{i,1}\iota)'\Phi^{-1}u_i) = 0$, $E(\iota'\Phi^{-1}u_i) = 0$, $E((x_i - x_{i,1}\iota)'\Phi^{-1}u_i) = 0$, $\{E((x_i - x_{i,-1})\iota'\Phi^{-1}u_i) = 0\}$, $E(tr((u_iu_i' - \Phi)\partial\Phi^{-1}/\partial\sigma^2)) = 0$, and $E(tr((u_iu_i' - \Phi)\partial\Phi^{-1}/\partial\widetilde{\sigma}_v^2)) = 0$, where $\Phi = E(u_iu_i') = \sigma^2 I + \widetilde{\sigma}_v^2\iota\iota'$.

Let us now assume that the $x_{i,t}$ are weakly exogenous. We show that in this case the QML method does not yield consistent estimators.[9] In the RE case one would replace μ_i in (22) by $\mu_i = \pi y_{i,1}\ \{+\gamma_3 x_{i,2}\} + v_i$. Let again $u_i = (1-\rho)v_i\iota + \varepsilon_i$. Then the REQMLE for ρ and β is inconsistent because the expected score vector is not equal to zero at the true value of the parameters. In particular, $E(y'_{i,-1}\Phi^{-1}u_i) \neq 0$, because generally $E(x_{i,s}\varepsilon_{i,t}) \neq 0$ when $s > t$. A similar argument applies to the FE case. However, when the $x_{i,t}$ are weakly exogenous one can still formulate consistent GMM estimators.

Let us define $z_{i,t} = y_{i,t-1} - \beta(1-\rho)\sum_{s=0}^{t-4}\rho^s x_{i,t-s-1}$, $t = 2, ..., T$, and $z_i = (z_{i,2}\ ...\ z_{i,T})'$. Then a consistent GMM estimator for the RE version of the model with weakly exogenous regressors exploits $E(z_i'\Phi^{-1}u_i) = 0$, $E(y_{i,1}\iota'\Phi^{-1}u_i) = 0$, $E(\iota'\Phi^{-1}u_i) = 0$, $E(\widetilde{x}_i^t u_{i,t}) = 0$, $t = 2, ..., T$, $E(tr((u_iu_i' - \Phi)\partial\Phi^{-1}/\partial\sigma^2)) = 0$, and $E(tr((u_iu_i' - \Phi)\partial\Phi^{-1}/\partial\widetilde{\sigma}_v^2)) = 0$, where $\widetilde{x}_i^t = (x_{i,2}\ ...\ x_{i,t})'$, $u_{i,t} = (1-\rho)v_i + \varepsilon_{i,t} = y_{i,t} - c(1-\rho) - \rho y_{i,t-1} - \beta(1-\rho)x_{i,t} - (1-\rho)\pi y_{i,1}$ $\{-(1-\rho)\gamma_2 x_{i,2}\}$, $t = 2, ..., T$, and $\Phi = E(u_iu_i') = \sigma^2 I + \widetilde{\sigma}_v^2\iota\iota'$. In this case consistency of the estimator also requires that the exogenous regressors are equi-correlated with the individual effects, i.e. $E(v_i\Delta x_{i,t}) = 0$, $t = 2, ..., T$. Since $E(v_i x_{i,2}) = 0$ (by construction if the term in $\{\}$ is included in the model), this assumption guarantees that $E(v_i x_{i,t}) = 0$, $t = 3, ..., T$.

Finally, a consistent GMM estimator for the FE model with weakly exogenous regressors exploits $E(\widetilde{z}_i'\Phi^{-1}u_i) = 0$, $E(\iota'\Phi^{-1}u_i) = 0$, $E((\widetilde{x}_i^t - x_{i,1}\iota)u_{i,t}) = 0$, $t = 2, ..., T$, $E\{tr((u_iu_i' - \Phi)\partial\Phi^{-1}/\partial\sigma^2)\} = 0$, and $E\{tr((u_iu_i' - \Phi)\partial\Phi^{-1}/\partial\widetilde{\sigma}_v^2)\} = 0$, where $u_{i,t} = (1-\rho)v_i + \varepsilon_{i,t} = y_{i,t} - y_{i,1} - c(1-\rho) - \rho(y_{i,t-1} - y_{i,1}) - \beta(1-\rho)(x_{i,t} - x_{i,1})\ \{-(1-\rho)\gamma_4(x_{i,2} - x_{i,1})\}$, $\widetilde{z}_{i,t} = y_{i,t-1} - y_{i,1} - \beta(1-\rho)\sum_{s=0}^{t-4}\rho^s(x_{i,t-s-1} - x_{i,1})$, $t = 2, ..., T$, $\widetilde{z}_i = (\widetilde{z}_{i,2}\ ...\ \widetilde{z}_{i,T})'$ and $\Phi = E(u_iu_i') = \sigma^2 I + \widetilde{\sigma}_v^2\iota\iota'$. Consistency of this estimator also requires that the exogenous regressors are equi-correlated with the individual effects, i.e. $E(v_i\Delta x_{i,t}) = 0$, $t = 3, ..., T$. Since $E(v_i\Delta x_{i,2}) = 0$ (by construction if the term in $\{\}$ is included in the model), this assumption implies that $E(v_i(x_{i,t} - x_{i,1})) = 0$, $t = 2, ..., T$. If the exogenous regressors are not equi-correlated with the individual effects, then one can still obtain a consistent FE GMM estimator by replacing $E((\widetilde{x}_i^t - x_{i,1}\iota)u_{i,t}) = 0$, $t = 2, ..., T$, by $E((\widetilde{x}_i^{t-1} - x_{i,1}\iota)\Delta u_{i,t}) = 0$, $t = 3, ..., T$.

[9] Thus the results in Hsiao et al. (2002) regarding this case are incorrect.

6 The finite sample performance of the estimators

In this section we compare through Monte Carlo simulations the finite sample properties of the REQMLE, the FEQMLE, the 2-step Optimal AB GMM estimator, the 2-step Optimal System estimator and the 2-step Optimal Linear GMM estimator due to Kruiniger (2003). In particular we study how the properties of these estimators are affected if we change (1) the conditional distributions of the differences between the initial conditions and the individual effects (the μ_i) given the individual effects, (2) the distributions of the idiosyncratic errors (the $\varepsilon_{i,t}$), and/or (3) the ratio of the variances of the error components.

In order to ascertain whether the QML estimators offer an attractive alternative to the AB GMM estimator we conducted the simulation experiments for $\rho = 0.5$, 0.8, 0.9 and 0.95. In most of the experiments $T = 6$ and $N = 100$. In one set of experiments, however, $T = 10$ and $N = 100$. For $\rho = 0.95$, we also conducted experiments where $T = 6$ and $N = 500$.

In all simulation experiments the error components have been drawn from normal distributions with zero means. For the distributions of the $\varepsilon_{i,t}$ we considered five different designs (indicated by a Roman number):

I Homogeneity and no autocorrelation of $\varepsilon_{i,t}$: $E(\varepsilon_i\varepsilon_i') = \sigma^2 I$ and $\sigma_{i,1}^2 = \sigma^2 = 1$.

II Arbitrary 'flat' heteroskedasticity but no autocorrelation of $\varepsilon_{i,t}$: $E(\varepsilon_i\varepsilon_i') = diag(\sigma_{i,t}^2)$ with $\sigma_{i,t} = \exp(-0.6+1.2U_i)\exp(-0.3+0.6V_{i,t})$, $t = 1, ..., T$, where $U_i \sim uniform[0,1]$, $V_{i,1} = 0.5$, and $V_{i,t} \sim uniform[0,1]$, $t = 2, ..., T$.

III Arbitrary 'non-flat' heteroskedasticity but no autocorrelation of $\varepsilon_{i,t}$: $E(\varepsilon_i\varepsilon_i') = diag(\sigma_{i,t}^2)$ with $\sigma_{i,t} = c_T((T+1)/T)^t \exp(-0.6 + 1.2U_i)\exp(-0.3 + 0.6V_{i,t})$, $t = 1, ..., T$, where $U_i \sim uniform[0,1]$, $V_{i,1} = 0.5$, $V_{i,t} \sim uniform[0,1]$, $t = 2, ..., T$, and $c_T = ((T+1)/T)^{-(T+1)/2}$.

IV Arbitrary 'flat' heteroskedasticity + individual specific MA(1): $\varepsilon_{i,t} = \omega_{i,t} + \phi_i\omega_{i,t-1}$, $t = 2, ..., T$, with $\phi_i = -0.6 + 1.2C_i$, $\omega_{i,t} \sim N(0, \sigma_{\omega,i,t}^2)$ and $\sigma_{\omega,i,t} = \exp(-0.6 + 1.2U_i)\exp(-0.3 + 0.6V_{i,t})$, where $C_i \sim uniform[0,1]$, $U_i \sim uniform[0,1]$, and $V_{i,t} \sim uniform[0,1]$, $t = 1, ..., T$. Furthermore $\sigma_{i,1}^2 = \sigma_{\omega,i,0}^2(1 + 2\rho\phi_i + \phi_i^2)$ with $\sigma_{\omega,i,0} = \exp(-0.6 + 1.2U_i)$.

V Arbitrary 'flat' heteroskedasticity + arbitrary MA(1): $\varepsilon_{i,t} = \omega_{i,t} + \phi_{i,t}\omega_{i,t-1}$, $t = 2, ..., T$, with $\phi_{i,t} = -0.6 + 1.2C_i - 0.15 + 0.3K_{i,t}$, $\omega_{i,t} \sim N(0, \sigma_{\omega,i,t}^2)$ and $\sigma_{\omega,i,t} = \exp(-0.6 + 1.2U_i)\exp(-0.3 + 0.6V_{i,t})$, where $C_i \sim uniform[0,1]$, $K_{i,t} \sim uniform[0,1]$, $t = 2, ..., T$, $U_i \sim$

$uniform[0,1]$, and $V_{i,t} \sim uniform[0,1]$, $t = 1, ..., T$. Furthermore $\sigma^2_{i,1} = \sigma^2_{\omega,i,0}(1+2\rho\phi_{i,1}+\phi^2_{i,1})$ with $\sigma_{\omega,i,0} = \exp(-0.6 + 1.2U_i)$ and $\phi_{i,1} = -0.6 + 1.2C_i$.

Note that in designs I, II, IV and V the cross-sectional average of the variances of the idiosyncratic errors is constant ('flat') over time, whereas in design III the cross-sectional average of these variances is increasing over time. Thus in designs I, II, IV and V the QML estimators suffer from a weak moment conditions problem when ρ is close to one, whereas in design III this is not the case.

For the individual effects we considered two scenarios: (IE1) $\sigma^2_\mu = 1$ and (IE2) $\sigma^2_\mu = 1/(1-\rho^2)$. Under design I, i.e. when $Var(\varepsilon_{i,t}) = \sigma^2$, scenario (IE2) keeps the ratio of the variances of the error components of $y_{i,t}$ constant across different values of ρ.

In order to assess how the assumptions with respect to $y_{i,1} - \mu_i$, $i = 1, ..., N$, affect the properties of the estimators, we conducted five different sets of experiments, which are identified by a capital: in one set the initial observations are drawn from 'stationary' distributions, (S), $(y_{i,1} - \mu_i)|\mu_i \sim N(0, \sigma^2_{i,1}/(1-\rho^2))$, whereas in the other four sets the initial observations are non-stationary. The four non-stationary cases considered are: (L): $y_{i,1} - \mu_i = 0$; (H): $(y_{i,1} - \mu_i)|\mu_i \sim N(0, 2\sigma^2_{i,1}/(1-\rho^2))$; (C): $(y_{i,1} - 2\mu_i)|\mu_i \sim N(0, \rho^2\sigma^2_{i,1}/(1-\rho^2))$; and (M): $(y_{i,1} - \mu_i)|\mu_i \sim N(\sigma_{i,1}/(1-\rho^2)^{0.5}, \sigma^2_{i,1}/(1-\rho^2))$.

Note that $(y_{i,t} - \mu_i)|\mu_i$ is stationary only under design I-S, i.e. when $\sigma^2_{i,t} = \sigma^2 = 1$, $t = 2, ..., T$, and $(y_{i,1} - \mu_i)|\mu_i \sim N(0, \sigma^2/(1-\rho^2))$. In both design L and design H the variance of $y_{i,1} - \mu_i$ is different from its variance under stationarity, while in design C non-stationarity of $(y_{i,t} - \mu_i)|\mu_i$ is due to the fact that $E(\mu_i(y_{i,1} - \mu_i)) \neq 0$. Note also that under design I-IE1-C $Var(y_{i,1} - \mu_i) = 1/(1-\rho^2)$, while under design I-IE2-C $Var(y_{i,1} - \mu_i) = (1+\rho^2)/(1-\rho^2)$. Finally, note that $E(y_{i,t} - y_{i,t-1}) = 0$ in all designs except design M.

When the data were generated according to design I, we imposed homoskedasticity on the likelihood functions. We also added the restrictions $s^2 > 0$ and $(T-1)\widetilde{s}^2_v + s^2 > 0$ to the likelihood functions to ensure that the estimates of $E(u_i u_i')$ were PDS. Under designs II-V we added instead the restrictions $s^2_t > 0$, $t = 2, ..., T$, and the stronger restriction $\widetilde{s}^2_v > 0$. Finally, in all cases we supposed the absence of a constant term and time dummies.

Tables 1-11 report the simulation results on the mean and standard deviations (SD) of the estimators. The tables differ with respect to the assumptions made about the distributions

of $\varepsilon_{i,t}$ and μ_i. Tables 1 and 2 correspond to design I-IE1, table 3 to design I-IE2, table 4 to design II-IE1, table 5 to design II-IE2, table 6 to design III-IE1, table 7 to design III-IE2, table 8 to design IV-IE1, table 9 to design IV-IE2, table 10 to design V-IE1, and table 11 corresponds to design V-IE2. In the tables the 2-step Optimal AB GMM estimator is labeled as ARBOND2.

Inspection of the results in tables 1-11 leads to the following conclusions: [10]

1. The AB GMM becomes severely biased and very imprecise when ρ approaches unity. The bias and variance of this estimator depend on the distribution of the $y_{i,1} - \mu_i$ as well as on the ratio of the variances of the error components.

2. When $\rho = 0.50$, the AB GMM estimator performs reasonably well when the $\varepsilon_{i,t}$ are not autocorrelated (designs I, II and III) but becomes substantially biased when the $\varepsilon_{i,t} \sim MA(1)$ (designs IV + V).

3. The REQMLE and FEQMLE display only a small bias and are more precise than the AB GMM estimator in almost all scenarios considered. The differences in precision between the QMLE and the AB GMM estimator increase considerably when ρ gets closer to unity.

4. When $\rho = 0.95$, the REQMLE and FEQMLE display a larger bias and are less precise in design II (weak moment conditions) than in design III (no weak moment conditions).

5. The System estimator (and the OLGMM estimator) perform(s) better than the QML estimators when all the moment conditions which are exploited by the former estimator(s) are valid; however, when mean-stationarity is violated (as in design C), then the System estimator can be substantially biased, whereas the QML estimators continue to perform well.

6. The differences in performance between the AB GMM and QML estimators and between the QML estimators and the System estimator are larger, when the variance of the initial conditions is smaller (as in design L).

[10] Some additional Monte Carlo evidence supporting conclusion 4, which is based on alternative designs, is available at http://alpha.qmul.ac.uk/~ugte185/

7. When the errors are heteroskedastic the REQMLE is more precise than the FEQMLE. On the other hand, the REQMLE has a larger bias than the FEQMLE when $\rho = 0.95$.

We have also constructed histograms for the AB GMM, SYS, REQML and FEQML estimators of ρ for designs I-IE1-S (weak moment conditions when ρ is close to unity) and III-IE1-S (no weak moment conditions for the QML estimators) and for $\rho = 0.50$ and $\rho = 0.95$. Each histogram is based on 10,000 simulations. When $\rho = 0.95$ the differences between the empirical distributions of the AB GMM estimator on the one hand and the other three estimators for ρ on the other hand are striking: the values of the first estimator are more dispersed and they are centered around a value well below 0.95. Furthermore, when the moment conditions are weak (design I-IE1-S and $\rho = 0.95$), the empirical distributions of the REMLE and FEMLE are different — the latter being bimodal — whereas in the other cases that we considered they are very similar.

We have also investigated the quality of some estimators for the standard errors of the QML estimators for ρ, and the bounds of 90% confidence intervals for ρ. We considered both traditional estimators based on first-order fixed parameter asymptotics and bootstrap estimators. The bootstrap estimators are based on reweighting the 'individual' contributions to the likelihood functions. We applied this bootstrap using 100 replications. When we used the bootstrap, we allowed for asymmetry of the CI's around the point estimates of ρ.

Tables 12 and 13 report the simulation results for designs I-IE1-S/C (weak instruments), while tables 14 and 15 report the simulation results for designs III-IE1-S/C (no weak instruments for the QML estimators). In the case of the CI's we have counted the number of times that the true value of ρ was outside the constructed CI's, that is, we have computed the rejection probabilities (RP). The results related to the AB GMM and SYS estimators have been included for comparison. For the latter estimators we have also reported corrected standard errors and CI's based on the method of Windmeijer (2005).

The findings can be summarized as follows:

1. Under design I the asymptotic standard errors, which are based on the 'sandwich formula' $H^{-1}GH^{-1}$, tend to underestimate the standard deviation of the REMLE. Furthermore when ρ is close to unity, the asymptotic standard errors tend to overestimate the standard deviation of the FEMLE. Under design III, the asymptotic standard

errors are nearly unbiased.

2. Under design I the bootstrap standard errors tend to underestimate the standard deviation of both the REMLE and the FEMLE when ρ is close to unity. This seems to be due to the weak instruments problem because in design III this problem disappears.

3. Under design I both the asymptotic CI's based on the MLE and the bootstrap CI's are too narrow when ρ is close to unity. Only the coverage probability $(= 1 - RP)$ of the bootstrap CI based on the REMLE, (84.6% when $\rho = 0.95$ and $N = 100$), is not very far from the nominal value of 90%. In the case of the FEMLE the asymptotic CI performs better than the bootstrap CI. The bootstrap CI based on the REMLE performs better than the asymptotic CI based on the FEMLE and the CI based on the AB GMM estimator and the corresponding Windmeijer corrected standard errors.

   Under design III the coverage probabilities of the bootstrap CI's based on the FEQMLE are close to the nominal value of 90%. In the case of the REQMLE the asymptotic CI performs better than the bootstrap CI. The bootstrap CI based on the FEQMLE performs better than the asymptotic CI based on the REQMLE and the CI based on the AB GMM estimator and the corresponding Windmeijer corrected standard errors.

Finally, we investigated the accuracy of the local-to-zero asymptotic distributions of the optimal AB GMM estimator and the FEMLE. According to theorem 4 the distributions of $\widehat{\rho}_{AB} - \rho$ are very similar for values of ρ that satisfy $\rho = 1 - \lambda N^{-0.5}$. Theorem 7 implies that when TSH (or TSH*) holds, the distributions of $N^{0.25}(\widehat{\rho}_{FEML} - \rho)$ are very similar for values of ρ that satisfy $\rho = 1 - \lambda N^{-0.25}$. Tables 16 and 17 report Monte Carlo results on the distributions of $\widehat{\rho}_{AB}$ and $\widehat{\rho}_{FEML}$ when ρ is close to unity. These results have been obtained for design I-IE1. Moreover, in all cases considered, i.e. for all values of ρ considered, the initial conditions have been drawn from the same distribution: $(y_{i,1} - \mu_i)|\mu_i \sim N(0, \sigma^2/[1-(0.9)^2])$. Noting that $4^{0.25} = 2^{0.5} \approx 0.7071$, it is easily seen from the tables that the biases and standard deviations of $\widehat{\rho}_{AB}$ and $\widehat{\rho}_{FEML}$ are in agreement with the aforementioned predictions of theorems 4 and 7.

7 Conclusions

In this paper we have shown that the Quasi ML estimation method yields a consistent estimator for the autoregression parameter ρ in the *conditional* panel AR(1) model (i.e. with arbitrary initial conditions) even when the errors are drawn from heterogenous distributions. We have compared both analytically and by means of Monte Carlo simulations the QML estimators with the GMM estimator proposed by Arellano and Bond, which ignores some of the moment conditions implied by the model. Unlike the AB GMM estimator, the QML estimators for ρ only suffer from a weak instruments problem when ρ is close to unity if the cross-sectional average of the variances of the errors is constant over time, e.g. under time-series homoskedasticity. However, even in this case the QML estimators are still consistent when ρ equals one and they display only a relatively small bias when ρ is close to one. On the other hand the AB GMM estimator is inconsistent when ρ equals one, and is severly biased when ρ is close to one. Moreover, our Monte Carlo results suggest that the local-to-zero asymptotic distributions of the AB GMM estimator and the FEQML estimator that we have derived in this paper give an accurate characterization of the distributional properties of these estimators when the moment condtions are weak. A panel unit root test based on the FEQML estimator is discussed in Kruiniger (2006b).

We have also examined the finite sample properties of two types of estimators for the standard errors of the QML estimators for ρ, and the bounds of QML based confidence intervals for ρ. The first estimator is based on first-order fixed parameter asymptotics while the second estimator is based on a simple bootstrap procedure. In a simulation study we found that the bootstrap CI based on the REQMLE performs better than alternative CI's when the moment conditions are weak, while the bootstrap CI based on the FEQMLE performs best when the moment conditions are not weak.

Finally, we have considered QML and GMM estimation of models that include exogenous regressors. We found that QML estimators are inconsistent when the model includes weakly exogenous regressors. Nonetheless we showed that for such models GMM estimators can be constructed which are not only closely related to the QML estimators but also consistent.

The results that we have obtained in this paper — in particular (the results related to) the QML based methods — should be useful when estimating dynamic panel data models with

persistent data, e.g. when estimating production functions with panel data (see Griliches and Mairesse, 1998, and Blundell and Bond, 2000). Firm data on factor inputs, especially data on capital stocks constructed by using the perpetual inventory method, tend to display a high degree of persistence.

A Proofs of the results in section 3

A.1 Proof of theorem 1 (Consistency of the RECGMM estimator for ρ in the conditional model when $-1 < \rho \leq 1$):

To prove consistency of the RECGMM estimator for ρ we need to verify that ρ is (uniquely) identified by the moment conditions in (12) and (13).

Tedious but straightforward algebra shows that the set of moment conditions given in (12) and (13) is equivalent to the following set:

$$\begin{aligned} E[m_{1,t}(\rho)] &= 0, \quad t = 3, ..., T, \quad \text{and} && (23) \\ E[m_{2,s,t}(\rho)] &= 0, \quad s = 2, ..., t-1, \quad \text{and} \quad t = 4, ..., T, && (24) \\ \text{where} \quad m_{1,t}(\rho) &= y_{i,1}(\Delta y_{i,t} - \rho \Delta y_{i,t-1}), \quad \text{and} \\ m_{2,s,t}(\rho) &= (y_{i,t} - \rho y_{i,t-1})(y_{i,s} - \rho y_{i,s-1}) - (y_{i,3} - \rho y_{i,2})(y_{i,2} - \rho y_{i,1}), \end{aligned}$$

It is clear and well-known that ρ is uniquely identified by $E[m_{1,t}(\rho)] = 0$, $t = 3, ..., T$, when $-1 < \rho < 1$. Therefore we focus on the unit root case, i.e. $\rho = 1$. Below we show that in this case ρ is uniquely identified by $E[m_{2,s,t}(\rho)] = 0, \quad s = 2, ..., t-1, \quad \text{and} \quad t = 4, ..., T$ if $T \geq 4$ unless $\overline{\sigma}^2_{i,s}/\overline{\sigma}^2_{i,s-1} = \overline{\sigma}^2_{i,3}/\overline{\sigma}^2_{i,2} \neq 1$, for all $s \in \{3, ..., T-1\}$.

When $\rho = 1$ the model in (1) reduces to $y_{i,t} = y_{i,t-1} + \varepsilon_{i,t}$ and we obtain the following results:

$m_{1,t}(r) = y_{i,1}(\Delta y_{i,t} - r\Delta y_{i,t-1}) = y_{i,1}(\Delta\varepsilon_{i,t} + (\rho - r)\varepsilon_{i,t-1})$ and $E[m_{1,t}(r)] = (\rho - r)E(y_{i,1}\varepsilon_{i,t-1}) = 0 \ \forall r \in (-1, 1]$. Therefore ρ is not identified by $E[m_{1,t}(\rho)] = 0$ when $\rho = 1$. In fact a similar result holds for all the moment conditions in (12).

$m_{2,s,t}(r) = (y_{i,t} - ry_{i,t-1})(y_{i,s} - ry_{i,s-1}) - (y_{i,3} - ry_{i,2})(y_{i,2} - ry_{i,1}) =$

$((\rho - r)y_{i,t-1} + \varepsilon_{i,t})((\rho - r)y_{i,s-1} + \varepsilon_{i,s}) - ((\rho - r)y_{i,2} + \varepsilon_{i,3})((\rho - r)y_{i,1} + \varepsilon_{i,2})$ and

$E[m_{2,s,t}(r)] = (\rho - r)^2 E(y_{i,s-1}^2 - y_{i,1}^2) + (\rho - r)(\sigma_{i,s}^2 - \sigma_{i,2}^2)$. There are two cases:

$s = 2 : E[m_{2,2,t}(r)] = 0 \; \forall r \in (-1, 1]$. Therefore ρ is not identified by $E[m_{2,2,t}(\rho)] = 0$ when $\rho = 1$.

$s > 2 : E[m_{2,s,t}(r)] = 0 \Rightarrow r = \rho \vee r = \rho + (E(y_{i,s-1}^2 - y_{i,1}^2))^{-1}(\sigma_{i,s}^2 - \sigma_{i,2}^2)$. Note that $E(y_{i,s-1}^2 - y_{i,1}^2) = \sum_{m=2}^{s-1} \sigma_{i,m}^2 > 0$. Therefore, when $\rho = 1$, ρ is uniquely identified by $E[m_{2,s,t}(\rho)] = 0$, $s = 3, ..., t-1$, and $t = 4, ..., T$ unless $(\sum_{m=2}^{s-1} \sigma_{i,m}^2)^{-1}(\sigma_{i,s}^2 - \sigma_{i,2}^2) = (\sigma_{i,2}^2)^{-1}(\sigma_{i,3}^2 - \sigma_{i,2}^2) \neq 0$ for all $s \in \{3, ..., T-1\}$ in which case $r = 1 \vee r = \sigma_{i,3}^2/\sigma_{i,2}^2$. It can be shown by induction that the latter condition is satisfied if and only if $\sigma_{i,s}^2/\sigma_{i,s-1}^2 = \sigma_{i,3}^2/\sigma_{i,2}^2 \neq 1$, for all $s \in \{3, ..., T-1\}$. When $T = 4$, the proof of this claim is trivial. Let $\sigma_{i,3}^2/\sigma_{i,2}^2 = c$. When $T = 5$ and $s = 4$, $(\sum_{m=2}^{3} \sigma_{i,m}^2)^{-1}(\sigma_{i,4}^2 - \sigma_{i,2}^2) = (\sigma_{i,2}^2)^{-1}(\sigma_{i,3}^2 - \sigma_{i,2}^2) \Leftrightarrow (\sum_{m=2}^{3} \sigma_{i,m}^2/\sigma_{i,2}^2)^{-1}(\sum_{m=3}^{4} \sigma_{i,m}^2/\sigma_{i,2}^2) = \sigma_{i,3}^2/\sigma_{i,2}^2 \Leftrightarrow \sigma_{i,4}^2/\sigma_{i,2}^2 = c^2 \Leftrightarrow \sigma_{i,4}^2/\sigma_{i,3}^2 = c$. Suppose now that the assertion holds for $T = n \geq 5$. Then it also holds for $T = n + 1$: the only new case is $s = n$ and $(\sum_{m=2}^{n-1} \sigma_{i,m}^2)^{-1}(\sigma_{i,n}^2 - \sigma_{i,2}^2) = (\sigma_{i,2}^2)^{-1}(\sigma_{i,3}^2 - \sigma_{i,2}^2) \Leftrightarrow (\sum_{m=2}^{n-1} \sigma_{i,m}^2/\sigma_{i,2}^2)^{-1}(\sum_{m=3}^{n} \sigma_{i,m}^2/\sigma_{i,2}^2) = c \Leftrightarrow (\sum_{m=2}^{n-1} c^{m-2})^{-1}(\sigma_{i,n}^2/\sigma_{i,2}^2 + \sum_{m=3}^{n-1} c^{m-2}) = c \Leftrightarrow \sigma_{i,n}^2/\sigma_{i,2}^2 = c^{n-2} \Leftrightarrow \sigma_{i,n}^2/\sigma_{i,n-1}^2 = c$.

The above implies that if $\rho = 1$ and $T \geq 4$, then ρ is uniquely identified by the moment conditions in (24) [or by those in (13)] unless $\overline{\sigma}_s^2/\overline{\sigma}_{s-1}^2 = \overline{\sigma}_3^2/\overline{\sigma}_2^2 \neq 1$ for all $s \in \{3, ..., T-1\}$. If $\rho = 1$ and the average variance $\overline{\sigma}_t^2$ changes at a constant rate between $t = 2$ and $t = T-1 \geq 3$, then ρ is only locally identified: $r = 1$ or $r = \overline{\sigma}_3^2/\overline{\sigma}_2^2$. However, ρ is potentially not uniquely identified only when $\rho = 1$: when $\rho \neq 1$ and $T \geq 3$, ρ is always uniquely identified. Therefore, when ρ is not uniquely identified, we know that ρ must be equal to one. We conclude that if $T \geq 4$ the RECGMM estimator for ρ is consistent even when $\rho = 1$. □

IDENTIFICATION BY THE 'HOMOSKEDASTICITY' MOMENT CONDITIONS IN (15):

Let $m_{3,t}(\rho) = (y_{i,t} - \rho y_{i,t-1})^2 - (y_{i,2} - \rho y_{i,1})^2$, $t = 3, ..., T$. If $\rho = 1$ and assumption TSH holds with $\sigma_i^2 > 0$, then for each $t \in \{3, ..., T\}$ we have $m_{3,t}(r) = ((\rho - r)y_{i,t-1} + \varepsilon_{i,t})^2 - ((\rho - r)y_{i,1} + \varepsilon_{i,2})^2$, $E[m_{3,t}(r)] = (\rho - r)^2 E(y_{i,t-1}^2 - y_{i,1}^2) = (\rho - r)^2(t-2)\sigma_i^2$, and $E[m_{3,t}(r)] = 0 \Rightarrow r = \rho$.

A.2 Proof of theorem 2 (Consistency of the RE and FE (Q)MLE's):

We first prove consistency of the FEQMLE for ρ.

The FE Quasi MLE for ρ in the conditional panel AR(1) model is based on the quasi likelihood function corresponding to the following auxiliary model

$$y_i - y_{i,1}\iota = \rho(y_{i,-1} - y_{i,1}\iota) + u_i, \tag{25}$$

where $u_i = -(1-\rho)v_{i,1}\iota + \varepsilon_i \sim N(0, \Phi)$, with $\Phi = \Phi(\varphi) = \widetilde{\sigma}_v^2 \iota\iota' + \Psi$, where φ is the vector comprising all (co-)variance parameters.

We have assumed that $-1 < \rho \leq 1$, and that assumptions SA, RFEA and B hold. To simplify matters we also assume that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \varepsilon_i\varepsilon_i' = \Psi$.

Let $\widetilde{\Delta}y_i = y_i - y_{i,1}\iota$ and $\widetilde{\Delta}y_{i,-1} = y_{i,-1} - y_{i,1}\iota$. Then the quasi log-likelihood function for the conditional panel AR(1) model with fixed effects is given by

$$\begin{aligned} \log L_F(r,F) &= -\frac{1}{2}N(T-1)\log 2\pi - \frac{N}{2}\log|F| \\ &\quad -\frac{1}{2}\sum_{i=1}^N (\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1})' F^{-1} (\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1}). \end{aligned} \tag{26}$$

To prove consistency of the FEQMLE for ρ, we will verify the conditions of theorem 4.1.1. in Amemiya (1985).

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

$$\widetilde{\Delta}y_{i,-1} = Pu_i = P\iota(\rho-1)v_{i,1} + P\varepsilon_i, \tag{27}$$

where

$$P = P(\rho) = \begin{pmatrix} 0 & . & . & 0 & 0 & 0 \\ 1 & 0 & & & 0 & 0 \\ \rho & 1 & 0 & & & 0 \\ . & \rho & 1 & 0 & & . \\ . & & \rho & 1 & 0 & . \\ \rho^{T-3} & . & . & \rho & 1 & 0 \end{pmatrix}. \tag{28}$$

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

$$\begin{aligned} &\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1} = (\rho - r)\widetilde{\Delta}y_{i,-1} + u_i = [(\rho-r)P + I]u_i = \\ &[(\rho-r)P\iota + \iota](\rho-1)v_{i,1} + [(\rho-r)P + I]\varepsilon_i. \end{aligned} \tag{29}$$

It can easily be seen from (29) that the quasi log-likelihood function, divided by N, converges uniformly in probability if our assumptions hold.

We note that one would obtain the same probability limit of the quasi log-likelihood function if the $\widetilde{v}_{i,1} = (\rho-1)v_{i,1}$ and ε_i were i.i.d. and Gaussian with $E(\widetilde{v}_{i,1}^2) = \widetilde{\sigma}_v^2$, $E(\widetilde{v}_{i,1}\varepsilon_i) = 0$, and $E(\varepsilon_i\varepsilon_i') = \Psi$. In that case we would have a standard ML estimation problem based on an error-component model. Identification of ρ (of the parameters) is discussed below. It immediately follows that the probability limit of the quasi log-likelihood function attains a maximum at the true value(s) of ρ (of the parameters).

Consistency of the REQMLE for ρ in the conditional panel AR(1) model can be proved along similar lines. However, instead of (29), one should use

$$\begin{aligned} &y_i - ry_{i,-1} - p(1-r)y_{i,1}\iota = \\ &[(\rho-r)P + I]u_i + (\rho-r)P\iota(1-\rho)\pi y_{i,1} + [\pi(1-\rho) - p(1-r)]y_{i,1}\iota. \end{aligned} \tag{30}$$

□

Identification of the parameters in (26):

To show that the parameters ρ and φ are identified when $-1 < \rho \leq 1$ one can proceed as follows:

Let $L_F(\widetilde{\Delta}y_i|\rho,\varphi)$ be the pdf of $\widetilde{\Delta}y_i$.

From (29) we obtain $(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1})'F^{-1}(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1}) = u_i'[(\rho-r)P+I]'F^{-1}[(\rho-r)P+I]u_i$.

Note that F^{-1} is PDS when $s_t^2 > 0$ for $t \geq 2$, and that $[(\rho-r)P+I]$ is nonsingular for any $-1 < \rho$, $r \leq 1$. Therefore $[(\rho-r)P+I]'F^{-1}[(\rho-r)P+I]$ is PDS.

Moreover, given the specific structure of F^{-1} and P, $[(\rho-r)P+I]'F^{-1}[(\rho-r)P+I] = \Phi^{-1}$ if and only if $f = \varphi$ and $r = \rho$.

It follows that $\Pr(\widetilde{\Delta}y_i : L_F(\widetilde{\Delta}y_i|r,F) \neq L_F(\widetilde{\Delta}y_i|\rho,\Phi)) = 1$ if $r \neq \rho$ and $f \neq \varphi$.

As an example, identification of ρ and φ can easily be verified for the homoskedastic case where $F = \widetilde{s}_v^2\iota\iota' + s^2 I$ with $s^2 > 0$ and $T \geq 3$. It follows from well-known results in the panel data literature that in this case $F^{-1} = s^{-2}Q + [s^2 + (T-1)\widetilde{s}_v^2]^{-1}\frac{1}{T-1}\iota\iota'$ with $Q = I_{T-1} - \frac{1}{T-1}\iota\iota'$.

A.3 Proof of theorem 3 (Limiting distribution of the FEMLE):

We consider the simple case where $E(u_i u_i') = \Phi = \widetilde{\sigma}_v^2 \iota\iota' + \sigma^2 I$. After noting that $\Phi = \sigma^2 Q + (\sigma^2 + (T-1)\widetilde{\sigma}_v^2)\frac{1}{T-1}\iota\iota'$, where $Q = I_{T-1} - \frac{1}{T-1}\iota\iota'$, it follows from some classical matrix algebra results in the panel data literature that $\Phi^{-1} = \sigma^{-2}Q + \widetilde{\sigma}^{-2}\frac{1}{T-1}\iota\iota'$ and $|\Phi| = \sigma^{2(T-2)}\widetilde{\sigma}^2$ where $\widetilde{\sigma}^2 = \sigma^2 + (T-1)\widetilde{\sigma}_v^2$ (see e.g. Hsiao, 1986). These results allow us to write the log-likelihood function of the FEMLE as:

$$\begin{aligned}\log L[(r\ s^2\ \widetilde{s}_v^2)'] &= -\frac{1}{2}N(T-1)\log 2\pi - \frac{N(T-2)}{2}\log s^2 - \frac{N}{2}\log(s^2 + (T-1)\widetilde{s}_v^2) \\ &\quad -\frac{1}{2s^2}\sum_{i=1}^{N}(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1})'Q(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1}) \qquad (31) \\ &\quad -\frac{1}{2(s^2 + (T-1)\widetilde{s}_v^2)}\frac{1}{T-1}\sum_{i=1}^{N}[\iota'(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1})]^2.\end{aligned}$$

The Fixed Effects (Quasi) ML estimator is defined as the global maximizer of the (quasi) log-likelihood function given in (31). The corresponding likelihood equations for ρ, σ^2 and $\widetilde{\sigma}_v^2$ are given by:

$$\frac{\partial \log L}{\partial r} = \sum_{i=1}^{N}\widetilde{\Delta}y_{i,-1}'F^{-1}(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1}) = 0, \qquad (32)$$

$$\begin{aligned}\frac{\partial \log L}{\partial s^2} &= -\frac{N(T-2)}{2s^2} - \frac{N}{2\widetilde{s}^2} + \frac{1}{2s^4}\sum_{i=1}^{N}(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1})'Q(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1}) \\ &\quad +\frac{1}{2\widetilde{s}^4}\frac{1}{T-1}\sum_{i=1}^{N}[\iota'(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1})]^2 = 0, \qquad (33)\end{aligned}$$

and

$$\frac{\partial \log L}{\partial \widetilde{s}_v^2} = -\frac{N(T-1)}{2\widetilde{s}^2} + \frac{1}{2\widetilde{s}^4}\sum_{i=1}^{N}[\iota'(\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1})]^2 = 0, \qquad (34)$$

where $F^{-1} = s^{-2}Q + \widetilde{s}^{-2}\frac{1}{T-1}\iota\iota'$ and $\widetilde{s}^2 = s^2 + (T-1)\widetilde{s}_v^2$.

To derive the limiting variance of the FEMLE, we will make use of the function $\xi(r) = \frac{1}{T-1}\sum_{t=1}^{T-2}\frac{T-1-t}{t}r^t$ and the fact that $\widetilde{\Delta}y_{i,-1} = Pu_i = P\iota(\rho-1)(y_{i,1} - \mu_i) + P\varepsilon_i$, cf (27). Then the following results are easily obtained:

$E[(\widetilde{\Delta}y_{i,-1})'\Phi^{-1}(\widetilde{\Delta}y_{i,-1})] = \sigma^2 tr(P'\Phi^{-1}P) + \widetilde{\sigma}_v^2\iota'P'\Phi^{-1}P\iota,$

$E([\iota'(\widetilde{\Delta}y_i - \rho\widetilde{\Delta}y_{i,-1})]^2) = (T-1)\widetilde{\sigma}^2$,

$E[(\widetilde{\Delta}y_i - \rho\widetilde{\Delta}y_{i,-1})'Q(\widetilde{\Delta}y_{i,-1})] = E[\varepsilon_i'QPu_i] = \sigma^2 tr(QP) = -\sigma^2\xi'(\rho)$, and

$E[\iota'(\widetilde{\Delta}y_{i,-1})(\widetilde{\Delta}y_i - \rho\widetilde{\Delta}y_{i,-1})'\iota] = E[\iota'Pu_iu_i'\iota] = [(T-1)\widetilde{\sigma}_v^2 + \sigma^2]\iota'P\iota = \widetilde{\sigma}^2\iota'P\iota =$

$\widetilde{\sigma}^2(T-1)\xi'(\rho)$.

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

$$[FEH]^{-1} = N[E(-\frac{\partial^2 \log L(\delta)}{\partial\delta\partial\delta'})]^{-1} = \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{12} & \frac{T-2}{2\sigma^4} + \frac{1}{2\widetilde{\sigma}^4} & \frac{T-1}{2\widetilde{\sigma}^4} \\ V_{13} & \frac{T-1}{2\widetilde{\sigma}^4} & \frac{(T-1)^2}{2\widetilde{\sigma}^4} \end{bmatrix}^{-1}, \tag{35}$$

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

When $|\rho| < 1$, $\sqrt{N}(\widehat{\delta}_{FEML} - \delta) \xrightarrow{d} N\left(0, [FEH]^{-1}\right)$.

When $\rho = 1$, $\widetilde{\sigma}_v^2 = 0$, $\widetilde{\sigma}^2 = \sigma^2$ and $\Phi = \sigma^2 I$. Furthermore $tr(P'P) = \frac{1}{2}(T-1)(T-2)$ and $\xi'(1) = \frac{1}{2}(T-2)$. It follows that when $\rho = 1$, FEH is a singular matrix. In this case the FEMLE for δ has a non-normal asymptotic distribution.

B Proofs of the results in section 4

B.1 Proof of theorem 5 (weak moment conditions GMM):

Let $E[m_{AB,s,t}(\rho)] = 0$ with $m_{AB,s,t}(\rho) = y_{i,s}(\Delta y_{i,t} - \rho \Delta y_{i,t-1})$ (where $s \leq t-2$) represent an arbitrary Arellano-Bond moment condition from (12), let $E[m_{AS,t}(\rho)] = 0$ with $m_{AS,t}(\rho) = (y_{i,T} - \rho y_{i,T-1})(\Delta y_{i,t-1} - \rho \Delta y_{i,t-2})$ represent an arbitrary Ahn-Schmidt moment condition from (13), and let $m_{HO,t}(\rho) = (y_{i,t} - \rho y_{i,t-1})^2 - (y_{i,2} - \rho y_{i,1})^2$ represent an arbitrary 'homoskedasticity' moment condition from (15). Then we obtain the following results:

$\frac{dm_{AB,s,t}}{d\rho} = -y_{i,s}\Delta y_{i,t-1}$ and $E(-y_{i,s}\Delta y_{i,t-1}) = (1-\rho)E[(y_{i,s} - \mu_i + \mu_i)(y_{i,t-2} - \mu_i)] = (1-\rho)\rho^{t-2-s}E[(y_{i,s} - \mu_i + \mu_i)(y_{i,s} - \mu_i)]$. It follows that $\lim_{\rho \uparrow 1} E(\frac{dm_{AB,s,t}}{d\rho}) = 0$;

$\frac{dm_{HO,t}}{d\rho} = -2y_{i,t-1}(y_{i,t} - \rho y_{i,t-1}) + 2y_{i,1}(y_{i,2} - \rho y_{i,1})$ and $E(\frac{dm_{HO,t}}{d\rho}|y_{i,1}, \mu_i) = E[-2y_{i,t-1}(1-\rho)\mu_i + 2y_{i,1}(1-\rho)\mu_i|y_{i,1}, \mu_i] = 2(1-\rho)(1-\rho^{t-2})\mu_i(y_{i,1} - \mu_i)$. Hence $\lim_{\rho \uparrow 1} E(\frac{dm_{HO,t}}{d\rho}) = 0$;

$\frac{dm_{AS,t}(\rho)}{d\rho} = -y_{i,T-1}\Delta\varepsilon_{i,t-1} - (\varepsilon_{i,T} + (1-\rho)\mu_i)\Delta y_{i,t-2}$ and $E(\frac{dm_{AS,t}(\rho)}{d\rho}) = -E[(y_{i,T-1} - \mu_i)\Delta\varepsilon_{i,t-1}] + (1-\rho)^2 E[\mu_i(y_{i,t-3} - \mu_i)] = -\rho^{T-t}E[(y_{i,t-1} - \mu_i)\Delta\varepsilon_{i,t-1}] + (1-\rho)^2 E[\mu_i(y_{i,t-3} - \mu_i)] = -\rho^{T-t}(\sigma^2_{i,t-1} - \rho\sigma^2_{i,t-2}) + (1-\rho)^2 E[\mu_i(y_{i,t-3} - \mu_i)]$. It follows that $\lim_{\rho \uparrow 1} E(\frac{dm_{AS,t}}{d\rho}) = -\sigma^2_{i,t-1} + \sigma^2_{i,t-2} = 0$ if and only if $\sigma^2_{i,t-1} = \sigma^2_{i,t-2}$.

We conclude that $E[m_{AB,s,t}(\rho)] = 0$ and $E[m_{HO,t}(\rho)] = 0$ are weak when ρ is close to one. However, $E[m_{AS,t}(\rho)] = 0$ is weak when ρ is close to one if and only if $\overline{\sigma}^2_{t-1} = \overline{\sigma}^2_{t-2}$. We remark that $E[m_{AB,s,t}(\rho)] = 0$, $E[m_{AS,t}(\rho)] = 0$ and $E[m_{HO,t}(\rho)] = 0$ are also weak under covariance stationarity (in the case of $E[m_{AB,s,t}(\rho)] = 0$, see Kruiniger (2006a)). □

B.2 Proof of theorem 6 (weak moment conditions QML):

There are two cases that we must consider. In the first case assumption TSH* does not hold while in the second case assumption TSH* does hold. In the former case we need to show that the Expected Hessian of the QMLE is nonsingular when $\rho = 1$, whereas in the latter case we need to show that the Expected Hessian of the QMLE is singular when $\rho = 1$ irrespective of whether TSH is embodied in the estimator or not.

We prove the theorem for the REQMLE's. In fact it is sufficient to consider the REMLE's only. Thus we assume that the ε_i and v_i are i.i.d. and Gaussian.

As noted at the end of section 3, when the ε_i and v_i are i.i.d. and Gaussian and assumption TSH has not been imposed on the estimator, the REMLE for ρ is asymptotically equivalent to the ORECGMM estimator for ρ. In fact, one can show that the REMLE for all the parameters of the conditional model is asymptotically equivalent to an Optimal GMM estimator that exploits all the second moment conditions (i.e. those involving all the elements of $E[(y_{i,1}\ y_i)(y_{i,1}\ y_i)']$) that are implied by this model. It follows that the Expected Hessian of the REMLE is equal to the inverse of the limiting variance of such an Optimal GMM estimator. Establishing (non-)singularity of the former matrix when $\rho = 1$ is therefore equivalent to proving (non-)singularity of the latter matrix when $\rho = 1$. The formula for the latter is given by $G(\theta_0)'(\Omega(\theta_0))^{-1}G(\theta_0)$, where $G(\theta_0) = E(dm(\theta)/d\theta|\theta_0)$ is the first derivative of the vector of moment conditions $E(m(\theta)) = 0$ with respect to the vector of parameters θ evaluated at the true values of the parameters, θ_0, and $\Omega(\theta_0) = E(m(\theta_0)m(\theta_0)')$ is the optimal weight matrix.

We first prove that $G(\theta_0)'(\Omega(\theta_0))^{-1}G(\theta_0)$ is nonsingular when $\rho = 1$ and TSH* does not hold. In this case consistency of the REMLE requires that $T \geq 4$ (cf theorem 1).

Let $T = 4$. Then $\theta_0 = (\sigma_{\eta y}\ \sigma_y^2\ \rho\ \sigma_\eta^2\ \sigma_2^2\ \sigma_3^2\ \sigma_4^2)'$ and the vector of all the $0.5T(T+1)$ available moment conditions $E(m(\theta_0)) = 0$ is defined by the $10-$dimensional vector:

$$m(\theta_0) = m_4(\theta_0) = \begin{bmatrix} y_{i,1}\Delta w_{i,3} \\ y_{i,1}\Delta w_{i,4} \\ y_{i,2}\Delta w_{i,4} \\ y_{i,1}\Delta y_{i,2} - \sigma_{\eta y} - (\rho-1)\sigma_y^2 \\ y_{i,1}^2 - \sigma_y^2 \\ w_{i,4}\Delta w_{i,3} \\ w_{i,2}w_{i,3} - \sigma_\eta^2 \\ w_{i,2}^2 - \sigma_\eta^2 - \sigma_2^2 \\ w_{i,3}^2 - \sigma_\eta^2 - \sigma_3^2 \\ w_{i,4}^2 - \sigma_\eta^2 - \sigma_4^2 \end{bmatrix}. \tag{36}$$

We can write $m(\theta_0) = (m_I'(\theta_0),\ m_{II}'(\theta_0))'$ where $\dim(m_{II}'(\theta_0)) = \dim(\theta) = 7$. It can easily be seen that $\det(E(dm_{II}(\theta)/d\theta|\theta_0)) \neq 0$ when $\rho = 1$ and TSH* does not hold. It follows immediately that $G(\theta_0)$ has full rank when $\rho = 1$ and TSH* does not hold.

We will now prove that $\Omega(\theta_0)$ is PDS when $\rho = 1$. Notice that in this case $\eta_i = 0$, $\Delta y_{i,2} = \varepsilon_{i,2}$ and $w_{i,t} = \varepsilon_{i,t}$, $t = 2, 3, 4$.

It is useful to define $m_y(\theta_0) = (y_{i,1}^2 - \sigma_y^2,\ \ y_{i,1}\Delta y_{i,2} - \sigma_{\eta y} - (\rho-1)\sigma_y^2,\ \ y_{i,1}\Delta w_{i,3},\ \ y_{i,1}\Delta w_{i,4})'$ and $m_{ny+}(\theta_0) = (w_{i,2}w_{i,3} - \sigma_\eta^2,\ \ m'_{ny}(\theta_0))'$, where the vector $m_{ny}(\theta_0)$ contains the remaining moment functions from $m(\theta_0)$ including $y_{i,2}\Delta w_{i,4}$. Notice that when $\rho = 1$ all the elements of $m_{ny+}(\theta_0)$ involve products of $\varepsilon_{i,s}$ and $\varepsilon_{i,t}$, with $s, t \in \{2, ..., T\}$, whereas the elements of $m_y(\theta_0)$ only involve products of $y_{i,1}$ and a second factor, namely $y_{i,1}$ or $\varepsilon_{i,t}$ where $t \in \{2, ..., T\}$. Notice also that when $\rho = 1$ the covariance matrix of $y_{i,2}\Delta w_{i,4}$ $(= y_{i,1}\Delta\varepsilon_{i,4} + \varepsilon_{i,2}\Delta\varepsilon_{i,4})$, $w_{i,4}\Delta w_{i,3}$ $(= \varepsilon_{i,4}\Delta\varepsilon_{i,3})$, and $w_{i,2}w_{i,3}$ $(= \varepsilon_{i,2}\varepsilon_{i,3})$ is nonsingular. It is then easily seen that both $E(m_y(\theta_0)m_y(\theta_0)')$ and $E(m_{ny+}(\theta_0)m_{ny+}(\theta_0)')$ are nonsingular when $\rho = 1$. From the preceding observations it follows immediately that $\Omega(\theta_0)$ is PDS when $\rho = 1$. We conclude that $G(\theta_0)'(\Omega(\theta_0))^{-1}G(\theta_0)$ is PDS when $\rho = 1$ and TSH* does not hold.

The above argument for showing that $\Omega(\theta_0)$ is nonsingular when $T = 4$ can easily be extended to the general case where $T \geq 4$. To that end we redefine the vector $m_y(\theta_0)$ by adding the elements $y_{i,1}\Delta w_{i,t}$, $t = 5, 6, ..., T$. All the other additional moment functions that are available from (12), (13) and (15) are included in $m_{ny}(\theta_0)$ and thereby in $m_{ny+}(\theta_0)$. Again it is easily seen that both $E(m_y(\theta_0)m_y(\theta_0)')$ and $E(m_{ny+}(\theta_0)m_{ny+}(\theta_0)')$ are nonsingular when $\rho = 1$. It follows again straightforwardly that $\Omega(\theta_0)$ is nonsingular.

When $T > 4$, we can formulate the vector of additional moment conditions $E(m_{ad}(\theta_0)) = 0$ in such a way that $E(m_4(\theta_0)m_{ad}(\theta_0)') = 0$. To see this, let the original set of additional moment conditions be given by $E(m_{in}(\theta_0)) = 0$, let $\Omega_{in,4}(\theta_0) = E(m_{in}(\theta_0)m_4(\theta_0)')$ and let $\Omega_4(\theta_0) = E(m_4(\theta_0)m_4(\theta_0)')$. Next define $m_{ad}(\theta_0) = m_{in}(\theta_0) - \Omega_{in,4}(\theta_0)(\Omega_4(\theta_0))^{-1}m_4(\theta_0)$. Then $E(m_4(\theta_0)m_{ad}(\theta_0)') = 0$.

Now let $G_4(\theta_0) = E(dm_4(\theta)/d\theta|\theta_0)$, let $G_{ad}(\theta_0) = E(dm_{ad}(\theta)/d\theta|\theta_0)$ and let $\Omega_{ad}(\theta_0) = E(m_{ad}(\theta_0)m_{ad}(\theta_0)')$. Then it follows from $E(m_4(\theta_0)m_{ad}(\theta_0)') = 0$ that $G(\theta_0)'(\Omega(\theta_0))^{-1}G(\theta_0) = G_4(\theta_0)'(\Omega_4(\theta_0))^{-1}G_4(\theta_0) + G_{ad}(\theta_0)'(\Omega_{ad}(\theta_0))^{-1}G_{ad}(\theta_0)$. Since $G_4(\theta_0)'(\Omega_4(\theta_0))^{-1}G_4(\theta_0)$ is PDS and $G_{ad}(\theta_0)'(\Omega_{ad}(\theta_0))^{-1}G_{ad}(\theta_0)$ is PSDS, it follows that $G(\theta_0)'(\Omega(\theta_0))^{-1}G(\theta_0)$ is PDS and therefore nonsingular. This completes the proof of the first case.

We will now consider the 'homoskedastic' case, i.e. the case where assumption TSH* does hold. We will prove that the Expected Hessian of the REMLE that embodies assumption TSH is singular when $\rho = 1$ and TSH* holds. We can prove a similar claim for the REMLE that does not embody TSH along similar lines. Note that when $\rho = 1$ the former REMLE

is consistent for $T \geq 3$, while the latter REMLE is only consistent for $T \geq 4$ (cf theorem 1).

Let us first define $E(m(\theta_0)) = 0$ for general T. This vector consists of all the moment conditions in (12) and (13), $E(w_{i,t}^2 - \sigma_\eta^2 - \sigma^2) = 0$, $t = 2, ..., T$, and the three remaining elements of $E(m_4(\theta_0)) = 0$. It is useful to replace $E(w_{i,t}^2 - \sigma_\eta^2 - \sigma^2) = 0$, $t = 2, ..., T$ in $E(m(\theta_0)) = 0$ by $E(w_{i,t}^2 - w_{i,t-1}^2) = 0$, $t = 3, ..., T$ (i.e. the moment conditions in (15)) and $E(w_{i,2}^2 - \sigma_\eta^2 - \sigma^2) = 0$. Now, when $\rho = 1$ and TSH* holds, $rank(G(\theta_0)) < \dim(\theta_0) = 5$ for any $T \geq 3$. This follows from the fact that the four columns of $G(\theta_0)$ corresponding to ρ, $\sigma_{\eta y}$, σ_η^2, and σ^2 are linearly dependent when $\rho = 1$. These columns consist of zeros except in the three rows corresponding to $E(y_{i,1}\Delta y_{i,2} - \sigma_{\eta y} - (\rho - 1)\sigma_y^2) = 0$, $E(w_{i,2}w_{i,3} - \sigma_\eta^2) = 0$, and $E(w_{i,2}^2 - \sigma_\eta^2 - \sigma^2) = 0$, because these are the only moment conditions that involve $\sigma_{\eta y}$, σ_η^2, and σ^2, and because theorem 5 implies that all the other moment conditions involving ρ, i.e. (12) and (13) and (15), are weak when TSH* holds and $\rho = 1$. Since $G(\theta_0)$ does not have full rank, $G(\theta_0)'(\Omega(\theta_0))^{-1}G(\theta_0)$ is singular and hence the Expected Hessian of the REMLE that embodies assumption TSH is singular when $\rho = 1$ and TSH* holds. □

B.3 Proof of theorem 7 (local-to-zero asymptotics FEMLE):

We first prove two lemmata. Some of the notation used below is defined in appendix A.3.

Lemma 8 *Let $T \geq 3$ and $f(\rho) = (T-2)tr(P'P) - 2(T-1)(\xi'(\rho))^2$, where $P = P(\rho)$ is given in (28). Then $f(1) = 0$, $f'(1) = 0$, and $f''(1) = (1/18))T(T-1)(T-1)^2(T-3)$.*

Proof of lemma 8:

Note that $tr(P'P) = \sum_{t=1}^{T-2} t\rho^{2(T-2-t)}$ and $(T-1)\xi'(\rho) = \sum_{t=1}^{T-2} t\rho^{(T-2-t)}$. Furthermore, note that $\sum_{t=1}^{T} t = (1/2)T(T+1)$, $\sum_{t=1}^{T} t^2 = (1/6)T(T+1)(2T+1)$, and $\sum_{t=1}^{T} t^3 = ((1/2)T(T+1))^2$. Then the claims of lemma 8 follow straightforwardly. □

Lemma 9 *If $-1 < \rho \leq 1$ and $T \geq 4$, then $\det(FEH(\delta(\rho))) = (1-\rho)^2 c(\rho)$ where $c(\rho)$ is a polynomial in ρ with $c(1) \neq 0$.*

Proof of lemma 9:

Note that $\Phi^{-1} = \sigma^{-2}\widetilde{\sigma}^{-2}(\sigma^2 I + (T-1)\widetilde{\sigma}_v^2 Q)$. Hence $V_{11} = \sigma^2 tr(P'\Phi^{-1}P) + \widetilde{\sigma}_v^2 \iota' P'\Phi^{-1}P\iota =$

$\sigma^{-2}\widetilde{\sigma}^{-2}\left(\sigma^4 h_1(\rho)+\sigma^2\widetilde{\sigma}_v^2(T-1)h_2(\rho)+\sigma^2\widetilde{\sigma}_v^2 h_3(\rho)+\widetilde{\sigma}_v^4(T-1)h_4(\rho)\right)$, where $h_1(\rho)=tr(P'P)$, $h_2(\rho)=tr(P'QP)$, $h_3(\rho)=\iota'P'P\iota$, and $h_4(\rho)=\iota'P'QP\iota$.

It follows that $\det(FEH(\delta(\rho)))=\left(4\sigma^6\widetilde{\sigma}^6\right)^{-1}(T-1)^2\left(\sigma^4 q_1(\rho)+\sigma^2\widetilde{\sigma}_v^2 q_2(\rho)+\widetilde{\sigma}_v^4 q_3(\rho)\right)$, where $q_1(\rho)=(T-2)h_1(\rho)-2(T-1)(\xi'(\rho))^2$, $q_2(\rho)=(T-2)(T-1)h_2(\rho)-2(T-1)^2(\xi'(\rho))^2+(T-2)h_3(\rho)$, and $q_3(\rho)=(T-2)(T-1)h_4(\rho)$.

Recall that $\widetilde{\sigma}_v^2=\widetilde{\sigma}_v^2(\rho)=(1-\rho)^2\sigma_v^2$. From lemma 8 it follows that if $T\geq 4$ then $q_1(\rho)=f(\rho)=(1-\rho)^2\widetilde{q}_1(\rho)$ where $\widetilde{q}_1(\rho)$ is a polynomial in ρ with $\widetilde{q}_1(1)\neq 0$. Hence $\det(FEH(\delta(\rho)))=(1-\rho)^2c(\rho)$ where $c(\rho)$ is a polynomial in ρ with $c(1)\neq 0$. By the way, if $T=3$ then $q_1(\rho)=q_2(\rho)=0$ and $q_3(\rho)=1/2$ so that $\det(FEH(\delta(\rho)))=(2\sigma^6\widetilde{\sigma}^6)^{-1}\widetilde{\sigma}_v^4=(2\sigma^6\widetilde{\sigma}^6)^{-1}(1-\rho)^4\sigma_v^4$. $\square$

Proof of theorem 7:

A Taylor expansion of $N^{-1}\frac{\partial \log L_F(\delta)}{\partial\delta}$ around $\widehat{\delta}_{FEML}$ yields $(\widehat{\delta}_{FEML}-\delta)=(FH(\check{\delta}))^{-1}\times N^{-1}\frac{\partial \log L_F(\delta)}{\partial\delta}$, where $\check{\delta}_k=\mu_k\widehat{\delta}_{FEML,k}+(1-\mu_k)\delta_k$ with $\mu_k\in[0,1]$ for $k=1,2,3$. Note that $(FH(\check{\delta}))^{-1}=(\det(FH(\check{\delta})))^{-1}\overline{FH}(\check{\delta})$.

Lemma 9 implies that $\lim_{\rho\uparrow 1}c_{\det}(\rho)=c_{\det}(1)\neq 0$. Moreover when $-1<\rho\leq 1$ we have $\det(FEH(\delta(\rho)))\geq 0$ and hence $c_{\det}(\rho)\geq 0$. It can also easily be verified that $\overline{FEH}_{11}(\delta(1))=(1/4)\sigma^{-8}(T-2)(T-1)^2\neq 0$ and $\overline{FEH}_{11}(\delta(\rho))\geq 0$ when $-1<\rho\leq 1$.

Note that under fixed parameter first-order asymptotics we have $\widehat{\delta}_{FEML}-\delta\xrightarrow{p}0$ and when $|\rho|<1$ we also have $N^{0.5}(\widehat{\delta}_{FEML}-\delta)\xrightarrow{d}N\left(0,(FEH)^{-1}\right)$ (cf appendix A.3).

Let $\rho=1-\lambda N^{-0.25}$ with $\lambda>0$. Then since 0.25 is less than the rate of convergence of $\widehat{\delta}_{FEML}$ under fixed parameter asymptotics (i.e. 0.5), we still have $\widehat{\delta}_{FEML}-\delta\xrightarrow{p}0$ and hence $FH(\check{\delta})\xrightarrow{p}FEH(\delta(1))$ and $\overline{FH}(\check{\delta})\xrightarrow{p}\overline{FEH}(\delta(1))$. Furthermore $N^{0.5}(FH(\check{\delta})-FEH(\delta))$ converges in distribution. Now, since $N^{0.5}\det(FEH(\delta(\rho)))=\lambda^2 c_{\det}(\rho)$, we obtain by using standard arguments that $N^{0.25}\overline{FEH}(\check{\delta})(N^{-0.5}\frac{\partial \log L_F(\delta)}{\partial\delta})\xrightarrow{d}N\left(0,\lambda^2 c_{\det}(1)\overline{FEH}(\delta(1))\right)$ and $N^{0.5}\det(FH(\check{\delta}))=N^{0.5}\det(FH(\check{\delta})-FEH(\delta)+FEH(\delta))\xrightarrow{d}\widetilde{Z}_2=\lambda^2 c_{\det}(1)+Z_{\det}(\lambda)$. Moreover $N^{0.25}(\widehat{\delta}_{FEML}-\delta)=(N^{0.5}\det(FH(\check{\delta})))^{-1}N^{0.25}\overline{FEH}(\check{\delta})(N^{-0.5}\frac{\partial \log L_F(\delta)}{\partial\delta})$.

We can use the above results to obtain the local-to-zero asymptotic distribution of $\widehat{\rho}_{FEML}$: $N^{0.25}(\widehat{\rho}_{FEML}-\rho)\xrightarrow{d}\widetilde{Z}_1/\widetilde{Z}_2$, where $\widetilde{Z}_1\sim N\left(0,\lambda^2\lim_{\rho\uparrow 1}c_{\det}(\rho)\overline{FEH}_{11}(\delta(1))\right)$. Moreover $E(\widetilde{Z}_1\widetilde{Z}_2)=\lim_{N\to\infty}E[N^{0.25}(1,\,0,\,0)\overline{FEH}(\check{\delta})(N^{-0.5}\frac{\partial \log L_F(\delta)}{\partial\delta})N^{0.5}\det(FH(\check{\delta}))]\neq 0$. $\square$

$T=10$	$N=100$	ARBOND2		REQMLE		FEQMLE		SYS2		OLGMM2	
model	ρ	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
L	0.50	0.475	0.057	0.499	0.046	0.499	0.046	0.508	0.041	0.549	0.043
L	0.80	0.729	0.084	0.805	0.070	0.805	0.069	0.802	0.029	0.837	0.027
L	0.95^a	0.597	0.195	0.935	0.068	0.929	0.065	0.950	0.021	0.966	0.015
L	0.95^b	0.818	0.101	0.946	0.044	0.943	0.043	0.951	0.009	0.989	0.005
S	0.50	0.473	0.061	0.500	0.042	0.500	0.042	0.516	0.047	0.518	0.046
S	0.80	0.750	0.069	0.805	0.059	0.809	0.069	0.801	0.042	0.804	0.040
S	0.95^a	0.820	0.110	0.962	0.065	0.937	0.060	0.945	0.038	0.946	0.032
S	0.95^b	0.923	0.043	0.956	0.039	0.949	0.041	0.949	0.017	0.951	0.016
H	0.50	0.478	0.055	0.500	0.038	0.500	0.038	0.518	0.044	0.536	0.043
H	0.80	0.769	0.055	0.801	0.047	0.804	0.055	0.804	0.045	0.823	0.040
H	0.95^a	0.878	0.076	0.958	0.060	0.948	0.064	0.941	0.045	0.951	0.033
H	0.95^b	0.935	0.031	0.951	0.030	0.953	0.038	0.947	0.028	0.977	0.016
C	0.50	0.480	0.052	0.499	0.042	0.499	0.043	0.642	0.063	0.635	0.060
C	0.80	0.758	0.064	0.803	0.056	0.808	0.069	0.783	0.045	0.793	0.043
C	0.95^a	0.825	0.109	0.963	0.065	0.936	0.061	0.942	0.040	0.945	0.033
C	0.95^b	0.925	0.043	0.955	0.038	0.950	0.041	0.944	0.019	0.950	0.017
M	0.50	0.477	0.054	0.499	0.038	0.499	0.038	0.517	0.044	0.539	0.043
M	0.80	0.770	0.055	0.801	0.047	0.803	0.053	0.805	0.045	0.827	0.040
M	0.95^a	0.880	0.077	0.958	0.060	0.949	0.064	0.942	0.045	0.952	0.032
M	0.95^b	0.936	0.031	0.952	0.030	0.954	0.038	0.947	0.028	0.979	0.016

Table 1: Estimators of ρ; Design I with $\sigma_\mu^2 = 1$; 10000 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		FEQMLE		SYS2		OLGMM2	
model	ρ	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
L	0.50	0.466	0.113	0.506	0.101	0.506	0.101	0.503	0.057	0.602	0.070
L	0.80	0.645	0.210	0.796	0.126	0.793	0.123	0.802	0.043	0.894	0.038
L	0.95^a	0.298	0.410	0.922	0.117	0.899	0.106	0.953	0.037	0.984	0.015
L	0.95^b	0.604	0.297	0.935	0.075	0.922	0.071	0.950	0.015	0.995	0.005
S	0.50	0.469	0.105	0.503	0.083	0.506	0.090	0.507	0.074	0.511	0.073
S	0.80	0.738	0.139	0.822	0.119	0.813	0.122	0.795	0.074	0.805	0.070
S	0.95^a	0.750	0.256	0.976	0.114	0.917	0.107	0.937	0.077	0.943	0.061
S	0.95^b	0.915	0.099	0.966	0.073	0.931	0.066	0.948	0.031	0.952	0.029
H	0.50	0.481	0.085	0.502	0.066	0.502	0.069	0.511	0.072	0.568	0.072
H	0.80	0.767	0.100	0.810	0.097	0.814	0.111	0.798	0.085	0.859	0.070
H	0.95^a	0.851	0.172	0.971	0.110	0.946	0.118	0.931	0.095	0.956	0.059
H	0.95^b	0.933	0.067	0.957	0.062	0.949	0.069	0.945	0.055	0.983	0.023
C	0.50	0.481	0.086	0.501	0.075	0.505	0.089	0.631	0.114	0.628	0.110
C	0.80	0.751	0.127	0.816	0.112	0.811	0.120	0.761	0.080	0.798	0.078
C	0.95^a	0.753	0.251	0.974	0.115	0.916	0.109	0.932	0.083	0.941	0.064
C	0.95^b	0.915	0.097	0.964	0.071	0.933	0.065	0.941	0.034	0.950	0.029
M	0.50	0.480	0.084	0.501	0.066	0.501	0.068	0.510	0.071	0.575	0.072
M	0.80	0.768	0.098	0.811	0.097	0.814	0.110	0.798	0.084	0.864	0.068
M	0.95^a	0.852	0.169	0.971	0.109	0.947	0.118	0.932	0.095	0.958	0.058
M	0.95^b	0.935	0.070	0.959	0.064	0.950	0.070	0.945	0.057	0.984	0.024

Table 2: Estimators of ρ; Design I with $\sigma_\mu^2 = 1$; 10000 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		FEQMLE		SYS2		OLGMM2	
model	ρ	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
L	0.50	0.466	0.112	0.506	0.101	0.506	0.100	0.503	0.060	0.592	0.078
L	0.80	0.644	0.210	0.795	0.127	0.791	0.124	0.802	0.047	0.871	0.056
L	0.95^a	0.294	0.414	0.922	0.116	0.899	0.107	0.952	0.040	0.970	0.038
L	0.95^b	0.599	0.304	0.929	0.074	0.923	0.071	0.951	0.016	0.990	0.011
S	0.50	0.466	0.111	0.503	0.085	0.506	0.092	0.509	0.077	0.512	0.075
S	0.80	0.711	0.167	0.837	0.132	0.811	0.122	0.800	0.078	0.811	0.072
S	0.95^a	0.617	0.335	0.971	0.118	0.916	0.106	0.941	0.079	0.949	0.060
S	0.95^b	0.878	0.135	0.978	0.076	0.932	0.066	0.947	0.034	0.953	0.030
H	0.50	0.477	0.089	0.501	0.067	0.501	0.068	0.514	0.073	0.571	0.073
H	0.80	0.758	0.112	0.826	0.112	0.815	0.109	0.812	0.089	0.868	0.071
H	0.95^a	0.802	0.216	0.985	0.114	0.945	0.118	0.944	0.093	0.963	0.058
H	0.95^b	0.923	0.083	0.964	0.070	0.951	0.070	0.949	0.061	0.983	0.025
C	0.50	0.482	0.082	0.500	0.072	0.503	0.084	0.756	0.103	0.750	0.095
C	0.80	0.767	0.102	0.809	0.097	0.815	0.114	0.905	0.079	0.935	0.057
C	0.95^a	0.839	0.177	0.974	0.111	0.942	0.117	0.968	0.071	0.979	0.040
C	0.95^b	0.930	0.072	0.957	0.065	0.948	0.069	0.992	0.037	0.995	0.011
M	0.50	0.478	0.088	0.500	0.065	0.501	0.067	0.514	0.073	0.578	0.072
M	0.80	0.757	0.114	0.826	0.113	0.816	0.111	0.811	0.090	0.871	0.070
M	0.95^a	0.803	0.212	0.983	0.112	0.947	0.117	0.946	0.094	0.964	0.058
M	0.95^b	0.923	0.083	0.964	0.070	0.950	0.070	0.949	0.062	0.983	0.025

Table 3: Estimators of ρ; Design I with $\sigma_\mu^2 = 1/(1-\rho^2)$; 10000 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		FEQMLE		SYS2		OLGMM2	
model	ρ	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
L	0.50	0.463	0.130	0.462	0.075	0.463	0.075	0.504	0.063	0.597	0.070
L	0.80	0.608	0.232	0.729	0.094	0.729	0.094	0.802	0.049	0.881	0.043
L	0.95^a	0.268	0.420	0.863	0.103	0.864	0.107	0.954	0.042	0.979	0.021
L	0.95^b	0.531	0.341	0.899	0.062	0.897	0.064	0.951	0.019	0.993	0.006
S	0.50	0.462	0.115	0.496	0.087	0.509	0.109	0.503	0.083	0.505	0.082
S	0.80	0.720	0.156	0.762	0.083	0.784	0.111	0.794	0.080	0.800	0.076
S	0.95^a	0.699	0.296	0.880	0.094	0.893	0.104	0.938	0.085	0.943	0.069
S	0.95^b	0.898	0.119	0.912	0.052	0.915	0.066	0.948	0.035	0.951	0.032
H	0.50	0.479	0.093	0.500	0.077	0.510	0.098	0.510	0.081	0.555	0.080
H	0.80	0.758	0.111	0.769	0.072	0.810	0.122	0.793	0.090	0.840	0.077
H	0.95^a	0.814	0.212	0.888	0.085	0.909	0.104	0.930	0.097	0.946	0.067
H	0.95^b	0.925	0.082	0.921	0.045	0.934	0.066	0.941	0.064	0.977	0.029
C	0.50	0.476	0.097	0.477	0.067	0.506	0.105	0.524	0.105	0.532	0.105
C	0.80	0.735	0.137	0.754	0.078	0.778	0.111	0.754	0.086	0.781	0.082
C	0.95^a	0.693	0.299	0.878	0.094	0.888	0.105	0.935	0.085	0.941	0.067
C	0.95^b	0.900	0.113	0.914	0.052	0.916	0.065	0.945	0.037	0.950	0.033
M	0.50	0.478	0.093	0.500	0.077	0.506	0.097	0.509	0.079	0.563	0.078
M	0.80	0.761	0.109	0.772	0.071	0.806	0.115	0.794	0.089	0.849	0.076
M	0.95^a	0.815	0.206	0.889	0.086	0.905	0.104	0.930	0.099	0.949	0.067
M	0.95^b	0.924	0.083	0.920	0.044	0.932	0.066	0.939	0.061	0.977	0.031

Table 4: Estimators of ρ; Design II with $\sigma_\mu^2 = 1$; 2500 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		FEQMLE		SYS2		OLGMM2	
model	ρ	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
L	0.50	0.458	0.132	0.460	0.072	0.461	0.073	0.504	0.064	0.589	0.074
L	0.80	0.610	0.244	0.731	0.093	0.730	0.094	0.802	0.052	0.867	0.055
L	0.95^a	0.282	0.419	0.863	0.101	0.864	0.104	0.952	0.045	0.969	0.037
L	0.95^b	0.533	0.333	0.899	0.064	0.896	0.065	0.951	0.019	0.991	0.011
S	0.50	0.465	0.121	0.501	0.090	0.511	0.107	0.508	0.082	0.510	0.081
S	0.80	0.699	0.174	0.765	0.093	0.780	0.112	0.798	0.083	0.807	0.079
S	0.95^a	0.579	0.350	0.877	0.099	0.890	0.105	0.939	0.086	0.945	0.068
S	0.95^b	0.869	0.146	0.914	0.057	0.916	0.065	0.946	0.038	0.952	0.033
H	0.50	0.476	0.097	0.500	0.079	0.507	0.095	0.512	0.080	0.558	0.080
H	0.80	0.747	0.128	0.779	0.081	0.816	0.121	0.805	0.096	0.847	0.079
H	0.95^a	0.764	0.248	0.887	0.091	0.904	0.105	0.939	0.095	0.954	0.065
H	0.95^b	0.918	0.097	0.925	0.050	0.936	0.066	0.946	0.067	0.977	0.032
C	0.50	0.479	0.092	0.478	0.065	0.509	0.105	0.634	0.120	0.641	0.118
C	0.80	0.757	0.114	0.763	0.070	0.794	0.114	0.846	0.097	0.892	0.082
C	0.95^a	0.799	0.217	0.887	0.086	0.904	0.102	0.956	0.085	0.970	0.056
C	0.95^b	0.924	0.085	0.921	0.045	0.929	0.066	0.973	0.056	0.989	0.021
M	0.50	0.476	0.094	0.501	0.078	0.509	0.097	0.513	0.077	0.566	0.076
M	0.80	0.751	0.127	0.779	0.081	0.808	0.116	0.804	0.093	0.854	0.076
M	0.95^a	0.764	0.244	0.888	0.093	0.904	0.106	0.940	0.098	0.956	0.067
M	0.95^b	0.914	0.098	0.922	0.050	0.931	0.066	0.942	0.061	0.977	0.032

Table 5: Estimators of ρ; Design II with $\sigma_\mu^2 = 1/(1-\rho^2)$; 2500 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		FEQMLE		SYS2		OLGMM2	
model	ρ	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
L	0.50	0.438	0.173	0.461	0.080	0.462	0.081	0.501	0.072	0.570	0.083
L	0.80	0.508	0.334	0.749	0.086	0.750	0.086	0.803	0.061	0.879	0.056
L	0.95^a	0.249	0.484	0.894	0.086	0.896	0.086	0.953	0.054	0.979	0.027
L	0.95^b	0.433	0.459	0.919	0.042	0.919	0.043	0.950	0.023	0.996	0.006
S	0.50	0.438	0.165	0.494	0.091	0.500	0.101	0.504	0.085	0.506	0.094
S	0.80	0.627	0.255	0.762	0.085	0.775	0.095	0.799	0.079	0.814	0.087
S	0.95^a	0.495	0.442	0.896	0.084	0.902	0.088	0.949	0.075	0.952	0.067
S	0.95^b	0.816	0.219	0.926	0.041	0.931	0.045	0.950	0.032	0.971	0.036
H	0.50	0.449	0.138	0.495	0.086	0.501	0.097	0.506	0.088	0.513	0.085
H	0.80	0.699	0.196	0.772	0.080	0.794	0.100	0.799	0.092	0.806	0.088
H	0.95^a	0.653	0.382	0.900	0.081	0.910	0.089	0.947	0.098	0.945	0.081
H	0.95^b	0.882	0.146	0.928	0.039	0.937	0.047	0.948	0.045	0.949	0.042
C	0.50	0.469	0.114	0.474	0.070	0.501	0.099	0.566	0.112	0.583	0.112
C	0.80	0.697	0.191	0.759	0.079	0.780	0.097	0.768	0.091	0.779	0.091
C	0.95^a	0.535	0.421	0.896	0.084	0.903	0.089	0.943	0.083	0.946	0.074
C	0.95^b	0.841	0.198	0.925	0.041	0.930	0.045	0.947	0.035	0.956	0.040
M	0.50	0.463	0.117	0.498	0.080	0.503	0.090	0.510	0.084	0.556	0.083
M	0.80	0.737	0.147	0.773	0.076	0.797	0.103	0.801	0.098	0.836	0.082
M	0.95^a	0.745	0.286	0.901	0.079	0.914	0.090	0.943	0.106	0.949	0.078
M	0.95^b	0.913	0.106	0.929	0.038	0.940	0.049	0.950	0.053	0.965	0.035

Table 6: Estimators of ρ; Design III with $\sigma_\mu^2 = 1$; 2500 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		FEQMLE		SYS2		OLGMM2	
model	ρ	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
L	0.50	0.441	0.175	0.461	0.084	0.462	0.084	0.503	0.077	0.562	0.089
L	0.80	0.521	0.337	0.752	0.084	0.753	0.084	0.804	0.064	0.860	0.070
L	0.95^a	0.240	0.486	0.895	0.086	0.897	0.087	0.954	0.057	0.968	0.048
L	0.95^b	0.434	0.450	0.924	0.042	0.924	0.042	0.951	0.024	0.990	0.015
S	0.50	0.431	0.172	0.494	0.095	0.500	0.103	0.505	0.089	0.506	0.095
S	0.80	0.571	0.298	0.768	0.089	0.775	0.095	0.801	0.081	0.812	0.088
S	0.95^a	0.381	0.481	0.900	0.086	0.903	0.088	0.948	0.079	0.953	0.068
S	0.95^b	0.698	0.319	0.929	0.043	0.931	0.045	0.951	0.033	0.964	0.037
H	0.50	0.447	0.144	0.497	0.089	0.502	0.099	0.508	0.091	0.518	0.089
H	0.80	0.668	0.220	0.776	0.087	0.790	0.102	0.802	0.096	0.811	0.089
H	0.95^a	0.538	0.420	0.900	0.086	0.906	0.092	0.947	0.099	0.948	0.079
H	0.95^b	0.845	0.194	0.931	0.043	0.936	0.048	0.950	0.045	0.953	0.041
C	0.50	0.472	0.105	0.475	0.069	0.502	0.098	0.662	0.114	0.681	0.108
C	0.80	0.744	0.140	0.765	0.072	0.794	0.100	0.858	0.098	0.889	0.084
C	0.95^a	0.747	0.285	0.902	0.079	0.913	0.090	0.965	0.102	0.965	0.067
C	0.95^b	0.913	0.106	0.929	0.038	0.940	0.048	0.975	0.057	0.982	0.028
M	0.50	0.463	0.123	0.500	0.083	0.503	0.091	0.510	0.085	0.556	0.084
M	0.80	0.716	0.174	0.783	0.082	0.799	0.102	0.809	0.099	0.842	0.084
M	0.95^a	0.648	0.354	0.906	0.083	0.914	0.091	0.950	0.106	0.954	0.078
M	0.95^b	0.890	0.132	0.933	0.041	0.940	0.048	0.951	0.055	0.965	0.038

Table 7: Estimators of ρ; Design III with $\sigma_\mu^2 = 1/(1-\rho^2)$; 2500 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		SYS2	
model	ρ	mean	SD	mean	SD	mean	SD
S	0.50	0.353	0.291	0.484	0.196	0.478	0.204
S	0.80	0.656	0.273	0.726	0.136	0.771	0.138
S	0.95^a	0.594	0.448	0.850	0.131	0.920	0.129
S	0.95^b	0.882	0.185	0.900	0.072	0.943	0.052
C	0.50	0.370	0.248	0.462	0.184	0.472	0.223
C	0.80	0.687	0.243	0.725	0.126	0.738	0.144
C	0.95^a	0.625	0.437	0.854	0.129	0.916	0.127
C	0.95^b	0.891	0.186	0.899	0.073	0.940	0.055

Table 8: Estimators of ρ; Design IV with $\sigma_\mu^2 = 1$; 2500 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		SYS2	
model	ρ	mean	SD	mean	SD	mean	SD
S	0.50	0.344	0.294	0.485	0.192	0.485	0.209
S	0.80	0.630	0.296	0.734	0.141	0.780	0.140
S	0.95^a	0.519	0.480	0.857	0.133	0.929	0.122
S	0.95^b	0.847	0.228	0.899	0.079	0.942	0.058
C	0.50	0.372	0.267	0.470	0.188	0.547	0.252
C	0.80	0.710	0.212	0.739	0.121	0.800	0.153
C	0.95^a	0.738	0.339	0.862	0.123	0.946	0.125
C	0.95^b	0.911	0.132	0.904	0.066	0.955	0.078

Table 9: Estimators of ρ; Design IV with $\sigma_\mu^2 = 1/(1-\rho^2)$; 2500 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		SYS2	
model	ρ	mean	SD	mean	SD	mean	SD
S	0.50	0.362	0.285	0.482	0.194	0.482	0.203
S	0.80	0.665	0.266	0.732	0.135	0.777	0.136
S	0.95^a	0.624	0.432	0.855	0.130	0.927	0.121
S	0.95^b	0.874	0.183	0.894	0.073	0.942	0.052
C	0.50	0.377	0.261	0.464	0.185	0.473	0.226
C	0.80	0.688	0.244	0.727	0.126	0.743	0.143
C	0.95^a	0.605	0.454	0.855	0.126	0.916	0.130
C	0.95^b	0.883	0.180	0.896	0.073	0.937	0.055

Table 10: Estimators of ρ; Design V with $\sigma_\mu^2 = 1$; 2500 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$	ARBOND2		REQMLE		SYS2	
model	ρ	mean	SD	mean	SD	mean	SD
S	0.50	0.334	0.303	0.482	0.202	0.481	0.204
S	0.80	0.629	0.309	0.738	0.138	0.780	0.136
S	0.95^a	0.513	0.494	0.860	0.131	0.923	0.130
S	0.95^b	0.841	0.222	0.898	0.076	0.943	0.054
C	0.50	0.369	0.259	0.462	0.186	0.532	0.246
C	0.80	0.700	0.221	0.735	0.127	0.799	0.158
C	0.95^a	0.738	0.337	0.862	0.124	0.940	0.125
C	0.95^b	0.911	0.130	0.904	0.067	0.955	0.078

Table 11: Estimators of ρ; Design V with $\sigma_\mu^2 = 1/(1-\rho^2)$; 2500 replications; SD: standard deviation; $a : N = 100, b : N = 500$.

$T=6$	$N=100$		ARBOND2		REQMLE		FEQMLE		SYS2	
model	ρ		LB	UB	LB	UB	LB	UB	LB	UB
S	0.50	E	0.298	0.636	0.377	0.634	0.376	0.643	0.384	0.628
		A	0.326	0.613	0.375	0.631	0.371	0.639	0.414	0.600
		B/W	0.299	0.643	0.378	0.660	0.382	0.688	0.388	0.625
S	0.80	E	0.507	0.966	0.643	1.037	0.630	1.023	0.669	0.909
		A	0.551	0.924	0.667	0.980	0.617	1.009	0.705	0.885
		B/W	0.511	0.964	0.658	0.982	0.669	0.992	0.677	0.910
S	0.95^a	E	0.309	1.109	0.773	1.145	0.733	1.063	0.796	1.050
		A	0.453	1.042	0.810	1.136	0.687	1.141	0.853	1.021
		B/W	0.355	1.134	0.799	1.097	0.789	1.065	0.819	1.056
S	0.95^b	E	0.756	1.073	0.848	1.080	0.823	1.024	0.895	0.995
		A	0.759	1.066	0.860	1.066	0.785	1.080	0.903	0.993
		B/W	0.752	1.074	0.855	1.051	0.857	1.029	0.897	0.998
C	0.50	E	0.335	0.623	0.381	0.629	0.376	0.648	0.447	0.818
		A	0.361	0.596	0.380	0.622	0.371	0.640	0.550	0.708
		B/W	0.342	0.620	0.382	0.637	0.383	0.694	0.495	0.769
C	0.80	E	0.539	0.953	0.644	1.015	0.629	1.021	0.626	0.889
		A	0.577	0.916	0.660	0.966	0.615	1.005	0.655	0.866
		B/W	0.541	0.955	0.654	0.969	0.671	0.993	0.628	0.896
C	0.95^a	E	0.298	1.117	0.776	1.143	0.732	1.068	0.781	1.052
		A	0.461	1.046	0.811	1.137	0.694	1.139	0.844	1.020
		B/W	0.360	1.133	0.797	1.096	0.784	1.064	0.806	1.059
C	0.95^b	E	0.759	1.075	0.851	1.076	0.824	1.024	0.882	0.993
		A	0.762	1.065	0.859	1.065	0.788	1.082	0.891	0.991
		B/W	0.755	1.075	0.851	1.052	0.857	1.033	0.885	0.998

Table 12: 90% Confidence Intervals; Design I with $\sigma_\mu^2 = 1$; 10000 Monte Carlo replications; E: CI based on empirical distribution; A: CI based on first-order fixed parameter asymptotic standard errors; B: asymmetric 90% bootstrap CI based on 100 replications (for QMLE only); W: CI based on Windmeijer's corrected asymptotic standard errors (for GMM estimators only); LB: lowerbound; UB: upperbound; $a : N = 100, b : N = 500$.

$T=6$	$N=100$		ARBOND2		REQMLE		FEQMLE		SYS2	
model	ρ		SE	RP	SE	RP	SE	RP	SE	RP
S	0.50	E	0.104		0.082		0.089		0.074	
		A	0.087	0.190	0.078	0.096	0.081	0.088	0.057	0.218
		B/W	0.105	0.116	0.087	0.114	0.095	0.115	0.072	0.125
S	0.80	E	0.139		0.119		0.121		0.073	
		A	0.113	0.221	0.095	0.242	0.119	0.194	0.055	0.224
		B/W	0.138	0.142	0.099	0.122	0.099	0.239	0.071	0.116
S	0.95^a	E	0.252		0.115		0.107		0.061	
		A	0.179	0.353	0.099	0.289	0.138	0.183	0.051	0.257
		B/W	0.237	0.210	0.092	0.154	0.086	0.158	0.072	0.117
S	0.95^b	E	0.097		0.072		0.065		0.031	
		A	0.093	0.153	0.062	0.252	0.090	0.156	0.028	0.132
		B/W	0.098	0.128	0.060	0.124	0.054	0.249	0.031	0.096
C	0.50	E	0.087		0.076		0.090		0.113	
		A	0.071	0.193	0.073	0.110	0.082	0.093	0.048	0.674
		B/W	0.085	0.124	0.078	0.102	0.097	0.115	0.083	0.471
C	0.80	E	0.125		0.112		0.122		0.080	
		A	0.103	0.218	0.093	0.226	0.118	0.202	0.064	0.245
		B/W	0.126	0.131	0.097	0.120	0.099	0.234	0.082	0.141
C	0.95^a	E	0.252		0.114		0.108		0.082	
		A	0.178	0.343	0.099	0.281	0.135	0.183	0.053	0.270
		B/W	0.235	0.207	0.092	0.153	0.087	0.154	0.077	0.118
C	0.95^b	E	0.096		0.070		0.064		0.034	
		A	0.092	0.147	0.063	0.226	0.089	0.143	0.030	0.137
		B/W	0.097	0.124	0.060	0.143	0.053	0.254	0.034	0.100

Table 13: Standard errors and size; Design I with $\sigma_\mu^2 = 1$; 10000 Monte Carlo replications; E: based on empirical distribution; A: based on first-order fixed parameter asymptotic distribution; B: based on bootstrap with 100 replications (for QMLE only); W: based on Windmeijer's corrected asymptotic standard errors (for GMM estimators only); SE: standard deviation/error; RP: rejection probability using the CI's defined in table 12 (nominal size is 10%); $a : N = 100, b : N = 500$.

$T=6$	$N=100$		ARBOND2		REQMLE		FEQMLE		SYS2	
model	ρ		LB	UB	LB	UB	LB	UB	LB	UB
S	0.50	E	0.153	0.712	0.337	0.641	0.338	0.671	0.364	0.644
		A	0.230	0.643	0.343	0.644	0.335	0.666	0.408	0.606
		B/W	0.165	0.707	0.333	0.619	0.334	0.649	0.370	0.644
S	0.80	E	0.193	1.004	0.613	0.884	0.610	0.911	0.662	0.917
		A	0.334	0.925	0.621	0.906	0.620	0.932	0.710	0.890
		B/W	0.215	1.045	0.597	0.867	0.599	0.894	0.674	0.926
S	0.95^a	E	-0.273	1.134	0.744	1.018	0.743	1.031	0.819	1.072
		A	0.070	0.929	0.759	1.038	0.756	1.048	0.867	1.032
		B/W	-0.172	1.170	0.733	1.000	0.735	1.012	0.829	1.070
S	0.95^b	E	0.454	1.128	0.849	0.982	0.848	0.991	0.896	1.003
		A	0.531	1.109	0.854	0.997	0.854	1.010	0.904	0.999
		B/W	0.478	1.162	0.844	0.976	0.845	0.986	0.899	1.004
C	0.50	E	0.280	0.650	0.353	0.585	0.349	0.681	0.378	0.755
		A	0.320	0.614	0.357	0.593	0.346	0.667	0.475	0.658
		B/W	0.282	0.651	0.339	0.569	0.344	0.659	0.417	0.716
C	0.80	E	0.381	0.990	0.616	0.875	0.610	0.922	0.610	0.912
		A	0.464	0.938	0.629	0.890	0.623	0.941	0.660	0.878
		B/W	0.388	1.014	0.604	0.857	0.606	0.901	0.619	0.919
C	0.95^a	E	-0.272	1.130	0.737	1.011	0.735	1.024	0.797	1.077
		A	0.122	0.947	0.759	1.032	0.751	1.057	0.856	1.033
		B/W	-0.072	1.169	0.731	0.997	0.732	1.010	0.816	1.081
C	0.95^b	E	0.514	1.122	0.848	0.979	0.845	0.990	0.888	1.003
		A	0.575	1.104	0.855	0.996	0.851	1.011	0.895	0.998
		B/W	0.531	1.147	0.846	0.976	0.846	0.987	0.889	1.004

Table 14: 90% Confidence Intervals; Design III with $\sigma_\mu^2 = 1$; 10000 Monte Carlo replications; E: CI based on empirical distribution; A: CI based on first-order fixed parameter asymptotic standard errors; B: asymmetric 90% bootstrap CI based on 100 replications (for QMLE only); W: CI based on Windmeijer's corrected asymptotic standard errors (for GMM estimators only); LB: lowerbound; UB: upperbound; $a : N = 100, b : N = 500$.

$T=6$	$N=100$		ARBOND2		REQMLE		FEQMLE		SYS2	
model	ρ		SE	RP	SE	RP	SE	RP	SE	RP
S	0.50	E	0.168		0.092		0.100		0.087	
		A	0.126	0.255	0.092	0.164	0.100	0.165	0.060	0.263
		B/W	0.165	0.140	0.086	0.095	0.095	0.100	0.083	0.130
S	0.80	E	0.259		0.083		0.095		0.078	
		A	0.180	0.334	0.086	0.162	0.095	0.205	0.055	0.252
		B/W	0.252	0.162	0.082	0.158	0.089	0.109	0.077	0.117
S	0.95^a	E	0.446		0.083		0.088		0.077	
		A	0.261	0.511	0.085	0.179	0.089	0.186	0.050	0.273
		B/W	0.408	0.288	0.081	0.188	0.084	0.153	0.073	0.112
S	0.95^b	E	0.216		0.042		0.044		0.032	
		A	0.176	0.213	0.043	0.126	0.047	0.153	0.029	0.131
		B/W	0.208	0.129	0.040	0.166	0.043	0.113	0.032	0.101
C	0.50	E	0.112		0.070		0.101		0.110	
		A	0.089	0.210	0.071	0.132	0.097	0.136	0.056	0.477
		B/W	0.112	0.123	0.069	0.145	0.095	0.105	0.091	0.262
C	0.80	E	0.190		0.080		0.099		0.092	
		A	0.144	0.267	0.080	0.160	0.097	0.217	0.066	0.272
		B/W	0.190	0.137	0.076	0.173	0.089	0.105	0.091	0.133
C	0.95^a	E	0.421		0.083		0.088		0.083	
		A	0.251	0.484	0.083	0.184	0.093	0.178	0.054	0.268
		B/W	0.377	0.265	0.080	0.206	0.084	0.164	0.081	0.118
C	0.95^b	E	0.191		0.040		0.045		0.036	
		A	0.161	0.210	0.043	0.143	0.049	0.160	0.031	0.153
		B/W	0.187	0.130	0.040	0.172	0.043	0.107	0.035	0.106

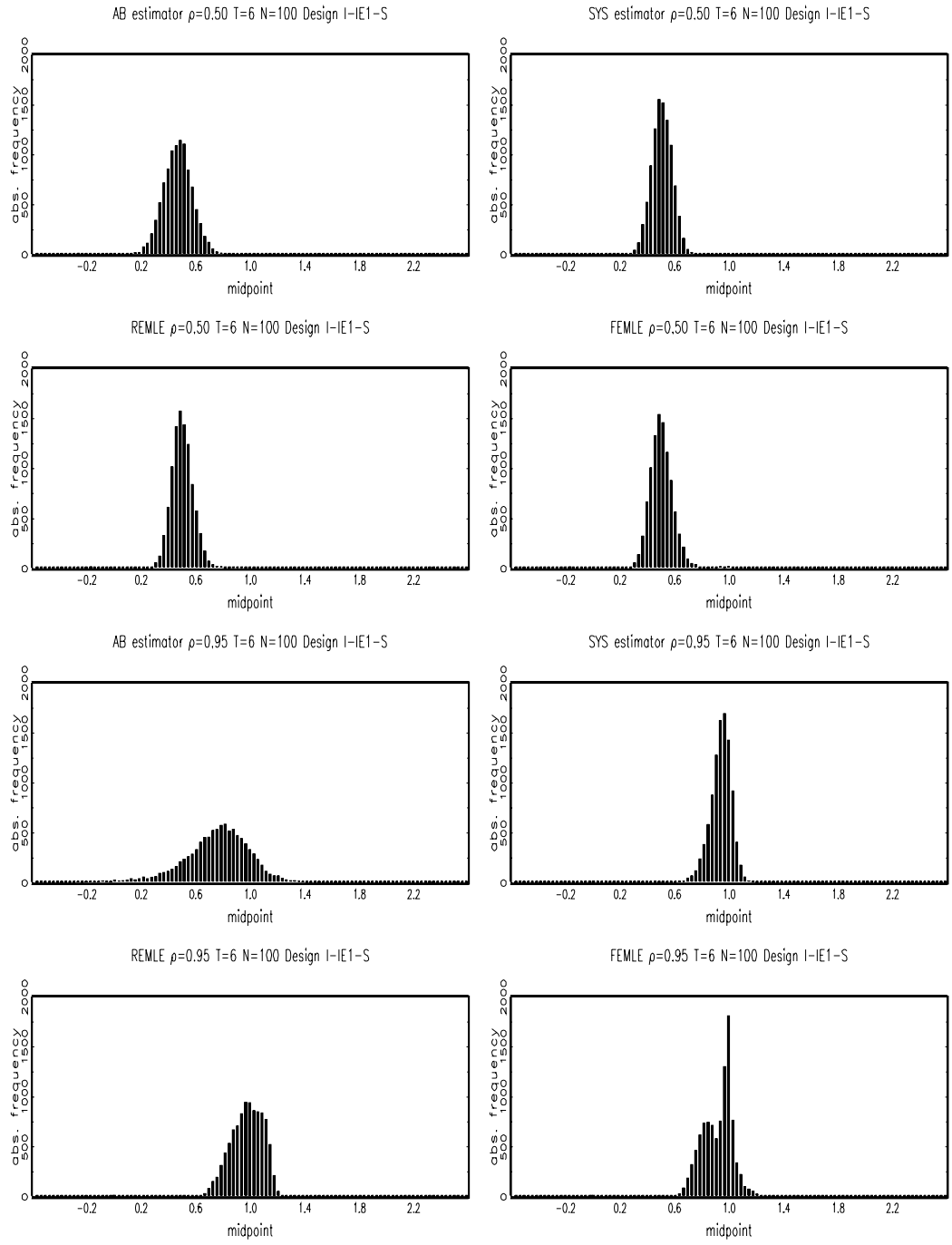
Table 15: Standard errors and size; Design III with $\sigma_\mu^2 = 1$; 10000 Monte Carlo replications; E: based on empirical distribution; A: based on first-order fixed parameter asymptotic distribution; B: based on bootstrap with 100 replications (for QMLE only); W: based on Windmeijer's corrected asymptotic standard errors (for GMM estimators only); SE: standard deviation/error; RP: rejection probability using the CI's defined in table 14 (nominal size is 10%); $a : N = 100, b : N = 500$.

ARBOND2	$N = 100$		$N = 200$		$N = 400$	
ρ	bias	SD	bias	SD	bias	SD
0.90	-0.110	0.180	-0.051	0.119		
0.92929	-0.202	0.256	-0.101	0.164	-0.048	0.113
0.95	-0.345	0.336	-0.195	0.238	-0.100	0.159
0.96465	-0.498	0.404	-0.340	0.328	-0.197	0.232
0.975	-0.618	0.424	-0.497	0.404	-0.320	0.336

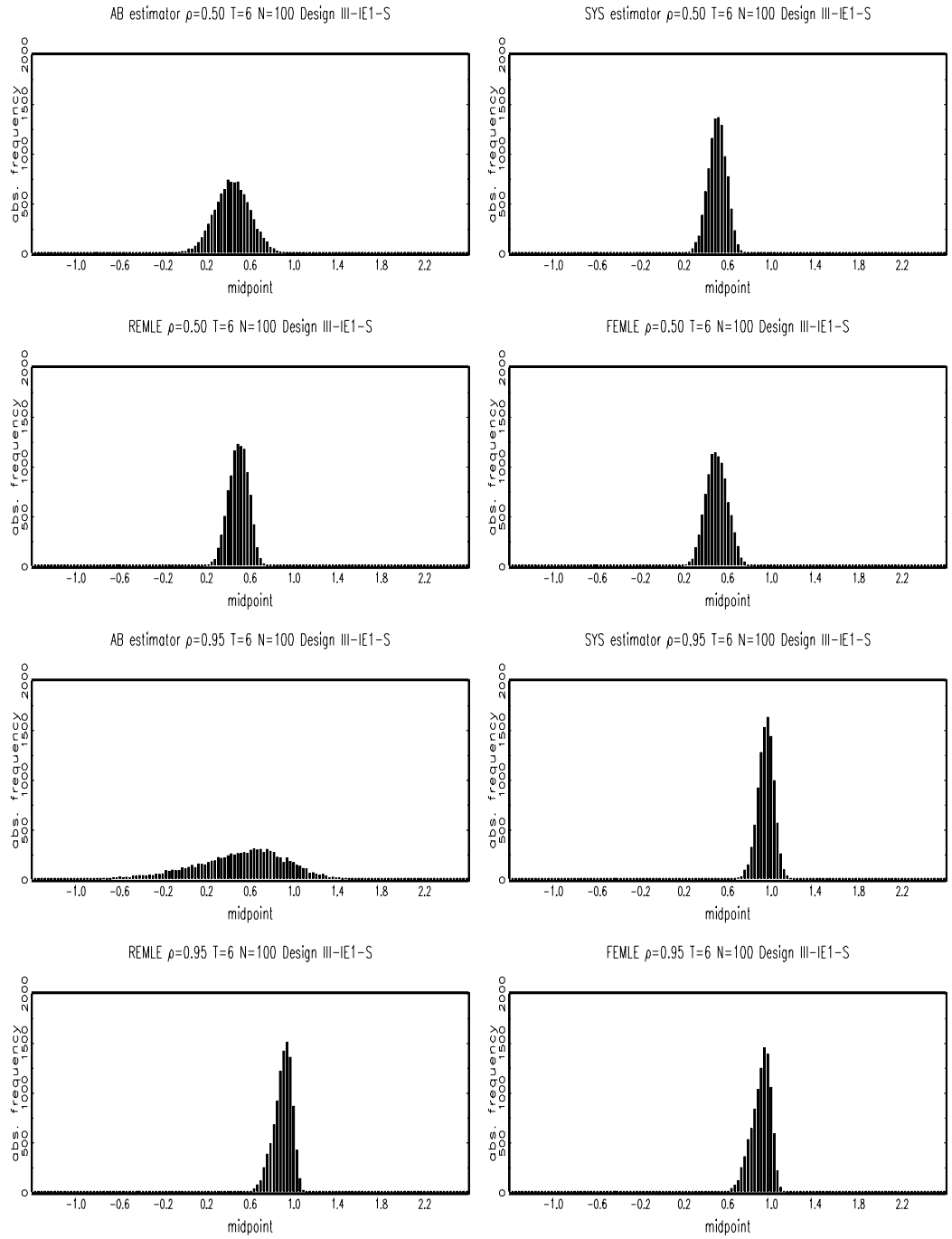
Table 16: Distributions of $\widehat{\rho}_{AB}$; Design I with $(y_{i,1}-\mu_i)|\mu_i \sim N(0, \sigma^2/[1-(0.9)^2])$ and $\sigma_\mu^2 = 1$; $T = 6$; 5000 replications; SD: standard deviation.

FEMLE	$N = 100$		$N = 400$	
ρ	bias	SD	bias	SD
0.90	-0.018	0.107		
0.92929	-0.034	0.105	-0.014	0.073
0.95	-0.045	0.108	-0.025	0.072
0.96465	-0.045	0.108	-0.031	0.071
0.975	-0.042	0.110	-0.033	0.071

Table 17: Distributions of $\widehat{\rho}_{FEML}$; Design I with $(y_{i,1} - \mu_i)|\mu_i \sim N(0, \sigma^2/[1 - (0.9)^2])$ and $\sigma_\mu^2 = 1$; $T = 6$; 5000 replications; SD: standard deviation.



Top four graphs for $\rho = 0.50$ and bottom four graphs for $\rho = 0.95$; in both groups of four graphs clockwise starting from top left: the AB GMM, the SYS, the FEML and the REML estimator for ρ; Design I-IE1-S, $T = 6$, $N = 100$; 10,000 replications.



Top four graphs for $\rho = 0.50$ and bottom four graphs for $\rho = 0.95$; in both groups of four graphs clockwise starting from top left: the AB GMM, the SYS, the FEML and the REML estimator for ρ; Design III-IE1-S, $T = 6$, $N = 100$; 10,000 replications.

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] Alonso-Borrego, C., and M. Arellano, 1996, Symmetrically normalised instrumental variable estimation using panel data, Journal of Business & Economic Statistics 17, 36-49.

[4] Alvarez, J., and M. Arellano, 2003, The time series and cross-section asymptotics of dynamic panel data estimators, Econometrica 71, 1121-1159.

[5] Amemiya, T., 1985, Advanced Econometrics, Harvard University Press.

[6] Anderson, T.W., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, Journal of econometrics 18, 47-82.

[7] Arellano, M., and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of economic studies 58, 277-297.

[8] Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, Journal of econometrics 68, 29-51.

[9] 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.

[10] Blundell, R.W., and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, Journal of econometrics 87, 115-143.

[11] Blundell, R.W., and S. Bond, 2000, GMM estimation with persistent panel data: an application to production functions, Econometric Reviews 19, 321-340.

[12] Blundell, R.W., and R. Smith, 1991, Initial conditions and efficient estimation in dynamic panel data models: an application to company investment behaviour, Annales d'économie et de la statistique 20/21, 109-123.

[13] Bound, J., D.A. Jaeger, and R.M. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, Journal of the American statistical association 90, 443-450.

[14] Chamberlain, G., 1980, Analysis of covariance with qualitative data, Review of economic studies XLVII, 225-238.

[15] Griliches, Z., and J. Mairesse, 1998, Production functions: the search for identification, in S. Strom (ed.), Essays in honour of Ragnar Frisch, Econometric Society Monograph Series, Cambridge University Press, Cambridge, United Kingdom.

[16] Hsiao, C., 1986, Analysis of panel data, Econometric Society Monographs, Cambridge University Press.

[17] Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel models covering short time periods, Journal of econometrics 109, 107-150.

[18] 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.

[19] Kruiniger, H., 2001, On the estimation of panel regression models with fixed effects, appeared in 2002 as working paper no. 450, Queen Mary, University of London, revised in 2003.

[20] Kruiniger, H., 2003, An efficient Linear GMM estimator for the covariance stationary AR(1)/UR model for panel data, forthcoming in Econometric Theory.

[21] Kruiniger, H., 2006a, GMM estimation and inference in dynamic panel data models with persistent data, available at http://alpha.qmul.ac.uk/∼ugte185/

[22] Kruiniger, H., 2006b, Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/UR model, available at http://alpha.qmul.ac.uk/∼ugte185/

[23] MaCurdy, T., 1981a, Multiple time series models applied to panel data, NBER working paper no. W0646, Cambridge, Massachusetts.

[24] MaCurdy, T., 1981b, Asymptotic properties of quasi-maximum likelihood estimators and test statistics, NBER Technical working paper no. T0014, Cambridge, Massachusetts.

[25] MaCurdy, T., 1982, The use of time series processes to model the time structure of earnings in a longitudinal data analysis, Journal of econometrics 18, 83-114.

[26] Neyman, J., and E.L. Scott, 1948, Consistent estimates based on partially consistent observations, Econometrica 16, 1-32.

[27] Sims, C., 2000, Using the likelihood perspective to sharpen econometric discourse: three examples, Journal of econometrics 95, 443-462.

[28] Stock, J.H., and J.H. Wright, 2000, GMM with weak identification, Econometrica 68, 1055-1096.

[29] Windmeijer, F., 2005, A finite sample correction for the variance of linear efficient two-step GMM estimators, Journal of econometrics 126, 25-51.

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

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