

An efficient Linear GMM estimator for the covariance stationary AR(1)/unit root model for panel data

Hugo Kruiniger*
Queen Mary, University of London

June 21, 2006

JEL classification: C12, C13, C23.

Keywords: dynamic panel data, GMM estimation, covariance stationarity, unit root, superconsistency, superefficiency.

*Address: H.KRUINIGER@QMUL.AC.UK; Department of Economics, Mile End Road, London E1 4NS. I thank Steve Bond and Frank Windmeijer for kindly making one of their computer programs available to me. I also thank two anonymous referees and a co-editor for very helpful comments. This research was funded by the ESRC under grant R000239139.

Abstract

This paper considers GMM estimation of the *inclusive* panel AR(1) model which contains the covariance stationary panel AR(1) model and the panel AR(1) model with a unit root as special cases. The paper presents a two-step Optimal Linear GMM (OLGMM) estimator for the inclusive model which is asymptotically equivalent to the optimal nonlinear GMM estimator of Ahn and Schmidt (1997) when the data are covariance stationary. Next the paper derives the asymptotic distribution of the OLGMM estimator when the model has a unit root under a variety of assumptions about the initial observations and the initial estimator. It is shown that in most cases the OLGMM estimator is superconsistent. In addition it is shown that the Iterated OLGMM estimator is superefficient when the variance of the initial observations is finite and fixed, i.e. small compared to the cross-sectional dimension of the panel. The paper also conducts a Monte Carlo study in which the finite sample properties of various GMM estimators for the inclusive panel AR(1) model are compared.

1 Introduction

In this paper we study GMM estimation of the *inclusive* panel AR(1) model which contains the covariance stationary panel AR(1) model and the panel AR(1) model with a unit root as special cases.

We first derive a linear GMM estimator for the autoregression parameter ρ in the inclusive panel AR(1) model which exhausts all the information in the second moments of the data. We show that when $|\rho| < 1$ an optimal version of our Linear GMM estimator is asymptotically equivalent to the optimal GMM estimator for the inclusive panel AR(1) model that was proposed by Ahn and Schmidt (1997). The latter estimator exploits a moment condition that is nonlinear in ρ .

Next we derive the asymptotic distribution of the two-step Optimal Linear GMM estimator when $\rho = 1$ under a variety of assumptions about the initial observations and the initial estimator. We find that this distribution is non-normal and involves the asymptotic distribution of the initial estimator. In most cases that we consider the OLGMM estimator is superconsistent. In addition we show that the Iterated OLGMM estimator is superefficient when the variance of the initial observations is finite and fixed, i.e. small compared to the cross-sectional dimension of the panel. These asymptotic properties of the OLGMM estimators are related to a moment condition which only identifies the unit root and whose sample counterpart holds deterministically when $\rho = 1$. As a result, this moment condition is extremely informative when $\rho = 1$.

We also conduct a Monte Carlo study of the finite sample behaviour of various GMM estimators for the inclusive panel AR(1) model. In the Monte Carlo study we will pay special attention to the gain in precision and reduction in the rmse that will be achieved when ρ is close to or equal to one by employing our preferred version of the Optimal Linear GMM estimator instead of other GMM estimators which use a different weight matrix and/or use less moment conditions, such as for instance the System estimator of Arellano and Bover (1995) and Blundell and Bond (1998).

The paper is organised as follows. Section 2 reviews GMM estimation of the covariance stationary panel AR(1) model. Section 3 introduces the Linear GMM estimator for the inclusive panel AR(1) model. Section 4 examines the asymptotic properties of the Optimal Linear GMM estimator and an iterative version of this estimator. Finally, section 5 compares the performance of various GMM estimators for the inclusive panel AR(1) model by means of a Monte Carlo study. The proofs are collected in the appendix.

2 The inclusive panel AR(1) model

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

$$y_{i,t} = \rho y_{i,t-1} + v_{i,t}, \quad (1)$$

$$v_{i,t} = \eta_i + \varepsilon_{i,t}, \quad \text{where } \eta_i = (1 - \rho)\mu_i, \quad (2)$$

for $i = 1, \dots, N$ and $t = S+1, \dots, T$, where $S \leq 1$ is the common starting date of the N processes. We assume that the number of ‘individuals’ in the panel, N , is large while the number of observations per individual, T , is fixed. Moreover $-1 < \rho \leq 1$.

The observations on $y = (y_{.,1} \dots y_{.,T})'$ are i.i.d. across the N individuals.¹ The composite error terms, the $v_{i,t}$, satisfy the following standard assumptions (cf Ahn and Schmidt, 1995, 1997):

$$E(\mu_i) = 0 \text{ and } E(\mu_i^2) = \sigma_\mu^2 < \infty, \text{ for } i = 1, \dots, N, \quad (3)$$

$$E(\varepsilon_{i,t}) = 0, \quad E(\varepsilon_{i,t}\eta_i) = 0, \text{ for } i = 1, \dots, N \text{ and } t = S+1, \dots, T, \quad (4)$$

$$E(\varepsilon_{i,t}y_{i,S}) = 0, \text{ for } i = 1, \dots, N \text{ and } t = S+1, \dots, T, \quad (5)$$

and

$$E(\varepsilon_{i,s}\varepsilon_{i,t}) = 0, \text{ for } i = 1, \dots, N \text{ and } s \neq t, \quad s, t = S+1, \dots, T. \quad (6)$$

We also assume that the idiosyncratic errors, the $\varepsilon_{i,t}$, are homoskedastic over time:

$$E(\varepsilon_{i,t}^2) = \sigma^2 < \infty, \text{ for } i = 1, \dots, N \text{ and } t = S+1, \dots, T. \quad (7)$$

Finally, the initial conditions are given by $y_{i,S} = \mu_i + (1 - \rho)\xi_{i,S}$, where $Var(\xi_{i,S}) < \infty$, for $i = 1, \dots, N$. Note that when $\rho = 1$ the initial conditions are equal to the individual effects, the μ_i .

When $|\rho| < 1$, we assume that $\{y_{i,t}\}$ is covariance stationary. Ahn and Schmidt (1997) have shown that given assumptions (3)-(7), $\{y_{i,t}\}$ is covariance stationary if and only if the initial observations satisfy the following assumptions:

$$E(y_{i,1} - \mu_i) = 0, \quad E((y_{i,1} - \mu_i)\mu_i) = 0 \text{ and} \quad (8)$$

$$Var(y_{i,1} - \mu_i) = \frac{\sigma^2}{1 - \rho^2}, \text{ for } |\rho| < 1 \text{ and } i = 1, \dots, N. \quad (9)$$

The first assumption, (8), is also known as mean-stationarity (cf Arellano and Bover, 1995, and Blundell and Bond, 1998).

When $\rho = 1$, we have $y_{i,1} = \mu_i + \sum_{t=0}^{-S} \varepsilon_{i,1-t}$ for $i = 1, \dots, N$. It follows that in this case $E(y_{i,1}) = E(\mu_i)$ as well. Therefore, when $\rho = 1$, $E(y_{i,1}) = 0$ and $\sigma_1^2 = \text{Var}(y_{i,1}^2) = E(y_{i,1}^2) = \sigma_\mu^2 + (1 - S)\sigma^2$. In the asymptotic analysis we allow $1 - S$ and thereby σ_1^2 to be related to N . That is, in the unit root case we assume that

$$E(y_{i,1}) = 0 \text{ and } \sigma_1^2 = E(y_{i,1}^2) = \bar{\sigma}_1^2 N^{2b} + o(N^{2b}), \text{ for } i = 1, \dots, N, \quad (10)$$

where $\bar{\sigma}_1^2$ is a constant and $0 \leq b \leq 1/2$. The parameter b determines the relation between σ_1^2 and N . If the $\{y_{i,t}\}$ actually began close to the date of the first observation and σ_1^2/σ^2 is not very large whereas N is large, then b is approximately zero; however, if the $\{y_{i,t}\}$ began a long time before $t = 1$ so that σ_1^2/σ^2 is large whereas N is not very large, then it is more reasonable to assume that b is closer to $1/2$.

We are now able to give a formal definition of the *inclusive* panel AR(1) model:

Definition 1 *The inclusive panel AR(1) model with random effects is characterized by (1)-(2), assumptions (3)-(7), and either assumptions (8) and (9) when $|\rho| < 1$, or assumption (10) when $\rho = 1$.*

If we restrict the parameter space to $|\rho| < 1$, we obtain the *covariance stationary* panel AR(1) model with random effects.

2.1 Moment conditions for the covariance stationary panel AR(1) model

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

Arellano and Bond (1991) have derived the following $(T - 1)(T - 2)/2$ basic moment conditions from assumptions (4)-(6):

$$E(y_{i,t-s} \Delta v_{i,t}) = 0 \text{ for } s = 2, \dots, t - 1 \text{ and } t = 3, \dots, T, \quad (11)$$

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

If mean-stationarity, (8), holds as well, we can add $T - 2$ moment conditions to the basic moment conditions (see Arellano and Bover, 1995):

$$E(v_{i,t} \Delta y_{i,t-1}) = 0 \text{ for } t = 3, \dots, T. \quad (12)$$

The estimator that exploits the moment conditions in both (11) and (12) is known as the System (SYS) estimator.

Adding the assumption of homoskedasticity, (7), implies another $T - 2$ moment conditions (see Ahn and Schmidt, 1995):

$$E(y_{i,t}v_{i,t} - y_{i,t-1}v_{i,t-1}) = 0 \text{ for } t = 3, \dots, T. \quad (13)$$

Notice that all the above moment conditions are linear in ρ .

Finally, adding the “covariance stationarity assumption” (9) yields one more moment condition (see Ahn and Schmidt, 1997): ²

$$E(y_{i,1}^2(1 - \rho)^2 + y_{i,2}\Delta v_{i,3}(1 - \rho)/(1 + \rho) - v_{i,3}v_{i,2}) = 0. \quad (14)$$

Notice that this moment condition is nonlinear in ρ . Therefore, computing the value of the GMM estimator of Ahn and Schmidt (1997) for the covariance stationary panel AR(1) model, which exploits all the moment conditions in (11)-(14), requires numerical optimization.

3 An efficient linear GMM estimator for the covariance stationary panel AR(1) model

We define the Linear GMM (LGMM) estimator for ρ in the covariance stationary panel AR(1) model as the estimator that exploits the $T(T + 1)/2 - 3$ moment conditions specified in (11)-(13), which are all linear in ρ , and a new moment condition, which is implied by covariance stationarity of $\{y_{i,t}\}$:

$$E((\rho - 1)(\Delta y_{i,2})^2 - 2\Delta y_{i,2}\Delta y_{i,3}) = 0. \quad (15)$$

Notice that this moment condition is also linear in ρ . Thus the LGMM estimator exploits (15) instead of the nonlinear moment condition (14), but otherwise exploits the same moment conditions as the Ahn-Schmidt estimator for ρ in the covariance stationary panel AR(1) model.

We have the following result:

Theorem 1 *The two-step Optimal Linear GMM estimator and the two-step Optimal Ahn-Schmidt GMM estimator for the covariance stationary panel AR(1) model are asymptotically equivalent.*

It should be noted that there is no feasible one-step optimal weight matrix for the moment conditions exploited by these estimators.

Clearly, the LGMM estimator is more easily calculated than the nonlinear GMM estimator due to Ahn and Schmidt, and its properties are more easily investigated. For instance, under normality it can easily be shown that (15) is not redundant relative to the moment conditions specified in (11)-(13) when an optimal weight matrix is used. Using arguments similar to those in Ahn and Schmidt (1995), it can also easily be shown that the Optimal LGMM (OLGMM) estimator attains the semi-parametric efficiency lowerbound for the stationary panel AR(1) model with random effects. Moreover, under normality, the OLGMM estimator is asymptotically equivalent to the Random Effects MLE for ρ in the stationary panel AR(1) model and attains the Cramér-Rao lowerbound when $|\rho| < 1$. We will see in section 4 that the LGMM estimator also exists for $\rho = 1$ unlike the Random Effects MLE.

As a computationally simpler alternative to their nonlinear estimator, Ahn and Schmidt (1997) proposed a linearized version of their estimator. The optimal version of this estimator is given by $\hat{\rho}_{OLASGMM} = \check{\rho} - [M_N(\check{\rho})' \hat{\Omega}_N^{-1} M_N(\check{\rho})]^{-1} M_N(\check{\rho})' \hat{\Omega}_N^{-1} m_N(\check{\rho})$, where $\check{\rho}$ is an initial \sqrt{N} -consistent estimator for ρ , $m_N(\rho) = N^{-1} \sum_{i=1}^N m_{N,i}(\rho)$ is the vector of sample moment conditions, $M_N(\check{\rho}) = (dm_N(\rho)/d\rho)|_{\check{\rho}}$ and $\hat{\Omega}_N = N^{-1} \sum_{i=1}^N m_{N,i}(\check{\rho}) m_{N,i}(\check{\rho})'$. The OLGMM estimator and the OLASGMM estimator are asymptotically equivalent when $|\rho| < 1$, but have different finite sample properties.

4 Asymptotic properties of the OLGMM estimator when $\rho = 1$

We first note some properties of the moment conditions that are implied by the inclusive panel AR(1) model when $\rho = 1$. Without loss of generality we assume that $T = 3$. Then each of (11)-(13) contains exactly one moment condition and the covariance stationary panel AR(1) model implies a total of four moment conditions which depend only on ρ :

$$E(m_k(\rho)) = 0, \quad k = 1, \dots, 4, \quad \text{where} \tag{16}$$

$$m_1(\rho) = y_{i,1} \Delta v_{i,3} = y_{i,1} (\Delta y_{i,3} - \rho \Delta y_{i,2}),$$

$$m_2(\rho) = (\varepsilon_{i,3} + (1 - \rho) \mu_i) \Delta y_{i,2} = (y_{i,3} - \rho y_{i,2}) \Delta y_{i,2},$$

$$m_3(\rho) = y_{i,3} (\varepsilon_{i,3} + (1 - \rho) \mu_i) - y_{i,2} (\varepsilon_{i,2} + (1 - \rho) \mu_i) = y_{i,3} (y_{i,3} - \rho y_{i,2}) - y_{i,2} (y_{i,2} - \rho y_{i,1}),$$

$$m_4(\rho) = (\rho - 1) (\Delta y_{i,2})^2 - 2 \Delta y_{i,2} \Delta y_{i,3}.$$

Let $m(\rho) = (m_1(\rho), m_2(\rho), m_3(\rho), m_4(\rho))'$. The moment conditions in $E(m(\rho)) = 0$ are also valid for $\rho = 1$. We have the following result on identification when $\rho = 1$ (cf Arellano and Bover, 1995):

Lemma 2 *All the moment conditions that are exploited by the Linear GMM estimator identify the unit root except for the moment conditions due to Arellano and Bond (1991) given in (11).*

However, some elements of the vector $E(\frac{dm(\rho)}{d\rho})$ are discontinuous at $\rho = 1$.

Lemma 3 *Let $\{y_{i,t}\}$ satisfy the inclusive model. Then $\lim_{\rho \uparrow 1} E(\frac{dm(\rho)}{d\rho}) \neq E(\frac{dm(\rho)}{d\rho}|_{\rho=1})$.*

We will see below that the discontinuity of $E(\frac{dm(\rho)}{d\rho})$ accounts for the discontinuity of the asymptotic distributions of the OLGMM estimator and the System estimator at $\rho = 1$.

Let us now consider $p_4(\rho) = 2m_2(\rho) + m_4(\rho) = (1 - \rho)(y_{i,2}^2 - y_{i,1}^2)$. Since $p_4(\rho)$ is a linear combination of $m_2(\rho)$ and $m_4(\rho)$, it must be true that $E(p_4(\rho)) = 0$ for $-1 < \rho \leq 1$. Indeed, when $|\rho| < 1$, $E(p_4(\rho)) = 0$ due to covariance stationarity of $\{y_{i,t}\}$. However, note that for $\rho = 1$, we not only have $E(p_4(\rho)) = 0$ but as a matter of fact $p_4(\rho) = 0$, i.e. $p_4(\rho) = 0$ holds deterministically at $\rho = 1$. Moreover, $E(dp_4(\rho)/d\rho|_{\rho=1}) = -E(y_{i,2}^2 - y_{i,1}^2|_{\rho=1}) = -\sigma^2 \neq 0$. Therefore, when $\rho = 1$, an estimator that gives all weight to $E(p_4(\rho)) = 0$ would be an optimal GMM estimator for ρ . The value of such an estimator equals 1 with probability one. However, when $|\rho| < 1$, $E(dp_4(\rho)/d\rho) = 0$ and therefore a GMM estimator that only exploits $E(p_4(\rho)) = 0$ would be inconsistent in this case. Nevertheless, GMM estimators that combine $E(p_4(\rho)) = 0$ with (some) other linearly independent moment condition(s) contained in $E(m(\rho)) = 0$ are consistent for both $|\rho| < 1$ and $\rho = 1$. It follows that one can construct a GMM estimator for ρ which is consistent for $-1 < \rho \leq 1$ and has an arbitrarily small asymptotic variance when $\rho = 1$ by choosing the weight for $E(p_4(\rho)) = 0$ sufficiently high relative to the weight given to other moment conditions. In particular, there exist GMM estimators for ρ that have an asymptotic variance that is smaller than the Cramér-Rao lowerbound for $\rho = 1$ (which is given in the proof of corollary 5). Such estimators are known as superefficient estimators. We will see that the Iterated OLGMM estimator (cf Hansen et al., 1996) is a superefficient estimator when $\rho = 1$ and $b = 0$, i.e. $\sigma_1^2 = E(y_{i,1}^2) = \bar{\sigma}_1^2$.

We now consider the asymptotic distribution of the OLGMM estimator when $\rho = 1$. We first note that the set of moment conditions contained in $E(m(\rho)) = 0$ is equivalent to the set of moment conditions contained in $E(p(\rho)) = 0$, where $p(\rho) = [p_1(\rho), p_2(\rho), p_3(\rho), p_4(\rho)]' = [m_1(\rho)/\sigma_1, m_2(\rho), m_3(\rho) - m_1(\rho), p_4(\rho)]'$. We have scaled $m_1(\rho)$ by σ_1 to guarantee that $N^{-1/2} \sum_{i=1}^N p_i(1)$ converges

in distribution regardless of our assumption about σ_1 , that is, about the value of b . We let $\widehat{\rho}_1$ denote an initial N^a consistent estimator for ρ which can be used to construct the asymptotically optimal weight matrix for the LGMM estimator: $N^a(\widehat{\rho}_1 - 1) \xrightarrow{d} \mathcal{K}$, with $a \geq 1/2$ and $\mathcal{K} \neq 0$. Examples of estimators that can be used as initial estimators for ρ are the System estimator and the First Difference MLE for the inclusive panel AR(1) model. Note that $b \leq a$. Finally, we strengthen the assumptions made in section 2 by including the following:

Additional Assumptions:

$$E(\varepsilon_{i,t}^4) = \kappa\sigma^4 < \infty, \text{ for } i = 1, \dots, N \text{ and } t = 2, \dots, T; \quad (17)$$

$$y_{i,1} - \mu_i, \mu_i, \varepsilon_{i,s} \text{ and } \varepsilon_{i,t} \text{ are jointly independent for } i = 1, \dots, N \text{ and } s \neq t, s, t = 2, \dots, T. \quad (18)$$

These additional assumptions are stronger than necessary but made for convenience. Let $F = \text{diag}\{1 \ 1 \ 1 \ \mathcal{K}\}$. Then we have the following result:

Theorem 4 *Let $\rho = 1$ and $T = 3$. Then $N^{-1/2} \sum_{i=1}^N p_i(1) \xrightarrow{d} \mathcal{A}$ and $N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho} \xrightarrow{d} \mathcal{B}$, where \mathcal{A} and \mathcal{B} are given in lemmata 6 and 7, respectively. $\mathcal{A}_4 = 0$ and $\mathcal{B}_1 = 0$. Furthermore, if $b < 1/2$, then $\mathcal{B}_4 = -\sigma^2$, while if $b = 1/2$, then $\mathcal{B}_4 \sim N(-\sigma^2, 4\overline{\sigma}_1^2\sigma^2)$. Finally, if $b < a$, then $N^{1/2+a-b}(\widehat{\rho}_{OLGMM} - 1) \xrightarrow{d} (\mathcal{B}_4\mathcal{V}^{4,4})^{-1}(\mathcal{V}^{4,\cdot}\mathcal{A})\mathcal{K}$, while if $b = a = 1/2$, then $N^{1/2}(\widehat{\rho}_{OLGMM} - 1) \xrightarrow{d} (\mathcal{B}'F^{-1}\mathcal{V}^{-1}F^{-1}\mathcal{B})^{-1}\mathcal{B}'F^{-1}\mathcal{V}^{-1}F^{-1}\mathcal{A}$, where the values of the elements of \mathcal{V} depend on the values of a and b . There are three cases: $b = 0$, $0 < b < a$, and $b = a$. The values of the elements of \mathcal{V} in these cases are implicitly given in the proof of lemma 8.*

Note that the asymptotic distribution of the OLGMM estimator depends on the asymptotic distribution of $\widehat{\rho}_1$, i.e. on \mathcal{K} , whatever the values of a and b .

When $\rho = 1$ and $b < a$, the OLGMM estimator is superconsistent. In addition, we have the following result for the Iterated OLGMM estimator:

Corollary 5 *When $\rho = 1$ and $b = 0$, the Iterated OLGMM estimator is superefficient.*

When $0 < b < 1/2$, the Iterated OLGMM estimator is *not* superefficient because in this case not only the asymptotic variance of the Iterated OLGMM estimator tends to zero when the number of iterations goes to infinity but also the Cramér-Rao lowerbound tends to zero since $\lim_{N \rightarrow \infty} E(y_{i,1}^2) = \infty$ when $b > 0$.

When $\rho = 1$ and $b < a$, it follows from lemma 8, which is given in the appendix, and the fact that $\lim_{N \rightarrow \infty} N^{b-a} = 0$ that $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N p_{k,i}(\hat{\rho}_1) p_{4,i}(\hat{\rho}_1) = 0$ for $k = 1, 2, 3, 4$.³ This implies that asymptotically the so-called optimal weight matrix for $E(p(\rho)) = 0$, i.e. $[N^{-1} \sum_{i=1}^N p_i(\hat{\rho}_1) p_i(\hat{\rho}_1)']^{-1}$, gives in principle all weight to $E(p_4(\rho)) = 0$ when $\rho = 1$ and $b < a$. Hence $N^{2(b-a)} [N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}]' \times [N^{-1} \sum_{i=1}^N p_i(\hat{\rho}_1) p_i(\hat{\rho}_1)']^{-1} [N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}] \xrightarrow{d} \mathcal{C}^2 \mathcal{V}^{4,4} \mathcal{K}^{-2}$ with $\mathcal{C} = -\sigma^2$ when $b < 1/2$, and $\mathcal{C} = \mathcal{B}_4$ when $b = 1/2$. However, since $p_{4,i}(1) = 0$ and hence $\mathcal{A}_4 = 0$, the first three elements of \mathcal{A} still enter the asymptotic distribution of the OLGMM estimator when $b < a$, that is $N^{(b-a)} [N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}]' [N^{-1} \sum_{i=1}^N p_i(\hat{\rho}_1) p_i(\hat{\rho}_1)']^{-1} [N^{-1/2} \sum_{i=1}^N p_i(1)] \xrightarrow{d} (\mathcal{C} \mathcal{V}^{4,4} \mathcal{A}) \mathcal{K}^{-1}$.

When $\rho = 1$ and $b = a = 1/2$, the weight matrix $[N^{-1} \sum_{i=1}^N p_i(\hat{\rho}_1) p_i(\hat{\rho}_1)']^{-1}$ is no longer asymptotically optimal because it also gives weight to moment conditions other than $E(p_4(\rho)) = 0$. This is a consequence of the fact that it makes use of an initial estimator of ρ , i.e. $\hat{\rho}_1$, rather than its true value.⁴

The results stated in lemma 8 in the appendix as well as the results discussed above, notably theorem 4 and corollary 5, can be generalized to allow for arbitrary value of T . When $T > 3$ the moment conditions in $E(p(\rho)) = 0$ and in particular $E(p_4(\rho)) = 0$ are still valid. However, there are $T(T+1)/2 - 6$ additional moment conditions which resemble $E(p_1(\rho)) = 0$, $E(p_2(\rho)) = 0$, and $E(p_3(\rho)) = 0$ but are different with respect to their timing. Therefore when $T > 3$ the results will be qualitatively the same as the aforementioned results for $T = 3$.

Finally, when $\rho = 1$ the two-step Optimal Linear GMM estimator and the two-step Optimal Ahn-Schmidt GMM estimator are no longer asymptotically equivalent. A proof for this assertion can be found at the end of the appendix. It turns out that adding $(\rho - 1)$ times a linear combination of p_1 , p_2 and p_3 to p_4 while keeping the other moment conditions will in general lead to a different non-normal asymptotic distribution of the resulting GMM estimator although doing so will not affect its rate of convergence. In particular, when $\rho = 1$ and $b = 0$, the Iterated Optimal Ahn-Schmidt GMM estimator is also superefficient.

5 Monte Carlo results

In a Monte Carlo study we have compared the finite sample properties of the following GMM estimators: the 2-step optimal System estimator (SYS2), two 2-step optimal LGMM estimators (LGMM2a/b), an optimal version of the Linearized GMM estimator due to Ahn and Schmidt (LASGMM2), an iterative 5-step optimal LGMM estimator (LGMM5b), and an iterative 10-step optimal LGMM estimator (LGMM10b).⁵

The SYS2 estimator uses the inverse of the sample covariance of the vector of moment conditions evaluated at a first-step System estimate of ρ (SYS1) as weight matrix in the second step, where the SYS1 estimator uses a weight matrix that would be asymptotically optimal if $\sigma_\mu^2 = 0$.

The full set of linear moment conditions for the inclusive model can be written as $E(Z_i' \tilde{v}_i(\rho)) = 0$, where $\tilde{v}_i(\rho) = ((\Delta v_i)' \quad v_i' \quad [(\rho - 1)(\Delta y_{i,2})^2 - 2\Delta y_{i,2}\Delta y_{i,3}])'$ with $\Delta v_i(\rho) = (\Delta v_{i,3} \dots \Delta v_{i,T})'$ and $v_i(\rho) = (v_{i,2} \dots v_{i,T})'$, and where Z_i is a matrix which contains the ‘instruments’ for the elements of $\tilde{v}_i(\rho)$.⁶ It follows that there exists no feasible 1-step optimal weight matrix for the LGMM estimator even when $\sigma_\mu^2 = 0$. However, one could use the LGMM estimator (LGMM1) which uses the weight matrix $(N^{-1} \sum_{i=1}^N Z_i' Z_i)^{-1}$, as a first-step estimator. The LGMM2a and LGMM2b estimators use $[N^{-1} \sum_{i=1}^N Z_i' \tilde{v}_i(\hat{\rho}_1) \tilde{v}_i'(\hat{\rho}_1) Z_i]^{-1}$ as weight matrix in the second step with $\hat{\rho}_1 = \hat{\rho}_{LGMM1}$ and $\hat{\rho}_1 = \hat{\rho}_{SYS1}$, respectively. The LASGMM2 estimator uses $\hat{\rho}_{SYS1}$ as the initial consistent estimator everywhere in its formula, that is both in the linear approximation and in the optimal weight matrix. Finally, the iterative estimators, LGMM5b and LGMM10b, continue from the LGMM2b estimator and are obtained by re-estimating the weight matrix using $\hat{\rho}$ from the previous iteration. To avoid numerical problems when computing the inverses of the weight matrices used by the iterative estimators, which might occur when the value of ρ is close to one, our computer program used moment conditions like $E(p(\rho)) = 0$ rather than $E(m(\rho)) = 0$. In addition, the factor $\hat{\rho} - 1$ in $p_{4,i}(\hat{\rho})$ was taken out and the remainder, i.e. $y_{i,2}^2 - y_{i,1}^2$, was scaled by its standard error before computing the inverses of the weight matrices. If $|\hat{\rho} - 1| < 0.0001$, the program put $\hat{\rho} = 0.9999$.

We have generated artificial data from an inclusive model with $\sigma^2 = \sigma_\mu^2 = 1$ for several values of ρ between 0 and 1 whilst assuming normality of the data. For the unit root case we have generated two sets of results: for one set the initial observations were drawn from the $N(0, 1)$ distribution and for the other set they were drawn from $N(0, 50)$. The first set corresponds to the case ‘ $b \approx 0$ ’ while the second set corresponds to the case ‘ $b \approx 1/2$ ’. The panels have dimensions $N = 100$ or

500 and $T = 4$ or 7 . The number of replications was 5000.

The Monte Carlo results are presented in table 1. We first discuss the results for the 2-step estimators and $\rho < 1$. They indicate that the LGMM2b estimator displays better finite sample properties than the SYS2 estimator and also performs better than the LGMM2a estimator, especially when $N = 100$, $T = 7$ and ρ is close to one. However, when $T = 7$ and ρ is close to one, the SYS2 estimator outperforms the LGMM2a estimator despite the fact that the latter exploits more moment conditions. The rmse of the LGMM2b estimator is up to 30% lower than the rmse of the SYS2 estimator when $0.9 \leq \rho < 1$. Finally, when $T = 4$ or when $T = 7$ and $N = 500$ the rmse (and in most cases also the bias) of the LASGMM2 estimator is larger than that of the LGMM2b estimator and in many cases, notably when $\rho < 0.9$, the rmse of the LASGMM2 estimator is also larger than the rmse of the SYS2 estimator.

We now turn to the Monte Carlo results for the unit root case and for the iterative LGMM5b and LGMM10b estimators. Table 1 shows that when $\rho = 1$ the LGMM2b estimator still has better finite sample properties than the LGMM2a and SYS2 estimators. Furthermore, the bias and the rmse of the LGMM estimators are larger when the variance of the initial observations is larger. The latter finding does not hold for the LASGMM2 estimator.

The superefficiency of the Iterated LGMM estimator when $b \approx 0$ can clearly be seen in the rows corresponding to $\rho = 1.00^*$. When $N = 500$, all the 5000 LGMM10b estimates lied in fact between 0.9997 and 1.0002. On the other hand when $\rho < 1$ or when $\rho = 1$ with $b \approx 1/2$ ($\rho = 1.00^\dagger$), the rmse of the iterated LGMM estimators are higher than the rmse of the LGMM2b estimator. In particular, when the number of iterations increases, the rmse of the iterated estimator increases and so does its mean when $T = 4$ or when $T = 7$ and $N = 500$. As a consequence the bias of the iterated estimator also increases unless ρ is close to one.

Finally the results in table 1 suggest that the rmse's of the SYS2 and the LGMM estimators are discontinuous at $\rho = 1$. These discontinuities are partly related to the fact that some of the elements of $E(\frac{dm(\rho)}{d\rho})$ are discontinuous at $\rho = 1$ (cf lemma 3). As a result, we have for instance $\lim_{\rho \uparrow 1} Asyvar(\hat{\rho}_{SYS2}) = 4 \times Asyvar(\hat{\rho}_{SYS2} | \rho = 1, b = 0) = 8/[(T - 1)(T - 2)]$. Moreover, when $|\rho| < 1$ the optimal LGMM estimators are \sqrt{N} -consistent and asymptotically normally distributed whereas when $\rho = 1$ they have a non-normal asymptotic distribution and are in fact superconsistent if $b < a$.

T/N	ρ	SYS2		LGMM2a		LASGMM2		LGMM2b		LGMM5b		LGMM10b	
		bias	rmse	bias	rmse	bias	rmse	bias	rmse	bias	rmse	bias	rmse
4 / 100	0.00	0.014	0.101	0.008	0.101	0.033	0.144	0.010	0.099	0.0084	0.1036	0.0089	0.1061
	0.30	0.009	0.114	0.014	0.118	0.007	0.139	0.015	0.114	0.0210	0.1276	0.0245	0.1381
	0.50	0.001	0.118	0.018	0.126	-0.006	0.135	0.016	0.120	0.0315	0.1441	0.0404	0.1624
	0.80	-0.013	0.126	0.026	0.129	-0.022	0.132	0.006	0.113	0.0318	0.1327	0.0441	0.1449
	0.90	-0.023	0.135	0.009	0.120	-0.024	0.129	-0.002	0.109	0.0159	0.1161	0.0241	0.1230
	0.95	-0.026	0.147	0.003	0.120	-0.025	0.134	-0.009	0.109	0.0039	0.1139	0.0097	0.1206
	0.98	-0.031	0.156	-0.007	0.117	-0.017	0.129	-0.013	0.111	-0.0058	0.1250	-0.0030	0.1444
	0.99	-0.027	0.157	-0.010	0.114	-0.018	0.124	-0.012	0.103	-0.0077	0.1227	-0.0078	0.1544
	1.00*	0.004	0.060	-0.003	0.025	-0.002	0.048	0.000	0.013	-0.0001	0.0012	0.0000	0.0007
	1.00†	-0.008	0.098	-0.008	0.090	-0.009	0.085	-0.006	0.077	-0.0048	0.0960	-0.0036	0.1176
4 / 500	0.00	0.002	0.043	0.001	0.041	0.007	0.072	0.001	0.040	0.0010	0.0400	0.0010	0.0400
	0.30	0.002	0.050	0.003	0.046	0.003	0.071	0.003	0.046	0.0027	0.0459	0.0027	0.0459
	0.50	0.000	0.052	0.005	0.049	-0.006	0.067	0.004	0.049	0.0047	0.0500	0.0048	0.0501
	0.80	-0.003	0.053	0.012	0.057	-0.009	0.062	0.008	0.054	0.0167	0.0639	0.0211	0.0722
	0.90	-0.004	0.055	0.015	0.058	-0.010	0.060	0.008	0.052	0.0204	0.0609	0.0259	0.0665
	0.95	-0.003	0.057	0.010	0.052	-0.010	0.063	0.004	0.049	0.0131	0.0513	0.0147	0.0530
	0.98	-0.007	0.064	0.004	0.050	-0.008	0.064	-0.001	0.049	0.0037	0.0492	0.0060	0.0509
	0.99	-0.012	0.076	0.000	0.052	-0.007	0.069	-0.004	0.051	0.0002	0.0511	0.0019	0.0549
	1.00*	0.002	0.026	0.000	0.004	0.000	0.021	0.000	0.002	-0.0001	0.0001	0.0000	0.0000
	1.00†	0.001	0.028	-0.001	0.027	-0.001	0.024	-0.001	0.022	-0.0010	0.0241	-0.0009	0.0246
7 / 100	0.00	0.016	0.061	-0.001	0.065	0.059	0.103	0.018	0.064	0.0035	0.0681	0.0020	0.0695
	0.30	0.013	0.065	-0.002	0.070	0.038	0.086	0.013	0.064	0.0053	0.0721	0.0044	0.0743
	0.50	0.009	0.066	-0.004	0.072	0.020	0.073	0.011	0.064	0.0083	0.0740	0.0084	0.0775
	0.80	-0.002	0.062	-0.009	0.079	-0.004	0.060	0.005	0.058	0.0129	0.0764	0.0191	0.0884
	0.90	-0.008	0.062	-0.015	0.080	-0.009	0.055	0.000	0.055	0.0077	0.0697	0.0128	0.0784
	0.95	-0.010	0.063	-0.023	0.080	-0.009	0.051	-0.006	0.051	-0.0030	0.0627	-0.0010	0.0721
	0.98	-0.011	0.062	-0.026	0.077	-0.008	0.044	-0.007	0.048	-0.0066	0.0642	-0.0072	0.0822
	0.99	-0.007	0.055	-0.026	0.071	-0.005	0.035	-0.006	0.041	-0.0061	0.0621	-0.0082	0.0882
	1.00*	0.002	0.028	-0.007	0.020	-0.001	0.023	-0.001	0.011	-0.0001	0.0017	0.0000	0.0006
	1.00†	-0.001	0.034	-0.016	0.052	-0.002	0.022	-0.002	0.028	-0.0029	0.0405	-0.0042	0.0561
7 / 500	0.00	0.001	0.023	0.000	0.024	0.016	0.044	0.001	0.023	0.0006	0.0233	0.0006	0.0233
	0.30	0.001	0.026	0.000	0.025	0.010	0.041	0.001	0.025	0.0010	0.0253	0.0010	0.0253
	0.50	0.001	0.027	0.000	0.025	0.004	0.037	0.001	0.025	0.0012	0.0256	0.0012	0.0256
	0.80	0.000	0.026	0.002	0.027	-0.003	0.030	0.003	0.025	0.0039	0.0264	0.0041	0.0270
	0.90	-0.002	0.026	0.004	0.031	-0.004	0.027	0.003	0.025	0.0067	0.0302	0.0090	0.0344
	0.95	-0.002	0.025	0.005	0.031	-0.004	0.025	0.002	0.024	0.0070	0.0283	0.0100	0.0314
	0.98	-0.002	0.027	0.001	0.030	-0.003	0.025	-0.001	0.023	0.0016	0.0249	0.0032	0.0266
	0.99	-0.003	0.029	-0.003	0.031	-0.003	0.024	-0.001	0.022	0.0005	0.0254	0.0017	0.0291
	1.00*	0.000	0.012	-0.001	0.003	0.000	0.010	0.000	0.001	-0.0001	0.0001	0.0000	0.0000
	1.00†	0.000	0.012	-0.001	0.013	0.000	0.009	0.000	0.008	-0.0003	0.0106	-0.0001	0.0112

Table 1: $\sigma^2 = \sigma_\mu^2 = 1$, 5000 replications; *: $y_{i,1} \sim N(0, 1)$; †: $y_{i,1} \sim N(0, 50)$.

A Appendix

Proof of theorem 1: Both the OLGMM estimator and the Optimal Ahn-Schmidt GMM estimator are defined for $T \geq 3$. For any value of $T \geq 3$, the only difference between these estimators is that the OLGMM estimator exploits the linear moment condition in (15) instead of the nonlinear moment condition in (14). Therefore, in order to prove that these estimators are asymptotically equivalent, it is sufficient to show for $T = 3$ that the set of moment conditions exploited by the OLGMM estimator can be obtained by using a one-to-one transformation of the set of moment conditions exploited by the optimal Ahn-Schmidt estimator.

When $T = 3$, the OLGMM estimator exploits $E(m(\rho)) = 0$, where $m_k(\rho)$, $k = 1, 2, 3, 4$, are defined in (16). Let us define $y_i^2 = [y_{i,1}^2 \ y_{i,1}y_{i,2} \ y_{i,1}y_{i,3} \ y_{i,2}^2 \ y_{i,2}y_{i,3} \ y_{i,3}^2]'$. Then there exists a (4×6) matrix G such that $m(\rho) = Gy_i^2$. Note that G depends on ρ and that $\text{rank}(G) = 4$ unless $\rho = 1$.

Let S_3 be a (3×4) selection matrix with $[S_3]_{i,i} = 1$, $i = 1, \dots, 3$, and zeros elsewhere. S_3 selects the first three rows of a matrix with four rows. When $T = 3$, the Ahn-Schmidt estimator exploits $E(S_3 Gy_i^2) = 0$ and in addition the moment condition in (14) which can be written as $E(hy_i^2) = 0$, where $h = [\rho^3 - \rho^2 - \rho + 1 \quad -\rho^3 - 2\rho^2 + \rho \quad \rho^2 + \rho \quad 2\rho^2 + \rho - 1 \quad -2\rho \quad 0]$. Thus when $T = 3$ the Ahn-Schmidt estimator exploits $E(Hy_i^2) = 0$, where $H' = [G'S_3' \ h']$.

Let $K' = [S_3' \ k']$, where $k = [\rho(\rho - 1) \quad -2(1 + \rho) \quad 0 \quad -1]$. It is easily verified that $\text{rank}(K) = 4$, $h = kG$, and thus $E(Hy_i^2) = KE(Gy_i^2)$, which completes the proof. \square

Proof of lemma 2: The result follows immediately from the fact that all these moment conditions in $m(\rho)$ are linear in ρ and $E(\frac{dm(\rho)}{d\rho}|_{\rho=1}) = [0 \quad -\sigma^2 \quad -\sigma^2 \quad \sigma^2]'$. \square

Proof of lemma 3: $E(\frac{dm(\rho)}{d\rho}|_{\rho=1}) = [0 \quad -\sigma^2 \quad -\sigma^2 \quad \sigma^2]'$ while under stationarity $E(\frac{dm(\rho)}{d\rho}) = [\frac{1}{1+\rho}\sigma^2 \quad -\frac{1}{1+\rho}\sigma^2 \quad 0 \quad \frac{2}{1+\rho}\sigma^2]'$. Hence $\lim_{\rho \uparrow 1} E(\frac{dm(\rho)}{d\rho}) \neq E(\frac{dm(\rho)}{d\rho}|_{\rho=1})$. \square

Lemma 6 *Let $\rho = 1$. Then*

$$\begin{aligned} N^{-1/2} \sum_{i=1}^N p_{1,i}(1) &= N^{-1/2} \sum_{i=1}^N (y_{i,1} \Delta \varepsilon_{i,3} / \sigma_1) \xrightarrow{d} N(0, 2\sigma^2), \\ N^{-1/2} \sum_{i=1}^N p_{2,i}(1) &= N^{-1/2} \sum_{i=1}^N (\varepsilon_{i,2} \varepsilon_{i,3}) \xrightarrow{d} N(0, \sigma^4), \\ N^{-1/2} \sum_{i=1}^N p_{3,i}(1) &= N^{-1/2} \sum_{i=1}^N (\varepsilon_{i,2} \varepsilon_{i,3} - \varepsilon_{i,2}^2 + \varepsilon_{i,3}^2) \xrightarrow{d} N(0, (2\kappa - 1)\sigma^4), \\ N^{-1/2} \sum_{i=1}^N p_{4,i}(1) &= 0. \end{aligned}$$

Proof of lemma 6: trivial.

Lemma 7 *Let $\rho = 1$. Then*

	<i>if $b < 1/2$:</i>	<i>if $b = 1/2$:</i>
$N^{-1/2} \sum_{i=1}^N dp_{1,i}(\rho)/d\rho \xrightarrow{d}$	$N(0, \sigma^2)$	$N(0, \sigma^2)$
$N^{-1} \sum_{i=1}^N dp_{2,i}(\rho)/d\rho \xrightarrow{d}$	$-\sigma^2$	$-\sigma^2 + N(0, \bar{\sigma}_1^2 \sigma^2)$
$N^{-1} \sum_{i=1}^N dp_{3,i}(\rho)/d\rho \xrightarrow{d}$	$-\sigma^2$	$-\sigma^2 + N(0, \bar{\sigma}_1^2 \sigma^2)$
$N^{-1} \sum_{i=1}^N dp_{4,i}(\rho)/d\rho \xrightarrow{d}$	$-\sigma^2$	$-\sigma^2 + N(0, 4\bar{\sigma}_1^2 \sigma^2)$

Proof of lemma 7: When $\rho = 1$, we have

$$\begin{aligned}
dp_{1,i}(\rho)/d\rho &= -y_{i,1}\Delta y_{i,2}/\sigma_1 = -y_{i,1}\varepsilon_{i,2}/\sigma_1, \\
dp_{2,i}(\rho)/d\rho &= -y_{i,2}\Delta y_{i,2} = -\varepsilon_{i,2}^2 - y_{i,1}\varepsilon_{i,2}, \\
dp_{3,i}(\rho)/d\rho &= -y_{i,3}y_{i,2} + y_{i,2}y_{i,1} + y_{i,1}\Delta y_{i,2} = -y_{i,1}\varepsilon_{i,3} - \varepsilon_{i,2}^2 - \varepsilon_{i,2}\varepsilon_{i,3}, \\
dp_{4,i}(\rho)/d\rho &= -(y_{i,2}^2 - y_{i,1}^2) = -\varepsilon_{i,2}^2 - 2y_{i,1}\varepsilon_{i,2}.
\end{aligned} \tag{19}$$

From this and $\sigma_1^2 = E(y_{i,1}^2) = \bar{\sigma}_1^2 N^{2b} + o(N^{2b})$ with $0 \leq b \leq 1/2$ the lemma follows immediately. \square

Lemma 8 *Let $\rho = 1$. Furthermore, let $F_1 = \text{diag}\{1 \ 1 \ 1 \ N^{b-a}\}$ and $\tilde{p}_i(\rho) = F_1^{-1}p_i(\rho)$. Then $N^{-1} \sum_{i=1}^N \tilde{p}_i(\hat{\rho}_1)\tilde{p}_i'(\hat{\rho}_1) \xrightarrow{d} \mathcal{V}_0 = F\mathcal{V}F$, where $F = \text{diag}\{1 \ 1 \ 1 \ \mathcal{K}\}$ and where the values of the elements of \mathcal{V} depend on the values of a and b . There are three cases: $b = 0$, $0 < b < a$, and $b = a$. In all three cases \mathcal{V} is positive definite symmetric. When $b = a$, \mathcal{V} is a random matrix, whereas when $b < a$, \mathcal{V} is a constant matrix.*

Proof of lemma 8: Recall that $a \geq 1/2$ and $\sigma_1^2 = \bar{\sigma}_1^2 N^{2b} + o(N^{2b})$ with $0 \leq b \leq 1/2$. Thus $b \leq a$.

Linearity of the moment conditions implies that $p_{k,i}(\hat{\rho}_1) = p_{k,i}(1) + (\hat{\rho}_1 - 1)dp_{k,i}(\rho)/d\rho$, $k = 1, 2, 3, 4$. From this we obtain expressions for $p_{k,i}(\hat{\rho}_1)$, $k = 1, 2, 3, 4$, when $\rho = 1$ by substituting the formulae for $p_{k,i}(1)$ and $dp_{k,i}(\rho)/d\rho$, $k = 1, 2, 3, 4$, which are given in lemma 6 and (19), respectively:

$$\begin{aligned}
p_{1,i}(\hat{\rho}_1) &= y_{i,1}\Delta\varepsilon_{i,3}/\sigma_1 + N^{-a}N^a(\hat{\rho}_1 - 1)(-y_{i,1}\varepsilon_{i,2})/\sigma_1, \\
p_{2,i}(\hat{\rho}_1) &= \varepsilon_{i,2}\varepsilon_{i,3} + N^{b-a}N^a(\hat{\rho}_1 - 1)(\bar{\sigma}_1 + o(1))(-y_{i,1}\varepsilon_{i,2} - \varepsilon_{i,2}^2)/\sigma_1, \\
p_{3,i}(\hat{\rho}_1) &= (\varepsilon_{i,2}\varepsilon_{i,3} - \varepsilon_{i,2}^2 + \varepsilon_{i,3}^2) + N^{b-a}N^a(\hat{\rho}_1 - 1)(\bar{\sigma}_1 + o(1))(-y_{i,1}\varepsilon_{i,3} - \varepsilon_{i,2}^2 - \varepsilon_{i,2}\varepsilon_{i,3})/\sigma_1, \\
N^{a-b}p_{4,i}(\hat{\rho}_1) &= N^a(\hat{\rho}_1 - 1)(\bar{\sigma}_1 + o(1))(-\varepsilon_{i,2}^2 - 2y_{i,1}\varepsilon_{i,2})/\sigma_1.
\end{aligned} \tag{20}$$

The expressions in (20) in turn enable us to establish that $N^{-1} \sum_{i=1}^N \tilde{p}_i(\hat{\rho}_1) \tilde{p}'_i(\hat{\rho}_1) = N^{-1} \sum_{i=1}^N F_1^{-1} p_i(\hat{\rho}_1) p'_i(\hat{\rho}_1) F_1^{-1} \xrightarrow{d} \mathcal{V}_0$ when $\rho = 1$, where the values of the elements of \mathcal{V}_0 depend on the values of a and b :

$$\begin{aligned} \text{if } b = 0, \text{ then } \mathcal{V}_0 &= \begin{pmatrix} 2\sigma^2 & 0 & 0 & 2\bar{\sigma}_1\sigma^2\mathcal{K} \\ 0 & \sigma^4 & \sigma^4 & 0 \\ 0 & \sigma^4 & (2\kappa - 1)\sigma^4 & (\kappa - 1)\sigma^4\mathcal{K} \\ 2\bar{\sigma}_1\sigma^2\mathcal{K} & 0 & (\kappa - 1)\sigma^4\mathcal{K} & (\kappa\sigma^4 + 4\bar{\sigma}_1^2\sigma^2)\mathcal{K}^2 \end{pmatrix}, \\ \text{if } 0 < b < a, \text{ then } \mathcal{V}_0 &= \begin{pmatrix} 2\sigma^2 & 0 & 0 & 2\bar{\sigma}_1\sigma^2\mathcal{K} \\ 0 & \sigma^4 & \sigma^4 & 0 \\ 0 & \sigma^4 & (2\kappa - 1)\sigma^4 & 0 \\ 2\bar{\sigma}_1\sigma^2\mathcal{K} & 0 & 0 & 4\bar{\sigma}_1^2\sigma^2\mathcal{K}^2 \end{pmatrix}, \text{ while} \\ \text{if } b = a, \text{ then } \mathcal{V}_0 &= \begin{pmatrix} 2\sigma^2 & \bar{\sigma}_1\sigma^2\mathcal{K} & -\bar{\sigma}_1\sigma^2\mathcal{K} & 2\bar{\sigma}_1\sigma^2\mathcal{K} \\ \bar{\sigma}_1\sigma^2\mathcal{K} & \sigma^4 + \bar{\sigma}_1^2\sigma^2\mathcal{K}^2 & \sigma^4 & 2\bar{\sigma}_1^2\sigma^2\mathcal{K}^2 \\ -\bar{\sigma}_1\sigma^2\mathcal{K} & \sigma^4 & (2\kappa - 1)\sigma^4 + \bar{\sigma}_1^2\sigma^2\mathcal{K}^2 & 0 \\ 2\bar{\sigma}_1\sigma^2\mathcal{K} & 2\bar{\sigma}_1^2\sigma^2\mathcal{K}^2 & 0 & 4\bar{\sigma}_1^2\sigma^2\mathcal{K}^2 \end{pmatrix}. \end{aligned}$$

Note that in all three cases the matrix \mathcal{V}_0 can be factorized as $F\mathcal{V}F$ where $F = \text{diag}\{1 \ 1 \ 1 \ \mathcal{K}\}$ and where the values of the elements of \mathcal{V} depend on the values of a and b . Furthermore, in all three cases \mathcal{V} is positive definite symmetric:

Leading principal minors of \mathcal{V} :			
dim	if $b = 0$:	if $0 < b < a$:	if $b = a$:
4	$2\sigma^{12}(\kappa - 1)((\kappa + 1)\sigma^2 + 4\bar{\sigma}_1^2)$	$8\sigma^{12}\bar{\sigma}_1^2(\kappa - 1)$	$8\sigma^{12}\bar{\sigma}_1^2(\kappa - 1)$
3	$\sigma^{10}(\kappa - 1)(\sigma^2(\kappa + 1) + 8\bar{\sigma}_1^2)$	$8\sigma^{10}\bar{\sigma}_1^2(\kappa - 1)$	$4\sigma^8\bar{\sigma}_1^2(2(\kappa - 1)\sigma^2 + \bar{\sigma}_1^2\mathcal{K}^2)$
2	$\sigma^6((\kappa^2 + \kappa - 1)\sigma^2 + 4(2\kappa - 1)\bar{\sigma}_1^2)$	$4\sigma^6\bar{\sigma}_1^2(2\kappa - 1)$	$4\sigma^4\bar{\sigma}_1^2((2\kappa - 1)\sigma^2 + \bar{\sigma}_1^2\mathcal{K}^2)$
1	$\kappa\sigma^4 + 4\sigma^2\bar{\sigma}_1^2$	$4\sigma^2\bar{\sigma}_1^2$	$4\sigma^2\bar{\sigma}_1^2$

Finally note that when $b = a$, \mathcal{V} is a random matrix, whereas when $b < a$, \mathcal{V} is a constant matrix. \square

Proof of theorem 4: Lemma 6 implies that $N^{-1/2} \sum_{i=1}^N p_i(1) \xrightarrow{d} \mathcal{A}$ with $\mathcal{A}_4 = 0$, while lemma 7 implies that $N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho} \Big|_{\rho=1} \xrightarrow{d} \mathcal{B}$ with $\mathcal{B}_4 = -\sigma^2$ if $b < 1/2$, and $\mathcal{B}_4 \sim N(-\sigma^2, 4\bar{\sigma}_1^2\sigma^2)$ if $b = 1/2$. Moreover, $\mathcal{B}_1 = 0$. Finally, when $\rho = 1$, lemma 8 implies that $N^{-1} \sum_{i=1}^N \tilde{p}_i(\hat{\rho}_1) \tilde{p}'_i(\hat{\rho}_1) \xrightarrow{d} \mathcal{V}_0 = F\mathcal{V}F$, where $\tilde{p}_i(\rho) = F_1^{-1}p_i(\rho)$ with $F_1 = \text{diag}\{1 \ 1 \ 1 \ N^{b-a}\}$. We can now easily verify the following results:

When $b < a$:

$$N^{2(b-a)} [N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}]' [N^{-1} \sum_{i=1}^N p_i(\hat{\rho}_1) p_i(\hat{\rho}_1)']^{-1} [N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}] \xrightarrow{d} \mathcal{B}_4^2 \mathcal{V}^{4,4} \mathcal{K}^{-2}.$$

$$N^{(b-a)} [N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}]' [N^{-1} \sum_{i=1}^N p_i(\hat{\rho}_1) p_i(\hat{\rho}_1)']^{-1} [N^{-1/2} \sum_{i=1}^N p_i(1)] \xrightarrow{d} (\mathcal{B}_4 \mathcal{V}^{4, \cdot} \mathcal{A}) \mathcal{K}^{-1}.$$

When $b = a = 1/2$:

$$[N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}]' [N^{-1} \sum_{i=1}^N p_i(\hat{\rho}_1) p_i(\hat{\rho}_1)']^{-1} [N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}] \xrightarrow{d} \mathcal{B}' F^{-1} \mathcal{V}^{-1} F^{-1} \mathcal{B}.$$

$$[N^{-1} \sum_{i=1}^N \frac{dp_i(\rho)}{d\rho}]' [N^{-1} \sum_{i=1}^N p_i(\hat{\rho}_1) p_i(\hat{\rho}_1)']^{-1} [N^{-1/2} \sum_{i=1}^N p_i(1)] \xrightarrow{d} \mathcal{B}' F^{-1} \mathcal{V}^{-1} F^{-1} \mathcal{A}.$$

From the above results, the claims in the theorem follow straightforwardly. \square

Proof of corollary 5: When $\rho = 1$ the asymptotic variance of the Iterated OLGMM estimator decreases after each iteration because $N^{1/2+a} > N^a$. Therefore the asymptotic variance of the Iterated OLGMM estimator tends to zero when the number of iterations tends to infinity. We need to compare this asymptotic variance with the Cramér-Rao (CR) lowerbound for $\rho = 1$. We derive the CR lowerbound for $\rho = 1$ from the likelihood function corresponding to the model $y_{i,t} = \rho y_{i,t-1} + \varepsilon_{i,t}$, $i = 1, \dots, N$, $t = 2, \dots, T$, where $y_{i,1} \sim i.i.d. N(0, \sigma_1^2)$ with $\sigma_1^2 = \bar{\sigma}_1^2 N^{2b} + o(N^{2b})$, and $\varepsilon_{i,t} | y_{i,1} \sim i.i.d. N(0, \sigma^2)$. The CR lowerbound can be derived from this likelihood function because the corresponding model is nested within any other panel AR model that reduces to the inclusive panel AR(1) model for $\rho = 1$. It is easily verified that the CR lowerbound for $\rho = 1$ is given by $CR = \lim_{N \rightarrow \infty} 2\sigma^2 / [(T-1)((T-2)\sigma^2 + 2\sigma_1^2)]$. Clearly $CR > 0$ when $b = 0$. \square

Non-equivalence of the OLGMM estimator and the Optimal Ahn-Schmidt GMM estimator for the inclusive panel AR(1) model when $\rho = 1$:

Let $E(m_5(\rho)) = 0$ denote the nonlinear moment condition in (14) and let $p_5(\rho) = -(1+\rho)(p_2(\rho) + m_5(\rho))$. Then it is easily verified that $p_5(1) = 0$, $p_5(\rho) = -\rho(\rho-1)\sigma_1 p_1(\rho) + (\rho-1)p_2(\rho) + p_4(\rho)$ and $N^{a-b} p_5(\hat{\rho}_1) = -\hat{\rho}_1 N^a (\hat{\rho}_1 - 1) (\bar{\sigma}_1 + o(1)) p_1(\hat{\rho}_1) + N^a (\hat{\rho}_1 - 1) N^{-b} p_2(\hat{\rho}_1) + N^{a-b} p_4(\hat{\rho}_1)$ when $\rho = 1$. The last expression can be used to derive the asymptotic distribution of the elements of the ‘optimal’ weight matrix corresponding to $E(p_5(\rho)) = 0$. Moreover, when $\rho = 1$, $N^{-1} \sum_{i=1}^N dp_{5,i}(\rho) / d\rho |_{\rho=\hat{\rho}_1} \xrightarrow{d} -\sigma^2$ if $b < 1/2$, while $N^{-1} \sum_{i=1}^N dp_{5,i}(\rho) / d\rho |_{\rho=\hat{\rho}_1} \xrightarrow{d} -\sigma^2 + N(0, 2\bar{\sigma}_1^2 \sigma^2)$ if $b = 1/2$. Therefore, if in $E(p(\rho)) = 0$, we replace $E(p_4(\rho)) = 0$ by $E(p_5(\rho)) = 0$, then we will obtain a different optimal weight matrix for $E(p(\rho)) = 0$, and if $b = 1/2$, $N^{-1} \sum_{i=1}^N dp_i(\rho) / d\rho |_{\rho=\hat{\rho}_1}$ will have a different asymptotic distribution. As a result, when $\rho = 1$, the optimal GMM estimator exploiting $E(p(\rho)) = 0$ with

$E(p_4(\rho)) = 0$ replaced by $E(p_5(\rho)) = 0$ will have a different non-normal asymptotic distribution than the OLGMM estimator but the same rate of convergence. \square

Notes

1. We assume identical distributions across the individuals for convenience.
2. Covariance stationarity of $\{y_{i,t}\}$ also implies restrictions on first moments: $E(\Delta y_{i,t}) = 0$, $t = 2, \dots, T$. However, they do not contain any information on ρ .
3. Note however that $E(p_4(\rho)) = 0$ is *not* redundant (otherwise $E(p_4(\rho)) = 0$ should be dropped). On the contrary, $p_4(\rho) = 0$ holds deterministically at $\rho = 1$ *and only at* $\rho = 1$ (so that $\text{Var}(p_4(1)) = 0$). Therefore $E(p_4(\rho)) = 0$ is extremely informative when $\rho = 1$ and should receive maximal weight rather than zero weight when $\rho = 1$.
4. The weight matrix $[N^{-1} \sum_{i=1}^N p_i(1)p_i(1)']^{-1}$ would still be asymptotically optimal for $\rho = 1$ when $b = a = 1/2$. However, an estimator for ρ that uses this weight matrix would be inconsistent if $|\rho| < 1$.
5. The first two estimators are identical to the GMM estimators for models 2 and 3 in Bond and Windmeijer (2002). That paper presents Monte Carlo simulation results on the finite sample properties of these estimators for the covariance stationary model with $\sigma^2 = (1 - \rho)^2 \sigma_\mu^2 = 1$.
6. The ‘instruments’ corresponding to the two errors in a typical moment condition in (13), i.e. $v_{i,t}$ and $v_{i,t-1}$, are $y_{i,t}$ and $-y_{i,t-1}$ respectively and the instrument corresponding to (15) is 1.

References

- [1] Ahn, S.C. & P. Schmidt (1995) Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5-28.
- [2] Ahn, S.C. & P. Schmidt (1997) Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation. *Journal of Econometrics* 76, 309-321.
- [3] Arellano, M. & 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.

- [4] Arellano, M. & O. Bover (1995) Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68, 29-51.
- [5] Blundell, R.W. & S. Bond (1998) Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.
- [6] Bond, S. & F. Windmeijer (2002) Projection estimators for autoregressive panel data models. *Econometrics Journal* 5.
- [7] Hansen, L.P., J. Heaton & A. Yaron (1996) Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics* 14, 262-280.