# Estimation of dynamic panel data models with a lot of heterogeneity

Hugo Kruiniger*
Durham University

This version: 12 February 2021

JEL classification: C12, C13, C23.

Keywords: Dynamic panel data, Fixed effects, GMM, Local asymptotics, Quasi ML, Redundant moment conditions, Weak moment conditions.

*Address: hugo.kruiniger@durham.ac.uk; Department of Economics, 1 Mill Hill Lane, Durham DH1 3LB, England. This paper was presented at the Econometric Study Group Conference in Bristol, England in July 2018. I would like to thank four referees for very helpful comments.

**Abstract**

The commonly used 1-step and 2-step System GMM estimators for the panel AR(1) model are inconsistent under mean stationarity when the ratio of the variance of the individual effects to the variance of the idiosyncratic errors is unbounded when $N \to \infty$. The reason for their inconsistency is that their weight matrices select moment conditions that do not identify the autoregressive parameter. This paper proposes a new 2-step System estimator that is still consistent in this case provided that $T > 3$. Unlike the commonly used 2-step System estimator, the new estimator uses an estimator of the optimal weight matrix that remains consistent in this case. We also show that the commonly used 1-step and 2-step Arellano-Bond GMM estimators and the Random Effects Quasi MLE remain consistent under the same conditions. To illustrate the usefulness of our new System estimator we revisit the growth study of Levine et al. (2000).

# 1 Introduction

It is well known that the GMM estimators that have been proposed by Arellano and Bond (1991, AB) for panel autoregressive models can have poor finite sample properties when the sum of the autoregressive parameters is close to unity due to a weak instruments problem, i.e., low correlation between the regressors in the first-differenced model, which are first-differences of the lagged dependent variable, and the instruments, which are lagged levels of the dependent variable. Arellano and Bover (1995, Arbov) and Blundell and Bond (1998) proposed the System estimator as a possible solution to this problem when mean stationarity holds. This estimator combines the AB style moment conditions with Arbov style moment conditions, which are based on the model in levels of the data and use lagged differences of the dependent variable as instruments. Simulation evidence in Blundell and Bond (1998) shows that under covariance stationarity the System estimator has much better finite sample properties, i.e., smaller bias and root mean squared error (rmse) than the AB GMM estimator.

The AB GMM estimator can also suffer from a weak instruments problem for a different reason than an autoregressive root being close to one. Considering only one AB style moment condition for a panel AR(1) model and assuming covariance stationarity, Blundell and Bond (1998) derived expressions for the probability limit of the estimator of the slope parameter of the first-stage regression and the so-called concentration parameter. They found that the corresponding AB GMM estimator also has a weak instruments problem when the ratio of the variance of the individual effects and the variance of the disturbances, henceforth the variance ratio, is large. Below we find that this is true more generally, namely under mean stationarity. On the other hand, Hayakawa (2009) found that when the data are not close to mean stationary and the variance ratio becomes large, then the AB GMM estimator that uses instruments in levels performs quite well. He argued that this is because the correlation between the lagged dependent variable and the instruments in levels gets larger owing to the unremoved individual effects, i.e., the instruments in levels become stronger. A high variance ratio is likely to occur in dynamic panel data models that are used for studying economic growth.

Hayakawa (2007) and Bun and Windmeijer (2010) showed that under covariance stationarity GMM estimators for the panel AR(1) model that only exploit Arbov style moment conditions also have a weak instruments problem when the variance ratio is large and the autoregressive parameter $\rho$ is not close to one. As the System estimator combines AB and Arbov style moment conditions, its properties are combinations of those of the AB and Arbov GMM estimators. Bun and Windmeijer (2010) provide simulations results for commonly used versions of the 1-step and 2-step System estimator for $\rho$ under covariance stationarity when $\rho = 0.8$, $N = 200$ and $T = 6$ or 15. They show that the biases of both System estimators increase substantially, to around 0.09, when the variance ratio increases from 1 to 4. We will see that both estimators are in fact inconsistent when the data are mean stationary and the variance ratio tends to infinity, because their weight matrices are such that these estimators effectively exploit moment conditions that depend on levels of the data and do not identify $\rho$.

In this paper we present a necessary condition for large $N$, fixed $T$ consistency of *any* random effects (RE) or fixed effects (FE) estimator for $\rho$. We use the label 'fixed effects' to indicate that we make minimal assumptions about the individual effects and the initial observations. The necessary condition for consistency requires the variance of the deviations of the initial observations from the individual effects to be finite. This condition is also sufficient for consistency of FE estimators for $\rho$, which only depend on differences of the data. However, this condition is in general not sufficient for consistency of RE estimators for $\rho$, which also depend on levels of the data. Nonetheless, we argue that even when the data are mean stationary and the variance ratio is infinite, RE GMM estimators will still be consistent *as long as* they use a suitable weight matrix that effectively only selects moment conditions that only depend on differences of the data, possibly by combining moment conditions that involve levels of the data. Specifically, we show that when $T > 3$, the 1-step AB GMM estimator, which is optimal under time-series homoskedasticity, and the 2-step optimal AB GMM estimator that uses the 1-step AB GMM estimator to estimate the weight matrix remain consistent in this case. Furthermore, when $T > 3$, the 2-step System estimator will still be consistent in this case if it uses a consistent estimate of the optimal weight matrix. The RE Quasi ML estimator for $\rho$ is also still consistent in this case. As the REQMLE has favorable properties, also when $\rho$ is near or equal to unity, cf. Kruiniger (2013), we propose using the REQMLE to estimate the optimal weight matrix of the System estimator. In this way we obtain a new, more robust version of the System estimator. When $\rho$ is not close to unity, one could also use the AB GMM estimator to estimate the optimal weight matrix of the System estimator leading to yet another version of the System estimator.

We also derive local asymptotic approximations to the finite sample distributions of

AB GMM estimators when the data are mean stationary and the variance ratio is large, and give conditions for redundancy of moment conditions that involve levels of the data.

The outline of the paper is as follows. Section 2 presents the necessary condition for the consistency of RE and FE estimators for the panel AR(1) model. It also derives the asymptotic properties of various GMM estimators when the variance ratio is large and the local asymptotic distributions for the RE AB GMM estimator when in addition the data are mean stationary, discusses conditions for redundancy of the additional moment conditions that are exploited by certain RE GMM estimators relative to those exploited by the corresponding optimal FE GMM estimators, and presents the new versions of the 2-step System estimator. Section 3 conducts a Monte Carlo study of the finite sample properties of the 2-step optimal RE AB GMM estimator, three versions of the 2-step System estimator and the RE and FE Quasi ML counterparts of a GMM estimator of Ahn and Schmidt (1995, AS) and conventional and Windmeijer (2005) corrected versions of asymptotic standard errors and confidence intervals related to the GMM estimators. It also investigates the properties of various tests for weak or underidentification due to Montiel Olea and Pflueger (2013) and Windmeijer (2018), respectively, and discusses how the former can be used to select a version of the System estimator that is (most) suitable for a particular application. Section 4 provides a real data application and section 5 concludes. An appendix contains all the proofs, some of the Monte Carlo results and some additional results related to the application.

# 2 Asymptotic properties of random and fixed effects GMM estimators for the panel AR(1) model

## 2.1 A necessary condition for consistency: the fixed effects assumption

The panel AR(1) model with arbitrary initial conditions is given by [1]

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

where $y_i = (y_{i,2} \ ... \ y_{i,T})'$, $y_{i,-1} = (y_{i,1} \ ... \ y_{i,T-1})'$, $\iota$ is a vector of ones, $\mu_i$ is the individual effect and $\varepsilon_i$ is the vector of (idiosyncratic) errors for all $i \in \mathcal{I} = \{1, 2, ..., N\}$. For

[1] Extending the analysis to models with strictly exogenous regressors, a constant and/or time dummies is straightforward.

each individual unit we have $T \geq 3$ observations on $y$, including the initial observation $y_{i,1}$. When considering the asymptotic properties of the estimators for this model we will assume that $N \to \infty$ while $T$ is fixed. Note that in the unit root case the individual effects disappear. The panel AR(1) model can be rewritten as

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

Let $\mu_i \equiv \sigma_\mu \widetilde{\mu}_i$. Furthermore, let $v_{i,1} \equiv y_{i,1} - \mu_i$, $\widetilde{v}_{i,1} \equiv (\rho - 1)v_{i,1}$ and $\mathcal{T} = \{2, ..., T\}$. We make the following assumptions.

**Standard Assumptions (SA):**

*i) $\varepsilon_i$ $i = 1, ..., N$ are independently distributed;*

*ii) $E(\varepsilon_i) = 0$ and $E(\varepsilon_i \varepsilon_i') = \sigma_i^2 I_{T-1}$ $\forall$ $i \in \mathcal{I}$;*

*iii) $E(|\varepsilon_{i,t}|^{2+\delta}) < \Delta_1 < \infty$ for some $\delta > 0$, some $\Delta_1 > 0$, $\forall$ $i \in \mathcal{I}$ and $\forall$ $t \in \mathcal{T}$;*

*iv) $E(|\widetilde{\mu}_i|^{2+\delta}) < \Delta_1 < \infty$ for some $\delta, \Delta_1 > 0$ and $\forall$ $i \in \mathcal{I}$; $plim_{N\to\infty} N^{-1} \sum_{i=1}^N \widetilde{\mu}_i^2 = 1$;*

*v) $E(\widetilde{v}_{i,1}) = 0$ and $y_{i,1}$ and $\mu_i$ are uncorrelated with the elements of $\varepsilon_j$ $\forall$ $i, j \in \mathcal{I}$;*

*vi) $\widetilde{v}_{i,1}\varepsilon_{i,t}$ and $\widetilde{v}_{j,1}\varepsilon_{j,t}$ are uncorrelated $\forall$ $i, j \in \mathcal{I}$ with $i \neq j$ and $\forall$ $t \in \mathcal{T}$;*

*vii) In case of a 2-step GMM estimator, $4 + \delta$ moments of the data exist for some $\delta > 0$.*

SA(ii) assumes unconditional homoskedasticity over time for presentational ease only; this assumption can be relaxed. We define $\sigma^2 \equiv \lim_{N\to\infty} N^{-1} \sum_{i=1}^N \sigma_i^2$ for later use. SA(iv) implies $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N \mu_i^2 = \sigma_\mu^2$. SA(v) allows for conditional heteroskedasticity over time and for cross-sectional dependence of the $y_{i,1}$ and $\mu_i$. W.l.o.g. it also assumes $E(\widetilde{v}_{i,1}) = 0$ and hence $E(y_{i,t} - y_{i,t-1}) = 0$ for all $t \in \mathcal{T}$. This assumption is equivalent to adding a constant to the model in (1) if necessary. SA(vi) is not required for the main results but affords a simplification if the data are not normally distributed (see the end of next paragraph) and is almost always satisfied. We will also use the following assumption.

**Fixed Effects Assumption (FEA):** $\widetilde{\sigma}_v^2 \equiv plim_{N\to\infty} N^{-1} \sum_{i=1}^N \widetilde{v}_{i,1}^2 < \infty$.

The assumption that $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N \widetilde{v}_{i,1}^2 < \infty$ is weaker than the assumption that $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N \mu_i^2 < \infty$ and $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N y_{i,1}^2 < \infty$, which is habitually made in papers that study random effects estimators for the panel AR(1) model, see e.g. Ahn and Schmidt (1995, 1997); it can still hold even if $\lim_{i\to\infty}E(y_{i,1}^2) = \infty$ and $\lim_{i\to\infty}E(\mu_i^2) = \infty$ when $\{E(y_{i,1}^2)\}$ and $\{E(\mu_i^2)\}$ are non-decreasing. Furthermore, under SA(ii) and covariance stationarity of $\{y_{i,t}\}$ we have $\lim_{\rho\uparrow 1}E((y_{i,1}-\mu_i)^2) = \lim_{\rho\uparrow 1}(\sigma_i^2/(1-\rho^2)) = \infty$ but $\lim_{\rho\uparrow 1}E(\widetilde{v}_{i,1}^2) = 0$ so that if in addition $\widetilde{v}_{i,1}$ $i = 1, ..., N$ are i.i.d., then FEA is satisfied when $\rho \uparrow 1$. FEA also allows for cross-sectional independence of the $\widetilde{v}_{i,1}$. In the appendix we show that SA and FEA imply that $\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}\widetilde{v}_{i,1}\varepsilon_{i,t} = 0$ for all $t \in \mathcal{T}$. The proof relies on SA(vi) in case the data are not normally distributed.

Below we will see that, given assumption SA, assumption FEA is, practically speaking, necessary for the consistency of any GMM or ML type estimator for $\rho$ and also sufficient for the consistency of GMM and ML type estimators for $\rho$ that only depend on differences of the data. For this reason the latter can be regarded as fixed effects estimators, where the label 'fixed effects' indicates that minimal assumptions are made about the individual effects and the initial observations.

An assumption that implies FEA when it is combined with SA and, like FEA, does not require that $E(y_{i,1}^2) < \infty$ and $E(\mu_i^2) < \infty$ for all $i \in \mathcal{I}$ is the following one.

**Fixed Effects Assumption* (FEA*):**

$\widetilde{v}_{i,1}$ $i = 1, ..., N$ *are i.h.d. and* $E(|\widetilde{v}_{i,1}|^{2+\delta}) < \Delta_1 < \infty$ *for some* $\delta, \Delta_1 > 0$ *and* $\forall$ $i \in \mathcal{I}$.

Unlike FEA, FEA* requires cross-sectional independence of the $\widetilde{v}_{i,1}$. When $\{y_{i,t}\}$ is stationary up to order 3 (or strictly stationary) for all $i \in \mathcal{I}$, $E(\varepsilon_{i,r}\varepsilon_{i,s}\varepsilon_{i,t}) = 0$ and $E(\varepsilon_{i,s}\varepsilon_{i,t}^2) = 0$ for any $r \leq s < t \in \mathcal{T}$ and for all $i \in \mathcal{I}$, we have $\lim_{\rho\uparrow 1}E(|\widetilde{v}_{i,1}|^3) = 0$ for all $i \in \mathcal{I}$ so that if in addition $\widetilde{v}_{i,1}$ $i = 1, ..., N$ are i.h.d., then FEA* is satisfied when $\rho \uparrow 1$.

It is useful to rewrite the panel AR(1) model in (1) as

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

where $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$. Notice that the differences of the data only depend on $\widetilde{v}_{i,1}$ and $\varepsilon_{i,t}$, $t = 2, ..., T$. In the appendix we prove the following lemma:

**Lemma 1** *Given SA, FEA holds iff* $\text{plim}_{N\to\infty}\left|N^{-1}\sum_{i=1}^{N}(\Delta y_{i,s}\Delta y_{i,t})\right| < \infty$ $\forall$ $s, t \in \mathcal{T}$.

Maximum Likelihood estimators for $\rho$ in (1) as well as any reasonable GMM estimator for $\rho$ depend on second-order sample moments. These estimators will be root-$N$ consistent only if these sample moments converge in probability (possibly after scaling) and if the probability limits of these sample moments allow for identification of $\rho$. [2] [3] The second-order sample moments can take the form of cross-sectional averages of products of levels of the data, cross-sectional averages of products of differences of the data or cross-sectional averages of products of levels and differences of the data. Lemma 1 implies that given assumption SA, assumption FEA about the $\widetilde{v}_{i,1}$ is sufficient and, practically speaking, also necessary for convergence in probability of the cross-sectional averages of the products of differences of the data. The reason for these results is that the differences of the data only depend on the $\widetilde{v}_{i,1}$ and the $\varepsilon_{i,t}$. To guarantee convergence in probability of second-order sample moments that involve levels of the data or consistency of estimators that depend on them, one also needs to add assumptions about the $\mu_i$, e.g. $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mu_i^2 < \infty$ and $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mu_i\varepsilon_{i,t} = 0$, $\forall t \in \mathcal{T}$ or, given FEA equivalently, assumptions about the $y_{i,1}$, e.g. $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} y_{i,1}^2 < \infty$ and $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} y_{i,1}\varepsilon_{i,t} = 0$, $\forall t \in \mathcal{T}$.

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

We note that FEA or $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}(\Delta y_{i,2})^2 < \infty$ is a reasonable assumption that is met in most applications, possibly after rescaling the data.

## 2.2 Asymptotic properties of RE and FE GMM estimators

Under SA and FEA we can derive expressions for the probability limits of the following second-order sample moments: $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}(\Delta y_{i,2})^2 =$

[2] Convergence in probability of these (scaled) sample moments is required in order to prove uniform convergence of the criterion function (see e.g. Newey and McFadden (1994)). Identification of $\rho$ requires that at least some of the probability limits of these (scaled) sample moments are different from zero.

[3] Under non-normality of the data, one may wish to consider GMM estimators for $\rho$ that also exploit information contained in third and higher order sample moments, cf Hahn (1997). In that case, issues similar to those discussed in this paper will arise. In particular, such an estimator is a consistent fixed effects estimator only if $\{\widetilde{v}_{i,1}\}$ satisfies a generalized version of FEA. Here we confine our attention to estimators that only exploit second moments of the data.

$(1 - \rho)^2\sigma_v^2 + \sigma^2$, $\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}(y_{i,1})^2 = \sigma_v^2 + 2\sigma_\mu\sigma_v corr(\mu, v_1) + \sigma_\mu^2$, and $\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}(y_{i,1}\Delta y_{i,2}) = (\rho - 1)(\sigma_v^2 + \sigma_\mu\sigma_v corr(\mu, v_1)) + \sigma_v\sigma corr(v_1, \varepsilon_2) + \sigma_\mu\sigma corr(\mu, \varepsilon_2)$, where $\sigma_v^2 \equiv \text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}v_{i,1}^2$, and $corr(X, Y) \equiv \text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}(X_iY_i)/(\sigma_X\sigma_Y)$. When $\sigma_\mu \to \infty$, we need to scale the levels of the data (e.g. $y_{i,1}$) by $\sigma_\mu$ for the sample moments that contain them to converge. Note that SA implies that $corr(\mu, \varepsilon_2) = 0$ and that under mean stationarity we also have $corr(\mu, v_1) = 0$ so in the latter case $\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}(y_{i,1}\Delta y_{i,2})/\sigma_\mu = 0$ when $\sigma_\mu \to \infty$. This result implies that $y_{i,1}$ is a weak instrument for $\Delta y_{i,2}$ when $corr(\mu, v_1) = 0$ and $\sigma_\mu/\sigma$ and $\sigma_\mu/\sigma_v$ are large. Furthermore, when $\{y_{i,t}\}$ is covariance stationary, then $\sigma_v^2 = \sigma^2/(1 - \rho^2)$, and if in addition $\rho \uparrow 1$, then $\sigma_v^2 \to \infty$. In that case, $\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}(y_{i,1}\Delta y_{i,2})$ may not exist and $\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}(y_{i,1})^2$ does not exist.

Below we assume that $\sigma_v^2$ and $\sigma^2$ are finite and of a similar order of magnitude and do not study what happens when $\rho \uparrow 1$ as this case has been studied in other papers, e.g. Kruiniger (2009) and Bun and Kleibergen (2017). Then the properties of the RE estimators for the panel AR(1) model — which involve levels of the data — depend crucially on the ratio of $\sigma_\mu^2$ and $\sigma^2$ (or $\sigma_\mu^2/\sigma_v^2$) and on the value of $corr(\mu, v_1)$.

### 2.2.1 Arellano-Bond GMM estimators

The RE Arellano-Bond (AB) GMM estimator, $\widehat{\rho}_{ABlev}$, exploits the following $m \equiv (T - 1)(T - 2)/2$ moment conditions:

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

This estimator uses lagged levels of the dependent variable as instruments.

In the appendix we prove the following result: [4] [5]

**Theorem 1** *Assume that SA and FEA* hold, $T = 3$, $|\rho| < 1$ and $\sigma_\mu^2 \to \infty$. Then $\widehat{\rho}_{ABlev}$ is $\sqrt{N}$-consistent if and only if $corr(\mu, v_1) \neq 0$.*

---

[4] $\sigma_\mu^2 \to \infty$ signifies that $\sigma_\mu^2$ approaches $\infty$ independently of $N$. Moreover, when deriving these asymptotic results, $\sigma_\mu^2 \to \infty$ first and then $N \to \infty$. Below we will also derive asymptotic results for $\widehat{\rho}_{ABlev}$ using the parameter sequence $\sigma_\mu^2/\sigma_v^2 = k_1N^p$.

[5] When $T = 3$ and $\sigma_\mu^2 < \infty$, $\widehat{\rho}_{ABlev}$ is inconsistent if $E(y_{i,1}\Delta y_{i,2}) = 0 \;\forall\; i \in \mathcal{I}$. $E(y_{i,1}\Delta y_{i,2}) = (\rho - 1)[E(y_{i,1}(y_{i,1} - \mu_i))] = (\rho - 1)[E(v_{i,1}^2) + E(\mu_i v_{i,1})]$. When $\{y_{i,t}\}$ is covariance stationary, $E(\mu_i v_{i,1}) = 0$ and $E(y_{i,1}(y_{i,1} - \mu_i)) = \sigma_i^2/(1 - \rho^2) \geq 0$. In general, $E(y_{i,1}(y_{i,1} - \mu_i)) = 0 \;\forall\; i \in \mathcal{I}$ is not very plausible unless $y_{i,1} = \mu_i$ or $y_{i,1} = 0 \;\forall\; i \in \mathcal{I}$.

The reason for this result is that $\text{plim}_{N\to\infty} \lim_{\sigma_\mu\to\infty}(N^{-1}\sum_i y_{i,1}\Delta y_{i,2})/\sigma_\mu = (\rho-1)\times corr(\mu, v_1)\sigma_v$, that is, when $corr(\mu, v_1) \neq 0$, the scaled instrument $y_1/\sigma_\mu$ is correlated with the lagged dependent variable $\Delta y_2$, whereas when $corr(\mu, v_1) = 0$, $y_{i,1}/\sigma_\mu$ is an invalid instrument for $\Delta y_{i,2}$.[6] In the latter case $\rho$ is not identified and the estimator $\widehat{\rho}_{ABlev}$ converges to $\rho$ plus a ratio of correlated normal random variables that have zero mean, $X_1/X_2$, which is defined in Theorem 3 below.

We will now extend the result in Theorem 1 to $T > 3$. Let $y_i^t = [y_{i,1} \ ... \ y_{i,t}]$ and let $Z_i = diag(y_i^1, ..., y_i^{T-2})$ be a $(T-2)\times m$ block-diagonal matrix. Then we can also write the set of AB moment conditions in (3) as $E(Z_i'\Delta\underline{\varepsilon}_i) = 0$ where $\Delta\underline{\varepsilon}_i \equiv \underline{\varepsilon}_i - \underline{\varepsilon}_{i,-1}$ with $\underline{\varepsilon}_i \equiv [\varepsilon_{i,3} \ ... \ \varepsilon_{i,T}]'$. Under our assumptions, $E(\Delta\underline{\varepsilon}_i\Delta\underline{\varepsilon}_i') = \sigma_i^2 H$, where $H = H_{T-2}$ is a $(T-2)$ band-diagonal matrix with 2'$s$ on the main diagonal, $-1$'$s$ on the first sub- and superdiagonal and zeros elsewhere. It follows that the RE AB GMM estimator which uses $W_{N,AB1} = (N^{-1}\sum_{i=1}^N Z_i'HZ_i)^{-1}$ as weight matrix is an optimal one-step GMM estimator. This estimator will be denoted as $\widehat{\rho}_{AB1lev}$. Let $w_{i,t} = (1-\rho)\mu_i + \varepsilon_{i,t}$, $w_i = (1-\rho)\mu_i\iota + \underline{\varepsilon}_i$ and $\underline{y}_i = [y_{i,3} \ ... \ y_{i,T}]'$. The two-step optimal RE AB GMM estimator uses $W_{N,AB2} = (N^{-1}\sum_{i=1}^N Z_i'\Delta\widehat{w}_i\Delta\widehat{w}_i'Z_i)^{-1}$ as weight matrix, where $\Delta\widehat{w}_i = \Delta\underline{y}_i - \widehat{\rho}_{AB1lev}\Delta\underline{y}_{i,-1}$, and will be denoted as $\widehat{\rho}_{AB2lev}$. Finally, let $\{W_N\}$ denote an arbitrary sequence of PD weight matrices with $\text{plim}_{N\to\infty}W_N = W$, where $W$ is PD. An RE AB GMM estimator that uses $W_N$ as weight matrix will be denoted as $\widehat{\rho}_{ABlev}$.

Let the $(T-2)\times(T-t-2)$ matrix $d_t$ be given by $d_t = [0 \ I_{T-t-2}]'$. Furthermore, let $Z_i^{AB} = [Z_i^I \ Z_i^D]$, where $Z_i^I = y_{i,1}I_{T-2}$ and $Z_i^D = [d_1\Delta y_{i,2} \ d_2\Delta y_{i,3} \ ... \ d_{T-3}\Delta y_{i,T-2}]$ is a $(T-2)\times[m-(T-2)]$ matrix. There exists a nonsingular constant matrix $K^{AB}$ such that $Z_i' = K^{AB}Z_i^{AB\prime}$. Thus we can restate $E(Z_i'\Delta\underline{\varepsilon}_i) = 0$ as $E(Z_i^{AB\prime}\Delta\underline{\varepsilon}_i) = 0$.

We can extend Theorem 1 to $T > 3$ by using Lemma 2 from the appendix.

**Theorem 2** *Assume that SA and FEA* hold, $T > 3$, $|\rho| < 1$ and $\sigma_\mu^2 \to \infty$. Then:*

*(i) $\widehat{\rho}_{AB1lev}$ and $\widehat{\rho}_{AB2lev}$ are $\sqrt{N}$-consistent.*

*Furthermore, if $\widehat{\rho}_{ABlev}$ exploits $E(Z_i^{AB\prime}\Delta\underline{\varepsilon}_i) = 0$ in lieu of $E(Z_i'\Delta\underline{\varepsilon}_i) = 0$, then*

*(ii) $\widehat{\rho}_{ABlev}$ is $\sqrt{N}$-consistent if and only if $corr(\mu, v_1) \neq 0$, and*

*(iii) if $corr(\mu, v_1) = 0$, $\widehat{\rho}_{ABlev} - \rho \xrightarrow{d} (X_{61}'W_{11}X_{61})^{-1}X_{61}'W_{11}X_{51}$ where $X_{51}$ and $X_{61}$ are Gaussian random vectors that are defined in Lemma 2.*

[6] When $corr(\mu, v_1) = 0$ and $\sigma_\mu^2 \to \infty$, then the value of the first stage regression coefficient is zero: $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty}(\sum_i(y_{i,1}\Delta y_{i,2})/(\sum_i y_{i,1}^2)) = 0$.

The proofs of Lemma 2 and Theorem 2 are similar to that of Theorem 1. According to Theorem 2, when $T > 3$ and $corr(\mu, v_1) = 0$ an RE AB GMM estimator can be consistent as long as it uses a suitable weight matrix. In particular, because the instruments in levels need to be scaled by $\sigma_\mu$ in order to achieve convergence of the averages of their cross products with the dependent variable and its lag, the blocks of its weight matrix that correspond to $E(Z_i^{I\prime}\Delta\underline{\varepsilon}_i/\sigma_\mu) = 0$ should provide this scaling in order to obtain a consistent estimator when $corr(\mu, v_1) = 0$. The weight matrix $W_{N,AB1}$ has this property but an arbitrary weight matrix $W_N$ does not. An RE AB GMM estimator that uses an arbitrary weight matrix $W_N$ will effectively only exploit the (scaled) moment conditions that involve an instrument in levels when $\sigma_\mu^2 \to \infty$, i.e., $E(Z_i^{I\prime}\Delta\underline{\varepsilon}_i/\sigma_\mu) = 0$, which do not identify $\rho$ when $corr(\mu, v_1) = 0$ (because $\text{plim}_{N\to\infty} \lim_{\sigma_\mu\to\infty} N^{-1}\sum_{i=1}^N Z_i^{I\prime}\Delta y_{i,-1}/\sigma_\mu = 0$), and hence will be inconsistent. In contrast, the RE AB GMM estimator that uses $W_{N,AB1}$ will continue to exploit both $E(Z_i^{I\prime}\Delta\underline{\varepsilon}_i/\sigma_\mu) = 0$ and $E(Z_i^{D\prime}\Delta\underline{\varepsilon}_i) = 0$ when $\sigma_\mu^2 \to \infty$, and the latter will still identify $\rho$ when $corr(\mu, v_1) = 0$. Of course, $E(Z_i^{D\prime}\Delta\underline{\varepsilon}_i) = 0$ will only weakly identify $\rho$ when $\rho$ is near unity and $corr(\mu, v_1)$ is (near) zero, cf. Kruiniger (2009).

When $corr(\mu, v_1) \neq 0$, $\text{plim}_{N\to\infty} \lim_{\sigma_\mu\to\infty} N^{-1}\sum_{i=1}^N Z_i^{I\prime}\Delta y_{i,-1}/\sigma_\mu \neq 0$ and an RE AB GMM estimator that uses an arbitrary weight matrix $W_N$ will remain consistent when $\sigma_\mu^2 \to \infty$. Hayakawa (2009) found that when the data are not close to mean stationary and the variance of the individual effects becomes large as compared to that of the disturbances, then the RE Arellano-Bond GMM estimator in fact performs quite well. He argued that this is because the correlation between the lagged dependent variable and the instruments in levels gets larger owing to the unremoved individual effects, i.e., the instruments in levels become strong. In contrast, we find that when the data are (close to) mean stationary and the variance of the individual effects becomes large relative to that of the disturbances, the instruments in levels become weaker.

When $T = 3$, $corr(\mu, v_1) = 0$ and the ratio of $\sigma_\mu^2$ and $\sigma_v^2$ is large relative to the sample size $N$, $y_1$ will be a weak instrument for $\Delta y_2$ and the standard fixed parameter first-order asymptotic approximation to the distribution of $\widehat{\rho}_{ABlev}$ will be poor. We may be able to obtain a better approximation by employing local asymptotics where the ratio $\sigma_\mu^2/\sigma_v^2$ is reparametrized as a non-decreasing function of the sample size, i.e. $\sigma_\mu^2/\sigma_v^2 = k_1 N^p$ with $p \geq 0$. The theorem below, which is proven in the appendix, describes how the limiting distribution of $\widehat{\rho}_{ABlev}$ changes as we vary the value of parameter $p$, which determines the quality of the instrument $y_{i,1}$:

**Theorem 3** *Let assumptions SA and FEA* hold, $T = 3$ and $|\rho| < 1$.*

*Furthermore, let $\sigma_\mu^2/\sigma_v^2 = k_1 N^p$ with $p \geq 0$, let $corr(\mu, v_1) = 0$,*

*and let $\lim_{N\to\infty} N^{-1-p} \sum_{i=1}^N E(\mu_i \varepsilon_{i,t})^2 = \zeta_1$, $t = 2, 3$ and*

$\lim_{N\to\infty} N^{-1-p} \sum_{i=1}^N E(\mu_i v_{i,1})^2 = \zeta_2$ *with $0 < \zeta_1, \zeta_2 < \infty$.*

*If $p = 0$, $0 < \lim_{N\to\infty} N^{-1} \sum_{i=1}^N E(y_{i,1}\Delta\varepsilon_{i,3})^2 = \zeta_0 < \infty$ and $\{y_{i,1}\Delta\varepsilon_{i,3}\}$ satisfies the Lindeberg condition, then $N^{0.5}\left(\widehat{\rho}_{ABlev} - \rho\right) \xrightarrow{d} N(0, [(\rho - 1)\sigma_v^2]^{-2}\zeta_0)$.*

*If $0 < p < 1$ and $\{\mu_i \Delta\varepsilon_{i,3}/N^{0.5p}\}$ satisfies the Lindeberg condition, then $N^{0.5(1-p)}\left(\widehat{\rho}_{ABlev} - \rho\right) \xrightarrow{d} N(0, [(\rho - 1)\sigma_v^2]^{-2} 2\zeta_1)$.*

*If $p \geq 1$, $\{\mu_i \varepsilon_{i,t}/N^{0.5p}\}$, $t = 2, 3$, and $\{\mu_i v_{i,1}/N^{0.5p}\}$ satisfy the Lindeberg condition,*

*and* $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2\zeta_1 & -\zeta_1 \\ -\zeta_1 & \zeta_1 + (\rho - 1)^2\zeta_2 \end{bmatrix}\right)$, *then*

*if $p = 1$, $\widehat{\rho}_{ABlev} - \rho \xrightarrow{d} \frac{X_1}{(\rho-1)\sigma_v^2 + X_2}$, while if $p > 1$, $\widehat{\rho}_{ABlev} - \rho \xrightarrow{d} \frac{X_1}{X_2}$.*

Generalizing this result to $T > 3$ is straightforward.

Under conditional homoskedasticity of the $\mu_i$ and the $\varepsilon_{i,t}$ and uncorrelatedness of $\mu_i^2$ and $v_{i,1}^2$ $\forall\ i \in \mathcal{I}$, we find that $\zeta_0 = 2(\sigma_\mu^2 + \sigma_v^2)\sigma^2$, $\zeta_1 = k_1\sigma_v^2\sigma^2$, and $\zeta_2 = k_1(\sigma_v^2)^2$.

For $p = 0$, one obtains the standard fixed parameter first-order limiting distribution of $\widehat{\rho}_{ABlev}$; when $0 < p < 1$, one obtains essentially the same limiting distribution because if $N \to \infty$, then $\sigma_\mu^2/\sigma_v^2 \to \infty$ and $E(y_{i,1}\Delta\varepsilon_{i,3}/\sigma_\mu)^2 \to E(\mu_i\Delta\varepsilon_{i,3}/\sigma_\mu)^2$; and when $p \geq 1$, $\widehat{\rho}_{ABlev} - \rho$ converges to a ratio of correlated normal r.v.'s, where $p = 1$ is a special case because in this case the mean of the denominator is different from zero. The limiting distribution for $p > 1$ is in fact equal to the sequential limiting distribution of $\widehat{\rho}_{ABlev}$ when $corr(\mu, v_1) = 0$ and first $\sigma_\mu^2 \to \infty$ and then $N \to \infty$. [7] [8]

[7] Alternatively, we may consider local asymptotics based on the sequence $corr(\mu, v_1) = cN^{-q}$, while $\sigma_\mu^2/\sigma_v^2 \to \infty$. Let us assume homoskedasticity of $\mu$, $v_1$ and $\varepsilon_t$ and let $\widetilde{Z}_1 = N^{-0.5}\sum_i y_{i,1}\Delta\varepsilon_{i,3}/\sigma_\mu$. It follows that $\widetilde{Z}_1 \xrightarrow{d} Z_1 \sim N(0, 2\sigma^2)$. Then for $0 \leq q < 0.5$, we obtain $N^{0.5-q}\left(\widehat{\rho}_{ABlev} - \rho\right) \xrightarrow{d} N(0, [(\rho - 1)c\sigma_v]^{-2}2\sigma^2)$. If $q > 0$, $\lim_{N\to\infty}E(\mu v_1/\sigma_\mu) = 0$. Let $\lim_{N\to\infty}\mathrm{Var}(\mu v_1/\sigma_\mu) = \overline{\sigma}_{\mu v}^2$ and let $\widetilde{Z}_2 = N^{-0.5}\sum_i y_{i,1}\Delta y_{i,2}/\sigma_\mu$. Then if $q = 0.5$, $\widetilde{Z}_2 \xrightarrow{d} Z_2 \sim N((\rho-1)c, (\rho-1)^2\overline{\sigma}_{\mu v}^2 + \sigma^2)$, while if $q > 0.5$, $\widetilde{Z}_2 \xrightarrow{d} Z_2 \sim N(0, (\rho-1)^2\overline{\sigma}_{\mu v}^2 + \sigma^2)$. It follows that if $q \geq 0.5$, $\left(\widehat{\rho}_{ABlev} - \rho\right) \xrightarrow{d} Z_1/Z_2$ where $\mathrm{Cov}(Z_1, Z_2) = -\sigma^2$. We can also consider combined local asymptotics based on the sequences $\sigma_\mu^2/\sigma_v^2 = k_1N^p$ and $corr(\mu, v_1) = cN^{-q}$. Then if $p = 1$ and $q = 0.5$, $\left(\widehat{\rho}_{ABlev} - \rho\right) \xrightarrow{d} Z_1/((\rho - 1)\sigma_v/k_1^{-0.5} + Z_2)$, while if $0 < p < 1$ and $q = 0.5p$, $N^{0.5(1-p)}\left(\widehat{\rho}_{ABlev} - \rho\right) \xrightarrow{d} N(0, [(\rho - 1)\sigma_v/k_1^{-0.5} + (\rho - 1)c\sigma_v]^{-2}2\sigma^2)$.

The parameters of the limiting distribution of $\widehat{\rho}_{ABlev}$ can be consistently estimated regardless of the value of $p$ : one only needs to know whether $p < 1$, $p = 1$, or $p > 1$. Note that the distribution of $X_1/X_2$ depends only on $\zeta_1/\widetilde{\zeta}_2$, where $\widetilde{\zeta}_2 = \zeta_1 + (\rho - 1)^2\zeta_2$. When $p > 1$, the ratio $\zeta_1/\widetilde{\zeta}_2$ can be consistently estimated by $0.5\sum_{i=1}^{N}(y_{i,1}[\Delta y_{i,3} - \widehat{\rho}_{ABdif}\Delta y_{i,2}])^2/\sum_{i=1}^{N}(y_{i,1}\Delta y_{i,2})^2$, where $\widehat{\rho}_{ABdif}$ is a consistent FE estimator that is defined below. When $p = 1$, one can estimate $\zeta_1$ by $0.5N^{-2}\sum_{i=1}^{N}(y_{i,1}[\Delta y_{i,3} - \widehat{\rho}_{ABdif}\Delta y_{i,2}])^2$, $\widetilde{\zeta}_2$ by $N^{-2}\sum_{i=1}^{N}(y_{i,1}\Delta y_{i,2})^2$, and $(1-\rho)\sigma_v^2$ by $[N^{-1}\sum_{i=1}^{N}(\Delta y_{i,t})^2 - 0.5N^{-1}\times \sum_{i=1}^{N}([\Delta y_{i,3} - \widehat{\rho}_{ABdif}\Delta y_{i,2}])^2]/(1-\widehat{\rho}_{ABdif})$ and when $p < 1$, one can estimate $\zeta_1$ by $0.5N^{-1-p}\sum_{i=1}^{N}(y_{i,1}[\Delta y_{i,3} - \widehat{\rho}_{ABdif}\Delta y_{i,2}])^2$.

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

$$E[\Delta y_{i,s}(\Delta y_{i,t} - \rho\Delta y_{i,t-1})] = 0, \quad 2 \leq s \leq t-2, \quad t = 4, ..., T. \tag{4}$$

This estimator, which uses an arbitrary PD weight matrix $W_N$ (with $\text{plim}_{N\to\infty}W_N = W$, where $W$ is PD) and only involves differences of the data, will be consistent for any sequence of individual effects as long as FEA (or FEA$^*$) holds.

**Theorem 4** *Assume that SA and FEA hold and $|\rho| < 1$. Then $\widehat{\rho}_{ABdif}$ is $\sqrt{N}$-consistent.*

The proof of this theorem is trivial. Recall that the first-differences of the data $\Delta y_{i,t}$, $t = 2, ..., T$, only depend on $y_{i,1}$ and $\mu_i$ through $v_{i,1}$.

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

**Moment Assumption 1 (MA1):** $0 < \overline{\sigma}^4 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sigma_i^4 < \infty$.

**Moment Assumption 2 (MA2):** $\sigma_{v,\varepsilon}^2 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E(v_{i,1}^2\varepsilon_{i,t}^2) < \infty\ \forall t \in \mathcal{T}$.

Then it is straightforward to prove the following result:

[8]The lagged values $y_{i,s}$, $s = 1, ..., t-2$, are also weak instruments for $\Delta y_{i,t-1}$ when $\rho$ is close to one. Kruiniger (2009) obtains local limiting distributions for $\widehat{\rho}_{ABlev}$ both under covariance stationarity and for the model with fixed initial conditions using the parameter sequences $\rho = 1 - (\lambda/N)$ and $\rho = 1 - (\lambda/N)^{0.5}$, respectively. These local limiting distributions depend on the localizing parameter $\lambda$ which cannot be consistently estimated when $T$ is fixed. Using parameter sequences for $\rho$ that tend to one at a faster rate would result in the standard (fixed parameter first-order) limiting distribution for $\rho = 1$ just as taking $p > 1$ results in the standard limiting distribution for $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) = 0$.

**Theorem 5** *Let assumptions SA, FEA$^*$, MA1, and MA2 hold, $T = 4$, and $|\rho| < 1$. Moreover, let $\{\Delta y_{i,2}\Delta\varepsilon_{i,4}\}$ satisfy the Lindeberg condition.*
*Then* $\sqrt{N}\left(\widehat{\rho}_{ABdif} - \rho\right) \xrightarrow{d} N(0, [(\rho-1)^2\rho\sigma_v^2 + (\rho-1)\sigma^2]^{-2}[2(\rho-1)^2\sigma_{v,\varepsilon}^2 + 2\overline{\sigma}^4])$.

Under conditional homoskedasticity of the $\varepsilon_{i,t}$, $\sigma_{v,\varepsilon}^2 = \sigma_v^2\sigma^2$ and $\overline{\sigma}^4 = \sigma^4$. Notice that the limiting distribution of $\widehat{\rho}_{ABdif}$ does not depend on $\sigma_\mu^2$ unlike the standard first-order fixed parameter limiting distribution of $\widehat{\rho}_{ABlev}$. It follows that $Var(\widehat{\rho}_{ABdif,\, T=4}) < Var(\widehat{\rho}_{ABlev,\, T=3})$ when $corr(\mu, v_1) = 0$ and $\sigma_\mu^2$ is large relative to $\sigma_v^2$ and $\sigma^2$.

### 2.2.2 Ahn-Schmidt GMM estimators and Quasi ML estimators

Just like the RE AB GMM estimator for $\rho$ does not exhaust the set of all the second moment conditions, the AB differences GMM estimator for $\rho$ does not exhaust the set of all the second moment conditions that follow from assumptions SA and FEA$^*$ and involve only differences of the data. The complete set of second moment conditions implied by the panel AR(1) model corresponds to

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

where $D$ is the $(T-2)\times(T-1)$ first difference matrix with $D_{k,k} = -1$ and $D_{k,k+1} = 1$, $k = 1, ..., T-2$, and zeros elsewhere. The RE HOmoskedastic Conditional (HOC) GMM estimator for $\rho$ exploits all these second moment conditions, or equivalently, the following moment conditions (cf Ahn and Schmidt, 1997):

$$\begin{aligned} & E(y_{i,1}(\Delta y_{i,t} - \rho\Delta y_{i,t-1})) = 0, \quad t = 3, ..., T, \\ & E((y_{i,t} - \rho y_{i,t-1})^2 - (y_{i,2} - \rho y_{i,1})^2) = 0, \quad t = 3, ..., T, \quad \text{and} \\ & E((y_{i,t} - \rho y_{i,t-1})(y_{i,s} - \rho y_{i,s-1}) - (y_{i,3} - \rho y_{i,2})(y_{i,2} - \rho y_{i,1})) = 0, \\ & \qquad\qquad\qquad\qquad\qquad\qquad 2 \le s < t, \quad t = 4, ..., T. \end{aligned} \tag{5}$$

Kruiniger (2013) has shown that when $-1 < \rho \le 1$ and $T \ge 3$, then $\rho$ is uniquely identified by them and the RE HOCGMM estimator for $\rho$ is consistent. The moment conditions in the second line of (5) depend on homoskedasticity over time. The RE Conditional GMM estimator for $\rho$ allows for heteroskedasticity over time and exploits the moment conditions in the first and third line of (5), cf. Ahn and Schmidt (1997).

As we have seen above, GMM estimators for $\rho$ that exploit moment conditions involving levels of the data, e.g. $y_{i,1}$, are not consistent for all sequences of fixed effects that satisfy FEA. On the other hand, Fixed Effects GMM estimators — which are consistent for any sequence of fixed effects that satisfies FEA — only exploit moment conditions that only involve differences of the observations. Noting that $\Delta y_{i,2} = \varepsilon_{i,2} - (1-\rho)v_{i,1}$ and $D\varepsilon_i = Dy_i - \rho Dy_{i,-1}$, SA and FEA imply

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\begin{pmatrix}\Delta y_{i,2}\\ Dy_i - \rho Dy_{i,-1}\end{pmatrix}\begin{pmatrix}\Delta y_{i,2} & (Dy_i - \rho Dy_{i,-1})'\end{pmatrix}\right] = \left[\begin{array}{cc}\sigma^2 + (1-\rho)^2\sigma_v^2 & -\sigma^2 \quad 0' \\ -\sigma^2 & \\ 0 & \sigma^2 H\end{array}\right], \tag{6}$$

where $H = DD'$.[9]

We will call the GMM estimator for $\rho$ in the panel AR(1) model that exploits all the 'moment conditions' in (6) the Fixed Effects HOCGMM estimator. This GMM estimator is the solution of a third-order polynomial in $\rho$ just like the Random Effects HOCGMM estimator for $\rho$.

The 'moment condition' $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}(\Delta y_{i,2})^2 = \sigma^2 + (1-\rho)^2\sigma_v^2$ is redundant for estimating $\rho$, because it is the only 'moment condition' in (6) that involves $\sigma_v^2$. [10] Cross-sectional heteroskedasticity of the $\varepsilon_{i,t}$ does not pose a problem for estimating $\rho$, since the $2+\delta$-th moments of the $\varepsilon_{i,t}$, $E(|\varepsilon_{i,t}|^{2+\delta})$, are bounded $\forall\ i \in \mathcal{I}$, and $\forall\ t \in \mathcal{T}$, and the cross-sectional average of the $\sigma_i^2$ converges.

Compared with the (Optimal) FE HOCGMM estimator for $\rho$, the (Optimal) RE HOCGMM estimator for $\rho$ exploits $T-2$ additional moment conditions, which involve levels of the data: $E(y_{i,1}(\Delta y_{i,t} - \rho\Delta y_{i,t-1})) = 0$, $t = 3, ..., T$. [11] In the appendix we prove the following result:

[9] We can extend these results to the panel AR(1) model with strictly exogenous regressors: $y_i = \rho y_{i,-1} + \iota\mu_i(1-\rho) + X_i\beta + \varepsilon_i$. In this case we define $v_{i,1} \equiv y_{i,1} - \mu_i - \overline{x}_i'\beta/(1-\rho)$ where $\overline{x}_i = X_i'\iota/(T-1)$. Then we obtain $\Delta y_{i,2} = (x_{i,2} - \overline{x}_i)'\beta + (\rho-1)v_{i,1} + \varepsilon_{i,2}$.

[10] See Breusch et al. (1999) for a definiton of redundancy of moment conditions.

[11] $E(y_{i,1}\Delta y_{i,2}) = (\rho-1)(\sigma_v^2 + \sigma_{\mu,v})$ and $E(y_{i,1}^2) = \sigma_v^2 + 2\sigma_{\mu,v} + \sigma_\mu^2$ are redundant for estimating $\rho$ because only they can be used to identify and estimate $\sigma_{\mu,v}$ and $\sigma_\mu^2$.

**Theorem 6** *Assume that SA and FEA$^*$ hold, $T \geq 3$, $|\rho| < 1$ and that $4+\delta$-th moments of the data exist for some $\delta > 0$.*

*(i) If $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) \neq 0$, then the Optimal Random Effects HOCGMM estimator for $\rho$, which exploits (5), is asymptotically more efficient than the Optimal Fixed Effects HOCGMM estimator for $\rho$, which exploits (6).*

*(ii) If $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) = 0$ and when optimal weighting is used, then all moment conditions involving levels of the data are redundant for estimating $\rho$ relative to the moment conditions in (6).*

**Corollary 1** *Assume that SA and FEA$^*$ hold, $T \geq 4$, $|\rho| < 1$ and that $4+\delta$-th moments of the data exist for some $\delta > 0$. If $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) = 0$ and when optimal weighting is used, then $E(y_{i,1}(\Delta y_{i,t} - \rho \Delta y_{i,t-1})) = 0$, $t = 3, ..., T$ are redundant for estimating $\rho$ relative to the Arellano-Bond moment conditions based on only differenced data that are given in (4).*

The proof of Corollary 1 is similar to the proof of Theorem 6.

The use of redundant moment conditions by the 2-step Optimal RE HOCGMM and Optimal RE Arellano-Bond GMM estimators does not affect their first-order asymptotic properties but does adversely affect their finite sample properties, cf e.g. Newey and Smith (2004).

When the data are i.i.d. and normal and $|\rho| < 1$, the Optimal RE HOCGMM estimator is asymptotically equivalent to the (correlated) RE ML estimator due to Chamberlain (1980) and Anderson and Hsiao (1982), and the Optimal FE HOCGMM estimator is asymptotically equivalent to the Transformed Maximum Likelihood (TML) estimator that has been proposed by Hsiao et al. (2002), see Kruiniger (2001, 2013). This TML estimator can also be viewed as the FE counterpart of (correlated) RE ML estimator. Furthermore, when $\rho = 1$, the REMLE and the FEMLE (i.e., the TMLE) are also aymptotically equivalent, cf Kruiniger (2013). When $\rho = 1$, $\rho$ is only second-order identified by (5) and (6), respectively, and the RE and FE HOCGMM estimators converge at rate $N^{1/4}$ and have a non-standard limiting distribution, cf Kruiniger (2013wp, 2017). We have the following counterpart of Theorem 6 for the MLEs:

**Theorem 7** *Assume that SA and FEA* hold,* $T \geq 3$, $|\rho| < 1$ *and that* $2+\delta$*-th moments of the data exist.*

*(i) If* $\sigma_\mu^2 \to \infty$ *and* $corr(\mu, v_1) \neq 0$, *then (a) the RE Quasi MLE for* $\rho$ *is consistent and (b) under normality of the data the REMLE for* $\rho$ *is asymptotically more efficient than the FEMLE for* $\rho$.

*(ii) If* $\sigma_\mu^2 \to \infty$ *and* $corr(\mu, v_1) = 0$, *then (a) the RE Quasi MLE for* $\rho$ *is consistent and (b) under normality of the data the REMLE for* $\rho$ *is asymptotically equivalent to the FEMLE for* $\rho$.

If the errors are heteroskedastic over time, then Theorem 7 only holds for $T > 3$.

### 2.2.3 Old and new Arellano-Bover and System GMM estimators

We will now consider the asymptotic properties of some other well-known GMM estimators for the panel AR(1) model when $\sigma_\mu^2 \to \infty$. Arellano and Bover (1995) noted that if mean stationarity holds as well, i.e., if $E(v_{i,1}) = 0$ and $corr(\mu, v_1) = 0$ also hold, then one can add $T-2$ moment conditions to those in (3):

$$E[(y_{i,t} - \rho y_{i,t-1})\Delta y_{i,t-1}] = 0 \text{ for } t = 3, ..., T. \tag{7}$$

A GMM estimator that exploits the moment conditions in (3) and (7) is known as a System (SYS) estimator.

The set of moment conditions in (3) and (7) is equivalent to a set that contains $T-2$ Arellano-Bond and $m$ Arellano-Bover type moment conditions:

$$E[y_{i,1}(\Delta y_{i,t} - \rho \Delta y_{i,t-1})] = 0 \text{ for } t = 3, ..., T, \tag{8}$$

and

$$E[(y_{i,t} - \rho y_{i,t-1})\Delta y_{i,s}] = 0 \text{ for } s = 2, ..., t-1 \text{ and } t = 3, ..., T. \tag{9}$$

A GMM estimator that only exploits the latter $m = (T-1)(T-2)/2$ moment conditions will be referred to as an Arellano-Bover (Arbov) estimator.

There exist no feasible optimal one-step weight matrices for the Arbov and SYS estimators, except when $\sigma_\mu^2 = 0$. Let $\Delta y_i^t = [\Delta y_{i,2} \ ... \ \Delta y_{i,t}]$, let $Z_i^{II} = diag(\Delta y_i^2, ..., \Delta y_i^{T-1})$ be a $(T-2)\times m$ block-diagonal matrix and let $Z_i^S = diag(Z_i^{II}, Z_i^I)$ be a $2(T-2) \times (m+T-2)$

block-diagonal matrix. The Arbov estimator exploits $E(Z_i^{II\prime}w_i) = 0$, whereas the System estimator exploits $E(Z_i^{S\prime}[w_i' \ \Delta w_i']') = 0$. When $\sigma_\mu^2 = 0$, optimal one-step weight matrices for the Arbov and SYS estimators are given by $W_{N,Arbov1} = (N^{-1}\sum_{i=1}^N Z_i^{II\prime}Z_i^{II})^{-1}$ and $W_{N,SYS1} = (N^{-1}\sum_{i=1}^N Z_i^{S\prime}AZ_i^S)^{-1}$, respectively, where $A$ is given by

$$A = \begin{bmatrix} I_{T-2} & C \\ C' & H \end{bmatrix},$$

where $C = C_{T-2} = E(\underline{\varepsilon}_i\Delta\underline{\varepsilon}_i')/\sigma_i^2$ is a $(T-2)\times(T-2)$ matrix with ones on the main diagonal, $-1$'$s$ on the first superdiagonal and zeros elsewhere. The one-step Arbov and SYS GMM estimators that use $W_{N,Arbov1}$ and $W_{N,SYS1}$, respectively, will be referred to as $\widehat{\rho}_{Arbov1}$ and $\widehat{\rho}_{SYS1}$, respectively. The two-step optimal Arbov and SYS estimators will be denoted as $\widehat{\rho}_{Arbov2}$ and $\widehat{\rho}_{SYS2}$, respectively. The Arbov and SYS estimators that use an arbitrary PD matrix $W_N$ as weight matrix will be denoted as $\widehat{\rho}_{Arbov}$ and $\widehat{\rho}_{SYS}$, respectively. In the appendix we prove the following result:

**Theorem 8** *Assume that SA and FEA* hold,* $E(v_{i,1}) = 0$, $corr(\mu, v_1) = 0$, $T = 3$, $|\rho| < 1$ *and* $\sigma_\mu^2 \to \infty$. *Then* $\widehat{\rho}_{Arbov}$, $\widehat{\rho}_{SYS}$ *and* $\widehat{\rho}_{SYS1}$ *are inconsistent.*

One can derive local asymptotic distributions of $\widehat{\rho}_{Arbov}$, $\widehat{\rho}_{SYS}$ and $\widehat{\rho}_{SYS1}$ similarly to those of $\widehat{\rho}_{ABlev}$.

We will now extend the results in Theorem 8 to $T > 3$. The set of $m$ Arbov moment conditions in $E(Z_i^{II\prime}w_i) = 0$ can be restated as $E(\tilde{Z}_i^{II\prime}[w_i' \ \Delta w_i']') = 0$ where $\tilde{Z}_i^{II} = diag(Z_i^L, Z_i^D)$ is a $2(T-2)\times m$ matrix with $Z_i^L = diag(\Delta y_{i,2}, ..., \Delta y_{i,T-1})$. We can extend the results given in Theorem 8 to $T > 3$ by using Lemma 3 from the appendix.

**Theorem 9** *Assume that SA and FEA* hold,* $E(v_{i,1}) = 0$, $corr(\mu, v_1) = 0$, $T > 3$, $|\rho| < 1$ *and* $\sigma_\mu^2 \to \infty$. *Assuming that* $\widehat{\rho}_{Arbov1}$ *and* $\widehat{\rho}_{Arbov}$ *exploit* $E(\tilde{Z}_i^{II\prime}[w_i' \ \Delta w_i']') = 0$ *in lieu of* $E(Z_i^{II\prime}w_i) = 0$, *then:*

*(i)* $\text{plim}_{N\to\infty}\widehat{\rho}_{Arbov1} = 1$ *and* $\text{plim}_{N\to\infty}\widehat{\rho}_{Arbov2} = \rho$; *(ii)* $\text{plim}_{N\to\infty}\widehat{\rho}_{Arbov} = 1$.

The proofs of Lemma 3 and Theorem 9 are similar to that of Theorem 8. The composite errors, $w_i$, and the lagged dependent variables, $y_{i,t-1}$, need to be scaled by $\sigma_\mu$ to achieve convergence of the averages of their cross products with the instruments. However, the blocks of the weight matrices of $\widehat{\rho}_{Arbov1}$ and $\widehat{\rho}_{Arbov}$ that correspond to

$E(Z_i^{L\prime}w_i/\sigma_\mu) = 0$ do not provide this scaling. These estimators will effectively only exploit the (scaled) moment conditions in $E(Z_i^{L\prime}w_i/\sigma_\mu) = 0$, which do not identify $\rho$ (because $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty}(N^{-1}\sum_{i=1}^N Z_i^{L\prime}y_{i,-1}/\sigma_\mu) = 0$), and hence will be inconsistent. When $T > 3$ and $\sigma_\mu^2 \to \infty$, the moment conditions in $E(Z_i^{D\prime}\Delta\underline{\varepsilon}_i) = 0$ still identify $\rho$ and a truly optimal Arbov estimator, e.g. $\widehat{\rho}_{Arbov2}$, is consistent and asymptotically equivalent to an optimal AB GMM estimator but, as noted above, no optimal 1-step Arbov estimator is available unless $\sigma_\mu^2 = 0$.

The 2-step optimal Arbov estimator, $\widehat{\rho}_{Arbov2}$, requires that the elements of the optimal weight matrix are consistently estimated. Any FE estimator can be used for this purpose, e.g. $\widehat{\rho}_{ABdif}$ or the Transformed ML estimator of Hsiao et al. (2002), but also $\widehat{\rho}_{AB1lev}$, $\widehat{\rho}_{AB2lev}$ or the RE Quasi ML estimator of Kruiniger (2013): when $T > 3$, the REQMLE for $\rho$ remains consistent when $\sigma_\mu^2/\sigma^2 \uparrow \infty$ and under normality of the data the REMLE for $\rho$ is asymptotically equivalent to the Transformed MLE for $\rho$ when $corr(\mu, v_1) = 0$, $\sigma_\mu^2/\sigma^2 \uparrow \infty$ and $N \to \infty$, see Theorem 7.

Note that the set of $m + T - 2$ moment conditions in $E(Z_i^{S\prime}[w_i' \; \Delta w_i']') = 0$ can be rewritten as $E(\tilde{Z}_i^{S\prime}[w_i' \; \Delta w_i']') = 0$, where $\tilde{Z}_i^S = diag(Z_i^L, Z_i^{AB})$ is a $2(T-2)\times$ $(T-2+m)$ matrix. Using Lemmas 2, 3 and 4 from the appendix we obtain the following results related to System estimators:

**Theorem 10** *Assume that SA and FEA* hold,* $E(v_{i,1}) = 0$, $corr(\mu, v_1) = 0$, $T > 3$, $|\rho| < 1$ *and* $\sigma_\mu^2 \to \infty$. *Let* $X_9 = ((1-\rho)X_{71}' \; X_{51}')'$ *and* $X_{10} = (X_{71}' \; X_{61}')'$ *where* $X_{51}$, $X_{61}$ *and* $X_{71}$ *are Gaussian r.v.'s that are defined in Lemmas 2 and 3. Assuming that* $\widehat{\rho}_{SYS1}$ *and* $\widehat{\rho}_{SYS}$ *exploit* $E(\tilde{Z}_i^{S\prime}[w_i' \; \Delta w_i']') = 0$ *in lieu of* $E(Z_i^{S\prime}[w_i' \; \Delta w_i']') = 0$, *then:*

*(i)* $\text{plim}_{N\to\infty}\widehat{\rho}_{SYS1} = 1$ *and* $\text{plim}_{N\to\infty}\widehat{\rho}_{SYS2} = \rho$;

*(ii)* $\widehat{\rho}_{SYS} - \rho \xrightarrow{d} (X_{10}'W_{11}X_{10})^{-1}X_{10}'W_{11}X_9$.

The proofs of Lemma 4 and Theorem 10 are similar to that of Theorem 8. The SYS1 (the SYS) estimator will effectively only exploit the (scaled) moment conditions in $E(Z_i^{L\prime}w_i/\sigma_\mu) = 0$ (and $E(Z_i^{I\prime}\Delta\underline{\varepsilon}_i/\sigma_\mu) = 0$), which do not identify $\rho$ (when $corr(\mu, v_1) = 0$), and hence will be inconsistent. Let $Z_i^{as} = diag(z_{i,4}^{as}, z_{i,5}^{as}, ..., z_{i,T}^{as}, 0)$ with $z_{i,t}^{as} = \Delta y_{i,2} + \sum_{s=3}^t \Delta w_{i,s}$. Note that $E(\tilde{Z}_i^{S\prime}[w_i' \; \Delta w_i']') = 0$ is equivalent to $E(Z_i^{AB\prime}\Delta\underline{\varepsilon}_i) = 0$, $E(Z_i^{as\prime}\Delta\underline{\varepsilon}_i) = 0$ and $E(w_{i,3}\Delta y_{i,2}) = 0$. When $T > 3$ and $\sigma_\mu^2 \to \infty$, $E(Z_i^{D\prime}\Delta\underline{\varepsilon}_i) = 0$ and $E(Z_i^{as\prime}\Delta\underline{\varepsilon}_i) = 0$ still identify $\rho$ and a truly optimal SYS estimator, e.g. $\widehat{\rho}_{SYS2}$, is consistent and asymptotically equivalent to the optimal FE GMM estimator for $\rho$ that exploits these moment conditions and also to the optimal FE GMM estimator for $\rho$ that exploits a version of the set of moment conditions in (6) that allows for heteroskedasticity over time. However, no optimal 1-step SYS estimator is available unless $\sigma_\mu^2 = 0$ and the 2-step 'optimal' System estimator that uses $\widehat{\rho}_{SYS1}$ to obtain an estimate of its weight matrix and is commonly used in the literature is inconsistent and asymptotically biased upwards. The 2-step optimal System estimator, $\widehat{\rho}_{SYS2}$, requires that the elements of the optimal weight matrix are consistently estimated. As the REQMLE remains consistent when $\sigma_\mu^2/\sigma^2 \uparrow \infty$ and has favorable properties, also when $\rho$ is near or equal to unity, cf. Kruiniger (2013), we propose a new System estimator that uses the REQMLE to estimate the optimal weight matrix. When $\rho$ is not close to unity, one could also use $\widehat{\rho}_{AB2lev}$ to estimate the optimal weight matrix of the System estimator leading to yet another version of the System estimator.

# 3 The finite sample performance of the estimators

Using Monte Carlo simulation, we study the finite sample properties of the 2-step optimal RE AB GMM estimator (AB2), the commonly used version of the 2-step 'optimal' System estimator (CSYS2; C for Conventional) of Arellano and Bover (1995) and Blundell and Bond (1998), and the RE and FE Quasi MLEs for $\rho$ in the model given in (1). The 2-step estimators AB2 and CSYS2 use the 1-step estimators $\widehat{\rho}_{AB1}$ and $\widehat{\rho}_{SYS1}$, respectively, to obtain an estimate of their weight matrix. We also consider the finite sample properties of the 2-step optimal System estimators that use the REQMLE or the AB2 estimator to obtain an estimate of the optimal weight matrix. These estimators are denoted as RSYS2 and ASYS2, respectively. Like the AB2 and SYS2 GMM estimators, the RE and FE Quasi MLEs that we consider allow for heteroskedasticity over time (and by doing so they actually allow for arbitrary heteroskedasticity).

In all simulation experiments the values of the error components, the $\mu_i$ and the $\varepsilon_{i,t}$, have been drawn from normal distributions with zero means and the $\varepsilon_{i,t}$ are homoskedastic and do not exhibit autocorrelation, i.e. $E(\varepsilon_i \varepsilon_i') = \sigma^2 I$.

We study how the properties of the estimators are affected if we change (1) the conditional distributions of the differences between the initial observations and the individual effects (the $v_{i,1}$) given the individual effects (the $\mu_i$), (2) the value of $N$, (3) the value of $\rho$, (4) the value of $\sigma_v^2/\sigma^2$ and/or (5) the value of $\sigma_\mu^2/\sigma^2$. W.l.o.g. we set $\sigma^2 = 1$.

We conducted the simulation experiments for $T = 6$, $N \in \{100, 500\}$, $\rho \in \{0.5, 0.9\}$, $\sigma_v^2 \in \{\frac{4}{3}, 5\frac{5}{19}\}$ and $\sigma_\mu^2 \in \{0, 1, 4, 10, 25\}$. Note that if $\rho = 0.5$, then $\sigma^2/(1-\rho^2) = \frac{4}{3}$, while if $\rho = 0.9$, then $\sigma^2/(1-\rho^2) = 5\frac{5}{19}$.

In order to assess how the conditional distribution of the $v_{i,1} = y_{i,1} - \mu_i$, $i = 1, ..., N$, affects the properties of the estimators, we conducted two types of experiments, which are identified by a capital letter: for one type of experiments the initial observations are drawn from 'mean stationary' distributions, (MS), $(y_{i,1} - \mu_i)|\mu_i \sim N(0, \sigma_v^2)$, whereas for the other type the initial observations are drawn from 'mean nonstationary' distributions, (MNS), $(y_{i,1} - (1 + \sqrt{0.2})\mu_i)|\mu_i \sim N(0, \sigma_v^2 - 0.2\sigma_\mu^2)$.

Under design MS, $\{y_{i,t}\}$ is mean stationary and $\{y_{i,t}\}$ is even strictly stationary if $\sigma_v^2$ is chosen equal to $\sigma^2/(1-\rho^2)$. Under design MNS, $\{y_{i,t}\}$ is nonstationary due to the fact that $E(\mu_i(y_{i,1} - \mu_i)) \neq 0$, although if $\sigma_v^2$ is chosen equal to $\sigma^2/(1-\rho^2)$, then $(y_{i,t} - \mu_i) \sim N(0, \sigma^2/(1-\rho^2))$ as is the case under design MS. In both designs $E(y_{i,t} - y_{i,t-1}) = 0$.

The AB GMM estimators and the QML estimators suffer from a weak moment conditions problem when $\rho$ is close to one.

When we ran the simulations, we did not impose homoskedasticity on the likelihood functions. However, we did maximize the likelihood functions subject to the restrictions $\sigma_t^2 > 0$ and $(T-1)\tilde{\sigma}_v^2 + \sigma_t^2 > 0$, $t = 2, ..., T$, on the variance parameters to ensure that the estimates of the covariance matrix of the composite errors were positive definite. We allowed for time effects by subtracting cross-sectional averages from the data.

Tables 1-4 report some of the simulation results on the properties of the estimators for $N = 100$ (and $T = 6$) in terms of the biases and mean squared errors (MSEs). These numbers have been multiplied by 100. The tables differ with respect to the assumptions made about the conditional distribution of the $v_{i,1}$ (design MS or MNS), the value of $\rho$ (0.5 or 0.9) and that of $\sigma_v^2$ (4/3 or $5\frac{5}{19}$). Appendix B reports further results. Inspection of the results in tables 1-4 and those in appendix B leads to the following conclusions:

1. The bias and the variance of the AB2 estimator depend on the conditional distribution of the $v_{i,1}$: they increase in $\sigma_\mu^2/\sigma^2$ under design MS (due to instruments becoming weaker) but decrease in $\sigma_\mu^2/\sigma^2$ under design MNS (due to instruments becoming stronger, cf. Hayakawa, 2009). The bias is larger when $\rho$ is closer to 1.

2. Similar to the bias and the variance of the AB2 estimator, those of the REQMLE increase in $\sigma_\mu^2/\sigma^2$ under design MS, but they decrease in $\sigma_\mu^2/\sigma^2$ under design MNS.

Table 1: MC results for estimators of $\rho$; $N = 100$, design MS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 0 | -1.85 | 0.76 | 0.16 | 0.62 | 0.86 | 1.10 | -0.45 | 0.46 | -0.37 | 0.49 | -0.98 | 0.53 |
| 1 | -3.37 | 1.23 | 0.32 | 0.78 | 0.90 | 1.15 | 0.45 | 0.56 | -0.03 | 0.58 | -1.28 | 0.67 |
| 4 | -5.38 | 2.09 | 0.71 | 1.00 | 0.95 | 1.15 | 3.22 | 0.85 | 0.87 | 0.74 | -1.60 | 0.93 |
| 10 | -6.53 | 2.64 | 1.01 | 1.14 | 1.02 | 1.16 | 7.64 | 1.68 | 1.20 | 0.89 | -2.28 | 1.22 |
| 25 | -7.26 | 3.17 | 1.32 | 1.24 | 1.23 | 1.19 | 17.6 | 5.01 | 1.62 | 1.11 | -3.26 | 1.84 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 2: MC results for estimators of $\rho$; $N = 100$, design MS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 0 | -8.94 | 3.44 | 2.65 | 1.81 | 2.70 | 2.18 | -1.22 | 0.55 | -0.24 | 0.75 | -4.80 | 1.36 |
| 1 | -10.8 | 4.23 | 3.69 | 2.00 | 3.17 | 2.22 | -1.04 | 0.58 | 0.16 | 0.80 | -5.73 | 1.60 |
| 4 | -16.1 | 7.74 | 4.97 | 2.23 | 3.22 | 2.14 | -0.59 | 0.60 | 0.89 | 0.96 | -8.89 | 2.91 |
| 10 | -25.4 | 14.7 | 5.66 | 2.45 | 3.13 | 2.20 | -0.08 | 0.64 | 1.39 | 1.12 | -15.0 | 5.91 |
| 25 | -37.8 | 26.9 | 5.68 | 2.50 | 3.48 | 2.20 | 1.87 | 0.66 | 2.43 | 1.41 | -25.2 | 13.2 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 3: MC results for estimators of $\rho$; $N = 100$, design MNS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 1 | -3.11 | 1.22 | 0.64 | 1.04 | 1.01 | 1.25 | -9.15 | 1.26 | -7.15 | 0.97 | -8.08 | 1.11 |
| 4 | -2.08 | 0.92 | 0.42 | 0.73 | 1.43 | 1.31 | 46.0 | 22.19 | 7.78 | 1.82 | 5.74 | 1.59 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 4: MC results for estimators of $\rho$; $N = 100$, design MNS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 1 | -36.9 | 24.2 | 4.30 | 2.41 | 3.54 | 2.33 | -0.41 | 0.19 | 0.34 | 0.36 | -14.9 | 5.02 |
| 4 | -28.7 | 16.5 | 4.43 | 2.31 | 3.56 | 2.29 | -2.66 | 0.45 | -1.15 | 0.70 | -16.2 | 5.55 |
| 10 | -18.8 | 9.12 | 4.22 | 2.23 | 3.80 | 2.28 | -2.86 | 0.91 | -0.97 | 1.26 | -14.3 | 4.88 |
| 25 | -9.68 | 3.74 | 2.89 | 1.81 | 3.23 | 2.17 | 11.1 | 1.68 | 5.71 | 1.64 | -4.02 | 2.48 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

The bias and the variance of the FEQMLE hardly change with $\sigma_\mu^2/\sigma^2$ under either design. Note that under design MS $Avar(\widehat{\rho}_{REML})/Avar(\widehat{\rho}_{FEML}) \uparrow 1$ as $\sigma_\mu^2/\sigma^2 \uparrow \infty$, where $Avar(\widehat{\rho})$ is the limiting variance of $\widehat{\rho}$ for $N \to \infty$, cf. Theorem 7. This result can be a bit 'misleading' when $\rho$ is not close to 1 but $\sigma_v^2/\sigma^2$ is large. In this case, when $\sigma_\mu^2/\sigma^2$ increases, the bias and the variance of the REQMLE first 'explode' and then decrease, whereas the FEQMLE, which is a restricted version of the REQMLE, is unaffected. Note that $\widehat{\rho}_{REML}$ is asy. equivalent to a GMM estimator that uses AB and AS moment conditions. Under design MS, if $\rho \ll 1$ and $\sigma_v^2/\sigma^2$ increases, then the former become stronger but are given (relatively) less weight while the latter become weaker. Nonetheless, once $\sigma_v^2/\sigma^2$ is fixed, if $\sigma_\mu^2/\sigma^2 \uparrow \infty$, both kinds of moment conditions become weaker and $\widehat{\rho}_{REML}$ eventually still tends to $\widehat{\rho}_{FEML}$.

3. The CSYS2 estimator performs better than the QMLEs when all the moment conditions that are exploited by the former estimator are valid and not weak (so the value of $\sigma_\mu^2/\sigma^2$ is not high) or when the value of $\rho$ is close to 1. However, if the value of $\rho$ is not close to 1 (so that (the effects of) the $\mu_i$ are not neutralized) and if mean stationarity is violated (as in design MNS) or the value of $\sigma_\mu^2/\sigma^2$ is high (so that some of the moment conditions exploited by the SYS estimators are weak and the 'optimal' weight matrix of the CSYS2 estimator is poorly estimated as a result of using $\widehat{\rho}_{SYS1}$), then the CSYS2 estimator is seriously biased (Bun and Windmeijer (2010) report similar findings when the value of $\sigma_\mu^2/\sigma^2$ is high), whereas the FEQMLE and often also the REQMLE (but see 2.) perform well.

4. When the value of $\rho$ is close to 1 (i.e., 0.9), both the CSYS2 estimator and the RSYS2 estimator usually perform well although the former performs (somewhat) better than the latter. The exception is when mean stationarity is violated, $\sigma_v^2/\sigma^2$ is high and $\sigma_\mu^2/\sigma^2$ is high. In this case, both estimators are biased and the RSYS2 estimator perform somewhat better than the CSYS2 estimator. When $\rho$ is close to 1, the ASYS2 estimator performs (much) worse than the CSYS2 and RSYS2 estimators, especially under design MS if $\sigma_\mu^2/\sigma^2$ is high and under design MNS (unless $\sigma_\mu^2/\sigma^2$ is very high) due to the poor performance of the AB2 estimator.

5. Unlike the CSYS2 estimator, the RSYS2 and ASYS2 estimators perform well under design MS when the value of $\rho$ is not close to 1 (e.g., 0.5) and the value of $\sigma_\mu^2/\sigma^2$ is high *if* the value of $\sigma_v^2/\sigma^2$ is moderate (e.g., 4/3). However, if the value of $\sigma_v^2/\sigma^2$

is high (e.g., $5\frac{5}{19}$), then, like the CSYS2 estimator, the RSYS2 estimator performs poorly under design MS when $\rho = 0.5$ and the value of $\sigma_\mu^2/\sigma^2$ is high due to the poor performance of the REQMLE. In this case the ASYS2 estimator still performs well.

6. When mean stationarity is violated, the value of $\rho$ is not close to 1 and the value of $\sigma_\mu^2/\sigma^2$ is not low, then the RSYS2 estimator and especially the ASYS2 estimator perform (much) better than the CSYS2 estimator, although the REQMLE is the preferred estimator in this case (because all the SYS estimators are inconsistent and hence biased in this case).

The reason for the last finding is that when instead of $\widehat{\rho}_{SYS1}$ a consistent estimator, such as $\widehat{\rho}_{AB2}$ or $\widehat{\rho}_{REML}$, is used to obtain an estimate of the 'optimal' weight matrix, then the weight given to the invalid Arbov moment conditions $E(Z_i^{L\prime} w_i) = 0$ drops relative to the weight given to the other, valid moment conditions that are exploited by an 'optimal' System estimator, i.e., $E(Z_i^{D\prime} \Delta \underline{\varepsilon}_i) = 0$ and $E(Z_i^{I\prime} \Delta \underline{\varepsilon}_i) = 0$. When a consistent estimator, e.g. $\widehat{\rho}_{AB2}$ or $\widehat{\rho}_{REML}$, is used to obtain an estimate of the 'optimal' weight matrix, an increase in the value of $\sigma_\mu^2/\sigma^2$ also leads to a drop in the relative weight given to the invalid Arbov moment conditions $E(Z_i^{L\prime} w_i) = 0$.

We also considered the usefulness of two versions of a test for weak instruments due to Montiel Olea and Pflueger (2013), and one test of underidentification due to Windmeijer (2018). The first two tests make use of cluster-robust F-test-statistics, $F_D$ and $F_L$, which are based on the first-stage regressions $\Delta y_{i,t-1} = \pi^D y_{i,t-2} + \omega_{i,t}^D$ and $y_{i,t-1} = \pi^L \Delta y_{i,t-1} + \omega_{i,t}^L$, respectively. $F_D$ ($F_L$) indicates weakness of the instruments for the model in differences (levels) when $F_D < 10$ ($F_L < 10$). As the models usually contain time dummies, we replace the observations $y_{i,s}$ by cross-sectionally demeaned versions $\widetilde{y}_{i,s} = y_{i,s} - N^{-1} \sum_{j=1}^{N} y_{j,s}$ when computing the values of $F_D$ and $F_L$. The underidentification test is only carried out for the model in differences and is based on Hansen's $J$-test-statistic for testing the validity of the 'orthogonality conditions' $E(\widetilde{y}_{i,1} \Delta \widetilde{y}_{i,t-1}) = 0$, $t = 3, ..., T$, where $\Delta \widetilde{y}_{i,t-1}$ are the 'errors' and $\widetilde{y}_{i,1}$ are the instruments. This test rejects underidentification in the model at the 5% level when $J > \chi^2_{0.95}(T-2)$.

Tables 5-8 report some of the simulation results on the properties of the tests for $N = 100$ (and $T = 6$) in terms of the average values of the test-statistics and the p-value of the $J$-test, $p_J$, as well as the relative frequency that some condition, e.g. non-rejection, is met. Appendix B again reports additional results.

Let $E_N(.)$ denote a sample average, e.g. $E_N(X) = N^{-1}\sum_{i=1}^{N} X_i$. Inspection of the results in tables 5-8 and those in appendix B leads to the following conclusions:

1. When using the rejection rules $p_J > 0.05$ and $F_D < 10$, the $J$ and the $F_D$ tests often lead to the same conclusion in the sense that $E_N(\mathbf{1}(p_J > 0.05)\mathbf{1}(F_D < 10))$ is close to $\min\{E_N(\mathbf{1}(p_J > 0.05)), E_N(\mathbf{1}(F_D < 10))\}$. Note also that usually $E_N(\mathbf{1}(F_D < 10)) < E_N(\mathbf{1}(p_J > 0.05))$. We focus below 'conservatively' only on the $F_D$ test.

2. Under design MS, $E_N(\mathbf{1}(F_D < 10))$ increases in the value of $\sigma_\mu^2/\sigma^2$, and for a given value of $\sigma_\mu^2/\sigma^2$, $E_N(\mathbf{1}(F_D < 10))$ is highest when $\rho$ is high ($\rho = 0.9$) and $\sigma_v^2/\sigma^2$ is low ($\sigma_v^2/\sigma^2 = 4/3$), while $E_N(\mathbf{1}(F_D < 10))$ is lowest when $\rho$ is low ($\rho = 0.5$) and $\sigma_v^2/\sigma^2$ is high ($\sigma_v^2/\sigma^2 = 5\frac{5}{19}$).

3. Under design MNS, if $\rho$ is low ($\rho = 0.5$), then $E_N(\mathbf{1}(F_D < 10)) = 0.00$, while if $\rho$ is high ($\rho = 0.9$), then $E_N(\mathbf{1}(F_D < 10))$ decreases in the value of $\sigma_\mu^2/\sigma^2$ from 0.87 for $\sigma_\mu^2/\sigma^2 = 1$ to 0.14 for $\sigma_\mu^2/\sigma^2 = 25$.

4. $E_N(\mathbf{1}(F_D < 10))$ is positively correlated with the bias of $\widehat{\rho}_{AB2}$ and when $\rho$ is high ($\rho = 0.9$), also with the bias of $\widehat{\rho}_{ASYS2}$.

5. Under design MS, $E_N(\mathbf{1}(F_L < 10))$ increases in the value of $\sigma_\mu^2/\sigma^2$, and for a given value of $\sigma_\mu^2/\sigma^2$, $E_N(\mathbf{1}(F_L < 10))$ is highest when $\rho$ is low ($\rho = 0.5$) and $\sigma_v^2/\sigma^2$ is high ($\sigma_v^2/\sigma^2 = 5\frac{5}{19}$), while $E_N(\mathbf{1}(F_L < 10))$ is lowest when $\rho$ is high ($\rho = 0.9$) and $\sigma_v^2/\sigma^2$ is low ($\sigma_v^2/\sigma^2 = 4/3$).

6. Under design MNS, if $\rho$ is low ($\rho = 0.5$), then $E_N(\mathbf{1}(F_L < 10))$ is very high when $\sigma_\mu^2/\sigma^2 = 4$ or 10 but $E_N(\mathbf{1}(F_L < 10)) = 0.00$ when $\sigma_\mu^2/\sigma^2 = 1$ or 25, while if $\rho$ is high ($\rho = 0.9$), then $E_N(\mathbf{1}(F_L < 10))$ is very high when $\sigma_\mu^2/\sigma^2 = 10$ or 25 but $E_N(\mathbf{1}(F_L < 10)) \leq 0.06$ when $\sigma_\mu^2/\sigma^2 = 1$ or 4.

A combination of the $F_D$ and $F_L$ tests can also be used to choose among $\widehat{\rho}_{CSYS2}$, $\widehat{\rho}_{RSYS2}$ and $\widehat{\rho}_{ASYS2}$ : if $\rho$ (e.g. $\widehat{\rho}_{FEQML}$) is high, then choose $\widehat{\rho}_{CSYS2}$ unless $F_D > 10$ and $F_L < 10$, in which case choose $\widehat{\rho}_{RSYS2}$, while if $\rho$ is not high and mean stationarity holds, then choose $\widehat{\rho}_{RSYS2}$ unless $F_D > 10$ and $F_L < 10$, in which case choose $\widehat{\rho}_{ASYS2}$.

We also investigated the quality of two estimators for the standard errors (SEs) of the GMM estimators for $\rho$, and for the bounds of 90% confidence intervals (CIs) for $\rho$. That is, we considered SEs based on first-order asymptotics and corrected SEs based on equation (2.6) in Windmeijer (2005).[12] The latter depend on the SE of the first-step estimator, which is given by another Windmeijer SE in the case of $\widehat{\rho}_{AB2}$ (the first-step estimator for $\widehat{\rho}_{ASYS2}$) and by a very simple non-parametric bootstrap estimate based on 100 replications in the case of $\widehat{\rho}_{REQML}$ (the first-step estimator for $\widehat{\rho}_{RSYS2}$).

Table 5: MC results for identification tests; $N = 100$, design MS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 29.4 | 0.00 | 0.00 | 0.00 | 137 | 0.00 | 136 | 0.00 | 0.00 | 0.00 |
| 1 | 22.5 | 0.00 | 0.00 | 0.00 | 69.5 | 0.00 | 69.0 | 0.00 | 0.00 | 0.00 |
| 4 | 14.0 | 0.03 | 0.17 | 0.46 | 28.1 | 0.01 | 28.1 | 0.01 | 0.01 | 0.01 |
| 10 | 9.12 | 0.14 | 0.57 | 0.84 | 13.0 | 0.31 | 13.0 | 0.32 | 0.27 | 0.30 |
| 25 | 6.45 | 0.29 | 0.80 | 0.95 | 5.75 | 0.88 | 5.82 | 0.88 | 0.77 | 0.12 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 6: MC results for identification tests; $N = 100$, design MS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 15.7 | 0.02 | 0.08 | 0.31 | 22.7 | 0.05 | 22.4 | 0.06 | 0.04 | 0.06 |
| 1 | 14.0 | 0.03 | 0.16 | 0.46 | 18.9 | 0.13 | 18.9 | 0.13 | 0.09 | 0.13 |
| 4 | 11.2 | 0.08 | 0.38 | 0.69 | 12.8 | 0.38 | 13.0 | 0.38 | 0.29 | 0.36 |
| 10 | 8.66 | 0.17 | 0.61 | 0.87 | 8.14 | 0.69 | 8.18 | 0.69 | 0.56 | 0.30 |
| 25 | 6.43 | 0.28 | 0.82 | 0.95 | 4.47 | 0.91 | 4.48 | 0.91 | 0.80 | 0.09 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 7: MC results for identification tests; $N$=100, design MNS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 15.2 | 0.02 | 0.10 | 0.35 | 53.5 | 0.00 | 67.9 | 0.00 | 0.00 | 0.00 |
| 4 | 26.3 | 0.00 | 0.00 | 0.00 | 53.2 | 0.00 | 5.14 | 0.90 | 0.00 | 0.90 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 8: MC results for identification tests; $N$=100, design MNS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4.72 | 0.42 | 0.92 | 0.99 | 5.37 | 0.87 | 86.8 | 0.00 | 0.83 | 0.00 |
| 4 | 6.59 | 0.28 | 0.80 | 0.95 | 6.28 | 0.82 | 21.3 | 0.06 | 0.74 | 0.05 |
| 10 | 9.74 | 0.12 | 0.51 | 0.80 | 9.70 | 0.59 | 3.43 | 0.96 | 0.45 | 0.41 |
| 25 | 15.1 | 0.02 | 0.11 | 0.36 | 18.3 | 0.14 | 2.23 | 0.98 | 0.08 | 0.84 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

[12]The derivation of the Windmeijer corrected SEs for the RSYS2 estimator and the ASYS2 estimator is provided in the appendix.

In the case of the CIs we report the number of times that the true value of $\rho$ was outside the CIs, that is, we report the rejection probabilities (RPs).

Tables 9 and 10 report the simulation results for design S with $\rho = 0.5$ and $\rho = 0.9$, respectively. For both tables, $\sigma_v^2 = \sigma^2/(1-\rho^2)$. Furthermore, $N = 100$ (and $T = 6$). The findings can be summarized as follows:

1. The conventional asymptotic SEs always seriously underestimate the standard deviations of the GMM estimators and are (much) lower than the Windmeijer corrected SEs. The latter are often close to the standard deviations of the AB2 estimator although they somewhat underestimate them when $\rho = 0.9$ and $\sigma_\mu^2/\sigma^2$ is high.

2. When $\rho = 0.5$, the Windmeijer corrected SEs somewhat underestimate the standard deviations of the RSYS2 and ASYS2 estimators and, when $\sigma_\mu^2/\sigma^2$ is not high, the CSYS2 estimator, but the Windmeijer corrected SE clearly underestimates the standard deviation of the CSYS2 estimator when $\sigma_\mu^2/\sigma^2$ is very high.

3. When $\rho = 0.9$, the Windmeijer corrected SEs somewhat underestimate the standard deviations of the CSYS2 and RSYS2 estimators and, when $\sigma_\mu^2/\sigma^2$ is not high, the ASYS2 estimator, but the Windmeijer corrected SE clearly underestimates the standard deviation of the ASYS2 estimator when $\sigma_\mu^2/\sigma^2$ is high.

4. When $\rho = 0.5$, the RPs of the CIs based on the RSYS2, the AB2 and the ASYS2 estimators, respectively, and a Windmeijer corrected SE are (much) closer to the nominal value of 10% than the RP of a similar CI based on the CSYS2 estimator and/or on a conventional SE, especially when $\sigma_\mu^2/\sigma^2$ is high.

5. When $\rho = 0.9$, the RP of the CI based on the CSYS2 estimator and the Windmeijer corrected SE is (much) closer to the nominal value of 10% than the RP of a similar CI based on any of the RSYS2, the AB2 and the ASYS2 estimators and/or on a conventional SE, which is always (much) too high, although the RP of the first CI is still somewhat high when $\sigma_\mu^2/\sigma^2 = 25$, namely about 0.17.

We conclude that under design MS with $\sigma_v^2 = \sigma^2/(1-\rho^2)$ the Windmeijer corrected SE corresponding to the preferred SYS2 estimator (RSYS2 when $\rho = 0.5$ and CSYS2 when $\rho = 0.9$) and the CI based on this estimator and this SE are close to being correct.

Table 9: MC results for st. errors and size; $N$=100, design MS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| | | AB2 | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | type | SE | RP | SE | RP | SE | RP | SE | RP |
| 0 | E | 0.085 | | 0.068 | | 0.070 | | 0.072 | |
| | A | 0.069 | 0.20 | 0.052 | 0.21 | 0.052 | 0.22 | 0.052 | 0.24 |
| | W | 0.083 | 0.12 | 0.065 | 0.13 | 0.066 | 0.13 | 0.067 | 0.14 |
| 1 | E | 0.106 | | 0.074 | | 0.076 | | 0.081 | |
| | A | 0.087 | 0.20 | 0.057 | 0.22 | 0.057 | 0.22 | 0.057 | 0.25 |
| | W | 0.104 | 0.12 | 0.072 | 0.12 | 0.074 | 0.13 | 0.077 | 0.13 |
| 4 | E | 0.134 | | 0.086 | | 0.085 | | 0.095 | |
| | A | 0.110 | 0.22 | 0.059 | 0.30 | 0.059 | 0.25 | 0.061 | 0.30 |
| | W | 0.133 | 0.14 | 0.083 | 0.16 | 0.082 | 0.14 | 0.090 | 0.15 |
| 10 | E | 0.149 | | 0.105 | | 0.094 | | 0.108 | |
| | A | 0.124 | 0.23 | 0.057 | 0.46 | 0.058 | 0.31 | 0.060 | 0.38 |
| | W | 0.151 | 0.14 | 0.095 | 0.25 | 0.090 | 0.13 | 0.105 | 0.15 |
| 25 | E | 0.163 | | 0.138 | | 0.104 | | 0.132 | |
| | A | 0.133 | 0.23 | 0.050 | 0.74 | 0.052 | 0.38 | 0.054 | 0.51 |
| | W | 0.163 | 0.15 | 0.103 | 0.50 | 0.099 | 0.12 | 0.124 | 0.14 |

Notes: 5000 Monte Carlo replications; E: based on empirical distribution; A: based on first-order asymptotic distribution; W: based on Windmeijer's corrected asymptotic standard errors; SE: standard deviation/error; RP: rejection probability (nominal size is 10%).

Table 10: MC results for st. errors and size; $N$=100, design MNS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| | | AB2 | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | type | SE | RP | SE | RP | SE | RP | SE | RP |
| 0 | E | 0.163 | | 0.073 | | 0.086 | | 0.106 | |
| | A | 0.130 | 0.26 | 0.052 | 0.23 | 0.053 | 0.31 | 0.058 | 0.36 |
| | W | 0.160 | 0.16 | 0.070 | 0.12 | 0.080 | 0.17 | 0.098 | 0.15 |
| 1 | E | 0.175 | | 0.075 | | 0.089 | | 0.113 | |
| | A | 0.142 | 0.26 | 0.053 | 0.24 | 0.053 | 0.34 | 0.060 | 0.39 |
| | W | 0.176 | 0.16 | 0.071 | 0.12 | 0.082 | 0.18 | 0.107 | 0.16 |
| 4 | E | 0.227 | | 0.077 | | 0.098 | | 0.146 | |
| | A | 0.168 | 0.32 | 0.053 | 0.26 | 0.052 | 0.40 | 0.063 | 0.49 |
| | W | 0.216 | 0.19 | 0.074 | 0.12 | 0.088 | 0.20 | 0.129 | 0.18 |
| 10 | E | 0.287 | | 0.080 | | 0.105 | | 0.191 | |
| | A | 0.198 | 0.39 | 0.052 | 0.30 | 0.052 | 0.46 | 0.066 | 0.59 |
| | W | 0.263 | 0.23 | 0.076 | 0.14 | 0.094 | 0.22 | 0.163 | 0.23 |
| 25 | E | 0.355 | | 0.079 | | 0.116 | | 0.261 | |
| | A | 0.233 | 0.48 | 0.047 | 0.37 | 0.049 | 0.53 | 0.066 | 0.75 |
| | W | 0.325 | 0.29 | 0.075 | 0.17 | 0.111 | 0.23 | 0.219 | 0.31 |

Notes: 5000 Monte Carlo replications; E: based on empirical distribution; A: based on first-order asymptotic distribution; W: based on Windmeijer's corrected asymptotic standard errors; SE: standard deviation/error; RP: rejection probability (nominal size is 10%).

# 4 Empirical example

To illustrate the importance of using a consistent estimator for the optimal weight matrix of the System estimator when the variance ratio is large, we revisit the economic growth study of Levine et al. (2000). They investigated the influence of financial intermediary development ($fid$) on growth while controlling for other possibly endogenous factors using a dynamic panel data model with time effects (see also Bazzi and Clemens (2013)):

$$\Delta \ln y_{i,t} = (fid_{i,t})\beta + (\ln y_{i,t-1})(\rho - 1) + x_{i,t}'\gamma + \delta_t + \eta_i + \varepsilon_{i,t},$$

where $y_{i,t}$ is GDP per capita. They used three alternative measures of $fid$: the ratio of Liquid Liabilities to GDP (LLY), the ratio of Commercial bank assets to commercial bank plus Central Bank Assets (CCBA) and the ratio of Credit issued to the Private sector to GDP (PRICR). The regressors they included in $x_{i,t}$ are: government size (GOV), openness to trade (TRADE), inflation (INFL), average years of secondary schooling (SEC) and black market premium (BMP), see Levine et al. (2000) for details.

We use an unbalanced panel dataset of Levine et al. (2000): it contains data for 74 countries and up to 7 five-year periods so that in total 437 observations are available. As the model contains a lag, 363 observations are available for the model. However, each moment condition exploited by the AB and SYS estimators involves observations from three periods so these estimators effectively exploit 289 observations. When the model is estimated with the Within estimator using the 363 (289) observations, the estimated variance ratio $\sigma_\eta^2/Var(\varepsilon_{i,t})$ is about 4.5 (25). Hence the variance ratio is fairly high.

In tables 11-13 we report estimation results based on the full dataset. In tables 30-32 in appendix C we report results based on a smaller dataset (of 253 observations) that contains data for at least three lags. In all tables we also report results for (Difference-in-) Hansen ($J$) tests of overidentifying restrictions, a heteroskedasticity and cluster robust Cragg-Donald ($CD$) underidentification test that is based on CU-GMM and discussed in Windmeijer (2018), and individual (i.e., per endogenous regressor) two-step GMM $J$ tests of underidentification in the spirit of Sanderson and Windmeijer (2016). In the tables LINIT denotes $\ln y_{i,t-1}$. We have excluded INFL and BMP from the models because the $CD$ tests for the SYS estimators for the models that include them do not reject underidentification. The "AB" (SYS) estimators use time dummies, $E(Z_{i,1}'\Delta w_i) = 0$ and $E(Z_{i,2}'w_i) = 0$ ($E(Z_{i,1}'\Delta w_i) = 0$ and $E(Z_{i,3}'w_i) = 0$) with $Z_{i,1}$, $Z_{i,2}$ and $Z_{i,3}$ defined below and $w_i = \eta_i + \varepsilon_i$. Note that the "AB estimator" is in reality also a System estimator;

it uses $E(Z_{i,2}'w_i) = 0$ to avoid underidentification due to poor predictability of $\Delta sec_{i,t}$. However, it does not rely on mean-stationarity of $\{\ln y_{i,t}\}$. Note also that $Z_{i,1}$, $Z_{i,2}$ and $Z_{i,3}$ are (partly) collapsed to avoid the problem of "too many instruments". The "ASYS2 estimator" uses the "AB2 estimator" to obtain an estimate of the optimal weight matrix.

$$Z_{i,1} = \begin{bmatrix} z_{i,1} & 0 & \tilde{z}_{i,1} & 0 & 0 & 0 \\ z_{i,2} & \Delta z_{i,2} & 0 & \tilde{z}_{i,2} & 0 & 0 \\ z_{i,3} & \Delta z_{i,3} & 0 & 0 & \tilde{z}_{i,3} & 0 \\ z_{i,4} & \Delta z_{i,4} & 0 & 0 & 0 & \tilde{z}_{i,4} \\ z_{i,5} & \Delta z_{i,5} & 0 & 0 & 0 & 0 \end{bmatrix}, \; Z_{i,2} = \begin{bmatrix} \Delta sec_{i,3} \\ \Delta sec_{i,4} \\ \Delta sec_{i,5} \\ \Delta sec_{i,6} \\ \Delta sec_{i,7} \end{bmatrix} \text{ and } Z_{i,3} = \begin{bmatrix} \Delta z_{i,2} \\ \Delta z_{i,3} \\ \Delta z_{i,4} \\ \Delta z_{i,5} \\ \Delta z_{i,6} \end{bmatrix}$$

with $z_{i,t} = (fid_{i,t}, \ln y_{i,t}, gov_{i,t}, trade_{i,t}, sec_{i,t+1})$ and $\tilde{z}_{i,t} = (fid_{i,t}, \ln y_{i,t})$.

The results in the tables show that the $CD$ tests of underidentification reject at the 5% or 10% level with one exception in table 31. The results for the individual $J$ tests suggest that the moment conditions have good predictive power for the regressors with one exception for $trade$ in table 13 and another exception for $sec$ in table 31. Furthermore, the $J$ tests of overidentifying restrictions and the $J$ difference tests of mean-stationarity do not reject. Therefore, we will focus on the CSYS2 and "ASYS2" estimates. These estimates have the expected sign in most cases except for a few insignificant ones, notably the estimates of the coefficients of $trade$ when $fid$ is measured by CCBA. The magnitude of the CSYS2 and ASYS2 estimates for the coefficient of $\ln y_{i,t-1}$ is also plausible: for the full sample these estimates lie in the interval $(-0.50, -0.34)$, while for the smaller sample they lie in the interval $(-0.54, -0.19)$ with one insignificant exception in table 30. The implied estimates for $\rho$ are well below unity. The main difference between the CSYS2 and ASYS2 estimation results is that all CSYS2 estimates of the effects of various measures of $fid$ on growth are positive and significant at the one-sided 5% level, whereas 50% of the ASYS2 estimates of these effects are insignificant. In addition, the ASYS2 estimates of the effect of $sec$ on growth tend to be significant unlike the CSYS2 estimates of that effect.

Summarizing, the CSYS2 estimation results may be unreliable as the variance ratio is high. In this situation the ASYS2 results may be more reliable. Although in some cases the ASYS2 estimates of the effects of $fid$ on growth are positive and significant and not very different from the CSYS2 estimates of these effects, in other cases the ASYS2 estimates of these effects are not significant unlike the CSYS2 estimates. In the latter cases, the ASYS2 estimates of the effect of $sec$ on growth are (positive and) significant, whereas the CSYS2 estimates of that effect are not significant. Overall, we conclude that the evidence for an effect of financial intermediary development on growth is mixed.

Table 11: Growth and Financial intermediation proxied by CCBA

| | "AB2" | | | CSYS2 | | | "ASYS2" | |
|---|---|---|---|---|---|---|---|---|
| | estim. | st. err. | $p_{un}$ | estim. | st. err. | $p_{un}$ | estim. | st. err. |
| CCBA | 3.17 | 1.62 | 0.00 | 3.88 | 1.49 | 0.00 | 3.76 | 1.45 |
| LINIT | -0.05 | 0.58 | 0.00 | -0.43 | 0.37 | 0.00 | -0.34 | 0.37 |
| GOV | -1.90 | 1.21 | 0.00 | -1.93 | 1.49 | 0.01 | -1.83 | 1.35 |
| TRADE | -1.35 | 1.53 | 0.00 | -1.19 | 1.61 | 0.00 | -1.27 | 1.52 |
| SEC | 0.99 | 0.45 | 0.00 | 1.24 | 0.51 | 0.02 | 1.07 | 0.52 |
| *CD test* | 25.91 (0.04) | | | 34.24 (0.02) | | | | |
| *J test* | 14.50 (0.41) | | | 20.24 (0.32) | | | 19.68 (0.35) | |
| *J-diff test* | 0.47 (0.49) | | | 5.74 (0.22) | | | 5.18 (0.27) | |

Notes: N=74; number of obs. = 289; time dummies are included; the estimators are defined in the text; apart from time dummies, the "AB" (SYS) estimators use 19 (23) instruments; Windmeijer robust standard errors are reported; $p_{un}$ is p-value of individual $J$ test of underidentification; the STATA command underid of Schaffer and Windmeijer (2020) was used to perform the underidentification tests; p-values are in parentheses; first $J$-diff test tests $E(Z'_{i,2}w_i) = 0$; second & third $J$-diff tests test $E(Z'_{i,3}w_i) = 0$ excluding $E(Z'_{i,2}w_i) = 0$.

Table 12: Growth and Financial intermediation proxied by LLY

| | "AB2" | | | CSYS2 | | | "ASYS2" | |
|---|---|---|---|---|---|---|---|---|
| | estim. | st. err. | $p_{un}$ | estim. | st. err. | $p_{un}$ | estim. | st. err. |
| LLY | 0.63 | 1.54 | 0.00 | 1.72 | 0.85 | 0.00 | 1.21 | 1.25 |
| LINIT | -0.40 | 0.65 | 0.00 | -0.33 | 0.32 | 0.01 | -0.43 | 0.46 |
| GOV | -1.32 | 2.10 | 0.01 | -1.33 | 1.88 | 0.03 | -1.55 | 1.93 |
| TRADE | 1.86 | 1.88 | 0.05 | 0.81 | 1.03 | 0.08 | 1.53 | 1.43 |
| SEC | 1.30 | 0.58 | 0.01 | 0.91 | 0.56 | 0.01 | 1.11 | 0.50 |
| *CD test* | 24.89 (0.05) | | | 30.44 (0.05) | | | | |
| *J test* | 19.87 (0.13) | | | 25.67 (0.11) | | | 23.94 (0.16) | |
| *J-diff test* | 0.12 (0.73) | | | 5.80 (0.21) | | | 4.07 (0.40) | |

Notes: see table 11.

Table 13: Growth and Financial intermediation proxied by PRICR

| | "AB2" | | | CSYS2 | | | "ASYS2" | |
|---|---|---|---|---|---|---|---|---|
| | estim. | st. err. | $p_{un}$ | estim. | st. err. | $p_{un}$ | estim. | st. err. |
| PRICR | 1.58 | 0.91 | 0.02 | 1.60 | 0.52 | 0.00 | 1.60 | 0.53 |
| LINIT | -0.61 | 0.62 | 0.00 | -0.34 | 0.32 | 0.02 | -0.50 | 0.42 |
| GOV | 0.27 | 2.30 | 0.00 | -0.38 | 1.46 | 0.01 | -0.14 | 1.58 |
| TRADE | 0.54 | 1.67 | 0.11 | 0.62 | 0.94 | 0.07 | 0.68 | 0.97 |
| SEC | 0.68 | 0.67 | 0.01 | 0.57 | 0.66 | 0.02 | 0.63 | 0.71 |
| *CD test* | 25.44 (0.04) | | | 32.32 (0.03) | | | | |
| *J test* | 16.74 (0.27) | | | 21.90 (0.24) | | | 20.90 (0.28) | |
| *J-diff test* | 2.55 (0.11) | | | 5.16 (0.27) | | | 4.16 (0.38) | |

Notes: see table 11.

# 5 Conclusions

In this paper we have studied estimation of the panel AR(1) model with arbitrary initial conditions and possibly heteroskedasticity in the time dimension. We have discussed necessary and sufficient conditions for consistency of FE and RE GMM estimators for this model, respectively. We found that a necessary condition for consistency of any GMM estimator for this model is that the average of the squared differences between the initial observations and the individual effects converges in probability. This condition can allow for cross-sectional dependence and heterogeneity of the data. A related but perhaps not very surprising result is that any consistent fixed effects estimator for the panel AR(1) model involves only differences of the data. In contrast, a random effects estimator also depends on levels of the data. When the data is mean stationary and the variance of the individual effects is infinite, then only moment conditions that only depend on differences of the data and moment conditions that can be combined to form such moment conditions help to identify the autoregressive parameter and a RE GMM estimator that exploits such moment conditions will be consistent provided that a suitable (e.g. optimal) weight matrix is used. In this situation the remaining moment conditions that involve levels of the data will be redundant. For instance, all the moment conditions that are exploited by the SYS estimator but not by the FE AB GMM estimator will be redundant. Furthermore, the 1-step System estimator does not use a suitable weight matrix and will be inconsistent, which in turn leads to inconsistency of the 2-step 'optimal' System estimator when the latter uses the former estimator to obtain an estimate of its weight matrix.

It follows that under heteroskedasticty over time and mean stationarity, for any RE GMM estimator to remain consistent when $\sigma_\mu^2/\sigma^2 \uparrow \infty$, one needs $T > 3$ and one needs to use (a consistent estimator of) a suitable (e.g. optimal) weight matrix such that the RE estimator converges to a FE estimator when $\sigma_\mu^2/\sigma^2 \uparrow \infty$ and $N \to \infty$. If $T > 3$ and $corr(\mu, v_1) = 0$, the REQMLE for $\rho$ is still consistent when $\sigma_\mu^2/\sigma^2 \uparrow \infty$ and, under the additional assumption of normality of the data, the RE MLE for $\rho$ is asymptotically equivalent to the FE MLE for $\rho$ when $\sigma_\mu^2/\sigma^2 \uparrow \infty$ and $N \to \infty$. If $T = 3$, $corr(\mu, v_1) = 0$ and the errors are heteroskedastic over time, then GMM and ML estimators for $\rho$ become inconsistent when $\sigma_\mu^2/\sigma^2 \uparrow \infty$.

When the data are mean stationary, $\rho$ is close to one and $\sigma_\mu^2/\sigma^2$ is not large, then the Arbov moment conditions $E(Z_i^{L\prime} w_i) = 0$ strongly identify $\rho$ but the AB moment conditions $E(Z_i^{I\prime} \Delta \underline{\varepsilon}_i) = 0$ and $E(Z_i^{D\prime} \Delta \underline{\varepsilon}_i) = 0$ weakly identify $\rho$, cf. Arellano and Bover (1995). However, when the data are mean stationary, $\rho$ is close to one and $\sigma_\mu^2/\sigma^2$ becomes large, then not only the (scaled) moment conditions in $E(Z_i^{I\prime} \Delta \underline{\varepsilon}_i/\sigma_\mu) = 0$ and $E(Z_i^{D\prime} \Delta \underline{\varepsilon}_i) = 0$ but also those in $E(Z_i^{L\prime} w_i/\sigma_\mu) = 0$ become weak, that is, in this case all moment conditions exploited by a SYS estimator seem to become weak. Nevertheless, even in this case the RSYS2 estimator, which uses a consistent estimator of the optimal weight matrix, and also the 'conventional' (C)SYS2 estimator, which uses the inconsistent and slightly upward biased SYS1 estimator to estimate the 'optimal' weight matrix, will still have a relatively small bias. This can be explained as follows. The set of moment conditions exploited by a SYS estimator contains/implies the set of moment conditions exploited by the non-linear GMM estimator of Ahn and Schmidt (1995, AS), i.e. those in lines one and three of (5), and a set of similar AS moment conditions that only involve differences of the data. Under normality of the data the optimal GMM estimators that exploit these sets of Ahn-Schmidt type moment conditions are asymptotically equivalent to the RE and FE MLE for $\rho$, respectively, which are consistent and have a convergence rate of $N^{1/4}$ close to and at the unit root, see Kruiniger (2013). Interestingly, the RSYS2 estimator also seems relatively robust to violations of mean stationarity.

Concluding, when ($T > 3$, $\sigma_\mu^2/\sigma^2$ may be large and) $\rho$ is close to unity, the preferred estimator is the 'conventional' (C)SYS2 estimator (unless $F_D > 10$ and $F_L < 10$, in which case it is the RSYS2 estimator) but if $\rho$ is not close to unity, e.g. around 0.5, then under mean stationarity the preferred estimator is the RSYS2 estimator (unless $F_D > 10$ and $F_L < 10$, in which case the FEQMLE is preferable) while under mean non-stationarity the preferred estimator is the REQMLE or perhaps its GMM counterpart, i.e., the non-linear (optimal) GMM estimator of Ahn and Schmidt (1995) if the data are very 'non-normal'. [13]

## REFERENCES

Ahn, S. C., Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68: 5-28.

[13] Unreported simulation results show that when the data are normal, then the finite sample properties of this (optimal) GMM estimator of Ahn and Schmidt (1995) are significantly worse than those of the RE(Q)MLE even though these estimators are asymptotically equivalent.

Ahn, S. C., Schmidt, P. (1997). Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation. *Journal of Econometrics* 76: 309-321.

Anderson, T. W., Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18: 47-82.

Arellano, M., Bond, S. R. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277-297.

Arellano, M., Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68: 29-51.

Bazzi, S., Clemens, M.A. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics* 5: 152-186.

Blundell, R.W., Bond, S. R. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115-143.

Breusch, T., Qian, H., Schmidt, P., Wyhowski, D. (1999). Redundancy of moment conditions. *Journal of Econometrics* 91: 89-111.

Bun, M.J.G., Kleibergen, F.R. (2017). Identification and inference in moments based analysis of linear dynamic panel data models. Unpublished manuscript, University of Amsterdam and Brown University.

Bun, M.J.G., Windmeijer, F. (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal* 13: 95-126.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies* 47: 225-238.

Hahn, J. (1997). Efficient estimation of panel data models with sequential moment restrictions. *Journal of Econometrics* 79: 1-21.

Hausman, J.A. (1978). Specification tests in econometrics. *Econometrica* 46: 1251–1271.

Hayakawa, K. (2007). Small sample bias properties of the system GMM estimator in dynamic panel data models. *Economics Letters* 95: 32-38.

Hayakawa, K. (2009). On the effect of mean-nonstationarity in dynamic panel data models. *Journal of Econometrics* 153: 133-135.

Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109: 107-150.

Kruiniger, H. (2001). On the estimation of panel regression models with fixed effects. Appeared in 2002 as WP no. 450, Queen Mary, Univ. of London. Revised in 2003.

Kruiniger, H. (2009). GMM estimation and inference in dynamic panel data models with persistent data. *Econometric Theory* 25: 1348-1391.

Kruiniger, H. (2013). Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. *Journal of Econometrics* 173: 175-188.

Kruiniger, H. (2013wp). Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. Working paper, Durham University, England.

Kruiniger, H. (2017). Uniform Quasi ML based inference for the panel AR(1) model with arbitrary initial conditions. Unpublished manuscript, Durham University.

Levine, R., Loayza, N., Beck, T. (2000). Financial intermediation and growth: causality and causes. *Journal of Monetary Economics* 46: 31-77.

Montiel Olea, J. L., and C. E. Pflueger (2013). A robust test for weak instruments. *Journal of Business and Economic Statistics* 31: 358–369.

Newey, N., McFadden, D. (1994). Large sample estimation and hypothesis testing. In R.F. Engle and D. McFadden (Eds.), Handbook of Econometrics, Volume IV, pp. 2111-2245. Amsterdam: North-Holland.

Newey, W.K., Smith, R.J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72: 219-255.

Sanderson, E., Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics* 190: 212-221.

Schaffer, M., and F. Windmeijer (2020). UNDERID: Stata module producing postestimation tests of under-and over-identification after linear IV estimation.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126: 25-51.

Windmeijer, F. (2018). Testing over- and underidentification in linear models, with applications to dynamic panel data and asset-pricing models. DP 18/696, University of Bristol, England.

# A Proofs

Proof of a claim in the text

We will show that SA and FEA imply that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\widetilde{v}_{i,1}\varepsilon_{i,t} = 0 \;\forall\, t \in \mathcal{T}$. FEA requires that for at most a finite number of individuals (indexed by $i$) in $\mathcal{I}$ it is true that $Var(\widetilde{v}_{i,1}) = O(N)$ but not $Var(\widetilde{v}_{i,1}) = O(1)$, and that for all other individuals in $\mathcal{I}$ it is true that $Var(\widetilde{v}_{i,1}) = O(1)$. SA implies that $E(\widetilde{v}_{i,1}\varepsilon_{i,t}) = 0 \;\forall\, i \in \mathcal{I}$. Furthermore, SA and FEA imply that the correlation between $\widetilde{v}_{i,1}\varepsilon_{i,t}$ and $\widetilde{v}_{j,1}\varepsilon_{j,t}$ is zero $\forall\, i,j \in \mathcal{I}$ with $i \neq j$ and $\sum_{i=1}^{N} Var(\widetilde{v}_{i,1}\varepsilon_{i,t}) = o(N^2)$ (Note that SA(vi) is redundant when the $\widetilde{v}_{i,1}$ and $\varepsilon_{i,t}$ are normally distributed). Then it follows from Chebyshev's LLN that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\widetilde{v}_{i,1}\varepsilon_{i,t} = 0 \;\forall\, t \in \mathcal{T}$. $\square$

Proof of Lemma 1

We will first show that SA and FEA imply that $\left|\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\widetilde{v}_{i,1}\varepsilon_{i,t}\right| < \infty \;\forall\, t \in \mathcal{T}$. Using SA and Markov's LLN we obtain that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\varepsilon_{i,t}^2 = \sigma^2 < \infty \;\forall\, t \in \mathcal{T}$. It follows from this result, FEA and the Cauchy-Schwarz inequality that $\left|\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\widetilde{v}_{i,1}\varepsilon_{i,t}\right| < \infty \;\forall\, t \in \mathcal{T}$.

"$\Longrightarrow$" From (2) we obtain $\Delta y_{i,t} = \rho^{t-2}\widetilde{v}_{i,1} + \rho^{t-2}\varepsilon_{i,2} + \sum_{s=3}^{t}\rho^{t-s}\Delta\varepsilon_{i,s}$, $\forall\, t \in \mathcal{T}$. Using these equalities, Markov's LLN and the Cauchy-Schwarz inequality, it follows from SA and FEA that $\text{plim}_{N\to\infty}\left|N^{-1}\sum_{i=1}^{N}(\Delta y_{i,s}\Delta y_{i,t})\right| < \infty$, $\forall\, s,t \in \mathcal{T}$.

"$\Longleftarrow$" Using $\Delta y_{i,2} = \widetilde{v}_{i,1} + \varepsilon_{i,2}$, Markov's LLN and SA gives $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\widetilde{v}_{i,1}(\widetilde{v}_{i,1} + 2\varepsilon_{i,2}) = \text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}(\Delta y_{i,2})^2 - \text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\varepsilon_{i,2}^2 < \infty$. It follows from this result that $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\widetilde{v}_{i,1}^2 < \infty$ and $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\widetilde{v}_{i,1}\varepsilon_{i,2} < \infty$ and hence FEA. $\square$

PROOF OF THEOREM 1

Recall that $v_{i,1} = y_{i,1} - \mu_i$. W.l.o.g. we have assumed that $T = 3$. Then the RE Arellano and Bond GMM estimator exploits only one moment condition, $E[y_{i,1}\Delta\varepsilon_{i,3}] = 0$, where $\Delta\varepsilon_{i,3} = \Delta y_{i,3} - \rho\Delta y_{i,2}$ :
$\widehat{\rho}_{ABlev} = [\sum_{i=1}^N y_{i,1}\Delta y_{i,2}]^{-1}[\sum_{i=1}^N y_{i,1}\Delta y_{i,3}] = \rho + [\sum_{i=1}^N y_{i,1}\Delta y_{i,2}]^{-1}[\sum_{i=1}^N y_{i,1}\Delta\varepsilon_{i,3}]$.
We can rewrite the numerator as $\sum_{i=1}^N y_{i,1}\Delta\varepsilon_{i,3} = \sum_{i=1}^N \mu_i\Delta\varepsilon_{i,3} + \sum_{i=1}^N v_{i,1}\Delta\varepsilon_{i,3}$
and the denominator as $\sum_{i=1}^N y_{i,1}\Delta y_{i,2} = \sum_{i=1}^N y_{i,1}[(\rho-1)(y_{i,1}-\mu_i) + \varepsilon_{i,2}] =$
$\sum_{i=1}^N[(\rho-1)v_{i,1}^2 + v_{i,1}\varepsilon_{i,2}] + \sum_{i=1}^N \mu_i[(\rho-1)v_{i,1} + \varepsilon_{i,2}]$.
Let us first consider the sums $N^{-0.5}\sum_{i=1}^N y_{i,1}\varepsilon_{i,t}$ , $t = 2, 3$. Note that $y_{i,1} = v_{i,1} + \mu_i$.
Hence $Var(N^{-0.5}\sum_{i=1}^N y_{i,1}\varepsilon_{i,t}) = O(N^{-1}\sum_{i=1}^N E(\mu_i^2))$. However, since $\sigma_\mu^2 \to \infty$, we have $N^{-1}\sum_{i=1}^N E(\mu_i^2) \to \infty$. As $\lim_{N\to\infty}\lim_{\sigma_\mu\to\infty} N^{-1}\sum_{i=1}^N E(\mu_i^2)/\sigma_\mu^2 = 1$, we will rescale the numerator and denominator of $\widehat{\rho}_{ABlev}$ by $\sigma_\mu$ in order to guarantee that they converge $N \to \infty$. Since $\lim_{\sigma_\mu\to\infty} E(y_{i,1}\Delta\varepsilon_{i,3}/\sigma_\mu) = 0\ \forall\ i \in \mathcal{I}$ and
$\lim_{N\to\infty}\lim_{\sigma_\mu\to\infty} Var(N^{-1}\sum_{i=1}^N y_{i,1}\Delta\varepsilon_{i,3}/\sigma_\mu) = 0$,
we obtain $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty} N^{-1}\sum_{i=1}^N y_{i,1}\Delta\varepsilon_{i,3}/\sigma_\mu = 0$.
After scaling by $\sigma_\mu$, the denominator becomes $\sum_{i=1}^N y_{i,1}\Delta y_{i,2}/\sigma_\mu$. It is easily seen that $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty} N^{-1}\sum_{i=1}^N[(\rho-1)v_{i,1}^2 + v_{i,1}\varepsilon_{i,2}]/\sigma_\mu = 0$ and that
$\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty} N^{-1}\sum_{i=1}^N \mu_i[(\rho-1)v_{i,1} + \varepsilon_{i,2}]/\sigma_\mu = (\rho-1)corr(\mu, v_1)\sigma_v$.
If $corr(\mu, v_1) \neq 0$, then $corr(\mu, v_1)\sigma_v \neq 0$ and it follows that $\widehat{\rho}_{ABlev}$ is consistent.
If $corr(\mu, v_1) = 0$, $(\lim_{\sigma_\mu\to\infty} N^{-0.5}\sum_{i=1}^N \mu_i[(\rho-1)v_{i,1} + \varepsilon_{i,2}]/\sigma_\mu) \xrightarrow{d} X_2$,
$\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty} N^{-0.5}\sum_{i=1}^N[(\rho-1)v_{i,1}^2 + v_{i,1}\varepsilon_{i,2}]/\sigma_\mu = 0$, and
$(\lim_{\sigma_\mu\to\infty} N^{-0.5}\sum_{i=1}^N y_{i,1}\Delta\varepsilon_{i,3}/\sigma_\mu) \xrightarrow{d} X_1$, where

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2\zeta_1 & -\zeta_1 \\ -\zeta_1 & \zeta_1 + (\rho-1)^2\zeta_2 \end{bmatrix}\right), \text{ with } 0 < \zeta_1, \zeta_2 < \infty.$$

Thus if $corr(\mu, v_1) = 0$, $\widehat{\rho}_{ABlev} \xrightarrow{d} X_1/X_2$ which implies that $\widehat{\rho}_{ABlev}$ is inconsistent. □

**Lemma 2** *Assume that SA and FEA* hold,* $T > 3$, $|\rho| < 1$ *and* $\sigma_\mu^2 \to \infty$. *Then:*

$N^{-1/2}\sum_{i=1}^N Z_i^{I\prime}\Delta\underline{\varepsilon}_i/\sigma_\mu \xrightarrow{d} X_{51}$;

*if* $corr(\mu, v_1) \neq 0$, $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N Z_i^{I\prime}\Delta y_{i,-1}/\sigma_\mu \neq 0$;

*if* $corr(\mu, v_1) = 0$, $N^{-1/2}\sum_{i=1}^N Z_i^{I\prime}\Delta y_{i,-1}/\sigma_\mu \xrightarrow{d} X_{61}$;

$N^{-1/2}\sum_{i=1}^{N} Z_i^{D\prime}\Delta\underline{\varepsilon}_i \xrightarrow{d} X_{52}$ *and* $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} Z_i^{D\prime}\Delta y_{i,-1} \neq 0$ *with* $(X_{51}'\ X_{61}')' \sim N(0,\Sigma_{56,11})$ *and* $X_{52}\sim N(0,\Sigma_{5,2})$.

*Furthermore,* $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} Z_i^{I\prime}HZ_i^{I}/\sigma_\mu^2 = \tilde{W}_{AB1}^{11}$;

*if* $corr(\mu,v_1)\neq 0$, $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} Z_i^{I\prime}HZ_i^{D}/\sigma_\mu = \tilde{W}_{AB1}^{12}\neq 0$;

*if* $corr(\mu,v_1) = 0$, $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} Z_i^{I\prime}HZ_i^{D}/\sigma_\mu = \tilde{W}_{AB1}^{12} = 0$;

$\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} Z_i^{D\prime}HZ_i^{D} = \tilde{W}_{AB1}^{22}$ *and* $\tilde{W}_{AB1}$ *is nonsingular.*

## Proof of Theorem 3

Recall that $\sigma_v^2 = \text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} v_{i,1}^2 < \infty$,

$\lim_{N\to\infty} N^{-1-p}\sum_{i=1}^{N} E(\mu_i\varepsilon_{i,t})^2 = \zeta_1$, $t=2,3$ with $0<\zeta_1<\infty$ and

$\lim_{N\to\infty} N^{-1-p}\sum_{i=1}^{N} E(\mu_i v_{i,1})^2 = \zeta_2$ with $0<\zeta_2<\infty$.

Then the following results can easily be verified:

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

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

If $0\leq p<1$, $\text{plim}_{N\to\infty} N^{-1}\sum \mu_i[(\rho-1)v_{i,1}+\varepsilon_{i,2}] = 0$.

If $0<\zeta_0 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E(y_{i,1}\Delta\varepsilon_{i,3})^2 < \infty$, and $\{y_{i,1}\Delta\varepsilon_{i,3}\}$ satisfies the Lindeberg condition, then $N^{-0.5}\sum_{i=1}^{N} y_{i,1}\Delta\varepsilon_{i,3} \xrightarrow{d} N(0,\zeta_0)$.

If $0\leq p<1$, then $\text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} y_{i,1}\Delta y_{i,2} = p\lim_{N\to\infty} N^{-1}\sum[(\rho-1)v_{i,1}^2 + v_{i,1}\varepsilon_{i,2}] + \text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mu_i[(\rho-1)v_{i,1}+\varepsilon_{i,2}] = (\rho-1)\sigma_v^2$.

So, if $p=0$, $0<\zeta_0 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E(y_{i,1}\Delta\varepsilon_{i,3})^2 < \infty$, and $\{y_{i,1}\Delta\varepsilon_{i,3}\}$ satisfies the Lindeberg condition, then $N^{0.5}(\widehat{\rho}_{ABlev}-\rho) \xrightarrow{d} N(0,[(\rho-1)\sigma_v^2]^{-2}\zeta_0)$.

If $p>0$ and $\{\mu_i\Delta\varepsilon_{i,3}/N^{0.5p}\}$ satisfies the Lindeberg condition, then $N^{-0.5(1+p)}\sum_{i=1}^{N} y_{i,1}\Delta\varepsilon_{i,3} = N^{-0.5(1+p)}\sum_{i=1}^{N} v_{i,1}\Delta\varepsilon_{i,3} + N^{-0.5(1+p)}\sum_{i=1}^{N}\mu_i\Delta\varepsilon_{i,3} \xrightarrow{d} X_1 \sim N(0,2\zeta_1)$; so, if $0<p<1$, $N^{0.5(1-p)}(\widehat{\rho}_{ABlev}-\rho) \xrightarrow{d} N(0,[(\rho-1)\sigma_v^2]^{-2}2\zeta_1)$;

Finally, if $p\geq 1$, and both $\{\mu_i\varepsilon_{i,2}/N^{0.5p}\}$ and $\{\mu_i v_{i,1}/N^{0.5p}\}$ satisfy the Lindeberg condition, then

if $p=1$, $N^{-0.5(1+p)}\sum_{i=1}^{N} y_{i,1}\Delta y_{i,2} \xrightarrow{d} (\rho-1)\sigma_v^2 + X_2$ with $X_2\sim N(0,\zeta_1+(\rho-1)^2\zeta_2)$;

if $p>1$, $N^{-0.5(1+p)}\sum_{i=1}^{N} y_{i,1}\Delta y_{i,2} \xrightarrow{d} X_2$. Furthermore, if $p\geq 0$, $Cov(X_1,X_2) = -\zeta_1$.

So, if $\{\mu_i\varepsilon_{i,t}/N^{0.5p}\}$, $t=2,3$, and $\{\mu_i v_{i,1}/N^{0.5p}\}$ satisfy the Lindeberg condition, then if $p = 1$, $(\widehat{\rho}_{ABlev} - \rho) \xrightarrow{d} \frac{X_1}{(\rho-1)\sigma_v^2+X_2}$, while if $p > 1$, $(\widehat{\rho}_{ABlev} - \rho) \xrightarrow{d} \frac{X_1}{X_2}$. $\square$

Proof of Theorem 6

Below we assume that the data are i.i.d. but the proof is similar for i.h.d. data.

The OREHOCGMM estimator for $\rho$ exploits all the moment conditions that are exploited by the OFEHOCGMM estimator for $\rho$, e.g. (6), plus the following moment conditions: $E(y_{i,1}\Delta\varepsilon_{i,t}) = E(y_{i,1}(\Delta y_{i,t} - \rho\Delta y_{i,t-1})) = 0$, $t = 3, ..., T$, $E(y_{i,1}\Delta y_{i,2}) = (\rho - 1)(\sigma_v^2 + \sigma_{\mu,v})$ and $E(y_{i,1}^2) = \sigma_v^2 + 2\sigma_{\mu,v} + \sigma_\mu^2$, where we have used $E(y_{i,1}\Delta y_{i,2}) = E((v_{i,1} + \mu_i)((\rho - 1)v_{i,1} + \varepsilon_{i,2}))$. The moment conditions corresponding to $E(y_{i,1}\Delta y_{i,2})$ and $E(y_{i,1}^2)$ are redundant for estimating $\rho$ because only they can be used to identify $\sigma_{\mu,v}$ and $\sigma_\mu^2$.

The $0.5(T-1)T$ moment conditions in (6) involve $\rho$, $\sigma^2$ and $\sigma_v^2$.

$E((\Delta y_{i,2})^2) = \sigma^2 + (1-\rho)^2\sigma_v^2$ is the only moment condition that involves $\sigma_v^2$ and is therefore redundant for estimating $\rho$. $E(\Delta y_{i,2}\Delta\varepsilon_{i,3}) = -\sigma^2$ can be used to remove $\sigma^2$ from the other moment conditions in (6). Thus the OFEHOCGMM estimator for $\rho$ is equal to a GMM estimator that optimally exploits $0.5(T-1)T - 2$ moment conditions which only involve $\rho$. Let us collect these moment conditions in the vector $E(m_2(\rho)) = 0$. Let $D_2 = E(\frac{dm_2(\rho)}{d\rho})$ and let $\Omega_{22} = E(m_2(\rho)m_2(\rho)')$. Then $Avar(\widehat{\rho}_{OFEHOCGMM}) = (D_2'\Omega_{22}^{-1}D_2)^{-1}$.

We will prove the first part of the theorem by showing that $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ is not redundant relative to the moment conditions that are optimally exploited by the FEHOCGMM estimator when $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) \neq 0$.

Consider the estimator $\widehat{\rho}_{OHOCGMM^+}$ which optimally exploits the moment conditions $E(m_\perp(\rho)) = 0$ and $E(m_2(\rho)) = 0$ where $m_\perp(r) = y_{i,1}(\Delta y_{i,3} - r\Delta y_{i,2}) - m_2'(r) \times \Omega_{22}^{-1}E(m_2(\rho)y_{i,1}\Delta\varepsilon_{i,3})$. Note that $E(m_\perp(\rho)) = 0$ and $E(m_2(\rho)) = 0$ are equivalent to $E(y_{i,1}(\Delta y_{i,3} - \rho\Delta y_{i,2})) = 0$ and $E(m_2(\rho)) = 0$. It is easily seen that $E(m_2(\rho)m_\perp(\rho)) = 0$ so the optimal weighting matrix used by $\widehat{\rho}_{OHOCGMM^+}$ is block-diagonal, i.e., $diag(w_{11}^{-1}, \Omega_{22}^{-1})$ with $w_{11} = E(m_\perp(\rho)m_\perp(\rho)')$ and $[Avar(\widehat{\rho}_{OHOCGMM^+})]^{-1} = D_2\Omega_{22}^{-1}D_2 + [E(\frac{dm_\perp(\rho)}{d\rho})]^2 w_{11}^{-1}$. It follows that $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ is not redundant relative to $E(m_2(\rho)) = 0$ if and only if $[E(\frac{dm_\perp(\rho)}{d\rho})]^2 w_{11}^{-1} > 0$.

Now $E(\frac{dm_\perp(\rho)}{d\rho}) = -E(y_{i,1}\Delta y_{i,2}) - E(\frac{dm_2'(\rho)}{d\rho})\Omega_{22}^{-1}E(m_2(\rho)y_{i,1}\Delta\varepsilon_{i,3})$, $E(m_2(\rho)y_{i,1}\Delta\varepsilon_{i,3}) = E(y_{i,1}\Delta y_{i,2}\Delta\varepsilon_{i,3}\Delta\varepsilon_{i,4},\ 0,\ 0,\ ...,\ 0,\ y_{i,1}\Delta y_{i,2}(\Delta\varepsilon_{i,3})^2 vech(H)')'$, $E(y_{i,1}\Delta y_{i,2}\Delta\varepsilon_{i,3}\Delta\varepsilon_{i,4}) = -\sigma^2 E(y_{i,1}\Delta y_{i,2})$ and $E(y_{i,1}\Delta y_{i,2}(\Delta\varepsilon_{i,3})^2) = 2\sigma^2 E(y_{i,1}\Delta y_{i,2})$. Note also that $\Delta y_{i,2} = (\rho - 1)(y_1 - \mu) + \varepsilon_2$.

When $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) \neq 0$, then $\lim_{\sigma_\mu\to\infty}([E(y_{i,1}\Delta y_{i,2})/\sigma_\mu]^2\sigma_\mu^2 w_{11}^{-1}) = [(\rho-1)corr(\mu,v_1)\sigma_v]^2 \lim_{\sigma_\mu\to\infty}(\sigma_\mu^2 w_{11}^{-1}) > 0$ and hence $\lim_{\sigma_\mu\to\infty}([E(\frac{dm_\perp(\rho)}{d\rho})]^2 w_{11}^{-1}) > 0$. This concludes the proof of the first part of the theorem.

We will now prove the second part of the theorem, where we assume $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) = 0$.

Let $m_1(\rho) = [y_{i,1}(\Delta y_{i,3} - \rho\Delta y_{i,2}),\ y_{i,1}(\Delta y_{i,4} - \rho\Delta y_{i,3}), ...,\ y_{i,1}(\Delta y_{i,T} - \rho\Delta y_{i,T-1})]'$ and let $m(\rho) = [m_1'(\rho)\ m_2'(\rho)]'$. The OREHOCGMM estimator for $\rho$ optimally exploits $E(m(\rho)) = 0$.

Let $D = [D_1'\ D_2']'$ with $D_i = E(\frac{dm_i(\rho)}{d\rho})$, and let $\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$ with $\Omega_{ij} = E(m_i(\rho)m_j(\rho)')$. Moreover, let $S_1 = \sigma_\mu^{-1} I_{T-2}$ and $S = diag(S_1, I_{(T-1)^2-2})$. Then $Avar(\widehat{\rho}_{OREHOCGMM}) = (D'\Omega^{-1}D)^{-1} = ((SD)'(S\Omega S)^{-1}SD)^{-1}$.

We can easily verify the following: $D_{1,k} = E(\frac{d}{d\rho}(y_{i,1}(\Delta y_{i,k+2} - \rho\Delta y_{i,k+1}))) = -E(y_{i,1}\Delta y_{i,k+1})$, $k = 1, ..., T-2$. Note that $\Delta y_{i,t} = \rho^{t-2}\Delta y_{i,2} + \sum_{s=3}^{t}\rho^{t-s}\Delta\varepsilon_{i,s}$ and $\Delta y_{i,2} = (\rho-1)v_{i,1} + \varepsilon_{i,2}$. Since $corr(\mu, v_1) = 0$, $\lim_{\sigma_\mu\to\infty} E(y_{i,1}\Delta y_{i,2}/\sigma_\mu) = \lim_{\sigma_\mu\to\infty}(\rho - 1)(\sigma_v^2/\sigma_\mu + corr(\mu, v_1)\sigma_v) = 0$. Moreover, $\lim_{\sigma_\mu\to\infty}[E(y_{i,1}\Delta\varepsilon_{i,t}/\sigma_\mu)] = 0$, $t = 3, ..., T$. It follows that $\lim_{\sigma_\mu\to\infty} S_1 D_1 = 0$.

We also have $\lim_{\sigma_\mu\to\infty} S_1\Omega_{11}S_1 = \widetilde{\Omega}_{11}$, where $\widetilde{\Omega}_{11,st} = \lim_{\sigma_\mu\to\infty} E(y_{i,1}^2\Delta\varepsilon_{i,s+2}\Delta\varepsilon_{i,s+2}/\sigma_\mu^2)$, $s, t = 1, ..., T-2$, that is $\widetilde{\Omega}_{11,st} = 2\sigma^2$ if $s = t$; $\widetilde{\Omega}_{11,st} = -\sigma^2$ if $|s-t| = 1$, and $\widetilde{\Omega}_{11,st} = 0$ if $|s-t| \geq 2$. It follows that $\widetilde{\Omega}_{11}$ is a finite PDS matrix.

We will now consider $\lim_{\sigma_\mu\to\infty} S_1\Omega_{12}$. We note that $m_2(\rho)$ only depends on squares and products of the differences of the data, $\Delta y_{i,t}$, $t = 2, ..., T$ and that $\Delta y_{i,t}$ can be written as sums of $v_{i,1}$ and $\varepsilon_{i,s}$, $s = 2, ..., t$, for all $t = 2, ..., T$.

Moreover, $\lim_{\sigma_\mu\to\infty} E(y_{i,1}\varepsilon_{i,t}/\sigma_\mu) = 0$, $t = 2, ..., T$, and $\lim_{\sigma_\mu\to\infty} E(y_{i,1}v_{i,1}/\sigma_\mu) = corr(\mu, v_1)\sigma_v = 0$ since $corr(\mu, v_1) = 0$. It follows that $\lim_{\sigma_\mu\to\infty} S_1\Omega_{12} = \widetilde{\Omega}_{12} = 0$.

It also follows that $\lim_{\sigma_\mu\to\infty} S\Omega S = \widetilde{\Omega}$ is a finite PDS matrix.

Finally, $\lim_{\sigma_\mu\to\infty}[Avar(\widehat{\rho}_{OREHOCGMM})]^{-1} = \lim_{\sigma_\mu\to\infty}((SD)'(S\Omega S)^{-1}SD) = [\lim_{\sigma_\mu\to\infty}(SD)'] \times [\lim_{\sigma_\mu\to\infty}(S\Omega S)^{-1}] \times [\lim_{\sigma_\mu\to\infty} SD] = [0\ \ D_2'](\widetilde{\Omega})^{-1}[0\ \ D_2']' = (D_2'\Omega_{22}^{-1}D_2) = [Avar(\widehat{\rho}_{OFEHOCGMM})]^{-1}$.

We conclude that if $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) = 0$, then $E(m_1(\rho)) = 0$ is redundant relative to $E(m_2(\rho)) = 0$ when the elements of the latter are optimally weighted by $\Omega_{22}^{-1}$. $\square$

Proof of Theorem 7

We only prove parts (i-a) and (ii). Part (i-b) follows from asymptotic equivalence of the MLEs to Optimal HOCGMM estimators and Theorem 6(i).

The RE (Quasi) MLE for $\theta_0 = (\pi,\ \rho,\ \sigma^2,\ \widetilde{\sigma}_v^2)$ in the AR(1) panel model is based on the (quasi) likelihood function corresponding to the following augmented model

$$\widetilde{\Delta} y_i = \rho \widetilde{\Delta} y_{i,-1} + \pi y_{i,1}\iota + u_i, \tag{10}$$

with $u_i = \widetilde{v}_{i,1}\iota + \varepsilon_i \sim N(0,\Sigma)$, where $\widetilde{\Delta} y_i = y_i - y_{i,1}\iota$, $\widetilde{\Delta} y_{i,-1} = y_{i,-1} - y_{i,1}\iota$ and $\Sigma = E(u_i u_i') = \widetilde{\sigma}_v^2 \iota\iota' + \sigma^2 I_{T-1}$. The FE (Quasi) MLE for $(\rho,\ \sigma^2,\ \widetilde{\sigma}_v^2)$ is based on the same model but without the term $\pi y_{i,1}\iota$. After dividing by $N$ and scaling $y_{i,1}$ by $1/\sigma_\mu$ whenever it is useful the (quasi) log-likelihood function is given by

$$\begin{aligned} \log L &= -\frac{1}{2}(T-1)\log 2\pi - \frac{1}{2}\log|S| \\ &\quad -\frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1} - p\sigma_\mu(\frac{y_{i,1}}{\sigma_\mu})\iota)' S^{-1}(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1} - p\sigma_\mu(\frac{y_{i,1}}{\sigma_\mu})\iota). \end{aligned} \tag{11}$$

Note that $S = s^2 Q + (s^2 + (T-1)\widetilde{s}_v^2)\frac{1}{T-1}\iota\iota'$, where $Q = I_{T-1} - \frac{1}{T-1}\iota\iota'$. It follows that $S^{-1} = \frac{1}{s^2}Q + \frac{1}{s^2+(T-1)\widetilde{s}_v^2}\frac{1}{T-1}\iota\iota'$ and $|S| = s^{2(T-2)}(s^2 + (T-1)\widetilde{s}_v^2)$.

The RE (Quasi) ML estimator is defined as the global maximizer of the (quasi) log-likelihood function. It is also a solution of the likelihood equations for $\pi$, $\rho$, $\sigma^2$ and $\widetilde{\sigma}_v^2$ which are given by:

$$\frac{\partial \log L}{\partial p} = \frac{1}{N}\sum_{i=1}^{N}[y_{i,1}\iota' S^{-1}(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1} - p y_{i,1}\iota)] = 0, \tag{12}$$

$$\frac{\partial \log L}{\partial r} = \frac{1}{N}\sum_{i=1}^{N}[\widetilde{\Delta} y_{i,-1}' S^{-1}(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1} - p y_{i,1}\iota)] = 0, \tag{13}$$

$$\begin{aligned} \frac{\partial \log L}{\partial s^2} &= -\frac{(T-2)}{2s^2} - \frac{1}{2\widetilde{s}^2} + \frac{1}{2s^4}\frac{1}{N}\sum_{i=1}^{N}[(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1})' Q(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1})] \\ &\quad + \frac{1}{2\widetilde{s}^4}\frac{1}{T-1}\frac{1}{N}\sum_{i=1}^{N}[\iota'(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1} - p y_{i,1}\iota)]^2 = 0, \end{aligned} \tag{14}$$

and

$$\frac{\partial \log L}{\partial \widetilde{s}_v^2} = -\frac{(T-1)}{2\widetilde{s}^2} + \frac{1}{2\widetilde{s}^4}\frac{1}{N}\sum_{i=1}^{N}[\iota'(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1} - p y_{i,1}\iota)]^2 = 0, \tag{15}$$

where $\widetilde{s}^2 = s^2 + (T-1)\widetilde{s}_v^2$.

Let $\widetilde{\pi} = \pi\sigma_\mu$, $\widetilde{\theta}_0 = (\widetilde{\pi},\ \rho,\ \sigma^2,\ \widetilde{\sigma}_v^2)$ and $\widetilde{p} = p\sigma_\mu$. By taking probability limits of the likelihood equations and replacing $\widetilde{\Delta} y_i$ in (12)-(15) by the RHS of (10) or $\rho\widetilde{\Delta} y_{i,-1} + \pi\sigma_\mu(\frac{y_{i,1}}{\sigma_\mu})\iota + u_i$, we can find the probability limit of the RE(Q)MLE for $\theta_0$. If $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) \neq 0$, we get $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty} N^{-1}\sum_{i=1}^{N}(y_{i,1}\widetilde{\Delta} y_t/\sigma_\mu) \neq 0$, $t = 2, ..., T$, $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty}\widehat{\widetilde{\theta}}_{REQML} = \widetilde{\theta}_0$ and $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty}\widehat{\pi}_{REQML} = 0$. Finally, if $\sigma_\mu^2 \to \infty$ and $corr(\mu, v_1) = 0$, we have $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty} N^{-1}\sum_{i=1}^{N}(y_{i,1}\widetilde{\Delta} y_t/\sigma_\mu) = 0$, $t = 2, ..., T$, $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty}\widehat{\widetilde{\theta}}_{REQML} = \widetilde{\theta}_0$ with $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty}\widehat{\widetilde{\pi}}_{REQML} = 0$ and hence $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty}\widehat{\pi}_{REQML} = 0$. If we also assume that the data are normally distributed, we find that $\lim_{\sigma_\mu\to\infty} E(\frac{\partial \log L}{\partial \widetilde{\theta}}\frac{\partial \log L}{\partial \widetilde{\theta}'}|_{\widetilde{\theta}_0}) = -\lim_{\sigma_\mu\to\infty} E(\frac{\partial^2 \log L}{\partial \widetilde{\theta}\partial \widetilde{\theta}'}|_{\widetilde{\theta}_0})$ is block-diagonal with a block corresponding to $\widetilde{\pi}$ and that $\widehat{\rho}_{REML}$ is asymptotically equivalent to $\widehat{\rho}_{FEML}$. Here we considered the asymptotic distribution of $\widehat{\widetilde{\theta}}$ ($\widehat{\widetilde{\pi}}$) rather than that of $\widehat{\theta}$ ($\widehat{\pi}$) only because $\widehat{\widetilde{\pi}}$ does not have a degenerate asymptotic distribution unlike $\widehat{\pi}$. Note that the asymptotic distribution of $\widehat{\theta}$ ($\widehat{\pi}$) can easily be obtained from that of $\widehat{\widetilde{\theta}}$ ($\widehat{\widetilde{\pi}}$): it only requires scaling of the asymptotic variance of $\widehat{\widetilde{\pi}}$ by $1/\sigma_\mu^2$. $\square$

Proof of Theorem 8

We first prove that $\widehat{\rho}_{Arbov}$ is inconsistent.

$\widehat{\rho}_{Arbov} = [\sum_{i=1}^{N}(y_{i,2}\Delta y_{i,2})]^{-1}\sum_{i=1}^{N}(y_{i,3}\Delta y_{i,2}) = \rho + [\sum_{i=1}^{N}(y_{i,2}\Delta y_{i,2})]^{-1} \times \sum_{i=1}^{N}(\Delta y_{i,2}((1-\rho)\mu_i + \varepsilon_{i,3}))$. Recall that $y_{i,1} = v_{i,1} + \mu_i$ and $\mu_i \equiv \sigma_\mu\widetilde{\mu}_i$. Note that $\sigma_\mu^2 \to \infty$ implies $N^{-1}\sum_{i=1}^{N} E(\mu_i^2) \to \infty$. We will rescale the numerator and the denominator of $\widehat{\rho}_{Arbov}$ by $\sigma_\mu$ to guarantee that they converge when $N \to \infty$.

Since $\lim_{\sigma_\mu\to\infty} E(\Delta y_{i,2}((1-\rho)\mu_i + \varepsilon_{i,3})/\sigma_\mu) = 0$, $\forall\ i \in \mathcal{I}$,

$\lim_{\sigma_\mu\to\infty} Var(\Delta y_{i,2}((1-\rho)\mu_i + \varepsilon_{i,3})/\sigma_\mu) < \infty$, $\forall\ i \in \mathcal{I}$,

$\lim_{\sigma_\mu\to\infty} E(y_{i,2}\Delta y_{i,2}/\sigma_\mu) = (\rho-1)corr(\mu, v_1)\sigma_v = 0$, $\forall\ i \in \mathcal{I}$ and

$\lim_{\sigma_\mu\to\infty} Var(y_{i,2}\Delta y_{i,2}/\sigma_\mu) < \infty$, $\forall\ i \in \mathcal{I}$, we obtain

$(\lim_{\sigma_\mu\to\infty} N^{-0.5}\sum_{i=1}^{N} y_{i,2}\Delta y_{i,2}/\sigma_\mu) \xrightarrow{d} X_{71}$ and

$(\lim_{\sigma_\mu\to\infty} N^{-0.5}\sum_{i=1}^{N} \Delta y_{i,2}((1-\rho)\mu_i + \varepsilon_{i,3})/\sigma_\mu) \xrightarrow{d} (1-\rho)X_{71}$, where

$X_{71} \sim N(0, \Sigma_{7,1})$ with $0 < \Sigma_{7,1} < \infty$. It follows that $\text{plim}_{N\to\infty}\lim_{\sigma_\mu\to\infty}\widehat{\rho}_{Arbov} =$

$\rho + (1-\rho) = 1$ so that $\widehat{\rho}_{Arbov}$ is inconsistent.

Using the above results and the results in the proof of Theorem 1 for the RE AB GMM estimator we can similarly show that $\widehat{\rho}_{SYS}$ and $\widehat{\rho}_{SYS1}$ are inconsistent. $\square$

**Lemma 3** *Assume that SA and FEA* hold, $E(v_{i,1}) = 0$, $corr(\mu, v_1) = 0$, $T > 3$, $|\rho| < 1$ and $\sigma_\mu^2 \to \infty$. Then:*

$N^{-1/2} \sum_{i=1}^N Z_i^{L\prime} w_i / \sigma_\mu \xrightarrow{d} (1-\rho) X_{71}$ *and*

$N^{-1/2} \sum_{i=1}^N Z_i^{L\prime} y_{i,-1} / \sigma_\mu \xrightarrow{d} X_{71}$ *with* $X_{71} \sim N(0, \Sigma_{7,1})$.

*Furthermore,* $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{L\prime} Z_i^L = \tilde{W}_{Arbov1}^{11}$;

$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{L\prime} C Z_i^D = \tilde{W}_{Arbov1}^{12}$;

$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{D\prime} H Z_i^D = \tilde{W}_{Arbov1}^{22}$ *and* $\tilde{W}_{Arbov1}$ *is nonsingular.*

**Lemma 4** *Assume that SA and FEA* hold, $E(v_{i,1}) = 0$, $corr(\mu, v_1) = 0$, $T > 3$, $|\rho| < 1$ and $\sigma_\mu^2 \to \infty$. Then:*

$(X_{51}' \ X_{61}' \ X_{71}')' \sim N(0, \Sigma_{567,111})$.

*Furthermore,* $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{L\prime} Z_i^L = \tilde{W}_{SYS1}^{11}$;

$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{L\prime} C Z_i^I / \sigma_\mu = \tilde{W}_{SYS1}^{12} = 0$;

$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{I\prime} H Z_i^I / \sigma_\mu^2 = \tilde{W}_{SYS1}^{22}$;

$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{L\prime} C Z_i^D = \tilde{W}_{SYS1}^{13}$;

$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{I\prime} H Z_i^D / \sigma_\mu = \tilde{W}_{SYS1}^{23} = 0$;

$\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N Z_i^{D\prime} H Z_i^D = \tilde{W}_{SYS1}^{33}$ *and* $\tilde{W}_{SYS1}$ *is nonsingular.*

### Derivation of Windmeijer (2005) corrected standard errors for the RSYS2 estimator and the ASYS2 estimator

We adopt the notation of Windmeijer (2005). Let $\widehat{\theta}_2$ be an efficient 2-step GMM estimator for $\theta_0$, e.g. a 2-step System estimator, based on the vector of linear moment conditions $E(g_i(\theta_0)) = 0$, and let $\widehat{\theta}_1$ be an initial consistent GMM estimator for $\theta_0$ that is based on a subset of $E(g_i(\theta_0)) = 0$ and used in an estimate of the optimal weight matrix, viz. $W_N^{-1}(\widehat{\theta}_1)$, where $W_N(\theta) = N^{-1} \sum_{i=1}^N g_i(\theta) g_i(\theta)'$. Let $\overline{g}(\theta) = N^{-1} \sum_{i=1}^N g_i(\theta)$

and $C(\theta) = \partial \overline{g}(\theta)/\partial \theta'$. From the bottom of p. 29 in Windmeijer (2005), we have

$$\begin{aligned} \widehat{\theta}_2 - \theta_0 &= -(C'W_N^{-1}(\widehat{\theta}_1)C)^{-1}C'W_N^{-1}(\widehat{\theta}_1)\overline{g}(\theta_0) \\ &= -(C'W_N^{-1}(\theta_0)C)^{-1}C'W_N^{-1}(\theta_0)\overline{g}(\theta_0) + D_{\theta_0,W_N(\theta_0)}(\widehat{\theta}_1 - \theta_0) + o_p(N^{-1}), \quad (16) \end{aligned}$$

where the matrix $D_{\theta_0,W_N(\theta_0)}$ is defined at the top of p. 30 in Windmeijer (2005). We note that $D_{\theta_0,W_N(\theta_0)}(\widehat{\theta}_1 - \theta_0) = O_p(N^{-1})$, cf. p. 29 in Windmeijer (2005). The variance of the first term in (16) can be estimated by $\frac{1}{N}(C'W_N^{-1}(\widehat{\theta}_1)C)^{-1}$. The covariance between the first and the second term in (16) can be approximated by $Cov(\widehat{\theta}_2 - \theta_0, \widehat{\theta}_1 - \theta_0)D'_{\widehat{\theta}_2,W_N(\widehat{\theta}_1)}$, where we have used that

$$-(C'W_N^{-1}(\theta_0)C)^{-1}C'W_N^{-1}(\theta_0)\overline{g}(\theta_0) = \widehat{\theta}_2 - \theta_0 + O_p(N^{-1})$$

and $\widehat{\theta}_2 - \theta_0 = O_p(N^{-1/2})$. As $\widehat{\theta}_1$ is less efficient than $\widehat{\theta}_2$, it follows from a well known result in Hausman (1978) that $Cov(\widehat{\theta}_2 - \theta_0, \widehat{\theta}_1 - \theta_0) = Var(\widehat{\theta}_2)$. The latter can also be estimated by $\frac{1}{N}(C'W_N^{-1}(\widehat{\theta}_1)C)^{-1}$. Noting that the variance of the second term in (16) is given by $D_{\theta_0,W_N(\theta_0)}Var(\widehat{\theta}_1)D'_{\theta_0,W_N(\theta_0)}$, we obtain that the corrected asymptotic variance of $\widehat{\theta}_2$ is given by equation (2.6) in Windmeijer (2005).

Next we note that the moment conditions exploited by the System estimator for $\rho$ imply those exploited by the Optimal RE Conditional GMM estimator for $\rho$ (i.e., those in lines one and three of (5)) and that the latter is at least as efficient as $\widehat{\rho}_{REQML}$. As the first-step estimators used by the RSYS2 estimator and the ASYS2 estimator for $\rho$, namely $\widehat{\rho}_{REQML}$ and $\widehat{\rho}_{AB2}$, respectively, exploit less information than the corresponding second-step estimators and hence are less efficient than them, we can conclude that we can use equation (2.6) in Windmeijer (2005) to compute the Windmeijer corrected SEs for the RSYS2 estimator and the ASYS2 estimator. □

# B SOME OF THE MONTE CARLO RESULTS.

Table 14: MC results for estimators of $\rho$; $N = 100$, design MS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 0 | -0.76 | 0.31 | -0.06 | 0.25 | 0.20 | 0.43 | -0.05 | 0.27 | 0.01 | 0.27 | -0.19 | 0.29 |
| 1 | -1.13 | 0.40 | -0.06 | 0.27 | 0.19 | 0.44 | 0.69 | 0.31 | 0.26 | 0.31 | -0.07 | 0.34 |
| 4 | -1.74 | 0.65 | 1.31 | 0.99 | 0.15 | 0.39 | 2.96 | 0.51 | 1.30 | 0.53 | 0.04 | 0.41 |
| 10 | -3.29 | 1.18 | 14.1 | 8.36 | 0.23 | 0.44 | 7.59 | 1.51 | 7.95 | 2.82 | -0.57 | 0.58 |
| 25 | -6.33 | 2.44 | 7.91 | 4.90 | 0.26 | 0.42 | 18.6 | 5.57 | 6.49 | 2.97 | -1.91 | 1.03 |
| 100 | -13.7 | 6.57 | 1.08 | 0.90 | 0.34 | 0.45 | 37.8 | 16.2 | 1.57 | 0.94 | -7.93 | 3.79 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 15: MC results for estimators of $\rho$; $N = 100$, design MS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 4/3$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 0 | -23.2 | 12.2 | 4.33 | 2.25 | 3.52 | 2.28 | 0.14 | 0.19 | 0.93 | 0.35 | -8.36 | 2.27 |
| 1 | -30.9 | 18.49 | 3.99 | 2.32 | 3.63 | 2.28 | 0.05 | 0.21 | 0.71 | 0.40 | -12.2 | 3.75 |
| 4 | -37.9 | 25.8 | 4.71 | 2.44 | 3.69 | 2.34 | 0.10 | 0.23 | 0.81 | 0.49 | -17.5 | 6.67 |
| 10 | -42.5 | 30.1 | 4.08 | 2.38 | 3.39 | 2.30 | -0.08 | 0.30 | 0.23 | 0.65 | -23.1 | 10.18 |
| 25 | -44.7 | 33.0 | 4.54 | 2.41 | 3.98 | 2.32 | 0.66 | 0.36 | 0.63 | 0.97 | -29.9 | 15.94 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 16: MC results for estimators of $\rho$; $N$=100, design MNS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 1 | -2.96 | 1.22 | 0.75 | 1.08 | 1.04 | 1.25 | -9.21 | 1.28 | -7.11 | 0.97 | -8.03 | 1.11 |
| 4 | -2.14 | 0.89 | 0.48 | 0.73 | 1.56 | 1.38 | 46.0 | 22.2 | 7.84 | 1.79 | 5.73 | 1.53 |
| 10 | -1.50 | 0.59 | 0.10 | 0.48 | 0.71 | 0.85 | 51.6 | 26.7 | 5.34 | 1.06 | 3.90 | 0.98 |
| 25 | -0.76 | 0.33 | -0.05 | 0.27 | 0.12 | 0.40 | 46.2 | 21.4 | 1.67 | 0.38 | 1.00 | 0.39 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 17: MC results for estimators of $\rho$; $N$=100, design MNS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 4/3$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 1 | -37.4 | 24.2 | 3.99 | 2.41 | 3.61 | 2.33 | -0.41 | 0.18 | 0.32 | 0.36 | -14.9 | 4.96 |
| 4 | -29.1 | 17.0 | 4.24 | 2.30 | 3.63 | 2.28 | -2.55 | 0.45 | -1.08 | 0.70 | -16.3 | 5.70 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 18: MC results for identification tests; $N = 100$, design MS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 34.4 | 0.00 | 0.00 | 0.00 | 273 | 0.00 | 16.0 | 0.28 | 0.00 | 0.28 |
| 1 | 32.3 | 0.00 | 0.00 | 0.00 | 163 | 0.00 | 10.8 | 0.53 | 0.00 | 0.53 |
| 4 | 27.1 | 0.00 | 0.00 | 0.00 | 73.6 | 0.00 | 5.85 | 0.83 | 0.00 | 0.83 |
| 10 | 21.1 | 0.00 | 0.00 | 0.04 | 36.2 | 0.00 | 3.43 | 0.95 | 0.00 | 0.95 |
| 25 | 14.2 | 0.03 | 0.15 | 0.43 | 16.7 | 0.20 | 2.02 | 0.98 | 0.12 | 0.80 |
| 100 | 8.05 | 0.20 | 0.67 | 0.88 | 5.04 | 0.88 | 1.24 | 0.99 | 0.66 | 0.12 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 19: MC results for identification tests; $N = 100$, design MS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 4/3$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 8.05 | 0.19 | 0.67 | 0.89 | 10.5 | 0.53 | 110 | 0.00 | 0.45 | 0.00 |
| 1 | 6.35 | 0.29 | 0.82 | 0.95 | 7.67 | 0.73 | 77.3 | 0.00 | 0.67 | 0.00 |
| 4 | 5.08 | 0.39 | 0.90 | 0.98 | 4.37 | 0.92 | 40.6 | 0.00 | 0.86 | 0.00 |
| 10 | 4.51 | 0.44 | 0.94 | 0.99 | 2.68 | 0.97 | 20.9 | 0.08 | 0.92 | 0.02 |
| 25 | 4.25 | 0.47 | 0.95 | 0.99 | 1.63 | 0.99 | 10.0 | 0.56 | 0.94 | 0.01 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 20: MC results for identification tests; $N$=100, design MNS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 15.3 | 0.02 | 0.11 | 0.34 | 53.7 | 0.00 | 67.9 | 0.00 | 0.00 | 0.00 |
| 4 | 26.3 | 0.00 | 0.00 | 0.01 | 53.0 | 0.00 | 5.10 | 0.91 | 0.00 | 0.91 |
| 10 | 31.5 | 0.00 | 0.00 | 0.00 | 73.2 | 0.00 | 4.55 | 0.92 | 0.00 | 0.92 |
| 25 | 34.2 | 0.00 | 0.00 | 0.00 | 120 | 0.00 | 40.4 | 0.00 | 0.00 | 0.00 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 21: MC results for identification tests; $N$=100, design MNS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 4/3$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4.70 | 0.42 | 0.93 | 0.99 | 5.27 | 0.87 | 86.9 | 0.00 | 0.83 | 0.00 |
| 4 | 6.46 | 0.28 | 0.81 | 0.95 | 6.17 | 0.83 | 21.4 | 0.06 | 0.75 | 0.05 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 22: MC results for estimators of $\rho$; $N = 500$, design MS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 0 | -0.30 | 0.13 | 0.01 | 0.11 | 0.05 | 0.14 | -0.07 | 0.08 | -0.10 | 0.08 | -0.11 | 0.08 |
| 1 | -0.62 | 0.21 | 0.01 | 0.12 | 0.03 | 0.13 | 0.11 | 0.10 | 0.01 | 0.10 | -0.06 | 0.10 |
| 4 | -0.98 | 0.34 | 0.00 | 0.13 | 0.02 | 0.14 | 0.43 | 0.13 | 0.18 | 0.12 | -0.00 | 0.13 |
| 10 | -1.33 | 0.44 | -0.05 | 0.13 | -0.05 | 0.13 | 0.82 | 0.16 | 0.22 | 0.13 | -0.09 | 0.14 |
| 25 | -1.55 | 0.54 | 0.04 | 0.13 | 0.04 | 0.13 | 2.58 | 0.35 | 0.33 | 0.16 | -0.08 | 0.19 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 23: MC results for estimators of $\rho$; $N = 500$, design MS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 0 | -1.89 | 0.49 | 0.42 | 0.45 | 1.31 | 0.81 | -0.24 | 0.09 | -0.24 | 0.10 | -0.49 | 0.11 |
| 1 | -2.10 | 0.60 | 0.87 | 0.54 | 1.26 | 0.80 | -0.20 | 0.10 | -0.17 | 0.11 | -0.55 | 0.12 |
| 4 | -2.98 | 0.88 | 2.15 | 0.74 | 1.62 | 0.81 | -0.14 | 0.10 | -0.00 | 0.13 | -0.83 | 0.16 |
| 10 | -5.01 | 1.56 | 3.89 | 1.01 | 1.60 | 0.84 | -0.18 | 0.13 | 0.20 | 0.18 | -1.71 | 0.28 |
| 25 | -9.35 | 3.28 | 5.21 | 1.22 | 1.69 | 0.84 | -0.03 | 0.15 | 0.54 | 0.25 | -3.92 | 0.67 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 24: MC results for estimators of $\rho$; $N = 500$, design MNS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 1 | -0.48 | 0.22 | 0.18 | 0.16 | 0.23 | 0.18 | -8.68 | 0.82 | -9.60 | 0.98 | -9.58 | 0.98 |
| 4 | -0.45 | 0.16 | -0.01 | 0.12 | 0.07 | 0.16 | 51.5 | 26.7 | 9.04 | 1.07 | 8.72 | 1.01 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 25: MC results for estimators of $\rho$; $N = 500$, design MNS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| | AB2 | | REQMLE | | FEQMLE | | CSYS2 | | RSYS2 | | ASYS2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_\mu^2$ | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE | bias | MSE |
| 1 | -9.54 | 3.23 | 3.02 | 1.00 | 2.82 | 0.99 | -0.46 | 0.03 | -0.27 | 0.04 | -1.29 | 0.08 |
| 4 | -6.14 | 1.93 | 2.77 | 0.93 | 3.13 | 1.01 | -2.70 | 0.13 | -2.28 | 0.13 | -3.42 | 0.24 |
| 10 | -3.62 | 1.08 | 1.80 | 0.77 | 3.02 | 1.03 | -6.93 | 0.66 | -6.94 | 0.72 | -7.76 | 0.84 |
| 25 | -1.67 | 0.53 | 0.70 | 0.48 | 1.61 | 0.83 | 15.7 | 2.56 | 9.96 | 1.43 | 8.44 | 1.23 |

Notes: 5000 replications; actual bias = bias/100 and actual MSE = MSE/100.

Table 26: MC results for identification tests; $N = 500$, design MS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 136 | 0.00 | 0.00 | 0.00 | 670 | 0.00 | 670 | 0.00 | 0.00 | 0.00 |
| 1 | 102 | 0.00 | 0.00 | 0.00 | 336 | 0.00 | 335 | 0.00 | 0.00 | 0.00 |
| 4 | 58.1 | 0.00 | 0.00 | 0.00 | 134 | 0.00 | 135 | 0.00 | 0.00 | 0.00 |
| 10 | 32.8 | 0.00 | 0.00 | 0.01 | 61.4 | 0.00 | 61.4 | 0.00 | 0.00 | 0.00 |
| 25 | 17.2 | 0.03 | 0.12 | 0.31 | 26.3 | 0.00 | 26.2 | 0.00 | 0.00 | 0.00 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 27: MC results for identification tests; $N = 500$, design MS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 67.6 | 0.00 | 0.00 | 0.00 | 107 | 0.00 | 106 | 0.00 | 0.00 | 0.00 |
| 1 | 59.5 | 0.00 | 0.00 | 0.00 | 89.1 | 0.00 | 89.0 | 0.00 | 0.00 | 0.00 |
| 4 | 44.3 | 0.00 | 0.00 | 0.00 | 59.6 | 0.00 | 59.8 | 0.00 | 0.00 | 0.00 |
| 10 | 30.0 | 0.00 | 0.00 | 0.01 | 36.1 | 0.00 | 36.3 | 0.00 | 0.00 | 0.00 |
| 25 | 17.8 | 0.02 | 0.11 | 0.28 | 18.6 | 0.13 | 18.4 | 0.13 | 0.09 | 0.13 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 28: MC results for identification tests; $N$=500, design MNS, $\rho = 0.5$ & $\sigma_v^2/\sigma^2 = 4/3$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 65.3 | 0.00 | 0.00 | 0.00 | 260 | 0.00 | 329 | 0.00 | 0.00 | 0.00 |
| 4 | 121 | 0.00 | 0.00 | 0.00 | 259 | 0.00 | 22.4 | 0.03 | 0.00 | 0.03 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

Table 29: MC results for identification tests; $N$=500, design MNS, $\rho = 0.9$ & $\sigma_v^2/\sigma^2 = 5\frac{5}{19}$.

| $\sigma_\mu^2$ | $J$ | $p_J$ | $p_J > 0.05$ | $p_J > 0.01$ | $F_D$ | $F_D < 10$ | $F_L$ | $F_L < 10$ | $p_J>0.05$ &$F_D<10$ | $F_D>10$ &$F_L<10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7.65 | 0.24 | 0.71 | 0.89 | 22.1 | 0.06 | 424 | 0.00 | 0.06 | 0.00 |
| 4 | 18.2 | 0.02 | 0.11 | 0.26 | 27.5 | 0.01 | 101 | 0.00 | 0.01 | 0.00 |
| 10 | 35.4 | 0.00 | 0.00 | 0.00 | 44.1 | 0.00 | 13.6 | 0.32 | 0.00 | 0.32 |
| 25 | 65.2 | 0.00 | 0.00 | 0.00 | 87.9 | 0.00 | 7.86 | 0.70 | 0.00 | 0.70 |

Notes: 5000 replications; $J$, $F_D$, $F_L$ tests are described in text; $p_J$ is p-value of $J$ test.

# C SOME ADDITIONAL RESULTS RELATED TO THE APPLICATION.

Table 30: Growth and Financial intermediation proxied by CCBA

| | "AB2" | | | CSYS2 | | | "ASYS2" | |
|---|---|---|---|---|---|---|---|---|
| | estim. | st. err. | $p_{un}$ | estim. | st. err. | $p_{un}$ | estim. | st. err. |
| CCBA | 2.96 | 1.29 | 0.01 | 3.42 | 1.36 | 0.00 | 3.30 | 1.41 |
| LINIT | 0.07 | 0.45 | 0.00 | -0.19 | 0.34 | 0.00 | -0.05 | 0.33 |
| GOV | -3.68 | 1.52 | 0.01 | -3.16 | 1.53 | 0.00 | -3.57 | 1.62 |
| TRADE | 0.03 | 1.68 | 0.01 | -0.38 | 1.70 | 0.03 | -0.11 | 1.91 |
| SEC | 0.88 | 0.54 | 0.03 | 0.96 | 0.62 | 0.03 | 0.88 | 0.60 |
| *CD test* | 22.59 (0.09) | | | 30.49 (0.05) | | | | |
| *J test* | 12.71 (0.55) | | | 18.23 (0.44) | | | 16.52 (0.56) | |
| *J-diff test* | 0.09 (0.77) | | | 5.52 (0.24) | | | 3.81 (0.43) | |

Notes: N=74; number of obs. = 253; time dummies are included; the estimators are defined in section 4; apart from time dummies, the "AB" (SYS) estimators use 19 (23) instruments; Windmeijer robust standard errors are reported; $p_{un}$ is p-value of individual $J$ test of underidentification; the STATA command underid of Schaffer and Windmeijer (2020) was used to perform the underidentification tests; p-values are in parentheses; first $J$-diff test tests $E(Z'_{i,2}w_i) = 0$; second & third $J$-diff tests test $E(Z'_{i,3}w_i) = 0$ excluding $E(Z'_{i,2}w_i) = 0$.

Table 31: Growth and Financial intermediation proxied by LLY

| | "AB2" | | | CSYS2 | | | "ASYS2" | |
|---|---|---|---|---|---|---|---|---|
| | estim. | st. err. | $p_{un}$ | estim. | st. err. | $p_{un}$ | estim. | st. err. |
| LLY | -0.32 | 1.36 | 0.01 | 2.20 | 0.66 | 0.00 | 0.95 | 1.28 |
| LINIT | 0.25 | 0.60 | 0.00 | -0.42 | 0.34 | 0.01 | -0.22 | 0.44 |
| GOV | -2.98 | 2.24 | 0.06 | -1.96 | 1.81 | 0.04 | -2.15 | 1.70 |
| TRADE | 2.56 | 1.83 | 0.06 | 1.02 | 1.02 | 0.08 | 1.77 | 1.41 |
| SEC | 1.31 | 0.52 | 0.18 | 0.82 | 0.58 | 0.06 | 1.12 | 0.56 |
| *CD test* | 19.21 (0.20) | | | 29.67 (0.06) | | | | |
| *J test* | 14.44 (0.42) | | | 21.53 (0.25) | | | 20.39 (0.31) | |
| *J-diff test* | 0.70 (0.40) | | | 7.09 (0.13) | | | 5.95 (0.20) | |

Notes: see table 30.

Table 32: Growth and Financial intermediation proxied by PRICR

| | "AB2" | | | CSYS2 | | | "ASYS2" | |
|---|---|---|---|---|---|---|---|---|
| | estim. | st. err. | $p_{un}$ | estim. | st. err. | $p_{un}$ | estim. | st. err. |
| PRICR | 0.72 | 0.95 | 0.01 | 1.60 | 0.54 | 0.01 | 1.24 | 0.80 |
| LINIT | -0.10 | 0.45 | 0.00 | -0.54 | 0.28 | 0.00 | -0.46 | 0.32 |
| GOV | -2.07 | 2.50 | 0.05 | -0.55 | 1.52 | 0.05 | -0.61 | 1.74 |
| TRADE | 1.92 | 1.99 | 0.03 | 1.27 | 1.21 | 0.05 | 1.28 | 1.35 |
| SEC | 0.92 | 0.54 | 0.03 | 0.84 | 0.61 | 0.06 | 1.07 | 0.62 |
| *CD test* | 23.65 (0.07) | | | 28.48 (0.07) | | | | |
| *J test* | 15.26 (0.36) | | | 19.18 (0.38) | | | 20.47 (0.31) | |
| *J-diff test* | 0.74 (0.39) | | | 3.92 (0.42) | | | 5.21 (0.27) | |

Notes: see table 30.