UNIVERSITÀ di VERONA | Department of ECONOMICS

Working Paper Series
Department of Economics
University of Verona

Testing Initial Conditions in Dynamic Panel Data Models

Giorgio Calzolari, Laura Magazzini

WP Number: 14 August 2013

ISSN: 2036-2919 (paper), 2036-4679 (online)

Testing Initial Conditions in Dynamic Panel Data Models

Laura Magazzini* Giorgio Calzolari†‡

This version: May 31, 2017

Abstract

We propose a new framework for testing the validity of initial conditions in dynamic panel data models. These initial conditions, also called "mean stationarity", are required for the consistency of the system GMM estimator. The empirical literature customarily applies the Sargan/Hansen test of overidentifying restrictions. In our set up the mean stationarity assumption is obtained as a parametric restriction in an extended set of moment conditions, allowing the use of a LM test to check its validity. Our framework provides a ranking in terms of power of the analyzed test statistics for the mean stationarity assumption. The approach we propose exhibits better power than the difference-in-Sargan/Hansen test that compares system GMM and difference GMM. The latter on its turn is more powerful than the Sargan/Hansen test based on the value of the minimized system GMM criterion function.

Keywords: panel data, dynamic model, GMM estimation, initial conditions, test of overidentifying restrictions.
JEL Codes: C23, C12.

*Corresponding author. Department of Economics, Università di Verona, Via Cantarane 24, 37129 Verona, Italy. Phone: +39 045 802 85 25. E-mail: <laura.magazzini@univr.it>

†Department of Statistics, Computer Science, Applications, Università di Firenze, <calzolar@disia.unifi.it>

‡We gratefully acknowledge comments and suggestions from Steve Bond, Massimiliano Cialdi and conference participants to the Fifth and Sixth Italian Congress of Econometrics and Empirical Economics, Genova and Salerno. This work has benefited from financial support of the project MIUR (PRIN): MISURA – Multivariate Models for Risk Assessment.

1 Introduction

Consider the dynamic linear panel data model on N units observed over $T \geq 2$ time periods $(i = 1, ..., N;\ t = 1, ..., T)$:[1]

$$y_{it} = \beta y_{it-1} + \alpha_i + e_{it} = \beta y_{it-1} + \varepsilon_{it} \tag{1}$$

where α_i captures individual 'unobserved heterogeneity', constant over time and different across units, and e_{it} is an idiosyncratic component changing both over time and across units.[2] In the following, we assume that y_{i0} is observed.[3]

Estimation of this model largely relies on the generalized method of moments (GMM; see Hansen, 1982): various estimators have been proposed in the literature, building on a different set of moment conditions (Arellano and Bond, 1991; Arellano and Bover, 1995; Ahn and Schmidt, 1995; Blundell and Bond, 1998). In all cases, it is assumed that α_i and e_{it} are uncorrelated for all i, and e_{it} is uncorrelated over time (including $t = 0$).

Ahn and Schmidt (1995) introduce the following "standard" assumptions that are exploited for the estimation of dynamic panel data models:

(A1) For all i, e_{it} is uncorrelated with y_{i0} for all t

(A2) For all i, e_{it} is uncorrelated with α_i for all t

(A3) For all i, e_{it} are mutually uncorrelated

that will be assumed valid all over this paper. When these assumptions are satisfied, a consistent estimator of β can be obtained by the difference GMM estimator (Arellano and Bond, 1991), usually employed in empirical applications, or by the non linear GMM estimator introduced by Ahn and Schmidt (1995).

Blundell and Bond (1998) make the additional assumption on the initial observation (see also Arellano and Bover, 1995):

(A4) $E(\Delta y_{i1} \varepsilon_{i2}) = 0$,

thus excluding cases in which y_0 is not 'mean stationary'.[4]

When condition (A4) is satisfied, the system GMM estimator proposed by Blundell and Bond (1998) outperforms all other GMM estimation methods in terms of asymptotic efficiency

[1] Our analysis focuses on micro panels where N is typically large and T is typically small. We further assume that $|\beta| < 1$.

[2] The more general specification of the composite error term includes three sources of variation (see, e.g. Baltagi, 2008, p. 33). Besides the individual component α_i and the idiosyncratic error term e_{it}, a time component τ_t is considered, which varies over time and is constant across units. Without loss of generality, we assume $\tau_t = 0$. As the time dimension is typically small, the time effects can be easily controlled for by including time dummies in the regression.

[3] $T + 1$ observations on y are therefore available for estimation.

[4] This 'mean-stationarity' is not an assumption of stationarity in the ordinary sense, but it only excludes correlation between deviations of y_{it} from its long run mean and the individual effect α_i (Roodman, 2009).

as well as small sample bias (especially when β approaches 1), but consistency relies on the validity of the moment restriction (A4). This is customarily tested using the Sargan/Hansen test of overidentifying restrictions, either based on the value of the GMM criterion function of the system GMM estimator, or on the difference between the GMM criterion function of the system GMM estimator and the GMM criterion function of the difference GMM estimator.

In this paper we propose an alternative framework for testing the mean stationarity assumption (A4) that shows better size and power properties than the testing procedures customarily employed in empirical applications.[5] The testing procedure we propose for the 'mean stationarity' assumption (A4) is based on a Lagrange Multiplier (LM) statistics that exploits an alternative parametrization of the estimator proposed by Blundell and Bond (1998). This alternative parametrization leads to the possibility of testing assumption (A4) as a parametric restriction. In the pure dynamic case, the approach is equivalent to a difference-in-Hansen (or difference-in-Sargan) test comparing the system GMM estimator proposed by Blundell and Bond (1998) and the estimator based on the alternative set of moment conditions, that is equivalent to the GMM estimator by Ahn and Schmidt (1995). The latter estimator efficiently exploits all the moment conditions spanning from the standard assumptions (A1)-(A3).

The paper proceeds as follows. The next section briefly describes the GMM estimators that have been proposed in the literature, the moment conditions they use and their relative performance. Section 3 describes the reparametrization of the moment conditions we propose and discuss the test procedure in a pure dynamic setting. The case of an endogenous regressor is considered in Section 4. In Section 5, we report a Monte Carlo experiment, showing the "better" performance of the proposed approach. Section 6 concludes.

2 GMM estimation of linear dynamic panel data models

Estimation of model (1) largely relies on GMM (Hansen, 1982).[6] Two main estimators are employed in the empirical analysis: difference GMM estimator, proposed by Arellano and Bond (1991), and the system GMM estimator (Arellano and Bover, 1995; Blundell and Bond, 1998).

The difference GMM estimator exploits the assumption of lack of serial correlation in e_{it}. First, differences of the data are considered in order to remove the individual effect α_i. Then, the moment conditions are based on the observation that, under the assumption of uncorrelated e_{it}, lag 2 and older of y_{it} can be used as instruments for Δy_{it-1}.[7] That is, the following $T(T-1)/2$

[5] See Newey and West (1987), Newey and MacFadden (1994), and Ruud (2000, ch. 22) for a general discussion of testing procedures in a GMM context.

[6] Other approaches have been considered for estimating linear dynamic panel data models. Hsiao, Pesaran, and Tahmiscioglu (2002) developed a transformed likelihood approach for the estimation of linear dynamic model within a FE framework. Kiviet (1995) proposed a method to correct the small sample bias of the least squares dummy variable estimator. The performance of simulation methods (indirect inference) has also been analysed (Gouriéroux, Phillips, and Yu, 2010). In this paper we focus on the GMM estimator.

[7] An earlier approach was based on the instrumental variable estimator, using only y_{it-2} as an instrument for Δy_{it-1} (Anderson and Hsiao, 1981, 1982).

moment conditions hold ($j, t \geq 2$):

$$E[y_{it-j}\Delta\varepsilon_{it}] = 0 \quad (2)$$

However, simulation studies have shown that the GMM estimator based on the moment conditions in (2) has large finite sample bias and poor performances especially when β approaches 1, or the variance of α_i is large with respect to the variance of e_{it} (Blundell and Bond, 1998). This is referred to as the 'weak instrument' problem of the difference GMM estimator (Stock and Wright, 2000).

Ahn and Schmidt (1995) raise the attention on the additional moment conditions that are available under the assumption of lack of serial correlation in e_{it} (and uncorrelation between e_{it} and α_i), and exploit the fact that the covariance between the initial condition y_{i0} and the composite error term is only driven by the presence of the individual effect α_i and is therefore constant over time. Thus, the following moment conditions can also be considered, leading to a non-linear estimator:

$$E(\Delta\varepsilon_{it}\varepsilon_{iT}) = 0 \quad t = 2, ..., T-1 \quad (3)$$

Ahn and Schmidt (1995) show that, by adding the $T-2$ moment conditions (3) to the moment conditions (2) of the difference GMM estimator proposed by Arellano and Bond (1991), gains in performance are achieved. The estimator makes efficient use of second moment information (Ahn and Schmidt, 1995).

Moment conditions in (2) and (3) exploit the assumption of lack of serial correlation in e_{it}, that can be tested using both the Sargan/Hansen test and the test for autocorrelation of first difference residuals proposed by Arellano and Bond (1991).[8] In Monte Carlo simulation experiments, Arellano and Bond (1991) show that the test for serial correlation of first difference residuals is more powerful than the Sargan/Hansen test.

Blundell and Bond (1998) introduce an additional assumption on the initial condition:

$$E(\Delta y_{i1}\varepsilon_{i2}) = 0 \quad (4)$$

labeled 'mean-stationarity' assumption. As e_{it} is uncorrelated over time, we can also state condition (4) in terms of α_i as

$$E(\Delta y_{i1}\alpha_i) = 0 \quad (5)$$

that makes more clear that differencing y_{i1} eliminates the correlation with α_i, thus excluding correlation between deviations of y_{i1} from its long run mean and the individual effect α_i (Roodman, 2009).

If the condition (4) on the initial observation is also satisfied, the following $T-1$ moment

[8] Indeed, if e_{it} is uncorrelated over time, first order autocorrelation is expected in Δe_{it}. On the contrary the null hypothesis of lack of second order autocorrelation in Δe_{it} should not be rejected.

conditions ($t \geq 2$) also hold:[9]

$$E[\Delta y_{it-1}\varepsilon_{it}] = 0 \tag{6}$$

When these are added to the moment conditions in (2), substantial efficiency gains are achieved, both with respect to the difference GMM estimator and to the estimator proposed by Ahn and Schmidt (1995). Furthermore, when the mean stationarity assumption is satisfied, these moment conditions imply the non-linear moment conditions (3), making these redundant for estimation (Blundell and Bond, 1998).

By looking at the semiparametric information bound, Hahn (1997) shows that when the 'mean-stationarity' assumption (4) is exploited in estimation, substantial efficiency gains arise with respect to the case when it is not. Besides gains in performance, the system GMM estimator that combines (2) and (6) also performs better in terms of small sample bias (Hayakawa, 2007). More recently, Bun and Windmeijer (2010) shows that a weak instrument problem can also arise for the equation in levels (6), confirming however the result of a smaller finite sample bias especially when the series are persistent.

Initial conditions have been also proven to be relevant for the difference GMM estimator, as the estimator may not suffer of the weak instrument problem when condition (4) is not satisfied even when data are persistent (Hayakawa, 2009; Hayakawa and Nagata, 2016).

3 An alternative testing procedure

We propose a different parametrization of the moment conditions customarily employed in GMM estimation of dynamic panel data models, allowing to test the mean stationarity assumption as a parametric restriction via a Lagrange Multiplier (LM) test. The test can be computed on the basis of the system GMM estimator of the coefficient of interest β.

We assume that the conditions for $\sqrt{N}$-consistency and asymptotic normality of the GMM estimator are satisfied.

The system GMM method is widely used in empirical analysis, and the mean stationarity assumption is customarily tested by relying on the Sargan/Hansen test by looking
(i) at the validity of the full set of moment conditions (2) and (6), or
(ii) by computing a difference-in-Sargan/Hansen test that takes into account the difference between the value of the minimized GMM criterion function of the system GMM estimator and of the difference GMM estimator (Bond, Bowsher, and Windmeijer, 2001; Hayakawa and Nagata, 2016).

Let us first consider the latter approach (ii). The test of overidentifying restrictions can be written as an LM test by artificially augmenting the set of moment conditions that characterizes

[9] Of course, longer lags could be considered, but these conditions are redundant when the moment conditions (2) are included for estimation (Arellano and Bover, 1995; Blundell and Bond, 1998).

the difference GMM estimator (2) as in Newey and MacFadden (1994):

$$E[\Delta y_{it-1}\varepsilon_{it}] + \psi_{t-1} = 0 \tag{7}$$

for $t = 2, ..., T$. As we are adding $T-1$ moment conditions and $T-1$ parameters, the unconstrained GMM estimator based on (2) and (7) produces the difference GMM estimator of β, whereas the system GMM estimator would be obtained as a constrained estimator under the null hypothesis

$$H_0^{(1)} : \psi_1 = \psi_2 = ... = \psi_{T-1} = 0. \tag{8}$$

A LM test could be considered for $H_0^{(1)}$ and, in this setting, the LM test would be equivalent to the test that considers the difference between the minimized GMM criterion function of the system GMM and of the difference GMM (Newey and MacFadden, 1994; Newey and West, 1987). This approach would lead no advantage with respect to the difference-in-Sargan/Hansen test customarily adopted in the literature. However, it allows a simpler evaluation of the power properties of the test statistic and comparison with the other test procedures analysed in the paper. Indeed, if we consider local alternative to the parametric null hypothesis of the form $H_1^{(1)} : (1/\sqrt{N})\delta$ for some $\delta \in R^{T-1}$, the test statistic converges in distribution to a noncentral chi-square random variable with $T-1$ degrees of freedom and noncentrality parameter equal to ν_1 that can be calculated as a function of δ, the constraints, and the variance of the unconstrained GMM estimator of $\theta_{(1)} = (\beta, \psi_1, ..., \psi_{T-1})'$ of dimension $T \times 1$ (Ruud, 2000, p. 590-591).

Our proposal is based on the fact that, given the assumption (A1)-(A3) on the data generating process for y_{it}, all the additional parameters $\psi_1, ..., \psi_{T-1}$ in (7) can be written as a function of β and ψ_1 only (Roodman, 2009):

$$\psi_{t-1} = \beta^{t-2}\psi_1$$

that is, we propose to "augment" the moment conditions that produce the difference GMM estimator (2) with the following set of moment conditions:

$$E[\Delta y_{it-1}\varepsilon_{it}] + \beta^{t-2}\psi_1 = 0 \tag{9}$$

Let $\theta_{(2)} = (\beta, \psi_1)'$ and $\mathbf{g}(\theta_{(\mathbf{2})})$ be the vector of moment conditions defined by (2), and (9). The empirical analogue is labeled $\mathbf{g}_N(\theta_{(2)})$.

The unconstrained GMM estimator based on (2) and (9) would now produce, for β, the non linear estimator developed by Ahn and Schmidt (1995),[10] and the system GMM estimator can obtained under the null hypothesis

$$H_0^{(2)} : \psi_1 = 0 \tag{10}$$

The system GMM estimator can therefore be obtained as a restricted GMM estimator, $\hat{\theta}_{RN}$,

[10]Note that, according to (9), we can write: $E(\Delta y_{it-1}\varepsilon_{it}) - \beta E(\Delta y_{it-2}\varepsilon_{it-1}) = 0$. Due to the moment conditions in (2), we have $E(\Delta y_{it-1}\varepsilon_{it}) - \beta E(\Delta y_{it-2}\varepsilon_{it-1}) = E(\Delta\varepsilon_{it-1}\varepsilon_{it}) = 0$, corresponding to the non linear moment conditions by Ahn and Schmidt (1995). See also Appendix A.

by minimizing the following quadratic form:

$$\min_{\boldsymbol{\theta_2}|\psi_1=0} \mathbf{g}_N(\boldsymbol{\theta_2})'\hat{\Omega}^{-1}\mathbf{g}_N(\boldsymbol{\theta_2}) = \min_{\boldsymbol{\theta_2}|\psi_1=0} Q_N(\boldsymbol{\theta_2})$$

with $\hat{\Omega}^{-1}$ the efficient weighting matrix, estimated on the basis of a first step GMM estimation (Hansen, 1982).

Various test statistics can be considered for testing $H_0^{(2)}$. Under the assumptions for $\sqrt{N}$-consistency and asymptotic normality of the GMM estimator, a Wald test, a LM test, and a test based on the difference of the constrained and unconstrained criterion function would be equivalent, at least asymptotically (Ruud, 2000; Newey and MacFadden, 1994; Newey and West, 1987). The first advantage of the LM test is that it only requires computation of the system GMM estimator, whereas the Wald test and the test based on the difference of the criterion function would also require computation of the unrestricted estimator.[11] Indeed, the LM test is based on the value of the gradient for the unconstrained criterion function evaluated at the restricted estimator (Ruud, 2000; Newey and West, 1987). Let $\mathbf{G}_N(\theta)$ be defined as $\partial\mathbf{g}_N(\theta)/\partial\theta'$. The test statistic can be computed as:

$$LM = N\mathbf{g}_N(\hat{\theta}_{RN})'\hat{\Omega}^{-1}\hat{\mathbf{G}}_N\left(\hat{\mathbf{G}}_N'\hat{\Omega}^{-1}\hat{\mathbf{G}}_N\right)^{-1}\hat{\mathbf{G}}_N'\hat{\Omega}^{-1}\mathbf{g}_N(\hat{\theta}_{RN})$$

with $\hat{\mathbf{G}}_N$ and $\hat{\Omega}^{-1}$ evaluated at the restricted estimate (i.e. the system GMM estimator of β and $\psi = 0$). Under conventional identification conditions in the GMM literature, the test has asymptotically a chi-square distribution with 1 degree of freedom (Ruud, 2000; Newey and West, 1987).

In this case, a local alternative to the parametric null hypothesis $H_0^{(2)}$ can be written as $H_1^{(2)}$: $(1/\sqrt{N})\delta$ for some $\delta \in R$. The test statistic converges in distribution to a noncentral chi-square random variable with 1 degrees of freedom and noncentrality parameter equal to ν_2, with ν_2 a function of δ, the constraint, and V_2, the variance of the unconstrained GMM estimator of $\theta_2 = (\beta, \psi_1)'$ of dimension 2×1. Note that, for β, this would be equal to the non linear GMM estimator by Ahn and Schmidt (1995).

In order to compare the two testing procedure, we need to reconcile the different size of the parameter vector under $H_1^{(1)}$ and $H_1^{(2)}$. This can be done by writing the alternative hypothesis in terms of a parameter that represents the deviation from the mean stationarity assumption. When this is done (see Appendix B for details), it is possible to show that the two noncentrality parameters ν_1 and ν_2 are equal and, therefore the power of the test statistic is driven by the number of degrees of freedom. As the power function of the noncentral chi-square distribution is decreasing in the number of degrees of freedom (see, e.g., Ruud, 2000, p. 921), the test of the null hypothesis $H_0^{(2)}$ is expected to be (locally) more powerful that $H_0^{(1)}$.

[11] However, in the Monte Carlo simulation we will also explore the performance of the testing procedure based on the difference between the value of the minimized (optimal) GMM criterion function of the system GMM estimator (Blundell and Bond, 1998) and the (optimal) non-linear estimator proposed by Ahn and Schmidt (1995).

Now consider the test strategy of the mean stationarity assumption based on the Sargan/Hansen test by looking (i) at the validity of the full set of moment conditions (2) and (6), that is the moment conditions that characterize the system GMM estimator. In the pure dynamic framework we are considering, under the null hypothesis of the validity of *all* moment conditions,[12] the test statistic is distributed as a chi-squared with $M-1$ degrees of freedom, with M equal to the number of moment conditions that characterizes the system GMM estimator, i.e. $M = T(T-1)/2 + (T-1)$. As in the previous case, this can be rewritten in terms of an artificial model, in which we partition the moment conditions in two sets: one moment condition $g^{(1)}(\beta) = 0$, and all the other moment conditions labeled $g^{(2)}(\beta) + \psi = 0$ with ψ of dimension $M-1$.[13] The constrained estimator, obtained under the null hypothesis $H_0^{(3)} : \psi = 0$ corresponds to the system GMM estimator, whereas the unconstrained estimator for β would be equivalent to the GMM estimator based on the singleton moment condition $g^{(1)}(\beta) = 0$. Again, by specifying a local alternative, written in terms of the parameter driving deviations from the mean stationarity assumption, we get a noncentral chi-square distribution, whose noncentrality parameter, labeled ν_3 is equal to ν_2 (and therefore also equal to ν_1). Loss in power is expected with respect to the test strategy in (ii) and to the testing procedure we propose due to the increase in the number of degrees of freedom (Ruud, 2000).[14] As a result, we can sort the three test procedures in terms of power, where the test procedure we propose is expected to be more powerful than the difference-in-Sargan/Hansen test, which in turn is expected to be more powerful than the Sargan-Hansen test based on the value of the minimized GMM criterion function.

To the best of our knowledge, the performance of the LM test in a GMM framework has been analysed by Bond and Windmeijer (2005) for testing hypothesis about β. Their Monte Carlo experiments show that the LM test performs well, except in cases where instruments are weak (Bond and Windmeijer, 2005).[15] More recently, Hayakawa and Nagata (2016) have explored the performance of the Sargan/Hansen tests for the mean stationarity assumption.

4 The case of an endogenous regressor

The proposed framework can be extended to the interesting case of a dynamic model with an additional endogenous regressor x:[16]

$$y_{it} = \beta_1 y_{it-1} + \beta_2 x_{it} + \varepsilon_{it} = \beta_1 y_{it-1} + \beta_2 x_{it} + \alpha_i + e_{it} \tag{11}$$

where we still assume $|\beta_1| < 1$, and we allow for correlation between x_{it} and both α_i and e_{is} $(s = 1, ..., t)$.

[12] Conditional on the assumption that one moment condition is actually valid.

[13] In this case, the formula of the test statistic is invariant with respect to the choice of the partition (Newey and MacFadden, 1994). As an example, we can consider $g^{(1)}(\beta) = E(y_{i0}\Delta\varepsilon_{i2}) = 0$.

[14] See Appendix B for details.

[15] See also Newey and Windmeijer (2009).

[16] In this Section we consider the case of a single endogenous regressor. Extension to the case of multiple regressors is straightforward.

Under assumptions (A1)-(A3), the following moment conditions can be included when computing the difference GMM estimator:

$$E(x_{it-j}\Delta\varepsilon_{it}) = 0 \tag{12}$$

for $j \geq 2$.

When the initial conditions are also satisfied, additional moment conditions can be defined, besides moment conditions (6):[17]

$$E[\Delta x_{it-1}\varepsilon_{it}] = 0 \tag{13}$$

spanning from the initial condition $E[\Delta x_{i1}\varepsilon_{i2}] = 0$. If this latter condition is not satisfied, we can write

$$E[\Delta x_{it-1}\varepsilon_{it}] + \xi_{t-1} = 0 \tag{14}$$

for $t = 1, ..., T-1$.[18]

An LM test can therefore be considered to check the validity of the initial moment conditions for both y and x as:

$$H_0 : \begin{cases} \psi = 0 \\ \xi_1 = 0 \\ \vdots \\ \xi_{T-1} = 0 \end{cases} \tag{15}$$

In this case, the approach is not equivalent to the comparison of the minimized value of the GMM criterion function of the system GMM estimator and the non linear estimator by Ahn and Schmidt (1995) as, when x is endogenous, validity of the moment condition (3) also requires $E[\Delta x_{it}\varepsilon_{iT}] = 0$ $(t = 2, ..., T-1)$.

In the next Section Monte Carlo experiments are reported to study the small sample performance of the proposed test procedures against the ones customarily employed in empirical applications.

5 Monte Carlo evidence

In this section an extensive Monte Carlo experiment is performed to explore the size and power properties of the LM test we propose (LM), and the difference-in-Sargan/Hansen test based on the difference between the minimized criterion function of the system GMM estimator and the minimized criterion function of the GMM estimator proposed by Ahn and Schmidt (1995; labeled $BBAS$). The two test procedures will be compared with the procedures customarily employed

[17]See Blundell, Bond and Windmeijer (2001) for an in-depth discussion of dynamic models with an endogenous regressor.

[18]No data generating process is therefore imposed in modeling x. Assumption of the dynamics in x can be easily introduced in the framework, however the test procedure would also requires estimation of the parameters the generate the x. We consider here the more general framework in order to avoid introducing additional assumptions on the model, and the estimation of additional parameters.

in empirical application, that is the Hansen test based on the minimized criterion function of the system GMM estimator (labeled BB), and the difference-in-Hansen test that compares the value of the minimized GMM criterion of the system GMM and of the difference GMM estimators, labeled BBAB. Simulation experiments are run using MatLab.

Following the empirical literature, the test will be computed on the basis of the two-step version of the GMM estimator.[19]

The dynamic panel data model we used is as follows:

$$y_{it} = \beta y_{it-1} + \varepsilon_{it} = \beta y_{it-1} + \alpha_i + e_{it} \tag{16}$$

with $|\beta| < 1$, and $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $e_{it} \sim N(0, \sigma_e^2)$ independent of each others. Heteroschedasticity is introduced in the design by setting the variance of e_{it} at time t equal to $1 + 0.1t$ $(t = 0, ..., T)$.

The initial observation, y_{i0} is generated as:

$$y_{i0} = \frac{\alpha_i}{1-\beta}\gamma + e_{i0} \tag{17}$$

with $e_{i0} = \sum_{j=0}^{\infty} \beta^j e_{i,-j}$, drawn from the covariance stationary distribution. The case $\gamma = 1$ corresponds to the case in which the 'mean stationarity' assumption (4) is satisfied, as well as moment conditions (6). By changing the parameter γ we deviate from a 'mean stationary' process; $\gamma = 0$ produces y_{i0} exogenous (i.e., uncorrelated with α_i).[20]

Experiments are run with $N = 100$, the number of time periods $T = 3, 7$;[21] and $\beta = 0.3, 0.6$. We experiment with $\sigma_\alpha^2 = 1/4, 1$, whereas $\sigma_e^2 = 1$ in all the experiments.[22] 20,000 Monte Carlo replications are run for each experiment with the same seed for the random number generator.[23]

Figure 1 reports the p-value discrepancy plots, and the power of the test procedure is represented in Figure 2 for the case $\sigma_\alpha^2 = 1$. On the x-axis the nominal size is reported (ranging from 1% to 20%); on the y-axis we report the difference between the share of rejections and the nominal size. The 5% confidence intervals computed on the basis of the Kolmogorov-Smirnov

[19]In unreported analysis, we also considered the continuously updated (CU) version (Hansen *et al.*, 1996). Results do not change substantially. In this case, Fortran 77 has been used for computations.

[20] Given the proposed set up, the correlation between the deviation of the initial condition y_{i0} from its long run steady state $\alpha_i/(1-\beta)$ and the level of the steady state itself is equal to

$$\frac{\gamma - 1}{\sqrt{(\gamma-1)^2 + \frac{\sigma_e^2}{\sigma_\alpha^2}\frac{1-\beta}{1+\beta}}}.$$

[21]The number of moment conditions in the system GMM estimator in the case $T = 3, 7$ is respectively 5 and 27.

[22]The choice of the two variance components draws on Bun and Windmeijer (2010). We also experimented with $\beta = 0.9$ and $\sigma_\alpha^2 = 4$. In these cases, however, the 'weak instrument' problem is particularly severe and the conditions for $T^{1/2}$-consistency and asymptotic normality of the GMM estimators are not satisfied (Stock and Wright, 2000). Accordingly, all test procedures analysed in this paper exhibit size distortion and very low power.

[23] As a result, in the assessment of size and power of the tests, a $\pm 0.962\%$ may be ascribed to experimental randomness (Davidson and MacKinnon, 1998). The value is computed by considering the critical value of the Kolmogorov-Smirnov test at the 5% significance level.

statistic are also reported (Davidson and MacKinnon, 1998).

When T is small ($T = 3$), the test procedures have the correct size, with the exception of $BBAS$ with $\beta = 0.6$. On the contrary, with $T = 7$ all test procedures exhibit size distortion, with $BBAS$ and LM showing smaller deviations with respect to BB and $BBAB$.

*** FIGURE 1 ABOUT HERE ***

Power properties of the test statistics for local deviations from the null hypothesis are displayed in Figure 2. Values of γ in the simulations have been chosen in order to have a value of the correlation between the deviation of the initial condition from its long run steady state (i.e., $y_{i0} - \alpha_i/(1-\beta)$) and the level of the steady state itself approximately in the range $[-0.7, 0.7]$ (see footnote 20).

On the x-axis, the value of γ is reported; on the y-axis we show the adjusted share of rejections. The power of the tests is calculated using as critical values the 95th percentiles of the distribution of the test statistics when $\gamma = 1$ (value that makes the assumption on the initial conditions satisfied, thus leading to consistency of the system GMM estimator).

In all the Monte Carlo experiments, the LM and $BBAS$ test that we propose has larger power to detect deviations from the 'mean stationarity' assumption than the tests based on the value of the minimized criterion function BB and $BBAB$; moreover, the advantage is larger with larger T, as expected.

*** FIGURE 2 ABOUT HERE ***

Results of the Monte Carlo experiments in the cases $\sigma^2_\alpha = 1/4$ are reported in Figure 3 and Figure 4.

*** FIGURE 3 ABOUT HERE ***

*** FIGURE 4 ABOUT HERE ***

Results substantially confirm what observed with $\sigma^2_\alpha = 1$. Size distortions emerge with $T = 7$, with the test procedures we propose exhibiting smaller deviations. On the contrary, all test procedures exhibit the correct size when $T = 3$. $BBAB$ generally exhibits the worst performance in terms of size.

When studying the power of the proposed test strategies, LM and $BBAS$ tests are better able to detect local deviations from the mean stationarity assumption, and, as expected, the two test procedures show very similar share of rejections.

Table 1 reports the mean and variance of the estimated values of β using the system GMM estimator in selected Monte Carlo experiments that differ for the values of γ. As expected, with $\gamma = 1$, no bias is detected in the system GMM estimator. On the contrary, when $\gamma \neq 1$ the mean-stationarity assumption required for the consistency of system GMM is not satisfied, and the estimator is biased. As in Hayakawa and Nagata (2016), the system GMM estimator shows smaller bias in the case $\gamma > 1$ than in the case $\gamma < 1$.

*** TABLE 1 ABOUT HERE ***

5.1 Simulation experiments with an endogenous x

We now consider the case of a dynamic model with an additional endogenous regressor x. The model is generated as in (11) with $|\beta_1| < 1$, $\beta_2 = 1$, and $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $e_{it} \sim N(0, \sigma_e^2)$ independent of each others. In these experiments, $\sigma_e^2 = \sigma_\alpha^2 = 1$.[24]

As a result, the variable x_{it} is generated as:

$$x_{it} = \beta_3 x_{it-1} + \lambda\alpha_i + \eta e_{it} + w_{it} \tag{18}$$

with $w_{it} \sim N(0, 0.16)$, $\lambda = 0.25$, and $\eta = -0.1$. The variable x_{it} is correlated both with the individual effect α_i and the idiosyncratic component at time t, e_{it}.

The initial observations, y_{i0} and x_{i0} are generated from the covariance stationary distribution as:

$$y_{i0} = \frac{\alpha_i + \beta_2 x_{i0}}{1 - \beta_1}\gamma + e_{i0}$$

and

$$x_{i0} = \frac{\lambda\alpha_i}{1 - \beta_3}\mu + w_{i0}$$

Deviations from the initial conditions for the validity of the 'mean stationarity' assumption arise when $\gamma \neq 1$ and $\mu \neq 1$. By changing the parameter γ and μ we deviate from a 'mean stationary' process respectively for y and x (or both).

In the experiments, we consider $\beta_1 = \beta_3 = 0.5$, $N = 100$, and $T = 3, 7$.

Results of the Monte Carlo simulations are reported in Tables 2, in which we consider different values for γ and μ. We report the mean and variance of estimated β_1 and β_2 using the system GMM estimator, as well as the share of rejections at the 5% level of significance for the test procedures analysed in this paper.

*** TABLE 2 ABOUT HERE ***

The system GMM estimator is biased in the cases in which $\gamma \neq 1$ or $\eta \neq 1$, thus when either y or x deviates from the mean stationarity assumption.

The relative performance of the $BBAB$ and LM test depends on the proposed set up. With $T = 3$, the two test statistics have similar behavior with LM showing better performance in terms of power when deviations from mean stationarity arise in x only. For longer panel data sets ($T = 7$), marked differences in power properties emerge; however, LM has larger size distortion than $BBAB$. Both test procedures outperform the BB test statistics.[25]

[24] The data generating process is based on the Monte Carlo experiments reported in Blundell, Bond and Windmeijer (2001).

[25] As in the pure dynamic case, all Monte Carlo replications are run using the same seed for the random number generator. As a result, in the assessment of size and power of the tests, a $\pm 0.962\%$ may be ascribed to experimental randomness (Davidson and MacKinnon, 1998). See footnote 23.

6 Conclusion

In this paper we have proposed an alternative procedure for testing the 'mean stationarity' assumption, that is needed for consistency of the system GMM estimator (Blundell and Bond, 1998; the most efficient linear GMM estimator when the assumption is satisfied).

The proposed approach exhibits better properties in terms of size and power with respect to the test based on the minimized value of the GMM criterion (Sargan/Hansen test statistic), customarily employed in empirical applications, especially when a pure dynamic model is considered. In this set up, the better properties are explained in terms of the different degrees of freedom that underly the analyzed test strategies. When an endogenous variable is included in the Monte Carlo experiments, relative performances depend on the characteristics of the data generating process, with the LM test procedure we propose exhibiting better/similar properties to $BBAB$ procedure. Both approaches outperform BB, the test procedure based on the value of the minimized GMM criterion function.

All in all, the test can better drive empirical researchers in the choice between the difference GMM and the system GMM estimators of dynamic panel data models.

A Appendix: The mean stationarity assumption as a parametric restriction

We follow Ahn and Schmidt (1995) and distinguish two sets of variables, i.e. those variables we make assumptions about, and those variables we make assumptions about and that are observable in the sense that can be written in terms of data and parameters. Referring to model (1), the assumptions of the difference GMM and system GMM estimators consider the vector $(e_{i1}, e_{i2}, ..., e_{iT}, y_{i0}, \alpha_i)'$ and we let:

$$\Sigma = cov \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \\ y_{i0} \\ \alpha_i \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1T} & \sigma_{10} & \sigma_{1\alpha} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2T} & \sigma_{20} & \sigma_{2\alpha} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \dots & \sigma_{TT} & \sigma_{T0} & \sigma_{T\alpha} \\ \sigma_{01} & \sigma_{02} & \dots & \sigma_{0T} & \sigma_{00} & \sigma_{0\alpha} \\ \sigma_{\alpha 1} & \sigma_{\alpha 2} & \dots & \sigma_{\alpha T} & \sigma_{\alpha 0} & \sigma_{\alpha\alpha} \end{pmatrix} \quad (19)$$

in which, of course, $\sigma_{ij} = \sigma_{ji}$ $(i, j = 0, 1, ..., T, \alpha)$.

The variance-covariance matrix of the vector of "observable" variables $(\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT}, y_{i0})$ is:

$$\Lambda = cov\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \\ y_{i0} \end{pmatrix} = cov\begin{pmatrix} \alpha_i + e_{i1} \\ \alpha_i + e_{i2} \\ \vdots \\ \alpha_i + e_{iT} \\ y_{i0} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1T} & \lambda_{10} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2T} & \lambda_{20} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{T1} & \lambda_{T2} & \dots & \lambda_{TT} & \lambda_{T0} \\ \lambda_{01} & \lambda_{02} & \dots & \lambda_{0T} & \lambda_{00} \end{pmatrix} \tag{20}$$

with $\lambda_{st} = \lambda_{ts} = \sigma_{\alpha\alpha} + \sigma_{st} + \sigma_{\alpha s} + \sigma_{\alpha t}$ for each $s, t = 1, ..., T$, $\lambda_{0t} = \lambda_{t0} = \sigma_{\alpha\alpha} + \sigma_{0t}$, and $\lambda_{00} = \sigma_{00}$.

The following constraints are customarily imposed on Λ:

- α_i and e_{it} are uncorrelated, so that $\sigma_{\alpha t} = 0$ for each $t = 1, ..., T$;
- e_{it} is uncorrelated over time, so that $\sigma_{st} = 0$ for each $s, t = 1, ..., T$ with $s \neq t$.

Moreover, as pointed out by Ahn and Schmidt (1995),

- $\sigma_{0t} = 0$ for each $t = 1, ..., T$, spanning from the fact that the covariance between the initial observation y_{i0} and ε_{it} is only driven by the presence of the individual effect α_i (therefore, $\lambda_{0t} = \sigma_{\alpha\alpha}$, that is $\sigma_{0t} = 0$).

As a results of these constraints, Λ simplifies to (Ahn and Schmidt, 1995):

$$\Lambda = \begin{pmatrix} (\sigma_{\alpha\alpha} + \sigma_{11}) & \sigma_{\alpha\alpha} & \dots & \sigma_{\alpha\alpha} & \sigma_{0\alpha} \\ \sigma_{\alpha\alpha} & (\sigma_{\alpha\alpha} + \sigma_{22}) & \dots & \sigma_{\alpha\alpha} & \sigma_{0\alpha} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{\alpha\alpha} & \sigma_{\alpha\alpha} & \dots & (\sigma_{\alpha\alpha} + \sigma_{TT}) & \sigma_{0\alpha} \\ \sigma_{0\alpha} & \sigma_{0\alpha} & \dots & \sigma_{0\alpha} & \sigma_{00} \end{pmatrix} \tag{21}$$

The non-linear estimator by Ahn and Schmidt (1995) therefore include additional moment conditions (with respect to the difference GMM estimator) to exploit the assumption that λ_{ts} ($s, t = 1, ..., T$ with $s \neq t$) is constant over time. In particular the set of moment conditions comprise

$$E(\Delta\varepsilon_{it}\varepsilon_{iT}) = 0$$

with $t = 2, ..., T - 1$.

The system GMM estimator exploits the same assumption on λ_{ts} ($s, t = 1, ..., T$ with $s \neq t$) writing the set of moment conditions in the form

$$E(\Delta\varepsilon_{it-1}\varepsilon_{it}) = 0$$

with $t = 3, ..., T$, or

$$E(\Delta\varepsilon_{it}\varepsilon_{it+1}) = 0$$

with $t = 2, ..., T-1$. However, the system GMM estimator also adds one additional moment condition to Λ, namely the mean stationarity assumption, that can also be written as (Arellano and Bover, 1995; Blundell and Bond, 1998):

$$\sigma_{0\alpha} = \frac{\sigma_{\alpha\alpha}}{1-\beta} \tag{22}$$

As a result, the mean stationarity assumption can also be tested by using a difference-in-Sargan/Hansen test that compare the non linear GMM estimator proposed by Ahn and Schmidt (1995) and the estimator proposed by Blundell and Bond (1998).

Using an alternative approach, consider the case $T = 3$. The moment conditions of the difference GMM estimator (Arellano and Bond, 1991) can be written as:

(c1) $E(y_{i0}\Delta\varepsilon_{i2}) = 0$ or $E(y_{i0}\varepsilon_{i2}) = E(y_{i0}\varepsilon_{i1})$;

(c2) $E(y_{i0}\Delta\varepsilon_{i3}) = 0$ or $E(y_{i0}\varepsilon_{i3}) = E(y_{i0}\varepsilon_{i2})$;

(c3) $E(y_{i1}\Delta\varepsilon_{i3}) = 0$ or $E(y_{i1}\varepsilon_{i3}) = E(y_{i1}\varepsilon_{i2})$;

The condition of the non linear estimator proposed by Ahn and Schmidt (1995) would be:

(c4) $E(\Delta\varepsilon_{i2}\varepsilon_{i3}) = 0$;

whereas the system GMM estimator would also consider:

(c5) $E(\Delta y_{i2}\varepsilon_{i3}) = 0$ or $E(y_{i2}\varepsilon_{i3}) = E(y_{i1}\varepsilon_{i3})$;

(c6) $E(\Delta y_{i1}\varepsilon_{i2}) = 0$ or $E(y_{i1}\varepsilon_{i2}) = E(y_{i0}\varepsilon_{i2})$.

We now show that, when conditions (c1)-(c3) are satisfied, conditions (c4) and (c6) are equivalent to conditions (c5) and (c6), making conditions (c1)-(c4) and (c6) equivalent to the moment condition of the system GMM estimator. As a result, the mean stationarity assumption (c6) can be tested by the difference-in-Sargan/Hansen test that we propose, comparing the non linear GMM estimator proposed by Ahn and Schmidt (1995) and the estimator proposed by Blundell and Bond (1998).

Start from condition (c4):

$$\begin{aligned} 0 &= E(\Delta\varepsilon_{i2}\varepsilon_{i3}) \\ &= E((\Delta y_{i2} - \beta\Delta y_{i1})\varepsilon_{i3}) \\ &= E(\Delta y_{i2})\varepsilon_{i3} - \beta E((y_{i1} - y_{i0})\varepsilon_{i3}) \end{aligned}$$

By using conditions (c2), (c3), and (c6) we can show that $E((y_{i1} - y_{i0})\varepsilon_{i3}) = 0$. Indeed, (i) for (c3) $E(y_{i1}\varepsilon_{i3}) = E(y_{i1}\varepsilon_{i2})$, and by using (c6), $E(y_{i1}\varepsilon_{i2}) = E(y_{i0}\varepsilon_{i2})$; (ii) using condition (c2) we also get $E(y_{i0}\varepsilon_{i3}) = E(y_{i0}\varepsilon_{i2})$. As a result $E((y_{i1} - y_{i0})\varepsilon_{i3}) = 0$. Therefore we get $0 = E(\Delta y_{i2}\varepsilon_{i3})$, condition (c5).

B Comparing the power of the different approaches

In this Appendix we explore the power of local deviations from the null hypothesis of the three test procedures explored in this paper:

(1) The test procedure based on the difference between the value of the minimized GMM function of the system GMM estimator and of the difference GMM estimator, labeled $BBAB$;

(2) The test procedure based on the LM statistic, that is equivalent to the difference between the value of the minimized GMM function of the system GMM estimator and of the non linear GMM estimator, labeled $LM/BBAS$;

(3) The test procedure based on the value of the minimized GMM function of the system GMM estimator, labeled BB.

Let us first introduce general notation. Let θ be the parameter vector to be estimated, and we will refer to the vector of moment conditions as $\mathbf{g}(\theta)$. The empirical analogue is labeled $\mathbf{g}_N(\theta)$. The GMM estimator is obtained as a solution to the following minimization problem:

$$\min_{\theta} \mathbf{g}_N(\theta)'\hat{\Omega}^{-1}\mathbf{g}_N(\theta) = \min_{\theta} Q_N(\theta)$$

We assume that the conditions for $\sqrt{N}$-consistency and asymptotic normality of the GMM estimator are satisfied, and that $\hat{\Omega}^{-1}$ is the efficient weighting matrix (Hansen, 1982). This can be estimated on the basis of a first step GMM estimation $\hat{\theta}_1$:

$$\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_N(\hat{\theta}_1)\mathbf{g}_N(\hat{\theta}_1)'$$

with probability limit equal to $\Omega = E(\mathbf{g}(\theta)\mathbf{g}(\theta)')$. Under conventional identification conditions in the GMM literature, the GMM estimator θ converges in distribution to a $N(\theta, V_0)$ with $V_0 = (G'\Omega G)^{-1}$ with $\mathbf{G}(\theta) = \partial\mathbf{g}(\theta)/\partial\theta'$.

As we have stated in the text, the three test procedures can be obtained as a LM test of an artificially augmented model. Let $MD = T(T-1)/2$, the number of moment conditions that characterizes the difference GMM estimator, and $ML = T-1$ the number of the moment conditions added on the level equations to obtained the system GMM estimator. We let $M = MD + ML = T(T-1)/2 + (T-1)$.

The test $BBAB$ can be obtained by verifying the null hypothesis $H_0^{(1)} : \psi_1 = \psi_2 = ... = \psi_{T-1} = 0$ for the moment conditions:[26]

[26] We order the moment conditions so that the first MD corresponds to the moment conditions (2) that characterize the difference GMM estimator, and the following ML moment conditions correspond to moment conditions on the level equation as in (6).

$$
\begin{aligned}
&g_1(\theta) = E(y_{i0}\Delta\varepsilon_{i2}) = 0 \\
&g_2(\theta) = E(y_{i0}\Delta\varepsilon_{i3}) = 0 \\
&\vdots \\
&g_{MD}(\theta) = E(y_{iT-2}\Delta\varepsilon_{iT}) = 0 \\
&g_{MD+1}(\theta) = E(\Delta y_{i1}\varepsilon_{i2}) + \psi_1 = 0 \\
&\vdots \\
&g_M(\theta) = E(\Delta y_{iT-1}\varepsilon_{iT}) + \psi_{T-1} = 0
\end{aligned} \tag{23}
$$

In this case, $\theta = (\beta, \psi_1, ..., \psi_{T-1})'$.

The test procedure $LM/BBAS$ consider the null hypothesis $H_0^{(2)} : \psi_1 = 0$ in the set of moment conditions:

$$
\begin{aligned}
&g_1(\theta) = E(y_{i0}\Delta\varepsilon_{i2}) = 0 \\
&g_2(\theta) = E(y_{i0}\Delta\varepsilon_{i3}) = 0 \\
&\vdots \\
&g_{MD}(\theta) = E(y_{iT-2}\Delta\varepsilon_{iT}) = 0 \\
&g_{MD+1}(\theta) = E(\Delta y_{i1}\varepsilon_{i2}) + \psi_1 = 0 \\
&\vdots \\
&g_M(\theta) = E(\Delta y_{iT-1}\varepsilon_{iT}) + \beta^{T-2}\psi_1 = 0
\end{aligned} \tag{24}
$$

In this case, $\theta = (\beta, \psi_1)'$.

Finally, the test procedure BB would consider the set of moment conditions:

$$
\begin{aligned}
&g_1(\theta) = E(y_{i0}\Delta\varepsilon_{i2}) = 0 \\
&g_2(\theta) = E(y_{i0}\Delta\varepsilon_{i3}) + \psi_1 = 0 \\
&\vdots \\
&g_{MD}(\theta) = E(y_{iT-2}\Delta\varepsilon_{iT}) + \psi_{MD-1} = 0 \\
&g_{MD+1}(\theta) = E(\Delta y_{i1}\varepsilon_{i2}) + \psi_{MD} = 0 \\
&\vdots \\
&g_M(\theta) = E(\Delta y_{iT-1}\varepsilon_{iT}) + \psi_{M-1} = 0
\end{aligned} \tag{25}
$$

The test statistics is equivalent to testing $H_0^{(3)} : \psi_1 = ... = \psi_{T-1} = \psi_T = ... = \psi_{M-1} = 0$. In this case, $\theta = (\beta, \psi_1, ..., \psi_{M-1})'$.

Note that, asymptotically, the inverse of the efficient weighting matrix $\Omega = E(\mathbf{g}(\theta)\mathbf{g}(\theta)')$ of the three GMM estimator is unchanged, so the variance covariance matrix of the estimator, labeled, respectively $V_{(1)}$, $V_{(2)}$, and $V_{(3)}$ differ due to the differences in the matrix G and the number of coefficients to be estimated (the size of θ differs in the three cases). In particular, let $G_\beta = \partial\mathbf{g}(\theta)/\partial\beta$, $\mathbf{0}_{[a\times b]}$ be a matrix of zeros of dimension $a \times b$, and I_a denote the identity matrix of dimension $a \times a$. Finally, let $B_\beta = (1\ \beta\ \beta^2\ ...\ \beta^{T-2})'$. The three matrices can be written as:

$$G_{(1)} = \left[G_\beta \left| \begin{pmatrix} \mathbf{0}_{[MD\times(T-1)]} \\ I_{T-1} \end{pmatrix} \right. \right]$$

$$G_{(2)} = \left[G_\beta \left| \begin{pmatrix} \mathbf{0}_{[MD\times 1]} \\ B_\beta \end{pmatrix} \right. \right]$$

$$G_{(3)} = \left[G_\beta \left| \begin{pmatrix} \mathbf{0}_{[1\times(M-1)]} \\ I_{M-1} \end{pmatrix} \right. \right]$$

Let us denote the (asymptotic) variance covariance matrix of the GMM estimators as $V_{(i)} = (G_i'\Omega G_i)^{-1}$, $i = 1, ..., 3$.

Consider local alternative to the parametric null hypothesis of the form $H_1 : (1/\sqrt{N})\delta$ for some δ in the appropriate space ($\delta \in R^{T-1}$ when considering $BBAB$; $\delta \in R$ for $LM/BBAS$; and $\delta \in R^{M-1}$ when considering BB). The test statistic converges in distribution to a noncentral chi-square random variable with a number of degrees of freedom equal to the size of δ and noncentrality parameter equal to $\nu = \delta'(RVR')^{-1}\delta$, with V the variance covariance matrix of the unconstrained GMM estimator of θ, R the first derivative of the constraints with respect to the parameters (Ruud, 2000, p. 590-591).

When comparing the power of the three test procedures, we therefore need to reconcile the different size of δ under the alternative hypothesis. This can be done by writing δ as a function of λ in (17). When $\lambda \neq 1$ in (17), the mean stationarity assumption is not satisfied, and the three test procedures can be compared in terms of the same deviation from the mean stationarity assumption.[27]

In particular, $E(\Delta y_{i1}\varepsilon_{i2}) = (1-\lambda)\sigma_\alpha^2$, so that we can let $\delta = (1-\lambda)\sigma_\alpha^2$ when considering (local) alternative hypothesis for $LM/BBAS$. In the case of $BBAB$, $E(\Delta y_{it-1}\varepsilon_{it}) = (1-\lambda)\sigma_\alpha^2\beta^{t-2}$ for $t = 2, ..., T$, so we let $\delta = (1-\lambda)\sigma_\alpha^2 B_\beta'$. Finally, for BB, changes in λ only affects the moment conditions based on the level equations, so we can let $\delta = (1-\lambda)\sigma_\alpha^2 \left(\mathbf{0}_{[1\times(MD-1)]}\ |B_\beta\right)'$. The three vectors δ defined in this way would allow us to compare the power of the three testing procedures against the same deviation from the mean stationarity assumption.

We can now compute the non centrality parameter of the three test procedures. The noncentrality parameter of the test $BBAB$ is:

$$\nu_1 = (1-\lambda)^2(\sigma_\alpha^2)^2 B_\beta' \left[\left(\mathbf{0}_{[(T-1)\times 1]} \ \middle|\ I_{(T-1)} \right) V_{(1)} \left(\frac{\mathbf{0}_{1\times(T-1)}}{I_{(T-1)}} \right) \right]^{-1} B_\beta \tag{26}$$

[27]This is actually what we did when running the Monte Carlo simulations in Section 5.

For the $LM/BBAS$ test, the noncentrality parameter is equal to:

$$\nu_2 = (1-\lambda)^2 (\sigma_\alpha^2)^2 \left[\begin{pmatrix} 0 & 1 \end{pmatrix} V_{(2)} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]^{-1} \tag{27}$$

Finally, the noncentrality parameter of BB is:

$$\nu_3 = (1-\lambda)^2 (\sigma_\alpha^2)^2 C_\beta' \left[\begin{pmatrix} \mathbf{0}_{(M-1)\times 1} & \big| & I_{M-1} \end{pmatrix} V_{(3)} \begin{pmatrix} \mathbf{0}_{1\times(M-1)} \\ \hline I_{M-1} \end{pmatrix} \right]^{-1} C_\beta \tag{28}$$

with $C_\beta = (\mathbf{0}_{[1\times MD-1]} \,|\, B_\beta)$.

By performing the computations for each $V_{(i)} = (G_i'\Omega G_i)^{-1}$, $i = 1, ..., 3$, it is possible to show that:

$$V_{(2)}^{-1} = B' V_{(1)}^{-1} B \tag{29}$$

By using this result it is possible to show that $\nu_1 = \nu_2$.

Let us first compute ν_2, by replacing $V_{(2)}$ in (27) with $B'V_{(1)}^{-1}B$ as in (29). Consider a partition of $V_{(1)}$ as

$$V_{(1)} = \begin{pmatrix} V_{\beta\beta} & V_{\beta\psi} \\ V_{\beta\psi}' & V_{\psi\psi} \end{pmatrix}$$

with $V_{\beta\beta}$ a scalar, corresponding to the variance of the difference GMM estimator of β, $V_{\beta\psi}$ of dimension $1 \times (T-1)$, and $V_{\psi\psi}$ the variance covariance matrix of the unrestricted GMM estimate of the parameters $\psi_1, ..., \psi_{T-1}$ from moment conditions (24). Using the formula of the inverse of a partitioned matrix, we can write:

$$V_{(1)}^{-1} = \begin{pmatrix} V_{\beta\beta}^{-1}(1 + V_{\beta\psi}(V_{\psi\psi} - V_{\beta\psi}'V_{\beta\beta}^{-1}V_{\beta\psi})^{-1}V_{\beta\psi}'V_{\beta\beta}^{-1}) & -V_{\beta\beta}^{-1}V_{\beta\psi}(V_{\psi\psi} - V_{\beta\psi}'V_{\beta\beta}^{-1}V_{\beta\psi}) \\ -V_{\beta\beta}^{-1}V_{\beta\psi}(V_{\psi\psi} - V_{\beta\psi}'V_{\beta\beta}^{-1}V_{\beta\psi}) & (V_{\psi\psi} - V_{\beta\psi}'V_{\beta\beta}^{-1}V_{\beta\psi})^{-1} \end{pmatrix}$$

or

$$V_{(1)}^{-1} = \begin{pmatrix} (V_{\beta\beta} - V_{\beta\psi}V_{\psi\psi}^{-1}V_{\beta\psi}')^{-1} & -(V_{\beta\beta} - V_{\beta\psi}V_{\psi\psi}^{-1}V_{\beta\psi}')^{-1}V_{\beta\psi}V_{\psi\psi}^{-1} \\ -(V_{\beta\beta} - V_{\beta\psi}V_{\psi\psi}^{-1}V_{\beta\psi}')^{-1}V_{\beta\psi}V_{\psi\psi}^{-1} & V_{\psi\psi}^{-1}(I + V_{\beta\psi}'(V_{\beta\beta} - V_{\beta\psi}V_{\psi\psi}^{-1}V_{\beta\psi}')^{-1}V_{\beta\psi}V_{\psi\psi}^{-1} \end{pmatrix}$$

We will denote:

$$V_{(1)}^{-1} = \begin{pmatrix} W_{11} & W_{12} \\ W_{12}' & W_{22} \end{pmatrix}$$

Therefore we have

$$BV_{(1)}^{-1}B' = \begin{pmatrix} W_{11} & W_{12}B_\beta' \\ B_\beta W_{12}' & B_\beta W_{22} B_\beta' \end{pmatrix} = \begin{pmatrix} W_{11} & X_{12} \\ X_{12}' & X_{22} \end{pmatrix}$$

so that

$$(BV_{(1)}^{-1}B')^{-1} = \begin{pmatrix} (W_{11} - X_{12}X_{22}^{-1}X_{12}')^{-1} & -W_{11}^{-1}X_{12}(X_{22} - X_{12}'W_{11}^{-1}X_{12})^{-1} \\ -(X_{22} - X_{12}'W_{11}^{-1}X_{12})^{-1}X_{12}'W_{11}^{-1} & (X_{22} - X_{12}'W_{11}^{-1}X_{12})^{-1} \end{pmatrix}$$

As a result, ν_2 is proportional to

$$\begin{aligned} \begin{pmatrix} 0 & 1 \end{pmatrix} (BV_{(1)}^{-1}B')^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= (X_{22} - X_{12}'W_{11}^{-1}X_{12})^{-1} \\ &= B_\beta W_{22} B_\beta' - W_{12}B_\beta' W_{11}^{-1} B_\beta W_{12}' \\ &\quad \text{as } B_\beta W_{12} \text{ is a scalar} \\ &= B_\beta W_{22} B_\beta' - B_\beta W_{12}' W_{11}^{-1} W_{12} B_\beta' \\ &= B_\beta V_{\psi\psi} B_\beta' \\ &\quad + B_\beta V_{\psi\psi}^{-1} V_{\beta\psi}' (V_{\beta\beta} - V_{\beta\psi}V_{\psi\psi}^{-1}V_{\beta\psi}')^{-1} V_{\beta\psi} V_{\psi\psi}^{-1} B_\beta' \\ &\quad - B_\beta V_{\psi\psi}^{-1} V_{\beta\psi}' (V_{\beta\beta} - V_{\beta\psi}V_{\psi\psi}^{-1}V_{\beta\psi}')^{-1} (V_{\beta\beta} - V_{\beta\psi}V_{\psi\psi}^{-1}V_{\beta\psi}') \\ &\qquad (V_{\beta\beta} - V_{\beta\psi}V_{\psi\psi}^{-1}V_{\beta\psi}')^{-1} V_{\beta\psi} V_{\psi\psi}^{-1} B_\beta' \\ &= B_\beta V_{\psi\psi} B_\beta' \end{aligned}$$

so that $\nu_2 = (1-\lambda)^2(\sigma_\alpha^2)^2 B_\beta V_{\psi\psi} B_\beta'$.

In order to compute ν_1, let us consider:

$$B_\beta' \left[\begin{pmatrix} \mathbf{0}_{[T-1\times 1]} & \big| & I_{T-1} \end{pmatrix} V_{(1)} \begin{pmatrix} \mathbf{0}_{1\times T-1} \\ \hline I_{T-1} \end{pmatrix} \right]^{-1} B_\beta = B_\beta' \left[\begin{pmatrix} \mathbf{0}_{[T-1\times 1]} & \big| & I_{T-1} \end{pmatrix} \begin{pmatrix} V_{\beta\beta} & V_{\beta\psi} \\ V_{\beta\psi}' & V_{\psi\psi} \end{pmatrix} \begin{pmatrix} \mathbf{0}_{1\times T-1} \\ \hline I_{T-1} \end{pmatrix} \right]^{-1} B_\beta$$

which is also equal to $B_\beta V_{\psi\psi}^{-1} B_\beta'$, proving that $\nu_1 = \nu_2$.

Analogous algebra can be employed to show that $\nu_2 = \nu_3$, by relying on the equality

$$V_{(2)}^{-1} = (C'V_{(3)}^{-1}C)^{-1} \tag{30}$$

with

$$C = \begin{pmatrix} 1 & \mathbf{0}_{1\times(MD-1)} & \mathbf{0}_{1\times(T-1)} \\ 0 & \mathbf{0}_{1\times(MD-1)} & B_\beta \end{pmatrix}$$

As the three noncentrality parameters are the same, the power of the test for local deviations from the mean stationarity assumption is only affected by the number of degrees of freedom, which is 1 in the case of $LM/BBAS$, $T-1$ for $BBAB$ and $M-1$ for BB (with $M = T(T-1)/2 + T - 1$). Differences in power are therefore expected to be larger for larger T.

References

Ahn, S.C. and Schmidt, P.: 1995, Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics* **68**(1), 5–27. DOI: 10.1016/0304-4076(94)01641-C

Anderson, T. and Hsiao, C.: 1981, Estimation of Dynamic Models with Error Components, *Journal of the American Statistical Association* **76**, 598–606. DOI: 10.1080/01621459.1981.10477691

Anderson, T. and Hsiao, C.: 1982, Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics* **18**, 47–82. DOI: 10.1016/0304-4076(82)90095-1

Arellano, M. and Bond, S.: 1991, Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *The Review of Economic Studies* **58**(2), 277–297. DOI: 10.2307/2297968

Arellano, M. and Bover, O.: 1995, Another Look at the Instrumental Variables Estimation of Error Components Models, *Journal of Econometrics* **68**, 29-51. DOI: 10.1016/0304-4076(94)01642-D

Baltagi, B.: 2008, *Econometric analysis of panel data*, John Wiley & Sons.

Blundell, S. and Bond, S.: 1998, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics* **87**(1), 115–143. DOI: 10.1016/S0304-4076(98)00009-8

Blundell, R., Bond, S. and Windmeijer, F.: 2001, Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator, in *Nonstationary Panels, Panel Cointegration, and Dynamic Panels* 15: 53–91.

Bond, S.R., Bowsher, C. and Windmeijer, F.: 2001, Criterion-Based Inference for GMM in Autoregressive Panel Data Models, *Economics Letters* **73**, 379–388. DOI: 10.1016/S0165-1765(01)00507-9

Bond, S. and Windmeijer, F.: 2005, Reliable Inference for GMM Estimators? Finite Sample Properties of Alternative Test Procedures in Linear Panel Data Models, *Econometric Reviews* **24**(1), 1–37. DOI: 10.1081/ETC-200049126

Bun, M.J.C. and Windmeijer, F.: 2010, The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models, *Econometrics Journal* **13**(1), 95–126. DOI: 10.1111/j.1368-423X.2009.00299.x

Davidson, R. and MacKinnon, J.G.: 1998, Graphic Methods For Investigating the Size and Power of Hypothesis Tests, *The Manchester School* **66**(1), 1–26. DOI: 10.1111/1467-9957.00086

Gouriéroux, C., Phillips, P.C.B. and Yu, J.: 2010, Inderect Inference for Dynamic Panel Models, *Journal of Econometrics* **157**(1), 68–77. DOI: 10.1016/j.jeconom.2009.10.024

Griliches, Z.: 1990, Patent Statistics as Economic Indicators: A Survey, *Journal of Economic Literature*, XXVIII, 1661–1707.

Guggenberger, P.: 2010, The Impact of a Hausman Pretest on the Asymptotic Size of a Hypothesis Test, *Econometric Theory* **26** 369–382. DOI: 10.1017/S0266466609100026

Hayakawa, K.: 2007, Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models, *Economics Letters* **95**, 32–38. DOI: 10.1016/j.econlet.2006.09.011

Hayakawa, K.: 2009, On the Effect of Mean-Nonstationarity in Dynamic Panel Data Models, *Journal of Econometrics* **153**(2), 133–135. DOI: 10.1016/j.jeconom.2009.04.008

Hayakawaa, K. and Nagata, S.: 2016, On the Behaviour of the GMM Estimator in Persistent Dynamic Panel Data Models with Unrestricted Initial Conditions, *Computational Statistics and Data Analysis* **100**, 262-303. DOI: 10.1016/j.csda.2015.03.007

Hahn, J.: 1999, How Informative is the Initial Condition in the Dynamic Panel Model with Fixed Effects?, *Journal of Econometrics* **93**, 309–326. DOI: 10.1016/S0304-4076(99)00013-5

Hansen, L.: 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* **50**, 1029-1054. DOI: 10.2307/1912775

Hansen, L.P., Heaton, J., and Yaron, A.: 1996, Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business & Economic Statistics* **14**(3), 262–280. DOI: 10.1080/07350015.1996.10524656

Hsiao, C., Pesaran, M.H. and Tahmiscioglu A.K.: 2002, Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, *Journal of Econometrics* **109**(1), 107–150. DOI: 10.1016/S0304-4076(01)00143-9

Kiviet, J.F.: 1995, On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models, *Journal of Econometrics* **68**(1), 53–78. DOI: 10.1016/0304-4076(94)01643-E

Kleibergen, F.: 2005, Testing Parameters in GMM without Assuming That They are Identified, *Econometrica* **73**, 1103–1123. DOI: 10.1111/j.1468-0262.2005.00610.x

Newey, W.K. and MacFadden, D.: 1994, Large sample estimation and hypothesis testing, in *Handbook of econometrics* 4, 2111–2245.

Newey, W.K. and West, K.D.: 1987, Hypothesis Testing with Efficient Method of Moment Estimation, *International Economic Review* **28**, 777–787. DOI: 10.2307/2526578

Newey, W.K. and F. Windmeijer: 2009, Generalized Method of Moments with Many Weak Moment Conditions, *Econometrica* **77** 687–719. DOI: 10.3982/ECTA6224

Roodman, D.: 2009, A Note on the Theme of Too Many Instruments, *Oxford Bulletin of Economics and Statistics* **71**(1), 135–158. DOI: 10.1111/j.1468-0084.2008.00542.x

Ruud, P.A. :2000, *An Introduction to Classical Econometric Theory*, Oxford University Press.

Stock, J. and J. Wright: 2000, GMM with Weak Identification, *Econometrica* **68**(5), 1055-1096.

Windmeijer, F.: 2000, A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators, *Journal of Econometrics* **126**(1), 25–51. DOI: 10.1016/j.jeconom.2004.02.005

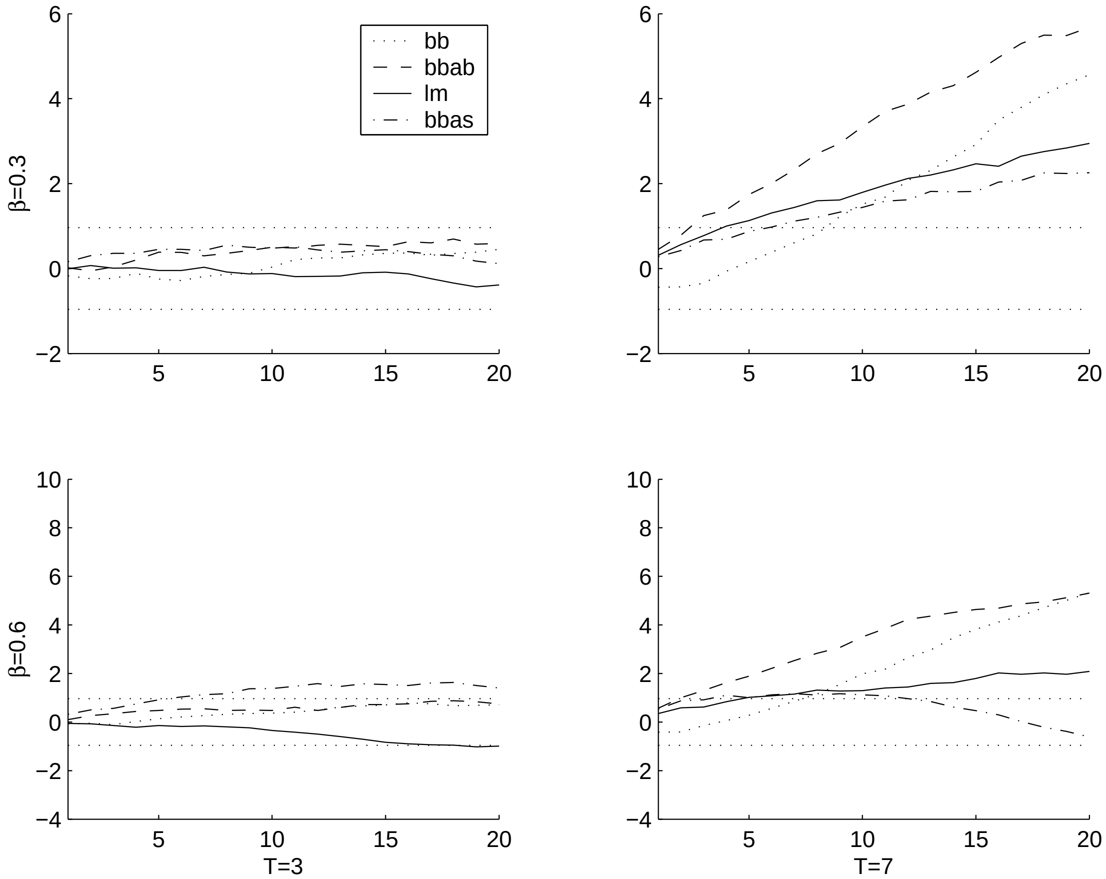

Figure 1: p-value discrepancy plots, $\sigma_\alpha^2 = 1$, (x-axis: nominal size (from 1% to 20%); y-axis: difference between the share of rejections and nominal size).

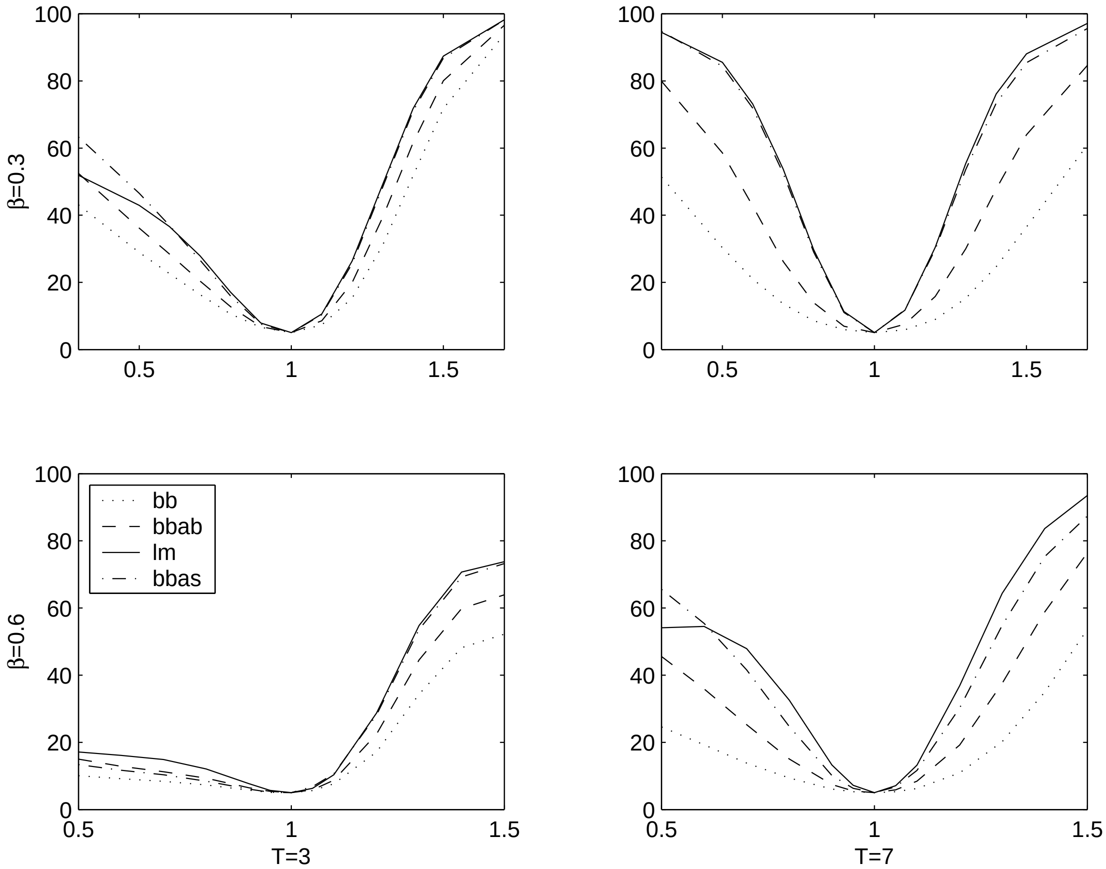

Figure 2: Power of the tests, $\sigma_\alpha^2 = 1$. On the x-axis: γ ($\gamma = 1$ implies 'mean stationarity'); on the y-axis: (adjusted) share of rejections with 5% nominal size.

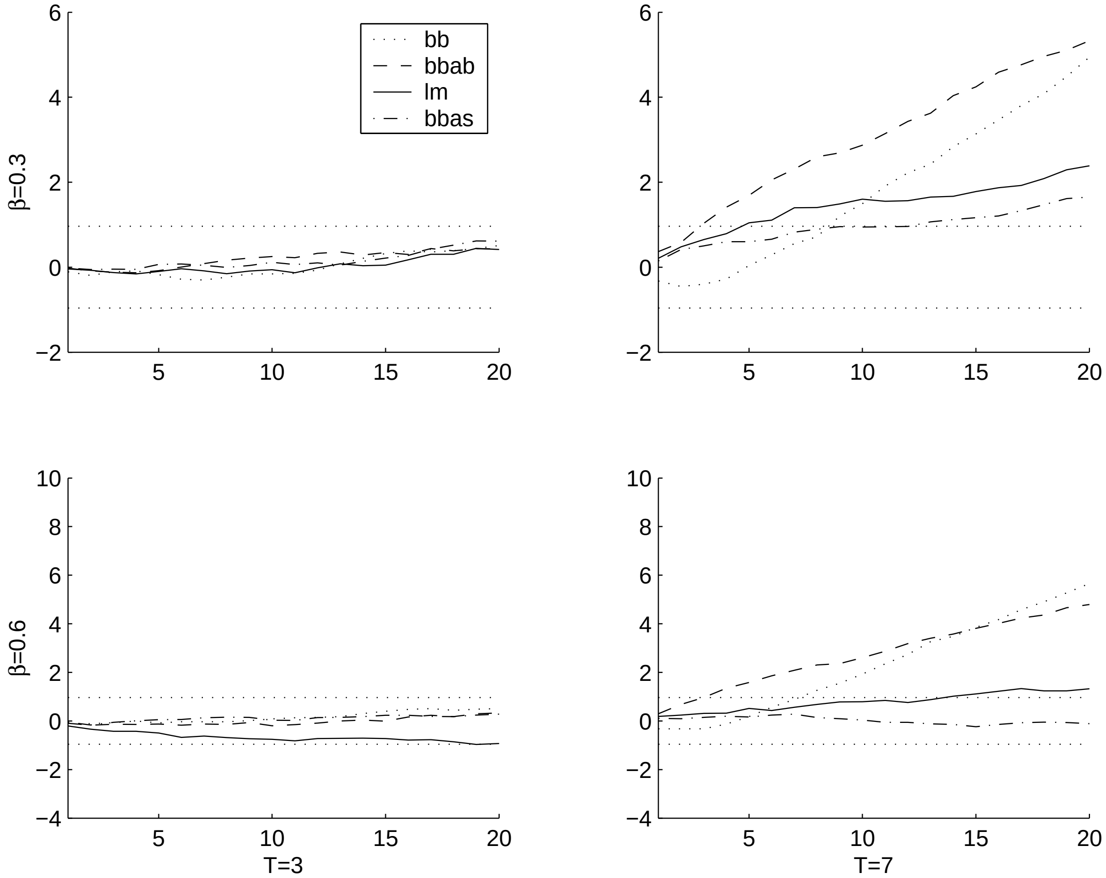

Figure 3: p-value discrepancy plots, $\sigma_\alpha^2 = 1/4$ (x-axis: nominal size (from 1% to 20%); y-axis: difference between the share of rejections and nominal size).

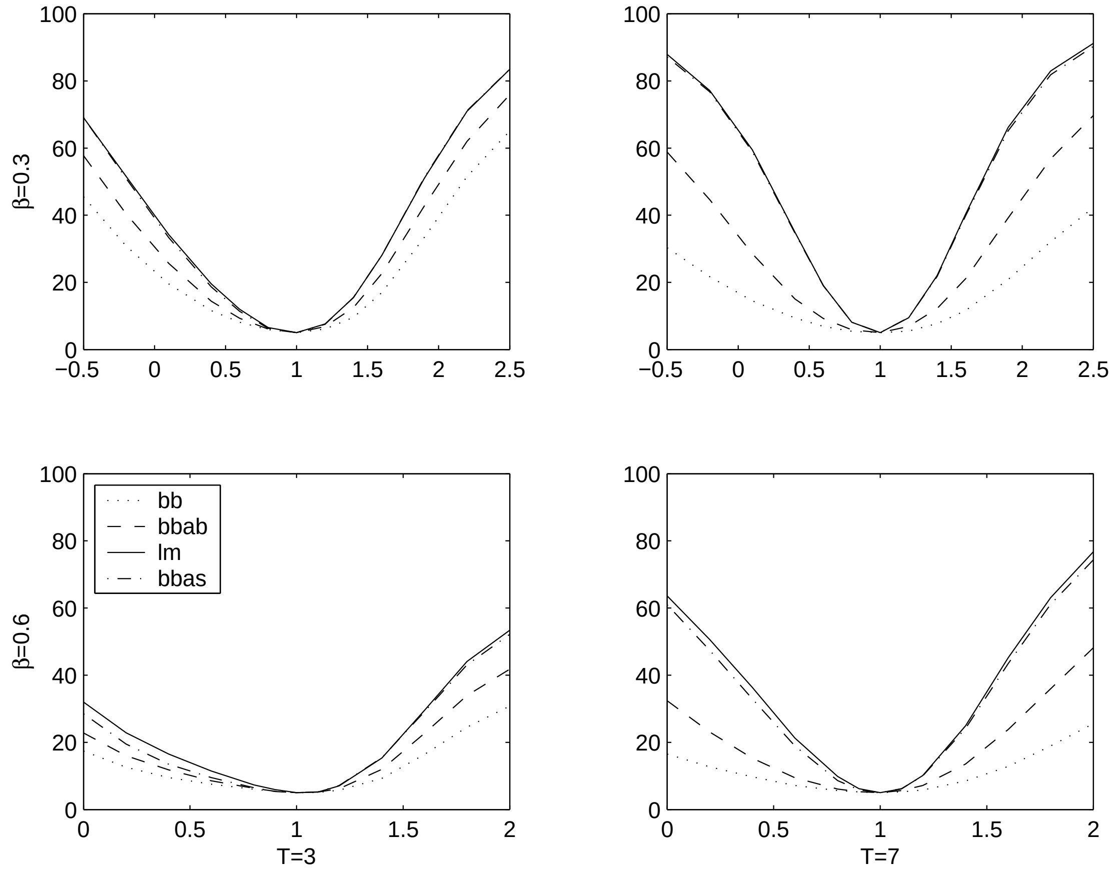

Figure 4: Power of the tests, $\sigma_\alpha^2 = 1/4$. On the x-axis: γ ($\gamma = 1$ implies 'mean stationarity'); on the y-axis: (adjusted) share of rejections with 5% nominal size.

β	γ	$\sigma_\alpha^2 = 1$			
		$T = 3$		$T = 7$	
		mean	variance ($\times 100$)	mean	variance ($\times 100$)
$\beta = 0.3$	0.30	.6440	1.063	.3981	.5001
	1.00	.3057	1.457	.2878	.3770
	1.70	.4518	1.931	.3064	.4019
$\beta = 0.6$	0.50	.8872	.4547	.7630	.3882
	1.00	.6001	1.884	.5858	.4723
	1.50	.9037	1.538	.6418	.6211

β	γ	$\sigma_\alpha^2 = 1/4$			
		$T = 3$		$T = 7$	
		mean	variance ($\times 100$)	mean	variance ($\times 100$)
$\beta = 0.3$	-0.50	.4394	1.185	.3365	.3831
	1.00	.2952	1.186	.2843	.3486
	2.50	.3728	1.303	.2959	.3541
$\beta = 0.6$	0.00	.7749	1.077	.6666	.4124
	1.00	.5881	1.542	.5754	.4053
	2.00	.7197	1.932	.5838	.4711

Table 1: 20,000 Monte Carlo replications: mean and variance of system GMM estimator $\hat{\beta}$.

DGP	system GMM		5%-level rejections		
	mean $\hat{\beta}_1$ (var. $\times 100$)	mean $\hat{\beta}_2$ (var. $\times 100$)	BB	BBAB	LM
		$T = 3$			
$\gamma = \mu = 1$	.5526	1.085	4.15	5.80	6.09
	(1.778)	(29.44)			
$\gamma = 1; \mu = 0.5$	.6201	1.542	8.71	14.26	18.54
	(1.225)	(20.629			
$\gamma = 1; \mu = 1.5$	.6646	1.255	13.28	20.71	24.69
	(.9586)	(32.88)			
$\gamma = 0.5; \mu = 1$	.7013	1.541	10.61	19.15	19.65
	(.7687)	(21.37)			
$\gamma = 1.5; \mu = 1$	.6620	1.645	16.21	25.28	27.22
	(.6088)	(20.06)			
$\gamma = \mu = 0.5$	.7169	1.570	6.48	12.07	10.48
	(.8591)	(21.30)			
$\gamma = \mu = 1.5$	.6404	1.599	20.83	32.36	30.75
	(.4542)	(20.37)			
		$T = 7$			
$\gamma = \mu = 1$	.5487	1.004	0.82	6.44	8.05
	(.4131)	(6.003)			
$\gamma = 1; \mu = 0.5$	.6002	1.204	2.16	16.16	24.07
	(.3563)	(5.137)			
$\gamma = 1; \mu = 1.5$	.5866	1.012	3.32	22.05	30.43
	(.3761)	(6.469)			
$\gamma = 0.5; \mu = 1$	.6669	1.293	3.35	24.20	29.07
	(.1962)	(5.067)			
$\gamma = 1.5; \mu = 1$	.6134	1.391	9.83	43.81	59.06
	(.1787)	(5.173)			
$\gamma = \mu = 0.5$	.6802	1.320	2.85	21.22	24.27
	(.2059)	(4.899)			
$\gamma = \mu = 1.5$	.6165	1.402	12.40	50.02	63.37
	(.1372)	(5.515)			

Table 2: Monte Carlo results, model with an endogenous x, $\beta_1 = \beta_3 = 0.5$