Financial Econometrics Series

SWP 2015/18

Unit Root Inference in Generally Trending and Cross-Correlated Fixed-T Panels

D. Robertson, V. Sarafidis, and J. Westerlund

The working papers are a series of manuscripts in their draft form. Please do not quote without obtaining the author's consent as these works are in their draft form. The views expressed in this paper are those of the author and not necessarily endorsed by the School or IBISWorld Pty Ltd.
UNIT ROOT INference in Generally Trending and 
CROSS-Correlated Fixed-T Panels

Donald Robertson
University of Cambridge

Vasilis Sarafidis
Monash University

Joakim Westerlund†
Lund University

and

Centre for Economics and Financial Econometrics Research
Deakin University

July 31, 2015

Abstract

This paper proposes a new panel unit root test based on the generalized method of moments approach for panels with a possibly small number of time periods, $T$, and a large number of cross-sectional units, $N$. In the model that we consider the deterministic trend function is essentially unrestricted and the errors obey a multi-factor structure that allows for rich forms of unobserved heterogeneity. In spite of these allowances, the GMM estimator considered is shown to be asymptotically unbiased, $\sqrt{N}$-consistent and asymptotically normal for all values of the autoregressive (AR) coefficient, $\rho$, including unity, making it an ideal candidate for unit root inference. Results from our Monte Carlo study suggest that the asymptotic properties are borne out well in small samples. The implementation of our approach is illustrated by using a large sample of US banking institutions to test Gibrat’s Law.

JEL Classification: C12; C13; C33; C36.

Keywords: Panel data; unit root test; unobserved heterogeneity; common factors; GMM.

* A previous version of this paper was presented at the 1st Annual IAAE Conference and the 20th International Panel Data Conference. The authors would like to thank seminar participants, and in particular Shakeeb Khan (Joint Editor), an Associate Editor and two anonymous referees for their very constructive comments. Westerlund thanks the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship, as well as the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014-0112:1.
† Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.
1 Introduction

There is a voluminous literature on panel unit root tests. The main motivation for using such procedures is that by considering not one but $N$ time series of length $T$ the power of panel-based tests can increase considerably relative to that achievable using univariate tests. The largest branch of the literature by far is that focusing on panels where both $N$ and $T$ are large (see Breitung and Pesaran, 2008, for an overview). A typical study assumes that $N, T \to \infty$ such that $N/T \to 0$. The main reason for this is the presence of cross-sectional heterogeneity, such as individual-specific effects, the consistent estimation of which requires $T \to \infty$. This induces an estimation error, which can be controlled, but typically only if $N/T \to 0$, for otherwise the bias is not eliminated as $N \to \infty$ (see Westerlund and Breitung, 2013, Section 5, for a detailed discussion). This requirement may put strain on the data. Indeed, as a large body of Monte Carlo evidence shows, while the large-$N$ requirement is usually not a problem, the large-$T$ requirement, and in particular the requirement that $T$ must be larger than $N$, poses a real restriction (see, for example, De Wachter et al., 2007; Hlouskova and Wagner, 2006), to the extent that researchers might well consider discarding data in order to have $N$ sufficiently small relative to $T$.\footnote{For example, while the Penn World Tables have $N \gg T$, by considering only the OECD countries, one can make $N \ll T$.} Moreover, in many panels, such as those frequently encountered in applied microeconometrics, $T (N)$ is simply too small (large) for such discarding practices to make sense, although the unit root hypothesis is still of considerable interest (see Bond et al., 2005).

Discussions like the one in the previous paragraph have motivated researchers to look for inferential procedures that are suitable in fixed-$T$ panels. Harris and Tzavalis (1999) were among the first. They proposed a fixed-$T$ panel unit root test based on the bias-corrected ordinary least squares (OLS) estimator of the autoregressive (AR) coefficient, $\rho$. Many other tests have since then been proposed (see De Blander and Dhaene, 2012, and the references provided therein).\footnote{See Breitung and Pesaran (2008) for a survey of the panel unit root and cointegration literatures.} The evidence reported so far suggests that in terms of small-sample performance, not requiring $T$ to be large can be a great advantage (see, for example, Harris and Tzavalis, 1999; Hadri and Larsson, 2005; Hlouskova and Wagner, 2006). In fact, fixed-$T$ tests often outperform large-$T$ tests and do so for a wide range of values of $T$. However, while much progress has been made, there are (at least) two important issues that have not received much attention. First, except for De Blander and Dhaene (2012), Harris and Tzavalis (1999, 2004), and Han and Phillips (2010), who consider the case of a linear trend, the fixed-$T$ literature has not yet ventured outside the...
fixed effects environment. This is noteworthy because if one admits to the possibility that time series might be trending, then the probability that the panel of multiple time series exhibits at least some trending behavior will tend to one as $N \to \infty$, in which case fixed effects-only tests are rendered invalid. Second, unlike several large-$T$ “second-generation” unit root tests that allow for cross-sectional dependence in the form of common factors (see, for example, Moon and Perron, 2004, Phillips and Sul, 2003, Bai and Ng, 2004, Pesaran, 2007, among others), as far as we are aware, there is presently no unit root test that is able to accommodate such dependence; that is, existing fixed-$T$ tests are “first-generation” tests (Baltagi, 2008, Chapter 12). A noteworthy exception is the recent working paper by Karavias and Tzavalis (2014), in which the authors allow for spatial correlation, representing “weak” cross-sectional dependence.\footnote{See Chudik et al. (2011) for a detailed treatment of the concepts of weak and strong cross-sectional dependence.}

The purpose of the current paper is to address both of the above mentioned issues. Specifically, a second-generation approach to unit roots in fixed-$T$ panels is proposed that allows for both cross-sectional dependence and generally trending behavior. This is accomplished by letting the data admit to a common factor structure, in which the unobserved factors are treated as unknown parameters to be estimated along with the other parameters of the model. This parametric treatment means that the factors are virtually unrestricted, apart from some mild regularity conditions. It also provides a way to control for (unobserved) deterministic trend terms, which in our model appear naturally as additional factors. In the terminology of Bai (2009), the model that we consider constitutes an “interactive effects” model. Interestingly, since factors are estimated, the usual problem in empirical work of deciding on which deterministic terms to include does not arise. Hence, the approach is not only general, but is in this sense also remarkably simple. This is in stark contrast to the existing large-$T$ literature, where the factors are incidental parameters, the number of which increases with $T$. Not only does this complicate estimation, but it also restricts quite substantially the types of factors that can be permitted.

The estimation is carried out by modifying the generalized method of moments (GMM) approach of Robertson and Sarafidis (2015). The new estimator is shown to have a number of desirable properties. First, it circumvents the usual incidental parameter bias problem, which arises because the number of factor loading parameters that has to be estimated increases with $N$. This is true both in the conventional fixed effects case and in the more general interactive effects model considered here. The reason is that we do not require estimates of the loadings themselves, but only estimation of their normalized second moment, whose dimension remains fixed as $N$ grows. Second, the estimator supports asymptotically normal inference for all val-
ues of $\rho$, including unity, and the well-known identification problems that arise when $\rho$ equals unity do not emerge (see, for example, Bun and Windmeijer, 2010, for a discussion of this issue). Hence, unlike most existing approaches, the limiting distribution of the GMM estimator considered here is continuous and has the same rate of consistency as $\rho$ passes through unity (see Han and Phillips, 2010, for a similar result). Finally, the estimator and the associated $t$-statistic for a unit root appear to have satisfactory small-sample properties.

The remainder of the paper is organized as follows. Section 2 presents the model and assumptions, which are used in Section 3 to derive the GMM estimator and its asymptotic distribution. The empirical usefulness of the new estimator is evaluated using both simulated and real data in Sections 4 and 5, respectively. Section 6 concludes.

2 Model

Consider the panel data variable $y_{i,t}$, which is observed over $t = 0, 1, \ldots, T$ time series and $i = 1, \ldots, N$ cross-sectional units. The data generating process (DGP) of this variable is assumed to be given by

$$ y_{i,t} = \rho y_{i,t-1} + u_{i,t}, $$

$$ u_{i,t} = \lambda_i' f_t + \varepsilon_{i,t}, $$

for $t = 1, \ldots, T$, where $\rho \in \mathbb{R}$, $f_t$ is an $r \times 1$ vector of common factors, $\lambda_i$ is the associated vector of factor loadings, and $\varepsilon_{i,t}$ is an idiosyncratic error term. The following assumptions are assumed to hold throughout the paper.

Assumption 1. $\varepsilon_{i,t} \sim \text{iid}(0, \sigma^2_{\varepsilon})$ and has finite moments up to fourth order.

Assumption 2. $\lambda_i \sim \text{iid}(0, \Sigma_\lambda)$ with finite moments up to fourth order. Furthermore, $\Sigma_\lambda$ is positive definite.

Assumption 3. $f_t$ is non-stochastic such that $\|f_t\| < \infty$.

Assumption 4. $y_{i,0}$ is iid and has finite moments up to fourth order.

Assumption 5. $E(\varepsilon_{i,t}|y_{i,0}, \ldots, y_{i,t-1}, \lambda_i') = 0$.

Assumptions 1–2 are mainly employed for simplicity and can be relaxed to allow for more general DGPs. For example, analogously to the existing dynamic panel GMM literature, arbitrary (unconditional) time series and cross-sectional heteroskedasticity in $\varepsilon_{i,t}$ can be allowed.
by using the standard sandwich formula of the covariance matrix of the GMM estimator.\textsuperscript{4} The independence assumption across $i$ can also be relaxed to allow for weak cross-sectional correlation. However, this would require replacing Assumption 1 with a “high-level” assumption like Assumption C in Bai and Ng (2004), which we would like to avoid. Similarly, while the conditional moments of $\lambda_i$ can be heterogeneously distributed, we avoid such generalizations in order to reduce the notational burden in the paper.

According to Assumption 3, $f_t$ is treated as a fixed parameter vector to be estimated along with the remaining parameters of the model. Such parametric treatments are common in the related large-$T$ literature; however, since in that literature $T$ grows, $f_t$ has to be restricted. For example, it is standard practice to assume that $f_t$ has some stationary distribution with mean zero and finite moments up to fourth order, which of course rules out the presence of deterministic constant and trend terms. Assumption 3 does not impose any distributional assumptions of this kind and is in this sense very general.

Assumption 4 basically imposes similar conditions for the initial observation as those that apply to the error components of the model. However, as will soon become clear, the functional form of $y_{i,0}$ remains unrestricted. For instance, one could set $y_{i,0} = \lambda_i^*f_0 + \varepsilon_{i,0}^*$. In this case, $y_{i,0}$ could be thought of as the reduced form of $y_{i,t}$ at period $t = 0$, which is assumed to be observed. Thus, the values of $\lambda_i^*$ and $\varepsilon_{i,0}^*$ are not necessarily identical to the values of $\lambda_i$ and $\varepsilon_{i,0}$ that would arise had $y_{i,0}$ been assumed to follow (1).

Assumption 5 implies that $\varepsilon_{i,t}$ is serially uncorrelated. This can also be relaxed in a straightforward way. In particular, a moving average (MA) process of order $q$ can be accommodated by replacing Assumption 5 with $E(\varepsilon_{i,t}|y_{i,0}, \ldots, y_{i,t-1-q}, \lambda_i^1) = 0$, for $t \geq 1 - q$. An autoregressive process (AR) in $\varepsilon_{i,t}$ can be accommodated by simply augmenting the right-hand side of (1) with more lags of $y_{i,t}$. This is discussed in detail in Remark 3, and then again in Section 5, where we implement our approach in the presence of MA errors. In addition, Assumption 5 implies no correlation between the idiosyncratic error and the factor loadings. This assumption is standard in dynamic panels and it could be relaxed, although the computational task becomes far more complex in this case.\textsuperscript{5} The assumption that $\varepsilon_{i,t}$ is uncorrelated with $\lambda_i$ is also standard in the large $T$-literature (see, for example, Moon and Perron, 2004; Bai and Ng, 2004; Pesaran, 2007).

\textsuperscript{4}Note that $\varepsilon_{i,t}$ is not required to be conditionally homoskedastic.

\textsuperscript{5}For instance, in the simple one-way error components model with $u_{i,t} = \eta_i + \varepsilon_{i,t}$, the GMM estimator proposed by Arellano and Bond (1991) employs moment conditions of the form $E(\varepsilon_{i,t}u_{i,s}) = 0$ for $s < t - 1$. It is easy to see that this generally requires $E(\varepsilon_{i,t}^2|\eta_i) = 0$, because otherwise $y_{i,s}$ will be correlated with the transformed error, since they are both functions of $\eta_i$. 

\begin{center}
5
\end{center}
The above DGP can be seen as a version of the type of components model commonly used in the large-$T$ (panel) unit root literature, in which the deterministic and stochastic elements of the model are separated explicitly. In order to illustrate this point, suppose that

\[
y_{i,t} = d_{i,t} + z_{i,t}, \\
z_{i,t} = \rho z_{i,t-1} + v_{i,t}, \\
v_{i,t} = \gamma'_i \mathbf{w}_t + \epsilon_{i,t},
\]

where $d_{i,t}$ represents the deterministic component of $y_{i,t}$, and $\mathbf{w}_t$ is an $m \times 1$ vector of “genuine” common factors. Multiplying the first lag of expression (3) by $\rho$ and subtracting from the original equation yields

\[
y_{t} = \rho y_{i,t-1} + d_{i,t} - \rho d_{i,t-1} + \gamma'_i \mathbf{w}_t + \epsilon_{i,t},
\]

which is equivalent to (1) and (2) with $\lambda'_i \mathbf{f}_t = d_{i,t} - \rho d_{i,t-1} + \gamma'_i \mathbf{w}_t$. To appreciate what implications the formulation in (6) has for the DGP in (1) and (2) it is useful to consider a few specific cases. The two most common specifications of $d_{i,t}$ by far are $d_{i,t} = \eta_i$ (fixed effects), and $d_{i,t} = \eta_i + \beta_i t$ (incidental trends). In the first specification, $d_{i,t} - \rho d_{i,t-1} = \eta_i(1 - \rho)$, whereas in the second, $d_{i,t} - \rho d_{i,t-1} = \eta_i(1 - \rho) + \beta_i \rho + \beta_i(1 - \rho)t$. One of the implications of (6) is therefore to make the order of the trend polynomial a function of $\rho$. In the trend case, for example, when $\rho = 1$ (6) corresponds to (1) and (2) with $r = m + 1$, $\lambda_i = (\beta_i, \gamma'_i)'$ and $\mathbf{f}_t = (1, \mathbf{w}_t)'$. On the other hand, when $|\rho| < 1$ the same model corresponds to (1) and (2) with $r = m + 2$, $\lambda_i = (\eta_i(1 - \rho) + \beta_i \rho, \beta_i(1 - \rho), \gamma'_i)'$ and $\mathbf{f}_t = (1, t, \mathbf{w}_t)'$. Hence, when cast in terms of (1) and (2), the fact that in equation (6) the order of the trend polynomial is reduced when $\rho = 1$ implies that one of the elements of $\lambda_i$ drops out. While this does not fit well with our assumption that $\Sigma_\lambda$ is positive definite, in practice there is an easy way out of this, which is to make the estimation procedure conditional on $r$. This treatment stands in sharp contrast with the standard dynamic panel data literature (see the detailed discussion in Section 3), as well as with the large-$T$ literature, in which the estimation is carried out conditional on knowing $d_{i,t}$. The fact is that the GMM approach developed here does not require $d_{i,t}$ to be pre-specified, which is a great advantage in practice, as the researcher is spared the problem of having to decide on which deterministic components to include. If $r$ is unknown, as is usually the case in practice, then an information criterion may be used to obtain a consistent estimator. In Section 3 we elaborate on this point.

\footnote{See Schmidt and Phillips (1992) for a discussion in the pure time series context.}
It is important to note that our approach is not limited to incidental intercept and trend terms only; on the contrary, it can accommodate virtually any trend function that is linear in the parameters, including polynomial trend functions, trigonometric functions and models of discrete and smooth structural shifts. This is again in stark contrast to the bulk of the existing large-$T$ literature, which has not yet ventured outside the linear trend environment (see Westerlund, 2015, for an exception). The reason is the presence of incidental parameter bias, the analytical complexity of which increases very quickly with both the number and the non-linearity of the trend terms. The implication is that once outside the linear trend environment, bias-correction is not really an attractive option. Against this background, the asymptotic unbiasedness of the GMM estimator developed here is clearly a great advantage, as it enables testing in situations that were previously not possible.

3 The GMM-based test approach

The DGP in (1) and (2) can be written in stacked vector form as

$$y_i = \rho y_{i-1} + F\lambda_i + \epsilon_i = (I_T \otimes \lambda'_i)f + \epsilon_i,$$  \hspace{1cm} (7)

where $y_i = (y_{i,1}, \ldots, y_{i,T})'$, $y_{i-1} = (y_{i,0}, \ldots, y_{i,T-1})'$ and $\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,T})'$ are $T \times 1$ vectors, $F = (f_1, \ldots, f_T)'$ a $T \times r$ matrix, and $f = \text{vec}(F')$.

Let $S_t$ be the $M_t \times T$ selector matrix of zeroes and ones that for each time period $t$ picks out the $M_t$ entries of $y_{i,-1}$ that are uncorrelated with $\epsilon_{i,t}$, i.e. $S_t$ is such that $E[(S_t y_{i,-1})\epsilon_{i,t}] = 0$.

Under Assumption 5 we have $S_t = (I_t, 0_{t \times (T-t)})$, a $t \times T$ matrix; therefore, at time $t$ the vector of valid instruments is simply given by $y_{i,0}, \ldots, y_{i,t-1}$. Define $S = \text{diag}(S_1, \ldots, S_T)$, a $M \times T^2$ block diagonal matrix, where $M = \sum_{t=1}^{T} M_t = T(T+1)/2$ in the present case. The $M \times T$ matrix of instruments can now be written as

$$Z'_i = S(I_T \otimes y_{i,-1}).$$  \hspace{1cm} (8)

It is easy to verify that $E(Z'_i \epsilon_i) = 0_{M \times 1}$. As a result, multiplying (7) by $Z'_i$ and taking expectations yields

$$m = \rho m_1 + S(I_T \otimes G)f,$$  \hspace{1cm} (9)

where $m = E(Z'_i y_i)$, $m_1 = E(Z'_i y_{i,-1})$ and $G = E(y_{i,-1}\lambda'_i)$, a $T \times r$ matrix.\(^7\)

\(^7\)The existence of $G$ is implied by Assumptions 1–4.
The dimension of $G$ can be reduced substantially by taking into account the AR structure of $y_{i,t}$. The latter implies that $y_{i,-1}$ can be written as follows:

$$y_{i,-1} = \Gamma e_{1} y_{i,0} + \Gamma \tilde{\epsilon}_{i,-1},$$

(10)

where $e_{t} = (0, ..., 0, 1, 0, ..., 0)'$ is a $T \times 1$ vector of zeros, except for a one sitting in position $t$, $\tilde{\epsilon}_{i} = (0_{T \times 1}, f_1, ..., f_{T-1})'$ and $\tilde{\epsilon}_{i,-1} = (0, \epsilon_{i,1}, ..., \epsilon_{i,T-1})'$. Also,

$$\Gamma = (I_{T} - \rho L)^{-1} = \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ \rho & 1 & \ldots & 0 & 0 \\ \rho^2 & \rho & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ \rho^{T-1} & \rho^{T-2} & \ldots & \rho & 1 \end{bmatrix}; \quad L = \begin{bmatrix} 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix},$$

both being $T \times T$ matrices. Using the expression in (10) to substitute for $y_{i,-1}$ in $E(y_{i,-1} \lambda_{i}')$ yields

$$E(y_{i,-1} \lambda_{i}') = E((\Gamma e_{1} y_{i,0} + \Gamma \tilde{\epsilon}_{i,-1} \lambda_{i}') = \Gamma e_{1} g_{0}' + \Gamma \tilde{\epsilon}_{i,-1} \Sigma_{\lambda},$$

(11)

where $g_{s} = E(y_{i,s} \lambda_{i})$ for any $s \geq 0$. This means that given $\rho$, $f$ and $\Sigma_{\lambda}$, the only unknown parameters in $G$ are those contained in the $r \times 1$ vector $g_{0}$. Thus, making use of (11) leads to a more parsimonious formulation of the model when compared to (9). As a result, noting that $f = (I_{T} \otimes F)'e$, where $e = \text{vec}(I_{T})$, the proposed vector of moment conditions is given by

$$m - \rho m_{1} - S[I_{T} \otimes (\Gamma e_{1} g_{0}' F' + \Gamma \tilde{\epsilon}_{i,-1} \Sigma_{\lambda} F')] e = 0_{M \times 1}.$$

(12)

Our model is not identified as it stands because the vector of moment conditions in (12) contains products of unknown parameters. In particular, $g_{0}' F'$ and $\tilde{\epsilon}_{i,-1} \Sigma_{\lambda} F'$ are not separately identifiable without normalizing restrictions. In what follows we therefore set, with a slight abuse of notation, $\tilde{\epsilon}_{i} = \tilde{\epsilon}_{i,-1} \Sigma_{\lambda}^{1/2}$, $F = \tilde{\epsilon}_{i,-1} \Sigma_{\lambda}^{1/2} F = \tilde{\epsilon}_{i,-1} \Sigma_{\lambda}^{-1/2} g_{0}$. Hence, using $f$ to denote the vectorized version of the normalized $F'$, the vector of parameters of interest is given by $\theta = (\rho, f', g_{0}')' = (\theta_1, \theta_2, \theta_3)'$, which is of order $\dim(\theta) \times 1$, where $\dim(\theta) = 1 + (T + 1)r$. The proposed GMM estimator of this parameter vector based on the above moment conditions is given by

$$\hat{\theta} = \arg \min_{\theta \in \Theta} Q_{N}(\theta),$$

where $\Theta$ is a compact subset of $\mathbb{R}^{\dim(\theta)}$ and

$$Q_{N}(\theta) = \bar{h}_{N}(\theta)' W_{N} \bar{h}_{N}(\theta),$$

$$\bar{h}_{N}(\theta) = \frac{1}{N} \sum_{i=1}^{N} h_{i}(\theta),$$

$$h_{i}(\theta) = Z_{i}' y_{i} - \rho Z_{i}' y_{i,-1} - S[I_{T} \otimes (\Gamma e_{1} g_{0}' F' + \Gamma \tilde{\epsilon}_{i,-1} F')] e,$$
where $W_N$ is an $M \times M$ weighting matrix.

Remark 1. Ignoring the structure of $G$ simplifies the estimation burden considerably because, letting $g = \text{vec}(G)$, (9) becomes $m = \rho m_1 + S(I_T \otimes F)g$. That is, for a given value of $F$, the model is linear and so it can be estimated using OLS. By contrast, the proposed GMM estimator exploits the non-linear restrictions in $G$ and is therefore non-linear. This extra complication is, however, well worthwhile, in the sense that ignoring the non-linear restrictions can lead to a substantial loss of efficiency (see, for example, Wansbeek and Bekker, 1996; Robertson and Sarafidis, 2015).

Remark 2. The moment conditions in (12) can be modified to allow for MA errors, or a more general AR structure for $y_{i,t}$. Suppose first that $\varepsilon_{i,t}$ follows an MA($q$) process. This case can be accommodated by setting $Z_i' = S(I_T \otimes y_{i,\cdot-q+1})$, where $y_{i,\cdot-q+1} = (y_{i,0}, \ldots, y_{i,T-1-q})'$ and the $t$-th diagonal element of $S$ is given by $S_t = (I_t, 0_{T \times (T-t-q)})$, which has dimension $t \times (T-q)$. In Section 5 we implement our approach for $q \in \{1, 2\}$.

Consider next the case when $\varepsilon_{i,t}$ is serially uncorrelated, but $y_{i,t}$ follows an AR(2) process;

$$y_{i,t} = \rho_1 y_{i,t-1} + \rho_2 y_{i,t-2} + u_{i,t},$$

where we assume for notational simplicity that $y_{i,0}$ and $y_{i,-1}$ are observed. This model can be written in vector form as

$$y_i = \rho_1 y_{i,-1} + \rho_2 y_{i,-2} + (I_T \otimes \lambda_i')f + \varepsilon_i,$$

where $y_{i,-2} = (y_{i,-1}, y_{i,0}, \ldots, y_{i,T-2})'$ and the remaining vectors are as before. In this case the vector of instruments is given by $y_{i,-1}^+ = (y_{i,-1}, y_{i,0}, \ldots, y_{i,T-1})'$, which can be written as

$$y_{i,-1}^+ = \Gamma e_2 y_{i,0} + \Pi e_1 y_{i,-1} + \Gamma \tilde{F}_{-1}^+ \lambda_i + \Gamma \tilde{\varepsilon}_{i,-1}^+,$$

where $\Gamma = (I_T - \rho_1 L - \rho_2 L^2)^{-1}$, $\tilde{F}_{-1}^+ = (0_{T \times 2}, f_1, \ldots, f_{T-1})'$, $\tilde{\varepsilon}_{i,-1}^+ = (0_{T \times 2}, \varepsilon_{i,1}, \ldots, \varepsilon_{i,T-1})'$ and $\Pi$ is a $T \times T$ matrix that is a function of $\rho_1$, $\rho_2$ and $T$. In this case, $Z_i' = S(I_T \otimes y_{i,-1}^+)$ with $S_t = (I_{t+1}, 0_{(t+1) \times (T-t)})$, a $(t + 1) \times (T + 1)$ matrix. Thus, pre-multiplying equation (13) by $Z_i'$, taking expectations (with the normalization enforced) and rearranging yields

$$m - \rho_1 m_1 - \rho_2 m_2 - S[I_T \otimes (\Gamma e_2 g_0'F' + \Pi e_1 g_{i,-1}'F' + \Gamma \tilde{F}_{-1}^+ F')]e = 0_{M \times 1},$$

where $m_2 = E(Z_i'y_{i,-2})$. 

9
Let
\[ \Delta = E[\nabla \theta h_i(\theta_0)], \quad \Omega = E[h_i(\theta_0)h_i(\theta_0)'], \]
where \( \theta_0 \) denotes the true value of \( \theta \). The following assumptions are sufficient to establish the asymptotic properties of our estimator:

**Assumption 6.**

(a) \( W_N \xrightarrow{p} W \) as \( N \to \infty \), where \( W \) is a positive definite matrix;

(b) \( \theta_0 \) belongs to the interior of \( \Theta \);

(c) \( E[h_i(\theta)] = 0_{M \times 1} \) if and only if \( \theta = \theta_0 \);

(d) \( \Delta \) and \( \Omega \) exist and both are full rank matrices.

**Remark 3.** One implication of Assumption 6 (d) is that \( M \), the number of moment conditions, is at least as large as the number of parameters in \( \theta \). Also, while Assumptions 1–6 (a)–(c) could be sufficient to ensure identification in the standard fixed effects and incidental trend cases, Assumption 6 (d) is required to ensure that identification holds in the more general interactive effects model considered here.

**Theorem 1.** Under Assumptions 1–6, as \( N \to \infty \),
\[ \sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0_{\text{dim}(\theta) \times 1}, \Sigma_{\hat{\theta}}), \]
where \( \Sigma_{\hat{\theta}} = (\Delta'W\Delta)^{-1}(\Delta'W\Omega W\Delta)(\Delta'W\Delta)^{-1} \).

**Proof.** See Appendix.

An analytical expression for \( \Delta \) is given in Appendix. If \( W = \Omega^{-1} \), the covariance matrix of \( \hat{\theta} \) reduces to \( \Sigma_{\hat{\theta}} = (\Delta'\Omega^{-1}\Delta)^{-1} \), which is optimal. The same holds if \( \Omega^{-1} \) is replaced with a consistent estimate. This can be achieved by setting
\[ \hat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} h_i(\hat{\theta}^1)h_i(\hat{\theta}^1)', \]
where \( \hat{\theta}^1 \) denotes any first-stage consistent GMM estimate of \( \theta \), such as the one obtained by setting \( W = I_M \). Notice that if interest lies only in testing the unit root null hypothesis, one can
compute \( \hat{\Omega} \) by setting the first entry in \( \hat{\theta} \) within \( h_1(\hat{\theta}) \) equal to unity. Alternatively, one can leave the first entry in \( \hat{\theta} \) unrestricted in the estimation and impose the unit restriction when making inference. This is the practice we have followed in our simulations.

**Remark 4.** According to Theorem 1 there is no asymptotic bias despite the generality of the DGP considered, that is, \((\hat{\theta} - \theta_0)\) is centered at zero even when scaled by \( \sqrt{N} \). The reason is that the GMM approach developed here does not require estimating \( \lambda_1, ..., \lambda_N \), which implies that the number of parameters that needs to be estimated remains fixed as \( N \) grows large (see Bai, 2013, for a similar approach based on maximum likelihood).

**Remark 5.** Theorem 1 holds for all values of \( \rho_0 \), and in this sense it presents a unified asymptotic result for the GMM estimator (see also the discussion provided at the end of this section). This is in contrast with the existing literature, in which the asymptotic distribution of dynamic panel estimators depends critically on whether \(|\rho_0| < 1, \rho_0 = 1\) or \( \rho_0 > 1 \). In fact, the only exceptions known to us are the GMM estimators of Han and Phillips (2010), and Kruiniger (2007, 2009, 2013), which have limit distributions that are continuous for \( \rho_0 \in (-1, 1) \), but not for \( \rho_0 > 1 \).

Let \( \hat{\theta} = (\hat{\rho}, \hat{f}', \hat{g}_0')' = (\hat{\theta}_1, \hat{\theta}_2', \hat{\theta}_3')' \), and denote by \( \hat{\Sigma}_b \) the associated GMM estimator of \( \Sigma_b \). The GMM-based \( t \)-statistic for testing \( H_0 : \rho_0 = c \) is given by

\[
t_{\hat{\rho}}(c) = \frac{\sqrt{N}(\hat{\rho} - c)}{\hat{\sigma}_{\hat{\rho}}},
\]

where \( \hat{\sigma}^2_{\hat{\rho}} \) is the first diagonal element of \( \hat{\Sigma}_b \). The local power of this test statistic is easily worked out using Theorem 1. Let us therefore assume that

\[
\rho_0 = c + \frac{\delta}{\sqrt{N}}, \tag{16}
\]

where \( \delta \in \mathbb{R} \). Since \( \hat{\sigma}^2_{\hat{\rho}} \) is consistent for \( \sigma^2_{\hat{\rho}} \), under the conditions of Theorem 1, we have

\[
t_{\hat{\rho}}(c) = \frac{\sqrt{N}(\hat{\rho} - c)}{\hat{\sigma}_{\hat{\rho}}} + \frac{\sqrt{N}(\rho_0 - c)}{\hat{\sigma}_{\hat{\rho}}} \xrightarrow{d} N(0, 1) + \frac{\delta}{\sigma_{\hat{\rho}}} \tag{17}
\]

as \( N \to \infty \).

The result in (17) nests the asymptotic results under both the null and the local alternative hypotheses. On the one hand, if \( \delta = 0 \), then \( H_0 \) is true and therefore \( t_{\hat{\rho}}(c) \xrightarrow{d} N(0, 1) \). If, on the other hand, \( \delta \neq 0 \), such that the local alternative is true, then the asymptotic distribution of \( t_{\hat{\rho}}(c) \) is no longer centered at zero, and therefore the test is unbiased. The magnitude of the power is driven by two parameters, \( \delta \) and \( \sigma^2_{\hat{\rho}} \); the smaller the uncertainty regarding \( \rho_0 \) and the
larger the deviation from the hypothesized value under $H_0$ are, the larger the power is (see Madsen, 2010, for a similar finding for some existing tests). However, given the generality of the DGP being considered, in terms of theoretical predictions this is as far as one can go without making more specific assumptions about the DGP. The fact that under our assumptions the test has power within the same shrinking neighborhood around the null hypothesis regardless of the specification of $f_t$ is consistent with the recent article of Karavias and Tzavalis (2015), who show that the well-known “incidental trends problem” need not arise in the current fixed-$T$ context.

**Remark 6.** The asymptotic distribution of most (if not all) unit root test statistics depends on the deterministic specification of the fitted test regression, which need not be equal to the true one. In time series, this implies that different deterministic specifications have their own critical values, whereas in panels, it implies that different specifications have their own mean and variance correction factors (see Westerlund and Breitung, 2013, Section 3). According to (17), the GMM-based $t$-statistic has a unique and practically very useful property in that, under the null hypothesis, it is asymptotically invariant to $f_t$ and hence to any trend function that it may contain. Therefore, mean and variance correction factors that depend on a particular deterministic specification are not required with our approach.

So far the number of factors has been treated as known, which need not be the case in practice. A natural approach towards this end is to treat the estimation of $r$ as a model selection problem, which can be handled using an information criterion. Let us denote by $r_0$ the “true” value of $r$, and let $\hat{\theta}(r)$ be the estimator of $\theta$ based on $r$ factors. The estimator of $r_0$ considered in the present paper is given by

$$\hat{r} = \arg \min_{r=0, \ldots, r_{\text{max}}} BIC(r)$$

with $r_0 \leq r_{\text{max}}$ and $BIC(r) = N \cdot Q_N(\hat{\theta}(r)) - \ln(N) \cdot b(r)$, where $Q_N(\hat{\theta}(r))$ is the value of the objective function evaluated at $\hat{\theta}(r)$, which is identically the value of the overidentifying restrictions $J$-statistic for the two-step estimator, and $b(r)$ is a penalty function that is strictly increasing in $r$. As usual, $b(r)$ is not given but has to be set by the researcher. In our simulations (see Section 4), we set $b(r) = \text{df}(r)/T^{0.3}$, where $\text{df}(r)$ is the degrees of freedom of the model based on $r$ factors.

---

8This criterion has been investigated extensively using Monte Carlo data by Ahn et al. (2013) and Robertson and Sarafidis (2015).
**Theorem 2.** Under the conditions of Theorem 1, as \( N \to \infty \),
\[
\hat{r} \overset{p}{\to} r_0.
\]

**Proof.** The proof of this theorem follows directly from the results in Robertson and Sarafidis (2015). It is therefore omitted.

Theorem 2 says that asymptotically knowing \( \hat{r} \) is as good as knowing \( r_0 \). This means that the result reported in Theorem 1 holds even if \( r_0 \) is replaced by \( \hat{r} \). To see that this must be the case, consider
\[
P(\sqrt{N}[\hat{\theta}(\hat{r}) - \theta_0] \leq \delta) = P(\sqrt{N}[\hat{\theta}(\hat{r}) - \theta_0] \leq \delta|\hat{r} = r_0)P(\hat{r} = r_0)
\]
\[
+ P(\sqrt{N}[\hat{\theta}(\hat{r}) - \theta_0] \leq \delta|\hat{r} \neq r_0)P(\hat{r} \neq r_0).
\]

Because \( P(\hat{r} = r_0) \to 1 \) and \( P(\hat{r} \neq r_0) \to 0 \) by Theorem 2, while the first term on the right-hand side converges to \( P(\sqrt{N}[\hat{\theta}(\hat{r}) - \theta_0] \leq \delta|\hat{r} = r_0) = P(\sqrt{N}[\hat{\theta}(r_0) - \theta_0] \leq \delta) \), the second term converges to zero. It follows that
\[
|P(\sqrt{N}[\hat{\theta}(\hat{r}) - \theta_0] \leq \delta) - P(\sqrt{N}[\hat{\theta}(r_0) - \theta_0] \leq \delta)| \to 0. \tag{19}
\]

**Remark 7.** Our testing procedure can be valid even if the DGP is different across individuals. Consider as an example the case when there are two clusters of cross-sectional units, \( C_1 \) and \( C_2 \). Let us denote by \( N_\ell \) the cardinality of \( C_\ell \), such that \( N_1 + N_2 = N \), and suppose that \( y_{i,t} = \rho y_{i,t-1} + \eta_i(1 - \rho) + \epsilon_{i,t} \) for \( i \in C_1 \), and \( y_{i,t} = \rho y_{i,t-1} + \eta_i(1 - \rho) + \beta_i(1 - \rho) + \epsilon_{i,t} \) for \( i \in C_2 \). Hence, if \( |\rho| < 1 \), then the number of cross-sectional units following a stationary process with a non-zero mean (trend) is given by \( N_1 \) (\( N_2 \)). Our procedure is consistent because the vector of moment conditions remains exactly as in (12), except that the value of \( \Sigma_\lambda \) will be different compared to the case where \( N_1 = 0 \). In the example considered here the last diagonal entry of \( \Sigma_\lambda \) will be equal to \( (1 - \pi)E(\beta_i^2)(1 - \rho_0)^2 \), where \( \pi = N_1 / N \). In Section 4 we use Monte Carlo simulations to investigate the effects of this type of clustering in small samples.

Under our assumptions, the model in (1) is identified for all values of \( \rho_0 \in \mathbb{R} \), including \( \rho_0 = 1 \). This is an important result that merits some discussion, as it is in contrast to known results for many existing GMM estimators for dynamic panel data models, such as the ones of Anderson and Hsiao (1981), and Arellano and Bond (1991), which can fail to identify \( \rho_0 \) when \( \rho_0 = 1 \) (see, for example, Bond et al., 2005, for a discussion). The reason for this is that
here the unobserved heterogeneity is absorbed by the factors, the dimension of which can be consistently estimated. That is, our GMM estimator is obtained conditional on \( \hat{r} \), which is again as good as knowing \( r_0 \) itself. To appreciate this point, suppose that (6) holds with \( T = 2, d_{i,t} = \eta_i \) and \( \gamma_i = 0_{m \times 1} \), which is (1) and (2) with \( f_i = 1 \) and \( \lambda_i = \eta_i(1 - \rho_0) \). As a result, there is no “genuine” factor structure. It follows that

\[
y_{i,t} = \eta_i(1 - \rho_0) + \rho_0 y_{i,t-1} + \epsilon_{i,t}.
\] (20)

The Arellano and Bond (1991) “first-differenced” GMM estimator of this model involves taking first-differences in order to eliminate \( \eta_i(1 - \rho_0) \);

\[
\Delta y_{i,2} = \rho_0 \Delta y_{i,1} + \Delta \epsilon_{i,2}.
\] (21)

When \( \rho_0 = 1 \), however, \( \eta_i(1 - \rho_0) = 0 \) and therefore first-differencing results in a lack of identification. Indeed, in the \( T = 2 \) case considered here, there is a single valid moment condition in (21), which is given by \( E(y_{i,0} \Delta \epsilon_{i,2}) = 0 \). Identification of \( \rho_0 \) requires \( E(y_{i,0} \Delta y_{i,1}) \neq 0 \). But if \( \rho_0 = 1 \), then

\[
E(y_{i,0} \Delta y_{i,1}) = E(y_{i,0}[(\rho_0 - 1)y_{i,0} + \eta_i(1 - \rho_0) + \epsilon_{i,1}]) = E(y_{i,0} \epsilon_{i,1}) = 0,
\]

and so \( \rho_0 \) is unidentified. However, notice that under \( \rho_0 = 1 \) the conventional pooled OLS estimator of (20) (without common factors) is still consistent, as it is based on a valid restriction \( (\eta_i(1 - \rho_0) = 0) \), although it becomes inconsistent for any other value of \( |\rho_0| < 1 \).

The discussion above suggests that it is desirable to have an estimator that is consistent regardless of the value taken by \( \rho_0 \). To see how such an estimator may be devised, let us consider the proposed GMM estimator, but without exploiting the restrictions described in (11) for \( G \), which are irrelevant for this argument. The appropriate moment conditions have the following form:

\[
\begin{bmatrix}
m_{01} \\
m_{02} \\
m_{12}
\end{bmatrix}
- \rho
\begin{bmatrix}
m_{00} \\
m_{01} \\
m_{11}
\end{bmatrix}
- g_0(1 - \rho_0)
\begin{bmatrix}
1 \\
1 \\
0
\end{bmatrix}
- g_1(1 - \rho_0)
\begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix}
= 0_{3 \times 1}.
\] (22)

where \( m_{st} = E(y_{i,s}y_{i,t}) \) and \( g_s = E(y_{i,s} \eta_i) \). These moment conditions actually correspond to those employed by Arellano and Bond (1991). The idea here is that the term involving \( g_0(1 - \rho_0) \) can be removed by simply subtracting the first row of (22) from the second, yielding\(^9\)

\[
(m_{02} - m_{01}) - \rho_0(m_{01} - m_{00}) = E[y_{i,0}(\Delta y_{i,2} - \rho \Delta y_{i,1})] = 0.
\] (23)

\(^9\)The last row in (22) can be ignored because it is a moment condition that is consumed to estimate an extra parameter that does not appear elsewhere, namely, \( g_1(1 - \rho) \).
Identification of $\rho_0$ based on (22) therefore requires $E(y_{i,0}\Delta y_{i,1}) \neq 0$, which is again violated if $\rho_0 = 1$. In terms of the general DGP considered in this paper, $\rho_0 = 1$ causes a violation of the full rank requirement of Assumption 6 (d). In order to see this, note that $m_{01} = E(y_{i,0}y_{i,1}) = E[y_{i,0}(\rho_0 y_{i,0} + \eta_i(1 - \rho_0) + \varepsilon_{i,1})] = E(y_{i,0}^2)$, implying

$$\Delta = - \begin{bmatrix} m_{00} & 1 & 0 \\ m_{01} & 1 & 0 \\ m_{11} & 0 & 1 \end{bmatrix} = - \begin{bmatrix} E(y_{i,0}^2) & 1 & 0 \\ E(y_{i,0}^2) & 1 & 0 \\ E(y_{i,1}^2) & 0 & 1 \end{bmatrix},$$

which has rank two only. But while GMM based on (22) cannot identify $\rho_0 = 1$, the corresponding restricted GMM estimator based on setting $\eta_i(1 - \rho_0) = 0$ can. But $\eta_i(1 - \rho_0) = 0$ is equivalent to setting $r_0 = 0$. Hence, provided that $r_0$ is consistently estimated, the proposed GMM estimator is consistent regardless of the value of $\rho_0$.

**Remark 8.** It is instructive to compare the approach considered here with the one of Bond et al. (2005), in which the popular “system” GMM estimator is employed to test for a unit root. The system GMM estimator is equivalent to the GMM estimator that results from imposing $g_0 = g_1$ on the moment conditions in (22). In the fixed effects example considered in (20) this restriction holds true, because

$$g_1 = E(y_{i,t} \eta_i) = E[(\rho y_{i,0} + (1 - \rho) \eta_i + \varepsilon_{i,1}) \eta_i] = \rho E(y_{i,0} \eta_i) + (1 - \rho) E(\eta_i^2)$$

$$= \rho E(\eta_i^2) + (1 - \rho) E(\eta_i^2) = E(\eta_i^2) = g_0.$$

The effect of this restriction can be appreciated by realizing that (22) reduces to

$$\begin{bmatrix} m_{01} \\ m_{02} \\ m_{12} \end{bmatrix} - \rho_0 \begin{bmatrix} m_{00} \\ m_{01} \\ m_{11} \end{bmatrix} - g_0(1 - \rho_0) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0_{3 \times 1},$$

with

$$\Delta = - \begin{bmatrix} E(y_{i,0}^2) & 1 \\ E(y_{i,0}^2) & 1 \\ E(y_{i,1}^2) & 1 \end{bmatrix}.$$
over $t$, which of course need not be the case in practice (see Bun and Sarafidis, 2015, for a detailed discussion). For example, if the model contains incidental trends, mean stationarity is obviously violated, causing the system GMM estimator to become invalid. Secondly, system GMM is not equipped to handle the presence of genuine factors in the error term.

4 Monte Carlo simulations

4.1 Design

Our DGP that we consider is of the same form as the one in (6), where we consider both the fixed effects ($d_{i,t} = \eta_i$) and incidental trends ($d_{i,t} = \eta_i + \beta_i t$) cases. Thus, while in the former (latter) case, under the unit root null hypothesis $y_{i,t}$ is a random walk without (with) a drift, under the alternative hypothesis, $y_{i,t}$ is stationary (trend-stationary). We also consider a “hybrid DGP”, in which half of the cross-sectional units are generated with fixed effects, while the other half are generated with incidental trends. Under the null, $\rho_0 = 1$, while under the alternative, $\rho_0 \in \{0.95, 0.99\}$. In both cases, $\eta_i \sim \text{iid } N(0, \sigma^2_{\eta,i})$ and $\beta_i \sim \text{iid } N(0, \sigma^2_{\beta,i})$, where $\sigma^2_{\eta,i} \sim \text{iid } U[0, 2]$ and $\sigma^2_{\beta,i} \sim \text{iid } U[0, 2]$. As for $\varepsilon_{i,t}$, we consider $\varepsilon_{i,t} \sim \text{iid } N(0, \sigma^2_{\varepsilon,i})$, where $\sigma^2_{\varepsilon,i} \sim \text{iid } U[0, 2]$, and $\varepsilon_{i,t} = v_{i,t} + 0.7 \cdot v_{i,t-1}$, $v_{i,t} \sim N(0, 1/(1 + q^2) \sigma^2_{v,i})$ for $t \geq 0$ and $\sigma^2_{v,i} \sim \text{iid } U[0, 2]$. Hence, not only is there substantial cross-sectional heterogeneity in the DGP, but we also allow $\varepsilon_{i,t}$ to be heteroskedastic and correlated across time. The initial observation is generated as $y_{i,0} = u_{i,0} = d_{i,0} + \gamma_i^t w_t + \varepsilon_{i,0}$, where $d_{i,0} = \eta_i$ in both the fixed effects and incidental trends cases and $\varepsilon_{i,0} \sim \text{iid } N(0, 1)$. We consider two values for $m$, the dimension of $w_t$, namely $m = 0$ and $m = 1$. Hence, while under the former parametrization, $\gamma_i^t w_t = 0$, under the latter, $\gamma_i^t w_t = \gamma_i^* w_t$, where $\gamma_i$ and $w_t$ are scalars satisfying $\gamma_i \sim \text{iid } N(0, \sigma^2_{\gamma,i})$, $w_t \sim \text{iid } N(0, 1)$ and $\sigma^2_{\gamma,i} \sim \text{iid } U[0, 2]$. Hence, in this DGP, $r_0 \in \{0, 1, 2, 3\}$; $r_0 = 0$ in the fixed effects case with $m = 0$ and $\rho_0 = 1$, $r_0 = 1$ in the fixed effects (incidental trends) case with $m = 0$ and $\rho_0 \leq 1$ ($\rho_0 = 1$), $r_0 = 2$ in the fixed effects (incidental trends) case with $\rho_0 < 1$ and $m = 1$ ($m = 0$), and $r_0 = 3$ in the incidental trends case with $\rho_0 \leq 1$ and $m = 1$.

We set $N \in \{100, 400, 1600\}$ and the number of effective time periods, $T^*$, is set to $T^* \in \{5, 7, 9, 11\}$.

All experiments are based on 5,000 replications.

\footnote{In absence of serial correlation in $\varepsilon_{i,t}$, $T^* = T$, whereas when $\varepsilon_{i,t}$ follows an MA process, $T^* = T + 1$. For the incidental trend and hybrid DGPs, we set the minimum value of $T^*$ equal to 6, in order to have enough moment conditions to be able to make inference based on the overidentifying restrictions $I$-statistic.}
4.2 Results

The following results are reported: (i) the mean of $\hat{\rho}$; (ii) the standard deviation (STD) of $\hat{\rho}$; (iii) the empirical rejection frequency of $t_\rho(1)$ at the nominal 5% level (for $\rho_0 \neq 1$ we report size-adjusted rejection frequency); (iv) the empirical rejection frequency of the overidentifying restrictions $J$-statistic (again at the nominal 5% level). To speed up the calculations, all results are based on setting $r_{\text{max}} = r_0 + 1$ with the estimation of $r_0$ carried out as explained in the previous section, using the $BIC(r)$ criterion.

Table 1 contains the results for the case when $m = 0$ and $T^* = 5$. The full set of moment conditions equals $M = 15$ in this case, noting that when $\varepsilon_{it}$ follows an MA(1) process only instruments up to $y_{it-2}$ are valid. As a rule of thumb, when $N/M \geq 10$ the results reported are based on the two-step estimator, whereas when $N/M < 10$ the results are based on the one-step estimator. This is in order to account for the well-known result that for small $N$ and large $M$, the size of two-step GMM-based test statistics tends to be distorted.

The performance of the proposed estimator and the corresponding $t_\rho(1)$-statistic is more than satisfactory. In particular, although for $N$ small there appears to be some small bias, the bias goes away as $N$ increases. Moreover, the size of the $t_\rho(1)$-statistic is close to the nominal level in all experiments considered. The power of the $t_\rho(1)$-statistic is satisfactory as well, and for $\rho_0 = .95$ it approaches unity rather quickly as $N$ increases.

In terms of STD, the results indicate that the estimator performs best in the fixed effects case, followed by the incidental trends case, and then the hybrid case. This is to be expected since the fixed effects-only model implies more degrees of freedom when compared to the other models.

The results reported in Table 2 for the case when $m = 1$ are very similar to those reported in Table 1 for $m = 0$. A noticeable difference is the dispersion of the estimator, which is much larger in Table 2 than in Table 1. This is due to the fact that there are more parameters to estimate with essentially the same amount of information.

Table 3 reports the empirical rejection frequency for several values of $T^*$. We see that, as long as $N$ is not too small, size remains close to the nominal level in all cases. On the other
hand, power appears to increase quite substantially with higher values of $T^*$. For instance, in
the fixed effects-only model when $m = 0$, $N = 1600$ and $\rho_0 = .99$, power increases from .589
when $T^* = 5$ to .993 when $T^* = 11$. The corresponding increase in power in the incidental
trends model is from .208 to .777.

Table 4 reports the power of the overidentifying restrictions $J$-statistic for the case when $\varepsilon_{i,t}$
follows an MA(1) process. This is an important empirical scenario, as serial correlation in $\varepsilon_{i,t}$
invalidates a subset of the proposed moment conditions. As we can see, in all cases considered
the power of the test is high.

5 Application

5.1 Gibrat’s Law

In this section we make use of our methodology in order to examine the empirical validity
of the well known “Law of Proportionate Effect”, or simply Gibrat’s Law (Gibrat, 1931) using
data from the US banking industry. Gibrat’s Law postulates that the growth rate of firms is
independent of their initial size. Analytically, we have

$$\Delta \ln s_{i,t} = \delta + (\rho - 1) \ln s_{i,t-1} + u_{i,t}, \quad (25)$$

where $s_{i,t}$ denotes the size of firm $i$ at time $t$ and $u_{i,t}$ is an error term. For $\rho < 1$ larger firms
tend to grow at a lower rate compared to smaller firms, while for $\rho > 1$ the process is explosive
and growth rate is proportional to firm size. For $\rho = 1$ Gibrat’s law holds true because firms’
growth rate is independent of their initial size. This means that Gibrat’s Law can be examined
by testing for a unit root in $\ln s_{i,t}$.

Gibrat’s law has become very popular because it provides an explanation for what has been
identified as an empirical regularity where the distribution of firms’ size is often highly skewed
across several industries. In particular, many sectors are characterised by a log-normal distri-
bution with a larger number of small to medium scale firms and relatively few large firms; see
Steindl (1965). Simon and Bonini (1958) argue that under (approximate) constant returns to
scale it is natural to expect that the probability for a given firm to increase/decrease in size in
proportion to its existing size is the same, on average, for all firms in the industry that lie above
a critical minimum size value. On the other hand, some of the more recent empirical evidence
(see, for example, Sutton, 1997; Caves, 1998) appears to suggest that while Gibrat’s law tends to
be confirmed in small subsamples of well-established, mature, large firms, this is not always the
case for larger samples that include small and young firms, since the latter often have higher
growth rate than their larger counterparts.

As the discussion in the above paragraph suggests, the relation between firm size and
growth rate remains an open issue. In this section we therefore provide new evidence based
on our newly developed tests, which has a number of advantages when compared to existing
tests. First, the test is valid for all values of $\rho$, including $\rho > 1$. Second, unlike many other
tests, the test developed here can be implemented without knowing the elements in $f_t$. This
advantage is particularly relevant in the present context, because the size of the firm is likely to
depend on its age, a variable that is not in our sample. As an illustration, let us denote by $a_{i,t}$
the age of firm $i$, and let $\beta_i$ denote the impact of age on the size of the firm. Since age increases
at the same rate for all firms, we may write $\beta_i a_{i,t} = \beta_i (a_{i,0} + t) = \gamma_i + \beta_i t$, where $\gamma_i = \beta_i + a_{i,0}$
with $a_{i,0}$ being the age of the firm at the beginning of the sample. Hence, under said conditions,
the effect of age can be captured using two factors, one is a constant, while the other is a linear
time trend. Of course, a priori we cannot say for sure that age has an effect, and so it is more
convenient to treat age as an unobserved factor to be estimated from the data. Another possi-
bility is that the factors represent in part common shocks due to for example the global financial
crisis.

5.2 Data

The data set consists of a panel of $N = 4,022$ depository financial institutions, each one ob-
served over nine years, in particular during the period 2003-2011. These data have been col-
clected from the electronic database maintained by the Federal Deposit Insurance Corporation
(FDIC).\textsuperscript{11} Two measures of bank size are considered; (i) fixed assets (FA), and (ii) number of
employees (EMP). Both variables are transformed by taking logs and FA is deflated using the
GDP deflator.

5.3 Results

As starting values for the factors we consider ($\sqrt{T}$ times) the eigenvectors corresponding to
the largest $r$ eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} (y_i - ay_{i,-1}) (y_i - ay_{i,-1})' / N$, where $a = \{0, .1, .2, ..., 1\}$, as well as fixed effects, linear trends, and a large number of random initializa-
tions from the normal and uniform distributions. We fit a maximum of $r_{max} = 3$ factors and
use the $BIC(r)$ criterion to pick the most appropriate number, given that it passes the $J$ test at

\textsuperscript{11}See http://www.fdic.gov.
the 5% level. To gauge against possible serial correlation in the errors, the GMM approach is
implemented assuming MA(q) errors, where \( q \in \{1, 2\} \). If the model is misspecified, this is
likely to show up in the \( J \) test.

Table 5 reports results obtained based on the two-step GMM estimator. The results are very
similar for EMP and FA. In particular, the point estimate of \( \rho_0 \) is below unity and the unit root
null hypothesis is rejected even at the 1% level, suggesting that Gibrat’s Law is not supported
by the data. The null hypothesis of instrument validity/correct model specification is also not
rejected. For EMP (FA) the best fitting model according to \( BIC(r) \) has one (two) factors. Ocular
inspection reveals that in case of FA while the first factor resembles a trend, the second factor
has a less clear cut shape. This demonstrates the importance of allowing for nonlinear effects,
casting doubt on existing results based on fixed effects-only unit root tests. The estimated factor
for EMP is very similar to the first factor for FA. In fact, the correlation between the two factors
is 0.915.

Our results imply that during the sampling period investigated, Gibrat’s Law is violated
implying that the growth rate of financial institutions is negatively correlated to their initial
size, that is, smaller institutions appear to grow faster than their larger counterparts.

6 Conclusion

This paper develops a GMM-based approach that enables unit root testing in panels where \( N \)
is large and \( T \) is finite. The assumption that \( T \) finite makes our test suitable for both micro and
small-\( T \) macro panels. The DGP considered is very general and accommodates an unrestricted
trend function and cross-sectional dependence in the form of common factors. These allowances
make the new approach one of the most general around. Indeed, as far as we are aware, this is
the only fixed-\( T \) unit root test approach that can be applied in the presence of cross-sectional de-
pendence and/or a potentially non-linear trend function. The approach is also relatively simple
to implement. In particular, since deterministic terms are treated as additional common factors,
which are estimated, there is no need to model the deterministic part. Our results show that the
new GMM-based unit root test statistic is asymptotically invariant to both the true and fitted
deterministic trend function. Hence, unlike existing tests, with the new test there is no need for
any mean and/or variance correction factors that reflect the fitted deterministic specification.
The limiting distribution of the GMM \( t \)-statistic is normal and this holds true regardless of the
value of the AR coefficient, \( \rho_0 \). Hence, again unlike most existing tests, with this test there is no
discontinuity in the asymptotic distribution at unity. The asymptotic properties are verified in small samples using both simulated and raw data.
References


Appendix

Proof of Theorem 1

The proof of Theorem 1 follows directly from Theorems 2.6 and 3.4 in Newey and McFadden (1994). In particular, it is straightforward to establish that the conditions listed therein are fulfilled under our Assumptions 1–5. For example, continuity of the moment functions holds, because \( h_i(\theta) \) is essentially a vector of elementary functions of \( \theta \), which belongs to the interior of \( \Theta \). Moreover, condition (iv) of Theorem 2.6 in Newey and McFadden (1994) holds, because of the compactness of \( \Theta \) and the fact that \( y_{i,t} \) has finite moments up to second order.

Derivatives

In this section we make heavy use of the results of Abadir and Magnus (2005). Readers are referred to this book for a more detailed treatment of the arguments used here.

Letting

\[
U(\theta_1, \theta_2, \theta_3) = S(I_T \otimes (\Gamma e_1 g_0' F' + \Gamma \bar{F}_1 F')) e,
\]

we have

\[
h_i(\theta) = Z'y_i - \rho Z'y_{i,-1} - U(\theta_1, \theta_2, \theta_3),
\]

\[
\bar{h}_N(\theta) = \hat{m} - \rho \hat{m}_1 - U(\theta_1, \theta_2, \theta_3).
\]

Clearly,

\[
U(\theta_1, \theta_2, \theta_3) = \text{vec}(U(\theta)) = (e' \otimes S) \text{vec}(I_T \otimes (\Gamma e_1 g_0' F' + \Gamma \bar{F}_1 F'))
\]

\[
= (e' \otimes S)(I_T \otimes K_{T,T} \otimes I_T) [\text{vec}(I_T) \otimes \text{vec}(\Gamma e_1 g_0' F' + \Gamma \bar{F}_1 F')]
\]

\[
= (e' \otimes S)(I_T \otimes K_{T,T} \otimes I_T) [\text{vec}(I_T) \otimes I_{T^2} \text{vec}(\Gamma e_1 g_0' F' + \Gamma \bar{F}_1 F')]
\]

\[
= (e' \otimes S)(I_T \otimes K_{T,T} \otimes I_T) [\text{vec}(I_T) \otimes I_{T^2} [1 \otimes \text{vec}(\Gamma e_1 g_0' F' + \Gamma \bar{F}_1 F')]]
\]

\[
= (e' \otimes S)(I_T \otimes K_{T,T} \otimes I_T) (e \otimes I_{T^2}) \text{vec}(\Gamma e_1 g_0' F' + \Gamma \bar{F}_1 F')
\]

\[
= A \text{vec}(\Gamma e_1 g_0' F' + \Gamma \bar{F}_1 F'),
\]

where \( A = (e' \otimes S)(I_T \otimes K_{T,T} \otimes I_T) (e \otimes I_{T^2}) \), a \( M \times T^2 \) matrix. Here, \( K_{kn,kn} \) is the \( kn \times kn \) commutation matrix of zeroes and ones such \( K_{kn,kn} \text{vec} \mathbf{H}^T = \text{vec} \mathbf{H} \) for any \( k \times n \) matrix \( \mathbf{H} \). As
a result,

\[
D_{\theta_1} U(\theta_1, \theta_2, \theta_3) = A[(F g_0 e_1' + \tilde{F} F_{-1}) \otimes I_T] D_{\rho} \Gamma(\rho);
\]

\[
D_{\theta_2} U(\theta_1, \theta_2, \theta_3) = A[I_T \otimes (\Gamma e_1 g_0' + \tilde{F} F_{-1}) + (F \otimes \Gamma) K_{T, r}(B \otimes I_r)];
\]

\[
D_{\theta_3} U(\theta_1, \theta_2, \theta_3) = A(F \otimes \Gamma e_1),
\]

where

\[
D_{\rho} \Gamma(\rho) = \text{vec}
\begin{bmatrix}
0 & 0 & \ldots & 0 & 0 \\
1 & 0 & \ldots & 0 & 0 \\
2\rho & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & 0 \\
(T - 1)\rho^{T-2} & (T - 2)\rho^{T-3} & 1 & 0
\end{bmatrix},
\]

and \(B\) is a \(T \times T\) matrix with zeros everywhere except in the first lower-diagonal that takes the value of one. We can therefore show that

\[
\nabla_{\theta} \bar{h}_N(\theta) = -[\bar{m}_1 + D_{\theta_1} U(\theta_1, \theta_2, \theta_3), D_{\theta_2} U(\theta_1, \theta_2, \theta_3), D_{\theta_3} U(\theta_1, \theta_2, \theta_3)].
\]
Table 1: Simulation results for the case when $m = 0$ and $T^* = 5$.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$N$</th>
<th>Fixed effects</th>
<th></th>
<th>Incidental trends</th>
<th></th>
<th>Hybrid</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>STD</td>
<td>$t_{\hat{\rho}}(1)$</td>
<td>$J$</td>
<td>Mean</td>
<td>STD</td>
<td>$t_{\hat{\rho}}(1)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\epsilon_{i,t}$ is serially uncorrelated</td>
<td></td>
<td>$\epsilon_{i,t}$ follows an MA(1) process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>.997</td>
<td>.021</td>
<td>.065</td>
<td>.059</td>
<td>.989</td>
<td>.054</td>
</tr>
<tr>
<td>1</td>
<td>400</td>
<td>1.00</td>
<td>.008</td>
<td>.069</td>
<td>.045</td>
<td>1.00</td>
<td>.025</td>
</tr>
<tr>
<td>1</td>
<td>1600</td>
<td>1.00</td>
<td>.003</td>
<td>.051</td>
<td>.047</td>
<td>1.00</td>
<td>.009</td>
</tr>
<tr>
<td>.99</td>
<td>100</td>
<td>.982</td>
<td>.046</td>
<td>.097</td>
<td>.043</td>
<td>.978</td>
<td>.073</td>
</tr>
<tr>
<td>.99</td>
<td>400</td>
<td>.987</td>
<td>.026</td>
<td>.203</td>
<td>.044</td>
<td>.987</td>
<td>.019</td>
</tr>
<tr>
<td>.99</td>
<td>1600</td>
<td>.989</td>
<td>.009</td>
<td>.589</td>
<td>.049</td>
<td>.989</td>
<td>.011</td>
</tr>
<tr>
<td>.95</td>
<td>100</td>
<td>.958</td>
<td>.026</td>
<td>.354</td>
<td>.071</td>
<td>.938</td>
<td>.072</td>
</tr>
<tr>
<td>.95</td>
<td>400</td>
<td>.956</td>
<td>.016</td>
<td>.967</td>
<td>.067</td>
<td>.948</td>
<td>.024</td>
</tr>
<tr>
<td>.95</td>
<td>1600</td>
<td>.953</td>
<td>.012</td>
<td>1.00</td>
<td>.061</td>
<td>.950</td>
<td>.010</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>.996</td>
<td>.025</td>
<td>.062</td>
<td>.043</td>
<td>.991</td>
<td>.057</td>
</tr>
<tr>
<td>1</td>
<td>400</td>
<td>1.00</td>
<td>.011</td>
<td>.067</td>
<td>.045</td>
<td>1.00</td>
<td>.029</td>
</tr>
<tr>
<td>1</td>
<td>1600</td>
<td>1.00</td>
<td>.003</td>
<td>.049</td>
<td>.048</td>
<td>1.00</td>
<td>.011</td>
</tr>
<tr>
<td>.99</td>
<td>100</td>
<td>.980</td>
<td>.051</td>
<td>.103</td>
<td>.042</td>
<td>.977</td>
<td>.077</td>
</tr>
<tr>
<td>.99</td>
<td>400</td>
<td>.986</td>
<td>.022</td>
<td>.194</td>
<td>.049</td>
<td>.988</td>
<td>.022</td>
</tr>
<tr>
<td>.99</td>
<td>1600</td>
<td>.987</td>
<td>.010</td>
<td>.569</td>
<td>.046</td>
<td>.988</td>
<td>.012</td>
</tr>
<tr>
<td>.95</td>
<td>100</td>
<td>.960</td>
<td>.030</td>
<td>.337</td>
<td>.044</td>
<td>.941</td>
<td>.076</td>
</tr>
<tr>
<td>.95</td>
<td>400</td>
<td>.956</td>
<td>.019</td>
<td>.946</td>
<td>.053</td>
<td>.947</td>
<td>.023</td>
</tr>
<tr>
<td>.95</td>
<td>1600</td>
<td>.954</td>
<td>.013</td>
<td>1.00</td>
<td>.051</td>
<td>.949</td>
<td>.011</td>
</tr>
</tbody>
</table>

Notes: “Mean” and “STD” refer to mean and standard deviation of $\hat{\rho}$, respectively. The results reported for the $J$-statistic are the empirical sizes. In the “hybrid” model, half of the units are generated with fixed effects, while the other half is generated with incidental trends.
Table 2: Simulation results for the case when $m = 1$ and $T^* = 6$.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$N$</th>
<th>Fixed effects</th>
<th>Incidental trends</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\bar{\rho}$</td>
<td>STD</td>
<td>$t_{\bar{\rho}}$</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>.995</td>
<td>.049</td>
<td>.043</td>
</tr>
<tr>
<td>1</td>
<td>400</td>
<td>1.00</td>
<td>.020</td>
<td>.065</td>
</tr>
<tr>
<td>1</td>
<td>1600</td>
<td>1.00</td>
<td>.006</td>
<td>.060</td>
</tr>
<tr>
<td>.99</td>
<td>100</td>
<td>.988</td>
<td>.051</td>
<td>.056</td>
</tr>
<tr>
<td>.99</td>
<td>400</td>
<td>.993</td>
<td>.020</td>
<td>.092</td>
</tr>
<tr>
<td>.99</td>
<td>1600</td>
<td>.993</td>
<td>.008</td>
<td>.307</td>
</tr>
<tr>
<td>.95</td>
<td>100</td>
<td>.961</td>
<td>.062</td>
<td>.227</td>
</tr>
<tr>
<td>.95</td>
<td>400</td>
<td>.964</td>
<td>.037</td>
<td>.798</td>
</tr>
<tr>
<td>.95</td>
<td>1600</td>
<td>.952</td>
<td>.013</td>
<td>.937</td>
</tr>
</tbody>
</table>

$\varepsilon_{i,t}$ is serially uncorrelated

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$N$</th>
<th>Fixed effects</th>
<th>Incidental trends</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\bar{\rho}$</td>
<td>STD</td>
<td>$t_{\bar{\rho}}$</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>.992</td>
<td>.053</td>
<td>.042</td>
</tr>
<tr>
<td>1</td>
<td>400</td>
<td>.996</td>
<td>.027</td>
<td>.057</td>
</tr>
<tr>
<td>1</td>
<td>1600</td>
<td>1.00</td>
<td>.008</td>
<td>.055</td>
</tr>
<tr>
<td>.99</td>
<td>100</td>
<td>.984</td>
<td>.054</td>
<td>.053</td>
</tr>
<tr>
<td>.99</td>
<td>400</td>
<td>.987</td>
<td>.027</td>
<td>.089</td>
</tr>
<tr>
<td>.95</td>
<td>100</td>
<td>.964</td>
<td>.069</td>
<td>.212</td>
</tr>
<tr>
<td>.95</td>
<td>400</td>
<td>.959</td>
<td>.043</td>
<td>.763</td>
</tr>
<tr>
<td>.95</td>
<td>1600</td>
<td>.948</td>
<td>.017</td>
<td>.894</td>
</tr>
</tbody>
</table>

Notes: See Table 1.
Table 3: Rejection frequencies for different values of $T^*$ when $\epsilon_{i,t}$ is serially uncorrelated.

<table>
<thead>
<tr>
<th>$\rho_0$</th>
<th>$N$</th>
<th>$T^* = 7$</th>
<th>$T^* = 9$</th>
<th>$T^* = 11$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>TR</td>
<td>HY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m = 0</td>
<td>m = 0</td>
<td>m = 0</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>.051</td>
<td>.050</td>
<td>.053</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.047</td>
<td>.084</td>
<td>.079</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>.050</td>
<td>.053</td>
<td>.057</td>
</tr>
<tr>
<td>.99</td>
<td>100</td>
<td>.081</td>
<td>.051</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.198</td>
<td>.102</td>
<td>.112</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>.857</td>
<td>.193</td>
<td>.186</td>
</tr>
<tr>
<td>.95</td>
<td>100</td>
<td>.441</td>
<td>.217</td>
<td>.224</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.836</td>
<td>.754</td>
<td>.768</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>1.00</td>
<td>.713</td>
<td>.736</td>
</tr>
<tr>
<td>m = 1</td>
<td>100</td>
<td>.038</td>
<td>.068</td>
<td>.057</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.050</td>
<td>.078</td>
<td>.066</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>.054</td>
<td>.075</td>
<td>.061</td>
</tr>
<tr>
<td>.99</td>
<td>100</td>
<td>.067</td>
<td>.056</td>
<td>.068</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.110</td>
<td>.094</td>
<td>.102</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>.819</td>
<td>.173</td>
<td>.193</td>
</tr>
<tr>
<td>.95</td>
<td>100</td>
<td>.701</td>
<td>.343</td>
<td>.368</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.983</td>
<td>.374</td>
<td>.381</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>1.00</td>
<td>.978</td>
<td>.998</td>
</tr>
</tbody>
</table>

Notes: “FE”, “TR” and “HY” refer to the fixed effects, the incidental trends and hybrid models, respectively.

Table 4: Power of the $J$-statistic when $\epsilon_{i,t}$ follows an MA(1) process.

<table>
<thead>
<tr>
<th>$\rho_0$</th>
<th>$N$</th>
<th>$\rho_0 = 1$</th>
<th>$\rho_0 = .99$</th>
<th>$\rho_0 = .95$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>TR</td>
<td>HY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m = 0</td>
<td>m = 0</td>
<td>m = 0</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>.983</td>
<td>.841</td>
<td>.843</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>1.00</td>
<td>1.00</td>
<td>.994</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.99</td>
<td>100</td>
<td>.857</td>
<td>.794</td>
<td>.782</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>.957</td>
<td>.938</td>
<td>.895</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Notes: The results for $m = 0$ ($m = 1$) are based on $T^* = 5$ ($T^* = 7$).
Table 5: Empirical results.

<table>
<thead>
<tr>
<th>Measure</th>
<th>$\hat{\rho}$</th>
<th>$\hat{\sigma}_\rho$</th>
<th>$t_\rho(1)$</th>
<th>p-value</th>
<th>$J$</th>
<th>p-value</th>
<th>BIC1</th>
<th>$\hat{r}$</th>
<th>$q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMP</td>
<td>.866</td>
<td>.001</td>
<td>-126.1</td>
<td>.000</td>
<td>22.08</td>
<td>.228</td>
<td>-57.9</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FA</td>
<td>.822</td>
<td>.001</td>
<td>-128.0</td>
<td>.000</td>
<td>3.89</td>
<td>.918</td>
<td>-36.1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: “$\hat{\rho}$” and “$\hat{\sigma}_\rho$” refer to the two-step GMM estimator of $\rho_0$ and its estimated standard error, “$t_\rho(1)$” refers to the unit root t-statistic, “$J$” refers to the Hansen–Sargan statistic, “BIC1” refers to the minimizing value of the BIC1, “$\hat{r}$” refers to the estimated number of factors using the BIC1, and “$q$” refers to the order of the assumed MA errors.