GMM Unit Root Inference in Generally Trending and Cross-Correlated Dynamic Panels

Robertson, Donald and Sarafidis, Vasilis and Westerlund, Joakim

Monash University

5 February 2014

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

Donald Robertson
University of Cambridge
UK

Vasilis Sarafidis†
Monash University
Australia

Joakim Westerlund
Deakin University
Australia

February 3, 2014

Abstract

This paper proposes a new panel unit root test based on the generalized method of moments approach for panels with a small number of time periods and a large number of cross-section units, N. In the model that we consider the deterministic trend function is essentially unrestricted and the errors are cross-sectionally correlated in a very general fashion. In spite of these allowances, the GMM-statistic is shown to be asymptotically unbiased, $\sqrt{N}$-consistent and asymptotically normal for all values of the autoregressive (AR) coefficient, ρ, including unity, making it an ideal candidate for unit root inference. Results from both simulated and real data are provided to suggest that the asymptotic properties are borne out well in small samples.

JEL Classification: C12; C13; C33; C36.

Keywords: Panel data; unit root test; cross-section dependence; common factors; GMM.

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 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, $T \to \infty$

*The authors would like to thank seminar participants at Monash University for useful comments and suggestions.

†Corresponding author. E-mail: vasilis.sarafidis@monash.edu.

such that $N/T \to \infty$. The main reason for this is the presence of cross-section heterogeneity, such as fixed effects, whose estimation requires $T \to \infty$. This induces an estimation error in T, which can only be controlled if $N/T \to 0$, for otherwise the accumulated effect as $N \to \infty$ will be unbounded (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 (see, for example, De Wachter et al., 2007; Hlouskova and Wagner, 2006), 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, pose a real restriction, to the point that researchers might well find themselves discarding data in order to have N sufficiently small relative to T. Moreover, in many panels, such as those frequently encountered in applied micro, 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).

The above issue has motivated researchers to look for inferential procedures that are suitable in fixed-T panels. Harris and Tzavalis (1999) proposed a panel unit root test based on the bias-corrected ordinary least squares (OLS) estimator of the autoregressive (AR) coefficient, ρ. Many other tests have since then been proposed (see De Blander and Dhaene, 2012, and the references provided therein).[1] The evidence reported so far (see, for example, Harris and Tzavalis, 1999; Hadri and Larsson, 2005; Hlouskova and Wagner, 2006) suggests that in terms of small-sample performance, not requiring T to be large can be a great advantage. 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 still plenty of important issues remaining unresolved in the fixed-T literature. First, except for Harris and Tzavalis (2004) and Han and Phillips (2010), who consider the case with 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 (in a potentially non-linear fashion), then the probability of the panel of multiple time series exhibiting at least some trending behavior will tend to one as $N \to \infty$, in which case fixed effects-only tests will be rendered invalid. Second, as far as we are aware, there is presently no test that is able to accommodate cross-section dependence, that is, existing fixed-T tests are "first-generation" tests (Baltagi, 2008, Chapter 12). This is again noteworthy because in practice such dependence is likely to be the rule rather than the exception, even in highly disaggregated data,

[1]See Breitung and Pesaran (2008) for a survey of the panel unit root and cointegration literatures.

because of herd behavior, fashions or fads.

The current paper addresses both issues. We develop a "second-generation" approach to unit roots in fixed-T panels characterized by both cross-section dependence and generally trending behavior. This is accomplished by assuming that the data admits to a common factor structure in which the 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 regulatory conditions. It also provides a means 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.

The estimation is carried out by modifying the generalized method of moments (GMM) approach of Robertson and Sarafidis (2013). The new estimator is shown to have a number of desirable properties. First, it is free from the otherwise common incidental parameter bias. This is true not only in the conventional fixed effects case, but also in the more general interactive effects model considered here. The reason for this is that we only require consistent estimation of the covariance matrix of the factor loadings, and not of the loadings themselves, thereby eliminating the incidental parameter problem. Second, the estimator supports asymptotically normal inference for all values of ρ, including unity, and the well-known weak instruments problem when ρ is in the vicinity of unity does not emerge (see, for example, Bun and Windmeijer, 2010). Hence, unlike most existing approaches, the limiting distribution of the GMM estimator considered here is continuous and has the same rate of consistency as ρ passes through unity (see Phillips and Han, 2010, for a similar result). Third, the estimator and the associated t-statistic for a unit root have excellent 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 small-sample accuracy of the asymptotic results are evaluated using both simulated and real data in Sections 4 and 5, respectively. Section 6 concludes.

2 Model and assumptions

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

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

$$u_{i,t} = \boldsymbol{\lambda}_i' \mathbf{f}_t + \varepsilon_{i,t}, \quad (2)$$

where $\rho \in \mathbb{R}$, $\mathbf{f}_t$ is an $r \times 1$ vector of common factors with $\boldsymbol{\lambda}_i$ being the associated vector of factor loadings, and $\varepsilon_{i,t}$ is an idiosyncratic error term. The following assumptions are made, where $\mathcal{F}_t$ is the sigma-field generated by $\{\varepsilon_{i,n}\}_{n=1}^{t}$, $\mathbf{1}_T = (1, ..., 1)'$ is a $T \times 1$ vector, and tr $\mathbf{A}$ and $||\mathbf{A}|| = \sqrt{\text{tr}\,(\mathbf{A}'\mathbf{A})}$ denote the trace and Frobenius (Euclidean) norm of the matrix $\mathbf{A}$, respectively.

Assumption ERR. $\varepsilon_{i,t}$ is independent across i with $E(\varepsilon_{i,t}|\mathcal{F}_{t-1}) = 0$, $\sum_{i=1}^{N} E(\varepsilon_{i,t}^2)/N \to \sigma_\varepsilon^2 > 0$ and $E(\varepsilon_{i,t}^4) < \infty$.

Assumption LAM. $\boldsymbol{\lambda}_i$ is a random coefficient vector such that $\sum_{i=1}^{N} E(\boldsymbol{\lambda}_i \boldsymbol{\lambda}_i')/N \to \boldsymbol{\Sigma}_\lambda$, an $r \times r$ positive definite matrix, $E(||\boldsymbol{\lambda}_i||^8) < \infty$, and $E(\boldsymbol{\lambda}_i \varepsilon_{j,t}) = \mathbf{0}_{r\times 1}$ for all i, j and t.

Assumption F.

(i) $\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$ is a non-random $T \times r$ matrix with full column rank;

(ii) Suppose that $\rho = 1$. If $\mathbf{1}_T$ is included in $\mathbf{F}$, then $r > 1$; otherwise, $r > 0$.

Assumption INI. $y_{i,0} = \boldsymbol{\lambda}_i' \mathbf{f}_0 + \varepsilon_{i,0}$, where $||\mathbf{f}_0|| < \infty$ and $\varepsilon_{i,0}$ is independent across i with $E(\varepsilon_{i,0}) = 0$, $\sum_{i=1}^{N} E(\varepsilon_{i,0}^2)/N \to \sigma_0^2 > 0$, $E(\varepsilon_{i,0}^4) < \infty$, and $E(\boldsymbol{\lambda}_i \varepsilon_{j,0}) = \mathbf{0}_{r\times 1}$ for all i and j.

Assumption MOM. $T(T+1)/2 > 1 + r\,[T + (1-r)\,/2]$.

Assumption EPS allows for cross-section and time series heteroskedasticity but implies no serial correlation in $\varepsilon_{i,t}$, the latter of which is a very common restriction in the fixed-T literature (see, for example, Bun and Sarafidis, 2013). One way to allow for more general forms of serial correlation is to consider more lags of $y_{i,t}$, such that the AR(1) in (1) becomes an AR(p) model (with $p \geq 1$). Another possibility is to put the serial correlation in $\varepsilon_{i,t}$ and to change the choice of instruments. This is discussed in detail in Remark 3, and then again in Section 5, where we show how to implement our approach in the presence of moving average (MA) errors. The random loading assumption can be relaxed in a relatively straightforward way, provided that $||\boldsymbol{\lambda}_i||^8 < \infty$ and $\lim_{N\to\infty} \sum_{i=1}^{N} \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i'/N$ is positive definite. Assumption F (i) is standard in panel data models with T fixed (see Sarafidis and Wansbeek, 2012). We note that the $\mathbf{F}$ could also be treated as stochastic by modifying the proofs accordingly. This would not change anything else that is of substance in the paper. As we explain in Remark 2, Assumption F (ii) is needed for identification of ρ in the unit root case. Assumption INI implies that the equation for $y_{i,0}$ can be thought of as the reduced form equation for $y_{i,t}$ at period time $t = 0$. Thus, $\mathbf{f}_0$ and $\varepsilon_{i,0}$ are not necessarily identical to the values that would arise had $y_{i,0}$ been assumed to follow (1). Assumption MOM is important because the number of factors needs to be small enough relative to the number of moment conditions, such that there are enough degrees of freedom to estimate the model.

A major difference when compared to the existing large-T second-generation panel unit root literature (see, for example, Bai and Ng, 2004; 2010; Moon and Perron, 2004) is that here $\mathbf{f}_t$ is treated as a fixed parameter vector to be estimated along with the other parameters of the model. Whether $\mathbf{f}_t$ has zero mean is therefore not an issue. It can also have arbitrary "dynamics". In terms of the terminology of Bai (2009), (1) and (2) constitute a fixed interactive effects model, which is more general than the models considered previously in the literature. Suppose, for example, that $\mathbf{f}_t = (1, \tau_t)'$ and $\boldsymbol{\lambda}_i = (\eta_i, 1)'$, such that $\boldsymbol{\lambda}_i'\mathbf{f}_t = \eta_i + \tau_t$. This means that the DGP reduces to

$$y_{i,t} = \rho y_{i,t-1} + \eta_i + \tau_t + \varepsilon_{i,t}.$$

This is the benchmark first-generation specification with incidental intercepts and time-specific fixed effects to account for cross-section dependence (see, for example, Im et al., 2003; Levin et al., 2002). Models with incidental trends and second-generation models with common factors can also be accommodated. For example, if $\mathbf{f}_t = (1, t, \mathbf{g}_t')'$, where $\mathbf{g}_t$ is an $(r-2) \times 1$ vector of common factors, and $\boldsymbol{\lambda}_i = (\eta_i, \beta_i, \boldsymbol{\delta}_i')'$, then

$$y_{i,t} = \rho y_{i,t-1} + \eta_i + \beta_i t + \boldsymbol{\delta}_i'\mathbf{g}_t + \varepsilon_{i,t}.$$

This specification is similar to those considered by, for example, Moon and Perron (2004), Pesaran (2007), and Phillips and Sul (2003) in the large-T case, and we will consider it again in our Monte Carlo study in Section 4. Note that while incidental trends can be allowed, this is by no means a restriction; the interactive effects model considered here can accommodate virtually any trend function that is linear in parameters, including polynomial trend functions, trigonometric functions and models of discrete and smooth structural shifts. The model considered for $\mathbf{f}_t$ is therefore very general indeed.

Moreover, while not necessary, the elements of $\mathbf{f}_t$ may be unknown. As we illustrate in Section 5, this means that the researcher is spared from the problem of having to decide on which deterministic components to include. For example, if structural shifts are present, then there is no need for any a priori knowledge regarding their locations, which are obtained as part of the estimation process. Hence, not only is the model very general, but the way that $\mathbf{f}_t$ is accommodated is also very convenient from an empirical point of view.

Remark 1. In this paper we assume that $\mathbf{f}_t$ enters via $u_{i,t}$. This is not necessary. As Bai and Ng (2010) discuss, when $\mathbf{f}_t$ is random a more general DGP is obtained by placing the common component directly under $y_{i,t}$, such that $y_{i,t} = \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{i,t}$, and then allow $\mathbf{f}_t$ and $u_{i,t}$ to have different dynamics. However, since in this paper $\mathbf{f}_t$ is fixed, the dynamics are driven by the idiosyncratic component only, and from this point of view it does not matter whether $\mathbf{f}_t$ enters via $u_{i,t}$ or $y_{i,t}$.

3 Main results

3.1 Moment conditions

Define the $T \times T$ matrix

$$\mathbf{M} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{M}_i,$$

where $\mathbf{M}_i = E(\mathbf{y}_i^0\mathbf{y}_i')$, $\mathbf{y}_i^0 = (y_{i,0}, ...y_{i,T-1})'$ and $\mathbf{y}_i = (y_{i,1}, ...y_{i,T})'$ are both $T \times 1$. We begin by deriving an expression for $\mathbf{M}$ in terms of the parameters of the DGP. This will then be used as a basis for formulating our moment conditions. Let us begin by defining the $T \times T$ lag matrix

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}, \tag{3}$$

and the $T \times 1$ vector $\mathbf{e}_t = (0, ..., 0, 1, 0, ..., 0)'$, where the one is at position t. The model for $y_{i,t}$ can now be written in vector form as

$$\mathbf{y}_i = \rho \mathbf{e}_1 y_{i,0} + \rho \mathbf{L} \mathbf{y}_i + \mathbf{F} \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i, \tag{4}$$

or

$$\mathbf{y}_i = \rho \boldsymbol{\Gamma} \mathbf{e}_1 y_{i,0} + \boldsymbol{\Gamma} \mathbf{F} \boldsymbol{\lambda}_i + \boldsymbol{\Gamma} \boldsymbol{\varepsilon}_i, \tag{5}$$

where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, ..., \varepsilon_{i,T})'$ is $T \times 1$,

$$\boldsymbol{\Gamma} = (\mathbf{I}_T - \rho \mathbf{L})^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \rho^{T-1} & \dots & \rho & 1 \end{bmatrix},$$

and recall that $\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_T)'$ is $T \times r$. Note that $\boldsymbol{\Gamma}$ is a function of ρ. In order to emphasize this, whenever appropriate we write $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\rho)$. By using Assumption INI to substitute for $y_{i,0}$ in (5), and then stacking $y_{i,0}$ and $\mathbf{y}_i$, we obtain

$$\begin{bmatrix} y_{i,0} \\ \mathbf{y}_i \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}_{1 \times T} \\ \rho \boldsymbol{\Gamma} \mathbf{e}_1 & \boldsymbol{\Gamma} \end{bmatrix} \begin{bmatrix} \mathbf{f}_0' \\ \mathbf{F} \end{bmatrix} \boldsymbol{\lambda}_i + \begin{bmatrix} 1 & \mathbf{0}_{1 \times T} \\ \rho \boldsymbol{\Gamma} \mathbf{e}_1 & \boldsymbol{\Gamma} \end{bmatrix} \begin{bmatrix} \varepsilon_{i,0} \\ \boldsymbol{\varepsilon}_i \end{bmatrix}.$$

For later use we also define $\mathbf{F}^+ = (\mathbf{f}_0, \mathbf{F}')'$, a $(T+1) \times r$ matrix. Note that, since $\rho \boldsymbol{\Gamma} \mathbf{e}_1 = (\rho, \rho^2, ..., \rho^T)'$, the matrix

$$\begin{bmatrix} 1 & \mathbf{0}_{1 \times T} \\ \rho \boldsymbol{\Gamma} \mathbf{e}_1 & \boldsymbol{\Gamma} \end{bmatrix}$$

is of the same form as $\boldsymbol{\Gamma}$, but of dimension $(T+1) \times (T+1)$ instead of $T \times T$. Hence, letting $\boldsymbol{\varepsilon}_i^0 = (\varepsilon_{i,0}, ..., \varepsilon_{i,T-1})'$ and $\mathbf{F}^0 = (\mathbf{f}_0, ..., \mathbf{f}_{T-1})'$, the following expression for $\mathbf{y}_i^0$ is obtained:

$$\mathbf{y}_i^0 = \boldsymbol{\Gamma} \mathbf{F}^0 \boldsymbol{\lambda}_i + \boldsymbol{\Gamma} \boldsymbol{\varepsilon}_i^0. \tag{6}$$

Note also that in this notation, $\mathbf{e}_1 y_{i,0} + \mathbf{L} \mathbf{y}_i = \mathbf{y}_i^0$, suggesting that (4) may be written as

$$\mathbf{y}_i = \rho \mathbf{y}_i^0 + \mathbf{F} \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i. \tag{7}$$

Let us now consider $\mathbf{M}$. Substitution of (7) yields

$$\mathbf{M}_i = E(\mathbf{y}_i^0 \mathbf{y}_i') = E[\mathbf{y}_i^0 (\rho \mathbf{y}_i^0 + \mathbf{F} \boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i)'] = \rho \mathbf{M}_{i,-1} + E(\mathbf{y}_i^0 \boldsymbol{\lambda}_i') \mathbf{F}' + E(\mathbf{y}_i^0 \boldsymbol{\varepsilon}_i'),$$

where $\mathbf{M}_{i,-1} = E[\mathbf{y}_i^0 (\mathbf{y}_i^0)']$. By using, in turn, (6) to substitute for $\mathbf{y}_i^0$, and then the fact that $\boldsymbol{\varepsilon}_i^0$ and $\boldsymbol{\lambda}_i$ are uncorrelated, $E(\mathbf{y}_i^0 \boldsymbol{\lambda}_i')$ may be written as

$$E(\mathbf{y}_i^0 \boldsymbol{\lambda}_i') = E[(\boldsymbol{\Gamma} \mathbf{F}_0 \boldsymbol{\lambda}_i + \boldsymbol{\Gamma} \boldsymbol{\varepsilon}_i^0) \boldsymbol{\lambda}_i'] = \boldsymbol{\Gamma} \mathbf{F}^0 E(\boldsymbol{\lambda}_i \boldsymbol{\lambda}_i') = \boldsymbol{\Gamma} \mathbf{F}_0 \boldsymbol{\Sigma}_\lambda. \tag{8}$$

Moreover, since $E(y_{i,s}\varepsilon_{i,t}) = 0$ for all $s \leq t-1$ and $t = 1, ..., T$, we have

$$E(\mathbf{y}_i^0 \boldsymbol{\varepsilon}_i') = \begin{bmatrix} 0 & 0 & \dots & 0 \\ E(y_{i,1}\varepsilon_{i,1}) & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ E(y_{T-1}\varepsilon_{i,1}) & \dots & E(y_{T-1}\varepsilon_{i,T-1}) & 0 \end{bmatrix}.$$

This matrix contain $T(T+1)/2$ zeroes. These are our moment conditions. A natural way of writing these conditions is as follows:

$$\text{vech}\,(\mathbf{M}') - \rho\text{vech}\,(\mathbf{M}'_{-1}) - \text{vech}\,[\mathbf{F}\boldsymbol{\Sigma}_\lambda(\mathbf{F}^0)'\boldsymbol{\Gamma}'] = \mathbf{0}_{T(T+1)/2\times 1}, \tag{9}$$

where $\mathbf{M}_{-1}$ is defined similarly to $\mathbf{M}$, and vech is the half-vec operator that when applied to a matrix $\mathbf{A}$ eliminates all supradiagonal elements of $\mathbf{A}$ from vec $\mathbf{A}$.

Unfortunately, the formulation in (9) is not very convenient to work with. Let us therefore denote by $\mathbf{S}_t$ ($t = 1, ..., T$) the $M_t \times T$ selection matrix of zeroes and ones that picks out the t entries of $\mathbf{y}_i^0$ which are valid at period t, that is, $\mathbf{S}_t$ is such that $E[(\mathbf{S}_t\mathbf{y}_i^0)\varepsilon_{i,t}] = \mathbf{0}_{t\times 1}$. Under Assumption EPS, we have $\mathbf{S}_t = (\mathbf{I}_t, \mathbf{0}_{t\times(T-t)})$, a $t \times T$ matrix. Hence, at time t the vector of valid instruments is given by $(y_{i,0}, ..., y_{i,t-1})'$. Define the $T(T+1)/2 \times T^2$ matrix $\mathbf{S} = \text{diag}(\mathbf{S}_1, ..., \mathbf{S}_T)$. The matrix of instruments can now be written as

$$\mathbf{Z}_i' = \mathbf{S}(\mathbf{I}_T \otimes \mathbf{y}_i^0), \tag{10}$$

which is $T(T+1)/2 \times T$. Note that $\mathbf{F}\boldsymbol{\lambda}_i = (\mathbf{I}_T \otimes \boldsymbol{\lambda}_i')\text{vec}\,(\mathbf{F}')$. By using this result, premultiplication of (7) by $\mathbf{Z}_i'$, and then taking expectations, we obtain

$$E(\mathbf{Z}_i'\mathbf{y}_i) = \rho E(\mathbf{Z}_i'\mathbf{y}_i^0) + E[\mathbf{Z}_i'(\mathbf{I}_T \otimes \boldsymbol{\lambda}_i')]\text{vec}\,(\mathbf{F}').$$

Combining (8) with (10), and using the fact that vec $(\mathbf{F}') = (\mathbf{I}_T \otimes \mathbf{F}')\mathbf{e}$, the moment condition in (9) can be written alternatively as

$$\mathbf{m} - \rho\mathbf{m}^0 - \mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e} = \mathbf{0}_{T(T+1)/2\times 1}, \tag{11}$$

where $\mathbf{m} = E(\mathbf{Z}_i'\mathbf{y}_i) = \text{vech}\,(\mathbf{M}')$ and $\mathbf{m}^0 = E(\mathbf{Z}_i'\mathbf{y}_i^0) = \text{vech}\,(\mathbf{M}'_{-1})$.

Remark 2. The moment condition in (11) hold for all values of ρ, provided that it is finite. This is in contrast to many of the existing GMM estimators of dynamic panel data models (such as those considered by, for example, Anderson and Hsiao, 1981; Arellano and Bond, 1991), which are known to suffer from a weak instrument problem when $\rho \approx 1$ (see Blundell and Bond, 1998). To appreciate this, suppose that $T = 3$, $\rho \in [0, 1]$, $r = 1$ and that $\mathbf{f}_t = f_t$ is known. In this case, $\mathbf{f} = (f_1, f_2, f_3)'$, $\mathbf{f}^0 = (f_0, f_1, f_2)'$ and

$$\mathbf{M} = \begin{bmatrix} m_{01} & m_{11} & m_{21} \\ m_{02} & m_{12} & m_{22} \\ m_{03} & m_{13} & m_{23} \end{bmatrix}, \qquad \mathbf{M}_{-1} = \begin{bmatrix} m_{00} & m_{10} & m_{20} \\ m_{01} & m_{11} & m_{21} \\ m_{03} & m_{12} & m_{22} \end{bmatrix},$$

suggesting that (11) can be written as

$$\begin{bmatrix} m_{01} \\ m_{02} \\ m_{12} \\ m_{03} \\ m_{13} \\ m_{23} \end{bmatrix} - \rho \begin{bmatrix} m_{00} \\ m_{01} \\ m_{11} \\ m_{02} \\ m_{12} \\ m_{22} \end{bmatrix} - \sigma_\lambda^2 \begin{bmatrix} \sum_{j=0}^{0} \rho^j f_j f_1 \\ \sum_{j=0}^{0} \rho^j f_j f_2 \\ \sum_{j=0}^{1} \rho^j f_j f_2 \\ \sum_{j=0}^{0} \rho^j f_j f_3 \\ \sum_{j=0}^{1} \rho^j f_j f_3 \\ \sum_{j=0}^{2} \rho^j f_j f_3 \end{bmatrix}. \tag{12}$$

Making use of the fact that $y_{i,t} = \lambda_i \sum_{j=1}^{t} \rho^j f_{t-j} + \sum_{j=1}^{t} \rho^j \varepsilon_{i,t-j}$, we can show that with $t \geq s$

$$m_{st} = E(y_{i,s} y_{i,t}) = \sigma_\lambda^2 \sum_{j=1}^{s} \sum_{n=1}^{t} \rho^{j+n} f_{s-j} f_{t-n} + \sigma_\varepsilon^2 \sum_{j=1}^{s} \rho^{t+s-2(j-1)},$$

where σ_λ^2 is the scalar version of $\mathbf{\Sigma}_\lambda$ and the second term on the right-hand side is equal to $\rho^{t-s}(1 - \rho^{2(s+1)})/(1 - \rho^2)$ for $\rho < 1$ and $(s + 1)$ for $\rho = 1$. It follows that, regardless of the value of ρ, there is enough variation across the rows in (12) to identify the unknown parameters, ρ and σ_λ^2. In the fixed effects case, however, $f_t = 1$ and therefore $\sum_{j=1}^{s} \sum_{n=1}^{t} \rho^{j+n} f_{s-j} f_{t-n}$ reduces to $(1 - \rho^t)(1 - \rho^s)/(1 - \rho)^2$ for $\rho < 1$ and $(s + 1)(t + 1)$ for $\rho = 1$. Thus, in this case ρ and σ_λ^2 are identified only for $\rho < 1$, because when $\rho = 1$ all rows in (12) become linear combinations of the first row, and so there is effectively a single informative moment condition based on which it is not possible to identify two parameters. In practice, fixed effects only are rather restrictive and, as acknowledged in the panel unit roots literature, unlikely to be able to capture all unobserved heterogeneity in the data (see, for example, Baltagi, 2008, Chapter 12).

Remark 3. The moment conditions in (11) can be modified to allow for error serial correlation, or an AR(p) for $y_{i,t}$. The case with MA errors is particularly easy. Suppose that $\varepsilon_{i,t}$ follows a MA(q) process. This case can be accommodated by setting $\mathbf{Z}_i' = \mathbf{S}(\mathbf{I}_T \otimes \mathbf{y}_i^q)$, where $\mathbf{y}_i^q = (y_{i,0}, ..., y_{i,T-1-q})'$ and the t-th diagonal element of $\mathbf{S}$ is given by $\mathbf{S}_t = (\mathbf{I}_t, \mathbf{0}_{t \times (T-t-q)})$, whose dimension is $t \times (T - q)$. In Section 5 we show how to implement our approach when

$q \in \{1,2\}$. Consider next the case when $\varepsilon_{i,t}$ is serially uncorrelated but that $y_{i,t}$ 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 can be written in vector form as

$$\mathbf{y}_i = \rho_1 \mathbf{\Gamma} \mathbf{e}_1 y_{i,0} + \rho_2 \mathbf{\Gamma}(\mathbf{e}_1 y_{i,-1} + \mathbf{e}_2 y_{i,0}) + \mathbf{\Gamma} \mathbf{F} \boldsymbol{\lambda}_i + \mathbf{\Gamma} \boldsymbol{\varepsilon}_i,$$

where $\mathbf{\Gamma} = (\mathbf{I}_T - \rho_1 \mathbf{L} - \rho_2 \mathbf{L}^2)^{-1}$ and $\mathbf{L}^2 = \mathbf{L}\mathbf{L}$. Alternatively, since $\mathbf{e}_1 y_{i,0} + \mathbf{L}\mathbf{y}_i = \mathbf{y}_i^0$, with $\mathbf{y}_i^{-1} = \mathbf{e}_1 y_{i,-1} + \mathbf{e}_2 y_{i,0} + \mathbf{L}^2 \mathbf{y}_i$,

$$\mathbf{y}_i = \rho_1 \mathbf{y}_i^0 + \rho_2 \mathbf{y}_i^{-1} + \mathbf{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i.$$

In this case, the matrix of instruments, $\mathbf{Z}_i'$, is still given by (10) but with $\mathbf{y}_i^0$ replaced by $(y_{i,-1}, y_{i,0}, ..., y_{i,T-1})'$ and $\mathbf{S}_t = (\mathbf{I}_{t+1}, \mathbf{0}_{(t+1)\times(T-t)})$, a $(t+1) \times (T+1)$ matrix. Thus, pre-multiplying the expression above by $\mathbf{Z}_i'$, taking expectations, and then rearranging yields

$$\mathbf{m} - \rho_1 \mathbf{m}^0 - \rho_2 \mathbf{m}^{-1} - \mathbf{S}(\mathbf{I}_T \otimes \mathbf{\Gamma} \mathbf{F}^{-1} \mathbf{\Sigma}_\lambda \mathbf{F}')\mathbf{e} = \mathbf{0}_{(T+T(T+1)/2)\times 1}, \tag{13}$$

where $\mathbf{m}^{-1} = E(\mathbf{Z}_i' \mathbf{y}_i^{-1})$ and $\mathbf{F}^{-1} = (\mathbf{f}_{-1}, \mathbf{f}_0, ..., \mathbf{f}_{T-1})'$.

3.2 Inference when $\mathbf{F}^+$ is known

Define $\boldsymbol{\theta} = [\rho, (\text{vech}\, \mathbf{\Sigma}_\lambda)']' = (\theta_1, \boldsymbol{\theta}_2')'$ and denote by $\boldsymbol{\theta}_0$ and ρ_0 the true values of $\boldsymbol{\theta}$ and ρ, respectively. The GMM estimator of this parameter vector, whose dimension is $(1 + r(r+1)/2) \times 1$, is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\theta \in \mathcal{C}_\theta} Q(\boldsymbol{\theta}),$$

where $\mathcal{C}_\theta$ is a compact subset of $\mathbb{R}^{1+r(r+1)/2}$ and

$$\begin{aligned} Q(\boldsymbol{\theta}) &= \mathbf{g}(\boldsymbol{\theta})' \mathbf{g}(\boldsymbol{\theta}), \\ g(\boldsymbol{\theta}) &= \mathbf{W}^{1/2}[\hat{\mathbf{m}} - \rho \hat{\mathbf{m}}^0 - \mathbf{S}(\mathbf{I}_T \otimes \mathbf{\Gamma} \mathbf{F}^0 \mathbf{\Sigma}_\lambda \mathbf{F}')\mathbf{e}], \end{aligned}$$

with $\hat{\mathbf{m}} = \sum_{i=1}^N \mathbf{Z}_i' \mathbf{y}_i / N$, $\hat{\mathbf{m}}^0 = \sum_{i=1}^N \mathbf{Z}_i' \mathbf{y}_i^0 / N$, and $\mathbf{W} = (\mathbf{W}^{1/2})' \mathbf{W}^{1/2}$ is a $T(T+1)/2 \times T(T+1)/2$ positive definite weight matrix. $\hat{\boldsymbol{\theta}}$ is the joint GMM estimator of θ_1 and $\boldsymbol{\theta}_2$. However, since we are primarily interested in $\theta_1 = \rho$, in this section we will consider a concentrated objective function, as is commonly done in the maximum likelihood literature. The GMM estimator of $\boldsymbol{\theta}_2$ given ρ is derived in (A6) and is given by

$$\hat{\boldsymbol{\theta}}_2 = \hat{\boldsymbol{\theta}}_2(\rho) = [\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)'\mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0),$$

where $\mathbf{R}(\rho) = \mathbf{W}^{1/2}(\mathbf{e}' \otimes \mathbf{S})(\mathbf{I}_T \otimes \mathbf{K}_{T^2} \otimes \mathbf{I}_T)[\text{vec}\,\mathbf{I}_T \otimes (\mathbf{I}_T \otimes \boldsymbol{\Gamma}(\rho))(\mathbf{F} \otimes \mathbf{F}^0)\mathbf{D}_r]$, $\mathbf{K}_{T^2}$ is the $T^2 \times T^2$ commutation matrix such that $\mathbf{K}_{mp}\text{vec}\,\mathbf{A} = \text{vec}\,\mathbf{A}'$, where $\mathbf{A}$ is $m \times p$, and $\mathbf{D}_r$ is the $r^2 \times r(r+1)/2$ duplication matrix such that $\text{vec}\,\mathbf{B} = \mathbf{D}_r\,\text{vech}\,\mathbf{B}$, where $\mathbf{B}$ is $r \times r$. The concentrated objective function, henceforth denoted as $Q_C(\rho)$, is obtained by replacing $\boldsymbol{\theta}_2$ in $\mathbf{g}(\boldsymbol{\theta})$ with $\hat{\boldsymbol{\theta}}_2$, that is,

$$Q_C(\rho) = \mathbf{g}_C(\rho)'\mathbf{g}_C(\rho),$$

where $\mathbf{g}_C(\rho) = \mathbf{g}(\rho, \hat{\boldsymbol{\theta}}_2(\rho))$. The estimator of ρ is given by

$$\hat{\rho} = \arg\min_{\rho \in \mathcal{C}_\rho} Q_C(\rho). \tag{14}$$

where $\mathcal{C}_\rho$ is a compact subset of $\mathbb{R}$.

Assumption IDE.

(i) $\rho_0 \in \mathcal{C}_\rho$;

(ii) $\mathbf{g}(\rho)$ is continuous in ρ and $E[\mathbf{g}(\rho)] = \mathbf{0}_{(1+r(r+1)/2)\times 1}$ implies $\rho = \rho_0$;

(iii) $\boldsymbol{\Sigma}_{g_C} = \lim_{N\to\infty} NE[\mathbf{g}_C(\rho_0)\mathbf{g}_C(\rho_0)']$ and $\gamma_1 = \lim_{N\to\infty} \partial\mathbf{g}_C(\rho_0)/\partial\rho$ are finite with $\boldsymbol{\Sigma}_{g_C}$ positive definite.

Theorem 1 provides the asymptotic distribution of $\hat{\rho}$.

Theorem 1. *Under Assumptions EPS, LAM, F, INI, MOM and IDE, as $N \to \infty$*

$$\sqrt{N}(\hat{\rho} - \rho_0) \xrightarrow{d} N(0, \sigma_\rho^2),$$

where $\xrightarrow{d}$ signifies convergence in distribution and $\sigma_\rho^2 = \gamma_1'\boldsymbol{\Sigma}_{g_C}\gamma_1/(\gamma_1'\gamma_1)^2$. Analytical expressions for $\boldsymbol{\Sigma}_{g_C}$ and γ_1 are given in Appendix.

Remark 4. According to Theorem 1 there is no asymptotic bias, despite the generality of the DGP considered; $(\hat{\rho} - \rho_0)$ is centered at zero even when scaled by $\sqrt{N}$. The reason for this is that the GMM approach considered here only requires an estimator of $\boldsymbol{\theta}_2 = \text{vech}\,\boldsymbol{\Sigma}_\lambda$; there is no need to estimate $\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_N$ themselves. This means that the number of parameters that needs to be estimated is substantially reduced, from Nr to $r(r+1)/2$, thereby eliminating the incidental parameter problem (see Bai, 2013, for a similar approach).

Remark 5. Theorem 1 holds for all values of ρ_0, and in this sense it presents a unified asymptotic result for the GMM estimator. This is in contrast to the existing literature, in which the asymptotic distribution of estimators depends critically on whether $|\rho_0| < 1$, $\rho_0 = 1$ or indeed $\rho_0 > 1$. 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 $\rho_0 > 1$.

An analytical expression for γ_1 is given in the Appendix. Define $\hat{\sigma}_\rho^2 = \hat{\gamma}_1' \hat{\boldsymbol{\Sigma}}_{gC} \hat{\gamma}_1 / (\hat{\gamma}_1' \hat{\gamma}_1)^2$, where $\hat{\boldsymbol{\Sigma}}_{gC} = N \mathbf{g}_C(\hat{\rho}) \mathbf{g}_C(\hat{\rho})'$ and $\hat{\gamma}_1 = \partial \mathbf{g}_C(\hat{\rho}) / \partial \rho$. The GMM t-statistic for testing $H_0 : \rho_0 = \rho^0$ is given by

$$t(\rho^0) = \frac{\sqrt{N}(\hat{\rho} - \rho^0)}{\hat{\sigma}_\rho},$$

and is the same regardless of the value of ρ^0. The local power of $t(\rho^0)$ is easily worked out using Theorem 1. Indeed, suppose that

$$\rho_0 = \rho^0 + \frac{c}{\sqrt{N}}, \tag{15}$$

where $c \in \mathbb{R}$, such that ρ_0 is local to ρ^0, the hypothesized value under the null. In this case, since $\hat{\sigma}_\rho^2 = \sigma_\rho^2 + o_p(1)$, we can show that

$$t(\rho^0) = \frac{\sqrt{N}(\hat{\rho} - \rho_0)}{\hat{\sigma}_\rho} + \frac{\sqrt{N}(\rho_0 - \rho^0)}{\hat{\sigma}_\rho} \xrightarrow{d} N(0, 1) + \frac{c}{\sigma_\rho}$$

as $N \to \infty$. Summarizing this, we have the following corollary to Theorem 1.

Corollary 1. *Under (15) and the conditions of Theorem 1, as $N \to \infty$*

$$t(\rho^0) \xrightarrow{d} N(0, 1) + \frac{c}{\sigma_\rho}.$$

Remark 6. Corollary 1 nests the asymptotic results under both the null and the local alternative hypotheses. On the one hand, if $c = 0$, then H_0 is true and therefore $t(\rho^0) \xrightarrow{d} N(0, 1)$. If, on the other hand, $c \neq 0$, such that the local alternative is true, then the asymptotic distribution of $t(\rho^0)$ has no longer mean at zero, and therefore the test is unbiased, as well as consistent under the local alternative when $|c/\sigma_\rho| \to \infty$. The extent of power is driven by two parameters, c and σ_ρ^2; the smaller the uncertainty regarding ρ_0 and the larger the deviation from the hypothesized value of ρ^0, the larger the power, as expected (see Madsen, 2010, for a similar finding for some existing tests). What is unexpected, however, is the fact that the appropriate rate of shrinking of the local alternative is the same regardless of the specification of the deterministic trend part of $\mathbf{f}_t$ (see Han and Phillips, 2010, Section 5.2, for a similar discussion).

Remark 7. The asymptotic distribution of most (if not all) unit root 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). Corollary 1 shows how the GMM-based t-statistic has the unique and practically very useful property that it is asymptotically invariant to $\mathbf{F}$, and hence to any trend function that it may contain. The standard fixed effects assumption is therefore not needed and the otherwise so common mean and variance correction factors reflecting the chosen deterministic specification can be completely avoided.

In (A7) in Appendix we show that

$$\mathbf{g}_C(\rho) = \mathbf{P}(\rho)\mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0),$$

where $\mathbf{P}(\rho) = \mathbf{I}_{T(T+1)/2} - \mathbf{R}(\rho)[\mathbf{R}(\rho)\mathbf{R}(\rho)']^{-1}\mathbf{R}(\rho)'$. This formulation of $\mathbf{g}_C(\rho)$ suggests a simple estimation approach that can be used also when $\mathbf{F}$ is unknown. In particular, while nonlinear in ρ, with $\mathbf{W}$ and $\mathbf{F}^+$ known, $\mathbf{R}(\rho)$ does not depend on any other parameters that are unknown. Hence, assuming for a moment that also ρ is known, then so is $\mathbf{R} = \mathbf{R}(\rho)$, in which case $Q_C(\rho)$ is just the (weighted) sum of squared residuals, and therefore the GMM estimator $\hat{\rho}$ is just the (weighted) OLS slope estimator in a regression of $\mathbf{P}\mathbf{W}^{1/2}\hat{\mathbf{m}}$ onto $\mathbf{P}\mathbf{W}^{1/2}\hat{\mathbf{m}}^0$:

$$\hat{\rho} = \hat{\rho}(\mathbf{P}) = \frac{(\hat{\mathbf{m}}^0)'(\mathbf{W}^{1/2})'\mathbf{P}'\mathbf{P}\mathbf{W}^{1/2}\hat{\mathbf{m}}}{(\hat{\mathbf{m}}^0)'(\mathbf{W}^{1/2})'\mathbf{P}'\mathbf{P}\mathbf{W}^{1/2}\hat{\mathbf{m}}^0},$$

where $\mathbf{P} = \mathbf{P}(\rho_0)$. Although $\mathbf{P}$ is not observed when estimating ρ_0, and vice versa, we can replace the unobserved quantities by initial estimates and iterate until convergence. Suppose we are interested in testing $H_0 : \rho_0 = \rho^0$. A natural initialization for ρ in this case is given by $\rho = \rho^0$. The GMM estimator of $\boldsymbol{\theta}_2$ can then be obtained as $\hat{\boldsymbol{\theta}}_2(\hat{\rho})$.

3.3 Inference when $\mathbf{F}^+$ is unknown

For $\mathbf{F}^+$ is unknown we define $\boldsymbol{\theta} = [\rho, (\text{vech}\,\boldsymbol{\Sigma}_\lambda)', (\text{vec}\,\mathbf{F}^+)']' = (\theta_1, \boldsymbol{\theta}_2', \boldsymbol{\theta}_3')'$, which is $(1 + r(r+1)/2 + (T+1)r) \times 1$. The estimation of this parameter vector can also proceed in an iterative fashion, as before. The only difference is that since now $\mathbf{F}^+$ is unknown, even if ρ was known, $\mathbf{R}$ and hence also $\mathbf{P}$ would still be unknown. In order to emphasize this dependence on $\mathbf{F}^+$ we write $\mathbf{P}(\rho, \mathbf{F}^+)$ for $\mathbf{P}$. The estimator $\hat{\mathbf{F}}^+$ of $\mathbf{F}^+$ may be obtained as follows:

1. Initialize ρ.
2. The last T rows of $\hat{\mathbf{F}}^+$ can be obtained as the eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^N (\mathbf{y}_i - \rho \mathbf{y}_i^0)(\mathbf{y}_i - \rho \mathbf{y}_i^0)'/N$. The first row of $\hat{\mathbf{F}}^+$ can be obtained as the first observation of each of the eigenvectors corresponding to the r largest eigenvalues of $\sum_{i=1}^N y_{i,0}^2/N$. Write $\hat{\mathbf{F}}^+(\rho)$ for $\hat{\mathbf{F}}^+$.
3. The estimator of ρ is given by $\hat{\rho}(\hat{\mathbf{P}})$, where $\hat{\mathbf{P}} = \hat{\mathbf{P}}(\rho) = \mathbf{P}(\rho, \hat{\mathbf{F}}^+(\rho))$.
4. Update $\hat{\mathbf{F}}^+(\hat{\rho})$, $\hat{\mathbf{P}}(\hat{\rho})$ and $\hat{\rho}(\hat{\mathbf{P}})$. Repeat until convergence.

Remark 8. So far we have assumed that the number of factors, r, is known. However, the asymptotic results also hold when r is replaced by a consistent estimator, $\hat{r}$ say. Write $\hat{\rho}(\mathbf{P}, r)$ for $\hat{\rho}$. To see that $\hat{\rho}(\mathbf{P}, \hat{r})$ has the same asymptotic distribution as $\hat{\rho}(\mathbf{P}) = \hat{\rho}(\mathbf{P}, r)$, consider

$$\begin{aligned} P(\sqrt{N}[\hat{\rho}(\mathbf{P}, \hat{r}) - \rho_0] \leq \delta) &= P(\sqrt{N}[\hat{\rho}(\mathbf{P}, \hat{r}) - \rho_0] \leq \delta | \hat{r} = r) P(\hat{r} = r) \\ &+ P(\sqrt{N}[\hat{\rho}(\mathbf{P}, \hat{r}) - \rho_0] \leq \delta | \hat{r} \neq r) P(\hat{r} \neq r). \end{aligned}$$

Because $P(\hat{r} = r) \to 1$ and $P(\hat{r} \neq r) \to 0$, the second term on the right-hand side converges to zero, and $P(\sqrt{N}[\hat{\rho}(\mathbf{P}, \hat{r}) - \rho_0] \leq \delta) = 1 + o(1)$. Moreover, conditional on $\hat{r} = r$, $\hat{\rho}(\mathbf{P}, \hat{r}) = \hat{\rho}(\mathbf{P}, r)$. Thus,

$$|P(\sqrt{N}[\hat{\rho}(\mathbf{P}, \hat{r}) - \rho_0] \leq \delta) - P(\sqrt{N}[\hat{\rho}(\mathbf{P}, r) - \rho_0] \leq \delta)| \to 0.$$

Ahn et al. (2013) consider the problem of consistent estimation of r in the context of a static panel data regression with factors, and make several suggestions toward this end. It is conjectured that these estimators are consistent also in the present setup. In Section 4 we examine the performance of our GMM approach when combined with BIC1 information criterion of Ahn et al. (2013).

4 Monte Carlo simulations

4.1 Design

The DGP is given by a restricted version of (1) and (2). Two specifications of $\varepsilon_{i,t}$ are considered. In the first, $\varepsilon_{i,t} \sim N(0,1)$, whereas in the second, $\varepsilon_{i,t}$ is generated as an MA(1) process; $\varepsilon_{i,t} = v_{i,t} + \phi v_{i,t-1}$, where $v_{i,t} \sim N(0, 1/(1-\phi^2))$ and $\phi = 0.7$. The common component is specified with $\boldsymbol{\lambda}_i = [\eta_i, \beta_i(1-\rho_0), \boldsymbol{\delta}_i']'$ and $\mathbf{f}_t = (1, t, \mathbf{g}_t')'$, such that

$$y_{i,t} = \rho_0 y_{i,t-1} + \eta_i + \beta_i(1-\rho_0)t + \boldsymbol{\delta}_i'\mathbf{g}_t + \varepsilon_{i,t}.$$

Here $\mathbf{g}_t \sim N(\mathbf{0}_{r_0\times 1}, \mathbf{I}_{r_0})$ is an $r_0 \times 1$ vector of unobserved common factors with loading $\boldsymbol{\delta}_i \sim N(\mathbf{0}_{r_0\times 1}, \sigma_\delta^2 \mathbf{I}_{r_0})$, where $r_0 \in \{1,2\}$ and $\sigma_\delta^2 \in \{0.5, 2.5\}$. When $\sigma_\delta^2 = 0.5$ the proportion of the variance of $u_{i,t}$ that is due to variations in $\boldsymbol{\delta}_i'\mathbf{g}_t$ is 33%, whereas when $\sigma_\delta^2 = 2.5$ this proportion is 72%. The intercept and trend slope, η_i and β_i, are both drawn from $N(0,1)$. Hence, in this DGP, while under the unit root null ($\rho_0 = 1$), $y_{i,t}$ follows a random walk with drift, under the alternative that $\rho_0 \in \{0.95, 0.99\}$, $y_{i,t}$ is trend stationary.[2] Finally, $y_{i,0} = \eta_i + u_{i,0}$, where $u_{i,0} \sim N(0,1)$. We set $T = 8$ and $N \in \{100, 400, 1600\}$. All experiments are based on 2000 replications.

4.2 Results

Two versions of our estimator are simulated, both based on (14). The first, denoted GMM1, is the one-step estimator that makes use of $\mathbf{W} = \mathbf{I}_{T(T+1)/2}$, while the second, GMM2, makes use of the optimal weighting matrix, that is, $\mathbf{W} = \hat{\boldsymbol{\Sigma}}_{g_C}$. In both cases, the number of unknown factors, r_0, is estimated using the BIC1 criterion of Ahn et al. (2013), which is of the form BIC1 $= J -$ penalty, where J denotes the value of the Hansen–Sargan statistic for overidentifying restrictions and the exact form of the penalty is given in Ahn et al. (2013). The

[2] We also considered the case when $\boldsymbol{\lambda}_i'\mathbf{f}_t = \eta_i(1-\rho_0) + \boldsymbol{\delta}_i'\mathbf{g}_t$; however, since the results were very similar, we only report the results for the specification with the trend included.

maximum number of factors considered is set to $r_{max} = r_0 + 1$. The following results are reported: (i) mean, standard deviation (SD) and root mean squared error (RMSE) for GMM1 and GMM2; (ii) size (nominal size is 5%) and power of $t(1)$ for the unit root t-statistic; (iii) the 5% size of the J-statistic for overidentifying restrictions, which is only relevant for GMM2; (iv) the correct selection frequency for BIC1 (based on GMM2).

Table 1 contains the results for the case when $\varepsilon_{i,t} \sim N(0,1)$. In this case, since there is no error serial correlation, the full set of $T(T+1)/2 = 36$ moment conditions is used. It is seen that the performance of the estimators and their t-statistics is more than satisfactory. In particular, the bias is small and it get closer to zero as N increases. As expected, unless $N = 100$, GMM2 is more efficient than GMM1. We also see that the size of the J-statistic is close to the nominal 5% level. This is reflected in the results for BIC1, which is very accurate. In fact, the correct selection frequency does not fall below 90% except one instance.

The size of the $t(1)$-statistic is close to the nominal level in all experiments considered. The only exception is when $N = 100$, in which case the GMM2-based statistic is oversized; however, the distortions vanishes rapidly as N increases. The highest power is obtained by using the GMM2-based $t(1)$-statistic, which is to be expected given that GMM2 is relatively more efficient. Naturally, the power of both statistics increases as ρ_0 deviates from unity and as N increases.

Larger values of σ_δ^2 are generally associated with increased performance; GMM1 and GMM2 tend to become more accurate, and the size accuracy and power of $t(1)$ improves. This is because a larger σ_δ^2 will make the common component easier to discern. Larger values of r, on the other hand, tend to push the results in the other direction, that is, performance is decreasing in the number of unknown factors. This latter effect is in accordance with our expectations, as the number of parameters increases while the information contained in the data (number of moment conditions) stays the same.

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. The model is the same as in (1), where $y_{i,t}$ is now the size of firm i at time t in logs. It is instructive to rewrite this model as

$$\Delta y_{i,t} = (\rho_0 - 1)y_{i,t-1} + u_{i,t}.$$

For $\rho_0 < 1$ larger firms tend to grow at a lower rate compared to smaller firms, while for $\rho_0 > 1$ the process is explosive and growth rate is proportional to firm size. For $\rho_0 = 1$ Gibrat's law holds true because firms' growth rate is independent of their initial size. An advantage of our methodology is that it remains valid throughout the range of possible values of ρ_0, including $\rho_0 > 1$. Testing Gibrat's law is therefore tantamount to testing for a unit root in $y_{i,t}$.

Gibrat's law has proved 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 characterized by a log-normal distribution 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 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 (see Sutton, 1997; Caves, 1998). Given that the relation between firm size and growth rate remains an open issue, it is useful to investigate this using data from the US banking industry.

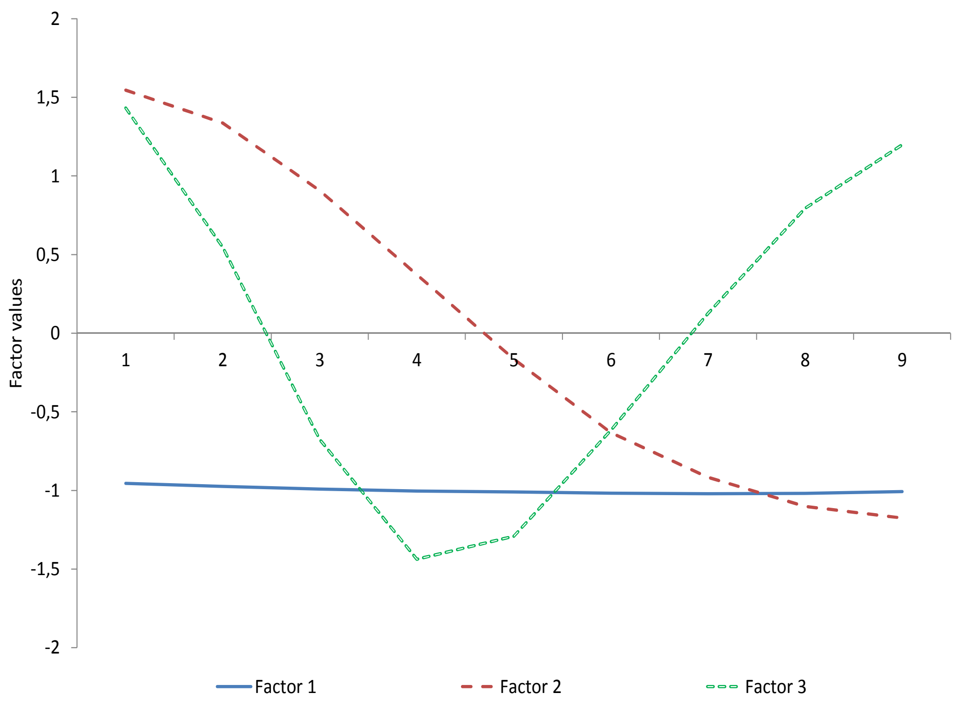
5.2 Data description and methodology

The data set consists of a panel of $N = 5593$ depository financial institutions, each one observed over a period of $T = 9$ years. These data have been collected from the electronic database maintained by the Federal Deposit Insurance Corporation (FDIC).[3] 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. In order

[3] See http://www.fdic.gov.

to account for common time effects, we further demean the variables with respect to their cross-section averages. Hence, in this application $y_{i,t}$ represents the demeaned log size of a firm.

Figure 1: Estimated factors for FA.



The factors are initiated by taking $\sqrt{T}$ times the eigenvectors corresponding to the largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i' / N$. To get a feeling for what the extracted factors look like, Figure 1 plots the values of the factors associated with the three largest eigenvalues for FA. The first factor is almost a straight line, suggesting that this factor is in fact capturing unit-specific fixed effects. The remaining two factors resemble a cubic line with a large smoothing parameter and a quadratic line. This demonstrates the importance of allowing for nonlinear effects, casting doubt on existing results based on fixed effects-only unit root tests. The factors for EMP are almost identical and are therefore not plotted. In fact, the correlation between the second (third) factor of the two variables is 0.99 (0.98). Hence, at least the common part of FA and EMP seems to be measuring the same thing.

In order to gauge against possible serial correlation in the errors, we implement our GMM approach assuming MA errors of order $q \in \{1, 2\}$. If the model is misspecified, this is likely to show up in the Hansen–Sargan test statistic. We fit a maximum of $r_{max} = 3$ factors and use the BIC1 criterion of Ahn et al. (2013) to pick the most appropriate number, given that it passes the Hansen–Sargan test at the 5% level.

5.3 Results

Table 3 reports results obtained based on the two-step GMM estimator, GMM2. The results are very similar for EMP and FA. In particular, the point estimate of ρ_0 is below unity, but the unit root null is not rejected even at the liberal 10% level, suggesting that Gibrat's law is supported by the data. The null hypothesis of instrument validity/correct model specification is also not rejected. For EMP the best fitting model according to BIC1 has two factors and allows for MA(1) errors, while for FA the preferred model includes one factor and MA(2) errors.

Our results are consistent with previous findings in the banking literature, which suggest the presence of constant returns to scale (see, for example, Robertson et al., 2013). As discussed by Simon and Bonini (1958) constant returns to scale corroborates Gibrat's law because in this case the probability of a given change in firm size (in proportion to current size) is likely to be the same for all firms in the industry.

6 Conclusion

This paper develops a GMM-based approach that enable 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 accommodate an unrestricted trend function and cross-section 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-section dependence and/or a potentially non-linear trend function. The approach is also very 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, ρ_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

Abadir, K. M., and J. R. Magnus (2005). *Matrix Algebra, Econometric Exercises 1*. Cambridge University Press, New York.

Anderson, T. W., and C. Hsiao (1981). Estimation of Dynamic Models with Error Components. *Journal of American Statistical Association* **76**, 598–606.

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

Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica* **77**, 1229–1279.

Bai, J. (2013). Fixed-Effects Dynamic Panel Models, a Factor Analytical Method. *Econometrica* **81**, 285–314.

Bai, J., and S. Ng (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica* **72**, 1127–1177.

Bai, J., and S. Ng (2010). Panel Unit Root Tests With Cross-Section Dependence: A Further Investigation. *Econometric Theory* **26**, 1088–1114.

Baltagi, B. (2008). *Econometric Analysis of Panel Data, Fourth Edition*. John Wiley & Sons, New York.

Blundell, R., and S. Bond (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* **87**, 115–143.

Bond, S., C. Nauges, and F. Windmeijer (2005). Unit Roots: Identification and Testing in Micro Panels. Unpublished manuscript.

Breitung, J., and H. Pesaran (2008). Unit Roots and Cointegration in Panels. In L. Matyas and P. Sevestre (ed.), *The Econometrics of Panel Data*.

Bun, M., and F. Windmeijer (2010). The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models. *Econometrics Journal* **13**, 95–126.

Bun, M. ,and V. Sarafidis (2013). Dynamic Panel Data Models. In B. H. Baltagi (ed.), *The Oxford Handbook of Panel Data*, Chapter 4.

Caves, R. E. (1998). Industrial Organization and New Findings on the Turnover and Mobility of Firms. *Journal of Economic Literature* **36**, 1947-1982.

Chesher, A. (1979). Testing the Law of Proportionate Effect. *Journal of Industrial Economics* **27**, 403–411.

De Blander, R., and G. Dhaene (2013). Unit Root Tests for Panel Data with AR(1) Errors and Small *T*. *Econometrics Journal* **15**, 101–124.

De Silva, S., K. Hadri and A. R. Tremayne (2009). Panel Unit Root Tests in the Presence of Cross-Sectional Dependence: Finite Sample Performance and an Application. *Econometrics Journal* **12**, 340–366.

De Wachter, S., R. Harris, and E. Tzavalis (2007). Panel Data Unit Roots Tests: The Role of Serial Correlation and the Time Dimension. *Journal of Statistical Planning and Inference 137*, 230–244.

Hadri, K., and R. Larsson (2005). Testing for Stationarity in Heterogeneous Panel Data where the Time Dimension is Finite. *Econometrics Journal* **8**, 55–69.

Han, C., and P. C. B. Phillips (2010). GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity. *Econometric Theory* **26**, 119–151.

Harris, R., and E. Tzavalis (1999). Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed. *Journal of Econometrics* **91**, 201–226.

Harris, R., and E. Tzavalis (2004). Testing for Unit Roots in Dynamic Panels in the Presence of a Deterministic Trend: Re-examining the Unit Root Hypothesis for Real Stock Prices and Dividends. *Econometric Reviews* **23**, 149–166.

Hlouskova, J., and M. Wagner (2006). The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation. *Econometric Reviews* **25**, 85–116.

Im, K. S., M. H. Pesaran and Y. Shin (2003). Testing for Unit Roots in Heterogeneous Panels. *Journal of Econometrics* **115**, 53–74.

Kruiniger, H. (2007). An Efficient Linear GMM Estimator for the Covariance Stationary AR(1)/Unit Root Model for Panel Data. *Econometric Theory* **23**, 519–535.

Kruiniger, H. (2009). GMM Estimation and Inference in Dynamic Panel Data Models with Persistent Data. *Econometric Theory* **25**, 1348–1391.

Kruiniger, H. (2013). Quasi ML Estimation of the Panel AR(1) Model with Arbitrary Initial Conditions. *Journal of Econometrics* **173**, 175–188.

Madsen, E. (2010). Unit Root Inference in Panel Data Models where the Time-Series Dimension is Fixed: A Comparison of Different Tests. *Econometrics Journal* **13**, 63–94.

Moon, H. R., and B. Perron (2004). Testing for Unit Root in Panels with Dynamic Factors. *Journal of Econometrics* **122**, 81–126.

Levin, A., C. Lin, and C.-J. Chu (2002). Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. *Journal of Econometrics* **108**, 1–24.

Pesaran, M. H. (2007). A Simple Panel Unit Root Test in the Presence of Cross Section Dependence. *Journal of Applied Econometrics* **22**, 265–312.

Phillips, P. C. B., and D. Sul (2003). Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence. *Econometrics Journal* **6**, 217–259.

Roberston, D., and V. Sarafidis (2013). IV Estimation of Panels with Factor Residuals. Unpublished manuscript.

Robertson, R., V. Sarafidis and T. Yamagata (2013). Estimation of Correlated Random Coefficient Models for Short Panels with a Multi-Factor Error Structure. Unpublished manuscript.

Sarafidis, V., and T. Wansbeek (2012). Cross-Sectional Dependence in Panel Data Analysis. *Econometric Reviews* **31**, 483–531.

Simon, H. A., and C. P. Bonini (1958). The Size Distribution of Business Firms. *American Economic Review* **58**, 607–617.

Steindl, J. (1965). *Random Processes and the Growth of Firms: a Study of the Pareto Law*. Hafner Publishing, New York.

Sutton, J. (1997). Gibrat's Legacy. *Journal of Economic Literature* **35**, 40–59.

Westerlund, J., and J. Breitung (2013). Lessons from a Decade of IPS and LLC. *Econometric Reviews* **32**, 547–591.

Appendix: Proofs

This Appendix makes heavy use of the results of Abadir and Magnus (2005). Here now we state some of the most frequently used results. Throughout $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are going to denote generic matrices of dimension $m \times p$, $n \times q$, $k \times r$ and $l \times s$, respectively. We also denote by $\mathbf{K}_{mp}$ the $mp \times mp$ commutation matrix of zeroes and ones such $\mathbf{K}_{mp}\text{vec}\,\mathbf{A} = \text{vec}\,\mathbf{A}'$. If $\mathbf{A}$ is $m \times m$ (square), then we denote by $\mathbf{D}_m$ the $m^2 \times m(m+1)/2$ duplication matrix of zeroes and ones such that $\mathbf{D}_m\text{vech}\,\mathbf{A} = \text{vec}\,\mathbf{A}$. The following results are going to be used frequently in the sequel:

- $\text{vec}\,(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}\,\mathbf{B}$;
- $\text{vec}\,(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{I}_p \otimes \mathbf{K}_{qm} \otimes \mathbf{I}_n)(\text{vec}\,\mathbf{A} \otimes \text{vec}\,\mathbf{B})$;
- $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$, if $\mathbf{AC}$ and $\mathbf{BD}$ are defined;
- $\mathbf{K}_{mn}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{qp}$.

We also define the matrix derivative operator D_x, which is such that if the matrix function $\mathbf{R}(x)$ is $m \times p$ and x is $n \times q$, then $\text{D}_x\,\mathbf{R}(x) = \partial\text{vec}\,\mathbf{R}(x)/\partial(\text{vec}\,x)'$ is $mp \times nq$. Hence, denoting by d the matrix differential, then we have $\text{d}\,\text{vec}\,\mathbf{R}(x) = \mathbf{A}(x)\text{d}\,\text{vec}\,x$, or $\text{D}_x\,\mathbf{R}(x) = \text{d}\,\text{vec}\,\mathbf{R}(x)/\text{d}\,\text{vec}\,x$. Also, if $\mathbf{R}(x)$ is a $m \times m$ and symmetric, then we define $\text{D}_x\,\mathbf{R}(x) = \partial\text{vec}\,\mathbf{R}(x)/\partial(\text{vech}\,x)'$. $\text{D}_x^q\,\mathbf{R}(x)$ means $\text{D}_x^q\,\mathbf{R}(x) = \partial^q\text{vec}\,\mathbf{R}(x)/[\partial(\text{vec}\,x)']$. Some important rules for differentials:

- $\text{d}\,[\mathbf{R}(x)\mathbf{G}(x)] = [\text{d}\,\mathbf{R}(x)]\mathbf{G}(x) + \mathbf{R}(x)\text{d}\,\mathbf{G}(x)$;
- $\text{d}\,[\mathbf{R}(x) \otimes \mathbf{G}(x)] = [\text{d}\,\mathbf{R}(x)] \otimes \mathbf{G}(x) + \mathbf{R}(x) \otimes \text{d}\,\mathbf{G}(x)$.

Proof of Theorem 1.

Note that $\text{vec}\,(\mathbf{F}^0\mathbf{\Sigma}_\lambda\mathbf{F}') = (\mathbf{F} \otimes \mathbf{F}^0)\text{vec}\,\mathbf{\Sigma}_\lambda$, where $\text{vec}\,\mathbf{\Sigma}_\lambda = \mathbf{D}_r\text{vech}\,\mathbf{\Sigma}_\lambda = \mathbf{D}_r\boldsymbol{\theta}_2$. It follows that $\text{vec}\,[\mathbf{\Gamma}(\rho)\mathbf{F}^0\mathbf{\Sigma}_\lambda\mathbf{F}'\mathbf{I}_T] = [\mathbf{I}_T \otimes \mathbf{\Gamma}(\rho)]\text{vec}\,(\mathbf{F}^0\mathbf{\Sigma}_\lambda\mathbf{F}') = [\mathbf{I}_T \otimes \mathbf{\Gamma}(\rho)](\mathbf{F} \otimes \mathbf{F}^0)\mathbf{D}_r\boldsymbol{\theta}_2$, where $\mathbf{\Gamma} = \mathbf{\Gamma}(\rho)$ has been written as a function of ρ. Making use of this result, it is clear that

$$\begin{aligned}
\mathbf{S}(\mathbf{I}_T \otimes \mathbf{\Gamma}(\rho)\mathbf{F}^0\mathbf{\Sigma}_\lambda\mathbf{F}')\mathbf{e} &= (\mathbf{e}' \otimes \mathbf{S})\text{vec}\,[\mathbf{I}_T \otimes \mathbf{\Gamma}(\rho)\mathbf{F}^0\mathbf{\Sigma}_\lambda\mathbf{F}'] \\
&= (\mathbf{e}' \otimes \mathbf{S})(\mathbf{I}_T \otimes \mathbf{K}_{T^2} \otimes \mathbf{I}_T)[\text{vec}\,\mathbf{I}_T \otimes \text{vec}\,(\mathbf{\Gamma}(\rho)\mathbf{F}^0\mathbf{\Sigma}_\lambda\mathbf{F}')] \\
&= (\mathbf{e}' \otimes \mathbf{S})(\mathbf{I}_T \otimes \mathbf{K}_{T^2} \otimes \mathbf{I}_T)[\text{vec}\,\mathbf{I}_T \otimes (\mathbf{I}_T \otimes \mathbf{\Gamma}(\rho))(\mathbf{F} \otimes \mathbf{F}^0)\mathbf{D}_r\boldsymbol{\theta}_2].
\end{aligned}$$

Moreover, since vec $\mathbf{I}_T$ has just one column,

$$\begin{aligned}[\operatorname{vec}\mathbf{I}_T \otimes (\mathbf{I}_T \otimes \boldsymbol{\Gamma}(\rho))(\mathbf{F}\otimes\mathbf{F}^0)\mathbf{D}_r\boldsymbol{\theta}_2] &= [\operatorname{vec}\mathbf{I}_T \otimes (\mathbf{I}_T \otimes \boldsymbol{\Gamma}(\rho))(\mathbf{F}\otimes\mathbf{F}^0)\mathbf{D}_r](1\otimes\boldsymbol{\theta}_2) \\ &= [\operatorname{vec}\mathbf{I}_T \otimes (\mathbf{I}_T \otimes \boldsymbol{\Gamma}(\rho))(\mathbf{F}\otimes\mathbf{F}^0)\mathbf{D}_r]\boldsymbol{\theta}_2.\end{aligned}$$

Hence,

$$\mathbf{W}^{1/2}\mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}(\rho)\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e} = \mathbf{R}(\rho)\boldsymbol{\theta}_2,$$

where $\mathbf{R}(\rho) = \mathbf{W}^{1/2}\mathbf{C}_1\mathbf{U}(\rho)$, $\mathbf{C}_1 = (\mathbf{e}'\otimes\mathbf{S})(\mathbf{I}_T\otimes\mathbf{K}_{T^2}\otimes\mathbf{I}_T)$ is $T(T+1)/2\times T^4$ and $\mathbf{U}(\rho) = [\operatorname{vec}\mathbf{I}_T \otimes (\mathbf{I}_T \otimes \boldsymbol{\Gamma}(\rho))(\mathbf{F}\otimes\mathbf{F}^0)\mathbf{D}_r]$ is $T^4\times r(r+1)/2$. $\mathbf{g}(\boldsymbol{\theta})$ can therefore be rewritten as

$$\begin{aligned}\mathbf{g}(\boldsymbol{\theta}) &= \mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0 - \mathbf{S}[\mathbf{I}_T\otimes\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}']\mathbf{e}) \\ &= \mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0) - \mathbf{W}^{1/2}\mathbf{S}[\mathbf{I}_T\otimes\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}']\mathbf{e} \\ &= \mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0) - \mathbf{R}(\rho)\boldsymbol{\theta}_2, \qquad \text{(A1)}\end{aligned}$$

suggesting that

$$\mathrm{d}\operatorname{vec}\mathbf{g}(\boldsymbol{\theta}) = \mathrm{d}\,\mathbf{g}(\boldsymbol{\theta}) = -\mathbf{R}(\rho)\mathrm{d}\,\boldsymbol{\theta}_2, \qquad \text{(A2)}$$

where the differential is taken with respect to $\boldsymbol{\theta}_2$ (treating ρ as fixed). The corresponding derivative is given by

$$\mathrm{D}_{\boldsymbol{\theta}_2}\,\mathbf{g}(\boldsymbol{\theta}) = -\mathbf{R}(\rho), \qquad \text{(A3)}$$

which in turn implies

$$\mathrm{D}_{\boldsymbol{\theta}_2}\,Q(\boldsymbol{\theta}) = 2\mathbf{g}(\boldsymbol{\theta})'\mathrm{D}_{\boldsymbol{\theta}_2}\,\mathbf{g}(\boldsymbol{\theta}) = -2\mathbf{g}(\boldsymbol{\theta})'\mathbf{R}(\rho). \qquad \text{(A4)}$$

The relevant first-order condition is therefore given by

$$[\mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0) - \mathbf{R}(\rho)\boldsymbol{\theta}_2]'\mathbf{R}(\rho) = \mathbf{0}_{1\times r(r+1)/2}, \qquad \text{(A5)}$$

or

$$\mathbf{R}(\rho)'\mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0) = \mathbf{R}(\rho)'\mathbf{R}(\rho)\hat{\boldsymbol{\theta}}_2.$$

The $r(r+1)/2\times r(r+1)/2$ matrix $\mathbf{R}(\rho)'\mathbf{R}(\rho)$ is nonsingular. Hence, defining $\hat{\mathbf{z}}_T(\rho) = \mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0)$, we have

$$\hat{\boldsymbol{\theta}}_2 = [\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)'\hat{\mathbf{z}}_T(\rho). \qquad \text{(A6)}$$

The concentrated version of $\mathbf{g}(\boldsymbol{\theta})$ is

$$\begin{aligned}\mathbf{g}_C(\rho) &= \mathbf{g}(\rho,\hat{\boldsymbol{\theta}}_2(\rho)) = \mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0) - \mathbf{R}(\rho)\hat{\boldsymbol{\theta}}_2(\rho) \qquad \text{(A7)} \\ &= \hat{\mathbf{z}}_T(\rho) - \mathbf{R}(\rho)\hat{\boldsymbol{\theta}}_2(\rho) = \mathbf{P}(\rho)\hat{\mathbf{z}}_T(\rho).\end{aligned}$$

where $\mathbf{P}(\rho) = \mathbf{I}_{T(T+1)/2} - \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)'$. Note that $\mathbf{g}(\boldsymbol{\theta}) = \mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0) - \mathbf{R}(\rho)\boldsymbol{\theta}_2 = \hat{\mathbf{z}}_T(\rho) - \mathbf{R}(\rho)\boldsymbol{\theta}_2$. Hence, since $\mathbf{P}(\rho)\mathbf{R}(\rho) = \mathbf{0}_{T(T+1)/2\times r(r+1)/2}$, we have

$$\mathbf{g}_C(\rho) = \mathbf{P}(\rho)\hat{\mathbf{z}}_T(\rho) = \mathbf{P}(\rho)[\hat{\mathbf{z}}_T(\rho) - \mathbf{R}(\rho)\boldsymbol{\theta}_2] = \mathbf{P}(\rho)\mathbf{g}(\boldsymbol{\theta}).$$

Consider $\sqrt{N}\mathbf{g}(\boldsymbol{\theta})$, which we can write as

$$\sqrt{N}\mathbf{g}(\boldsymbol{\theta}) = \sqrt{N}\mathbf{W}^{1/2}[(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0) - \mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}] = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{g}_i(\boldsymbol{\theta}).$$

where $\mathbf{g}_i(\boldsymbol{\theta}) = \mathbf{W}^{1/2}[\mathbf{Z}_i'(\mathbf{y}_i - \rho\mathbf{y}_i^0) - \mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}]$. By using the fact that $\mathbf{y}_i = \rho\mathbf{y}_i^0 + \mathbf{u}_i$, where $\mathbf{u}_i = (u_{i,1}, ..., u_{i,T})' = \mathbf{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i$ with $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, ..., \varepsilon_{i,T})'$, $\mathbf{g}_i(\boldsymbol{\theta})$ can be written as

$$\begin{aligned}\mathbf{g}_i(\boldsymbol{\theta}) &= \mathbf{W}^{1/2}[\mathbf{Z}_i'(\mathbf{y}_i - \rho\mathbf{y}_i^0) - \mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}] \\ &= \mathbf{W}^{1/2}[\mathbf{Z}_i'\mathbf{u}_i - \mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}],\end{aligned}$$

or

$$\mathbf{g}_i(\boldsymbol{\theta}) = \mathbf{W}^{1/2}\begin{bmatrix}\mathbf{S}_1(\mathbf{y}_i^0\mathbf{u}_i' - \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}_1 \\ \vdots \\ \mathbf{S}_T(\mathbf{y}_i^0\mathbf{u}_i' - \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}_T\end{bmatrix}.$$

From $\mathbf{y}_i^0 = \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\lambda}_i + \boldsymbol{\Gamma}\boldsymbol{\varepsilon}_i^0$, we obtain

$$\begin{aligned}\mathbf{y}_i^0\mathbf{u}_i' - \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}' &= (\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\lambda}_i + \boldsymbol{\Gamma}\boldsymbol{\varepsilon}_i^0)(\mathbf{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i)' - \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}' \\ &= \boldsymbol{\Gamma}\mathbf{F}^0(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' - \boldsymbol{\Sigma}_\lambda)\mathbf{F}' + \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i' + \boldsymbol{\Gamma}\boldsymbol{\varepsilon}_i^0\boldsymbol{\lambda}_i'\mathbf{F}' + \boldsymbol{\Gamma}\boldsymbol{\varepsilon}_i^0\boldsymbol{\varepsilon}_i'.\end{aligned}$$

It is easy to see that premultiplying the above four terms by $\mathbf{S}_t$, post-multiplying by $\mathbf{e}_t$ and taking expectations yields zero for all t. Furthermore, these terms are independent across i with bounded fourth-order moments. Hence, by a central limit theorem, we have

$$\sqrt{N}\mathbf{g}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{g}_i(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}_{T(T+1)/2\times 1}, \boldsymbol{\Sigma}_g)$$

as $N \to \infty$, where $\boldsymbol{\Sigma}_g = \lim_{N\to\infty}\sum_{i=1}^{N} E[\mathbf{g}_i(\boldsymbol{\theta}_0)\mathbf{g}_i(\boldsymbol{\theta}_0)']/N$ and $\xrightarrow{d}$ signifies convergence in distribution. $E[\mathbf{g}_i(\boldsymbol{\theta})\mathbf{g}_i(\boldsymbol{\theta})']$ may be expanded in the following fashion:

$$\begin{aligned}&E[\mathbf{g}_i(\boldsymbol{\theta})\mathbf{g}_i(\boldsymbol{\theta})'] \\ =\ & \mathbf{W}^{1/2}E[(\mathbf{Z}_i'\mathbf{u}_i - \mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e})(\mathbf{Z}_i'\mathbf{u}_i - \mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e})'](\mathbf{W}^{1/2})' \\ =\ & \mathbf{W}^{1/2}E(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i)(\mathbf{W}^{1/2})' - \mathbf{W}^{1/2}\mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}E(u_i'\mathbf{Z}_i)(\mathbf{W}^{1/2})' \\ -\ & \mathbf{W}^{1/2}E(\mathbf{Z}_i'\mathbf{u}_i)\mathbf{e}'(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{S}'(\mathbf{W}^{1/2})' \\ +\ & \mathbf{W}^{1/2}\mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}\mathbf{e}'(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{S}'(\mathbf{W}^{1/2})' \\ =\ & \mathbf{W}^{1/2}E(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i)(\mathbf{W}^{1/2})' - \mathbf{W}^{1/2}\mathbf{S}(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}\mathbf{e}'(\mathbf{I}_T \otimes \boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')'\mathbf{S}'(\mathbf{W}^{1/2})'.\end{aligned}$$

Here

$$\begin{aligned}
&E(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i)\\
&= E[\mathbf{Z}_i'(\mathbf{F}\boldsymbol{\lambda}_i+\boldsymbol{\varepsilon}_i)(\mathbf{F}\boldsymbol{\lambda}_i+\boldsymbol{\varepsilon}_i)'\mathbf{Z}_i]\\
&= E(\mathbf{Z}_i'\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\mathbf{Z}_i)+E(\mathbf{Z}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\lambda}_i'\mathbf{F}'\mathbf{Z}_i)+E(\mathbf{Z}_i'\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\varepsilon}_i'\mathbf{Z}_i)+E(\mathbf{Z}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{Z}_i)\\
&= E(\mathbf{Z}_i'\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\mathbf{Z}_i)+E[\mathbf{Z}_i'E(\boldsymbol{\varepsilon}_i|\boldsymbol{\lambda}_i,\mathbf{Z}_i)\boldsymbol{\lambda}_i'\mathbf{F}'\mathbf{Z}_i]+E[\mathbf{Z}_i'\mathbf{F}\boldsymbol{\lambda}_iE(\boldsymbol{\varepsilon}_i|\boldsymbol{\lambda}_i,\mathbf{Z}_i)'\mathbf{Z}_i]+E[\mathbf{Z}_i'E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\mathbf{Z}_i)\mathbf{Z}_i]\\
&= E(\mathbf{Z}_i'\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\mathbf{Z}_i)+\sigma_\varepsilon^2E(\mathbf{Z}_i'\mathbf{Z}_i),
\end{aligned}$$

where the third equality holds because $\boldsymbol{\varepsilon}_i$ is uncorrelated with $\boldsymbol{\lambda}_i$ and $\mathbf{Z}_i$. The fourth equality follows from the fact that $E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')=\sigma_\varepsilon^2\mathbf{I}_T$. As for $E(\mathbf{Z}_i'\mathbf{Z}_i)$, we have

$$E(\mathbf{Z}_i'\mathbf{Z}_i)=E[\mathbf{S}(\mathbf{I}_T\otimes\mathbf{y}_i^0)(\mathbf{I}_T\otimes\mathbf{y}_i'^0)\mathbf{S}']=\mathbf{S}(\mathbf{I}_T\otimes\boldsymbol{\Sigma}_{y^0})\mathbf{S}',$$

where $\boldsymbol{\Sigma}_{y^0}=E[\mathbf{y}_i^0(\mathbf{y}_i^0)']$. Similarly, by using the fact that $\boldsymbol{\varepsilon}_i^0$ is uncorrelated with $\boldsymbol{\lambda}_i$ we have

$$\begin{aligned}
E(\mathbf{Z}_i'\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\mathbf{Z}_i) &= E[\mathbf{S}(\mathbf{I}_T\otimes\mathbf{y}_i^0)\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'(\mathbf{I}_T\otimes(\mathbf{y}_i^0)')\mathbf{S}']\\
&= \mathbf{S}E[(\mathbf{I}_T\otimes\mathbf{y}_i^0)\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'(\mathbf{I}_T\otimes(\mathbf{y}_i^0)')]\mathbf{S}'\\
&= \mathbf{S}E[(\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\otimes\mathbf{y}_i^0)(\mathbf{I}_T\otimes(\mathbf{y}_i^0)')]\mathbf{S}'\\
&= \mathbf{S}E[\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\otimes\mathbf{y}_i^0(\mathbf{y}_i^0)']\mathbf{S}'\\
&= \mathbf{S}E[\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\otimes(\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\lambda}_i+\boldsymbol{\Gamma}\boldsymbol{\varepsilon}_i^0)(\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\lambda}_i+\boldsymbol{\Gamma}\boldsymbol{\varepsilon}_i^0)']\mathbf{S}'\\
&= \mathbf{S}[E(\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\otimes\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\left(\mathbf{F}^0\right)'\boldsymbol{\Gamma}')+E(\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\otimes\boldsymbol{\Gamma}\boldsymbol{\varepsilon}_i^0(\boldsymbol{\varepsilon}_i^0)'\boldsymbol{\Gamma}')]\mathbf{S}',
\end{aligned}$$

and therefore

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^N E(\mathbf{Z}_i'\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\mathbf{Z}_i) &= \mathbf{S}\left(\frac{1}{N}\sum_{i=1}^N E(\mathbf{F}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\mathbf{F}'\otimes\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'(\mathbf{F}^0)'\boldsymbol{\Gamma}')+\mathbf{F}\boldsymbol{\Sigma}_\lambda\mathbf{F}'\otimes\boldsymbol{\Gamma}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}'\right)\mathbf{S}'\\
&= \mathbf{S}\boldsymbol{\Sigma}_{F\lambda y^0}\mathbf{S}',
\end{aligned}$$

with an implicit definition of $\boldsymbol{\Sigma}_{F\lambda y^0}$. It is straightforward to see that under our assumptions, $\boldsymbol{\Sigma}_{F\lambda y^0}$ exists and is positive definite.

Thus, putting everything together,

$$\boldsymbol{\Sigma}_g=\mathbf{W}^{1/2}\mathbf{S}\boldsymbol{\Sigma}_{F\lambda y^0}\mathbf{S}'(\mathbf{W}^{1/2})'-\mathbf{W}^{1/2}\mathbf{S}(\mathbf{I}_T\otimes\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')\mathbf{e}\mathbf{e}'(\mathbf{I}_T\otimes\boldsymbol{\Gamma}\mathbf{F}^0\boldsymbol{\Sigma}_\lambda\mathbf{F}')'\mathbf{S}'(\mathbf{W}^{1/2})'.$$

The above results suggest that if we let $\boldsymbol{\Sigma}_{gc}=\mathbf{P}(\rho_0)\boldsymbol{\Sigma}_g\mathbf{P}(\rho_0)'$, then

$$\sqrt{N}\mathbf{g}_C(\rho_0)=\mathbf{P}(\rho_0)\mathbf{g}(\boldsymbol{\theta}_0)\xrightarrow{d}N(\mathbf{0}_{T(T+1)/2\times1},\boldsymbol{\Sigma}_{gC}).$$

This result will be used as a basis for deriving the asymptotic distribution of $\hat{\rho}$. In order to appreciate how, consider $Q_C(\rho) = \mathbf{g}_C(\rho)'\mathbf{g}_C(\rho)$. By Taylor expansion about $\rho = \rho_0$, we have

$$Q_C(\rho) = Q_C(\rho_0) + \sum_{q=1}^{\infty} \mathrm{D}_\rho^q\, Q_C(\rho_0)\frac{(\rho-\rho_0)^q}{q!},$$

suggesting that

$$\begin{aligned} N[Q_C(\rho) - Q_C(\rho_0)] &= \sqrt{N}\mathrm{D}_\rho\, Q_C(\rho_0)\sqrt{N}(\rho-\rho_0) + \mathrm{D}_\rho^2\, Q_C(\rho_0)\frac{[\sqrt{N}(\rho-\rho_0)]^2}{2} \\ &+ O_p\left(\frac{\mathrm{D}_\rho^3\, Q_C(\rho_0)[\sqrt{N}(\rho-\rho_0)]^3}{\sqrt{N}}\right). \end{aligned}$$

Later on we show that $\mathrm{D}_\rho^3\, Q_C(\rho_0) = O_p(1)$. The last term is therefore $O_p(1/\sqrt{N}\,) = o_p(1)$. $\hat{\rho}$ is the minimizer of $N[Q_C(\rho) - Q_C(\rho_0)]$. Thus, treating this as a function of $\sqrt{N}(\rho-\rho_0)$, we obtain the following first order condition:

$$\sqrt{N}\mathrm{D}_\rho\, Q_C(\rho_0) + \mathrm{D}_\rho^2\, Q_C(\rho_0)\sqrt{N}(\hat{\rho}-\rho_0) + o_p(1) = 0,$$

or

$$\sqrt{N}(\hat{\rho}-\rho_0) = -\frac{\sqrt{N}\mathrm{D}_\rho\, Q_C(\rho_0)}{\mathrm{D}_\rho^2\, Q_C(\rho_0)} + o_p(1).$$

Thus, in order to work out the asymptotic distribution of $\sqrt{N}(\hat{\rho}-\rho_0)$, we need $\mathrm{D}_\rho\, Q_C(\rho_0)$ and $\mathrm{D}_\rho^2\, Q_C(\rho_0)$, which, by the product rule, can be written as

$$\begin{aligned} \mathrm{D}_\rho\, Q_C(\rho) &= 2\mathbf{g}_C(\rho)'\mathrm{D}_\rho\, \mathbf{g}_C(\rho), \\ \mathrm{D}_\rho^2\, Q_C(\rho) &= 2\mathbf{g}_C(\rho)'\mathrm{D}_\rho^2\, \mathbf{g}_C(\rho) + 2\mathrm{D}_\rho\, \mathbf{g}_C(\rho)'\mathrm{D}_\rho\, \mathbf{g}_C(\rho), \\ \mathrm{D}_\rho^3\, Q_C(\rho) &= 4\mathbf{g}_C(\rho)'\mathrm{D}_\rho^3\, \mathbf{g}_C(\rho) + 6\mathrm{D}_\rho\, \mathbf{g}_C(\rho)'\mathrm{D}_\rho^2\, \mathbf{g}_C(\rho). \end{aligned}$$

Consider $\mathrm{D}_\rho\, \mathbf{g}_C(\rho)$. By the product rule,

$$\begin{aligned} \mathrm{d}\,\mathbf{g}_C(\rho) &= \mathbf{I}_{T(T+1)/2}\mathrm{d}\,\mathbf{P}(\rho)\hat{\mathbf{z}}_T(\rho) + \mathbf{P}(\rho)\mathrm{d}\,\hat{\mathbf{z}}_T(\rho) \\ &= [\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{d}\,\mathrm{vec}\,\mathbf{P}(\rho) + \mathbf{P}(\rho)\mathrm{d}\,\hat{\mathbf{z}}_T(\rho), \end{aligned}$$

giving the following derivative:

$$\mathrm{D}_\rho\, \mathbf{g}_C(\rho) = [\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\, \mathbf{P}(\rho) + \mathbf{P}(\rho)\mathrm{D}_\rho\, \hat{\mathbf{z}}_T(\rho).$$

$\mathrm{D}_\rho\, \hat{\mathbf{z}}_T(\rho)$ is particularly easy and is given by

$$\mathrm{D}_\rho\, \hat{\mathbf{z}}_T(\rho) = \mathrm{D}_\rho\left(\mathbf{W}^{1/2}(\hat{\mathbf{m}} - \rho\hat{\mathbf{m}}^0)\right) = -\mathbf{W}^{1/2}\hat{\mathbf{m}}^0.$$

$\mathrm{D}_\rho\, \mathbf{P}(\rho)$ requires more work. We begin by noting that

$$\begin{aligned}
\mathrm{d}\,\mathbf{P}(\rho) &= -\mathrm{d}\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)') \\
&= -(\mathrm{d}\,\mathbf{R}(\rho))[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)' - \mathbf{R}(\rho)(\mathrm{d}\,[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1})\mathbf{R}(\rho)' \\
&- \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}[\mathrm{d}\,\mathbf{R}(\rho)]' \\
&= -(\mathrm{d}\,\mathbf{R}(\rho))[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)' \\
&+ \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}(\mathrm{d}\,[\mathbf{R}(\rho)'\mathbf{R}(\rho)])[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)' \\
&- \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}[\mathrm{d}\,\mathbf{R}(\rho)]' \\
&= -(\mathrm{d}\,\mathbf{R}(\rho))[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)' \\
&+ \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}([\mathrm{d}\,\mathbf{R}(\rho)]'\mathbf{R}(\rho) + \mathbf{R}(\rho)'[\mathrm{d}\,\mathbf{R}(\rho)])[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)' \\
&- \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}[\mathrm{d}\,\mathbf{R}(\rho)]' \\
&= -\mathbf{P}(\rho)(\mathrm{d}\,\mathbf{R}(\rho))[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)' - \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}[\mathrm{d}\,\mathbf{R}(\rho)]'\mathbf{P}(\rho),
\end{aligned}$$

suggesting

$$\begin{aligned}
&\mathrm{d}\,\mathrm{vec}\,\mathbf{P}(\rho) \\
&= -[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]\mathrm{d}\,\mathrm{vec}\,\mathbf{R}(\rho) - [\mathbf{P}(\rho) \otimes \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}]\mathrm{d}\,\mathrm{vec}\,\mathbf{R}(\rho)' \\
&= -[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]\mathrm{d}\,\mathrm{vec}\,\mathbf{R}(\rho) \\
&- [\mathbf{P}(\rho) \otimes \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}]\mathbf{K}_{Tr(T+1)(r+1)/4}\mathrm{d}\,\mathrm{vec}\,\mathbf{R}(\rho) \\
&= -[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]\mathrm{d}\,\mathrm{vec}\,\mathbf{R}(\rho) \\
&- \mathbf{K}_{T^2(T+1)^2/4}[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]\mathrm{d}\,\mathrm{vec}\,\mathbf{R}(\rho) \\
&= -(\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]\mathrm{d}\,\mathrm{vec}\,\mathbf{R}(\rho).
\end{aligned}$$

Here

$$\mathrm{d}\,\mathrm{vec}\,\mathbf{R}(\rho) = \mathrm{vec}\,[\mathbf{W}^{1/2}\mathbf{C}_1\mathrm{d}\,\mathbf{U}(\rho)] = [\mathbf{I}_{r(r+1)/2} \otimes \mathbf{W}^{1/2}\mathbf{C}_1]\mathrm{d}\,\mathrm{vec}\,\mathbf{U}(\rho),$$

and therefore

$$\mathrm{D}_\rho\,\mathbf{R}(\rho) = [\mathbf{I}_{r(r+1)/2} \otimes \mathbf{W}^{1/2}\mathbf{C}_1]\mathrm{D}_\rho\,\mathbf{U}(\rho).$$

Consider $\mathrm{D}_\rho\,\mathbf{U}(\rho)$. A direct calculation reveals that

$$
\begin{aligned}
&\mathrm{d}\,\mathrm{vec}\,\mathbf{U}(\rho)\\
&= \mathrm{vec}\,[\mathrm{vec}\,\mathbf{I}_T \otimes (\mathbf{I}_T \otimes \mathrm{d}\,\boldsymbol{\Gamma}(\rho))(\mathbf{F}\otimes\mathbf{F}^0)\mathbf{D}_r]\\
&= (\mathbf{K}_{T^2r(r+1)/2}\otimes\mathbf{I}_{T^2})[\mathrm{vec}\,\mathbf{I}_T\otimes\mathrm{vec}\,((\mathbf{I}_T\otimes\mathrm{d}\,\boldsymbol{\Gamma}(\rho))(\mathbf{F}\otimes\mathbf{F}^0)\mathbf{D}_r)]\\
&= (\mathbf{K}_{T^2r(r+1)/2}\otimes\mathbf{I}_{T^2})[\mathrm{vec}\,\mathbf{I}_T\otimes(\mathbf{D}_r'(\mathbf{F}\otimes\mathbf{F}^0)'\otimes\mathbf{I}_{T^2})\mathrm{vec}\,(\mathbf{I}_T\otimes\mathrm{d}\,\boldsymbol{\Gamma}(\rho))]\\
&= (\mathbf{K}_{T^2r(r+1)/2}\otimes\mathbf{I}_{T^2})[\mathrm{vec}\,\mathbf{I}_T\otimes\mathbf{D}_r'(\mathbf{F}\otimes\mathbf{F}^0)'\otimes\mathbf{I}_{T^2})(\mathbf{I}_T\otimes\mathbf{K}_{T^2}\otimes\mathbf{I}_T)(\mathrm{vec}\,\mathbf{I}_T\otimes\mathrm{d}\,\mathrm{vec}\,\boldsymbol{\Gamma}(\rho))].
\end{aligned}
$$

Clearly,

$$
\mathrm{D}_\rho\,\boldsymbol{\Gamma}(\rho) = \mathrm{vec}\begin{bmatrix} 0 & 0 & \dots & 0\\ 1 & 0 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ \rho^{T-2} & \dots & 1 & 0 \end{bmatrix},
$$

which we can use, together with the fact that $\mathrm{vec}\,\mathbf{I}_T$ is vector, to show that

$$
\begin{aligned}
\mathrm{D}_\rho\,\mathbf{U}(\rho) &= (\mathbf{K}_{T^2r(r+1)/2}\otimes\mathbf{I}_{T^2})\\
&\times [\mathrm{vec}\,\mathbf{I}_T\otimes\mathbf{D}_r'(\mathbf{F}\otimes\mathbf{F}^0)'\otimes\mathbf{I}_{T^2})(\mathbf{I}_T\otimes\mathbf{K}_{T^2}\otimes\mathbf{I}_T)(\mathrm{vec}\,\mathbf{I}_T\otimes\mathrm{D}_\rho\,\boldsymbol{\Gamma}(\rho))]\\
&= \mathbf{C}_2\mathrm{D}_\rho\,\boldsymbol{\Gamma}(\rho),
\end{aligned}
$$

where

$$
\mathbf{C}_2 = (\mathbf{K}_{T^2r(r+1)/2}\otimes\mathbf{I}_{T^2})[\mathrm{vec}\,\mathbf{I}_T\otimes\mathbf{D}_r'(\mathbf{F}\otimes\mathbf{F}^0)'\otimes\mathbf{I}_{T^2})(\mathbf{I}_T\otimes\mathbf{K}_{T^2}\otimes\mathbf{I}_T)(\mathrm{vec}\,\mathbf{I}_T\otimes\mathbf{I}_{T^2})].
$$

This can be substituted back into the expression for $\mathrm{D}_\rho\,\mathbf{R}(\rho)$, giving

$$
\mathrm{D}_\rho\,\mathbf{R}(\rho) = [\mathbf{I}_{r(r+1)/2}\otimes\mathbf{W}^{1/2}\mathbf{C}_1]\mathrm{D}_\rho\,\mathbf{U}(\rho) = [\mathbf{I}_{r(r+1)/2}\otimes\mathbf{W}^{1/2}\mathbf{C}_1]\mathbf{C}_2\mathrm{D}_\rho\,\boldsymbol{\Gamma}(\rho),
$$

which in turn implies

$$
\begin{aligned}
\mathrm{D}_\rho\,\mathbf{P}(\rho) &= -(\mathbf{I}_{T^2(T+1)^2/4}+\mathbf{K}_{T^2(T+1)^2/4})[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\otimes\mathbf{P}(\rho)]\mathrm{D}_\rho\,\mathbf{R}(\rho)\\
&= -(\mathbf{I}_{T^2(T+1)^2/4}+\mathbf{K}_{T^2(T+1)^2/4})[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\otimes\mathbf{P}(\rho)]\\
&\times [\mathbf{I}_{r(r+1)/2}\otimes\mathbf{W}^{1/2}\mathbf{C}_1]\mathbf{C}_2\mathrm{D}_\rho\,\boldsymbol{\Gamma}(\rho). \qquad \text{(A8)}
\end{aligned}
$$

We had

$$
\mathrm{D}_\rho\,\mathbf{g}_C(\rho) = [\hat{\mathbf{z}}_T(\rho)'\otimes\mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\,\mathbf{P}(\rho) + \mathbf{P}(\rho)\mathrm{D}_\rho\,\hat{\mathbf{z}}_T(\rho).
$$

Since $\mathrm{D}_\rho \mathbf{P}(\rho)$, $\mathbf{P}(\rho)$ and $\mathbf{W}^{1/2}$ are just constant matrices, in order to work out the limit of $\mathrm{D}_\rho \mathbf{g}_C(\rho)$ we only need to consider $\hat{\mathbf{z}}_T(\rho)$ and $\mathrm{D}_\rho \hat{\mathbf{z}}_T(\rho)$. The limits of these terms are simple consequences of the law of large numbers. Indeed, letting $\mathbf{z}_T(\rho) = \mathbf{W}^{1/2}(\mathbf{m} - \rho\mathbf{m}^0)$, we have

$$\begin{aligned} \hat{\mathbf{z}}_T(\rho) &= \mathbf{z}_T(\rho) + o_p(1), && \text{(A9)} \\ \mathrm{D}_\rho \hat{\mathbf{z}}_T(\rho) &= -\mathbf{W}^{1/2}\mathbf{m}^0 + o_p(1) = \mathrm{D}_\rho \mathbf{z}_T(\rho) + o_p(1), && \text{(A10)} \end{aligned}$$

from which we obtain

$$\mathrm{D}_\rho \mathbf{g}_C(\rho) = \gamma_1(\rho) + o_p(1),$$

where $\gamma_1(\rho) = [\mathbf{z}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho \mathbf{P}(\rho) + \mathbf{P}(\rho)\mathrm{D}_\rho \mathbf{z}_T(\rho)$.

Next, consider $\mathrm{D}^2_\rho \mathbf{g}_C(\rho)$. In view of the result for $\mathrm{D}_\rho \mathbf{g}_C(\rho)$, which is already in vector form, it is clear that

$$\begin{aligned} &\mathrm{d}\,\mathrm{D}_\rho \mathbf{g}_c(\rho) \\ &= \mathrm{d}\left([\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\mathbf{P}(\rho) + \mathbf{P}(\rho)\mathrm{D}_\rho\hat{\mathbf{z}}_T(\rho)\right) \\ &= [\mathrm{d}\,\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\mathbf{P}(\rho) + [\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{d}\,\mathrm{D}_\rho\mathbf{P}(\rho) \\ &+ \mathrm{d}\,\mathbf{P}(\rho)\mathrm{D}_\rho\hat{\mathbf{z}}_T(\rho) + \mathbf{P}(\rho)\mathrm{d}\,\mathrm{D}_\rho\hat{\mathbf{z}}_T(\rho) \\ &= [\mathrm{D}_\rho\mathbf{P}(\rho) \otimes \mathbf{I}_{T(T+1)/2}](\mathbf{I}_{T(T+1)/2} \otimes \mathbf{K}_{T(T+1)/2} \otimes \mathbf{I}_{T(T+1)/2})[\mathrm{d}\,\mathrm{vec}\,\hat{\mathbf{z}}_T(\rho) \otimes \mathrm{vec}\,\mathbf{I}_{T(T+1)/2}] \\ &+ [\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{d}\,\mathrm{vec}\,\mathrm{D}_\rho\mathbf{P}(\rho) \\ &+ [\mathrm{D}_\rho\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{d}\,\mathrm{vec}\,\mathbf{P}(\rho) + \mathbf{P}(\rho)\mathrm{d}\,\mathrm{vec}\,\mathrm{D}_\rho\hat{\mathbf{z}}_T(\rho), \end{aligned}$$

where we have made use of

$$\begin{aligned} &[\mathrm{d}\,\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\mathbf{P}(\rho) \\ &= \mathrm{vec}\left(\mathbf{I}_{T(T+1)/2}[\mathrm{d}\,\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\mathbf{P}(\rho)\right) \\ &= [\mathrm{D}_\rho\mathbf{P}(\rho) \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{vec}\,[\mathrm{d}\,\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}] \\ &= [\mathrm{D}_\rho\mathbf{P}(\rho) \otimes I_{T(T+1)/2}](\mathbf{I}_{T(T+1)/2} \otimes \mathbf{K}_{T(T+1)/2} \otimes \mathbf{I}_{T(T+1)/2})[\mathrm{vec}\,\mathrm{d}\,\hat{\mathbf{z}}_T(\rho)' \otimes \mathrm{vec}\,\mathbf{I}_{T(T+1)/2}] \\ &= [\mathrm{D}_\rho\mathbf{P}(\rho) \otimes \mathbf{I}_{T(T+1)/2}](\mathbf{I}_{T(T+1)/2} \otimes \mathbf{K}_{T(T+1)/2} \otimes \mathbf{I}_{T(T+1)/2})[\mathrm{d}\,\mathrm{vec}\,\hat{\mathbf{z}}_T(\rho) \otimes \mathrm{vec}\,\mathbf{I}_{T(T+1)/2}], \end{aligned}$$

and

$$\mathrm{d}\,\mathbf{P}(\rho)\mathrm{D}_\rho\hat{\mathbf{z}}_T(\rho) = \mathrm{vec}\,[\mathbf{I}_{T(T+1)/2}\mathrm{d}\,\mathbf{P}(\rho)\mathrm{D}_\rho\hat{\mathbf{z}}_T(\rho)] = [\mathrm{D}_\rho\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{d}\,\mathrm{vec}\,\mathbf{P}(\rho).$$

Hence,

$$
\begin{aligned}
&\mathrm{D}^2_\rho\, \mathbf{g}_C(\rho) \\
&= [\mathrm{D}_\rho\, \mathbf{P}(\rho) \otimes \mathbf{I}_{T(T+1)/2}](\mathbf{I}_{T(T+1)/2} \otimes \mathbf{K}_{T(T+1)/2} \otimes \mathbf{I}_{T(T+1)/2})[\mathrm{D}_\rho\, \hat{\mathbf{z}}_T(\rho) \otimes \operatorname{vec} \mathbf{I}_{T(T+1)/2}] \\
&+ [\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}^2_\rho\, \mathbf{P}(\rho) + [\mathrm{D}_\rho\, \hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\, \mathbf{P}(\rho) \\
&+ \mathbf{P}(\rho)\mathrm{D}^2_\rho\, \hat{\mathbf{z}}_T(\rho),
\end{aligned}
\tag{A11}
$$

where all terms but $\mathrm{D}^2_\rho\, \mathbf{P}(\rho)$ and $\mathrm{D}^2_\rho\, \hat{\mathbf{z}}_T(\rho)$ are known from before. We start with $\mathrm{D}^2_\rho\, \mathbf{P}(\rho)$, whose differential is given by

$$
\begin{aligned}
\mathrm{d}\,\mathrm{D}_\rho\, \mathbf{P}(\rho) &= -(\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})\mathrm{d}\,([\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]\mathrm{D}_\rho\, \mathbf{R}(\rho)) \\
&= -(\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})(\mathrm{d}\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)])\mathrm{D}_\rho\, \mathbf{R}(\rho) \\
&- (\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]\mathrm{d}\,\mathrm{D}_\rho\, \mathbf{R}(\rho).
\end{aligned}
$$

Consider $\mathrm{d}\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]$, which can be expanded in the following manner:

$$
\begin{aligned}
&\mathrm{d}\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)] \\
&\quad= \mathrm{d}\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \otimes \mathbf{P}(\rho) + \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathrm{d}\,\mathbf{P}(\rho),
\end{aligned}
$$

where

$$
\begin{aligned}
&\mathrm{d}\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \\
&= (\mathrm{d}\,\mathbf{R}(\rho))[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} + \mathbf{R}(\rho)(\mathrm{d}\,[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \\
&= (\mathrm{d}\,\mathbf{R}(\rho))[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} - \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}(\mathrm{d}\,[\mathbf{R}(\rho)'\mathbf{R}(\rho)])[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \\
&= (\mathrm{d}\,\mathbf{R}(\rho))[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \\
&\quad -\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}([\mathrm{d}\,\mathbf{R}(\rho)]'\mathbf{R}(\rho) + \mathbf{R}(\rho)'\mathrm{d}\,\mathbf{R}(\rho))[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}.
\end{aligned}
$$

We now put this in vector form. Note in particular how

$$
\begin{aligned}
&\operatorname{vec}\,([\mathrm{d}\,\mathbf{R}(\rho)]'\mathbf{R}(\rho) + \mathbf{R}(\rho)'\mathrm{d}\,\mathbf{R}(\rho)) \\
&= \operatorname{vec}\,(\mathbf{I}_{r(r+1)/2}[\mathrm{d}\,\mathbf{R}(\rho)]'\mathbf{R}(\rho)) + \operatorname{vec}\,(\mathbf{R}(\rho)'\mathrm{d}\,\mathbf{R}(\rho)\mathbf{I}_{r(r+1)/2}) \\
&= (\mathbf{R}(\rho)' \otimes \mathbf{I}_{r(r+1)/2})\operatorname{vec} \mathrm{d}\,\mathbf{R}(\rho)' + (\mathbf{I}_{r(r+1)/2} \otimes \mathbf{R}(\rho)')\mathrm{d}\operatorname{vec} \mathbf{R}(\rho) \\
&= [(\mathbf{R}(\rho)' \otimes \mathbf{I}_{r(r+1)/2})\mathbf{K}_{Tr(T+1)(r+1)/4} + \mathbf{I}_{r(r+1)/2} \otimes \mathbf{R}(\rho)']\mathrm{d}\operatorname{vec} \mathbf{R}(\rho) \\
&= (\mathbf{K}_{r^2(r+1)^2/4} + \mathbf{I}_{r^2(r+1)^2/4})[\mathbf{I}_{r(r+1)/2} \otimes \mathbf{R}(\rho)']\mathrm{d}\operatorname{vec} \mathbf{R}(\rho),
\end{aligned}
$$

giving

$$
\begin{aligned}
& \operatorname{vec} \mathrm{d}\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \\
= \;& ([\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{I}_{T(T+1)/2}) \mathrm{d} \operatorname{vec} \mathbf{R}(\rho) \\
- \;& ([\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \operatorname{vec}\,([\mathrm{d}\,\mathbf{R}(\rho)]'\mathbf{R}(\rho) + \mathbf{R}(\rho)'\mathrm{d}\,\mathbf{R}(\rho)) \\
= \;& ([\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{I}_{T(T+1)/2}) \mathrm{d} \operatorname{vec} \mathbf{R}(\rho) \\
- \;& ([\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1})(\mathbf{K}_{r^2(r+1)^2/4} + \mathbf{I}_{r^2(r+1)^2/4}) \\
\times \;& [\mathbf{I}_{r(r+1)/2} \otimes \mathbf{R}(\rho)'] \mathrm{d} \operatorname{vec} \mathbf{R}(\rho).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{D}_\rho\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) &= ([\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{I}_{T(T+1)/2}) \mathrm{D}_\rho\,\mathbf{R}(\rho) \\
&- ([\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \\
&\times (\mathbf{K}_{r^2(r+1)^2/4} + \mathbf{I}_{r^2(r+1)^2/4})[\mathbf{I}_{r(r+1)/2} \otimes \mathbf{R}(\rho)'] \mathrm{D}_\rho\,\mathbf{R}(\rho). \quad \text{(A12)}
\end{aligned}
$$

It follows that since

$$
\begin{aligned}
& \operatorname{vec} \mathrm{d}\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)] \\
= \;& \operatorname{vec}\,[\mathrm{d}\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \otimes \mathbf{P}(\rho)] + \operatorname{vec}\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathrm{d}\,\mathbf{P}(\rho)] \\
= \;& (\mathbf{I}_{r(r+1)/2} \otimes \mathbf{K}_{T^2(T+1)^2/4} \otimes \mathbf{I}_{T(T+1)/2}) \\
\times \;& [\mathrm{d} \operatorname{vec}\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \otimes \operatorname{vec} \mathbf{P}(\rho) + \operatorname{vec}\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \otimes \mathrm{d} \operatorname{vec} \mathbf{P}(\rho)],
\end{aligned}
$$

we can show that

$$
\begin{aligned}
& \mathrm{D}_\rho\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)] \\
= \;& (\mathbf{I}_{r(r+1)/2} \otimes \mathbf{K}_{T^2(T+1)^2/4} \otimes \mathbf{I}_{T(T+1)/2}) \\
\times \;& [\mathrm{D}_\rho\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \otimes \operatorname{vec} \mathbf{P}(\rho) + \operatorname{vec}\,(\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1}) \otimes \mathrm{D}_\rho\,\mathbf{P}(\rho)]. \quad \text{(A13)}
\end{aligned}
$$

Because

$$
\begin{aligned}
\mathrm{d}\,\mathrm{D}_\rho\,\mathbf{P}(\rho) &= -(\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})(\mathrm{d}\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]) \mathrm{D}_\rho\,\mathbf{R}(\rho) \\
&- (\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)] \mathrm{d}\,\mathrm{D}_\rho\,\mathbf{R}(\rho) \\
&= -[\mathrm{D}_\rho\,\mathbf{R}(\rho)' \otimes (\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})] \mathrm{d} \operatorname{vec}\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)] \\
&- (\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)] \mathrm{d} \operatorname{vec} \mathrm{D}_\rho\,\mathbf{R}(\rho),
\end{aligned}
$$

this implies

$$
\begin{aligned}
\mathrm{D}^2_\rho\,\mathbf{P}(\rho) &= -[\mathrm{D}_\rho\,\mathbf{R}(\rho)' \otimes (\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})]\mathrm{D}_\rho\,[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)] \\
&- (\mathbf{I}_{T^2(T+1)^2/4} + \mathbf{K}_{T^2(T+1)^2/4})[\mathbf{R}(\rho)[\mathbf{R}(\rho)'\mathbf{R}(\rho)]^{-1} \otimes \mathbf{P}(\rho)]\mathrm{D}^2_\rho\,\mathbf{R}(\rho). \qquad \text{(A14)}
\end{aligned}
$$

The only term missing here is $\mathrm{D}^2_\rho\,\mathbf{R}(\rho)$, which, in view of the expression for $\mathrm{D}_\rho\,\mathbf{R}(\rho)$, is given by

$$
\mathrm{D}^2_\rho\,\mathbf{R}(\rho) = [\mathbf{I}_{r(r+1)/2} \otimes \mathbf{W}^{1/2}\mathbf{C}_1]\mathbf{C}_2\mathrm{D}^2_\rho\,\mathbf{\Gamma}(\rho),
$$

where

$$
\mathrm{D}^2_\rho\,\mathbf{\Gamma}(\rho) = \mathrm{vec}\begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \rho^{T-3} & \dots & 1 & 0 & 0 \end{bmatrix}.
$$

By using this and the fact that

$$
\mathrm{D}^2_\rho\,\hat{\mathbf{z}}_T(\rho) = -\mathrm{D}^2_\rho\,\mathbf{W}^{1/2}\hat{\mathbf{m}}^0 = \mathbf{0}_{T(T+1)/2\times 1},
$$

we can show that

$$
\begin{aligned}
&\mathrm{D}^2_\rho\,\mathbf{g}_C(\rho) \\
&= [\mathrm{D}_\rho\,\mathbf{P}(\rho) \otimes \mathbf{I}_{T(T+1)/2}](\mathbf{I}_{T(T+1)/2} \otimes \mathbf{K}_{T(T+1)/2} \otimes \mathbf{I}_{T(T+1)/2})[\mathrm{D}_\rho\,\hat{\mathbf{z}}_T(\rho) \otimes \mathrm{vec}\, I_{T(T+1)/2}] \\
&+ [\hat{\mathbf{Z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}^2_\rho\,\mathbf{P}(\rho) + [\mathrm{D}_\rho\,\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\,\mathbf{P}(\rho) + \mathbf{P}(\rho)\mathrm{D}^2_\rho\,\hat{\mathbf{z}}_T(\rho) \\
&= [\mathrm{D}_\rho\,\mathbf{P}(\rho) \otimes \mathbf{I}_{T(T+1)/2}](\mathbf{I}_{T(T+1)/2} \otimes \mathbf{K}_{T(T+1)/2} \otimes \mathbf{I}_{T(T+1)/2})[\mathrm{D}_\rho\,\hat{\mathbf{z}}_T(\rho) \otimes \mathrm{vec}\,\mathbf{I}_{T(T+1)/2}] \\
&+ [\hat{\mathbf{Z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}^2_\rho\,\mathbf{P}(\rho) + [\mathrm{D}_\rho\,\hat{\mathbf{z}}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\,\mathbf{P}(\rho). \qquad \text{(A15)}
\end{aligned}
$$

whose limit is given by

$$
\mathrm{D}^2_\rho\,\mathbf{g}_C(\rho) = \gamma_2(\rho) + o_p(1),
$$

where

$$
\begin{aligned}
&\gamma_2(\rho) \\
&= [\mathrm{D}_\rho\,\mathbf{P}(\rho) \otimes \mathbf{I}_{T(T+1)/2}](\mathbf{I}_{T(T+1)/2} \otimes \mathbf{K}_{T(T+1)/2} \otimes \mathbf{I}_{T(T+1)/2})[\mathrm{D}_\rho\,\mathbf{z}_T(\rho) \otimes \mathrm{vec}\,\mathbf{I}_{T(T+1)/2}] \\
&+ [\mathbf{z}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}^2_\rho\,\mathbf{P}(\rho) + [\mathrm{D}_\rho\,\mathbf{z}_T(\rho)' \otimes \mathbf{I}_{T(T+1)/2}]\mathrm{D}_\rho\,\mathbf{P}(\rho).
\end{aligned}
$$

The results for $\mathrm{D}_\rho\,\mathbf{g}_C(\rho)$ and $\mathrm{D}^2_\rho\,\mathbf{g}_C(\rho)$ imply, together with the fact that $\sqrt{N}\mathbf{g}_C(\rho_0) = O_p(1)$,

$$
\sqrt{N}\mathrm{D}_\rho\,Q_C(\rho_0) = 2\sqrt{N}\mathbf{g}_C(\rho_0)'\mathrm{D}_\rho\,\mathbf{g}_C(\rho_0) \xrightarrow{d} N(0, 4\gamma_1'\mathbf{\Sigma}_{gc}\gamma_1)
$$

as $N \to \infty$, where $\gamma_1 = \gamma_1(\rho_0)$. By using this and

$$\begin{aligned}
\mathrm{D}_\rho^2\, Q_C(\rho_0) &= 2\mathbf{g}_C(\rho_0)'\mathrm{D}_\rho^2\, \mathbf{g}_C(\rho_0) + 2\mathrm{D}_\rho\, \mathbf{g}_C(\rho_0)'\mathrm{D}_\rho\, \mathbf{g}_C(\rho_0) \\
&= 2\mathrm{D}_\rho\, \mathbf{g}_C(\rho_0)'\mathrm{D}_\rho\, \mathbf{g}_C(\rho_0) + o_p(1) \\
&= 2\gamma_1'\gamma_1 + o_p(1), \qquad \text{(A16)}
\end{aligned}$$

we obtain

$$\sqrt{N}(\hat{\rho} - \rho_0) = -\frac{\sqrt{N}\mathrm{D}_\rho\, Q_C(\rho_0)}{\mathrm{D}_\rho^2\, Q_C(\rho_0)} + o_p(1) \xrightarrow{d} \frac{\sqrt{\gamma_1'\mathbf{\Sigma}_{gC}\gamma_1}}{\gamma_1'\gamma_1} N(0,1),$$

as was to be shown. Note also that, letting $\gamma_2 = \gamma_2(\rho_0)$,

$$\begin{aligned}
\mathrm{D}_\rho^3\, Q_C(\rho_0) &= 4\mathbf{g}_C(\rho_0)'\mathrm{D}_\rho^3\, \mathbf{g}_C(\rho_0) + 6\mathrm{D}_\rho\, \mathbf{g}_C(\rho_0)'\mathrm{D}_\rho^2\, \mathbf{g}_C(\rho_0) = 6\mathrm{D}_\rho\, \mathbf{g}_C(\rho_0)'\mathrm{D}_\rho^2\, \mathbf{g}_C(\rho_0) + o_p(1) \\
&= 6\gamma_1'\gamma_2 + o_p(1),
\end{aligned}$$

verifying that $\mathrm{D}_\rho^3\, Q_C(\rho_0)$ is indeed $O_p(1)$. ■

Table 1: Monte Carlo results for the case when $\varepsilon_{i,t} \sim N(0,1)$.

		GMM1				GMM2					
ρ_0	N	Mean	SD	RMSE	$t(1)$	Mean	SD	RMSE	$t(1)$	J	BIC1
					$\sigma_\delta^2 = 0.5, r_0 = 1$						
1	100	.9977	.0219	.0220	.051	.9989	.0230	.0230	.109	.056	.939
1	400	.9993	.0096	.0096	.049	.9998	.0092	.0092	.063	.069	.968
1	1600	1.000	.0048	.0048	.048	1.000	.0042	.0042	.056	.067	.996
.99	100	.9875	.0229	.0230	.117	.9885	.0258	.0258	.302	.057	.948
.99	400	.9908	.0146	.0146	.220	.9907	.0133	.0133	.329	.069	.941
.99	1600	.9909	.0071	.0071	.432	.9905	.0064	.0064	.514	.067	.982
.95	100	.9471	.0248	.0250	.795	.9487	.0263	.0264	.892	.055	.962
.95	400	.9491	.0115	.0116	.982	.9498	.0109	.0109	.994	.068	.979
.95	1600	.9500	.0057	.0057	.998	.9500	.0052	.0052	1.000	.064	.996
					$\sigma_\delta^2 = 2.5, r_0 = 1$						
1	100	1.010	.0203	.0226	.065	1.006	.0222	.0223	.102	.037	.944
1	400	1.002	.0093	.0093	.059	1.001	.0089	.0089	.076	.041	.973
1	1600	1.000	.0041	.0041	.061	1.000	.0037	.0037	.073	.044	1.00
.99	100	.9915	.0226	.0227	.146	.9908	.0220	.0220	.313	.035	.974
.99	400	.9911	.0101	.0102	.345	.9910	.0096	.0097	.447	.042	.999
.99	1600	.9907	.0049	.0051	.518	.9906	.0039	.0040	.792	.051	.998
.95	100	.9531	.0204	.0206	.517	.9528	.0238	.0239	.684	.031	.974
.95	400	.9521	.0101	.0103	.876	.9506	.0095	.0095	.959	.052	.983
.95	1600	.9504	.0049	.0049	.997	.9509	.0044	.0045	1.00	.055	1.00
					$\sigma_\delta^2 = 0.5, r_0 = 2$						
1	100	.9929	.0263	.0272	.077	.9926	.0279	.0289	.128	.040	.910
1	400	.9969	.0105	.0110	.055	.9977	.0102	.0105	.074	.055	.963
1	1600	.9996	.0058	.0058	.060	1.000	.0051	.0051	.058	.053	.984
.99	100	.9979	.0268	.0279	.047	.9949	.0256	.0261	.081	.033	.927
.99	400	.9869	.0151	.0154	.251	.9891	.0119	.0119	.293	.055	.963
.99	1600	.9895	.0081	.0081	.610	.9901	.0076	.0076	.650	.053	.996
.95	100	.9685	.0333	.0397	.331	.9655	.0312	.0397	.482	.035	.896
.95	400	.9581	.0132	.0155	.611	.9556	.0124	.0137	.779	.040	.993
.95	1600	.9486	.0059	.0061	.969	.9505	.0052	.0052	.985	.054	.999
					$\sigma_\delta^2 = 2.5, r_0 = 2$						
1	100	.9986	.0249	.0249	.071	.9990	.0253	.0253	.111	.039	.948
1	400	.9987	.0120	.0121	.065	.9992	.0118	.0118	.067	.041	.971
1	1600	.9998	.0051	.0051	.056	.9997	.0045	.0045	.065	.042	1.00
.99	100	.9945	.0231	.0235	.081	.9962	.0233	.0241	.130	.047	.938
.99	400	.9935	.0242	.0245	.298	.9918	.0116	.0117	.213	.045	.987
.99	1600	.9905	.0069	.0069	.692	.9903	.0064	.0064	.749	.056	.998
.95	100	.9556	.0310	.0315	.312	.9551	.0254	.0259	.464	.032	.946
.95	400	.9527	.0228	.0230	.631	.9518	.0123	.0124	.868	.048	.988
.95	1600	.9501	.0115	.0115	.865	.9508	.0070	.0070	.997	.054	.999

Notes: "GMM1" and "GMM2" refer to the one- and two-step GMM estimators, "Mean", "SD" and "RMSE" refer to the mean standard deviation, and root mean squared error of the estimators, "$t(1)$" and "J" refer to the rejection frequency of the unit root t-statistic and Hansen–Sargan statistics, and "BIC1" refers to the correct selection frequency of the BIC1, respectively.

Table 2: Monte Carlo results for the case when $\varepsilon_{i,t}$ follows an MA(1) process.

		GMM1				GMM2					
ρ_0	N	Mean	SD	RMSE	$t(1)$	Mean	SD	RMSE	$t(1)$	J	BIC1
					$\sigma_\delta^2 = 0.5, r_0 = 1$						
1	100	.9964	.0257	.0260	.041	.9971	.0251	.0253	.091	.051	.912
1	400	.9991	.0136	.0136	.053	.9991	.0131	.0131	.057	.071	.959
1	1600	1.005	.0068	.0084	.047	1.004	.0057	.0070	.051	.063	.981
.99	100	.9871	.0274	.0276	.101	.9869	.0271	.0273	.286	.051	.963
.99	400	.9910	.0173	.0173	.194	.9887	.0162	.0163	.311	.063	.952
.99	1600	.9903	.0099	.0099	.417	.9897	.0085	.0085	.519	.066	.979
.95	100	.9473	.0284	.0285	.761	.9469	.0291	.0293	.912	.059	.958
.95	400	.9496	.0153	.0153	.967	.9479	.0140	.0142	.983	.061	.976
.95	1600	.9510	.0084	.0085	.983	.9505	.0080	.0080	1.000	.054	.994
					$\sigma_\delta^2 = 2.5, r_0 = 1$						
1	100	1.013	.0231	.0284	.060	1.011	.0240	.0241	.083	.032	.924
1	400	1.008	.0113	.0139	.054	1.006	.0097	.0097	.068	.039	.962
1	1600	1.003	.0062	.0069	.058	1.009	.0042	.0042	.061	.048	.994
.99	100	.9924	.0260	.0261	.139	.9937	.0243	.0246	.284	.038	.967
.99	400	.9917	.0145	.0146	.340	.9924	.0129	.0131	.419	.041	.985
.99	1600	.9911	.0081	.0082	.501	.9912	.0067	.0069	.804	.056	1.00
.95	100	.9538	.0260	.0263	.525	.9542	.0248	.0252	.648	.037	.961
.95	400	.9526	.0133	.0137	.861	.9523	.0115	.0117	.975	.042	.968
.95	1600	.9511	.0069	.0072	.984	.9513	.0058	.0059	.996	.043	.989
					$\sigma_\delta^2 = 0.5, r_0 = 2$						
1	100	.9932	.0298	.0306	.071	.9944	.0304	.0309	.101	.047	.894
1	400	.9956	.0141	.0148	.059	.9958	.0135	.0141	.068	.059	.957
1	1600	.9991	.0077	.0078	.061	.9987	.0064	.0064	.053	.054	.979
.99	100	.9864	.0317	.0319	.062	.9851	.0314	.0318	.073	.033	.924
.99	400	.9872	.0204	.0206	.236	.9876	.0194	.0196	.249	.037	.954
.99	1600	.9887	.0116	.0117	.601	.9889	.0106	.0107	.641	.042	.974
.95	100	.9445	.0341	.0345	.319	.9441	.0332	.0337	.467	.037	.906
.95	400	.9479	.0168	.0169	.631	.9476	.0159	.0161	.762	.041	.984
.95	1600	.9486	.0109	.0110	.954	.9491	.0093	.0093	.977	.046	.993
					$\sigma_\delta^2 = 2.5, r_0 = 2$						
1	100	1.019	.0262	.0324	.069	1.023	.0270	.0355	.093	.033	.936
1	400	1.013	.0129	.0183	.061	1.016	.0114	.0197	.062	.037	.959
1	1600	1.006	.0069	.0091	.058	1.004	.0059	.0071	.055	.041	.994
.99	100	.9927	.0284	.0285	.082	.9944	.0294	.0270	.104	.037	.942
.99	400	.9921	.0178	.0179	.274	.9931	.0162	.0165	.195	.042	.975
.99	1600	.9915	.0094	.0095	.668	.9916	.0081	.0083	.716	.051	.997
.95	100	.9576	.0309	.0318	.305	.9578	.0304	.0314	.469	.037	.959
.95	400	.9541	.0144	.0150	.657	.9547	.0140	.0148	.832	.043	.991
.95	1600	.9528	.0093	.0097	.926	.9521	.0081	.0084	.983	.041	1.00

Notes: See table 1 for an explanation.

Table 3: Empirical results.

Measure	$\hat{\rho}$	$\hat{\sigma}_\rho$	$t(1)$	p-value	J	p-value	BIC1	$\hat{r}$	q
EMP	.879	.091	-1.32	.186	10.23	.176	4.26	2	1
FA	.843	.105	-1.49	.135	11.91	.104	5.93	1	2

Notes: "$\hat{\rho}$" and "$\hat{\sigma}_\rho$" refer to the GMM2 estimator of ρ_0 and its estimated standard error, "$t(1)$" 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.