ESTIMATION AND INFERENCE IN SHORT PANEL VECTOR AUTOREGRESSIONS WITH UNIT ROOTS AND COINTEGRATION

Michael Binder, Cheng Hsiao
and M. Hashem Pesaran

Banco de España - Servicio de Estudios
Documento de Trabajo nº 0005

ESTIMATION AND INFERENCE IN SHORT PANEL VECTOR AUTOREGRESSIONS WITH UNIT ROOTS AND COINTEGRATION (*)

Michael Binder, Cheng Hsiao
and M. Hashem Pesaran

(*) We are grateful to Manuel Arellano, Jinyong Hahn, Ingmar Prucha, and Peter Schmidt for helpful comments. The first author also gratefully acknowledges financial support though a Bank of Spain research fellowship. Correspondence addresses are: Department of Economics, University of Maryland, Tydings Hall, College Park, MD 20742; Department of Economics, University of Southern California, University Park, Los Angeles, CA. 90089; Faculty of Economics and Politics, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DD.

Banco de España - Servicio de Estudios
Documento de Trabajo nº 0005

In publishing this series the Banco de España seeks to disseminate studies of interest that will help acquaint readers better with the Spanish economy.

The analyses, opinions and findings of these papers represent the views of their authors; they are not necessarily those of the Banco de España.

The Banco de España disseminates its main reports and most of its publications via the INTERNET at the following website: http://www.bde.es

ISSN: 0213-2710
ISBN: 84-7793-708-7
Depósito legal: M. 26034-2000
Imprenta del Banco de España

Abstract

This paper considers estimation and inference in panel vector autoregressions (PVARs) with fixed effects when the time dimension of the panel is finite, and the cross-sectional dimension is large. A Maximum Likelihood (ML) estimator based on a transformed likelihood function is proposed and shown to be consistent and asymptotically normally distributed irrespective of the unit root and cointegrating properties of the underlying PVAR model. The transformed likelihood framework is also used to derive unit root and cointegration tests in panels with short time dimension; these tests have the attractive feature that they are based on standard chi-square and normal distributed statistics. Examining Generalized Method of Moments (GMM) estimation as an alternative to our proposed ML estimator, it is shown that conventional GMM estimators based on standard orthogonality conditions break down if the underlying time series contain unit roots. Also, the implementation of extended GMM estimators making use of variants of homoskedasticity and stationarity restrictions as suggested in the literature in a univariate context is subject to difficulties. Monte Carlo evidence is adduced suggesting that the ML estimator and parameter hypothesis and cointegration tests based on it perform well in small sample; this is in marked contrast to the small sample performance of the GMM estimators.

1 Introduction

Vector autoregressive models (VARs), whether postulated as reduced form econometric models or built to examine long-run restrictions suggested by economic theory while allowing for unrestricted short-run dynamics have become a widely used modelling tool in economics (for example, Hsiao, 1979, 1982, Sims, 1980, King, Plosser, Stock and Watson, 1991, Pesaran and Shin, 1999). It is well known, though, that time-series inference on VARs critically depends on whether the underlying processes are (trend) stationary or integrated, or cointegrated, and, if they are cointegrated, the rank of cointegration (for example, Sims, Stock, and Watson, 1990, Phillips, 1991, Johansen, 1995, and Pesaran, Shin, and Smith, 2000).

It is also widely recognized that panel data have the attractive feature that often a large number of cross-sectional observations are available over time. The application of VARs to panel data is often hampered by two issues, however: (i) The time dimension of the panel may be short. (ii) Cross-sectional heterogeneity is present. The first issue gives raise to the problem of modeling the initial observations (for example, Anderson and Hsiao, 1981, 1982, Bhargava and Sargan, 1983, Blundell and Smith, 1991, and Nerlove, 1999). The second issue gives raise to the classical incidental parameters problem, Neyman and Scott (1948), if the cross-sectional heterogeneity is modelled through fixed effects. Both issues can lead to the breakdown of the classical Maximum Likelihood (ML) estimator.

Because of the difficulties with implementing likelihood based inference, Generalized Method of Moments (GMM) estimators have been popular for the estimation of (predominantly univariate) dynamic panel data models (for example, Holtz-Eakin, Newey, and Rosen, 1988, Arellano and Bond, 1991, Ahn and Schmidt, 1995, and Arellano and Bover, 1995). However, conventional GMM estimators based on standard orthogonality conditions have proved to have rather poor finite sample properties in the case of autoregressive panel data models with near unit roots (for example, Blundell and Bond, 1998).[1] In this paper we show that for Panel VARs (PVARs), conventional GMM estimators in fact break down altogether if the underlying time series contain unit roots. We then investigate the possibility of obtaining GMM estimators that could be applied to PVAR models when the unit root properties of the model are not known *a priori*. To this end we augment the standard orthogonality conditions with homoskedasticity and stationarity implied moment conditions proposed initially by Ahn and Schmidt (1995) and Arellano and Bover (1995) in the context of univariate dynamic panel data models. Unfortunately, the implementation of extended GMM estimators incorporating variants of these additional moment conditions is subject to difficulties.

[1] Monte Carlo experiments reported in Blundell and Bond (1998), Hsiao, Pesaran, and Tahmisciogl̆u (1999), and in Section 10 below suggest that the finite sample properties of conventional GMM estimators of autoregressive panel data models are poor even in the case of models with moderately large sized roots. Conventional GMM estimation of autoregressive panel data models thus tends to be problematic for a wide range of applications.

Given the problems with GMM estimation, in this paper we propose a likelihood based estimation and inference approach. We derive a transformed unconditional likelihood function for PVARs with fixed effects that does not involve incidental parameters and provides a model-consistent way of formulating the distribution of the initial observations. We show that the ML estimator associated with our transformed likelihood function always exists and is (as the cross-sectional dimension of the panel, N, tends to infinity) consistent and asymptotically normally distributed, irrespective of whether the underlying time series are (trend) stationary, integrated of order one, I(1), or I(1) and cointegrated. We also suggest a consistent Minimum Distance (MD) estimator derived from the transformed likelihood framework to be used as the initial estimator for the ML iterations. Furthermore, we consider hypothesis testing based on the transformed likelihood function. In particular, Wald- and likelihood ratio-based panel testing procedures for the presence of unit roots and for the cointegration rank are derived. 'An attractive feature of these tests is that they are based on standard chi-square and normal distributed statistics. We also show how our likelihood based estimation and inference approach can be extended to PVARs where under the long-run forcing restriction a subset of the variables is modelled in terms of the remaining variables. Finally, some Monte Carlo evidence regarding the small sample biases and root mean square errors of the conventional GMM, the MD, and the ML estimators is adduced. The Monte Carlo simulations are also employed to investigate the size and power properties of parameter hypothesis tests based on the conventional GMM and the ML estimators, and to investigate the size and power properties of cointegration rank tests based on the ML estimator. The Monte Carlo evidence suggests that the ML estimator in small sample performs remarkably well, clearly outperforming the conventional GMM estimators, and that the various ML estimator based hypothesis tests tend to have good size and power properties. The Monte Carlo simulations also further document difficulties with the implementation of extended GMM estimation.

The remainder of this paper is organized as follows: Section 2 introduces the PVAR under fixed effects. Section 3 discusses how the incidental parameters problem arising in the estimation of this model may be resolved by working with transformed model equations, and provides a general expression for the unconditional joint probability distribution of the first differences of the observations. Section 4 considers GMM estimation based on the transformed model equations, showing that conventional GMM estimators break down in the presence of unit roots, and demonstrating that the use of homoskedasticity and stationarity implied moment conditions to construct an extended GMM estimator is subject to problems of its own. Section 5 presents the proposed ML estimator, showing that it is consistent and asymptotically normally distributed irrespective of whether the underlying time series are (trend) stationary, pure I(1), or I(1) and cointegrated. Section 6 derives a MD estimator that we use as the initial estimator for the ML iterations. Section 7 discusses the estimation of long-run relations within the PVAR framework, and examines model

identification when the variance-covariance matrix entering the transformed likelihood function is treated as unrestricted. Hypothesis testing on the basis of the transformed likelihood function is discussed in Section 8, which in particular proposes new tests for unit roots and cointegration rank in panels with short time dimension. Section 9 considers the issue of how to carry out inference in conditional models where a subset of the variables in the PVAR is modelled in terms of the remaining variables. In Section 10, Monte Carlo simulation results are reported assessing the finite sample performance of the GMM, MD, and ML estimators. Section 11 concludes and provides some suggestions for future research. The extension of formulae derived in the body of the paper for first-order PVARs to p-th order PVARs is described in an appendix.

2 A PVAR Under Fixed Effects

Let $\mathbf{w}_{it}$ be an $m \times 1$ vector of random variables for the i-th cross-sectional unit at time t. We assume that the $\mathbf{w}_{it}$'s are generated by the following PVAR model of order p, PVAR(p):

$$\Phi(L)(\mathbf{w}_{it} - \boldsymbol{\mu}_i - \boldsymbol{\gamma} t) = \boldsymbol{\varepsilon}_{it}, \qquad i = 1, 2, \ldots, N, \tag{2.1}$$

where

$$\Phi(L) = \mathbf{I}_m - \sum_{j=1}^{p} \Phi_j L^j, \tag{2.2}$$

Φ_j, $j = 1, 2, \ldots, p$, denote $m \times m$ matrices of slope coefficients, $\boldsymbol{\mu}_i$ is an $m \times 1$ vector of fixed effects,[2] $\boldsymbol{\gamma}$ is an $m \times 1$ vector providing the common deterministic trends/drifts across cross-sectional units,[3] $\boldsymbol{\varepsilon}_{it}$ is an $m \times 1$ vector of disturbances, $\mathbf{I}_m$ denotes the identity matrix of dimension $m \times m$, and L the lag operator, $L\mathbf{w}_{it} \equiv \mathbf{w}_{i,t-1}$. We make the following assumptions:

(A1) The disturbance vectors $\boldsymbol{\varepsilon}_{it}$ are independently and identically distributed across i and over t with $E(\boldsymbol{\varepsilon}_{it}) = \mathbf{0}_{m \times 1}$ and $Var(\boldsymbol{\varepsilon}_{it}) = \Omega$, Ω being a positive definite matrix.

(A1) is a common assumption in the dynamic panel data literature as reviewed by Arellano and Honoré (2000). The coefficients making up the vectors $\boldsymbol{\mu}_i$, $i = 1, 2, \ldots, N$, and $\boldsymbol{\gamma}$, and the matrices Φ_j, $j = 1, 2, \ldots, p$, and Ω are assumed to be unknown.

(A2) The time dimension of the panel is finite with $T > p$, and the available observations are $\mathbf{w}_{i0}$, $\mathbf{w}_{i1}$, $\ldots$, $\mathbf{w}_{iT}$.

[2] As a variety of fixed effects specifications have been considered in the literature, it may be worthwhile to spell out what we mean by fixed effects in this paper. We view the individual-specific effects, $\boldsymbol{\mu}_i$, as fixed effects if they are generated by a probability distribution function where the number of parameters characterizing this function is allowed to increase at the same rate as the number of cross-sectional observations in the panel. As will be seen below, for our proposed ML estimator to be valid, no further assumptions on the $\boldsymbol{\mu}_i$'s are needed.

[3] The common deterministic trend/drift $\boldsymbol{\gamma} t$ in (2.1) could be replaced by more general time-specific effects $\boldsymbol{\gamma}_t$ without adding any conceptual problems for estimation and inference.

(A3) The roots of the determinantal equation

$$|\Phi(\varrho)| = 0 \tag{2.3}$$

are either equal to unity or fall outside the unit circle.

As our interest in this paper concerns (trend-) stationary, pure unit root, and cointegrated PVAR models, the PVAR(p) model (2.1) ensures that if $\boldsymbol{\gamma} = \mathbf{0}_{m\times 1}$, then the model intercepts are restricted as otherwise $\{\mathbf{w}_{it}\}$ would exhibit no trend growth across i if all roots of the determinantal equation (2.3) fell outside the unit circle, and $\{\mathbf{w}_{it}\}$ (or some of its linear combinations) would exhibit differential trend growth across i if one or more roots of the determinantal equation (2.3) were equal to unity. Similarly, the PVAR(p) model (2.1) ensures that if $\boldsymbol{\gamma} \neq \mathbf{0}_{m\times 1}$, then the trend coefficients are restricted as otherwise $\{\mathbf{w}_{it}\}$ (or some of its linear combinations) would exhibit quadratic trends if one or more roots of the determinantal equation (2.3) were equal to unity.[4]

A number of well known time-series specifications can be obtained from the PVAR(p) model (2.1) as special cases. We set out some of these for the reader's convenience.

Case 1: Stationary PVAR with fixed effects

Let $\boldsymbol{\gamma} = \mathbf{0}_{m\times 1}$, and let all roots of the determinantal equation (2.3) fall outside the unit circle. Then (2.1) becomes

$$\Phi(L)\,\mathbf{w}_{it} = \mathbf{b}_{i0} + \boldsymbol{\varepsilon}_{it}, \qquad i = 1, 2, \ldots, N, \tag{2.4}$$

where $\mathbf{b}_{i0} = -\Pi\boldsymbol{\mu}_i$, with

$$\Pi = -\left(\mathbf{I}_m - \sum_{j=1}^{p}\Phi_j\right). \tag{2.5}$$

Case 2: Trend-stationary PVAR with fixed effects

Let all roots of the determinantal equation (2.3) fall outside the unit circle and $\boldsymbol{\gamma} \neq \mathbf{0}_{m\times 1}$. Then (2.1) becomes

$$\Phi(L)\,\mathbf{w}_{it} = \mathbf{a}_{i0} + \mathbf{a}_1 t + \boldsymbol{\varepsilon}_{it}, \qquad i = 1, 2, \ldots, N, \tag{2.6}$$

where $\mathbf{a}_{i0} = -\Pi\boldsymbol{\mu}_i + (\Gamma + \Pi)\,\boldsymbol{\gamma}$, with

$$\Gamma = -\Pi + \sum_{j=1}^{p} j\Phi_j, \tag{2.7}$$

[4]See Pesaran, Shin, and Smith (2000) for a further discussion of this issue in a time-series context. It may also be worth noting that while we do not pursue the possibility of incidental deterministic trends/drifts, $\boldsymbol{\gamma}_i$, in this paper, these could be accommodated by working with the unconditional joint probability distribution function of the second differences of the observations. This will become clear in Section 3 below.

and where $\mathbf{a}_1 = -\Pi\gamma$.

Case 3: PVAR with unit roots (but non-cointegrated) and fixed effects
In this case $\sum_{j=1}^{p} \Phi_j = \mathbf{I}_m$. Then (2.1) becomes

$$\Phi^* (L) \triangle \mathbf{w}_{it} = -\Pi^* \gamma + \varepsilon_{it}, \qquad i = 1, 2, \ldots, N, \tag{2.8}$$

where $\triangle \equiv 1 - L$,

$$\Phi^* (L) = \mathbf{I}_m - \sum_{j=1}^{p-1} \Phi_j^* L^j, \tag{2.9}$$

with $\Phi_j^* = -\left(\mathbf{I}_m - \sum_{l=1}^{j} \Phi_l\right)$, $j = 1, 2, \ldots, p-1$, and where $\Pi^* = -\left(\mathbf{I}_m - \sum_{j=1}^{p-1} \Phi_j^*\right)$.

Case 4: Cointegrated PVAR with fixed effects
In the case where some (but not all) roots of the determinantal equation (2.3) are equal to unity, $rank\,(\Pi) = r$, $0 < r < m$, and (2.1) may be rewritten in the form of a panel vector error correction model as

$$\triangle \mathbf{w}_{it} = \Gamma\gamma - (p-1)\,\Pi\gamma + \Pi\xi_{i,t-p} + \sum_{j=1}^{p-1} \Gamma_j \triangle \mathbf{w}_{i,t-j} + \varepsilon_{it}, \qquad i = 1, 2, \ldots, N, \tag{2.10}$$

where

$$\Gamma_j = -\mathbf{I}_m + \sum_{l=1}^{j} \Phi_l, \qquad j = 1, 2, \ldots, p-1. \tag{2.11}$$

For ease of exposition, in what follows we shall motivate our remaining assumptions to be made for estimation and inference on the PVAR(p) model (2.1) first for the case where $p = 1$, and then generalize our discussion to the case where $p \geq 1$. Under $p = 1$, the PVAR model (2.1) reduces to

$$(\mathbf{I}_m - \Phi L)\,(\mathbf{w}_{it} - \mu_i - \gamma t) = \varepsilon_{it}, \qquad i = 1, 2, \ldots, N. \tag{2.12}$$

When the time dimension of the panel, T, is finite, in estimating (2.12) there arises a classical incidental parameters problem as the number of fixed effects vectors, μ_i, to be estimated increases proportionately to the cross-sectional dimension of the panel, and there arises the problem of the initialization of $\{\mathbf{w}_{it}\}$. Our formulation of the initial observation in the sample, $\mathbf{w}_{i0}$, makes use of the fact that the PVAR(1) model in ξ_{it}, defined by

$$\xi_{it} = \mathbf{w}_{it} - \mu_i - \gamma_t, \tag{2.13}$$

does not involve the fixed effects, μ_i, although it is still subject to the initial value problem:

$$(\mathbf{I}_m - \Phi L)\,\xi_{it} = \varepsilon_{it}, \qquad i = 1, 2, \ldots, N. \tag{2.14}$$

Suppose that the PVAR(1) model (2.12) started at time $t = -M$, $M \geq 0$. Then from (2.14) we obtain

$$\boldsymbol{\xi}_{it} = \Phi^{t+M}\boldsymbol{\xi}_{i,-M} + \sum_{j=0}^{t+M-1} \Phi^{j}\boldsymbol{\varepsilon}_{i,t-j}, \qquad t = -M+1, -M+2, \ldots, T. \tag{2.15}$$

In practice where the time-series properties of $\{\boldsymbol{\xi}_{it}\}$ are not known *a priori*, to characterize the distribution of $\{\boldsymbol{\xi}_{it}\}_{t=-M+1}^{T}$ we need to distinguish between three main cases; namely $\{\boldsymbol{\xi}_{it}\}$ being covariance stationary, pure I(1), or I(1) and cointegrated.

Suppose that the initial deviations, $\boldsymbol{\xi}_{i,-M}$, are independently and identically distributed across i with mean vector zero and a finite variance-covariance matrix. In the case where all roots of the determinantal equation (2.3) fall outside the unit circle, it is then clear from (2.15) and (*A*1) that the variance-covariance matrix of $\boldsymbol{\xi}_{it}$ will also be finite over the sample period, $t = 0, 1, \ldots, T$, irrespective of whether

(*i*) $\{\boldsymbol{\xi}_{it}\}$ has been in operation for a long period of time, namely $M \to \infty$, and the process has reached covariance stationarity, so that $E(\boldsymbol{\xi}_{it}) = \mathbf{0}_{m\times 1}$ and $Var(\boldsymbol{\xi}_{it}) = \sum_{j=0}^{\infty} \Phi^{j}\Omega\Phi^{j\prime}$, $t = 0, 1, \ldots, T$,

or

(*ii*) $\{\boldsymbol{\xi}_{it}\}$ has started in a finite period in the past not too far back from $t = 0$.

We shall pursue possibility (*i*) in this paper;[5] and note that under (*i*) (2.15) becomes

$$\boldsymbol{\xi}_{it} = \sum_{j=0}^{\infty} \Phi^{j}\boldsymbol{\varepsilon}_{i,t-j}, \qquad t \leq T. \tag{2.16}$$

In the case where all roots of the determinantal equation (2.3) are equal to unity, to ensure that the variance-covariance matrix of $\boldsymbol{\xi}_{it}$ is finite over the sample period, $t = 0, 1, \ldots, T$, $\{\boldsymbol{\xi}_{it}\}$ must have started in a finite period in the past. It follows that (2.15) in the pure unit root case becomes

$$\boldsymbol{\xi}_{it} = \boldsymbol{\xi}_{i,-M} + \sum_{j=0}^{t+M-1} \boldsymbol{\varepsilon}_{i,t-j}, \qquad t = -M+1, -M+2, \ldots, T. \tag{2.17}$$

It remains to consider the intermediate case where some (but not all) roots of the determinantal equation (2.3) are equal to unity, and the remaining ones fall outside the unit circle. For this case, it is helpful to work with the error correction representation of the PVAR(1) model (2.12) given by

$$\Delta \mathbf{w}_{it} = \boldsymbol{\gamma} + \Pi\boldsymbol{\xi}_{i,t-1} + \boldsymbol{\varepsilon}_{it}, \qquad t = -M+1, -M+2, \ldots, T, \tag{2.18}$$

where

$$\Pi = -(\mathbf{I}_m - \Phi). \tag{2.19}$$

[5] See Hsiao, Pesaran, and Tahmiscioglu (1999) for an analysis (in a univariate context) of possibility (*ii*).

As is well known from the cointegration literature, if $rank\,(\Pi) = r$, $r = 1, 2, \ldots, m-1$, then there exist $m \times r$ matrices α and β such that

$$\Pi = \alpha\beta', \tag{2.20}$$

where α and β have full column rank, and are commonly referred to as the error correction coefficients and the cointegrating vectors, respectively. The following assumption is standard in the cointegration literature:[6]

(A4) If $rank\,(\Pi) = r$, $r = 1, 2, \ldots, m-1$, *then* $\alpha'_\perp \beta_\perp$ *is of rank* $m - r$, *where* $\alpha_\perp$ *and* $\beta_\perp$ *are* $m \times (m-r)$ *matrices of full column rank such that* $\alpha'\alpha_\perp = 0_{r\times(m-r)}$ *and* $\beta'\beta_\perp = 0_{r\times(m-r)}$.

Under Assumptions (*A*3) and (*A*4) the elements of ξ_{it} are either I(0) or I(1). To ensure that the variance-covariance matrix of ξ_{it} is finite over the sample period, $t = 0, 1, \ldots, T$, the $m - r$ pure unit root components in $\{\xi_{it}\}$ must have started in a finite period in the past. To separate the pure unit root components in $\{\xi_{it}\}$ from the covariance-stationary ones, it will be useful to introduce the matrix $\mathbf{C}$ defined as

$$\mathbf{C} = \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp. \tag{2.21}$$

It is well known from the cointegration literature that under $rank\,(\Pi) = r$, $r = 0, 1, \ldots, m$, the matrix $\mathbf{C}$ has rank $m - r$, and there are $m - r$ pure unit root components in ξ_{it}, $t = -M+1, -M+2, \ldots, T$.[7] For consistency with (2.16), the r cointegrating relations $\beta'\xi_{it}$, $t = -M+1, -M+2, \ldots, T$, must have reached covariance stationarity. The following final assumption ensures both that the variance-covariance matrix of ξ_{it} is finite over the sample period, and that the cointegrating relations $\beta'\xi_{it}$ have reached covariance stationarity, irrespective of the number of pure unit root components in $\{\xi_{it}\}$.

(A5) The initial deviations $\xi_{i,-M}$ *are given by*

$$\xi_{i,-M} = \sum_{j=0}^{\infty} \left(\Phi^j - \mathbf{C}\right) \varepsilon_{i,-M-j} + \mathbf{C}\upsilon_i, \tag{2.22}$$

where ε_{it}, $i = 1, 2, \ldots, N$; $t \leq T$, *are independently and identically distributed across* i *and over* t, *and* υ_i, $i = 1, 2, \ldots, N$, *are independently and identically distributed across* i, *with*

$$E\begin{pmatrix} \varepsilon_{it} \\ \upsilon_i \end{pmatrix} = 0_{2m\times 1}, \quad \textit{and} \quad Var\begin{pmatrix} \varepsilon_{it} \\ \upsilon_i \end{pmatrix} = \begin{pmatrix} \Omega & \mathfrak{L} \\ \mathfrak{L}' & \mathfrak{F} \end{pmatrix}, \tag{2.23}$$

$\mathfrak{F}$ *being a positive definite matrix.*

[6] See, for example, Johansen (1995, Ch. 4).

[7] See, for example, Johansen (1995, Ch. 4).

Substituting (2.22) back into (2.15), and noting from the definition of $\mathbf{C}$ that $\mathbf{\Pi C} = \mathbf{0}_m$, and thus $\mathbf{C} = \mathbf{\Phi C}$, we have that

$$\boldsymbol{\xi}_{it} = \sum_{j=0}^{\infty} \left(\mathbf{\Phi}^j - \mathbf{C}\right) \boldsymbol{\varepsilon}_{i,t-j} + \mathbf{C} \left(\sum_{j=0}^{t+M-1} \boldsymbol{\varepsilon}_{i,t-j} \right) + \mathbf{C}\mathfrak{v}_i, \tag{2.24}$$

$t = -M+1, -M+2, \ldots, T$, and therefore also

$$\boldsymbol{\beta}'\boldsymbol{\xi}_{it} = \boldsymbol{\beta}' \sum_{j=0}^{\infty} \mathbf{\Phi}^j \boldsymbol{\varepsilon}_{i,t-j}, \tag{2.25}$$

$E\left(\boldsymbol{\beta}'\boldsymbol{\xi}_{it}\right) = \mathbf{0}_{m\times 1}$, and $Var\left(\boldsymbol{\beta}'\boldsymbol{\xi}_{it}\right) = \boldsymbol{\beta}' \left(\sum_{j=0}^{\infty} \mathbf{\Phi}^j \mathbf{\Omega} \mathbf{\Phi}^{j\prime}\right) \boldsymbol{\beta}$, $t = -M+1, -M+2, \ldots, T$. It is thus seen that $(A5)$ indeed implies that the cointegrating relations $\boldsymbol{\beta}'\boldsymbol{\xi}_{it}$ have reached covariance stationarity, irrespective of the number of pure unit root components in $\{\boldsymbol{\xi}_{it}\}$. Furthermore, in the case where all roots of the determinantal equation (2.3) fall outside the unit circle, $\mathbf{C} = \mathbf{0}_m$, and (2.24) reduces to (2.16). In the case where all roots of (2.3) are equal to unity, $\mathbf{C} = \mathbf{\Phi} = \mathbf{I}_m$, and (2.24) reduces to (2.17), with $\boldsymbol{\xi}_{i,-M} = \mathfrak{v}_i$. Also note from (2.24) that since $\left\{\mathbf{\Phi}^j - \mathbf{C}\right\}_{j=0}^{\infty}$ is absolutely summable irrespective of the number of eigenvalues of $\mathbf{\Phi}$ that are equal to unity, the variance-covariance matrix of $\boldsymbol{\xi}_{it}$, $t = -M+1, -M+2, \ldots, T$, is finite. Furthermore, it is worth noting that there are $(m-r)(m-r+1)/2$ free parameters in the variance-covariance matrix of $\mathbf{C}\mathfrak{v}_i$, which corresponds directly to the number of pure unit root components in $\{\boldsymbol{\xi}_{it}\}$. Finally, since the pure unit root components in $\{\boldsymbol{\xi}_{it}\}$ have started in a finite period in the past, there are m individual-specific intercepts present in $\{\mathbf{w}_{it}\}$ irrespective of whether $\{\boldsymbol{\xi}_{it}\}$ is covariance stationary, pure I(1), or I(1) and cointegrated.

Our initialization of the PVAR(1) model (2.12) can be readily extended to the PVAR(p) model (2.1) for $p \geq 1$. It is now assumed that $\boldsymbol{\alpha}_{\perp}'\mathbf{\Gamma}\boldsymbol{\beta}_{\perp}$ is of rank $m-r$, with $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_{\perp}$, $\boldsymbol{\beta}$, and $\boldsymbol{\beta}_{\perp}$ defined as before, and that the initial p deviations in the sample, namely $\boldsymbol{\xi}_{it}$, $t = 0, 1, \ldots, p-1$, are given by

$$\boldsymbol{\xi}_{it} = \mathbf{C}^*(L)\boldsymbol{\varepsilon}_{it} + \mathbf{C}\mathbf{s}_{it} + \mathbf{C}\boldsymbol{\zeta}_i, \qquad i = 1, 2, \ldots, N;\ t = 0, 1, \ldots, p-1, \tag{2.26}$$

where $\mathbf{s}_{i0} = \mathbf{0}_{m\times 1}$, $\mathbf{s}_{it} = \sum_{l=1}^{t} \boldsymbol{\varepsilon}_{il}$, $\mathbf{C}^*(L) = \sum_{j=0}^{\infty} \mathbf{C}_j^* L^j$, $\mathbf{C}_0^* = \mathbf{I}_m - \mathbf{C}$, $\mathbf{C}_j^* = \mathbf{C}_j + \mathbf{C}_{j-1}^*$, $j \geq 1$, $\mathbf{C}(L) = \sum_{j=0}^{\infty} \mathbf{C}_j L^j$, $\mathbf{C} = \sum_{j=0}^{\infty} \mathbf{C}_j$, $\mathbf{C}_0 = \mathbf{I}_m$, $\mathbf{C}_1 = -(\mathbf{I}_m - \mathbf{\Phi}_1)$, $\mathbf{C}_j = \sum_{l=1}^{p} \mathbf{C}_{j-l}\mathbf{\Phi}_l$, $j \geq 2$ (with $\mathbf{C}_j = \mathbf{0}_m$ for $j < 0$),[8] and $\boldsymbol{\zeta}_i$ is independently distributed across i with $E(\boldsymbol{\zeta}_i) = \mathbf{0}_{m\times 1}$, and $Var(\boldsymbol{\zeta}_i) = \mathbb{F}$, $\mathbb{F}$ being a positive definite matrix. It is readily verified that when $p = 1$, the specification (2.26) of the initial sample deviations is equivalent to the one implied by $(A5)$, with $\boldsymbol{\zeta}_i = \mathfrak{v}_i + \sum_{j=0}^{M-1} \boldsymbol{\varepsilon}_{i,-j}$, and $\mathbb{F} = \mathfrak{F} + M(\mathfrak{Q} + \mathfrak{L})$. The specification (2.26) again ensures that the

[8] Johansen (1995, Ch. 4) discusses the properties of the matrix polynomials $\mathbf{C}(L)$ and $\mathbf{C}^*(L)$ in some detail, and in particular shows that both $\{\mathbf{C}_j\}_{j=0}^{\infty}$ and $\{\mathbf{C}_j^*\}_{j=0}^{\infty}$ are absolutely summable. See also Pesaran, Shin, and Smith (2000).

variance-covariance matrix of $\boldsymbol{\xi}_{it}$ is finite over the sample period, $t = 0, 1, \ldots, T$, and that the r cointegrating relations $\boldsymbol{\beta}'\boldsymbol{\xi}_{it}$, $t = 0, 1, \ldots, T$, have reached covariance stationarity, irrespective of the number of pure unit root components in $\{\boldsymbol{\xi}_{it}\}$. The same processes generating the covariance-stationary components of $\{\boldsymbol{\xi}_{it}\}_{t=p}^{T}$ also generate those of $\{\boldsymbol{\xi}_{it}\}_{t=0}^{p-1}$, and the same processes generating the pure unit root components of $\{\boldsymbol{\xi}_{it}\}_{t=p}^{T}$ also generate those of $\{\boldsymbol{\xi}_{it}\}_{t=0}^{p-1}$.

3 Resolving the Incidental Parameters Problem

To overcome the incidental parameters problem due to the fixed effects and the initial observations when T is finite, following the analysis in Hsiao, Pesaran, and Tahmiscioglu (1999) of a univariate dynamic panel data model we work with the unconditional joint probability distribution of the first differences $\Delta\mathbf{w}_{i1}, \Delta\mathbf{w}_{i2}, \ldots, \Delta\mathbf{w}_{ip}, \Delta\mathbf{w}_{i,p+1}, \ldots, \Delta\mathbf{w}_{iT}$, which we decompose as[9]

$$\Pr(\Delta\mathbf{w}_{i1}, \Delta\mathbf{w}_{i2}, \ldots, \Delta\mathbf{w}_{ip}, \Delta\mathbf{w}_{i,p+1}, \ldots, \Delta\mathbf{w}_{iT}) \tag{3.1}$$

$$= \Pr(\Delta\mathbf{w}_{i1}, \Delta\mathbf{w}_{i2}, \ldots, \Delta\mathbf{w}_{ip}) \cdot \Pr(\Delta\mathbf{w}_{i,p+1}, \Delta\mathbf{w}_{i,p+2}, \ldots, \Delta\mathbf{w}_{iT} | \Delta\mathbf{w}_{i1}, \Delta\mathbf{w}_{i2}, \ldots, \Delta\mathbf{w}_{ip}).$$

To obtain the first component on the right-hand side of (3.1), namely the unconditional joint probability distribution of $\Delta\mathbf{w}_{i1}, \Delta\mathbf{w}_{i2}, \ldots, \Delta\mathbf{w}_{ip}$, we invoke the specification (2.26) of the initial deviations in the sample. Combining the covariance-stationary representations in first differences of the observations dated $t = 1, 2, \ldots, p-1$ implied by (2.26) with the error correction representation of the PVAR for $t = p$ one obtains the unconditional joint probability distribution of $\Delta\mathbf{w}_{i1}, \Delta\mathbf{w}_{i2}, \ldots, \Delta\mathbf{w}_{ip}$ free of the incidental parameters and initial value problems. From (2.26) it is readily verified that the following first-difference representation holds for $t = 1, 2, \ldots, p-1$:

$$\Delta\mathbf{w}_{it} - \boldsymbol{\gamma} = \mathbf{C}(L)\,\boldsymbol{\varepsilon}_{it}, \qquad t = 1, 2, \ldots, p-1. \tag{3.2}$$

Also, substituting for $\boldsymbol{\xi}_{i0}$ from (2.26) into the error correction representation (2.10) for $t = p$, after re-arranging terms and noting that $\boldsymbol{\Pi}\mathbf{C} = \mathbf{0}_m$ one obtains that

$$\Delta\mathbf{w}_{ip} - \boldsymbol{\gamma} = \sum_{j=1}^{p-1} \boldsymbol{\Gamma}_j \left(\Delta\mathbf{w}_{i,p-j} - \boldsymbol{\gamma}\right) + \boldsymbol{\eta}_{ip}, \tag{3.3}$$

where

$$\boldsymbol{\eta}_{ip} = \boldsymbol{\Pi}\mathbf{C}^{*}(L)\,\boldsymbol{\varepsilon}_{i0} + \boldsymbol{\varepsilon}_{ip}. \tag{3.4}$$

Collecting the disturbance terms in (3.2) and (3.3),

$$\mathbf{e}_i = vec\left[\ \mathbf{C}(L)\,\boldsymbol{\varepsilon}_{i1},\ \ \mathbf{C}(L)\,\boldsymbol{\varepsilon}_{i2},\ \ \ldots,\ \ \mathbf{C}(L)\,\boldsymbol{\varepsilon}_{i,p-1},\ \ \boldsymbol{\eta}_{ip}\ \right], \tag{3.5}$$

[9]Recall that the available observations are $\mathbf{w}_{i0}, \mathbf{w}_{i1}, \ldots, \mathbf{w}_{iT}$, irrespective of the order of the PVAR, $p < T$.

we have that $E(\mathbf{e}_i) = \mathbf{0}_{mp\times 1}$, and $Var(\mathbf{e}_i) = \Sigma_e$, where

$$\Sigma_e = \begin{bmatrix} \mathcal{G}_0 & \mathcal{G}_1 & \mathcal{G}_2 & \cdots & \mathcal{G}_{p-2} & \mathcal{H}_1 \\ \mathcal{G}_1' & \mathcal{G}_0 & \mathcal{G}_1 & \cdots & \mathcal{G}_{p-3} & \mathcal{H}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{G}_{p-2}' & \mathcal{G}_{p-3}' & \mathcal{G}_{p-4}' & \cdots & \mathcal{G}_0 & \mathcal{H}_{p-1} \\ \mathcal{H}_1' & \mathcal{H}_2' & \mathcal{H}_3' & \cdots & \mathcal{H}_{p-1}' & \Psi \end{bmatrix}, \tag{3.6}$$

with

$$\mathcal{G}_l = \sum_{j=0}^{\infty} \mathbf{C}_j \Omega \mathbf{C}_{j+l}', \; l = 0, 1, \ldots, p-2; \quad \mathcal{H}_l = \left(\sum_{j=0}^{\infty} \mathbf{C}_{j+l} \Omega \mathbf{C}_j^{*\prime} \right) \Pi', \; l = 1, 2, \ldots, p-1, \tag{3.7}$$

and

$$\Psi = \Pi \left(\sum_{j=0}^{\infty} \mathbf{C}_j^* \Omega \mathbf{C}_j^{*\prime} \right) \Pi' + \Omega. \tag{3.8}$$

The variance-covariance matrices $\mathcal{G}_l$, $l = 0, 1, \ldots, p-2$, $\mathcal{H}_l$, $l = 1, 2, \ldots, p-1$, and Ψ under our assumptions are well defined irrespective of whether some or all of the roots of the determinantal equation (2.3), $|\Phi(\varrho)| = 0$, are equal to unity. For the pure unit root case, note that $\Pi = \mathbf{0}_m$, and thus Ψ becomes equal to Ω, and $\mathcal{H}_l$ becomes equal to $\mathbf{0}_m$, $l = 1, 2, \ldots, p-1$. For the cases of trend-stationary and cointegrated PVARs where some but not all roots of $|\Phi(\varrho)| = 0$ may be equal to unity, it should be noted that $\{\mathbf{C}_j\}_{j=0}^{\infty}$ as well as $\left\{\mathbf{C}_j^*\right\}_{j=0}^{\infty}$ are absolutely summable. Also note that all elements of Σ_e are fully specified by Φ_j, $j = 1, 2, \ldots, p$, and/or Ω. The parameters of the distribution of $\boldsymbol{\zeta}_i$, $i = 1, 2, \ldots, N$, do not enter Σ_e for any $r = 0, 1, \ldots, m$. Observing that the Jacobian of the transformation from $\mathbf{e}_i$ to $\left(\triangle \mathbf{w}_{i1}', \triangle \mathbf{w}_{i2}', \ldots, \triangle \mathbf{w}_{ip}'\right)'$ is unity, $\Pr(\triangle \mathbf{w}_{i1}, \triangle \mathbf{w}_{i2}, \ldots, \triangle \mathbf{w}_{ip})$ can be derived, and is, for normally distributed errors, provided in the appendix.

The second component on the right-hand side of (3.1), namely the conditional joint probability distribution of $\triangle \mathbf{w}_{i,p+1}, \triangle \mathbf{w}_{i,p+2}, \ldots, \triangle \mathbf{w}_{iT}$ given $\triangle \mathbf{w}_{i1}, \triangle \mathbf{w}_{i2}, \ldots, \triangle \mathbf{w}_{ip}$, can be derived using the first differences of (2.1) for $t = p+1, p+2, \ldots, T$:

$$\triangle \mathbf{w}_{it} - \boldsymbol{\gamma} = \sum_{j=1}^{p} \Phi_j \left(\triangle \mathbf{w}_{i,t-j} - \boldsymbol{\gamma} \right) + \triangle \boldsymbol{\epsilon}_{it}, \qquad t = p+1, p+2, \ldots, T. \tag{3.9}$$

Note that the first-differencing eliminates the fixed effects for the last $T-p$ time-series observations in the panel. Collecting the disturbance terms in (3.9),

$$\mathbf{f}_i = \left(\; \triangle \boldsymbol{\epsilon}_{i,p+1}', \quad \triangle \boldsymbol{\epsilon}_{i,p+2}', \quad \ldots, \quad \triangle \boldsymbol{\epsilon}_{iT}' \; \right)', \tag{3.10}$$

using linear projection we have that $E(\mathbf{f}_i|\mathbf{e}_i) = Cov(\mathbf{f}_i, \mathbf{e}_i)\Sigma_{\mathbf{e}}^{-1}\mathbf{e}_i$, which simplifies to $E(\mathbf{f}_i|\mathbf{e}_i) = \left[\; -(\Omega \mathcal{R} \mathbf{e}_i)', \quad \mathbf{0}_{1\times m(T-p-1)} \; \right]'$, where $\mathcal{R}$ denotes the $m \times mp$ matrix formed by the last m rows of

Σ_{ϵ}^{-1}, and that $Var(\mathfrak{f}_i|\mathfrak{e}_i) = Var(\mathfrak{f}_i) - Cov(\mathfrak{f}_i, \mathfrak{e}_i)\Sigma_{\epsilon}^{-1}Cov(\mathfrak{e}_i, \mathfrak{f}_i)$, which simplifies to $Var(\mathfrak{f}_i|\mathfrak{e}_i) = \Sigma_{\mathfrak{f}|\epsilon}$, where

$$\Sigma_{\mathfrak{f}|\epsilon} = \begin{pmatrix} \mathfrak{B} & -\Omega & & & & \\ -\Omega & 2\Omega & -\Omega & & \mathbf{0} & \\ & \ddots & \ddots & \ddots & & \\ & \mathbf{0} & & -\Omega & 2\Omega & -\Omega \\ & & & & -\Omega & 2\Omega \end{pmatrix}, \tag{3.11}$$

with

$$\mathfrak{B} = 2\Omega - \Omega\mathcal{S}\Omega, \tag{3.12}$$

and $\mathcal{S}$ denoting the $m \times m$ matrix formed by the last m columns of $\mathcal{R}$. Observing that the Jacobian of the transformation from $\mathfrak{f}_i$ to $\left(\triangle \mathbf{w}'_{i,p+1}, \triangle \mathbf{w}'_{i,p+2}, \ldots, \triangle \mathbf{w}'_{iT}\right)'$ is unity,

$$\Pr\left(\triangle \mathbf{w}_{i,p+1}, \triangle \mathbf{w}_{i,p+2}, \ldots, \triangle \mathbf{w}_{iT} | \triangle \mathbf{w}_{i1}, \triangle \mathbf{w}_{i2}, \ldots, \triangle \mathbf{w}_{ip}\right)$$

can be derived, and is, for normally distributed errors, provided in the appendix.

The unconditional joint probability distribution function of the transformed observations, $\triangle \mathbf{w}_{i1}$, $\triangle \mathbf{w}_{i2}, \ldots, \triangle \mathbf{w}_{ip}, \triangle \mathbf{w}_{i,p+1}, \ldots, \triangle \mathbf{w}_{iT}$, can now be assembled as in (3.1), and in contrast to the probability distribution function of the untransformed model, (2.1), is not subject to the incidental parameters problem. Before establishing the properties of the resultant ML estimator, we shall consider the use of the transformed model equations (3.2), (3.3), and (3.9) for GMM estimation. Having discussed the difficulties with GMM estimation even after the incidental parameters problem has been resolved, in Section 5 we shall establish the properties of the ML estimator and discuss its computational implementation.

4 GMM Estimation

Most discussions of GMM estimation of dynamic panel data models are within a single equation context (for example, Arellano and Bond, 1991, Ahn and Schmidt, 1995, Arellano and Bover, 1995, Blundell and Bond, 1998).[10] However, just like Three Stage Least Squares (3SLS) estimation of a system of equations can be more efficient than the single equation based Two Stage Least Squares (2SLS), we shall generalize GMM estimation to a systems context, and then show that if the PVAR contains unit roots, the conventional GMM approach using instruments that are orthogonal to the disturbance terms of the model equations breaks down. We then investigate if the use of variants

[10] An exception is Holtz-Eakin, Newey, and Rosen (1988), who study a multivariate setting also, but in a framework different than ours.

of homoskedasticity and stationarity implied moment conditions as suggested by Ahn and Schmidt (1995) and Arellano and Bover (1995) can help resolve the issue.

To simplify the exposition, for much of the remainder of this paper we set $p = 1$, and consider the PVAR(1) model (2.12). The counterparts of the transformed model equations (3.3) and (3.9) for the PVAR(1) model (2.12) are given by[11]

$$\triangle \mathbf{w}_{i1} - \boldsymbol{\gamma} = \boldsymbol{\eta}_{i1}, \tag{4.1}$$

where $\boldsymbol{\eta}_{i1}$ is given by (3.4) for $p = 1$, namely

$$\boldsymbol{\eta}_{i1} = \boldsymbol{\varepsilon}_{i1} - (\mathbf{I}_m - \Phi) \sum_{j=0}^{\infty} \Phi^j \boldsymbol{\varepsilon}_{i,-j}, \tag{4.2}$$

and

$$\triangle \mathbf{w}_{it} - \boldsymbol{\gamma} = \Phi (\triangle \mathbf{w}_{i,t-1} - \boldsymbol{\gamma}) + \triangle \boldsymbol{\varepsilon}_{it}, \qquad t = 2, 3, \ldots, T. \tag{4.3}$$

Also, from (4.1) it follows that for the PVAR(1) model (2.12) it holds that

$$\triangle \mathbf{w}_{i1} - \boldsymbol{\gamma} \overset{iid}{\sim} (\mathbf{0}_{m \times 1}, \Psi), \tag{4.4}$$

where Ψ as before is given by (3.8), which in the case where $p = 1$ can be written as

$$\Psi = (\mathbf{I}_m - \Phi) \left(\sum_{j=0}^{\infty} \Phi^j \Omega \Phi^{j\prime} \right) (\mathbf{I}_m - \Phi)' + \Omega. \tag{4.5}$$

The appendix provides many of the results derived in the remainder of the paper for $p = 1$ for the case where $p > 1$.

4.1 Conventional GMM Estimation

The conventional GMM estimator, developed for (trend-) stationary models, employs instruments that are orthogonal to the model disturbances. For the first-differenced PVAR(1) model (4.3), such instruments are given by a constant and levels of the dependent variables, $\mathbf{w}_{it}$, lagged two or more periods (which are predetermined). The standard orthogonality conditions are thus given by

$$E\left\{ [(\triangle \mathbf{w}_{it} - \boldsymbol{\gamma}) - \Phi (\triangle \mathbf{w}_{i,t-1} - \boldsymbol{\gamma})] \, \mathbf{q}_{it}' \right\} = \mathbf{0}_{m \times [(t-1)m+1]}, \qquad t = 2, 3, \ldots, T, \tag{4.6}$$

where

$$\mathbf{q}_{it} = \left(\; 1, \;\; \mathbf{w}_{i0}', \;\; \mathbf{w}_{i1}', \;\; \ldots, \;\; \mathbf{w}_{i,t-2}' \; \right)'. \tag{4.7}$$

[11] Note that (3.2) need only be considered in the case where $p > 1$.

To obtain the conventional GMM estimator using the moment conditions (4.6), note that the first-differenced PVAR(1) model (4.3) can in stacked form be written as

$$\mathbf{S}_i = \mathbf{R}_i\Lambda + \mathbf{E}_i, \tag{4.8}$$

$i = 1,2,\ldots,N$, where

$$\mathbf{S}_i = \left(\begin{array}{cccc} \triangle\mathbf{w}_{i2}, & \triangle\mathbf{w}_{i3}, & \ldots, & \triangle\mathbf{w}_{iT} \end{array} \right)', \qquad \mathbf{E}_i = \left(\begin{array}{cccc} \triangle\boldsymbol{\varepsilon}_{i2}, & \triangle\boldsymbol{\varepsilon}_{i3}, & \ldots, & \triangle\boldsymbol{\varepsilon}_{iT} \end{array} \right)', \tag{4.9}$$

$$\mathbf{R}_i = \left(\begin{array}{cc} \mathbf{P}_i, & \iota_{T-1} \end{array} \right), \qquad \mathbf{P}_i = \left(\begin{array}{cccc} \triangle\mathbf{w}_{i1}, & \triangle\mathbf{w}_{i2}, & \ldots, & \triangle\mathbf{w}_{i,T-1} \end{array} \right)', \tag{4.10}$$

$$\Lambda = \left(\begin{array}{cc} \Phi, & \mathbf{a}_1 \end{array} \right)', \qquad \mathbf{a}_1 = (\mathbf{I}_m - \Phi)\,\boldsymbol{\gamma}, \tag{4.11}$$

and ι_{T-1} denotes a $(T-1) \times 1$ vector of ones.

Pre-multiplying (4.8) by the $(mT/2+1)(T-1) \times (T-1)$ block-diagonal instrumental variable matrix $\mathbf{Q}_i$,

$$\mathbf{Q}_i = \begin{pmatrix} \mathbf{q}_{i2} & \mathbf{0}_{(m+1)\times 1} & \mathbf{0}_{(m+1)\times 1} & & \\ \mathbf{0}_{(2m+1)\times 1} & \mathbf{q}_{i3} & \mathbf{0}_{(2m+1)\times 1} & & \mathbf{0} \\ \mathbf{0}_{(3m+1)\times 1} & \mathbf{0}_{(3m+1)\times 1} & \mathbf{q}_{i4} & & \\ & & & \ddots & \\ & \mathbf{0} & & & \mathbf{q}_{iT} \end{pmatrix}, \tag{4.12}$$

one obtains

$$\mathbf{Q}_i\mathbf{S}_i = \mathbf{Q}_i\mathbf{R}_i\Lambda + \mathbf{Q}_i\mathbf{E}_i, \tag{4.13}$$

the transpose of which in vectorized form becomes

$$(\mathbf{Q}_i \otimes \mathbf{I}_m)\, vec\left(\mathbf{S}_i'\right) = (\mathbf{Q}_i\mathbf{R}_i \otimes \mathbf{I}_m)\boldsymbol{\lambda} + (\mathbf{Q}_i \otimes \mathbf{I}_m)\, vec\left(\mathbf{E}_i'\right), \tag{4.14}$$

where $\boldsymbol{\lambda} = vec(\Lambda')$. The conventional GMM estimator of $\boldsymbol{\lambda}$ is obtained by minimizing

$$\left[\sum_{i=1}^{N} (\mathbf{Q}_i \otimes \mathbf{I}_m)\, vec\left(\mathbf{S}_i'\right) - (\mathbf{Q}_i\mathbf{R}_i \otimes \mathbf{I}_m)\,\boldsymbol{\lambda}\right]' \left[\sum_{i=1}^{N} (\mathbf{Q}_i \otimes \mathbf{I}_m)\,\overline{\Sigma}\,(\mathbf{Q}_i \otimes \mathbf{I}_m)'\right]^{-1}$$
$$\cdot \left[\sum_{i=1}^{N} (\mathbf{Q}_i \otimes \mathbf{I}_m)\, vec\left(\mathbf{S}_i'\right) - (\mathbf{Q}_i\mathbf{R}_i \otimes \mathbf{I}_m)\,\boldsymbol{\lambda}\right], \tag{4.15}$$

where

$$\overline{\Sigma} = \begin{pmatrix} 2\Omega & -\Omega & & & & & \\ -\Omega & 2\Omega & -\Omega & & & \mathbf{0} & \\ & \ddots & \ddots & \ddots & & & \\ & & & & -\Omega & 2\Omega & -\Omega \\ & \mathbf{0} & & & & -\Omega & 2\Omega \end{pmatrix}. \tag{4.16}$$

The moment conditions relevant to the estimation of Ω are given by

$$E\left\{\left[\Delta \mathbf{w}_{it} - \boldsymbol{\gamma} - \Phi\left(\Delta \mathbf{w}_{i,t-1} - \boldsymbol{\gamma}\right)\right]\left[\Delta \mathbf{w}_{it} - \boldsymbol{\gamma} - \Phi\left(\Delta \mathbf{w}_{i,t-1} - \boldsymbol{\gamma}\right)\right]' - 2\Omega\right\} = 0_m, \tag{4.17}$$

for $t = 2,3,...,T$. Also note that in the case of a trend-stationary PVAR, upon estimation of $\mathbf{a}_1$ from (4.15), $\boldsymbol{\gamma}$ may be consistently estimated as

$$\widehat{\boldsymbol{\gamma}} = \left(\mathbf{I}_m - \widehat{\Phi}\right)^{-1} \widehat{\mathbf{a}}_1. \tag{4.18}$$

The conventional GMM estimator is consistent if all eigenvalues of Φ fall inside the unit circle, but breaks down if some eigenvalues of Φ are equal to unity.[12] Note that a necessary condition for the GMM estimator (4.15) to exist is that $rank\left(N^{-1}\sum_{i=1}^{N}\mathbf{Q}_i\mathbf{R}_i\right) = m+1$ as $N \to \infty$. In the case where $\Phi = \mathbf{I}_m$, $rank\left(N^{-1}\sum_{i=1}^{N}\mathbf{Q}_i\mathbf{R}_i\right)$ as $N \to \infty$ is less than $m+1$, however. This is because when $\Phi = \mathbf{I}_m$, $\Delta \mathbf{w}_{it} = \boldsymbol{\gamma} + \boldsymbol{\epsilon}_{it}$, and $\mathbf{w}_{it} = \mathbf{w}_{i0} + \boldsymbol{\gamma} t + \mathbf{s}_{it}$, with $\mathbf{s}_{it}$ as before given by $\mathbf{s}_{it} = \sum_{l=1}^{t}\boldsymbol{\epsilon}_{il}$, and thus it follows that for $t = 2,3,\ldots,T$, $j = 0,1,\ldots,t-2$, as $N \to \infty$

$$\frac{1}{N}\sum_{i=1}^{N}\Delta \mathbf{w}_{i,t-1}\mathbf{w}_{ij}' = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\gamma} + \boldsymbol{\epsilon}_{i,t-1}\right)\left(\mathbf{w}_{i0} + \boldsymbol{\gamma} j + \mathbf{s}_{ij}\right)' \xrightarrow{p} \boldsymbol{\gamma}\left(\mathbf{w}_{i0} + \boldsymbol{\gamma} j\right)' \tag{4.19}$$

(where $\xrightarrow{p}$ denotes convergence in probability), which is of rank one. In other words, when $\Phi = \mathbf{I}_m$, the elements of $\mathbf{q}_{it}$ are not legitimate instruments. For the PVAR(p) model (2.1) (with $p > 1$), although $N^{-1}\sum_{i=1}^{N}\Delta \mathbf{w}_{i,t-1}\mathbf{w}_{ij}'$, $t = 2,3,\ldots,T$, $j = 0,1,\ldots,t-2$, may be of full rank as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_{it}\left[\ \left(\Delta \mathbf{w}_{i,t-1} - \boldsymbol{\gamma}\right)',\ \left(\Delta \mathbf{w}_{i,t-2} - \boldsymbol{\gamma}\right)',\ \ldots,\ \left(\Delta \mathbf{w}_{i,t-p} - \boldsymbol{\gamma}\right)'\ \right], \tag{4.20}$$

[12]Note that if there are no restrictions on the deterministic trend terms, that is, the PVAR(1) model is written as $\left(\mathbf{I}_m - \Phi L\right)\mathbf{w}_{it} = \boldsymbol{\mu}_i + \boldsymbol{\gamma} t + \boldsymbol{\varepsilon}_{it}$, then when the roots of $|\mathbf{I}_m - \Phi\varrho| = 0$ are equal to unity, minimizing (4.15) can yield consistent estimates of $\boldsymbol{\lambda}$ if $m = 1$, but not for $m > 1$. Note that if there are no restrictions on the deterministic trend terms, then when the roots of $|\mathbf{I}_m - \Phi\varrho| = 0$ are equal to unity, the model contains both a linear and a quadratic deterministic trend, whereas it only contains a linear deterministic trend when all roots of $|\mathbf{I}_m - \Phi\varrho| = 0$ fall outside the unit circle.

$t = 1,2,\ldots,T$, will have deficient rank as $N \to \infty$. To see this, note from (2.8) that if $\sum_{j=1}^{p} \Phi_j = \mathbf{I}_m$, then (2.1) may be written as

$$\Phi^*(L)(\Delta \mathbf{w}_{it} - \boldsymbol{\gamma}) = \boldsymbol{\varepsilon}_{it}, \tag{4.21}$$

with $\Phi^*(L)$ given by (2.9). Post-multiplying (4.20) by $\left(\mathbf{I}_m, \ -\Phi_1^*, \ -\Phi_2^*, \ \ldots, \ -\Phi_{p-1}^* \right)'$, it follows that for $t = 2,3,\ldots,T$, as $N \to \infty$

$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{q}_{it} \boldsymbol{\varepsilon}_{i,t-1}' \xrightarrow{p} \mathbf{0}_{[m(t-1)+1] \times m}. \tag{4.22}$$

In the event that the PVAR(p) model (2.1) contains unit roots, the conventional GMM estimator breaks down.

4.2 Conventional GMM Estimation Incorporating Initial Conditions

In the case of models with non-zero deterministic trends the conventional GMM moment conditions (4.6) can for estimation of the trend coefficients ($\boldsymbol{\gamma}$) be augmented with additional moment conditions based on (4.1), namely

$$E(\Delta \mathbf{w}_{i1} - \boldsymbol{\gamma}) = \mathbf{0}_{m \times 1}. \tag{4.23}$$

Stacking (4.1) and (4.3), we have

$$\widetilde{\mathbf{S}}_i = \widetilde{\mathbf{R}}_i \Lambda + \widetilde{\mathbf{E}}_i, \tag{4.24}$$

$i = 1,2,\ldots,N$, where

$$\widetilde{\mathbf{S}}_i = \left(\ \Delta \mathbf{w}_{i1}, \ \ \mathbf{S}_i' \ \right)', \qquad \widetilde{\mathbf{E}}_i = \left[\ \boldsymbol{\varepsilon}_{i1} - (\mathbf{I}_m - \Phi) \textstyle\sum_{j=0}^{\infty} \Phi^j \boldsymbol{\varepsilon}_{i,-j}, \ \ \mathbf{E}_i' \ \right]', \tag{4.25}$$

$$\widetilde{\mathbf{R}}_i = \left[\ \widetilde{\mathbf{P}}_i, \ \ \iota_T, \ \ \begin{pmatrix} 0 \\ \iota_{T-1} \end{pmatrix} \ \right], \qquad \widetilde{\mathbf{P}}_i = \left(\ \mathbf{0}_{m \times 1}, \ \ \mathbf{P}_i' \ \right)', \tag{4.26}$$

and

$$\widetilde{\Lambda} = \left(\ \Phi, \ \ \boldsymbol{\gamma}, \ \ -\Phi\boldsymbol{\gamma} \ \right)'. \tag{4.27}$$

Pre-multiplying (4.24) by the $[m(T-1)/2+1]\,T \times T$ block-diagonal instrumental variable matrix $\widetilde{\mathbf{Q}}_i$,

$$\widetilde{\mathbf{Q}}_i = \begin{pmatrix} \mathbf{q}_{i1} & 0 & 0 & & \\ \mathbf{0}_{(m+1)\times 1} & \mathbf{q}_{i2} & \mathbf{0}_{(m+1)\times 1} & & \mathbf{0} \\ \mathbf{0}_{(2m+1)\times 1} & \mathbf{0}_{(2m+1)\times 1} & \mathbf{q}_{i3} & & \\ & & & \ddots & \\ & \mathbf{0} & & & \mathbf{q}_{iT} \end{pmatrix}, \tag{4.28}$$

and vectorizing the transpose of the resultant expression one obtains

$$\left(\tilde{\mathbf{Q}}_i \otimes \mathbf{I}_m\right) vec\left(\tilde{\mathbf{S}}_i'\right) = \left(\tilde{\mathbf{Q}}_i \tilde{\mathbf{R}}_i \otimes \mathbf{I}_m\right) \tilde{\boldsymbol{\lambda}} + \left(\tilde{\mathbf{Q}}_i \otimes \mathbf{I}_m\right) vec\left(\tilde{\mathbf{E}}_i'\right), \tag{4.29}$$

where $\tilde{\boldsymbol{\lambda}} = vec\left(\tilde{\boldsymbol{\Lambda}}'\right)$. The conventional GMM estimator of $\tilde{\boldsymbol{\lambda}}$ is obtained by minimizing

$$\left[\sum_{i=1}^{N}\left(\tilde{\mathbf{Q}}_i \otimes \mathbf{I}_m\right) vec\left(\tilde{\mathbf{S}}_i'\right) - \left(\tilde{\mathbf{Q}}_i \tilde{\mathbf{R}}_i \otimes \mathbf{I}_m\right) \tilde{\boldsymbol{\lambda}}\right]' \left[\sum_{i=1}^{N}\left(\tilde{\mathbf{Q}}_i \otimes \mathbf{I}_m\right) \Sigma \left(\tilde{\mathbf{Q}}_i \otimes \mathbf{I}_m\right)'\right]^{-1}$$
$$\cdot \left[\sum_{i=1}^{N}\left(\tilde{\mathbf{Q}}_i \otimes \mathbf{I}_m\right) vec\left(\tilde{\mathbf{S}}_i'\right) - \left(\tilde{\mathbf{Q}}_i \tilde{\mathbf{R}}_i \otimes \mathbf{I}_m\right) \tilde{\boldsymbol{\lambda}}\right], \tag{4.30}$$

where

$$\Sigma = \begin{pmatrix} \Psi & -\Omega & & & & \\ -\Omega & 2\Omega & -\Omega & & \mathbf{0} & \\ & \ddots & \ddots & \ddots & & \\ & \mathbf{0} & & -\Omega & 2\Omega & -\Omega \\ & & & & -\Omega & 2\Omega \end{pmatrix}, \tag{4.31}$$

with Ψ given by (4.5). The moment conditions relevant to the estimation of Ω are again given by (4.17).

The conventional GMM estimator of Φ incorporating the initial conditions will be more efficient than the conventional GMM estimator of Φ not incorporating the initial conditions. Note, though, that (4.30) introduces nonlinear restrictions between Φ and $\boldsymbol{\gamma}$. Both the conventional GMM estimator of Φ based on (4.15) and that based on (4.30) break down if some eigenvalues of Φ are equal to unity.

4.3 Extended GMM Estimation Using Homoskedasticity and Stationarity Implied Moment Conditions

The above discussion shows that in the presence of unit roots the consistent estimation of the PVAR model requires further moment conditions. One possibility would be to augment the standard orthogonality conditions (4.6) and (4.23) by homoskedasticity and stationarity implied moment conditions as derived by Ahn and Schmidt (1995) and Arellano and Bover (1995) in the context of univariate dynamic panel data models. In the context of the PVAR(1) model (2.12), the Ahn and Schmidt (1995) homoskedasticity implied moment conditions can be written as

$$E\left(\mathbf{w}_{it} \triangle \boldsymbol{\varepsilon}_{i,t+1}' - \mathbf{w}_{i,t+1} \triangle \boldsymbol{\varepsilon}_{i,t+2}'\right) = \mathbf{0}_m, \qquad t = 1, 2, \ldots, T-2, \tag{4.32}$$

and

$$E\left[\left(T^{-1}\sum_{t=1}^{T}\mathbf{d}_{it}\right) \triangle \boldsymbol{\varepsilon}_{i,t+1}'\right] = \mathbf{0}_m, \qquad t = 1, 2, \ldots, T-1, \tag{4.33}$$

where

$$\mathbf{d}_{it} = \mathbf{w}_{it} - \Phi \mathbf{w}_{i,t-1} - (\mathbf{I}_m - \Phi)\,\gamma t - \Phi\gamma. \tag{4.34}$$

The Arellano and Bover (1995) stationarity implied moment conditions can in the context of the PVAR(1) model (2.12) be written as[13]

$$E\left[(\Delta \mathbf{w}_{i,t-1} - \gamma)\,\mathbf{d}_{it}'\right] = \mathbf{0}_m, \qquad t = 2,3,\ldots,T. \tag{4.35}$$

Following Ahn and Schmidt (1995), the moment conditions (4.32), (4.33), and (4.35) after removing duplicate information may in combined form be written as

$$E\left[\mathbf{d}_{iT}\,(\Delta \mathbf{w}_{it} - \gamma)'\right] = \mathbf{0}_m, \qquad t = 1,2,\ldots,T-1, \tag{4.36}$$

and

$$E\left\{\mathbf{d}_{it}\,(\mathbf{w}_{it} - \gamma t)' - \mathbf{d}_{i,t-1}\,[\mathbf{w}_{i,t-1} - \gamma\,(t-1)]'\right\} = \mathbf{0}_m, \qquad t = 2,3,\ldots,T. \tag{4.37}$$

Since the moment conditions (4.36) and (4.37) involve the fixed effects, μ_i, it is necessary to verify whether they are also applicable to the PVAR(1) model (2.12) as we have specified it.

Noting that the counterpart of the first-difference representation (3.2) for $p = 1$ and $t = 1,2,\ldots,T$ is given by

$$\Delta \mathbf{w}_{it} - \gamma = \varepsilon_{it} - (\mathbf{I}_m - \Phi)\sum_{j=0}^{\infty} \Phi^j \varepsilon_{i,t-j}, \tag{4.38}$$

and that $\mathbf{d}_{iT} = (\mathbf{I}_m - \Phi)\,\mu_i + \varepsilon_{iT}$, (4.36) can be written as

$$E\left\{[(\mathbf{I}_m - \Phi)\,\mu_i + \varepsilon_{iT}]\left[\varepsilon_{it} - (\mathbf{I}_m - \Phi)\sum_{j=0}^{\infty}\Phi^j \varepsilon_{i,t-j}\right]'\right\} = \mathbf{0}_m, \qquad t = 1,2,\ldots,T-1, \tag{4.39}$$

which is valid under the assumption that $(\mathbf{I}_m - \Phi)\,\mu_i$ and ε_{it}, $t \leq T-1$, are uncorrelated. Note that this assumption does not rule out correlation between μ_i and initial disturbances, say $\varepsilon_{i,-M}$; if μ_i, for example, was generated by

$$\mu_i = \mathbf{C}\mathfrak{M}_i \varepsilon_{i,-M} + \mathbf{u}_i, \qquad i = 1,2,\ldots,N, \tag{4.40}$$

where $\mathfrak{M}_i$ is an $m \times m$ matrix of parameters, and $\mathbf{u}_i$, $i = 1,2,\ldots,N$, is an $m \times 1$ vector of random variables independently distributed across i, and also independent of ε_{it}, $t \leq T-1$, then $(\mathbf{I}_m - \Phi)\,\mu_i$ and ε_{it}, $t \leq T-1$, would still be uncorrelated.

[13] The Arellano and Bover (1995) moment conditions in the literature are referred to as stationarity implied moment conditions, as they may be motivated by a stationarity restriction on the cross product of the regressand and the fixed effects vector.

Turning to (4.37), noting that under (*A*5) $\{\mathbf{w}_{it}\}_{t=1}^{T}$ is generated from

$$\mathbf{w}_{it} = \boldsymbol{\mu}_i + \boldsymbol{\gamma} t + \sum_{j=0}^{\infty} \left(\boldsymbol{\Phi}^j - \mathbf{C}\right) \boldsymbol{\varepsilon}_{i,t-j} + \mathbf{C} \left(\sum_{j=0}^{t+M-1} \boldsymbol{\varepsilon}_{i,t-j} \right) + \mathbf{C} \mathfrak{v}_i, \tag{4.41}$$

it is readily seen that (4.37) will be valid if $\boldsymbol{\mu}_i$ and $\boldsymbol{\varepsilon}_{it}$, $t = 1,2,\ldots,T$, as well as $(\mathbf{I}_m - \boldsymbol{\Phi})\,\boldsymbol{\mu}_i$ and $\boldsymbol{\varepsilon}_{it}$, $t \leq 0$, are uncorrelated, and $(\mathbf{I}_m - \boldsymbol{\Phi})\,E\left(\boldsymbol{\mu}_i \boldsymbol{\mu}_i'\right)$ as well as $(\mathbf{I}_m - \boldsymbol{\Phi})\,E\left(\boldsymbol{\mu}_i \mathfrak{v}_i'\right)\mathbf{C}'$ exist.

The validity of our proposed ML estimator will be unaffected by whether the assumptions needed for the homoskedasticity and stationarity implied moment conditions (4.36) and (4.37) to be valid do hold, as it will be based on the unconditional joint probability distribution of the transformed observations, $\Delta\mathbf{w}_{i1}$, $\Delta\mathbf{w}_{i2}$, ..., $\Delta\mathbf{w}_{iT}$, which, as seen in Section 3, does not involve $\boldsymbol{\mu}_i$. It should also be noted that even when the moment conditions (4.36) and (4.37) are valid, the variance-covariance matrix associated with these moment conditions does depend on the individual-specific effects $\boldsymbol{\mu}_i$. Because of this dependence, without further (homogeneity) restrictions on the distribution of $\boldsymbol{\mu}_i$ it is not possible to derive the optimal weighting matrix associated with the homoskedasticity and stationarity implied moment conditions (4.36) and (4.37), and the orthogonality conditions (4.6) and (4.23). The extended GMM estimator based on a non-optimal weighting matrix will in general not be as efficient as our proposed ML estimator.

To overcome the dependence of the variance-covariance matrix associated with the moment conditions (4.36) and (4.37) on the individual-specific effects $\boldsymbol{\mu}_i$, an alternative way of making use of the homoskedasticity restrictions on $\boldsymbol{\varepsilon}_{it}$ would be to combine (4.32) with (4.6), and consider the moment conditions

$$E\left(\Delta\mathbf{w}_{it} \triangle \boldsymbol{\varepsilon}_{i,t+1}' - \Delta\mathbf{w}_{i,t+1} \triangle \boldsymbol{\varepsilon}_{i,t+2}'\right) = \mathbf{0}_m, \qquad t = 2,3,\ldots,T-2. \tag{4.42}$$

Note that

$$\begin{aligned} & E\left(\Delta\mathbf{w}_{it} \triangle \boldsymbol{\varepsilon}_{i,t+1}' - \Delta\mathbf{w}_{i,t+1} \triangle \boldsymbol{\varepsilon}_{i,t+2}'\right) \\ = \; & E\left(\mathbf{w}_{it} \triangle \boldsymbol{\varepsilon}_{i,t+1}' - \mathbf{w}_{i,t+1} \triangle \boldsymbol{\varepsilon}_{i,t+2}'\right) - E\left(\mathbf{w}_{i,t-1} \triangle \boldsymbol{\varepsilon}_{i,t+1}'\right) + E\left(\mathbf{w}_{it} \triangle \boldsymbol{\varepsilon}_{i,t+2}'\right), \end{aligned} \tag{4.43}$$

and thus it is readily seen that (4.32) and (4.6) imply (4.42).[14] However, the moment conditions (4.42) do not lead to a workable extended GMM estimator. To see this, let

$$\mathbf{V}_{it}(\boldsymbol{\gamma}) = \Delta\mathbf{w}_{it}\left(\Delta\mathbf{w}_{i,t+1} - \boldsymbol{\gamma}\right)' - \Delta\mathbf{w}_{i,t+1}\left(\Delta\mathbf{w}_{i,t+2} - \boldsymbol{\gamma}\right)', \tag{4.44}$$

and

$$\mathbf{Z}_{it}(\boldsymbol{\gamma}) = \Delta\mathbf{w}_{it}\left(\Delta\mathbf{w}_{it} - \boldsymbol{\gamma}\right)' - \Delta\mathbf{w}_{i,t+1}\left(\Delta\mathbf{w}_{i,t+1} - \boldsymbol{\gamma}\right)'. \tag{4.45}$$

[14] It may be verified that under Assumption (*A*5) (4.42) is applicable for $t = 1$ as well.

Then (4.42) can be rewritten as

$$E\left\{vec\left[\Delta \mathbf{w}_{it}\,\Delta\,\varepsilon'_{i,t+1} - \Delta \mathbf{w}_{i,t+1}\,\Delta\,\varepsilon'_{i,t+2}\right]\right\} = E\left\{vec\left[\mathbf{V}_{it}(\boldsymbol{\gamma})\right] - \left[\mathbf{I}_m \otimes \mathbf{Z}_{it}(\boldsymbol{\gamma})\right]\right\} vec\left(\Phi'\right). \tag{4.46}$$

However, it is readily seen that

$$\operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\Delta \mathbf{w}_{it}\left(\Delta \mathbf{w}_{i,t+1} - \boldsymbol{\gamma}\right)'\right] = \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\Delta \mathbf{w}_{i,t+1}\left(\Delta \mathbf{w}_{i,t+2} - \boldsymbol{\gamma}\right)'\right]. \tag{4.47}$$

Similarly,

$$\operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\Delta \mathbf{w}_{it}\left(\Delta \mathbf{w}_{it} - \boldsymbol{\gamma}\right)'\right] = \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\Delta \mathbf{w}_{i,t+1}\left(\Delta \mathbf{w}_{i,t+1} - \boldsymbol{\gamma}\right)'\right]. \tag{4.48}$$

Thus, the moment conditions (4.42) are of no use for the estimation of Φ.

Another possibility of making use of the homoskedasticity restrictions would be to consider the second moments directly, taking account of the correlation of $\Delta \mathbf{w}_{i,t-1}$ with $\Delta \boldsymbol{\varepsilon}_{it}$. Using (4.1) and (4.3), it is readily seen that

$$E\left\{\left[\Delta \mathbf{w}_{i2} - \boldsymbol{\gamma} - \Phi\left(\Delta \mathbf{w}_{i1} - \boldsymbol{\gamma}\right)\right]\left(\Delta \mathbf{w}_{i1} - \boldsymbol{\gamma}\right)'\right\} = E\left[\Delta \boldsymbol{\varepsilon}_{i2}\left(\Delta \mathbf{w}_{i1} - \boldsymbol{\gamma}\right)'\right] = -\Omega. \tag{4.49}$$

Thus, the following moment conditions must hold:

$$E\left\{\left[\Delta \mathbf{w}_{i2} - \boldsymbol{\gamma} - \Phi\left(\Delta \mathbf{w}_{i1} - \boldsymbol{\gamma}\right)\right]\left(\Delta \mathbf{w}_{i1} - \boldsymbol{\gamma}\right)' + \Omega\right\} = \mathbf{0}_m. \tag{4.50}$$

Furthermore, using (4.3) and (4.38) it is readily seen that the following $T-2$ sets of moment conditions must also hold:

$$E\left\{\left[\Delta \mathbf{w}_{it} - \boldsymbol{\gamma} - \Phi\left(\Delta \mathbf{w}_{i,t-1} - \boldsymbol{\gamma}\right)\right]\left(\Delta \mathbf{w}_{i,t-1} - \boldsymbol{\gamma}\right)' + \Omega\right\} = \mathbf{0}_m, \qquad t = 3, 4, \ldots, T. \tag{4.51}$$

The $T-1$ sets of moment conditions (4.50) and (4.51) hold regardless of the unit root and cointegrating properties of the PVAR(1) model (2.12), and could augment the standard orthogonality conditions (4.6) and (4.23). The standard orthogonality conditions (4.6) will not be contributing to the estimation of Φ if it happens that $\Phi = \mathbf{I}_m$, and will contribute only partially under cointegration. As some of the roots of Φ approach unity the relative contribution of the moment conditions (4.6) diminishes, and the $\mathbf{q}_{it}$'s become weak instruments in the terminology of Staiger and Stock (1997).

The problem with the homoskedasticity implied moment conditions (4.50) and (4.51) is that they do not imply unique solutions for Φ, in general. To see this, note that upon substituting (4.17)

into (4.50) and (4.51) to express (4.50) and (4.51) as a function of Φ alone, the resulting moment conditions become quadratic in Φ, and it is not clear how to choose among the multiple solutions of the extended GMM moment conditions, in general.

To construct an extended GMM estimator, let us stack the standard orthogonality conditions (4.6) and (4.23), and the homoskedasticity implied moment conditions (4.50) and (4.51) as

$$E\left[\mathbf{m}(\Delta\mathbf{w}_i, \boldsymbol{\kappa})\right] = \mathbf{0}_{\{m[m(T-1)/2+1]T+m^2(T-1)\}\times 1}, \tag{4.52}$$

where $\boldsymbol{\kappa} = (\boldsymbol{\gamma}', \boldsymbol{\phi}')'$,

$$\boldsymbol{\phi} = vec\left(\Phi\right), \tag{4.53}$$

$$\Delta\mathbf{w}_i = vec\Big(\ \Delta\mathbf{w}_{i1},\ \ \Delta\mathbf{w}_{i2},\ \ \ldots,\ \ \Delta\mathbf{w}_{iT}\ \Big), \tag{4.54}$$

and

$$\mathbf{m}(\Delta\mathbf{w}_i, \boldsymbol{\kappa}) = \begin{pmatrix} \Delta\mathbf{w}_{i1} - \boldsymbol{\gamma} \\ vec\{[(\Delta\mathbf{w}_{i2} - \boldsymbol{\gamma}) - \Phi\,(\Delta\mathbf{w}_{i1} - \boldsymbol{\gamma})]\,\mathbf{q}_{i2}'\} \\ vec\{[(\Delta\mathbf{w}_{i3} - \boldsymbol{\gamma}) - \Phi\,(\Delta\mathbf{w}_{i2} - \boldsymbol{\gamma})]\,\mathbf{q}_{i3}'\} \\ \vdots \\ vec\{[(\Delta\mathbf{w}_{iT} - \boldsymbol{\gamma}) - \Phi\,(\Delta\mathbf{w}_{i,T-1} - \boldsymbol{\gamma})]\,\mathbf{q}_{iT}'\} \\ vec[\mathbf{U}_{i2}(\boldsymbol{\gamma}, \Phi)] \\ vec[\mathbf{U}_{i3}(\boldsymbol{\gamma}, \Phi)] \\ \vdots \\ vec[\mathbf{U}_{iT}(\boldsymbol{\gamma}, \Phi)] \end{pmatrix}, \tag{4.55}$$

with

$$\begin{aligned} \mathbf{U}_{it}(\boldsymbol{\gamma}, \Phi) \;=\; & 2\left[\Delta\mathbf{w}_{it} - \boldsymbol{\gamma} - \Phi\,(\Delta\mathbf{w}_{i,t-1} - \boldsymbol{\gamma})\right](\Delta\mathbf{w}_{i,t-1} - \boldsymbol{\gamma})' \\ & + \left[\Delta\mathbf{w}_{it} - \boldsymbol{\gamma} - \Phi\,(\Delta\mathbf{w}_{i,t-1} - \boldsymbol{\gamma})\right]\left[\Delta\mathbf{w}_{it} - \boldsymbol{\gamma} - \Phi\,(\Delta\mathbf{w}_{i,t-1} - \boldsymbol{\gamma})\right]', \end{aligned} \tag{4.56}$$

$i = 1, 2, \ldots, N$.[15] Using familiar results from the literature on GMM estimation developed by Hansen (1982) and noting that by assumption $\Delta\mathbf{w}_i$ are independently and identically distributed across i, the extended GMM estimators of $\boldsymbol{\kappa}$ can be computed by solving the following minimization problem:

$$\min_{\boldsymbol{\kappa}} \left[\mathbf{M}_N'(\boldsymbol{\kappa})\mathbf{A}_N'(\boldsymbol{\kappa})\mathbf{A}_N(\boldsymbol{\kappa})\mathbf{M}_N(\boldsymbol{\kappa})\right], \tag{4.57}$$

[15] In our Monte Carlo simulations in Section 10 below, we shall also consider the extended GMM estimator constructed by stacking the standard orthogonality conditions (4.6) and (4.23), and the homoskedasticity and stationarity implied moment conditions (4.36) and (4.37).

where

$$\mathbf{M}_N(\boldsymbol{\kappa}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{m}(\Delta\mathbf{w}_i,\boldsymbol{\kappa}), \qquad \mathbf{A}_N(\boldsymbol{\kappa}) = \mathbf{D}_N'(\boldsymbol{\kappa})\mathbf{W}_N^{-1}(\boldsymbol{\kappa}), \tag{4.58}$$

$$\mathbf{D}_N(\boldsymbol{\kappa}) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial\mathbf{m}(\Delta\mathbf{w}_i,\boldsymbol{\kappa})}{\partial\boldsymbol{\kappa}'}, \quad \text{and} \quad \mathbf{W}_N(\boldsymbol{\kappa}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{m}(\Delta\mathbf{w}_i,\boldsymbol{\kappa})\mathbf{m}'(\Delta\mathbf{w}_i,\boldsymbol{\kappa}). \tag{4.59}$$

Finally, the moment conditions relevant to the estimation of Ψ and Ω are the same as for the conventional GMM estimator incorporating the initial conditions.

It may also be noted that the solution of (4.57) with $\mathbf{W}_N(\boldsymbol{\kappa})$ given in (4.59) is not as efficient as the extended GMM estimator replacing $\mathbf{W}_N(\boldsymbol{\kappa})$ by $E[\mathbf{W}_N(\boldsymbol{\kappa})]$, that is, using a weighting matrix that takes into account all model restrictions.[16] However, the derivation of $E[\mathbf{W}_N(\boldsymbol{\kappa})]$ is not a trivial exercise, and will not be pursued here. Rather, in light of the various problems with GMM estimation that we have discussed, we revert our attention to the unconditional joint probability distribution of the first differences of the observations derived in Section 3, and establish the properties of the ML estimator based on it.

5 ML Estimation

The unconditional joint probability distribution of $\Delta\mathbf{w}_{i1}, \Delta\mathbf{w}_{i2}, \ldots, \Delta\mathbf{w}_{ip}, \Delta\mathbf{w}_{i,p+1}, \ldots, \Delta\mathbf{w}_{iT}$ was derived in Section 3. In this section, we again set $p = 1$ to simplify the exposition and facilitate comparison with the GMM estimators derived in Section 4.[17] In the case where $p = 1$ it may be simpler not to invoke the decomposition (3.1), but to work directly with the unconditional joint probability distribution of $\Delta\mathbf{w}_{i1}, \Delta\mathbf{w}_{i2}, \ldots, \Delta\mathbf{w}_{iT}$. From (4.1) and (4.3) it readily follows that under the assumption that ϵ_{it}, $t \leq T$, are normally distributed the unconditional joint probability distribution function of

$$\mathbf{r}_i = vec\Big(\ \Delta\mathbf{w}_{i1} - \boldsymbol{\gamma},\quad \Delta\mathbf{w}_{i2} - \boldsymbol{\gamma},\quad \cdots,\quad \Delta\mathbf{w}_{iT} - \boldsymbol{\gamma}\ \Big) \tag{5.1}$$

is given by

$$\prod_{i=1}^{N}(2\pi)^{-k/2}\left|\Sigma\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\mathbf{r}_i - \mathbf{H}_i\boldsymbol{\phi}\right)'\Sigma^{-1}\left(\mathbf{r}_i - \mathbf{H}_i\boldsymbol{\phi}\right)\right], \tag{5.2}$$

where $k = mT$,

$$\mathbf{H}_i = \mathbf{G}_i' \otimes \mathbf{I}_m, \tag{5.3}$$

[16] See also Amemiya (1973).

[17] Once again, the appendix discusses the more general case where $p > 1$.

$$\mathbf{G}_i = \left(\mathbf{0}_{m\times 1},\ \ \Delta\mathbf{w}_{i1} - \boldsymbol{\gamma},\ \ \Delta\mathbf{w}_{i2} - \boldsymbol{\gamma},\ \ \ldots,\ \ \Delta\mathbf{w}_{i,T-1} - \boldsymbol{\gamma} \right), \tag{5.4}$$

$\boldsymbol{\phi}$ is given by (4.53), $\boldsymbol{\phi} = vec(\Phi)$, and Σ is given by (4.31),

$$\Sigma = \begin{pmatrix} \Psi & -\Omega & & & & \\ -\Omega & 2\Omega & -\Omega & & \mathbf{0} & \\ & \ddots & \ddots & \ddots & & \\ & \mathbf{0} & & -\Omega & 2\Omega & -\Omega \\ & & & & -\Omega & 2\Omega \end{pmatrix},$$

with Ψ given by (4.5),

$$\Psi = (\mathbf{I}_m - \Phi)\left(\sum_{j=0}^{\infty} \Phi^j \Omega \Phi^{j\prime}\right)(\mathbf{I}_m - \Phi)' + \Omega.$$

The ML estimator of (5.2) is equivalent to finding $\boldsymbol{\gamma}$, Φ, and Ω to maximizing the unconditional log-likelihood function $l(\boldsymbol{\theta})$,

$$l(\boldsymbol{\theta}) = -\frac{Nk}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{H}_i\boldsymbol{\phi})'\,\Sigma^{-1}\,(\mathbf{r}_i - \mathbf{H}_i\boldsymbol{\phi}), \tag{5.5}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\gamma}',\ \ \boldsymbol{\phi}',\ \ \boldsymbol{\sigma}' \right)'$, with $\boldsymbol{\sigma} = vech(\Omega)$.[18]

Under the assumption that T is finite, the likelihood function associated with $\Delta\mathbf{w}_{it}$, $i = 1,2,\ldots,N$; $t = 1,2,\ldots,T$, is well defined irrespective of the location of the eigenvalues of Φ, and depends on a fixed number of parameters.[19] We then have the following proposition:

Proposition 5.1 *Under Assumptions (A1), (A2), and (A5),*[20] *and assuming that $\boldsymbol{\theta} \in \Theta$, where Θ is a compact subset of $\Re^{3m(m+1)/2}$ and the true parameter vector, $\boldsymbol{\theta}_0$, falls in the interior of Θ, the ML estimator associated with the likelihood function for $\Delta\mathbf{w}_{it}$, $i = 1,2,\ldots,N$; $t = 1,2,\ldots,T$, $\widehat{\boldsymbol{\theta}}_{ML}$, is consistent and asymptotically normally distributed when N tends to infinity.*

Note that normality assumptions on ε_{it}, $i = 1,2,\ldots,N$; $t \leq T$, are not needed for large N asymptotics.[21]

[18] The appendix in (A.16) provides the unconditional log-likelihood function derived using the decomposition (3.1). It may be verified that the ML estimator of $\boldsymbol{\theta}$ based on (5.5) is equivalent to the ML estimator of $\boldsymbol{\theta}$ based on (A.16) under $p = 1$.

[19] While in our exposition we have for simplicity not allowed for any of the roots of the determinantal equation (2.3) to fall inside the unit circle, the latter possibility can be readily accommodated by augmenting (A5) with the assumption that the explosive components in $\{\xi_{it}\}$ have started in a finite period in the past, with an independent and identical distribution across i, and that the covariance-stationary and pure unit root components in $\{\xi_{it}\}$ are generated by the relevant counterpart of (2.22). A note discussing this issue in further detail is available from the authors upon request.

[20] See also Footnote 19.

[21] For certain non-normal distributions generating $\{\varepsilon_{it}\}$, further regularity conditions may need to be verified. See, for example, van der Vaart (1998) for further discussion.

To compute the ML estimator, in the case of a (trend-) stationary PVAR it is useful to rewrite the infinite sum on the right-hand side of (4.5) yielding Ψ, and in all three cases it is useful to invoke an efficient scheme for the computation of the determinant and the inverse of Σ. To express Ψ in the case of a trend-stationary PVAR in non-recursive form, note from (4.5) that

$$\begin{aligned}\Phi\Psi\Phi' &= \Phi\Omega\Phi' + \sum_{j=0}^{\infty}\Phi^{j+1}\left(\mathbf{I}_m - \Phi\right)\Omega\left(\mathbf{I}_m - \Phi\right)'\Phi^{j+1\prime} \\ &= \Phi\Omega\Phi' + \Psi - \Omega - \left(\mathbf{I}_m - \Phi\right)\Omega\left(\mathbf{I}_m - \Phi\right)'. \end{aligned} \tag{5.6}$$

Vectorizing (5.6) it therefore follows that

$$vech\left(\Psi\right) = \mathbf{D}_m^{+}\left(\mathbf{I}_{m^2} - \Phi\otimes\Phi\right)^{-1}\mathbf{D}_m vech\left(2\Omega - \Phi\Omega - \Omega\Phi'\right), \tag{5.7}$$

where $\mathbf{D}_m$ is a $m^2 \times m\left(m+1\right)/2$ dimensional matrix, known as the duplication matrix, defined such that $vec\left(\mathbb{M}\right) = \mathbf{D}_m vech\left(\mathbb{M}\right)$ for any symmetric $m \times m$ dimensional matrix $\mathbb{M}$, and $\mathbf{D}_m^{+}$ is the generalized inverse of $\mathbf{D}_m$.[22] To compute the determinant of Σ, one may make use of the block-tridiagonal structure of Σ. Applying the block LDL' factorization to Σ, the latter may be factorized as[23]

$$\Sigma = \mathbb{A}_L\mathbb{A}_D\mathbb{A}_L', \tag{5.8}$$

where $\mathbb{A}_D$ is a block-diagonal matrix with j-th diagonal block given by $\mathbb{A}_D^{(1)} = \Psi^{-1}$, $\mathbb{A}_D^{(j)} = \left(2\Omega - \Omega\mathbb{A}_D^{(j-1)}\Omega\right)^{-1}$, $j = 2,3,\ldots,T$, and where $\mathbb{A}_L = \mathbb{A}_D^{-1} - \mathbb{A}_U$, with $\mathbb{A}_U$ being a block-subdiagonal matrix with all subdiagonal blocks equal to Ω. It then follows that

$$\begin{aligned}\det\left(\Sigma\right) &= \left[\det\left(\mathbb{A}_L\right)\right]^2\det\left(\mathbb{A}_D\right) = \frac{1}{\det\left(\mathbb{A}_D\right)} \\ &= \frac{1}{\det\left(\Psi\right)}\prod_{j=2}^{T}\det\left(\mathbb{A}_D^{(j)}\right). \end{aligned} \tag{5.9}$$

To compute the inverse of Σ, one may again make use of the block-tridiagonal structure of Σ. An efficient scheme is to adapt the recursions based on Bowden's procedure in Binder and Pesaran (2000) to Σ. The inverse of Σ can then be computed as

$$\Sigma_{jl}^{-1} = \left(T+1-l\right)\left[\left(j-1\right)\Omega^{-1}\Psi - \left(j-2\right)\mathbf{I}_m\right]\left[T\Psi - \left(T-1\right)\Omega\right]^{-1}, \qquad l \geq j, \tag{5.10}$$

[22] In practice, one will want to avoid computing the generalized inverse, $\mathbf{D}_m^{+}$, of $\mathbf{D}_m$ for large values of m, and instead determine $vech\left(\Psi\right)$ from $vec\left(\Psi\right)$ by selecting the appropriate elements of $vec\left(\Psi\right)$ through a simple element selection loop. In the case of a cointegrated PVAR, $\left(\mathbf{I}_{m^2} - \Phi\otimes\Phi\right)$ is singular; Ψ may then be computed using numerical recursions invoking an appropriate stopping rule to replace the infinite sum on the right-hand side of (4.5) by a finite sum.

[23] For further details see, for example, Binder and Pesaran (2000), who in the context of the solution of multivariate linear rational expectations models discuss the block LDU factorization, of which the block LDL' factorization is a special case.

and

$$\Sigma_{jl}^{-1} = (T+1-j)\left[T\Psi - (T-1)\Omega\right]^{-1}\left[(l-1)\Psi\Omega^{-1} - (l-2)\mathbf{I}_m\right], \qquad j > l, \tag{5.11}$$

where Σ_{jl}^{-1} denotes the jl-th block of Σ^{-1}, $j,l = 1,2,\ldots,T$.

A note describing further details of the numerical algorithm we use to compute the ML estimates is available from the authors upon request.

6 MD Estimation

In this section, we suggest a MD estimator based on the transformed model equations.[24] The MD estimator will be used as an initial estimator for the ML iterations.

Consider again the PVAR(1) model (2.12). Conditional on Σ, the ML estimator of Φ and γ is equivalent to the MD estimator that minimizes

$$\sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{H}_i\phi)' \Sigma^{-1} (\mathbf{r}_i - \mathbf{H}_i\phi), \tag{6.1}$$

where $\mathbf{r}_i$, $\mathbf{H}_i$, ϕ, and Σ are defined in (5.1), (5.3), (4.53), and (4.31), respectively. Therefore, conditional on γ and Σ, the MD estimator of Φ is given by

$$\widehat{\phi} = \left(\sum_{i=1}^{N} \mathbf{H}_i'\Sigma^{-1}\mathbf{H}_i\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{H}_i'\Sigma^{-1}\mathbf{r}_i\right), \tag{6.2}$$

and, conditional on Φ and Σ, the MD estimator of γ is equal to

$$\widehat{\gamma} = \left(N\mathbf{J}\Sigma^{-1}\mathbf{J}'\right)^{-1}\left[\sum_{i=1}^{N} \mathbf{J}\Sigma^{-1}(\Delta\mathbf{w}_i - \mathbf{L}_i\phi)\right], \tag{6.3}$$

where

$$\mathbf{J} = \left(\mathbf{I}_m,\ \mathbf{I}_m - \Phi',\ \mathbf{I}_m - \Phi',\ \ldots,\ \mathbf{I}_m - \Phi'\right), \tag{6.4}$$

$$\mathbf{L}_i = \mathbf{K}_i' \otimes \mathbf{I}_m, \quad \text{and} \quad \mathbf{K}_i = \left(\mathbf{0}_{m\times 1},\ \Delta\mathbf{w}_{i1},\ \Delta\mathbf{w}_{i2},\ \ldots,\ \Delta\mathbf{w}_{i,T-1}\right). \tag{6.5}$$

The variance-covariance matrix Σ may be (block) diagonalized as

$$\mathbb{U}\Sigma\mathbb{U}' = \mathcal{D} = \begin{pmatrix} \mathcal{D}_1 & & & \mathbf{0} \\ & \mathcal{D}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathcal{D}_T \end{pmatrix}, \tag{6.6}$$

[24] See Chamberlain (1984) for an early discussion of MD estimation of panel data models.

where $\mathcal{D}_1 = \mathbf{I}_m$, $\mathcal{D}_j = \mathbb{U}_{j+1}\Omega\mathbb{U}_j' = \left[j\Psi^{1/2}\Omega^{-1} - (j-1)\,\Psi^{-1/2}\right]\Omega\left[(j-1)\,\Psi^{1/2}\Omega^{-1} - (j-2)\,\Psi^{-1/2}\right]$, $j = 2,3,\ldots,T$, and

$$\mathbb{U} = \begin{pmatrix} \mathbb{U}_1 & & & & \\ \mathbb{U}_1 & \mathbb{U}_2 & & & \mathbf{0} \\ \mathbb{U}_1 & \mathbb{U}_2 & \mathbb{U}_3 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \mathbb{U}_1 & \mathbb{U}_2 & \mathbb{U}_3 & \cdots & \mathbb{U}_T \end{pmatrix}, \tag{6.7}$$

with $\mathbb{U}_1 = \Psi^{-1/2}$, $\mathbb{U}_2 = \Psi^{1/2}\Omega^{-1}$, $\mathbb{U}_j = 2\mathbb{U}_{j-1} - \mathbb{U}_{j-2} = (j-1)\,\Psi^{1/2}\Omega^{-1} - (j-2)\,\Psi^{-1/2}$, $j = 3,4,\ldots,T$.[25] Therefore, one may consider the following iterative scheme to obtain the MD estimator:

Step 1: Form initial estimates of γ, Φ, and Ω from the moment conditions $E\left(\Delta\mathbf{w}_{it} - \gamma\right) = \mathbf{0}_{m\times 1}$, $t = 1,2,\ldots,T$, (4.50), (4.51), and (4.17) as follows:

$$\widehat{\gamma}^{(0)} = \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\Delta\mathbf{w}_{it}, \tag{6.8}$$

$$\frac{1}{(T-1)N}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[\Delta\mathbf{w}_{it} - \widehat{\gamma}^{(0)} - \widehat{\Phi}^{(0)}\left(\Delta\mathbf{w}_{i,t-1} - \widehat{\gamma}^{(0)}\right)\right]\left(\Delta\mathbf{w}_{i,t-1} - \widehat{\gamma}^{(0)}\right)' + \widehat{\Omega}^{(0)} = \mathbf{0}_m, \tag{6.9}$$

or

$$\begin{aligned}\widehat{\Phi}^{(0)} &= \left[\sum_{t=2}^{T}\sum_{i=1}^{N}\left(\Delta\mathbf{w}_{it} - \widehat{\gamma}^{(0)}\right)\left(\Delta\mathbf{w}_{i,t-1} - \widehat{\gamma}^{(0)}\right)' + \widehat{\Omega}^{(0)}\right] \\ &\quad\cdot\left[\sum_{t=2}^{T}\sum_{i=1}^{N}\left(\Delta\mathbf{w}_{i,t-1} - \widehat{\gamma}^{(0)}\right)\left(\Delta\mathbf{w}_{i,t-1} - \widehat{\gamma}^{(0)}\right)'\right]^{-1},\end{aligned} \tag{6.10}$$

and

$$\begin{aligned}\widehat{\Omega}^{(0)} &= \frac{1}{2N(T-1)}\sum_{t=2}^{T}\sum_{i=1}^{N}\left[\Delta\mathbf{w}_{it} - \widehat{\gamma}^{(0)} - \widehat{\Phi}^{(0)}\left(\Delta\mathbf{w}_{i,t-1} - \widehat{\gamma}^{(0)}\right)\right] \\ &\quad\cdot\left[\Delta\mathbf{w}_{it} - \widehat{\gamma}^{(0)} - \widehat{\Phi}^{(0)}\left(\Delta\mathbf{w}_{i,t-1} - \widehat{\gamma}^{(0)}\right)\right]'.\end{aligned} \tag{6.11}$$

It is clear from (6.10) and (6.11) that $\widehat{\Phi}^{(0)}$ needs to be computed by iterative techniques. To initialize the iterations, one will need to make an initial guess for $\widehat{\Omega}$.[26]

[25] There are alternative procedures available to achieve such a block diagonalization of Σ. One alternative is to employ the block LDL' factorization used in Section 5 to compute the determinant of Σ. Using $\mathbf{A}_D$ and $\mathbf{A}_L$ as defined in Section 5, we have that $\mathbf{A}_D = \mathbf{A}_L^{-1}\Sigma(\mathbf{A}_L')^{-1}$, where $\mathbf{A}_L^{-1}$ is lower triangular.

[26] As was discussed in Section 4, the solution of the homoskedasticity implied moment conditions (4.50) and (4.51) is in general not unique. But for the purpose of initializing the MD iterations in our Monte Carlo simulations reported in Section 10 below we found that if all elements of the initial estimate of $\widehat{\Omega}$ were chosen sufficiently small, then the eigenvalues of $\widehat{\Phi}^{(0)}$ computed by iterating on (6.10) and (6.11) tend to fall inside the unit circle if the eigenvalues of Φ fall inside the unit circle.

Step 2: From the initial estimates of Φ, γ, and Ω transform $\mathbf{r}_i$ into $\mathbf{r}_i^* = \mathcal{D}^{-1/2}\mathbb{U}\mathbf{r}_i$, and $vec(\mathbf{G}_i)$ into $vec(\mathbf{G}_i^*) = \mathcal{D}^{-1/2}\mathbb{U}vec(\mathbf{G}_i)$, and estimate Φ and γ by minimizing the Ordinary Least Squares objective function

$$\sum_{i=1}^{N} (\mathbf{r}_i^* - \mathbf{H}_i^*\phi)'(\mathbf{r}_i^* - \mathbf{H}_i^*\phi), \tag{6.12}$$

where $\mathbf{H}_i^* = \mathbf{G}_i^{*\prime} \otimes \mathbf{I}_m$. Numerically, this may be achieved in fast and accurate manner by solving two unit triangular equation systems each for Φ and γ. For example, factorizing $\sum_{i=1}^{N} \mathbf{H}_i'\Sigma^{-1}\mathbf{H}_i$ as

$$\sum_{i=1}^{N} \mathbf{H}_i'\Sigma^{-1}\mathbf{H}_i = \mathbb{B}_L\mathbb{B}_D\mathbb{B}_L', \tag{6.13}$$

where $\mathbb{B}_L$ is unit lower triangular and $\mathbb{B}_D$ diagonal (see, for example, Golub and van Loan, 1996, for an algorithm to achieve such a factorization), to obtain an estimate of Φ, one needs to solve the unit upper triangular equation system

$$\mathbb{B}_L'\phi = \mathbb{B}_D^{-1}\mathbf{z}, \tag{6.14}$$

with z being the solution to the unit lower triangular equation system

$$\mathbb{B}_L\mathbf{z} = \sum_{i=1}^{N} \left(\mathcal{D}^{-1/2}\mathbb{U}\mathbf{H}_i\right)' \left(\mathcal{D}^{-1/2}\mathbb{U}\mathbf{r}_i\right). \tag{6.15}$$

Step 3: Re-estimate Ω based on the revised estimates of Φ and γ, using the moment condition (4.17).

Step 4: Repeat Steps 2 to 3 until the estimates converge.

The MD estimator is consistent. The consistency of $\widehat{\gamma}$ can be easily seen from (6.8). Therefore, we shall treat γ as known and consider the consistency of $\widehat{\Phi}$. Note that

$$\widehat{\phi} - \phi = \left(\sum_{i=1}^{N} \mathbf{H}_i'\Sigma^{-1}\mathbf{H}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{H}_i'\Sigma^{-1}\vartheta_i\right), \tag{6.16}$$

where

$$\vartheta_i = vec\left(\ \eta_{i1},\ \ \triangle\varepsilon_{i2},\ \ \triangle\varepsilon_{i3},\ \ \ldots,\ \ \triangle\varepsilon_{iT}\ \right), \tag{6.17}$$

with η_{i1} given by (4.2). The numerator of the second term of (6.16) divided by N has the form

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}(\mathbb{U}_{t+1}\Omega\mathbb{U}_t')^{-1} \left(\sum_{j=1}^{t-1}\mathbb{U}_j \triangle \mathbf{w}_{ij}\right) \left(\mathbb{U}_1\eta_{i1} + \sum_{j=2}^{t}\mathbb{U}_j \triangle \varepsilon_{ij}\right)'. \tag{6.18}$$

Substituting $\Delta \mathbf{w}_{it} - \boldsymbol{\gamma} = \Phi^{t-1}(\Delta \mathbf{w}_{i1} - \boldsymbol{\gamma}) + \sum_{j=0}^{t-2} \Phi^j \Delta \boldsymbol{\varepsilon}_{i,t-j}$, and noting that $\mathbb{U}_{j+1} - 2\mathbb{U}_j + \mathbb{U}_{j-1} = \mathbf{0}_m$, one can show that

$$E\left(\sum_{j=1}^{t-1} \mathbb{U}_j \Delta \mathbf{w}_{ij}\right)\left(\mathbb{U}_1 \boldsymbol{\eta}_{i1} + \sum_{j=2}^{t} \mathbb{U}_j \Delta \boldsymbol{\varepsilon}_{ij}\right)' = \mathbf{0}_m. \tag{6.19}$$

Therefore, by a law of large numbers, (6.16) converges to $\mathbf{0}_{m^2 \times 1}$ as $N \to \infty$. The MD estimator does not require normality assumptions on $\boldsymbol{\varepsilon}_{it}$, $i = 1,2,\ldots,N$; $t \leq T$, to remain valid.

Furthermore, conditional on $\boldsymbol{\gamma}$ and Σ, the MD estimator of $\boldsymbol{\phi}$ is identical to the ML estimator of $\boldsymbol{\phi}$ with asymptotic variance-covariance matrix given by

$$\left[\sum_{i=1}^{N} (\mathbf{G}_i \otimes \mathbf{I}_m) \Sigma^{-1} (\mathbf{G}_i' \otimes \mathbf{I}_m)\right]^{-1}. \tag{6.20}$$

When Σ is unknown, the asymptotic variance-covariance matrices of the ML and MD estimators of Φ do not converge to (6.20) because when lagged dependent variables appear as regressors, the estimation of Φ and Σ is not asymptotically independent (Amemiya and Fuller, 1967). The asymptotic variance-covariance matrix of the feasible MD estimator of Φ is equal to the sum of (6.20) and a positive semi-definite matrix attributable to the estimation error in $\widehat{\Sigma}$ (Hsiao, Pesaran, and Tahmiscioglu, 1999).

7 Estimating Long-Run Relations

In this section we consider ML and MD estimation of the PVAR(1) model (2.12) when $rank(\Pi) = r$, $r = 1,2,\ldots,m-1$, and there are thus r cointegrating relations present. Subtracting $\Delta \mathbf{w}_{i,t-1}$ from both sides of (4.3) yields

$$\Delta^2 \mathbf{w}_{it} = \Pi(\Delta \mathbf{w}_{i,t-1} - \boldsymbol{\gamma}) + \Delta \boldsymbol{\varepsilon}_{it}, \qquad t = 2,3,\ldots,T, \tag{7.1}$$

with Π given by (2.19), $\Pi = -(\mathbf{I}_m - \Phi)$. Combining (7.1) with (4.1),

$$\Delta \mathbf{w}_{i1} - \boldsymbol{\gamma} = \boldsymbol{\varepsilon}_{i1} + \Pi \sum_{j=0}^{\infty} (\Pi + \mathbf{I}_m)^j \boldsymbol{\varepsilon}_{i,-j},$$

we have the logarithm of the joint probability distribution function of

$$\mathbf{s}_i = vec\left(\ \Delta \mathbf{w}_{i1} - \boldsymbol{\gamma},\ \ \Delta^2 \mathbf{w}_{i2},\ \ \Delta^2 \mathbf{w}_{i3},\ \ \ldots,\ \ \Delta^2 \mathbf{w}_{iT}\ \right) \tag{7.2}$$

as

$$l(\boldsymbol{\theta}) = -\frac{Nk}{2} \ln(2\pi) - \frac{N}{2} \ln|\Sigma| - \frac{1}{2} \sum_{i=1}^{N} (\mathbf{s}_i - \mathbf{H}_i \boldsymbol{\pi})' \Sigma^{-1} (\mathbf{s}_i - \mathbf{H}_i \boldsymbol{\pi}), \tag{7.3}$$

where $\theta = \left(\gamma', \ \pi', \ \sigma' \right)'$, $\pi = vec(\Pi)$, and $\mathbf{H}_i$ and Σ are defined by (5.3) and (4.31), respectively.

Decomposing the matrix Π as $\Pi = \alpha\beta'$, where α and β as before are $m \times r$ matrices of full column rank, since $\alpha\beta' = \alpha\mathbb{K}\mathbb{K}^{-1}\beta'$ for any $r \times r$ nonsingular matrix $\mathbb{K}$, one needs r^2 restrictions to restrict $\mathbb{K} = \mathbf{I}_r$. (The maximum of the likelihood function under $rank(\Pi) = r$ is invariant to the choice of $\mathbb{K}$.) A convenient form for the identification of β is to let

$$\beta = \mathbb{H}\delta + \mathbf{b}, \tag{7.4}$$

where $\mathbb{H}$ and $\mathbf{b}$ are, respectively, $m \times q$ and $m \times r$ matrices, both with known coefficients, and δ is a $q \times r$ matrix with unknown coefficients. For example, if one chooses the Phillips (1991) normalization restriction that

$$\beta = \left(\mathbf{I}_r, \ \tilde{\beta}' \right)', \tag{7.5}$$

where $\tilde{\beta}'$ is an $r \times (m-r)$ matrix with unrestricted coefficients, then $\mathbb{H} = \left(\mathbf{0}_{(m-r)\times r}, \ \mathbf{I}_{m-r} \right)'$, $\mathbf{b} = \left(\mathbf{I}_r, \ \mathbf{0}_{r\times(m-r)} \right)'$, and $\delta = \tilde{\beta}$.

As for MD estimation, we note that conditional on β, γ, and Σ, the ML estimator of α is equal to

$$vec(\hat{\alpha}) = \left(\sum_{i=1}^{N} \tilde{\mathbf{H}}_i' \Sigma^{-1} \tilde{\mathbf{H}}_i \right)^{-1} \left(\sum_{i=1}^{N} \tilde{\mathbf{H}}_i' \Sigma^{-1} \mathbf{s}_i \right), \tag{7.6}$$

where

$$\tilde{\mathbf{H}}_i = \mathbf{G}_i'\beta \otimes \mathbf{I}_m, \tag{7.7}$$

with $\mathbf{G}_i$ given by (5.4). Substituting $vec(\beta) = (\mathbf{I}_m \otimes \mathbb{H})\, vec(\delta) + vec(\mathbf{b})$ into (7.3), the ML estimator of δ conditional on α, γ, and Σ is equal to

$$vec\left(\hat{\delta}'\right) = \left[\sum_{i=1}^{N} \left(\mathbb{H}'\mathbf{G}_i \otimes \alpha' \right) \Sigma^{-1} \left(\mathbf{G}_i'\mathbb{H} \otimes \alpha \right) \right]^{-1} \left[\sum_{i=1}^{N} \left(\mathbb{H}'\mathbf{G}_i \otimes \alpha' \right) \Sigma^{-1} \mathbf{s}_i^* \right], \tag{7.8}$$

where

$$\mathbf{s}_i^* = \mathbf{s}_i - \left(\mathbf{G}_i' \otimes \alpha \right) vec\left(\mathbf{b}' \right). \tag{7.9}$$

Therefore, one may obtain MD estimates of α and β as initial estimates for the ML iterations by first estimating β using r cross-section regressions in the elements of $\mathbf{w}_{it}$ under the normalization restriction (7.5), and then iterating on (7.6) and (7.8), revising the estimates of γ and Ω in each iteration step as in Section 6.

The estimation of the cointegrating relations as derived above is based on the assumption that the time dimension of the panel, T, exceeds for a general PVAR model the latter's autoregressive order, p. Sometimes, however, there may not be sufficiently many time-series observations available for T to exceed p. In the event that T is less than p, one cannot consistently estimate the autoregressive coefficients in Φ_j, $j = 1,2,\ldots,p$. A natural question to pose, however, is whether it would still be possible to estimate the cointegrating relations. Unfortunately, this is not the case. It may be verified that when $T < p$, then the variance-covariance matrix of the disturbances of the transformed model equations is unrestricted, which in turn leads to a lack of identification. This is perhaps most easily seen by examining identification if the true data generating process is a PVAR(p) with $p > 1$, but estimation is based on a PVAR(1). The unconditional log-likelihood function (5.5) then applies, except that the variance-covariance matrix Σ is unrestricted.

Proposition 7.1 *The PVAR(1) model (2.12) with unrestricted variance-covariance matrix Σ is not identified.*

Proof of Proposition 7.1: The unconditional log-likelihood function of the PVAR(1) model (2.12) with unrestricted variance-covariance matrix Σ is given by

$$l^*(\theta) = -\frac{Nk}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{H}_i\phi)'\Sigma^{-1}(\mathbf{r}_i - \mathbf{H}_i\phi), \tag{7.10}$$

where $\mathbf{r}_i$ and $\mathbf{H}_i$ are given by (5.1) and (5.3), respectively, and $\theta = \left(\gamma', \ \phi', \ \sigma^{*\prime}\right)'$, with $\sigma^* = vech(\Sigma)$. Concentrating this log-likelihood function in terms of γ and ϕ, one obtains

$$l^*(\gamma,\phi) = -\frac{Nk}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma(\gamma,\phi)| - \frac{N}{2}, \tag{7.11}$$

where

$$\Sigma(\gamma,\phi) = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \mathbf{H}_i\phi)(\mathbf{r}_i - \mathbf{H}_i\phi)'. \tag{7.12}$$

The concentrated log-likelihood function, $l^*(\gamma,\phi)$, does not depend on ϕ, and hence it will not be possible to identify ϕ from the sample observations. To see this, first note that the variance-covariance matrix Σ can be written as

$$\Sigma(\gamma,\phi) = \begin{pmatrix} \sigma^*_{11} & \sigma^*_{12} & \sigma^*_{13} & \cdots & \sigma^*_{1T} \\ \cdot & \sigma^*_{22} & \sigma^*_{23} & \cdots & \sigma^*_{2T} \\ & & & \ddots & \\ \cdot & \cdot & \cdot & \cdots & \sigma^*_{TT} \end{pmatrix}, \tag{7.13}$$

where

$$\sigma^*_{11} = \varpi_{11}, \tag{7.14}$$

$$\sigma^*_{1l} = \varpi_{1l} - \varpi_{1,l-1}\Phi', \qquad l = 2,3,\ldots,T, \tag{7.15}$$

$$\sigma^*_{jl} = \varpi_{jl} - \varpi_{j,l-1}\Phi' - \Phi\varpi_{j-1,l} + \Phi\varpi_{j-1,l-1}\Phi', \qquad j = 2,3,\ldots,T;\ l = j, j+1,\ldots,T, \tag{7.16}$$

and

$$\varpi_{jl} = N^{-1}\sum_{i=1}^{N}(\triangle \mathbf{w}_{ij} - \boldsymbol{\gamma})(\triangle \mathbf{w}_{il} - \boldsymbol{\gamma})', \qquad j = 1,2,\ldots,T;\ l = 1,2,\ldots,T. \tag{7.17}$$

Σ may now be factorized as

$$\Sigma(\boldsymbol{\gamma}, \boldsymbol{\phi}) = \mathbb{L}(\boldsymbol{\phi})\,\Upsilon(\boldsymbol{\gamma})\,\mathbb{L}(\boldsymbol{\phi})', \tag{7.18}$$

where

$$\Upsilon(\boldsymbol{\gamma}) = \begin{pmatrix} \varpi_{11} & \varpi_{12} & \varpi_{13} & \cdots & \varpi_{1T} \\ \varpi_{21} & \varpi_{22} & \varpi_{23} & \cdots & \varpi_{2T} \\ & & & \ddots & \\ \varpi_{T1} & \varpi_{T2} & \varpi_{T3} & \cdots & \varpi_{TT} \end{pmatrix}, \tag{7.19}$$

and

$$\mathbb{L}(\boldsymbol{\phi}) = \begin{pmatrix} \mathbf{I}_m & & & \\ -\Phi & \mathbf{I}_m & & \mathbf{0} \\ & \ddots & \ddots & \\ \mathbf{0} & & -\Phi & \mathbf{I}_m \end{pmatrix}. \tag{7.20}$$

Note that $|\mathbb{L}(\boldsymbol{\phi})| = 1$, and $|\Upsilon(\boldsymbol{\gamma})|$ is a function of $\boldsymbol{\gamma}$ only. Therefore,

$$|\Sigma(\boldsymbol{\gamma}, \boldsymbol{\phi})| = |\mathbb{L}(\boldsymbol{\phi})|\,|\Upsilon(\boldsymbol{\gamma})|\,|\mathbb{L}(\boldsymbol{\phi})'| = |\Upsilon(\boldsymbol{\gamma})|, \tag{7.21}$$

and

$$l^*(\boldsymbol{\gamma}, \boldsymbol{\phi}) = -\frac{Nk}{2}\ln(2\pi) - \frac{N}{2}\ln|\Upsilon(\boldsymbol{\gamma})| - \frac{N}{2}, \tag{7.22}$$

which does not depend on $\boldsymbol{\phi}$. ■

8 Hypothesis Testing

In this section we consider the issue of hypothesis testing for the PVAR(1) model (2.12). The time-series properties of $\mathbf{w}_{it}$ that are of interest, for instance, whether $\mathbf{w}_{it}$ is (trend) stationary or I(1), and whether $\mathbf{w}_{it} - \boldsymbol{\mu}_i - \boldsymbol{\gamma}t$, if I(1), contains cointegrating relations, can be formulated in

terms of restrictions on Φ. Since the ML estimator of Φ is asymptotically normally distributed irrespective of whether the elements of $\mathbf{w}_{it}$ are I(1) or I(0), standard testing procedures such as the likelihood ratio, the Wald and the Lagrange multiplier methods can be applied.

In order to carry out cointegration analysis and be able to interpret the rank of the matrix $\Pi = -(\mathbf{I}_m - \Phi)$ as the number of linearly independent cointegrating relations, it is necessary to know whether each of the variables in $\mathbf{w}_{it}$ follows an I(1) process. Our framework can be easily adapted to test for unit roots in short panel univariate autoregressive models, namely $m = 1$. In this case, for $p = 1$ we have that (see (2.12))

$$w_{it} = (1-\phi)\,\mu_i + \phi\gamma + (1-\phi)\,\gamma t + \phi w_{i,t-1} + \varepsilon_{it}, \qquad \varepsilon_{it} \overset{iid}{\sim} \left(0, \sigma^2\right), \tag{8.1}$$

where w_{it} is now a scalar variable, $i = 1,2,\ldots,N$.[27] For testing the unit root hypothesis

$$H_0 : \phi = 1 \qquad \text{vs.} \qquad H_1 : \phi < 1, \tag{8.2}$$

the log-likelihood function is given by (see (5.5))

$$l(\boldsymbol{\theta}_1) = -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{f}_i - \mathbf{g}_i\phi)'\,\Sigma^{-1}\,(\mathbf{f}_i - \mathbf{g}_i\phi), \tag{8.3}$$

where $\boldsymbol{\theta}_1 = \left(\ \gamma,\ \ \phi,\ \ \sigma^2\ \right)'$,

$$\mathbf{f}_i = vec\left(\ \Delta w_{i1} - \gamma,\ \ \Delta w_{i2} - \gamma,\ \ \Delta w_{i3} - \gamma,\ \ \ldots,\ \ \Delta w_{iT} - \gamma\ \right), \tag{8.4}$$

$$\mathbf{g}_i = vec\left(\ 0,\ \ \Delta w_{i1} - \gamma,\ \ \Delta w_{i2} - \gamma,\ \ \ldots,\ \ \Delta w_{i,T-1} - \gamma\ \right), \tag{8.5}$$

and

$$\Sigma = \sigma^2 \begin{pmatrix} 2/(1+\phi) & -1 & & & & \\ -1 & 2 & -1 & & 0 & \\ & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 \\ & 0 & & & -1 & 2 \end{pmatrix}. \tag{8.6}$$

Under H_0 we have

$$l(\boldsymbol{\theta}_0) = -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln\left|\Sigma_{|\phi=1}\right| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{f}_i - \mathbf{g}_i)'\left(\Sigma_{|\phi=1}\right)^{-1}(\mathbf{f}_i - \mathbf{g}_i), \tag{8.7}$$

[27] As for unit root testing in the time-series context, the appropriate order of augmentation of w_{it} is important for the validity of the test. In practice one may therefore need to consider higher-order cases as well. Here we confine ourselves to $p = 1$ for simplicity.

where $\theta_0 = \left(\gamma, \ \sigma^2 \right)'$. Denoting by LL^u the maximum of the log-likelihood under H_1, and by LL^r the maximum of the log-likelihood under H_0, $2(LL^u - LL^r)$ is asymptotically chi-square distributed with one degree of freedom.[28] A Wald type statistic of testing H_0 versus H_1 will be

$$t_\phi = \frac{\widehat{\phi} - 1}{\text{se}(\widehat{\phi})}, \tag{8.8}$$

where $\text{se}(\widehat{\phi})$ denotes the standard error of $\widehat{\phi}$; t_ϕ is asymptotically distributed as a standard normal variate, and it can thus be taken into account that the alternative hypothesis, H_1, is one-sided.

In contrast to the Least Squares Dummy Variables based unit root test for dynamic panel data models with finite time dimension proposed by Harris and Tzavalis (1999), note that the unit root test statistics we propose here are not based on an inconsistent estimator, and thus standard chi-square and normal distribution theory are applicable. It may also be noted that the tests we propose here are readily extended to the case where $p > 1$, whereas such an extension for the unit root test proposed by Harris and Tzavalis (1999) would require computing the necessary bias adjustments for the test statistics for each value of p separately.[29]

The natural next step is to test for cointegration. Consider again the PVAR(1) model (2.12) in the m variables $\mathbf{w}_{it}$, now assumed to be I(1). The hypothesis that $\mathbf{w}_{it} - \boldsymbol{\mu}_i - \boldsymbol{\gamma} t$ is cointegrated with rank r versus rank $r+1$, $r = 0, 1, \ldots, m-1$, can be formulated as

$$H_r : \Phi = \mathbf{I}_m + \boldsymbol{\alpha}_r \boldsymbol{\beta}_r' \qquad \text{vs.} \qquad H_{r+1} : \Phi = \mathbf{I}_m + \boldsymbol{\alpha}_{r+1} \boldsymbol{\beta}_{r+1}', \tag{8.9}$$

where $\boldsymbol{\alpha}_r$ and $\boldsymbol{\beta}_r$ are $m \times r$ matrices of full column rank r. As discussed in Section 7, to achieve identification of $\boldsymbol{\beta}_r$ one may employ the Phillips (1991) normalization restriction (7.5),

$$\boldsymbol{\beta}_r = \left(\mathbf{I}_r, \ \widetilde{\boldsymbol{\beta}}_r' \right),$$

where $\widetilde{\boldsymbol{\beta}}_r'$ is an $r \times (m-r)$ matrix with unrestricted coefficients. The likelihood ratio test statistic of H_r versus H_{r+1} is asymptotically chi-square distributed with $(m-r)^2 - (m-r-1)^2 = 2(m-r) - 1$ degrees of freedom. (Imposing Π to be of rank r leaves $m^2 - (m-r)^2$ unrestricted coefficients in Π.)[30]

[28] A drawback of the likelihood ratio test is that it does not allow one to readily take into account that the alternative hypothesis, H_1, is one-sided.

[29] For panel unit root tests under $p = 1$ where the deterministic trend coefficients are allowed to differ across cross-sectional units, see, for example, Im, Pesaran, and Shin (1997) and Harris and Tzavalis (1999). As noted in Footnote 4, differential deterministic trend coefficients could be accommodated in our setting also. For panel unit roots tests when the time dimension is large and possible slope heterogeneity can be allowed for, see, for example, Levin and Lin (1993) and Im, Pesaran, and Shin (1997).

[30] Note that in the special case where $m = 1$ and $r = 0$, (8.9) reduces to the likelihood ratio based unit root test discussed above.

Additional parameter restrictions or overidentifying restrictions can be formulated in terms of

$$vec(\Phi) = \mathbb{G}\varkappa + \mathbf{f}, \tag{8.10}$$

where $\mathbb{G}$ is an $m^2 \times q$ matrix and $\mathbf{f}$ an $m^2 \times 1$ vector, both with known elements, and $\varkappa$ is a $q \times 1$ vector of free parameters. A likelihood ratio test of (8.10) will be asymptotically chi-square distributed with $m^2 - q$ degrees of freedom.

9 Conditional Inference

In applied economic problems the focus of the analysis is often on conditional models where a subset of the variables in $\mathbf{w}_{it}$ are modelled in terms of the remaining variables. In this section we consider the conditions under which such an analysis is valid for the PVAR(1) model (2.12), and show how the methodology developed in the previous sections can be adapted to the analysis of conditional PVAR models.

In the time-series context, where the sample sizes typically are large, the conditions under which the marginal distribution of $\mathbf{x}_t$ can be ignored when estimating parameters of interest that enter the conditional distribution of $\mathbf{y}_t$ given $\mathbf{x}_t$ are discussed in detail, for example, by Engle, Hendry, and Richard (1983). These conditions are, however, obtained for given initial values, $\mathbf{w}_0 = \left(\mathbf{y}_0', \ \mathbf{x}_0' \right)'$. As should be clear from our earlier discussion, in the context of dynamic panel data models with small T the effect of the initial values, $\mathbf{w}_{i0}$, cannot be ignored and unless appropriately accounted for will make the exogeneity analysis subject to an incidental parameters problem. To formulate an appropriate notion of conditioning (exogeneity) for the PVAR(1) model (2.12), we consider the transformed version of this model which is free of the incidental parameters problem. For this purpose we work with the following system of equations: for the initial observations we have (see (4.1))

$$\Delta\mathbf{w}_{i1} - \boldsymbol{\gamma} = \boldsymbol{\varepsilon}_{i1} - (\mathbf{I}_m - \Phi)\,\boldsymbol{\xi}_{i0}, \qquad i = 1,2,\ldots,N,$$

and for the subsequent periods, $t = 2,3,\ldots,T$, we have (see (4.3))

$$\Delta\mathbf{w}_{it} - \boldsymbol{\gamma} = \Phi\,(\Delta\mathbf{w}_{i,t-1} - \boldsymbol{\gamma}) + \Delta\boldsymbol{\varepsilon}_{it}, \qquad i = 1,2,\ldots,N.$$

Partitioning $\Delta\mathbf{w}_{it}$ into $\Delta\mathbf{w}_{it} = \left(\Delta\mathbf{y}_{it}' \quad \Delta\mathbf{x}_{it}' \right)'$, where $\mathbf{y}_{it}$ is $m_y \times 1$ and $\mathbf{x}_{it}$ is $m_x \times 1$, and $m_y + m_x = m$, it is of interest to determine under what conditions knowledge of the marginal probability distribution function of $\Delta\mathbf{x}_{it}$, $f\left[\Delta\mathbf{x}_{it}|\boldsymbol{\kappa}_2(\boldsymbol{\theta})\right]$, $t = 1,2,\ldots,T$; $i = 1,2,\ldots,N$, is redundant for purposes of inference regarding the parameters of interest of the conditional probability distribution function of $\Delta\mathbf{y}_{it}$ given $\Delta\mathbf{x}_{it}$, $f\left[\Delta\mathbf{y}_{it}|\Delta\mathbf{x}_{it},\boldsymbol{\kappa}_1(\boldsymbol{\theta})\right]$, $t = 1,2,\ldots,T$; $i = 1,2,\ldots,N$. In other words, for purposes of maximizing

$$\prod_{i=1}^{N}\left\{ f\left(\Delta\mathbf{y}_{i1},\Delta\mathbf{x}_{i1}|\boldsymbol{\theta}\right)\left[\prod_{t=2}^{T} f\left(\Delta\mathbf{y}_{it},\Delta\mathbf{x}_{it}|\mathcal{I}_{i,t-1},\boldsymbol{\theta}\right)\right]\right\} \tag{9.1}$$

with respect to $\kappa_1(\theta)$, under what conditions is it sufficient to consider

$$\prod_{i=1}^{N}\left(f\left[\Delta\mathbf{y}_{i1}|\Delta\mathbf{x}_{i1},\kappa_1(\theta)\right]\left\{\prod_{t=2}^{T}f\left[\Delta\mathbf{y}_{it}|\Delta\mathbf{x}_{it},\mathcal{I}_{i,t-1},\kappa_1(\theta)\right]\right\}\right), \tag{9.2}$$

where $\mathcal{I}_{i,t-1}=\left(\Delta\mathbf{w}_{i,t-1},\Delta\mathbf{w}_{i,t-2},\ldots,\Delta\mathbf{w}_{i1}\right)$.

Partitioning the disturbance vector $\boldsymbol{\varepsilon}_{it}$ and its variance-covariance matrix Ω conformably with the partition of $\Delta\mathbf{w}_{it}$ into $\Delta\mathbf{y}_{it}$ and $\Delta\mathbf{x}_{it}$ as

$$\boldsymbol{\varepsilon}_{it}=\begin{pmatrix}\boldsymbol{\varepsilon}_{yit}\\ \boldsymbol{\varepsilon}_{xit}\end{pmatrix},\qquad\text{and}\qquad\Omega=\begin{pmatrix}\Omega_{yy}&\Omega_{yx}\\ \Omega_{xy}&\Omega_{xx}\end{pmatrix},$$

one may write $\boldsymbol{\varepsilon}_{yit}$ conditionally in terms of $\boldsymbol{\varepsilon}_{xit}$ as

$$\boldsymbol{\varepsilon}_{yit}=\mathbf{B}_y\boldsymbol{\varepsilon}_{xit}+\boldsymbol{\upsilon}_{it}, \tag{9.3}$$

where

$$\mathbf{B}_y=\Omega_{yx}\Omega_{xx}^{-1}, \tag{9.4}$$

and $\boldsymbol{\upsilon}_{it}\overset{iid}{\sim}\left(0_{m_y},\Omega_{\upsilon}\right)$, with

$$\Omega_{\upsilon}=\Omega_{yy}-\mathbf{B}_y\Omega_{xy}. \tag{9.5}$$

Note that $\boldsymbol{\upsilon}_{it}$ is independent of $\boldsymbol{\varepsilon}_{xit}$ across i and over t, $i=1,2,\ldots,N$; $t\leq T$. Substituting (9.3) into the error correction representation (2.10) under $p=1$ yields

$$\begin{pmatrix}\Delta\mathbf{y}_{it}-\mathbf{B}_y\Delta\mathbf{x}_{it}\\ \Delta\mathbf{x}_{it}\end{pmatrix}=\begin{pmatrix}\mathbf{a}_y\\ \boldsymbol{\gamma}_x\end{pmatrix}+\begin{pmatrix}\Pi_y-\mathbf{B}_y\Pi_x\\ \Pi_x\end{pmatrix}\boldsymbol{\xi}_{i,t-1}+\begin{pmatrix}\boldsymbol{\upsilon}_{it}\\ \boldsymbol{\varepsilon}_{xit}\end{pmatrix}, \tag{9.6}$$

$i=1,2,\ldots,N$; $t=1,2,\ldots,T$, where

$$\mathbf{a}_y=\boldsymbol{\gamma}_y+\mathbf{B}_y\boldsymbol{\gamma}_x, \tag{9.7}$$

and where we have partitioned $\Pi=\begin{pmatrix}\Pi_y',&\Pi_x'\end{pmatrix}'$ and $\boldsymbol{\gamma}=\begin{pmatrix}\boldsymbol{\gamma}_y',&\boldsymbol{\gamma}_x'\end{pmatrix}'$ conformably with the partition of $\mathbf{w}_{it}$ into $\mathbf{y}_{it}$ and $\mathbf{x}_{it}$. To examine under what conditions the last m_x equations in (9.6) are redundant for purposes of inference regarding the parameters of interest in the first m_y equations in (9.6), let us write these equations separately for $t=1$ and $t=2,3,\ldots,T$. For $t=1$ we have the conditional model for $\Delta\mathbf{y}_{i1}$ given $\Delta\mathbf{x}_{i1}$ and the marginal model for $\Delta\mathbf{x}_{i1}$,

$$\Delta\mathbf{y}_{i1}-\mathbf{B}_y\,\Delta\,\mathbf{x}_{i1}=\mathbf{a}_y+\left(\Pi_y-\mathbf{B}_y\Pi_x\right)\boldsymbol{\xi}_{i0}+\boldsymbol{\upsilon}_{i1}, \tag{9.8}$$

and

$$\Delta\mathbf{x}_{i1}=\boldsymbol{\gamma}_x+\Pi_x\boldsymbol{\xi}_{i0}+\boldsymbol{\varepsilon}_{xi1}. \tag{9.9}$$

For $t = 2,3,\ldots,T$ we take first differences of (9.6) to obtain the conditional model for $\Delta^2\mathbf{y}_{it}$ given $\Delta^2\mathbf{x}_{it}$ and $\Delta\mathbf{w}_{i,t-1}$, and the marginal model for $\Delta^2\mathbf{x}_{it}$ given $\Delta\mathbf{w}_{i,t-1}$,

$$\Delta^2\mathbf{y}_{it} - \mathbf{B}_y\,\Delta^2\,\mathbf{x}_{it} = -\left(\mathbf{\Pi}_y - \mathbf{B}_y\mathbf{\Pi}_x\right)\boldsymbol{\gamma} + \left(\mathbf{\Pi}_y - \mathbf{B}_y\mathbf{\Pi}_x\right)\Delta\,\mathbf{w}_{i,t-1} + \Delta\boldsymbol{\upsilon}_{it}, \tag{9.10}$$

$$\Delta^2\mathbf{x}_{it} = -\mathbf{\Pi}_x\boldsymbol{\gamma} + \mathbf{\Pi}_x\,\Delta\,\mathbf{w}_{i,t-1} + \Delta\boldsymbol{\epsilon}_{xit}, \tag{9.11}$$

$t = 2,3,\ldots,T$, $i = 1,2,\ldots,N$.

From (9.10) it is clear that for $t = 2,3,\ldots,T$, $\Delta^2\mathbf{x}_{it}$ in (9.10) is long-run forcing if $\mathbf{\Pi}_x = \mathbf{0}_{m_x\times m}$.[31] Under this restriction, the information in (9.11) is redundant for purposes of inference regarding the parameters of interest in (9.10), namely the long-run coefficients in $\mathbf{\Pi}_y$ and the short-run coefficients in $\mathbf{B}_y$. As for $t = 1$, note from (9.8) and (9.9) that under $\mathbf{\Pi}_x = \mathbf{0}_{m_x\times m}$, $\Delta\mathbf{y}_{i1} - \mathbf{B}_y\,\Delta\,\mathbf{x}_{i1} - \mathbf{a}_y$ is independent of $\Delta\mathbf{x}_{i1} - \boldsymbol{\gamma}_x$, and we have that

$$E\left(\Delta\mathbf{y}_{i1} - \mathbf{B}_y\Delta\mathbf{x}_{i1} - \mathbf{a}_y|\Delta\mathbf{x}_{i1}\right) = \mathbf{0}_{m_y\times 1}, \quad \text{and} \quad Var\left(\Delta\mathbf{y}_{i1} - \mathbf{B}_y\Delta\mathbf{x}_{i1} - \mathbf{a}_y|\Delta\mathbf{x}_{i1}\right) = \mathbf{\Psi}_{yy},$$

where

$$\mathbf{\Psi}_{yy} = \boldsymbol{\alpha}_y\mathcal{Z}_y\boldsymbol{\alpha}_y' + \mathbf{\Omega}_{\upsilon}, \tag{9.12}$$

with

$$\mathcal{Z}_y = \boldsymbol{\beta}'\left(\sum_{j=0}^{\infty}\mathbf{\Phi}^j\mathbf{\Omega}\mathbf{\Phi}^{j\prime}\right)\boldsymbol{\beta}, \tag{9.13}$$

and $\mathbf{\Pi}_y = \boldsymbol{\alpha}_y\boldsymbol{\beta}'$, $\boldsymbol{\alpha}_y$ being of dimension $m_y\times r$, and $\boldsymbol{\beta}$ as before being an $m\times r$ matrix of full column rank, $r = 1,2,\ldots,m-1$. Note that the $r\,(r+1)/2$ distinct parameters in $\mathcal{Z}_y$ are unconstrained. In the special case of a PVAR(1) with unit roots but no cointegrating relations where $\mathbf{\Phi}_{yy} = \mathbf{I}_{m_y}$ and $\mathbf{\Phi}_{yx} = \mathbf{0}_{m_y\times m_x}$, $\mathcal{Z}_y = \mathbf{0}_{m_y}$, and $\mathbf{\Psi}_{yy}$ becomes equal to $\mathbf{\Omega}_{\upsilon}$. The marginal distribution of $\Delta\mathbf{x}_{i1}$ is redundant for inference on the coefficients in $\mathbf{\Pi}_y$ and $\mathbf{B}_y$.[32] We have thus established that under $\mathbf{\Pi}_x = \mathbf{0}_{m_x\times m}$ efficient inference on the coefficients in $\mathbf{\Pi}_y$ and $\mathbf{B}_y$ may be carried out by maximizing the conditional likelihood function associated with (9.8) and (9.10). Note that the panel long-run forcing restriction $\mathbf{\Pi}_x = \mathbf{0}_{m_x\times m}$ does not preclude feedbacks from $\mathbf{y}_{it}$ onto $\mathbf{x}_{it}$ in the short run. While in the set-up of (9.6) the panel long-run forcing restriction imposes the constraint that there

[31] See, for example, Granger and Lin (1995) and Pesaran, Shin, and Smith (2000) for a further discussion of the long-run forcing restriction in a time-series context.

[32] If the intercept term $\boldsymbol{\gamma}_y$ was also of interest, it could be consistently (though in general not efficiently) estimated even if it was not identified from (9.8) and (9.10) by first obtaining an estimate of $\boldsymbol{\gamma}_x$ from (9.9) and/or (9.11), and then retrieving an estimate of $\boldsymbol{\gamma}_y$ from

$$\boldsymbol{\gamma}_y = \mathbf{a}_y - \mathbf{B}_y\boldsymbol{\gamma}_x.$$

are no cointegrating relations among the I(1) forcing variables $\mathbf{x}_{it} - \boldsymbol{\mu}_{xi} - \boldsymbol{\gamma}_x t$, it is clear that similar arguments as given above could be advanced regarding conditional inference in the case where $\Delta \mathbf{x}_{it}$ is a trend-stationary process with homogeneous slope coefficients.[33]

To derive the likelihood function associated with (9.8) and (9.10), note that

$$Cov\left[\left(\Pi_y - \mathbf{B}_y \Pi_x\right) \boldsymbol{\xi}_{i0} + \boldsymbol{\upsilon}_{i1}, \Delta \boldsymbol{\upsilon}_{i2}\right] = -\Omega_{\upsilon},$$

and that

$$Cov\left[\left(\Pi_y - \mathbf{B}_y \Pi_x\right) \boldsymbol{\xi}_{i0} + \boldsymbol{\upsilon}_{i1}, \Delta \boldsymbol{\upsilon}_{it}\right] = \mathbf{0}_{m_y},$$

for $t = 3,4,\ldots,T$. The conditional likelihood function associated with (9.8) and (9.10) is given by

$$\prod_{i=1}^{N} (2\pi)^{-k/2} \left|\Sigma_{\upsilon}\right|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{w}_i^{*\prime} \Sigma_{\upsilon}^{-1} \mathbf{w}_i^{*}\right), \tag{9.14}$$

where

$$\mathbf{w}_i^{*} = \begin{pmatrix} \Delta \mathbf{y}_{i1} - \mathbf{B}_y \Delta \mathbf{x}_{i1} - \mathbf{a}_y \\ \Delta^2 \mathbf{y}_{i2} - \mathbf{B}_y \Delta^2 \mathbf{x}_{i2} - \Pi_y \Delta \mathbf{w}_{i1} - \mathbf{c}_y \\ \vdots \\ \Delta^2 \mathbf{y}_{iT} - \mathbf{B}_y \Delta^2 \mathbf{x}_{iT} - \Pi_y \Delta \mathbf{w}_{i,T-1} - \mathbf{c}_y \end{pmatrix}, \tag{9.15}$$

$\mathbf{c}_y = -\Pi_y \boldsymbol{\gamma}$, and

$$\Sigma_{\upsilon} = \begin{pmatrix} \Psi_{yy} & -\Omega_{\upsilon} & & & & \\ -\Omega_{\upsilon} & 2\Omega_{\upsilon} & -\Omega_{\upsilon} & & \mathbf{0} & \\ & \ddots & \ddots & \ddots & & \\ & \mathbf{0} & & -\Omega_{\upsilon} & 2\Omega_{\upsilon} & -\Omega_{\upsilon} \\ & & & & -\Omega_{\upsilon} & 2\Omega_{\upsilon} \end{pmatrix}. \tag{9.16}$$

The ML estimator of (9.14) is equivalent to finding $\mathbf{a}_y$, $\mathbf{c}_y$, Π_y, $\mathbf{B}_y$, $\mathcal{Z}_y$, and Ω_{υ} to maximize

$$l(\boldsymbol{\theta}) = -\frac{Nk}{2} \ln(2\pi) - \frac{N}{2} \ln\left|\Sigma_{\upsilon}\right| - \frac{1}{2} \sum_{i=1}^{N} \mathbf{w}_i^{*\prime} \Sigma_{\upsilon}^{-1} \mathbf{w}_i^{*}, \tag{9.17}$$

where $\boldsymbol{\theta} = \left(\ \mathbf{a}_y',\ \mathbf{c}_y',\ \boldsymbol{\pi}_y',\ \mathbf{b}_y',\ \boldsymbol{\jmath}_y',\ \boldsymbol{\sigma}_{\upsilon}'\ \right)'$, with $\boldsymbol{\pi}_y = vec(\Pi_y)$, $\mathbf{b}_y = vec(\mathbf{B}_y)$, $\boldsymbol{\jmath}_y = vech(\mathcal{Z}_y)$, and $\boldsymbol{\sigma}_{\upsilon} = vech(\Omega_{\upsilon})$.[34] The likelihood function (9.17) is well defined and depends on a fixed number of parameters.

[33] See Hsiao, Pesaran, and Tahmiscioglu (1999) for a discussion of this in the case where $m_y = m_x = 1$.

[34] In the case where $m_y \geq m_x$, $\boldsymbol{\gamma}_y$ and $\boldsymbol{\gamma}_x$ may be identified from (9.17), and $\boldsymbol{\theta}$ becomes

$$\boldsymbol{\theta} = \left(\ \boldsymbol{\gamma}',\ \boldsymbol{\pi}_y',\ \mathbf{b}_y',\ \boldsymbol{\jmath}_y',\ \boldsymbol{\sigma}_{\upsilon}'\ \right)'.$$

Tests of cointegration rank or other hypotheses of interest on elements of Φ_{yy} and Φ_{yx} can be carried out in rather similar fashion as described for unconditional PVARs in Section 8. Presuming that there are no cointegrating relations among $\mathbf{x}_{it} - \boldsymbol{\mu}_{xi} - \boldsymbol{\gamma}_x t$, the hypothesis that $\mathbf{w}_{it} - \boldsymbol{\mu}_i - \boldsymbol{\gamma} t$ is cointegrated with rank r versus rank $r+1$, $r = 0,1,\ldots,m-1$, can be formulated as

$$H_r : \Pi_y = \boldsymbol{\alpha}_{yr}\boldsymbol{\beta}_r' \qquad \text{vs.} \qquad H_{r+1} : \Pi_y = \boldsymbol{\alpha}_{y,r+1}\boldsymbol{\beta}_{r+1}'. \tag{9.18}$$

The likelihood ratio test statistic of H_r versus H_{r+1} is asymptotically chi-square distributed with $m_y + m - r$ degrees of freedom. (Imposing Π_y to be of rank r, there are $(m_y + m)r - r^2$ unrestricted coefficients in Π_y, and $r(r+1)/2$ unrestricted coefficients in $\mathcal{Z}_y$.) Identification of $\boldsymbol{\beta}_r$ may be achieved as discussed in Sections 7 and 8.

To test the long-run forcing restriction, one will, of course, need to consider the marginal likelihood function associated with (9.9) and (9.11).

10 Finite Sample Evidence

In this section we provide evidence on the finite sample properties of the GMM, the MD, and the ML estimators. This analysis is necessarily limited in scope and is intended as an illustration of the type of finite sample results that can be obtained rather than a comprehensive study. Nevertheless, we conjecture that our conclusions are likely to be of more general validity.

10.1 Monte Carlo Design

We consider four designs (experiments) in our Monte Carlo study. Common to all designs we set $m = 2$, $p = 1$, $\boldsymbol{\gamma} = \begin{pmatrix} 0.02, & 0.02 \end{pmatrix}'$, $N = (50,250)$, $T = (3,10)$, and use 1,000 replications. As a partial analysis of the robustness of the ML estimator to the normality assumption, we consider two schemes for generating the error terms:

$$\boldsymbol{\varepsilon}_{it} \overset{iid}{\sim} N(\mathbf{0}_{m\times 1}, \Omega), \tag{10.1}$$

and

$$\boldsymbol{\varepsilon}_{it} = \frac{1}{2}\mathbf{P}' \begin{pmatrix} \varsigma_{1it,1}^2 + \varsigma_{2it,1}^2 - 2 \\ \varsigma_{1it,2}^2 + \varsigma_{2it,2}^2 - 2 \\ \\ \varsigma_{1it,m}^2 + \varsigma_{2it,m}^2 - 2 \end{pmatrix}, \tag{10.2}$$

where $\mathbf{P}$ is the (upper triangular) Cholesky factor of Ω, $\varsigma_{lit,j} \overset{iid}{\sim} N(0,1)$, $l = 1,2$, $j = 1,2,\ldots,m$, with $\varsigma_{lit,j}$ being independent for all l and j. For all experiments the fixed effects are generated as:

$$\boldsymbol{\mu}_i = \mathbf{C}\boldsymbol{\varepsilon}_{i0} + \mathbf{t}_i, \qquad i = 1,2,\ldots,N, \tag{10.3}$$

where

$$\begin{aligned} \mathbf{t}_i &\overset{iid}{\sim} N(\upsilon_{1t}, \Sigma_{it}), \qquad i = 1,2,\ldots,\frac{N}{2}, \\ \mathbf{t}_i &\overset{iid}{\sim} N(\upsilon_{2t}, \Sigma_{it}), \qquad i = \frac{N}{2}+1, \frac{N}{2}+2,\ldots,N, \end{aligned} \tag{10.4}$$

with $\upsilon_{1t} = \left(0.25, \quad 0.25 \right)'$, $\upsilon_{2t} = \left(0.5, \quad 0.5 \right)'$, the two diagonal elements of the Cholesky factor $\mathbf{P}_{it}$ of Σ_{it} being generated as chi-square variates with one degree of freedom, and the upper triangular element of $\mathbf{P}_{it}$ being generated as minus one plus a chi-square variate with one degree of freedom, all chi-square variates being independent for each given i, as well as independent across i, $i = 1,2,\ldots,N$.

The four designs distinguish between trend-stationary, pure unit root, and cointegrated PVAR models. For all designs the $\mathbf{w}_{it}$-process is generated using (2.26), assuming that ζ_i and the ϵ_{it}'s follow the same distribution function.

To give an indication of the degree of fit of the various designs, below we also report (where appropriate) the population R^2 values associated with each design.

Design 1: Trend-stationary first-order PVAR with maximum eigenvalue of Φ *equal to* 0.6

$$\Phi = \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{pmatrix}.$$

The other eigenvalue of Φ is 0.2, and the population R^2 values are given by $R^2_{\Delta w_{1it}} = R^2_{\Delta w_{2it}} =$ 0.2471, where

$$\begin{aligned} R^2_{\Delta w_{lit}} &= 1 - \frac{Var(\Delta w_{lit} | \Delta \mathbf{w}_{i,t-1})}{Var(\Delta w_{lit})} \\ &= 1 - \frac{[\Omega]_{ll}}{\left[\sum_{j=0}^{\infty} \Phi^j \Omega (\Phi^j)'\right]_{ll}}, \qquad l = 1,2, \end{aligned} \tag{10.5}$$

with $[\mathbb{S}]_{ll}$ denoting the element in the l-th row and l-th column of $\mathbf{S}$.

Design 2: Trend-stationary first-order PVAR with maximum eigenvalue of Φ *equal to* 0.8

$$\Phi = \begin{pmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0.1 & -0.08 \\ -0.08 & 0.1 \end{pmatrix}.$$

The other eigenvalue of Φ is 0.4, and the population R^2 values are given by $R^2_{\Delta w_{1it}} = R^2_{\Delta w_{2it}} =$ 0.2588, where $R^2_{\Delta w_{lit}}$, $l = 1,2$, are computed as in (10.5).

Design 3: First-order PVAR with unit roots (but non-cointegrated)

$$\Phi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{pmatrix}.$$

Design 4: Cointegrated first-order PVAR

$$\Phi = \begin{pmatrix} 0.4 & 0.6 \\ -0.2 & 1.2 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0.06 & 0.02 \\ 0.02 & 0.01 \end{pmatrix}.$$

The eigenvalues of Φ in this case are given by 1 and 0.6, and the implied vectors/matrices α, β, and Π are given by

$$\alpha = \begin{pmatrix} -0.6 \\ -0.2 \end{pmatrix}, \qquad \beta = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \Pi = \begin{pmatrix} -0.6 & 0.6 \\ -0.2 & 0.2 \end{pmatrix}.$$

The population R^2 values are given by $R^2_{\Delta y_{1it}} = 0.2195$ and $R^2_{\Delta y_{2it}} = 0.1579$, where

$$\begin{aligned} R^2_{\Delta y_{lit}} &= 1 - \frac{Var\left(\Delta y_{lit} | \mathbf{w}_{i,t-1}\right)}{Var\left(\Delta y_{lit}\right)} \\ &= 1 - \frac{[\Omega]_{ll}}{\left[\sum_{j=0}^{\infty} \mathbf{C}_j \Omega \mathbf{C}_j'\right]_{ll}}, \qquad l = 1,2, \end{aligned} \tag{10.6}$$

with $\mathbf{C}_0 = \mathbf{I}_m$, $\mathbf{C}_1 = -(\mathbf{I}_m - \Phi)$, and $\mathbf{C}_j = \mathbf{C}_{j-1}\Phi$, $j = 2,3,\ldots$.[35]

In what follows we compare the conventional GMM, the MD, and the ML estimators of the various parameters in terms of their biases and root mean square errors (RMSEs). The finite sample performance of a number of tests based on the conventional GMM and the ML estimators is also investigated,[36] as is the possibility of employing extended GMM estimation. For Designs 1-3 the various estimators are computed with Π unrestricted, and for Design 4 the MD and ML estimators are computed both with and without imposing rank restrictions on Π.

The various estimators and their variance-covariance matrices are computed as follows: The conventional GMM estimator not incorporating the initial conditions (in the tables denoted by '*GMM$_a$*') is obtained by minimizing (4.15) via feasible Generalized Least Squares iterations employing factorizations similar to the ones discussed in Section 6.1, with Ω in (4.16) replaced by

$$\widehat{\Omega} = \frac{1}{2N(T-1)} \sum_{t=2}^{T} \sum_{i=1}^{N} \left[\Delta \mathbf{w}_{it} - \widehat{\gamma} - \widehat{\Phi}\left(\Delta \mathbf{w}_{i,t-1} - \widehat{\gamma}\right)\right] \left[\Delta \mathbf{w}_{it} - \widehat{\gamma} - \widehat{\Phi}\left(\Delta \mathbf{w}_{i,t-1} - \widehat{\gamma}\right)\right]'. \tag{10.7}$$

The variance-covariance matrix of the resulting conventional GMM estimator of Φ is computed using

$$\left\{ \left(\sum_{i=1}^{N} \mathbf{G}_i \overline{\mathbf{Q}}_i' \otimes \mathbf{I}_m \right) \left[\sum_{i=1}^{N} \left(\overline{\mathbf{Q}}_i \otimes \mathbf{I}_m \right) \Sigma \left(\overline{\mathbf{Q}}_i' \otimes \mathbf{I}_m \right) \right]^{-1} \left(\sum_{i=1}^{N} \overline{\mathbf{Q}}_i \mathbf{G}_i' \otimes \mathbf{I}_m \right) \right\}^{-1}, \tag{10.8}$$

[35] See Pesaran, Shin, and Smith (2000) for a discussion of the computation of R^2 values for cointegrated VARs.

[36] As we had discussed in Section 6, derivation of the asymptotic variance-covariance matrix of the MD estimator is not a trivial task, and is not pursued in this paper. We thus confine ourselves to reporting bias and RMSE for the MD estimator.

where $\overline{\mathbf{Q}}_i = \left(0_{(mT/2+1)(T-1)\times 1} \quad \mathbf{Q}_i \right)$, $\mathbf{Q}_i$ is given by (4.12), and replacing γ by $\widehat{\gamma}$, and Σ by $\widehat{\Sigma}$. The conventional GMM estimator incorporating the initial conditions (in the tables denoted by 'GMM_b') is computed by minimizing (4.30). The extended GMM estimator is obtained by solving (4.57). Initial estimates of γ, Φ, and Ω needed to initialize the computation of the various GMM estimators are obtained using (6.8), (6.10), and (6.11), respectively.[37] The MD estimator is computed using Steps 1-4 as described in Section 6.1. (Under rank restrictions on Π, we use (7.6) and (7.8) to estimate Φ.) The ML estimator is computed by maximizing the log-likelihood function (5.5). As initial estimates for the ML algorithm we use the MD estimates. The variance-covariance matrix of the ML estimator of Φ is computed using the Hessian matrix evaluated at the estimates in the final iteration. The numerical optimization routine used for computation of the conventional GMM estimator incorporating the initial conditions, of the extended GMM estimator, and of the ML estimator employs a trust region method based algorithm, and is described in some detail in a note available from the authors upon request.

10.2 Monte Carlo Evidence

We begin by discussing the Monte Carlo evidence under normality of the error terms, as summarized in Tables 1-8. Tables 1 and 2 present the biases and RMSEs of the conventional GMM estimator (not incorporating the initial conditions), and the MD and ML estimators of γ and Φ for different values of T and N under Designs 1-3.[38] Table 2 demonstrates that in the trend-stationary case with maximum eigenvalue of Φ equal to 0.6 (Design 1), the biases and RMSEs of the conventional GMM, MD, and ML estimators all behave reasonably well for relatively large N and moderately sized T (namely, $N = 250$ and $T = 10$). But even in this case the MD and ML estimators dominate the conventional GMM estimator, both in terms of bias and RMSE. As to be expected, the differences between the conventional GMM estimator and the ML and MD estimators in the trend-stationary cases are most pronounced for smaller values of N and/or T. For $N = 50$ and $T = 3$, for example, even when the maximum eigenvalue of Φ is as small as 0.6, the conventional GMM estimator does poorly, with its performance deteriorating further as the magnitude of the maximum eigenvalue of Φ increases (0.8 for Design 2). For Design 1 the performance of the conventional GMM estimator

[37] For the extended GMM estimator, different initial estimates were also explored. See the discussion below.

[38] In constructing Tables 1-6 we have removed across all the estimation procedures for the various cases considered those (very few) replications where the ML estimation routine, due to poor initial conditions (the MD estimates) did not converge globally, that is, where the likelihood value at the final iteration ML estimates was lower in value than for a set of coefficients in the neighborhood of the true parameter values. In other words, the total number of replications for each case was such that after eliminating those replications where the ML estimation routine did not converge globally, there were 1,000 replications left. The following experiments for this reason involved more than 1,000 total replications: Design 2, $N = 50$, $T = 3$: 1,004 replications; and Design 2, $N = 250$, $T = 3$: 1,001 replications.

improves relatively quickly with N and/or T, while the magnitudes of the biases and RMSEs continue to be relatively large for Design 2 when $N = 250$ and $T = 3$, and when $N = 50$ and $T = 10$. Table 2 also demonstrates that the conventional GMM estimator breaks down in the presence of a unit root; for the coefficients ϕ_{11} and ϕ_{22}, both equal to unity under Design 3, the biases and RMSEs continue to be large even when $N = 250$ and $T = 10$. The ML and MD estimators, in contrast, perform quite well in the trend-stationary as well as in the unit root case, with a slightly superior performance of the ML estimator.

Tables 3-6 report rejection frequencies for tests involving the parameter values ϕ_{jk}, where ϕ_{jk} denotes the element in the j-th row and k-th column of Φ. All tests have the nominal size of 5%. Table 3 demonstrates that the power and size properties of tests of ϕ_{11} based on the conventional GMM estimator (not incorporating the initial conditions) are very poor, except for the case where the maximum eigenvalue of Φ is equal to 0.6, $N = 250$, and $T = 10$. When the maximum eigenvalue of Φ is equal to 0.8, tests based on the conventional GMM estimator are highly over-sized. For example, even for $N = 250$ and $T = 10$, the size of the test based on the conventional GMM estimator is equal to 45.2%. Also, as expected, tests based on conventional GMM estimators break down altogether in the presence of unit roots. (See the third panel of Table 3.) In contrast the ML estimator based tests perform remarkably well even for small N and T and irrespective of how close the eigenvalues of Φ are to unity.

Tables 4-6 present test results applied to the remaining three coefficients of Φ, namely ϕ_{12}, ϕ_{21}, and ϕ_{22}. The conventional GMM based tests do relatively better for these coefficients (with further variations across these coefficients) as compared to the tests involving ϕ_{11}. However, the overall performance of the GMM based tests is still rather poor and the tests cannot be recommended for N and/or T relatively small and when the maximum eigenvalue of Φ is relatively large. Overall, the tests based on the conventional GMM estimator are all dominated by the ML estimator based tests both in terms of size and power.[39] The ML estimator based tests display (very) good size properties for all designs and coefficients, even when N is as small as 50, and T is as small as 3. It should also be noted, though, that specifically for the trend-stationary designs the power properties of the ML estimator based tests may not be very strong when $N = 50$ and $T = 3$. However, these power properties improve quickly when N and/or T is increased: when $N = 50$ and $T = 10$ and when $N = 250$ and $T = 3$, the power properties are very good outside the immediate neighborhood of the true parameter values. (There is one exception to this, namely tests involving ϕ_{12} under Design 2.)

[39]Power comparisons of tests are only meaningful when the tests under consideration have empirical sizes that are close to their nominal size. Therefore, the power of the conventional GMM and ML estimator based tests can be meaningfully compared in the case of experiments where N is relatively large, T is moderately sized, and the maximum eigenvalue of Φ is relatively small. See, for example, the results in the first panels of Tables 4 and 5 under $N = 250$ and $T = 10$.

Table 7 reports evidence on bias, RMSEs, and power and size of hypothesis tests under Design 4, which involves one cointegrating relation.[40] The results reported in Tables 7a-7d are based on ML (and MD) estimates computed under the rank restriction $rank\,(\Pi) = 1$. The results in these tables suggest that as under Designs 1-3, the differences between the ML and MD estimators are small with a slightly superior performance of the ML estimator. When N is relatively large (say, $N = 250$) and when there are a fair number of time-series observations available (say, $T = 10$), both the ML and the MD estimator perform well in terms of bias and RMSE, and the size and power properties of the ML estimator based parameter tests are very good. Table 7e reports the likelihood ratio tests (8.9) for cointegration rank.[41] The results are uniformly very good; possibly with the exception of the test in the case of $N = 50$ and $T = 3$ which is slightly undersized.

Table 8 provides evidence on the relative performance of the conventional GMM estimator incorporating the initial conditions compared with its counterpart not incorporating initial conditions.[42] Using Design 1, Table 8 reveals that while the conventional GMM estimator incorporating the initial conditions has, as to be expected, lower biases and RMSEs for the coefficients in γ than its counterpart not incorporating the initial conditions, the biases and RMSEs of the two estimators are rather similar for the coefficients in Φ. Also, comparing Table 8a to Tables 1 and 2, the biases and RMSEs of the coefficients in Φ of the conventional GMM estimator not incorporating the initial conditions are little affected by whether 'outlier' replications need to be eliminated from the sample. This underlines the robustness of the conventional GMM estimator results reported in Tables 1 and 2. Tables 8b shows that the size properties of tests involving ϕ_{11} based on the conventional GMM estimator incorporating the initial conditions are (quite) a bit worse than those based on the conventional GMM estimator not incorporating the initial conditions when $N = 250$ and $T = 10$.[43] Overall, Table 8 suggests that little is to be gained from incorporating the initial

[40]In constructing Table 7 we have removed those replications where at least one of the ML estimation routines (under $rank\,(\Pi)$ being constrained to be equal to zero or one, or left unconstrained) did not converge globally. The primary reason for such non-convergence was that one of the eigenvalues of the implied $\widehat{\Phi}$ under $rank\,(\Pi) = 1$ during the iterations started to fall outside the unit circle. Our simulations therefore involved the following number of replications: $N = 50$, $T = 3$: 1,274 replications; $N = 50$, $T = 10$: 1,013 replications; $N = 250$, $T = 3$: 1,035 replications; and $N = 250$, $T = 10$: 1,011 replications. A potential remedy to the convergence problems would be to employ constrained versions of the numerical algorithm we use to compute the estimates that enforce the restriction that none of the eigenvalues of Φ fall outside the unit circle. However, this would likely be at the cost of decreased speed of computations when there is convergence even without imposing constraints on the eigenvalues of Φ.

[41]We are not aware of any GMM based tests for cointegration rank in dynamic panel data models where the time dimension of the panel is finite, and thus confine ourselves to the likelihood ratio tests (8.9).

[42]In constructing Table 8, we have removed across all the estimation procedures those replications where the optimization routine for the conventional GMM estimator incorporating the initial conditions did not converge globally, that is, where the objective function value at the final iteration estimates was higher in value than for a set of coefficients in the neighborhood of the true parameter values. The following cases for this reason involved more than 1,000 total replications: $N = 50$, $T = 3$: 1,050 replications; and $N = 250$, $T = 3$: 1,013 replications.

[43]Since the results for tests involving ϕ_{12}, ϕ_{21}, and ϕ_{22} are qualitatively similar to those involving ϕ_{11} in Table

conditions in conventional GMM estimation, unless estimation of the trend coefficients is critical. In fact, taking into account the size properties of tests, and taking into account that the conventional GMM estimator incorporating the initial conditions is computationally significantly more involved than its counterpart not incorporating the initial conditions, the latter estimator seems preferable, if conventional GMM estimation is to be carried out at all. The Monte Carlo simulations we have carried out also suggest that the conclusions we have drawn from Table 8 carry over to Designs 2 and 3 as well.

Table 9 provides some evidence on the properties of the conventional GMM estimator (not incorporating the initial conditions), and the MD and ML estimators when the error terms are drawn from (10.2) instead of the normal distribution, (10.1). Table 9 indicates that for Design 1 both estimators continue to perform well in small sample when the error terms are not normally distributed, specifically when they are drawn from (10.2). The ML estimator continues to slightly outperform the MD estimator in terms of bias and RMSE. Tables 9b-9e reveal that the power and size properties of tests based on the ML estimator remain largely unaffected by the change in the scheme generating the error terms, except for a couple of cases where the tests become slightly oversized. The Monte Carlo simulations we have carried out also suggest that the conclusions we have drawn from Table 9 carry over to Designs 2-4 as well.

We have also used Monte Carlo simulations to further investigate the possibility of employing extended GMM estimation. As discussed in Section 4, the extended GMM estimator (4.57) involving the homoskedasticity implied moment conditions (4.50) and (4.51) is not unique. The Monte Carlo simulations that we carried out illustrated the difficulties involved in choosing among the multiple solutions to the moment conditions (4.52) that exist in general. The final iteration extended GMM estimates that we obtained under the different designs were sensitive to the initial estimates employed. Specifically, we obtained different extended GMM estimates when the initial estimates were obtained using (6.8), (6.10), and (6.11), than when the MD estimates were employed as initial estimates. The Monte Carlo simulations also revealed a further difficulty with extended GMM estimation using the homoskedasticity implied moment conditions (4.50) and (4.51): The variance-covariance matrix $\mathbf{W}_N(\kappa)$, given in (4.59), tends to have a small reciprocal condition number even for m and T as small as $m = 2$ and $T = 3$, under both $N = 50$ and $N = 250$. The values of $\mathbf{A}_N(\kappa)$ computed in the course of the iterations are thus subject to potentially large numerical approximation errors. In fact, for designs with $m = 4$ and $T = 10$, attempts to compute $\mathbf{A}_N(\kappa)$ broke down altogether, even when methods for the inversion of ill-conditioned matrices were employed. These difficulties were yet more accentuated for the extended GMM estimator augmenting the standard orthogonality conditions with the homoskedasticity and stationarity implied moment

8b, we have omitted the tables for size and power properties of tests involving ϕ_{12}, ϕ_{21}, and ϕ_{22}. These tables are available from the authors upon request.

conditions (4.36) and (4.37) rather than the homoskedasticity implied moment conditions (4.50) and (4.51). For all designs and for all combinations of N and T considered the reciprocal condition number of $\mathbf{W}_N(\boldsymbol{\kappa})$ was too small to be able to compute $\mathbf{A}_N(\boldsymbol{\kappa})$ with any reliable degree of numerical accuracy. This was true irrespective of whether the Σ_{it} matrices were heteroskedastic across i (as implied by (10.4)), or were restricted to be homoskedastic across i. The various difficulties just described with implementing extended GMM estimation appear to further strengthen the appeal of the ML estimator.

11 Conclusion

Using panel data to estimate time-series models has the virtue that one can often invoke conventional central limit theorems across individuals irrespective of the time-series properties of the variables under consideration, in particular irrespective of whether they are (trend) stationary, pure I(1), or I(1) and cointegrated. However, the virtue of having a large number of cross-sectional observations available often appears to be offset by the need to introduce individual-specific effects to capture the unobserved heterogeneity among the cross-sectional units. Fixed effects and, when the time dimension of the panel is short, initial observations introduce the classical incidental parameters problem into conventional likelihood based estimation and inference, which lead to a violation of the regularity conditions for the consistency of the conventional ML estimator.

To overcome these problems, GMM estimation has been suggested in the literature. However, conventional GMM estimators based on instruments that are orthogonal to the disturbances of the first-differenced form of the model are not efficient (for example, Ahn and Schmidt, 1995, Arellano and Bover, 1995, and Hahn, 1999) and suffer from the defect of weak instruments if one or more roots of the underlying PVAR model approach unity (for example, Blundell and Bond, 1998, and Hsiao, Pesaran and Tahmiscioglu, 1999). As we have shown in this paper, conventional GMM estimators of fixed effects PVARs break down completely if the underlying model contains unit roots, and thus also cannot be used to construct cointegration rank tests. To obtain a consistent GMM estimator when the data series contain unit roots the standard orthogonality conditions have to be augmented by further moment conditions. In this paper, we have argued that the approach of augmenting the standard orthogonality conditions by homoskedasticity and stationarity implied moment conditions for fixed effects PVARs is subject to difficulties. In particular, it may not be possible to derive the optimal weighting matrix for the resultant extended GMM estimators, or the resultant extended GMM estimators are not unique, in general.

Given the problems with GMM estimation, we have suggested a latent variable framework for the analysis of fixed effects PVARs. This framework provides a model-consistent way of formulating the initial observations and leads to a transformed likelihood function that no longer involves

incidental parameters. We have shown that the resulting ML estimator exists, is consistent, and is asymptotically normally distributed when the cross-sectional dimension of the panel approaches infinity, irrespective of whether the underlying time series are (trend) stationary, pure I(1), or I(1) and cointegrated. We have also shown how this ML estimator can be used in conditional models where a subset of the variables in the PVAR is modelled in terms of the remaining variables. Furthermore, we have proposed procedures for conducting tests for unit roots and cointegration rank in panels with short time dimension. We have shown that the limiting distributions of the test statistics follow standard chi-square and normal distributions.

Finally, Monte Carlo studies have been carried out to compare the small sample performance of the ML and conventional and extended GMM estimators. We have found that the conventional GMM estimator performs poorly even when the largest root is moderately sized only. Furthermore, implementation of extended GMM estimation faces difficult computational problems. The ML estimator and parameter hypothesis as well as cointegration tests based on it, on the other hand, perform remarkably well even when the sample size is small and the data are generated by (some type of) non-normal disturbances.

Appendix: Conventional GMM, ML, and MD Estimation of a p-th Order PVAR ($p > 1$)

In this appendix we extend the formulae provided in Sections 4 to 6 for the conventional GMM, ML, and MD estimation of a first-order PVAR to a p-th order PVAR, $p > 1$. We continue to invoke the assumptions for the PVAR(p) model (2.1) put forward in Section 2.

The transformed model (3.2), (3.3), and (3.9) can be written as

$$\Delta \mathbf{w}_{it} - \boldsymbol{\gamma} = \mathbf{C}(L)\,\boldsymbol{\varepsilon}_{it}, \qquad t = 1,2,\ldots,p-1, \tag{A.1}$$

$$\Delta \mathbf{w}_{ip} - \boldsymbol{\gamma} = \sum_{j=1}^{p-1} \boldsymbol{\Gamma}_j \left(\Delta \mathbf{w}_{i,p-j} - \boldsymbol{\gamma}\right) + \boldsymbol{\Pi}\mathbf{C}^*(L)\,\boldsymbol{\varepsilon}_{i0} + \boldsymbol{\varepsilon}_{ip}, \tag{A.2}$$

$$\Delta \mathbf{w}_{it} - \boldsymbol{\gamma} = \boldsymbol{\Phi}\left[\Delta \mathbf{z}_{i,t-1} - (\boldsymbol{\iota}_p \otimes \boldsymbol{\gamma})\right] + \Delta \boldsymbol{\varepsilon}_{it}, \qquad t = p+1, p+2, \ldots, T, \tag{A.3}$$

where

$$\mathbf{z}_{it} = \left(\mathbf{w}_{it}', \quad \mathbf{w}_{i,t-1}', \quad \ldots, \quad \mathbf{w}_{i,t-p+1}' \right)', \tag{A.4}$$

$$\boldsymbol{\Phi} = \left(\boldsymbol{\Phi}_1, \quad \boldsymbol{\Phi}_2, \quad \ldots, \quad \boldsymbol{\Phi}_p \right), \tag{A.5}$$

$\boldsymbol{\Gamma}_j$, $j = 1,2,\ldots,p-1$, is given by (2.11), $\boldsymbol{\Pi}$ by (2.5), $\boldsymbol{\iota}_p$ denotes a $p \times 1$ vector of ones, and, as defined in Section 2, $\mathbf{C}(L) = \sum_{j=0}^{\infty} \mathbf{C}_j L^j$, $\mathbf{C}_0 = \mathbf{I}_m$, $\mathbf{C}_1 = -(\mathbf{I}_m - \boldsymbol{\Phi}_1)$, $\mathbf{C}_j = \sum_{l=1}^{p} \mathbf{C}_{j-l}\boldsymbol{\Phi}_l$, $j \geq 2$ (with $\mathbf{C}_j = \mathbf{0}_m$ for $j < 0$), $\mathbf{C}^*(L) = \sum_{j=0}^{\infty} \mathbf{C}_j^* L^j$, $\mathbf{C}_0^* = \mathbf{I}_m - \mathbf{C}$, $\mathbf{C}_j^* = \mathbf{C}_j + \mathbf{C}_{j-1}^*$, $j \geq 1$, and $\mathbf{C} = \sum_{j=0}^{\infty} \mathbf{C}_j$.

Conventional GMM Estimation

The first-differenced PVAR(p) model (A.3) can in stacked form be written as

$$\overline{\mathbf{S}}_i = \overline{\mathbf{R}}_i \overline{\boldsymbol{\Lambda}} + \overline{\mathbf{E}}_i, \tag{A.6}$$

$i = 1,2,\ldots,N$, where

$$\overline{\mathbf{S}}_i = \left(\Delta \mathbf{w}_{i,p+1}, \quad \Delta \mathbf{w}_{i,p+2}, \quad \ldots, \quad \Delta \mathbf{w}_{iT} \right)', \quad \overline{\mathbf{E}}_i = \left(\Delta \boldsymbol{\varepsilon}_{i,p+1}, \quad \Delta \boldsymbol{\varepsilon}_{i,p+2}, \quad \ldots, \quad \Delta \boldsymbol{\varepsilon}_{iT} \right)', \tag{A.7}$$

$$\overline{\mathbf{R}}_i = \left(\overline{\mathbf{P}}_i, \quad \boldsymbol{\iota}_{T-1} \right), \qquad \overline{\mathbf{P}}_i = \left(\Delta \mathbf{z}_{ip}, \quad \Delta \mathbf{z}_{i,p+1}, \quad \ldots, \quad \Delta \mathbf{z}_{i,T-1} \right)', \tag{A.8}$$

and

$$\overline{\Lambda} = \left(\Phi, \quad \gamma - \Phi \left(\iota_p \otimes \gamma \right) \right)', \tag{A.9}$$

with Φ given by (A.5).

Premultiplying (A.6) by the $[m(T+p-1)/2+1](T-p)\times(T-p)$ block-diagonal instrumental variable matrix $\overline{\mathbf{Q}}_i$,

$$\overline{\mathbf{Q}}_i = \begin{pmatrix} \overline{\mathbf{q}}_{i,p+1} & \mathbf{0}_{(mp+1)\times 1} & \mathbf{0}_{(mp+1)\times 1} & & \mathbf{0} \\ \mathbf{0}_{[m(p+1)+1]\times 1} & \overline{\mathbf{q}}_{i,p+2} & \mathbf{0}_{[m(p+1)+1]\times 1} & & \\ \mathbf{0}_{[m(p+2)+1]\times 1} & \mathbf{0}_{[m(p+2)+1]\times 1} & \overline{\mathbf{q}}_{i,p+3} & & \\ & & & \ddots & \\ & \mathbf{0} & & & \overline{\mathbf{q}}_{iT} \end{pmatrix}, \tag{A.10}$$

where $\overline{\mathbf{q}}_{i,p+1} = \left(1, \quad \mathbf{z}'_{i,p-1} \right)'$, and $\overline{\mathbf{q}}_{it} = \left(1, \quad \mathbf{w}'_{i,t-2}, \quad \mathbf{w}'_{i,t-3}, \quad \ldots, \quad \mathbf{w}'_{ip}, \quad \mathbf{z}'_{i,p-1} \right)'$, $t = p+2, p+3, \ldots, T$, one obtains

$$\overline{\mathbf{Q}}_i \overline{\mathbf{S}}_i = \overline{\mathbf{Q}}_i \overline{\mathbf{R}}_i \overline{\Lambda} + \overline{\mathbf{Q}}_i \overline{\mathbf{E}}_i, \tag{A.11}$$

the transpose of which in vectorized form becomes

$$\left(\overline{\mathbf{Q}}_i \otimes \mathbf{I}_m \right) vec\left(\overline{\mathbf{S}}'_i \right) = \left(\overline{\mathbf{Q}}_i \overline{\mathbf{R}}_i \otimes \mathbf{I}_m \right) \overline{\boldsymbol{\lambda}} + \left(\overline{\mathbf{Q}}_i \otimes \mathbf{I}_m \right) vec\left(\overline{\mathbf{E}}'_i \right), \tag{A.12}$$

where $\overline{\boldsymbol{\lambda}} = vec\left(\overline{\Lambda}' \right)$. The conventional GMM estimator of $\overline{\boldsymbol{\lambda}}$ (not incorporating initial conditions) is obtained by minimizing

$$\left[\sum_{i=1}^{N} \left(\overline{\mathbf{Q}}_i \otimes \mathbf{I}_m \right) vec\left(\overline{\mathbf{S}}'_i \right) - \left(\overline{\mathbf{Q}}_i \overline{\mathbf{R}}_i \otimes \mathbf{I}_m \right) \overline{\boldsymbol{\lambda}} \right]' \left[\frac{1}{N} \sum_{i=1}^{N} \left(\overline{\mathbf{Q}}_i \otimes \mathbf{I}_m \right) \overline{\Sigma} \left(\overline{\mathbf{Q}}_i \otimes \mathbf{I}_m \right)' \right]^{-1}$$
$$\cdot \left[\sum_{i=1}^{N} \left(\overline{\mathbf{Q}}_i \otimes \mathbf{I}_m \right) vec\left(\overline{\mathbf{S}}'_i \right) - \left(\overline{\mathbf{Q}}_i \overline{\mathbf{R}}_i \otimes \mathbf{I}_m \right) \overline{\boldsymbol{\lambda}} \right], \tag{A.13}$$

where $\overline{\Sigma}$ is of the same form as (4.16),

$$\overline{\Sigma} = \begin{pmatrix} 2\Omega & -\Omega & & & & \\ -\Omega & 2\Omega & -\Omega & & \mathbf{0} & \\ & \ddots & \ddots & \ddots & & \\ & \mathbf{0} & & -\Omega & 2\Omega & -\Omega \\ & & & & -\Omega & 2\Omega \end{pmatrix}. \tag{A.14}$$

As for the case where $p = 1$, the conventional GMM moment conditions underlying (A.13) may be augmented with additional moment conditions based on (A.1) and (A.2) to construct a conventional GMM estimator incorporating initial conditions. We do not go into the details here.

A2

ML Estimation

The unconditional log-likelihood function associated with (A.1), (A.2), and (A.3) is given by[44]

$$\begin{aligned} l(\boldsymbol{\theta}) &= -\frac{Nk}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma_{\epsilon}| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{r}_{\epsilon i} - \mathbf{H}_{\epsilon i}\phi)'\,\Sigma_{\epsilon}^{-1}\,(\mathbf{r}_{\epsilon i} - \mathbf{H}_{\epsilon i}\phi) \\ &\quad -\frac{N}{2}\ln|\Sigma_{f|\epsilon}| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{r}_{fi} - \mathbf{H}_{fi}\phi)'\,\Sigma_{f|\epsilon}^{-1}\,(\mathbf{r}_{fi} - \mathbf{H}_{fi}\phi)\,, \end{aligned} \tag{A.16}$$

where $k = mpT$,

$$\mathbf{r}_{\epsilon i} = vec\Big[\ \triangle\mathbf{w}_{i1} - \boldsymbol{\gamma},\ \ \triangle\mathbf{w}_{i2} - \boldsymbol{\gamma},\ \ \ldots,\ \ \triangle\mathbf{w}_{i,p-1} - \boldsymbol{\gamma},\ \ \textstyle\sum_{j=0}^{p-1}(\triangle\mathbf{w}_{i,p-j} - \boldsymbol{\gamma})\ \Big], \tag{A.17}$$

$$\mathbf{H}_{\epsilon i} = \mathbf{G}_{\epsilon i}' \otimes \mathbf{I}_m, \tag{A.18}$$

$$\mathbf{G}_{\epsilon i} = \Big(\ \mathbf{0}_{mp\times(p-1)},\ \ \boldsymbol{\tau}_{\epsilon i}\ \Big), \tag{A.19}$$

$$\boldsymbol{\tau}_{\epsilon i} = vec\Big[\ \textstyle\sum_{j=1}^{p-1}(\triangle\mathbf{w}_{i,p-j} - \boldsymbol{\gamma}),\ \ \sum_{j=2}^{p-1}(\triangle\mathbf{w}_{i,p-j} - \boldsymbol{\gamma}),\ \ \ldots,\ \ \triangle\mathbf{w}_{i1} - \boldsymbol{\gamma},\ \ \mathbf{0}_{m\times 1}\ \Big], \tag{A.20}$$

$$\mathbf{r}_{fi} = vec\Big(\ \boldsymbol{\tau}_{fi},\ \ \triangle\mathbf{w}_{i,p+2} - \boldsymbol{\gamma},\ \ \triangle\mathbf{w}_{i,p+3} - \boldsymbol{\gamma},\ \ \ldots,\ \ \triangle\mathbf{w}_{iT} - \boldsymbol{\gamma}\ \Big), \tag{A.21}$$

$$\boldsymbol{\tau}_{fi} = \triangle\mathbf{w}_{i,p+1} - \boldsymbol{\gamma} + \Omega\mathcal{R}\mathbf{r}_{\epsilon i}, \tag{A.22}$$

$$\mathbf{H}_{fi} = \mathbf{G}_{f1,i}' \otimes \mathbf{I}_m + \mathbf{G}_{f2,i}' \otimes \Omega\mathcal{S}, \tag{A.23}$$

$$\mathbf{G}_{f1,i} = \Big[\ \triangle\mathbf{z}_{ip} - (\boldsymbol{\iota}_p \otimes \boldsymbol{\gamma}),\ \ \triangle\mathbf{z}_{i,p+1} - (\boldsymbol{\iota}_p \otimes \boldsymbol{\gamma}),\ \ \ldots,\ \ \triangle\mathbf{z}_{i,T-1} - (\boldsymbol{\iota}_p \otimes \boldsymbol{\gamma})\ \Big], \tag{A.24}$$

$$\mathbf{G}_{f2,i} = \Big(\ \boldsymbol{\tau}_{\epsilon i},\ \ \mathbf{0}_{mp\times(T-p-1)}\ \Big), \tag{A.25}$$

[44] To derive (A.16), we have rewritten (A.2) using (2.11) as

$$\sum_{j=0}^{p-1}(\triangle\mathbf{w}_{i,p-j} - \boldsymbol{\gamma}) = \sum_{l=1}^{p-1}\Phi_l\sum_{j=l}^{p-1}(\triangle\mathbf{w}_{i,p-j} - \boldsymbol{\gamma}) + \Pi\mathbf{C}^{*}(L)\varepsilon_{i0} + \varepsilon_{ip}. \tag{A.15}$$

$$\phi = vec(\Phi)\,, \tag{A.26}$$

Φ is given by (A.5), Σ_{ϵ} by (3.6), $\Sigma_{f|\epsilon}$ by (3.11), $\mathcal{R}$ denotes the $m \times mp$ matrix formed by the last m rows of Σ_{ϵ}^{-1}, $\mathcal{S}$ denotes the $m \times m$ matrix formed by the last m columns of $\mathcal{R}$, and

$$\theta = \left(\gamma', \quad \phi', \quad \sigma' \right)',$$

with $\sigma = vech(\Omega)$. The matrices $\mathcal{G}_l$, $l = 0,1,\ldots,p-2$, $\mathcal{H}_l$, $l = 1,2,\ldots,p-1$, and Ψ may be computed using numerical recursions invoking an appropriate stopping rule to replace the infinite sums in their definitions by finite sums. As $\Sigma_{f|\epsilon}$ has a block-tridiagonal structure, its determinant and inverse may be computed using similar formulae as given in (5.9), (5.10), and (5.11), respectively. A computationally attractive alternative to the formulation (A.16) of the unconditional log-likelihood function might in some cases be to factorize $\Pr(\triangle \mathbf{w}_{i1}, \triangle \mathbf{w}_{i2}, \ldots, \triangle \mathbf{w}_{ip})$ as

$$\Pr(\triangle \mathbf{w}_{i1}, \triangle \mathbf{w}_{i2}, \ldots, \triangle \mathbf{w}_{ip}) = \Pr(\triangle \mathbf{w}_{i1}) \cdot \Pr(\triangle \mathbf{w}_{i2} | \triangle \mathbf{w}_{i1})$$

$$\cdot \ldots \cdot \Pr(\triangle \mathbf{w}_{i,p-1} | \triangle \mathbf{w}_{i1}, \triangle \mathbf{w}_{i2}, \ldots, \triangle \mathbf{w}_{i,p-2}) \cdot \Pr(\triangle \mathbf{w}_{ip} | \triangle \mathbf{w}_{i1}, \triangle \mathbf{w}_{i2}, \ldots, \triangle \mathbf{w}_{i,p-1})\,.$$

We do not go into the details here. A note describing further details of the numerical algorithm used to compute the ML estimates is available from the authors upon request.

MD Estimation

Conditional on Σ_{ϵ} and $\Sigma_{f|\epsilon}$ (given by (3.6) and (3.11), respectively), the ML estimator of Φ (defined by (A.5)) and of γ is equivalent to the MD estimator that minimizes

$$\sum_{i=1}^{N} \left[(\mathbf{r}_{\epsilon i} - \mathbf{H}_{\epsilon i}\phi)' \Sigma_{\epsilon}^{-1} (\mathbf{r}_{\epsilon i} - \mathbf{H}_{\epsilon i}\phi) + (\mathbf{r}_{fi} - \mathbf{H}_{fi}\phi)' \Sigma_{f|\epsilon}^{-1} (\mathbf{r}_{fi} - \mathbf{H}_{fi}\phi) \right], \tag{A.27}$$

where $\mathbf{r}_{\epsilon i}$, $\mathbf{H}_{\epsilon i}$, $\mathbf{r}_{fi}$, and $\mathbf{H}_{fi}$ are defined in (A.17), (A.18), (A.21), and (A.23), respectively, and ϕ is given by (A.26). Therefore, conditional on γ, Σ_{ϵ}, and $\Sigma_{f|\epsilon}$, the MD estimator of Φ is given by

$$\phi = \left(\sum_{i=1}^{N} \mathbf{H}_{\epsilon i}' \Sigma_{\epsilon}^{-1} \mathbf{H}_{\epsilon i} + \mathbf{H}_{fi}' \Sigma_{f|\epsilon}^{-1} \mathbf{H}_{fi} \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{H}_{\epsilon i}' \Sigma_{\epsilon}^{-1} \mathbf{r}_{\epsilon i} + \mathbf{H}_{fi}' \Sigma_{f|\epsilon}^{-1} \mathbf{r}_{fi} \right), \tag{A.28}$$

and, conditional on Φ, Σ_{ϵ}, and $\Sigma_{f|\epsilon}$, the MD estimator of γ is equal to

$$\gamma = \left(N \mathbf{J}_{\epsilon} \Sigma_{\epsilon}^{-1} \mathbf{J}_{\epsilon}' + N \mathbf{J}_{f} \Sigma_{f|\epsilon}^{-1} \mathbf{J}_{f}' \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{J}_{\epsilon} \Sigma_{\epsilon}^{-1} \mathbf{l}_{\epsilon i} + \mathbf{J}_{f} \Sigma_{f|\epsilon}^{-1} \mathbf{l}_{fi} \right), \tag{A.29}$$

where

$$\mathbf{J}_{\epsilon} = \left[\mathbf{I}_m, \quad \mathbf{I}_m, \quad \ldots, \quad \mathbf{I}_m, \quad \Gamma' - (p-1)\,\Pi' \right]', \tag{A.30}$$

$$\mathbf{J}_{\mathrm{f}} = \Big(\; -\Pi', \quad -\Pi', \quad \ldots, \quad -\Pi' \; \Big)', \tag{A.31}$$

$$\mathbf{l}_{ei} = vec\Big(\; \triangle \mathbf{w}_{i1}, \quad \triangle \mathbf{w}_{i2}, \quad \ldots, \quad \triangle \mathbf{w}_{i,p-1}, \quad \textstyle\sum_{j=0}^{p-1} \triangle \mathbf{w}_{i,p-j} - \sum_{l=1}^{p-1} \Phi_l \sum_{j=l}^{p-1} \triangle \mathbf{w}_{i,p-j} \; \Big), \tag{A.32}$$

and

$$\mathbf{l}_{fi} = vec\Big(\; \triangle \mathbf{w}_{i,p+1} - \Phi \triangle \mathbf{z}_{ip}, \quad \triangle \mathbf{w}_{i,p+2} - \Phi \triangle \mathbf{z}_{i,p+1}, \quad \ldots, \quad \triangle \mathbf{w}_{iT} - \Phi \triangle \mathbf{z}_{i,T-1} \; \Big). \tag{A.33}$$

Using (A.28) and (A.29), an iterative scheme similar to that discussed in Section 6 can be employed for the computation of the MD estimates.

Table 1[1]
Bias and RMSE of Alternative Estimators of Panel VAR
N = 50

	Estimator	γ_1		γ_2		ϕ_{11}		ϕ_{12}		ϕ_{21}		ϕ_{22}	
Trend-Stationary PVAR (λ_{max}=0.6)	*True Value*	0.02		0.02		0.4		0.2		0.2		0.4	
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
T=3	GMM_a	0.0074	0.1962	0.0024	0.1507	0.3340	0.4948	0.1179	0.4011	0.1280	0.3997	0.3135	0.4751
	MD	0.0002	0.0232	-0.0004	0.0223	0.0246	0.1551	0.0037	0.1355	0.0081	0.1367	0.0130	0.1607
	ML	0.0001	0.0234	-0.0005	0.0226	0.0202	0.1441	0.0017	0.1275	0.0055	0.1276	0.0082	0.1507
T=10	GMM_a	0.0005	0.0088	0.0002	0.0088	0.1168	0.1356	0.0359	0.0802	0.0383	0.0819	0.1168	0.1349
	MD	0.0002	0.0066	0.0000	0.0063	0.0068	0.0552	-0.0004	0.0496	0.0032	0.0493	0.0071	0.0555
	ML	0.0003	0.0066	0.0001	0.0064	0.0073	0.0521	-0.0010	0.0470	0.0027	0.0475	0.0078	0.0517
Trend-Stationary PVAR (λ_{max}=0.8)	*True Value*	0.02		0.02		0.6		0.2		0.2		0.6	
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
T=3	GMM_a	-0.0071	0.2411	-0.0052	0.2710	0.7955	1.0269	-0.0365	0.3459	0.7201	1.2742	0.1851	0.5612
	MD	0.0011	0.0281	-0.0001	0.0142	0.0442	0.2052	-0.0074	0.0819	0.0417	0.3343	-0.0017	0.1530
	ML	0.0010	0.0287	-0.0001	0.0144	0.0367	0.1985	-0.0056	0.0781	0.0364	0.3310	-0.0003	0.1475
T=10	GMM_a	0.0003	0.0110	0.0003	0.0060	0.2786	0.3011	-0.0237	0.0562	0.2924	0.3436	0.0437	0.0934
	MD	-0.0001	0.0086	0.0000	0.0042	0.0133	0.0640	-0.0029	0.0261	0.0159	0.1077	0.0017	0.0458
	ML	0.0000	0.0087	0.0000	0.0043	0.0127	0.0607	-0.0030	0.0247	0.0121	0.1046	0.0029	0.0446
PureUnitRoot PVAR(λ_{max}=1)	*True Value*	0.02		0.02		1.0		0		0		1.0	
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
T=3	GMM_a	0.0025	0.1565	-0.0135	0.3779	0.9301	1.0613	-0.0353	0.5339	-0.0088	0.5424	0.9298	1.0529
	MD	0.0003	0.0274	-0.0003	0.0253	0.0140	0.0858	-0.0014	0.0842	-0.0007	0.0865	0.0105	0.0837
	ML	0.0004	0.0286	-0.0005	0.0267	0.0132	0.0855	-0.0013	0.0840	-0.0007	0.0861	0.0097	0.0835
T=10	GMM_a	0.0001	0.0191	0.0002	0.0193	0.3954	0.4085	-0.0019	0.1019	-0.0004	0.1054	0.4055	0.4187
	MD	-0.0007	0.0142	0.0003	0.0139	0.0057	0.0229	-0.0015	0.0222	0.0012	0.0214	0.0068	0.0237
	ML	-0.0005	0.0152	0.0002	0.0150	0.0054	0.0228	-0.0015	0.0221	0.0012	0.0213	0.0065	0.0236

[1] For details of the Monte Carlo design, see Section 10.1. The data generating process is given by $(I_2 - \Phi L)(w_{it} - \mu_i - \gamma t) = \epsilon_{it}$, where μ_i is generated from (10.3), and ϵ_{it} is generated from (10.1). λ_{max} denotes the maximum eigenvalue of Φ, γ_j the j-th element of γ, $j = 1, 2$, and ϕ_{jk} the element in the j-th row and k-th column of Φ, $j, k = 1, 2$. 'RMSE' denotes the root mean square error, 'GMM_a' the conventional GMM estimator not incorporating initial conditions, 'MD' the Minimum Distance estimator, and 'ML' the Maximum Likelihood estimator.

Table 2[2]
Bias and RMSE of Alternative Estimators of Panel VAR
$N = 250$

	Estimator	γ_1		γ_2		ϕ_{11}		ϕ_{12}		ϕ_{21}		ϕ_{22}	
Trend-Stationary PVAR (λ_{max}=0.6)	*True Value*	0.02		0.02		0.4		0.2		0.2		0.4	
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
T=3	GMM_a	0.0269	0.8021	0.0300	0.9139	0.1170	0.2528	0.0646	0.2351	0.0695	0.2277	0.1081	0.2299
	MD	-0.0005	0.0100	0.0004	0.0100	0.0056	0.0750	0.0014	0.0577	0.0005	0.0595	0.0057	0.0740
	ML	-0.0004	0.0101	0.0004	0.0102	0.0027	0.0698	0.0010	0.0539	0.0005	0.0560	0.0044	0.0688
T=10	GMM_a	-0.0001	0.0044	0.0000	0.0042	0.0322	0.0490	0.0127	0.0381	0.0126	0.0391	0.0336	0.0492
	MD	0.0000	0.0028	0.0001	0.0029	0.0014	0.0242	0.0000	0.0212	-0.0010	0.0216	0.0019	0.0244
	ML	-0.0000	0.0029	0.0001	0.0029	0.0014	0.0227	-0.0004	0.0205	-0.0013	0.0205	0.0022	0.0229
Trend-Stationary PVAR (λ_{max}=0.8)	*True Value*	0.02		0.02		0.6		0.2		0.2		0.6	
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
T=3	GMM_a	0.0744	2.3086	0.0227	0.5933	0.4989	0.7402	-0.0003	0.2906	0.5751	0.9576	0.0543	0.4145
	MD	-0.0007	0.0123	0.0003	0.0066	0.0108	0.0939	-0.0014	0.0362	0.0074	0.1425	0.0025	0.0696
	ML	-0.0008	0.0125	0.0003	0.0067	0.0089	0.0893	-0.0007	0.0340	0.0056	0.1413	0.0030	0.0675
T=10	GMM_a	-0.0002	0.0062	-0.0001	0.0032	0.1184	0.1371	-0.0075	0.0340	0.1389	0.1753	0.0088	0.0510
	MD	-0.0000	0.0040	0.0001	0.0019	0.0024	0.0270	-0.0002	0.0109	-0.0002	0.0462	0.0012	0.0212
	ML	-0.0000	0.0041	0.0000	0.0019	0.0024	0.0255	-0.0004	0.0106	-0.0012	0.0442	0.0014	0.0206
Pure Unit Root PVAR(λ_{max}=1)	*True Value*	0.02		0.02		1.0		0		0		1.0	
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
T=3	GMM_a	-0.0039	0.0785	0.0002	0.0605	0.9454	1.0747	0.0230	0.5455	-0.0100	0.5626	0.9795	1.0941
	MD	0.0000	0.0119	0.0003	0.0118	0.0048	0.0381	0.0012	0.0373	-0.0010	0.0374	0.0032	0.0370
	ML	0.0000	0.0125	0.0004	0.0126	0.0043	0.0381	0.0012	0.0372	-0.0010	0.0374	0.0030	0.0369
T=10	GMM_a	0.0002	0.0092	0.0000	0.0091	0.3951	0.4087	0.0020	0.1017	-0.0041	0.1049	0.3939	0.4068
	MD	0.0003	0.0066	-0.0001	0.0064	0.0008	0.0095	0.0003	0.0095	-0.0003	0.0098	0.0007	0.0097
	ML	0.0002	0.0070	-0.0001	0.0069	0.0008	0.0095	0.0003	0.0095	-0.0003	0.0098	0.0006	0.0096

[2] See the footnote to Table 1 for a description of the data generating process and the notation used in this table.

Table 3[3]
Size and Power Properties of Tests for ϕ_{11}

Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$[4]

	Estimator	ϕ_{11}=0.1	ϕ_{11}=0.2	ϕ_{11}=0.3	ϕ_{11}=0.4	ϕ_{11}=0.5	ϕ_{11}=0.6	ϕ_{11}=0.7
N=50, T=3	GMM_a	0.0960	0.1290	0.1860	0.2590	0.3470	0.4470	0.5610
	ML	0.4220	0.1830	0.0670	0.0470	0.1200	0.3260	0.5800
N=50, T=10	GMM_a	0.7960	0.2630	0.0750	0.4310	0.9060	0.9980	1.0000
	ML	1.0000	0.9640	0.4540	0.0460	0.5480	0.9830	1.0000
N=250, T=3	GMM_a	0.1570	0.1010	0.0980	0.1620	0.2840	0.4570	0.6030
	ML	0.9920	0.8270	0.2920	0.0640	0.3410	0.8240	0.9870
N=250, T=10	GMM_a	1.0000	0.9930	0.4820	0.1600	0.9560	1.0000	1.0000
	ML	1.0000	1.0000	0.9950	0.0460	0.9960	1.0000	1.0000

Trend-Stationary Panel VAR with $\lambda_{max} = 0.8$[5]

	Estimator	ϕ_{11}=0.3	ϕ_{11}=0.4	ϕ_{11}=0.5	ϕ_{11}=0.6	ϕ_{11}=0.7	ϕ_{11}=0.8	ϕ_{11}=0.9
N=50, T=3	GMM_a	0.2330	0.2790	0.3290	0.3910	0.4610	0.5140	0.5710
	ML	0.2500	0.1130	0.0630	0.0570	0.1020	0.2110	0.3770
N=50, T=10	GMM_a	0.1310	0.1640	0.4630	0.8110	0.9720	0.9970	1.0000
	ML	0.9950	0.8950	0.3330	0.0600	0.4840	0.9610	1.0000
N=250, T=3	GMM_a	0.1960	0.2190	0.2610	0.3060	0.3680	0.4390	0.5060
	ML	0.7720	0.4730	0.1500	0.0580	0.2000	0.5390	0.8040
N=250, T=10	GMM_a	0.7740	0.2690	0.0650	0.4520	0.9030	0.9950	1.0000
	ML	1.0000	1.0000	0.9630	0.0460	0.9770	1.0000	1.0000

Pure Unit Root Panel VAR with $\lambda_{max} = 1$[6]

	Estimator	ϕ_{11}=0.7	ϕ_{11}=0.8	ϕ_{11}=0.9	ϕ_{11}=1	ϕ_{11}=1.1	ϕ_{11}=1.2	ϕ_{11}=1.3
N=50, T=3	GMM_a	0.3410	0.4350	0.5160	0.6060	0.6880	0.7520	0.8120
	ML	0.9050	0.6320	0.2140	0.0550	0.2580	0.7250	0.9800
N=50, T=10	GMM_a	0.2950	0.6810	0.9550	1.0000	1.0000	1.0000	1.0000
	ML	1.0000	1.0000	0.9720	0.0640	1.0000	1.0000	1.0000
N=250, T=3	GMM_a	0.3410	0.4290	0.5260	0.6060	0.6880	0.7640	0.8200
	ML	1.0000	0.9990	0.7280	0.0480	0.7960	1.0000	1.0000
N=250, T=10	GMM_a	0.3140	0.6690	0.9470	0.9970	1.0000	1.0000	1.0000
	ML	1.0000	1.0000	1.0000	0.0510	1.0000	1.0000	1.0000

[3] See the footnote to Table 1 for a description of the data generating process and the notation used in this table.
[4] The table reports the fraction of rejections for tests of H_0: $\phi_{11} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, versus two-sided alternatives. The true value of ϕ_{11} is equal to 0.4.
[5] The table reports the fraction of rejections for tests of H_0: $\phi_{11} = \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, versus two-sided alternatives. The true value of ϕ_{11} is equal to 0.6.
[6] The table reports the fraction of rejections for tests of H_0: $\phi_{11} = \{0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3\}$, versus two-sided alternatives. The true value of ϕ_{11} is equal to 1

Table 4[7]
Size and Power Properties of Tests for ϕ_{12}

Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$[8]

	Estimator	ϕ_{12}=-0.1	ϕ_{12}=0	ϕ_{12}=0.1	ϕ_{12}=0.2	ϕ_{12}=0.3	ϕ_{12}=0.4	ϕ_{12}=0.5
N=50, T=3	GMM_a	0.1220	0.0750	0.0690	0.1000	0.1580	0.2470	0.3770
	ML	0.6370	0.3390	0.1270	0.0580	0.1320	0.3600	0.6670
N=50, T=10	GMM_a	0.9670	0.6760	0.1810	0.1220	0.5640	0.9350	0.9990
	ML	1.0000	0.9670	0.5510	0.0500	0.5950	0.9920	1.0000
N=250, T=3	GMM_a	0.2570	0.1420	0.0880	0.1090	0.1970	0.3550	0.5370
	ML	0.9990	0.9430	0.3950	0.0440	0.4160	0.9390	0.9990
N=250, T=10	GMM_a	1.0000	0.9980	0.6790	0.0710	0.8690	1.0000	1.0000
	ML	1.0000	1.0000	0.9980	0.0450	0.9970	1.0000	1.0000

Trend-Stationary Panel VAR with $\lambda_{max} = 0.8$[9]

	Estimator	ϕ_{12}=-0.1	ϕ_{12}=0	ϕ_{12}=0.1	ϕ_{12}=0.2	ϕ_{12}=0.3	ϕ_{12}=0.4	ϕ_{12}=0.5
N=50, T=3	GMM_a	0.1650	0.1720	0.1940	0.2190	0.2450	0.2760	0.3150
	ML	0.1350	0.0930	0.0800	0.0610	0.0900	0.1420	0.2140
N=50, T=10	GMM_a	0.0840	0.1280	0.2510	0.4600	0.6910	0.8570	0.9540
	ML	0.7870	0.4630	0.1320	0.0530	0.2090	0.5360	0.8540
N=250, T=3	GMM_a	0.1720	0.1940	0.2190	0.2580	0.2890	0.3250	0.3710
	ML	0.4410	0.2480	0.0940	0.0510	0.1100	0.2600	0.4830
N=250, T=10	GMM_a	0.3660	0.1090	0.0810	0.2870	0.6630	0.9010	0.9860
	ML	1.0000	0.9930	0.5870	0.0450	0.5930	0.9870	1.0000

Pure Unit Root Panel VAR with $\lambda_{max} = 1$[10]

	Estimator	ϕ_{12}=-0.3	ϕ_{12}=-0.2	ϕ_{12}=-0.1	ϕ_{12}=0	ϕ_{12}=0.1	ϕ_{12}=0.2	ϕ_{12}=0.3
N=50, T=3	GMM_a	0.1450	0.0960	0.0720	0.0530	0.0570	0.0910	0.1430
	ML	0.9490	0.6430	0.2280	0.0560	0.2240	0.6640	0.9270
N=50, T=10	GMM_a	0.9160	0.7000	0.3530	0.1520	0.3400	0.6900	0.9290
	ML	1.0000	1.0000	0.9920	0.0450	0.9930	1.0000	1.0000
N=250, T=3	GMM_a	0.1200	0.0820	0.0510	0.0550	0.0710	0.0960	0.1350
	ML	1.0000	0.9990	0.7860	0.0570	0.7680	1.0000	1.0000
N=250, T=10	GMM_a	0.9300	0.7160	0.3670	0.1610	0.3420	0.6770	0.9120
	ML	1.0000	1.0000	1.0000	0.0460	1.0000	1.0000	1.0000

[7] See the footnote to Table 1 for a description of the data generating process and the notation used for this table.
[8] The table reports the fraction of rejections for tests of H_0: $\phi_{12} = \{-0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, versus two-sided alternatives. The true value of ϕ_{12} is equal to 0.2.
[9] The table reports the fraction of rejections for tests of H_0: $\phi_{12} = \{-0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, versus two-sided alternatives. The true value of ϕ_{12} is equal to 0.2.
[10] The table reports the fraction of rejections for tests of H_0: $\phi_{12} = \{-0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3\}$, versus two-sided alternatives. The true value of ϕ_{12} is equal to 0.

Table 5[11]
Size and Power Properties of Tests for ϕ_{21}

Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$[12]

	Estimator	ϕ_{21}=-0.1	ϕ_{21}=0	ϕ_{21}=0.1	ϕ_{21}=0.2	ϕ_{21}=0.3	ϕ_{21}=0.4	ϕ_{21}=0.5
N=50, T=3	GMM_a	0.1230	0.0860	0.0700	0.0790	0.1350	0.2260	0.3370
	ML	0.6560	0.3880	0.1400	0.0650	0.1340	0.3790	0.6730
N=50, T=10	GMM_a	0.9700	0.6860	0.1890	0.1090	0.5620	0.9290	0.9950
	ML	1.0000	0.9900	0.5700	0.0600	0.5830	0.9890	1.0000
N=250, T=3	GMM_a	0.2530	0.1350	0.0890	0.1170	0.2140	0.3600	0.5070
	ML	1.0000	0.9460	0.4120	0.0490	0.1330	0.9530	1.0000
N=250, T=10	GMM_a	1.0000	1.0000	0.6950	0.0770	0.8820	1.0000	1.0000
	ML	1.0000	1.0000	0.9970	0.0500	1.0000	1.0000	1.0000

Trend-Stationary Panel VAR with $\lambda_{max} = 0.8$[13]

	Estimator	ϕ_{21}=-0.1	ϕ_{21}=0	ϕ_{21}=0.1	ϕ_{21}=0.2	ϕ_{21}=0.3	ϕ_{21}=0.4	ϕ_{21}=0.5
N=50, T=3	GMM_a	0.2990	0.2110	0.1580	0.1220	0.1400	0.1720	0.2580
	ML	0.9610	0.7260	0.2600	0.0540	0.2340	0.6810	0.9450
N=50, T=10	GMM_a	1.0000	0.9930	0.7350	0.1110	0.3880	0.9420	0.9980
	ML	1.0000	1.0000	0.9890	0.0620	0.9740	1.0000	1.0000
N=250, T=3	GMM_a	0.3490	0.2410	0.1900	0.1580	0.1830	0.2570	0.3570
	ML	0.9840	0.9820	0.7980	0.0530	0.7830	0.9830	0.9850
N=250, T=10	GMM_a	1.0000	1.0000	0.9130	0.0700	0.8190	1.0000	1.0000
	ML	1.0000	1.0000	1.0000	0.0480	1.0000	1.0000	1.0000

Pure Unit Root Panel VAR with $\lambda_{max} = 1$[14]

	Estimator	ϕ_{21}=-0.3	ϕ_{21}=-0.2	ϕ_{21}=-0.1	ϕ_{21}=0	ϕ_{21}=0.1	ϕ_{21}=0.2	ϕ_{21}=0.3
N=50, T=3	GMM_a	0.1680	0.1160	0.0710	0.0520	0.0570	0.0910	0.1330
	ML	0.9460	0.6870	0.2610	0.0560	0.2270	0.6700	0.9470
N=50, T=10	GMM_a	0.9360	0.7250	0.3550	0.1460	0.3440	0.7030	0.9350
	ML	1.0000	1.0000	0.9990	0.0540	0.9960	1.0000	1.0000
N=250, T=3	GMM_a	0.1380	0.0980	0.0720	0.0570	0.0700	0.0990	0.1490
	ML	1.0000	1.0000	0.7740	0.0560	0.7770	1.0000	1.0000
N=250, T=10	GMM_a	0.9270	0.6940	0.3560	0.1640	0.3640	0.7120	0.9350
	ML	1.0000	1.0000	1.0000	0.0500	1.0000	1.0000	1.0000

[11] See the footnote to Table 1 for a description of the data generating process and the notation used for this table.
[12] The table reports the fraction of rejections for tests of H_0: $\phi_{21} = \{-0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, versus two-sided alternatives. The true value of ϕ_{21} is equal to 0.2.
[13] The table reports the fraction of rejections for tests of H_0: $\phi_{21} = \{-0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, versus two-sided alternatives. The true value of ϕ_{21} is equal to 0.2.
[14] The table reports the fraction of rejections for tests of H_0: $\phi_{21} = \{-0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3\}$, versus two-sided alternatives. The true value of ϕ_{21} is equal to 0.

Table 6[15]
Size and Power Properties of Tests for ϕ_{22}

Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$[16]

	Estimator	$\phi_{22}=0.1$	$\phi_{22}=0.2$	$\phi_{22}=0.3$	$\phi_{22}=0.4$	$\phi_{22}=0.5$	$\phi_{22}=0.6$	$\phi_{22}=0.7$
N=50, T=3	GMM_a	0.0940	0.1120	0.1530	0.2490	0.3430	0.4730	0.5780
	ML	0.4810	0.2140	0.0820	0.0620	0.1240	0.3020	0.5420
N=50, T=10	GMM_a	0.8030	0.2600	0.0630	0.4380	0.9080	1.0000	1.0000
	ML	1.0000	0.9660	0.4470	0.0540	0.5510	0.9850	1.0000
N=250, T=3	GMM_a	0.1630	0.0850	0.0800	0.1440	0.2660	0.4310	0.6030
	ML	0.9960	0.8230	0.2730	0.0540	0.3380	0.8250	0.9980
N=250, T=10	GMM_a	1.0000	0.9980	0.4570	0.1700	0.9650	1.0000	1.0000
	ML	1.0000	1.0000	0.9910	0.0520	0.9920	1.0000	1.0000

Trend-Stationary Panel VAR with $\lambda_{max} = 0.8$[17]

	Estimator	$\phi_{22}=0.3$	$\phi_{22}=0.4$	$\phi_{22}=0.5$	$\phi_{22}=0.6$	$\phi_{22}=0.7$	$\phi_{22}=0.8$	$\phi_{22}=0.9$
N=50, T=3	GMM_a	0.1320	0.1160	0.1250	0.1480	0.1870	0.2370	0.2820
	ML	0.4900	0.2670	0.0980	0.0590	0.0950	0.2520	0.4960
N=50, T=10	GMM_a	0.8950	0.5470	0.1480	0.0970	0.4800	0.8790	0.9840
	ML	1.0000	0.9660	0.5850	0.0510	0.6170	0.9940	1.0000
N=250, T=3	GMM_a	0.2140	0.1670	0.1470	0.1450	0.1770	0.2250	0.2860
	ML	0.9560	0.7720	0.2860	0.0510	0.3110	0.8110	0.9540
N=250, T=10	GMM_a	1.0000	0.9660	0.4800	0.0640	0.9920	0.9850	1.0000
	ML	1.0000	1.0000	0.9970	0.0550	1.0000	1.0000	1.0000

PureUnit Root Panel VAR with $\lambda_{max} = 1$[18]

	Estimator	$\phi_{22}=0.7$	$\phi_{22}=0.8$	$\phi_{22}=0.9$	$\phi_{22}=1$	$\phi_{22}=1.1$	$\phi_{22}=1.2$	$\phi_{22}=1.3$
N=50, T=3	GMM_a	0.3370	0.4370	0.5300	0.6260	0.7080	0.7740	0.8250
	ML	0.9250	0.6280	0.1990	0.0520	0.2480	0.7280	0.9660
N=50, T=10	GMM_a	0.3480	0.7140	0.9640	0.9980	1.0000	1.0000	1.0000
	ML	1.0000	1.0000	0.9670	0.0510	1.0000	1.0000	1.0000
N=250, T=3	GMM_a	0.3580	0.4500	0.5480	0.6380	0.7080	0.7640	0.8060
	ML	1.0000	1.0000	0.7470	0.0540	0.8010	1.0000	1.0000
N=250, T=10	GMM_a	0.2970	0.6860	0.9480	0.9990	1.0000	1.0000	1.0000
	ML	1.0000	1.0000	1.0000	0.0460	1.0000	1.0000	1.0000

[15] See the footnote to Table 1 for a description of the data generating process and the notation used for this table.
[16] The table reports the fraction of rejections for tests of H_0: $\phi_{22} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, versus two-sided alternatives. The true value of ϕ_{22} is equal to 0.4.
[17] The table reports the fraction of rejections for tests of H_0: $\phi_{22} = \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, versus two-sided alternatives. The true value of ϕ_{22} is equal to 0.6.
[18] The table reports the fraction of rejections for tests of H_0: $\phi_{22} = \{0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3\}$, versus two-sided alternatives. The true value of ϕ_{22} is equal to 1.

Table 7a[19]
Bias and RMSE of Alternative Estimators of Cointegrated Panel VAR

	Estimator	γ_1=0.02		γ_2=0.02		α_1=-0.6		α_2=-0.2		β_2=-1	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=50, T=3	MD	0.0003	0.0190	0.0002	0.0082	0.0362	0.1957	0.0019	0.0317	-0.0561	0.4625
	ML	0.0003	0.0190	0.0002	0.0082	0.0423	0.1423	0.0046	0.0281	-0.0209	0.3447
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=50, T=10	MD	0.0003	0.0086	0.0002	0.0039	0.0005	0.1013	-0.0008	0.0154	-0.0220	0.2252
	ML	0.0003	0.0086	0.0002	0.0039	0.0070	0.0682	0.0005	0.0129	-0.0025	0.1509
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=250, T=3	MD	-0.0002	0.0086	-0.0001	0.0037	0.0029	0.1006	-0.0003	0.0155	-0.0166	0.2236
	ML	-0.0002	0.0086	-0.0000	0.0037	0.0073	0.0693	0.0007	0.0131	-0.0004	0.1599
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=250, T=10	MD	-0.0001	0.0030	-0.0001	0.0024	0.0016	0.0263	0.0003	0.0050	0.0035	0.0433
	ML	-0.0001	0.0030	-0.0001	0.0024	0.0020	0.0211	0.0002	0.0045	0.0030	0.0401

Table 7b[20]
Size and Power Properties of Tests for α_1 under Cointegrated Panel VAR

	Estimator	α_1=-0.9	α_1=-0.8	α_1=-0.7	α_1=-0.6	α_1=-0.5	α_1=-0.4	α_1=-0.3
N=50, T=3	ML	0.3620	0.1680	0.0600	0.0410	0.1260	0.3290	0.5690
N=50, T=10	ML	1.0000	0.9530	0.4830	0.0630	0.5810	0.9920	1.0000
N=250, T=3	ML	0.8620	0.6880	0.2590	0.0600	0.3200	0.7590	0.8850
N=250, T=10	ML	1.0000	1.0000	0.9920	0.0460	0.9970	1.0000	1.0000

[19] See the footnote to Table 1 for a description of the data generating process, where now $\Phi = I_2 + \alpha\beta'$, with $\alpha = (\alpha_1, \alpha_2)'$ and $\beta = (\beta_1, \beta_2)'$; α and β are commonly referred to as the error correction coefficients and the cointegrating vectors, respectively. The remaining notation is as described in the footnote to Table 1.

[20] The table reports the fraction of rejections for tests of H_0: $\alpha_1 = \{-0.9, -0.8, -0.7, -0.6, -0.5, -0.4, -0.3\}$, versus two-sided alternatives. The true value of α_1 is equal to -0.6.

Table 7c[21]

Size and Power Properties of Tests for α_2 under Cointegrated Panel VAR

	Estimator	α_2=-0.5	α_2=-0.4	α_2=-0.3	α_2=-0.2	α_2=-0.1	α_2=0	α_2=0.1
N=50, T=3	ML	0.9770	0.9760	0.8800	0.0410	0.9220	0.9740	0.9770
N=50, T=10	ML	1.0000	1.0000	1.0000	0.0480	1.0000	1.0000	1.0000
N=250, T=3	ML	0.9740	0.9740	0.9740	0.0570	0.9740	0.9740	0.9740
N=250, T=10	ML	1.0000	1.0000	1.0000	0.0490	1.0000	1.0000	1.0000

Table 7d[22]

Size and Power Properties of Tests for β_2 under Cointegrated Panel VAR

	Estimator	β_2=-1.3	β_2=-1.2	β_2=-1.1	β_2=-1	β_2=-0.9	β_2=-0.8	β_2=-0.7
N=50, T=3	ML	0.1170	0.0720	0.0610	0.0670	0.0730	0.1090	0.1630
N=50, T=10	ML	0.9180	0.6140	0.1940	0.0370	0.1840	0.5750	0.8840
N=250, T=3	ML	0.3930	0.2180	0.1110	0.0650	0.1270	0.2460	0.4280
N=250, T=10	ML	1.0000	1.0000	0.6510	0.0530	0.7060	0.9990	1.0000

Table 7e[23]

Size and Power Properties of Cointegration Test

	Estimator	H_0 vs. H_1	H_1 vs. H_2
N=50, T=3	ML	1.0000	0.0300
N=50, T=10	ML	1.0000	0.0570
N=250, T=3	ML	1.0000	0.0500
N=250, T=10	ML	1.0000	0.0470

[21] The table reports the fraction of rejections for tests of H_0: $\alpha_2 = \{-0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1\}$, versus two-sided alternatives. The true value of α_2 is equal to -0.2.
[22] The table reports the fraction of rejections for tests of H_0: $\beta_2 = \{-1.3, -1.2, -1.1, -1, -0.9, -0.8, -0.7\}$, versus two-sided alternatives. The true value of β_2 is equal to -1.
[23] The table reports the fraction of rejections for tests of H_r: $rank(\Pi) = r$ versus H_{r+1}: $rank(\Pi) = r+1$, $r = 0, 1$, where the true rank of Π is equal to 1.

Table 8a[24]

Bias and RMSE of Alternative Conventional GMM Estimators under Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$

	Estimator	γ_1=0.02		γ_2=0.02		ϕ_{11}=0.4		ϕ_{12}=0.2		ϕ_{21}=0.2		ϕ_{22}=0.4	
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
N=50, T=3	GMM_a	-0.0028	0.0572	-0.0008	0.0540	0.3562	0.4941	0.1270	0.3908	0.1483	0.3940	0.3262	0.4764
	GMM_b	-0.0010	0.0458	-0.0001	0.0236	0.3652	0.5043	0.1339	0.4133	0.1493	0.4025	0.3370	0.4848
	MD	0.0003	0.0233	-0.0003	0.0223	0.0257	0.1569	0.0043	0.1362	0.0087	0.1375	0.0163	0.1600
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
N=50, T=10	GMM_a	0.0005	0.0088	0.0002	0.0088	0.1168	0.1356	0.0359	0.0802	0.0383	0.0819	0.1168	0.1349
	GMM_b	0.0002	0.0066	0.0000	0.0064	0.1002	0.1290	0.0278	0.0792	0.0299	0.0833	0.0997	0.1288
	MD	0.0002	0.0066	0.0000	0.0063	0.0068	0.0552	-0.0004	0.0496	0.0032	0.0493	0.0071	0.0555
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
N=250,T=3	GMM_a	0.0015	0.0398	0.0007	0.0397	0.1203	0.2537	0.0708	0.2339	0.0732	0.2274	0.1131	0.2288
	GMM_b	-0.0003	0.0105	0.0001	0.0141	0.1286	0.2536	0.0779	0.2360	0.0787	0.2310	0.1212	0.2291
	MD	-0.0004	0.0100	0.0004	0.0100	0.0057	0.0752	0.0017	0.0575	0.0008	0.0595	0.0056	0.0734
		Bias	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
N=250,T=10	GMM_a	-0.0001	0.0044	0.0000	0.0042	0.0322	0.0490	0.0127	0.0381	0.0126	0.0391	0.0336	0.0492
	GMM_b	0.0000	0.0029	0.0001	0.0029	0.0310	0.0483	0.0123	0.0380	0.0123	0.0390	0.0324	0.0484
	MD	0.0000	0.0028	0.0001	0.0029	0.0014	0.0242	0.0000	0.0212	-0.0010	0.0216	0.0019	0.0214

Table 8b[25]

Size and Power Properties of Tests for ϕ_{11} under Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$

	Estimator	ϕ_{11}=0.1	ϕ_{11}=0.2	ϕ_{11}=0.3	ϕ_{11}=0.4	ϕ_{11}=0.5	ϕ_{11}=0.6	ϕ_{11}=0.7
N=50, T=3	GMM_a	0.1070	0.1350	0.1930	0.2610	0.3450	0.4400	0.5540
	GMM_b	0.1770	0.2260	0.2870	0.3690	0.4610	0.5820	0.6730
N=50, T=10	GMM_a	0.7960	0.2630	0.0750	0.4310	09060	0.9980	1.0000
	GMM_b	0.8390	0.4780	0.1700	0.5100	0.8580	0.9110	0.9250
N=250, T=3	GMM_a	0.2050	0.1320	0.1200	0.1520	0.2780	0.4210	0.5870
	GMM_b	0.2850	0.2010	0.1830	0.2660	0.3970	0.5800	0.7100
N=250, T=10	GMM_a	1.0000	0.9930	0.4820	0.1600	0.9560	1.0000	1.0000
	GMM_b	1.0000	0.9990	0.7100	0.3090	0.9820	1.0000	1.0000

[24] See the footnote to Table 1 for a description of the data generating process. 'GMM_b' denotes the conventional GMM estimator incorporating initial conditions. The remaining notation is as described in the footnote to Table 1.

[25] The table reports the fraction of rejections for tests of H_0: $\phi_{11} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, versus two-sided alternatives. The true value of ϕ_{11} is equal to *0.4*.

Table 9a[26]

Bias and RMSE of Alternative Estimators under Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$ and Non-Normal Disturbances

	Estimator	γ_1=0.02		γ_2=0.02		ϕ_{11}=0.4		ϕ_{12}=0.2		ϕ_{21}=0.2		ϕ_{22}=0.4	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=50, T=3	GMM_a	-0.0087	0.5155	-0.0002	0.2590	0.3716	0.5353	0.1344	0.4174	0.1405	0.4052	0.3088	0.4901
	MD	0.0000	0.0219	-0.0000	0.0229	0.0169	0.1966	0.0000	0.1386	0.0112	0.1386	0.0256	0.1969
	ML	-0.0008	0.0222	-0.0008	0.0230	0.0138	0.1809	0.0013	0.1321	0.0112	0.1322	0.0169	0.1811
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=50, T=10	GMM_a	0.0002	0.0089	0.0003	0.0088	0.1216	0.1401	0.0350	0.0784	0.0353	0.0798	0.1146	0.1336
	MD	-0.0002	0.0065	-0.0001	0.0062	0.0101	0.0679	0.0016	0.0488	0.0017	0.0506	0.0096	0.0664
	ML	-0.0002	0.0066	-0.0001	0.0063	0.0089	0.0543	0.0006	0.0459	0.0009	0.0477	0.0090	0.0540
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=250, T=3	GMM_a	0.0008	0.2104	0.0111	0.3238	0.0976	0.2412	0.0655	0.2346	0.0452	0.2245	0.1118	0.2431
	MD	0.0003	0.0101	0.0001	00097	0.0044	0.0898	-0.0025	0.0603	0.0012	0.0600	0.0033	0.0928
	ML	0.0001	0.0102	-0.0000	0.0098	0.0024	0.0808	-0.0026	0.0562	0.0006	0.0562	-0.0002	0.0839
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=250,T=10	GMM_a	-0.0000	0.0042	-0.0000	0.0042	0.0334	0.0485	0.0138	0.0393	0.0140	0.0382	0.0327	0.0486
	MD	-0.0001	0.0029	-0.0001	0.0028	0.0018	0.0287	0.0008	0.0221	-0.0004	0.0223	-0.0001	0.0308
	ML	-0.0001	0.0030	-0.0001	0.0029	0.0015	0.0229	0.0003	0.0209	-0.0003	0.0204	0.0007	0.0240

Table 9b[27]

Size and Power Properties of Tests for ϕ_{11} under Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$ and Non-Normal Disturbances

	Estimator	ϕ_{11}=0.1	ϕ_{11}=0.2	ϕ_{11}=0.3	ϕ_{11}=0.4	ϕ_{11}=0.5	ϕ_{11}=0.6	ϕ_{11}=0.7
N=50, T=3	GMM_a	0.0910	0.1340	0.1940	0.2710	0.3730	0.4730	0.5800
	ML	0.4430	0.2360	0.1260	0.0960	0.2000	0.3540	0.5640
N=50, T=10	GMM_a	0.7740	0.2360	0.0780	0.4630	0.9080	0.9990	1.0000
	ML	1.0000	0.9600	0.4300	0.0570	0.5620	0.9810	1.0000
N=250, T=3	GMM_a	0.1790	0.1060	0.0850	0.1390	0.2510	0.4070	0.5650
	ML	0.9980	0.7800	0.3030	0.0990	0.3690	0.7800	0.9690
N=250, T=10	GMM_a	1.0000	0.9980	0.4550	0.1500	0.9620	1.0000	1.0000
	ML	1.0000	1.0000	0.9930	0.0450	0.9910	1.0000	1.0000

[26] For details of the Monte Carlo design, see Section 10.1. The data generating process is given by $(I_2 - \Phi L)(w_{it} - \mu_i - \gamma t) = \epsilon_{it}$, where μ_i is generated from (10.3), and ϵ_{it} is generated from (10.2). See the footnote to Table 1 for a description of the notation used in this table.

[27] The table reports the fraction of rejections for tests of H_0: $\phi_{11} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, versus two-sided alternatives. The true value of ϕ_{11} is equal to 0.4.

Table9c[28]

Size and Power Properties of Tests for ϕ_{12} under Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$ and Non-Normal Disturbances

	Estimator	ϕ_{12}=-0.1	ϕ_{12}=0	ϕ_{12}=0.1	ϕ_{12}=0.2	ϕ_{12}=0.3	ϕ_{12}=0.4	ϕ_{12}=0.5
N=50, T=3	GMM_a	0.1400	0.0860	0.0670	0.1060	0.1760	0.2670	0.3800
	ML	0.6210	0.3410	0.1220	0.0620	0.1490	0.3910	0.6810
N=50, T=10	GMM_a	0.9650	0.6890	0.2090	0.1030	0.5510	0.9330	0.9970
	ML	1.0000	0.9920	0.5400	0.0430	0.6000	0.9860	1.0000
N=250, T=3	GMM_a	0.2970	0.1580	0.0950	0.1070	0.1820	0.3190	0.4910
	ML	0.9980	0.9300	0.4080	0.0460	0.4150	0.9330	1.0000
N=250, T=10	GMM_a	1.0000	1.0000	0.6680	0.0580	0.8850	1.0000	1.0000
	ML	1.0000	1.0000	0.9970	0.0350	0.9980	1.0000	1.0000

Table9d[29]

Size and Power Properties of Tests for ϕ_{21} under Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$ and Non-Normal Disturbances

	Estimator	ϕ_{21}=-0.1	ϕ_{21}=0	ϕ_{21}=0.1	ϕ_{21}=0.2	ϕ_{21}=0.3	ϕ_{21}=0.4	ϕ_{21}=0.5
N=50, T=3	GMM_a	0.1240	0.0880	0.0700	0.0960	0.1470	0.2240	0.3310
	ML	0.6640	0.3700	0.1660	0.0670	0.1380	0.3800	0.6730
N=50, T=10	GMM_a	0.9730	0.7030	0.2000	0.1060	0.5620	0.9320	0.9990
	ML	1.0000	0.9900	0.5860	0.0430	0.5830	0.9940	1.0000
N=250, T=3	GMM_a	0.2840	0.1400	0.0880	0.1020	0.2000	0.3550	0.5600
	ML	0.9990	0.9400	0.4510	0.0500	0.4070	0.9260	0.9990
N=250, T=10	GMM_a	1.0000	0.9990	0.6850	0.0850	0.8970	1.0000	1.0000
	ML	1.0000	1.0000	0.9980	0.0550	0.9980	1.0000	1.0000

Table 9e[30]

Size and Power Properties of Tests for ϕ_{22} under Trend-Stationary Panel VAR with $\lambda_{max} = 0.6$ and Non-Normal Disturbances

	Estimator	ϕ_{22}=0.1	ϕ_{22}=0.2	ϕ_{22}=0.3	ϕ_{22}=0.4	ϕ_{22}=0.5	ϕ_{22}=0.6	ϕ_{22}=0.7
N=50, T=3	GMM_a	0.0950	0.1230	0.1920	0.2600	0.3510	0.4460	0.5720
	ML	0.4270	0.2160	0.1060	0.1100	0.1980	0.3830	0.5710
N=50, T=10	GMM_a	0.8160	0.2720	0.0800	0.4380	0.8830	0.9970	1.0000
	ML	1.0000	0.9680	0.4200	0.0600	0.5840	0.9750	1.0000
N=250, T=3	GMM_a	0.1770	0.1050	0.1050	0.1660	0.2940	0.4410	0.6170
	ML	0.9800	0.7950	0.3290	0.1100	0.3000	0.7760	0.9570
N=250, T=10	GMM_a	1.0000	0.9900	0.4980	0.1690	0.9640	1.0000	1.0000
	ML	1.0000	1.0000	0.9860	0.0630	0.9850	1.0000	1.0000

[28] The table reports the fraction of rejections for tests of H_0: $\phi_{12} = \{-0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, versus two-sided alternatives. The true value of ϕ_{12} is equal to 0.2.

[29] The table reports the fraction of rejections for tests of H_0: $\phi_{21} = \{-0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5\}$, versus two-sided alternatives. The true value of ϕ_{21} is equal to 0.2.

[30] The table reports the fraction of rejections for tests of H_0: $\phi_{22} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, versus two-sided alternatives. The true value of ϕ_{22} is equal to 0.4.

References

Ahn, S.C., and P. Schmidt (1995): Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-27.

Amemiya, T. (1973): Generalized Least Squares with an Estimated Autocovariance Matrix, *Econometrica*, 41, 723-732.

Amemiya, T., and W. Fuller (1967): A Comparative Study of Alternative Estimators in a Distributed-Lag Model, *Econometrica*, 35, 509-529.

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

Anderson, T.W., and C. Hsiao (1982): Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, 18, 47-82.

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

Arellano, M., and O. Bover (1995): Another Look at the Instrumental Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-51.

Arellano, M., and B. Honoré (2000): Panel Data Models: Some Recent Developments, in: J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Vol. 5, Amsterdam: North-Holland (Forthcoming).

Bhargava, A., and J.D. Sargan (1983): Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods, *Econometrica*, 51, 1635-1659.

Binder, M., and M.H. Pesaran (2000): Solution of Finite-Horizon Multivariate Linear Rational Expectations Models and Sparse Linear Systems, *Journal of Economic Dynamics and Control*, 24, 325-346.

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

Blundell, R., and R.J. Smith (1991): Initial Conditions and Efficient Estimation in Dynamic Panel Data Models, *Annales d'Economie et de Statistique*, 20/21, 109-123.

Chamberlain, G. (1984): Panel Data, in: Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, Vol. 2, Amsterdam: North-Holland, 1247-1318.

Engle, R.F., D. Hendry, and J.-F. Richard (1983): Exogeneity, *Econometrica*, 51, 277-304.

Golub, G.H., and C.F. Van Loan (1996): *Matrix Computations*, 3rd Ed., Baltimore: Johns Hopkins University Press.

Granger, C.W.J., and J.-L. Lin (1995): Causality in the Long Run, *Econometric Theory*, 11, 530-536.

Hahn, J. (1999): How Informative is the Initial Condition in the Dynamic Panel Data Model with Fixed Effects?, *Journal of Econometrics*, 93, 309-326.

Hansen, L.P. (1982): Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029-1054.

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

Holtz-Eakin, D., W. Newey, and H.S. Rosen (1988): Estimating Vector Autoregressions with Panel Data, *Econometrica*, 56, 1371-1395.

Hsiao, C. (1979): Autoregressive Modelling of Canadian Money and Income Data, *Journal of the American Statistical Association*, 74, 553-560.

Hsiao, C. (1982): Autoregressive Modelling and Causal Ordering of Economic Variables, *Journal of Economic Dynamics and Control*, 4, 243-259.

Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu (1999): Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, Mimeo, University of Southern California, University of Cambridge, and University of Wisconsin.

Im, K.S., M.H. Pesaran, and Y. Shin (1997): Testing for Unit Roots in Heterogeneous Panels, Mimeo, University of Cambridge.

Johansen, S. (1995): *Likelihood Based Inference on Cointegration in the Vector Autoregressive Model*, Oxford: Oxford University Press.

King, R.G., C. Plosser, J.H. Stock, and M.W. Watson (1991): Stochastic Trends and Economic Fluctuations, *American Economic Review*, 81, 819-840.

Levin, A., and C.F. Lin (1993): Unit Root Tests in Panel Data: New Results, Mimeo, University of California-San Diego.

Nerlove, M. (1999): Likelihood Inference for Dynamic Panel Data Models, *Annales d'Economie et de Statistique*, 55/56, 370-410.

Neyman, J., and E. Scott (1948): Consistent Estimates Based on Partially Consistent Observations, *Econometrica*, 16, 1-32.

Pesaran, M.H., and Y. Shin (1999): Long-Run Structural Modelling, Mimeo, University of Cambridge and University of Edinburgh.

Pesaran, M.H., Y. Shin, and R.J. Smith (2000): Structural Analysis of Vector Error Correction Models with Exogenous I(1) Variables, *Journal of Econometrics* (Forthcoming).

Phillips, P.C.B. (1991): Optimal Inference in Cointegrated Systems, *Econometrica*, 59, 283-306.

Sims, C.A. (1980): Macroeconomics and Reality, *Econometrica*, 48, 1-48.

Sims, C.A., J.H. Stock, and M.W. Watson (1990): Inference in Linear Time Series Models with Some Unit Roots, *Econometrica*, 58, 113-144.

Staiger, D., and J.H. Stock (1997): Instrumental Variables Regression with Weak Instruments, *Econometrica*, 65, 557-586.

Vaart, A.W. van der (1998): *Asymptotic Statistics*, Cambridge: Cambridge University Press.

R2

WORKING PAPERS (1)

9923 **Juan Ayuso and Roberto Blanco:** Has financial market integration increased during the nineties?

9924 **Ignacio Fuentes and Teresa Sastre:** Merkers and acquisitions in the Spanish Banking industry: some empirical evidence.

0001 **Georges Siotis:** Market power, total factor productivity growth, and structural change. An illustration for Spain, 1983-1996.

0002 **Alberto Cabrero:** Seasonal adjustment in economic time series: The experience of the Banco de España (with the model-based method).

0003 **Luis Gordo and Pablo Hernández de Cos:** The financing arrangements for the regional (autonomous) governments for the period 1997-2001. (The Spanish original of this publication has the same number.)

0004 **J. Andrés, F. Ballabriga and J. Vallés:** Monetary Policy and Exchange Rate Behavior in the Fiscal Theory of the Price Level.

0005 **Michael Binder, Cheng Hsiao and M. Hashem Pesaran:** Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration.

(1) Previously published Working Papers are listed in the Banco de España publications catalogue.

Queries should be addressed to: Banco de España
Sección de Publicaciones. Negociado de Distribución y Gestión
Telephone: 91 338 5180
Alcalá, 50. 28014 Madrid