Asymptotic distribution of a simple linear estimator for VARMA models in echelon form *

Jean-Marie Dufour † and Tarek Jouini ‡
Université de Montréal

First version: September 2000
Revised: October 2003, March 2004
This version: July 18, 2004
Compiled: July 18, 2004, 11:29pm

This paper is forthcoming in *Statistical Modeling and Analysis for Complex Data Problems*, edited by Pierre Duchesne and Bruno Rémillard, Kluwer, The Netherlands.

---

* The authors thank Denis Pelletier, an anonymous referee, and the Editor Pierre Duchesne for several useful comments. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Institut de Finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca . Web page: http://www.fas.umontreal.ca/SCECO/Dufour

‡ CIRANO, CIREQ, and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL.: 1 (514) 343-6111, ext. 1814; FAX: 1 (514) 343 5831. E-mail: jouinit@magellan.umontreal.ca

ABSTRACT

In this paper, we study the asymptotic distribution of a simple two-stage (Hannan-Rissanen-type) linear estimator for stationary invertible vector autoregressive moving average (VARMA) models in the echelon form representation. General conditions for consistency and asymptotic normality are given. A consistent estimator of the asymptotic covariance matrix of the estimator is also provided, so that tests and confidence intervals can easily be constructed.

Keywords : Time series; VARMA; Stationary; Invertible; Echelon Form; Estimation; Asymptotic normality; Bootstrap; Hannan-Rissanen.

Contents

List of assumptions, propositions and theorems

1. Introduction

Multivariate time series analysis is widely based on vector autoregressive models (VAR), especially in econometric studies [see Lütkepohl (1991, 2001) and Hamilton (1994, Chapter 11)]. One reason for this popularity is that VAR models are easy to estimate and can account for relatively complex dynamic phenomena. On the other hand, very large numbers of parameters are often required to obtain a good fit, and the class of VAR models is not robust to disaggregation: if a vector process satisfies a VAR scheme, its subvectors (such as individual components) do not follow VAR processes. Instead, the subvectors of VAR processes follow vector autoregressive moving average (VARMA) processes. The latter class, indeed, includes VAR models as a special case, and can reproduce in a parsimonious way a much wider class of autocovariance structures. So they can lead to improvements in estimation and forecast precision. Further, VARMA modelling is theoretically consistent, in the sense that the subvectors of a VARMA model also satisfy VARMA schemes (usually of different order). Similarly, the VARMA class of models is not affected by temporal aggregation, while a VAR model may cease to be a VAR after it has been aggregated over time [see Lütkepohl (1987)].

VARMA modelling has been proposed a long time ago [see Hillmer and Tiao (1979), Tiao and Box (1981), Lütkepohl (1991), Boudjellaba, Dufour and Roy (1992, 1994), Reinsel (1997)], but has remained little used in practical work. Although the process of building VARMA models is, in principle, similar to the one associated with univariate ARMA modelling, the difficulties involved are compounded by the multivariate nature of the data.

At the specification level, new identification issues (beyond the possible presence of common factors) arise and must be taken into account to ensure that unique parameter values can be associated with a given autocovariance structure (compatible with a VARMA model); see Hannan (1969, 1970, 1976*b*, 1979), Deistler and Hannan (1981), Hannan and Deistler (1988, Chapter 2), Lütkepohl (1991, Chapter 7) and Reinsel (1997, Chapter 3). An important finding of this work is the importance of the concepts of dynamic dimension and Kronecker indices in the formulation of identifiable VARMA structures. Further, specifying such models involves the selection of several autoregressive and moving average orders: in view of achieving both identifiability and efficiency, it is important that a reasonably parsimonious model be formulated. Several methods for that purpose have been proposed. The main ones include: (1) techniques based on canonical variate analysis [Akaike (1976), Cooper and Wood (1982), Tiao and Tsay (1985, 1989), Tsay (1989*a*)]; (2) methods which specify an echelon form through the estimation of Kronecker indices [Hannan and Kavalieris (1984*b*), Tsay (1989*b*), Nsiri and Roy (1992, 1996), Poskitt (1992), Lütkepohl and Poskitt (1996), Bartel and Lütkepohl (1998)]; (3) scalar-component models [Tiao and Tsay (1989), Tsay (1991)].

At the estimation level, once an identifiable specification has been formulated, the most widely proposed estimation method is maximum likelihood (ML) derived under the assumption of i.i.d. (independent and identically distributed) Gaussian innovations; see Hillmer and Tiao (1979), Tiao and Box (1981), Shea (1989), Mauricio (2002), and the review of Mélard, Roy and Saidi (2002). This is mainly due to the presence of a moving average part in the model, which makes the latter fundamentally nonlinear. For example, in the Gaussian case, maximizing the likelihood function of a VARMA(p, q) model is typically a burdensome numerical exercise, as soon as the model includes a moving average part. Even numerical convergence may be problematic. Note also that, in the case of weak white noise innovations, quasi-maximum likelihood estimates may not be consistent. These problems also show up (at a smaller scale) in the estimation of univariate ARMA models.

From the viewpoint of making VARMA modelling, it appears crucial to have estimation methods that are both quick and simple to implement with standard statistical software, even if this may involve an efficiency cost. Another reason for putting a premium on such estimation methods is that large-sample distributional theory tends to be quite unreliable in high-dimensional dynamic models, so that tests and confidence sets based on asymptotic approximations are also unreliable (for example, the actual size of test procedures may be far larger than their nominal size). This suggests that simulation-based procedures – for example, bootstrap techniques – should be used, but simulation may be impractical if calculation of the estimators involved is difficult or time consuming.

In the case of univariate ARMA models, a relatively simple estimation procedure was originally proposed by Hannan and Rissanen (1982); see also Durbin (1960), Hannan and Kavalieris (1984*a*), Zhao-Guo (1985), Hannan, Kavalieris and Mackisack (1986), Poskitt (1987), Koreisha and Pukkila (1990*a*, 1990*b*, 1995), Pukkila, Koreisha and Kallinen (1990) and Galbraith and Zinde-Walsh (1994, 1997). This approach is based on estimating (by least squares) the innovations of the process through a long autoregression; after that, the lagged innovations are replaced by the corresponding residuals in the ARMA equation, which may then be also estimated by least squares.

Extensions of this method to VARMA models have been studied by Hannan and Kavalieris (1984*b*, 1986), Hannan and Deistler (1988), Koreisha and Pukkila (1989), Huang and Guo (1990), Poskitt (1992), Poskitt and Lütkepohl (1995), Lütkepohl and Poskitt (1996), Lütkepohl and Claessen (1997) and Flores de Frutos and Serrano (2002). Work on VARMA estimation has focused on preliminary use of such linear estimators for model selection purposes. It is then suggested that other estimation procedures (such as ML) be used. Although consistency is proved, the asymptotic distribution of the basic two-step estimator has not apparently been supplied.

In this paper, we consider the problem of estimating the parameters of stationary VARMA models in echelon form using only linear least squares methods. The echelon form is selected because it tends to deliver relatively parsimonious parameterizations. In particular, we study a simple two-step estimator that can be implemented only through single equation linear regressions and thus is remarkably simple to apply. Such an estimator was previously considered in the above mentioned work on linear VARMA estimation, but its asymptotic distribution has not apparently been established. Given the Kronecker indices of the VARMA process, we derive the asymptotic distribution of this estimator under standard regularity conditions. In particular, we show that the latter has an asymptotic normal distribution (which entails its consistency), and we provide a simple consistent estimator for its asymptotic covariance matrix, so that asymptotically valid tests and confidence tests can be built for the parameters of the model.

The paper is organized as follows. In section 2, we formulate the background model, where the echelon form VARMA representation is considered to ensure unique parametrization, and we define the assumptions which will be used in the rest of the paper. The two-step linear estimation procedure studied in the paper is described in section 3, and we derive its asymptotic distribution in section 4. We conclude in section 5. The proofs of the propositions and theorems appear in the Appendix.

2. Framework

In this section, we describe the theoretical framework and the assumptions we will consider in the sequel. We will first define the standard VARMA representation. As the latter may involve identification problems, we will then define the echelon form on the VARMA model, which ensures uniqueness of model parameters. Finally, we shall formulate the basic regularity assumptions we shall consider.

2.1. Standard form

A k-dimensional regular vector process $\{Y_t : t \in \mathbb{Z}\}$ has a VARMA(p, q) representation if it satisfies an equation of the form:

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + u_t + \sum_{j=1}^{q} B_j u_{t-j} \, , \tag{2.1}$$

for all t, where $Y_t = (Y_{1,t}, \ldots , Y_{k,t})'$, p and q are non-negative integers (respectively, the autoregressive and moving average orders), A_i and B_j the $k \times k$ coefficient matrices, and $\{u_t : t \in \mathbb{Z}\}$ is a (second order) white noise $WN[0, \varSigma_u]$, where $\varSigma_u$ is a $k \times k$ positive definite symmetric matrix. Under the stationary and invertibility conditions the coefficients A_i and B_j satisfy the constraints

$$\det \{A(z)\} \neq 0 \text{ and } \det \{B(z)\} \neq 0 \text{ for all } |z| \leq 1 \tag{2.2}$$

where z is a complex number, $A(z) = I_k - \sum_{i=1}^{p} A_i z^i$ and $B(z) = I_k + \sum_{j=1}^{q} B_j z^j$. This process has the following autoregressive and moving average representations:

$$Y_t = \sum_{\tau=1}^{\infty} \varPi_\tau Y_{t-\tau} + u_t \, , \tag{2.3}$$

$$Y_t = u_t + \sum_{\tau=1}^{\infty} \varPsi_\tau u_{t-\tau}, \; t = 1, \ldots , T \, , \tag{2.4}$$

where

$$\varPi(z) = B(z)^{-1} A(z) = I_k - \sum_{\tau=1}^{\infty} \varPi_\tau z^\tau \, , \tag{2.5}$$

$$\varPsi(z) = A(z)^{-1} B(z) = I_k + \sum_{\tau=1}^{\infty} \varPsi_\tau z^\tau \, , \tag{2.6}$$

$$\det\{\varPi(z)\} \neq 0 \text{ and } \det\{\varPsi(z)\} \neq 0 \, , \text{ for all } |z| \leq 1 \, . \tag{2.7}$$

Note also that we can find real constants $C > 0$ and $\rho \in (0, 1)$ such that

$$\|\varPi_\tau\| \leq C\rho^\tau \text{ and } \|\varPsi_\tau\| \leq C\rho^\tau \, , \tag{2.8}$$

hence

$$\sum_{\tau=1}^{\infty}\|\varPi_{\tau}\|<\infty\,,\quad \sum_{\tau=1}^{\infty}\|\varPsi_{\tau}\|<\infty\,, \tag{2.9}$$

where $\|.\|$ is the Schur norm for a matrix [see Horn and Johnson (1985, section 5.6)], *i.e.*

$$\|M\|^{2}=\operatorname{tr}\left(M'M\right)\,. \tag{2.10}$$

2.2. Echelon form

It is well known that the standard VARMA(p,q) representation given by (2.1) is not unique, in the sense that different sets of coefficients A_i and B_j may represent the same autocovariance structure. To ensure a unique parameterization, we shall consider the stationary invertible VARMA(p,q) process in echelon form representation. Such a representation can be defined as follows:

$$\varPhi\left(L\right)Y_t=\varTheta\left(L\right)u_t\,, \tag{2.11}$$

$$\varPhi\left(L\right)=\varPhi_0-\sum_{i=1}^{\bar{p}}\varPhi_iL^i\,,\quad \varTheta\left(L\right)=\varTheta_0+\sum_{j=1}^{\bar{p}}\varTheta_jL^j\,, \tag{2.12}$$

where L denotes the lag operator, $\varPhi_i=\left[\phi_{lm,i}\right]_{l,m=1,\ldots,k}$ and $\varTheta_j=\left[\theta_{lm,j}\right]_{l,m=1,\ldots,k}$, $\bar{p}=\max\left(p,q\right)$, $\varTheta_0=\varPhi_0$, and $\varPhi_0$ is a lower-triangular matrix whose diagonal elements are all equal to one. The VARMA representation (2.11) has an echelon form if $\varPhi\left(L\right)=\left[\phi_{lm}\left(L\right)\right]_{l,m=1,\ldots,k}$ and $\varTheta\left(L\right)=\left[\theta_{lm}\left(L\right)\right]_{l,m=1,\ldots,k}$ satisfy the following conditions: given a vector of orders $(p_1,\ldots,p_k)$ called the *Kronecker indices*, the operators $\phi_{lm}\left(L\right)$ and $\theta_{lm}\left(L\right)$ on any given row l of $\varPhi\left(L\right)$ and $\varTheta\left(L\right)$ have the same degree p_l $(1\leq l\leq k)$ and

$$\begin{aligned}\phi_{lm}\left(L\right)&=1-\sum_{i=1}^{p_l}\phi_{ll,i}L^i &&\text{if } l=m\,,\\ &=-\sum_{i=p_l-p_{lm}+1}^{p_l}\phi_{lm,i}L^i &&\text{if } l\neq m\,,\end{aligned} \tag{2.13}$$

$$\theta_{lm}\left(L\right)=\sum_{j=0}^{p_l}\theta_{lm,j}L^j\quad\text{with } \varTheta_0=\varPhi_0\,, \tag{2.14}$$

for $l,m=1,\ldots,k$, where

$$\begin{aligned}p_{lm}&=\min\left(p_l+1,\,p_m\right) &&\text{for } l\geq m\,,\\ &=\min\left(p_l,\,p_m\right) &&\text{for } l<m\,.\end{aligned} \tag{2.15}$$

Clearly, $p_{ll}=p_l$ is the order of the polynomial (*i.e.*, the number of free coefficients) on the l-th diagonal element of $\varPhi\left(L\right)$ as well as the order of the polynomials on the corresponding row of $\varTheta\left(L\right)$, while p_{lm} specifies the number of free coefficients in the operator $\phi_{lm}\left(L\right)$ for $l\neq m$. The sum of the Kronecker indices $\sum_{l=1}^{k}p_l$ is called the McMillan degree. The P matrix formed by the Kronecker indices associated with the model is $P = [p_{lm}]_{l,m=1,\dots,k}$. This leads to $\sum_{l=1}^{k}\sum_{m=1}^{k} p_{lm}$ autoregressive and $k\sum_{l=1}^{k} p_l$ moving average free coefficients, respectively. Obviously, for the VARMA orders we have $\bar{p} = \max(p_1, \dots, p_k)$. Note that this identified parameterization for VARMA(p, q) models ensures the uniqueness of left-coprime operators $\Phi(L)$ and $\Theta(L)$. Although other identifiable parameterizations could be used – such as the final equations form – the echelon form tends to be more parsimonious and can lead to efficiency gains. For proofs of the uniqueness of the echelon form and for other identification conditions, the reader should consult to Hannan (1969, 1970, 1976*a*, 1979), Deistler and Hannan (1981), Hannan and Deistler (1988) and Lütkepohl (1991, Chapter 7).

The stationarity and invertibility conditions for echelon form of (2.11) are the same as usual, namely

$$\det\{\Phi(z)\} \neq 0 \quad \text{for all } |z| \leq 1\,, \tag{2.16}$$

for stationarity, and

$$\det\{\Theta(z)\} \neq 0 \quad \text{for all } |z| \leq 1\,, \tag{2.17}$$

for invertibility, where

$$\Phi(z) = \Phi_0 - \sum_{i=1}^{\bar{p}} \Phi_i z^i\,, \quad \Theta(z) = \Theta_0 + \sum_{j=1}^{\bar{p}} \Theta_j z^j\,, \tag{2.18}$$

with $\Pi(z) = \Theta(z)^{-1}\Phi(z)$ and $\Psi(z) = \Phi(z)^{-1}\Theta(z)$. It will be useful to observe that (2.11) can be rewritten in the following form:

$$Y_t = (I_k - \Phi_0)V_t + \sum_{i=1}^{\bar{p}} \Phi_i Y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j} + u_t \tag{2.19}$$

where

$$V_t = Y_t - u_t = \Phi_0^{-1}\Big[\sum_{i=1}^{\bar{p}} \Phi_i Y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j}\Big]\,. \tag{2.20}$$

Note that V_t is a function of *lagged* values of Y_t and u_t, so that the error term u_t in (2.19) is uncorrelated with all the other variables on the right-hand side of the equation.

Set

$$X_t = \big[V_t', Y_{t-1}', \dots, Y_{t-\bar{p}}', u_{t-1}', \dots, u_{t-\bar{p}}'\big]'\,, \tag{2.21}$$

$$D = [I_k - \Phi_0, \Phi_1, \dots, \Phi_{\bar{p}}, \Theta_1, \dots, \Theta_{\bar{p}}]'\,. \tag{2.22}$$

The vector X_t has dimension $(kh)\times 1$ where $h = 2\bar{p}+1$ while D is a $(kh)\times k$ matrix of coefficients. In view of (2.20), it is clear the covariance matrix of X_t is singular, so it is crucial that (identifying) restrictions be imposed on model coefficients. Under the restrictions of the echelon form (2.12) - (2.15), we can find a unique $(k^2h)\times\nu$ full rank matrix R such that $\beta = R\eta$, where η is a $\nu\times 1$

vector of free coefficients and $\nu < k^2h$. Thus Y_t in (2.19) can be expressed as

$$Y_t = D'X_t + u_t = \left(I_k \otimes X_t'\right) R\eta + u_t \,. \tag{2.23}$$

The structure of R is such that

$$\beta = \mathrm{vec}(D) = R\eta \,, \tag{2.24}$$

$$R = \mathrm{diag}(R_1, \dots, R_k) = \begin{bmatrix} R_1 & 0 & \cdots & 0 \\ 0 & R_2 & \cdots & \vdots \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & R_k \end{bmatrix}, \tag{2.25}$$

where R_i, $i = 1, 2, \dots, k$, are $(kh) \times \nu_i$ full-rank selection (zero-one) matrices, each one of which selects the non-zero elements of the corresponding equation, and ν_i is the number of freely varying coefficients present in the i-th equation. The structure of R_i is such that $R_i'R_i = I_{\nu_i}$ and $\beta_i = R_i\eta_i$ where β_i and η_i are respectively a $(kh) \times 1$ and $\nu_i \times 1$ vectors so that β_i is the unconstrained parameter vector in the i-th equation of (2.19) – on which zero restrictions are imposed – and η_i is the corresponding vector of free parameters:

$$\beta = \left(\beta_1', \beta_2', \dots, \beta_k'\right)', \quad \eta = \left(\eta_1', \eta_2', \dots, \eta_k'\right)'. \tag{2.26}$$

Note also that successful identification entails that

$$\mathrm{rank}\left\{\mathsf{E}\left[R'\left(I_k \otimes X_t\right)\left(I_k \otimes X_t'\right)R\right]\right\} = \mathrm{rank}\left\{R'\left(I_k \otimes \Gamma\right)R\right\} = \nu \tag{2.27}$$

where $\Gamma = \mathsf{E}(X_tX_t')$, or equivalently

$$\mathrm{rank}\left\{\mathsf{E}\left[R_i'X_tX_t'R_i\right]\right\} = \mathrm{rank}\left\{R_i'\Gamma R_i\right\} = \nu_i \,, \quad i = 1, \dots, k \,. \tag{2.28}$$

Setting

$$\begin{aligned} X(T) &= [X_1, \dots, X_T]' \,, & (2.29)\\ Y(T) &= [Y_1, \dots, Y_T]' = [y_1(T), \dots, y_k(T)], & (2.30)\\ U(T) &= [u_1, \dots, u_T]' = [U_1(T), \dots, U_k(T)] \,, & (2.31)\\ y(T) &= \mathrm{vec}[Y(T)] \,, \quad u(T) = \mathrm{vec}[U(T)] \,, & (2.32) \end{aligned}$$

(2.23) can be put in any one of the two following matrix forms:

$$\begin{aligned} Y(T) &= X(T)D + U(T) \,, & (2.33)\\ y(T) &= [I_k \otimes X(T)]\, R\eta + u(T) \,, & (2.34) \end{aligned}$$

where $[I_k \otimes X(T)]\, R$ is a $(kT) \times \nu$ matrix. In the sequel, we shall assume that

$$\operatorname{rank}\left([I_k \otimes X(T)]\, R\right) = \nu \text{ with probability 1.} \tag{2.35}$$

Under the assumption that the process is a regular process with continuous distribution, it is easy that the latter must hold.

To see better how the echelon restrictions should be written, consider the following VARMA$(2,1)$ model in echelon form:

$$\begin{aligned} Y_{1,t} &= \phi_{11,1} Y_{1,t-1} + \phi_{11,2} Y_{1,t-2} + u_{1,t}\,, && (2.36)\\ Y_{2,t} &= \phi_{21,0}\left(Y_{1,t} - u_{1,t}\right) + \phi_{21,1} Y_{1,t-1} + \phi_{22,1} Y_{2,t-1} + \theta_{22,1} u_{2,t-1} + u_{2,t}\,. && (2.37) \end{aligned}$$

In this case, we have:

$$\begin{aligned} \varPhi(L) &= \begin{bmatrix} 1 - \phi_{11,1} L - \phi_{11.2} L^2 & -\phi_{12,2} L^2 \\ -\phi_{21,0} - \phi_{21,1} L & 1 - \phi_{22,1} L \end{bmatrix}, && (2.38)\\ \varTheta(L) &= \begin{bmatrix} 1 + \theta_{11,1} L + \theta_{11,2} L^2 & \theta_{12,1} L + \theta_{12,2} L^2 \\ \theta_{21,1} L & 1 + \theta_{22,1} L \end{bmatrix}, && (2.39) \end{aligned}$$

with $\phi_{12,2} = 0$, $\theta_{11,1} = 0$, $\theta_{11,2} = 0$, $\theta_{12,1} = 0$, $\theta_{12,2} = 0$, $\theta_{21,1} = 0$, so that the Kronecker indices are $p_1 = p_{11} = 2$, $p_2 = p_{22} = 1$, $p_{21} = 2$ and $p_{12} = 1$. Setting $X_t = \left[V_t', Y_{t-1}', Y_{t-2}', u_{t-1}'\right]'$, $V_t = (V_{1,t}, V_{2,t})'$, $V_{1,t} = (Y_{1,t} - u_{1,t})$ and $V_{2,t} = (Y_{2,t} - u_{2,t})$, we can then write:

$$\begin{aligned} \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ \phi_{21,0} & 0 \end{bmatrix} \begin{bmatrix} V_{1,t} \\ V_{2,t} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & 0 \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} \\ &\quad + \begin{bmatrix} \phi_{11,2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{1,t-2} \\ Y_{2,t-2} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \theta_{22,1} \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}. \end{aligned} \tag{2.40}$$

Here we have:

$$\beta = \left(0,\, 0,\, \phi_{11,1},\, 0,\, \phi_{11,2},\, 0,\, 0,\, 0,\, \phi_{21,0},\, 0,\, \phi_{21,1},\, \phi_{22,1},\, 0,\, 0,\, 0,\, \theta_{22,1}\right)', \tag{2.41}$$

$$\eta = \left(\phi_{11,1},\, \phi_{11,2},\, \phi_{21,0},\, \phi_{21,1},\, \phi_{22,1},\, \theta_{22,1}\right)', \tag{2.42}$$

$$\left[I_k \otimes X_t'\right] R = \begin{bmatrix} Y_{1,t-1} & Y_{1,t-2} & 0 & 0 & 0 & 0 \\ 0 & 0 & V_{1,t} & Y_{1,t-1} & Y_{2,t-1} & u_{2,t-1} \end{bmatrix}, \tag{2.43}$$

and

$$
[I_k \otimes X(T)]\, R = \begin{bmatrix}
Y_{1,0} & Y_{1,-1} & 0 & 0 & 0 & 0 \\
0 & 0 & V_{1,1} & Y_{1,0} & Y_{2,0} & u_{2,0} \\
Y_{1,1} & Y_{1,0} & 0 & 0 & 0 & 0 \\
0 & 0 & V_{1,2} & Y_{1,1} & Y_{2,1} & u_{2,1} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
Y_{1,T-1} & Y_{1,T-2} & 0 & 0 & 0 & 0 \\
0 & 0 & V_{1,T} & Y_{1,T-1} & Y_{2,T-1} & u_{2,T-1}
\end{bmatrix}. \tag{2.44}
$$

The appropriate matrix R is given by:

$$
R' = \begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}. \tag{2.45}
$$

2.3. Regularity assumptions

In order to establish the asymptotic distribution of the linear estimator defined below, we will need further assumptions on the innovation process and the truncation lag of the first step autoregression. We now state the assumptions we shall consider.

Assumption 2.1 STRONG WHITE NOISE INNOVATIONS. *The vectors u_t, $t \in \mathbb{Z}$, are independent and identically distributed $(i.i.d.)$ with mean zero, covariance matrix Σ_u and continuous distribution.*

Assumption 2.2 UNIFORM BOUNDEDNESS OF FOURTH MOMENTS. *There is a finite constant m_4 such that, for all $1 \leq i,\ j,\ r,\ s \leq k$ and for all t,*

$$
\mathsf{E}\,|u_{it}u_{jt}u_{rt}u_{st}| \leq m_4 < \infty\,.
$$

Assumption 2.3 AUTOREGRESSIVE TRUNCATION LAG OF ORDER LESS THAN $T^{1/2}$. *n_T is a function of T such that*

$$
n_T \to \infty \text{ and } n_T^2/T \to 0 \text{ as } T \to \infty \tag{2.46}
$$

and, for some $c > 0$ and $0 < \bar{\delta} < 1/2$,

$$
n_T \geq cT^{\bar{\delta}} \text{ for } T \text{ sufficiently large.} \tag{2.47}
$$

Assumption 2.4 DECAY RATE OF TRUNCATED AUTOREGRESSIVE COEFFICIENTS. *The coefficients of the autoregressive* (2.3) *representation*

$$n_T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T \to \infty\,. \tag{2.48}$$

Assumption **2.1** means that we have a strong VARMA process, while Assumption **2.2** on moments of order four will ensure the empirical autocovariances of the process have finite variances. Assumption **2.3** implies that n_T goes to infinity at a rate slower than $T^{1/2}$; for example, the assumption is satisfied if $n_T = cT^{\delta}$ with $0 < \bar{\delta} \leq \delta < 1/2$. Assumption **2.4** characterizes the rate of decay of autoregressive coefficients in relation with n_T.

Although the above assumptions are sufficient to show consistency of the two-stage linear estimator, we will need another assumption to show that the asymptotic distribution is normal with a distribution which is unaffected by the use of estimated innovations.

Assumption 2.5 AUTOREGRESSIVE TRUNCATION LAG OF ORDER LESS THAN $T^{1/4}$. n_T *is a function of* T *such that*

$$n_T \to \infty \text{ and } n_T^4/T \to 0 \text{ as } T \to \infty\,. \tag{2.49}$$

The latter assumption means that n_T goes to infinity at a rate slower than $T^{1/4}$; for example, it is satisfied if $n_T = cT^{\delta}$ with $0 < \bar{\delta} \leq \delta < 1/4$. It is easy to see that the condition (2.49) entails (2.46). Finally, it is worthwhile to note that (2.48) holds for VARMA processes whenever $n_T = cT^{\delta}$ with $c > 0$ and $\delta > 0$, *i.e.*

$$T^{\delta} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T \to \infty\,, \quad \text{for all } \delta > 0\,. \tag{2.50}$$

This is easy to see from the exponential decay property of VARMA processes [see (2.8)].

3. Two-step linear estimation

In this section, we describe a simple estimation procedure for a VARMA models in echelon form with known order. The Kronecker indices characterizing the echelon form VARMA model are taken as given, and we focus our attention on the estimation of the autoregressive and moving average coefficients.

Let $(Y_{-n_T+1}, \ldots, Y_T)$ be a random sample of size $T + n_T$, where n_T goes to infinity as T goes to infinity. We consider first a "long" multivariate linear vector autoregression:

$$Y_t = \sum_{\tau=1}^{n_T} \Pi_\tau Y_{t-\tau} + u_t(n_T)\,, \quad t = 1, \ldots, T\,, \tag{3.1}$$

and the corresponding least squares estimates:

$$\tilde{\Pi}\,(n_T) = \left[\tilde{\Pi}_1(n_T),\, \dots\,,\, \tilde{\Pi}_{n_T}(n_T)\right]. \tag{3.2}$$

Such an estimation can be performed by running k separate univariate linear regressions (one for each variable in Y_t). Yule-Walker estimates of the corresponding theoretical coefficients Π_τ could also be considered. Then, under model (2.3) and the assumptions **2.1** to **2.4**, it follows from the results of Paparoditis (1996, Theorem 2.1) and Lewis and Reinsel (1985, proof of Theorem 1) that:

$$\|\tilde{\Pi}\,(n_T) - \Pi\,(n_T)\,\| = O_p(n_T^{1/2}/T^{1/2}) \tag{3.3}$$

where

$$\Pi\,(n_T) = \left[\Pi_1,\, \dots\,,\, \Pi_{n_T}\right]. \tag{3.4}$$

As usual, for any sequence of random variables Z_T and positive numbers r_T, $T = 1, 2, \dots$, the notation $Z_T = O_p(r_T)$ means that Z_T/r_T is asymptotically bounded in probability (as $T \to \infty$), while $Z_T = o_p(r_T)$ means that Z_T/r_T converges to zero in probability. When Y_t satisfies a VARMA scheme, the assumptions **2.3** and **2.4** are satisfied by any truncation lag of the form $n_T = cT^\delta$ with $c > 0$ and $0 < \delta < 1/2$. If, furthermore, the assumptions **2.3** and **2.4** are replaced by stronger ones, namely

$$n_T \to \infty \text{ and } n_T^3/T \to 0 \text{ as } T \to \infty\,, \tag{3.5}$$

$$T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T \to \infty\,, \tag{3.6}$$

then asymptotic normality also holds:

$$T^{1/2}\, l\,(n_T)'\left[\tilde{\pi}\,(n_T) - \pi\,(n_T)\,\right] \underset{T\to\infty}{\longrightarrow} \mathrm{N}\left[0,\, l\,(n_T)'\, Q(n_T) l\,(n_T)\right]\,, \tag{3.7}$$

where $l\,(n_T)$ is a sequence of $k^2 n_T \times 1$ vectors such that $0 < M_1 \le \|l\,(n_T)\| \le M_2 < \infty$ for $n_T = 1, 2, \dots$, and

$$\tilde{\pi}\,(n_T) - \pi\,(n_T) = \mathrm{vec}\left[\tilde{\Pi}\,(n_T) - \Pi\,(n_T)\right]\,, \tag{3.8}$$

$$Q(n_T) = \Gamma(n_T)^{-1} \otimes \Sigma_u\,,\; \Gamma(n_T) = \mathsf{E}[Y_t(n_T) Y_t(n_T)']\,, \tag{3.9}$$

$$Y_t(n_T) = \left[Y'_{t-1},\, Y'_{t-2},\, \dots\,,\, Y'_{t-n_T}\right]'. \tag{3.10}$$

Note that a possible choice for the sequence n_T that satisfies both $n_T^3/T \to 0$ and $T^{1/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\| \to 0$ is for example $n_T = T^{1/\varepsilon}$ with $\varepsilon > 3$. On the other hand $n_T = \ln(\ln T)$, as suggested by Hannan and Kavalieris (1984*b*), is not a permissible choice because in general $T^{1/2}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|$ does not approach zero as $T \to \infty$.

Let

$$\tilde{u}_t(n_T) = Y_t - \sum_{\tau=1}^{n_T} \tilde{\Pi}_\tau(n_T) Y_{t-\tau} = Y_t - \tilde{\Pi}\,(n_T)\, Y_t(n_T) \tag{3.11}$$

be the estimated residuals obtained from the first stage estimation procedure,

$$\tilde{\Sigma}_u(n_T) = \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t(n_T)\tilde{u}_t(n_T)' \tag{3.12}$$

the corresponding estimator of the innovation covariance matrix, and

$$\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^{T} u_t u_t' \tag{3.13}$$

the covariance "estimator" based on the true innovations. Then, we have the following equivalences and convergences.

Proposition 3.1 INNOVATION COVARIANCE ESTIMATOR CONSISTENCY. *Let* $\{Y_t : t \in \mathbb{Z}\}$ *be a k-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15). *Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$\Big\|\frac{1}{T} \sum_{t=1}^{T} u_t[\tilde{u}_t(n_T) - u_t]'\Big\| = O_p(\frac{n_T}{T})\,, \tag{3.14}$$

$$\frac{1}{T} \sum_{t=1}^{T} \|\tilde{u}_t(n_T) - u_t\|^2 = O_p\left(\frac{n_T^2}{T}\right)\,, \tag{3.15}$$

$$\Big\|\frac{1}{T} \sum_{t=1}^{T} [\tilde{u}_t(n_T) - u_t][\tilde{u}_t(n_T) - u_t]'\Big\| = O_p\left(\frac{n_T^2}{T}\right)\,, \tag{3.16}$$

$$\|\tilde{\Sigma}_u(n_T) - \hat{\Sigma}_T\| = O_p\left(\frac{n_T^2}{T}\right)\,, \quad \|\tilde{\Sigma}_u(n_T) - \Sigma_u\| = O_p\left(\frac{n_T^2}{T}\right)\,. \tag{3.17}$$

The asymptotic equivalence between $\tilde{u}_t(n_T)$ and u_t stated in the above proposition suggests we may be able to consistently estimate the parameters of the VARMA model in (2.19) after replacing the unobserved lagged innovations $u_{t-1}, \ldots, u_{t-\bar{p}}$ with the corresponding residuals $\tilde{u}_{t-1}(n_T), \ldots, \tilde{u}_{t-\bar{p}}(n_T)$ from the above long autoregression. So, in order to estimate the coefficients Φ_i and Θ_j of the VARMA process, we consider a linear regression of the form

$$Y_t = \sum_{i=1}^{\bar{p}} \Phi_i Y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j \tilde{u}_{t-j}(n_T) + e_t(n_T) \tag{3.18}$$

imposing the (exclusion) restrictions associated with the echelon form. Setting

$$\tilde{V}_t(n_T) = Y_t - \tilde{u}_t(n_T)\,, \tag{3.19}$$

this regression can also be put in a regression form similar to (2.19):

$$Y_t = (I_k - \Phi_0)\,\tilde{V}_t(n_T) + \sum_{i=1}^{\bar{p}} \Phi_i Y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j \tilde{u}_{t-j}(n_T) + e_t(n_T) \tag{3.20}$$

where

$$e_t(n_T) = \tilde{u}_t(n_T) + \sum_{j=0}^{\bar{p}} \Theta_j [u_{t-j} - \tilde{u}_{t-j}(n_T)]\,. \tag{3.21}$$

Note that (3.20) can be written as

$$Y_t = \big[I_k \otimes \tilde{X}_t(n_T)'\big]R\eta + e_t(n_T)\,, \quad t = 1, \ldots, T\,, \tag{3.22}$$

where

$$\tilde{X}_t(n_T) = \big[\tilde{V}_t(n_T)',\, Y'_{t-1}, \ldots, Y'_{t-\bar{p}},\, \tilde{u}_{t-1}(n_T)', \ldots, \tilde{u}_{t-\bar{p}}(n_T)'\big]'\,. \tag{3.23}$$

Therefore the second step estimators $\tilde{\eta}$ can be obtained by running least squares on the equations (3.22). Setting

$$\tilde{X}(n_T) = \big[\tilde{X}_1(n_T), \tilde{X}_2(n_T), \ldots, \tilde{X}_T(n_T)\big]' \tag{3.24}$$

we get, after some manipulations,

$$\begin{aligned} \tilde{\eta} &= \{R'\big[I_k \otimes \tilde{X}(n_T)'\tilde{X}(n_T)\,\big]R\}^{-1} R'[I_k \otimes \tilde{X}(n_T)']y(T) \\ &= \big(\tilde{\eta}'_1, \tilde{\eta}'_2, \ldots, \tilde{\eta}'_k\big)' \end{aligned} \tag{3.25}$$

where

$$\tilde{\eta}_i = [R'_i \tilde{X}(n_T)'\tilde{X}(n_T)R_i]^{-1} R'_i \tilde{X}(n_T)' y_i(T)\,. \tag{3.26}$$

$\tilde{\eta}$ can be easily obtained by stacking the single equation LS estimators $\tilde{\eta}_i$ which are obtained by regressing y_i on $\tilde{X}(n_T)R_i$.

4. Asymptotic distribution

We will now study the asymptotic distribution of the linear estimator described in the previous section. For that purpose, we note first that the estimator $\tilde{\eta}$ in (3.25) can be expressed as

$$\tilde{\eta} = \{R'[I_k \otimes \tilde{\Gamma}(n_T)]\}R\}^{-1}\big\{\frac{1}{T}\sum_{t=1}^{T} R'[I_k \otimes \tilde{X}_t(n_T)]Y_t\big\} \tag{4.1}$$

where

$$\tilde{\Gamma}(n_T) = \frac{1}{T}\sum_{t=1}^{T} \tilde{X}_t(n_T)\tilde{X}_t(n_T)' \,. \tag{4.2}$$

Let also

$$\begin{aligned}
\tilde{\Upsilon}(n_T) &= I_k \otimes \tilde{\Gamma}(n_T)\,, \quad \tilde{Q}(n_T) = [R'\tilde{\Upsilon}(n_T)R]^{-1}\,, & (4.3)\\
\tilde{\Omega}(n_T) &= \frac{1}{T}\sum_{t=1}^{T} R'[I_k \otimes \tilde{X}_t(n_T)]e_t(n_T)\,. & (4.4)
\end{aligned}$$

It is then easy to see that

$$\tilde{\eta} - \eta = \tilde{Q}(n_T)\tilde{\Omega}(n_T) \tag{4.5}$$

hence

$$\|\tilde{\eta} - \eta\| \leq \|\tilde{Q}(n_T)\|_1 \|\tilde{\Omega}(n_T)\| \leq \|\tilde{Q}(n_T)\| \|\tilde{\Omega}(n_T)\| \tag{4.6}$$

where $\|A\|_1 = \sup_{x \neq 0} \left\{ \frac{\|Ax\|}{\|x\|} \right\}$ stands for the largest eigenvalue of $A'A$ and we used the inequality $\|AB\|^2 \leq \|A\|_1^2 \|B\|^2$ for any two conformable matrices A and B [see Horn and Johnson (1985, section 5.6)].

Define

$$\begin{aligned}
\Gamma &= \mathsf{E}\left(X_t X_t'\right)\,, \quad \Upsilon = I_k \otimes \Gamma\,, \quad Q = (R'\Upsilon R)^{-1}\,, & (4.7)\\
\Gamma_T &= \frac{1}{T}\sum_{t=1}^{T} X_t X_t'\,, \quad \Upsilon_T = I_k \otimes \Gamma_T = \frac{1}{T}\sum_{t=1}^{T} I_k \otimes X_t X_t'\,, & (4.8)\\
Q_T &= (R'\Upsilon_T R)^{-1}\,, \quad \Omega_T = \frac{1}{T}\sum_{t=1}^{T} R'(I_k \otimes X_t)u_t\,. & (4.9)
\end{aligned}$$

Note that $R'\Upsilon R$ is positive definite by the regularity assumption. To study the convergence and distributional properties of $\tilde{\eta} - \eta$, we need first to establish the following proposition.

Proposition 4.1 *Let $\{Y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15). *Then, under the assumptions* **2.1** *to* **2.4***, we have the following equivalences:*

$$\frac{1}{T}\|\tilde{X}(n_T) - X(T)\|^2 = O_p\left(\frac{n_T^2}{T}\right)\,, \tag{4.10}$$

$$\|\tilde{\Gamma}(n_T) - \Gamma_T\| = O_p\left(\frac{n_T}{T^{1/2}}\right)\,, \tag{4.11}$$

$$\|\tilde{\Upsilon}(n_T) - \Upsilon_T\| = O_p\left(\frac{n_T}{T^{1/2}}\right)\,, \tag{4.12}$$

$$\|\tilde{Q}(n_T)^{-1} - Q^{-1}\| = O_p\left(\frac{n_T}{T^{1/2}}\right)\,, \tag{4.13}$$

$$\|\tilde{Q}(n_T) - Q\| = O_p\left(\frac{n_T}{T^{1/2}}\right). \tag{4.14}$$

The latter proposition shows that the matrices $\tilde{\Gamma}(n_T)$, $\tilde{\Upsilon}(n_T)$, $\tilde{Q}(n_T)^{-1}$ and $\tilde{Q}(n_T)$ – based on approximate innovations (estimated from a long autoregression) – are all asymptotically equivalent to the corresponding matrices based on true innovations, according to the rate $n_T/T^{1/2}$. Similarly the norm of the difference between the approximate regressor matrix $\tilde{X}(n_T)$ and $X(T)$ has order $O_p(n_T/T^{1/2})$. This suggests that $\tilde{\eta}$ converges to η, and we give the appropriate rate of convergence in the following theorem.

Theorem 4.1 CONSISTENCY OF SECOND STEP HR ESTIMATES. *Let $\{Y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15). *Then, under the assumptions* **2.1** *to* **2.4**, *we have*

$$\|\Omega_T\| = O_p\left(\frac{1}{T^{1/2}}\right), \quad \|\tilde{\Omega}(n_T) - \Omega_T\| = O_p\left(\frac{n_T^2}{T}\right), \tag{4.15}$$

$$\|\tilde{\eta} - \eta\| = O_p\left(\frac{1}{T^{1/2}}\right) + O_p\left(\frac{n_T^2}{T}\right). \tag{4.16}$$

If, furthermore,

$$n_T^4/T \to 0 \text{ as } T \to \infty, \tag{4.17}$$

then

$$\|\tilde{\eta} - \eta\| = O_p\left(\frac{1}{T^{1/2}}\right). \tag{4.18}$$

The latter theorem shows that $\tilde{\eta}$ is a consistent estimator. If furthermore, $n_T^4/T \to 0$ as $T \to \infty$, then $\tilde{\eta}$ converges at the rate $T^{-1/2}$ which is typically expected to get asymptotic normality. In order to derive an asymptotic distribution for $\tilde{\eta}$, we shall establish that the following random matrices

$$\tilde{S}(n_T) = T^{1/2}\tilde{Q}(n_T)\tilde{\Omega}(n_T), \quad S_T = T^{1/2}Q\Omega_T, \tag{4.19}$$

are asymptotically equivalent.

Proposition 4.2 ASYMPTOTIC EQUIVALENCE. *Let $\{Y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA echelon representation given by* (2.11) - (2.15). *Then, under the assumptions* **2.1** *to* **2.4**, *the following equivalence holds*

$$\|\tilde{S}(n_T) - S_T\| = O_p\left(\frac{n_T^2}{T^{1/2}}\right).$$

Finally, we can give the asymptotic distribution of $\sqrt{T}\,(\tilde{\eta} - \eta)$.

Theorem 4.3 ASYMPTOTIC DISTRIBUTION OF TWO-STAGE ESTIMATOR. *Let $\{Y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA echelon representation* *given by* (2.11) - (2.15). *If the assumptions* **2.1** *to* **2.5** *are satisfied, then the asymptotic distribution of the estimator* $\tilde{\eta}$ *is the following:*

$$\sqrt{T}\big(\tilde{\eta}-\eta\big) \underset{T\to\infty}{\longrightarrow} \mathrm{N}[0,\, \varSigma_{\eta}]$$

where

$$\varSigma_{\eta} = Q\varSigma_{Xu}Q' , \quad \varSigma_{Xu} = R'\left[\varSigma_{u}\otimes\varGamma\right]R\,, \tag{4.20}$$

$$Q = (R'\varUpsilon R)^{-1} , \quad \varUpsilon = I_{k}\otimes\varGamma\,, \quad \varGamma = \mathsf{E}\left(X_{t}X_{t}'\right)\,, \tag{4.21}$$

$X_{t} = \left[V_{t}', Y_{t-1}',\, \ldots\,, Y_{t-\bar{p}}', u_{t-1}',\, \ldots\,,\, u_{t-\bar{p}}'\right]'$ *and* $V_{t} = Y_{t} - u_{t}$.

An important consequence of the above theorem is the fact that the asymptotic distribution of $\tilde{\eta}$ is the same as in the case where the innovations $u_{t-1}',\, \ldots\,,\, u_{t-\bar{p}}'$ are known rather than approximated by a long autoregression. Furthermore, the covariance matrix $\varSigma_{\eta}$ can be consistently estimated by

$$\hat{\varSigma}_{\eta} = \tilde{Q}(n_{T})\{R'[\tilde{\varSigma}_{u}(n_{T})\otimes\tilde{\varGamma}(n_{T})]R\}\tilde{Q}(n_{T})'\,, \tag{4.22}$$

where

$$\tilde{Q}(n_{T}) = [R'\tilde{\varUpsilon}(n_{T})R]^{-1}, \quad \tilde{\varUpsilon}(n_{T}) = I_{k}\otimes\tilde{\varGamma}(n_{T})\,, \tag{4.23}$$

$$\tilde{\varGamma}(n_{T}) = \frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{t}(n_{T})\tilde{X}_{t}(n_{T})'. \tag{4.24}$$

Standard t and F-type tests may then be performed in the usual way.

5. Conclusion

In this paper, we have provided the asymptotic distribution of a simple two-stage estimator for VARMA models in echelon form. The estimator is consistent when the auxiliary long autoregression used to generate first step estimates of model innovations has an order n_{T} which increases to infinity at a rate inferior to T^{δ} with $0 < \delta_{0} \leq \delta < 1/2$. Further, it has an asymptotic normal distribution provided n_{T} increases at a rate inferior to T^{δ} with $0 < \delta_{0} \leq \delta < 1/4$. In the latter case, the asymptotic distribution is not affected by the fact that estimated lagged residuals are used.

The above results can be exploited in several ways. First, the two-stage estimates and the associated distributional theory can be directly used for inference on the VARMA model. In particular, they can be used for model selection purposes and to simplify the model (e.g., by eliminating insignificant coefficients). Second, two-stage estimates can be exploited to get more efficient estimators, such as ML estimators or estimators that are asymptotically to ML. This can be done, in particular, to achieve efficiency with Gaussian innovations. Note, however, that such gains of efficiency may not obtain if the innovations are not Gaussian. Thirdly, because of its simplicity, the two-stage linear estimator is especially well adapted for being used in the context of simulation-based inference procedures, such as bootstrap tests. Further, the asymptotic distribution provided above can be useful in order to improve the validity of the bootstrap. Several of these issues will be studied in a subsequent paper.

A. Appendix: Proofs

PROOF OF PROPOSITION **3.1** Let us write:

$$\|\tilde{\Sigma}_u(n_T) - \Sigma_u\| = \|\tilde{\Sigma}_u(n_T) - \hat{\Sigma}_T\| + \|\hat{\Sigma}_T - \Sigma_u\| \tag{A.1}$$

where

$$\hat{\Sigma}_T - \Sigma_u = \frac{1}{T}\sum_{t=1}^{T}[u_t u_t' - \Sigma_u] \,, \tag{A.2}$$

$$\begin{aligned}
\tilde{\Sigma}_u(n_T) - \hat{\Sigma}_T &= \frac{1}{T}\sum_{t=1}^{T}\left\{\tilde{u}_t(n_T)\tilde{u}_t(n_T)' - u_t u_t'\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{[\tilde{u}_t(n_T) - u_t]\tilde{u}_t(n_T)^{'} + u_t[\tilde{u}_t(n_T) - u_t]^{'}\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{[\tilde{u}_t(n_T) - u_t]u_t^{'} + u_t[\tilde{u}_t(n_T) - u_t]^{'} + [\tilde{u}_t(n_T) - u_t][\tilde{u}_t(n_T) - u_t]^{'}\right\} .
\end{aligned} \tag{A.3}$$

By the assumptions **2.1** and **2.2**,

$$\hat{\Sigma}_T - \Sigma_u = \frac{1}{T}\sum_{t=1}^{T}[u_t u_t' - \Sigma_u] = O_p\left(\frac{1}{T}\right) , \tag{A.4}$$

$$\frac{1}{T}\sum_{t=1}^{T}\|u_t\| = O_p(1) \,, \quad \frac{1}{T}\sum_{t=1}^{T}\|u_t\|^2 = O_p(1) \,. \tag{A.5}$$

Now

$$\tilde{u}_t(n_T) - u_t = [\Pi\,(n_T) - \tilde{\Pi}\,(n_T)]Y_t(n_T) + \sum_{\tau=n_T+1}^{\infty}\Pi_\tau Y_{t-\tau} \,, \tag{A.6}$$

hence

$$\frac{1}{T}\sum_{t=1}^{T}[\tilde{u}_t(n_T) - u_t]u_t^{'} = [\Pi\,(n_T) - \tilde{\Pi}\,(n_T)]C_{Yu}(n_T) + S_{Yu}(n_T) \tag{A.7}$$

where $Y_t(n_T) = \left[Y_{t-1}', \ldots, Y_{t-n_T}'\right]'$, and

$$C_{Yu}(n_T) = \frac{1}{T}\sum_{t=1}^{T}Y_t(n_T)u_t^{'} = [C_{Yu}(1, T)', \ldots, C_{Yu}(n_T, T)']' \,, \tag{A.8}$$

$$C_{Yu}(\tau, T) = \frac{1}{T}\sum_{t=1}^{T}Y_{t-\tau}u_t^{'} \,, \tag{A.9}$$

$$S_{Yu}(n_T) \;=\; \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty} \Pi_\tau Y_{t-\tau}u_t^{'} . \tag{A.10}$$

Using the fact that u_t is independent of $X_t,\, u_{t-1},\, \ldots\, ,\, u_1$, we see that

$$\begin{aligned}
\mathsf{E}\|C_{Yu}(\tau,\, T)\|^2 &= \mathsf{E}[C_{Yu}(\tau,\, T)C_{Yu}(\tau,\, T)'] = \frac{1}{T^2}\sum_{t=1}^{T}\mathsf{E}[\mathrm{tr}(Y_{t-\tau}u_t^{'}u_tY_{t-\tau}')] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}[\mathsf{E}(u_t^{'}u_t)\mathsf{E}(Y_{t-\tau}'Y_{t-\tau})] = \frac{1}{T}\mathrm{tr}(\Sigma_u)\mathrm{tr}[\Gamma(0)]\,, && \text{(A.11)}\\
\mathsf{E}[S_{Yu}(n_T)] &= 0\,, && \text{(A.12)}
\end{aligned}$$

where $\Gamma(0) = \mathsf{E}(Y_tY_t')$, hence

$$\begin{aligned}
\mathsf{E}\|C_{Yu}(n_T)\|^2 &= \mathsf{E}[C_{Yu}(n_T)'C_{Yu}(n_T)] = \sum_{\tau=1}^{n_T}\mathsf{E}\|C_{Yu}(\tau,\, T)\|^2 \\
&= \frac{n_T}{T}\mathrm{tr}(\Sigma_u)\mathrm{tr}[\Gamma(0)]\,, && \text{(A.13)}\\
\sum_{\tau=1}^{n_T}\|C_{Yu}(\tau,\, T)\|^2 &= O_p\left(\frac{n_T}{T}\right)\,, && \text{(A.14)}
\end{aligned}$$

and

$$\|[\tilde{\Pi}\,(n_T) - \Pi\,(n_T)]C_{Yu}(n_T)\| \le \|\tilde{\Pi}\,(n_T) - \Pi\,(n_T)\,\|\|C_{Yu}(n_T)\| = O_p\left(\frac{n_T}{T}\right)\,. \tag{A.15}$$

Using the stationarity of Y_t and (2.8), we have:

$$\begin{aligned}
\mathsf{E}\left[\|S_{Yu}(n_T)\|\right] &\le \mathsf{E}\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\,\|Y_{t-\tau}\|\,\|u_t\|\Big)\Big] \\
&\le \big[\mathsf{E}\big(\|Y_t\|^2\big)\big]^{1/2}\big[\mathsf{E}\big(\|u_t\|^2\big)\big]^{1/2}\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\| \\
&\le \big[\mathsf{E}\big(\|Y_t\|^2\big)\big]^{1/2}\big[\mathsf{E}\big(\|u_t\|^2\big)\big]^{1/2}\frac{C}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\rho^\tau \\
&\le \big[\mathsf{E}\big(\|Y_t\|^2\big)\big]^{1/2}\big[\mathsf{E}\big(\|u_t\|^2\big)\big]^{1/2}\frac{C}{T}\sum_{t=1}^{T}\frac{\rho^{n_T+1}}{1-\rho} \\
&= \big[\mathsf{E}\big(\|Y_t\|^2\big)\big]^{1/2}\big[\mathsf{E}\big(\|u_t\|^2\big)\big]^{1/2}\left(\frac{C\,\rho}{1-\rho}\right)\rho^{n_T} = O(\rho^{n_T}) && \text{(A.16)}
\end{aligned}$$

hence

$$\big\|S_{Yu}(n_T)\big\| = O_p(\rho^{n_T})\,. \tag{A.17}$$

Consequently,

$$\begin{aligned}\Big\|\frac{1}{T}\sum_{t=1}^{T} u_t[\tilde{u}_t(n_T)-u_t]'\Big\| &= \Big\|\frac{1}{T}\sum_{t=1}^{T}[\tilde{u}_t(n_T)-u_t]u_t'\Big\| \\ &\leq \|[\tilde{\mathit{\Pi}}\,(n_T)-\mathit{\Pi}\,(n_T)]C_{Yu}(n_T)\| + \big\|S_{Yu}(n_T)\big\| \\ &= O_p\left(\frac{n_T}{T}\right), \end{aligned} \tag{A.18}$$

and (3.14) is established. Finally,

$$\begin{aligned}\Big\|\frac{1}{T}\sum_{t=1}^{T}[\tilde{u}_t(n_T)-u_t][\tilde{u}_t(n_T)-u_t]'\Big\| &\leq \frac{1}{T}\sum_{t=1}^{T}\big\|[\tilde{u}_t(n_T)-u_t][\tilde{u}_t(n_T)-u_t]'\big\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\big\|\tilde{u}_t(n_T)-u_t\big\|^2 \end{aligned} \tag{A.19}$$

where

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\|\tilde{u}_t(n_T)-u_t\|^2 &\leq \frac{3}{T}\sum_{t=1}^{T}\Big\{\|\tilde{\mathit{\Pi}}\,(n_T)-\mathit{\Pi}\,(n_T)\,\|^2\,\|Y_t(n_T)\|^2 \\ &\quad +\Big(\sum_{\tau=n_T+1}^{\infty}\|\mathit{\Pi}_\tau\|\,\|Y_{t-\tau}\|\Big)^2\Big\} \\ &\leq 3\,\|\tilde{\mathit{\Pi}}\,(n_T)-\mathit{\Pi}\,(n_T)\,\|^2\frac{1}{T}\sum_{t=1}^{T}\|Y_t(n_T)\|^2 \\ &\quad +\frac{3}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\mathit{\Pi}_\tau\|\,\|Y_{t-\tau}\|\Big)^2. \end{aligned} \tag{A.20}$$

Since

$$\mathsf{E}\Big[\frac{1}{T}\sum_{t=1}^{T}\|Y_t(n_T)\|^2\Big] = \mathsf{E}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{n_T}\|Y_{t-\tau}\|^2\Big] = n_T\mathsf{E}\big(\,\|Y_t\|^2\,\big)\,, \tag{A.21}$$

we have

$$\frac{1}{T}\sum_{t=1}^{T}\|Y_t(n_T)\|^2 = O_p(n_T)\,. \tag{A.22}$$

Further,

$$\begin{aligned}
\mathsf{E}\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\,\|Y_{t-\tau}\|\Big)\Big] &= \mathsf{E}\,\|Y_t\|\,\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\| \\
&\le \mathsf{E}\,\|Y_t\|\,\frac{C}{T}\sum_{t=1}^{T}\frac{\rho^{n_T+1}}{1-\rho} = \left(\frac{C\,\mathsf{E}\,\|Y_t\|\,\rho}{1-\rho}\right)\rho^{n_T} \\
&= O(\rho^{n_T})\,, \qquad \text{(A.23)}
\end{aligned}$$

hence

$$\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\,\|Y_{t-\tau}\|\Big) = O_p(\rho^{n_T})\,, \qquad \text{(A.24)}$$

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\,\|Y_{t-\tau}\|\Big)^2 &\le T\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\,\|Y_{t-\tau}\|\Big)\Big]^2 \\
&= O_p(T\rho^{2n_T})\,. \qquad \text{(A.25)}
\end{aligned}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\big\|\tilde{u}_t(n_T)-u_t\big\|^2 \le O_p\left(\frac{n_T}{T}\right)O_p(n_T)+O_p(T\rho^{2n_T}) = O_p\left(\frac{n_T^2}{T}\right), \qquad \text{(A.26)}$$

$$\big\|\frac{1}{T}\sum_{t=1}^{T}[\tilde{u}_t(n_T)-u_t][\tilde{u}_t(n_T)-u_t]'\big\| = O_p\left(\frac{n_T^2}{T}\right). \qquad \text{(A.27)}$$

We can thus conclude that

$$\|\tilde{\Sigma}_u(n_T)-\hat{\Sigma}_T\| = O_p(\frac{n_T}{T})+O_p\left(\frac{n_T^2}{T}\right) = O_p\left(\frac{n_T^2}{T}\right), \qquad \text{(A.28)}$$

$$\|\tilde{\Sigma}_u(n_T)-\Sigma_u\| = O_p\left(\frac{n_T^2}{T}\right). \qquad \text{(A.29)}$$

PROOF OF PROPOSITION **4.1** Using (4.2) and (4.8), we see that

$$\begin{aligned}
\tilde{\Gamma}(n_T)-\Gamma_T &= \frac{1}{T}\sum_{t=1}^{T}\big[\tilde{X}_t(n_T)\tilde{X}_t(n_T)'-X_tX_t'\big] \\
&= \frac{1}{T}\sum_{t=1}^{T}\big\{[\tilde{X}_t(n_T)-X_t]X_t'+X_t[\tilde{X}_t(n_T)-X_t]'\big\} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T}\big\{[\tilde{X}_t(n_T)-X_t][\tilde{X}_t(n_T)-X_t]'\big\} \qquad \text{(A.30)}
\end{aligned}$$

hence, using the triangular and Cauchy-Schwarz inequalities,

$$\begin{aligned}
\|\tilde{\mathit{\Gamma}}(n_T) - \mathit{\Gamma}_T\| &\leq 2\Big(\frac{1}{T}\sum_{t=1}^{T}\|X_t\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T}\|\tilde{X}_t(n_T) - X_t\|^2\Big)^{1/2} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\|\tilde{X}_t(n_T) - X_t\|^2 \\
&= 2\Big(\frac{1}{T}\|X(T)\|^2\Big)^{1/2}\Big(\frac{1}{T}\|\tilde{X}(n_T) - X(T)\|^2\Big)^{1/2} \\
&\quad + \frac{1}{T}\|\tilde{X}(n_T) - X(T)\|^2 \qquad \text{(A.31)}
\end{aligned}$$

where

$$\tilde{X}_t\left(n_T\right) - X_t = \begin{bmatrix} u_t - \tilde{u}_t\left(n_T\right) \\ 0 \\ \vdots \\ 0 \\ \tilde{u}_{t-1}\left(n_T\right) - u_{t-1} \\ \vdots \\ \tilde{u}_{t-\tilde{p}}\left(n_T\right) - u_{t-\tilde{p}} \end{bmatrix}, \qquad \text{(A.32)}$$

$$\begin{aligned}
\frac{1}{T}\|\tilde{X}(n_T) - X(T)\|^2 &= \frac{1}{T}\sum_{t=1}^{T}\|\tilde{X}_t(n_T) - X_t\|^2 \\
&= \sum_{j=0}^{\bar{p}}\Big[\frac{1}{T}\sum_{t=1}^{T}\|\tilde{u}_{t-j}(n_T) - u_{t-j}\|^2\Big] = O_p\left(\frac{n_T^2}{T}\right) \qquad \text{(A.33)}
\end{aligned}$$

and, by the stationarity assumption,

$$\frac{1}{T}\|X(T)\|^2 = \frac{1}{T}\sum_{t=1}^{T}\|X_t\|^2 = O_p\left(1\right). \qquad \text{(A.34)}$$

It follows from the above orders that

$$\|\tilde{\mathit{\Gamma}}(n_T) - \mathit{\Gamma}_T\| = O_p\left(\frac{n_T}{T^{1/2}}\right). \qquad \text{(A.35)}$$

Consequently, we have:

$$\begin{aligned}
\|\tilde{\mathit{\Upsilon}}(n_T) - \mathit{\Upsilon}_T\| &= \|I_k \otimes \tilde{\mathit{\Gamma}}(n_T) - I_k \otimes \mathit{\Gamma}_T\| \\
&= \|I_k \otimes \big(\tilde{\mathit{\Gamma}}(n_T) - \mathit{\Gamma}_T\big)\| \\
&= k^{1/2}\|\tilde{\mathit{\Gamma}}(n_T) - \mathit{\Gamma}_T\| = O_p\left(\frac{n_T}{T^{1/2}}\right), \qquad \text{(A.36)}
\end{aligned}$$

$$
\begin{aligned}
\|\tilde{Q}(n_T)^{-1} - Q_T^{-1}\| &= \|R'\big[\tilde{\Upsilon}(n_T) - \Upsilon_T\big]R\| \\
&\leq \|R\|^2\|\tilde{\Upsilon}(n_T) - \Upsilon_T\| = O_p\left(\frac{n_T}{T^{1/2}}\right) . \qquad (A.37)
\end{aligned}
$$

Further, since

$$
\|\tilde{Q}(n_T)^{-1} - Q^{-1}\| \leq \|\tilde{Q}(n_T)^{-1} - Q_T^{-1}\| + \|Q_T^{-1} - Q^{-1}\| \qquad (A.38)
$$

and

$$
\begin{aligned}
\|Q_T^{-1} - Q^{-1}\| &= \left\|R'\left(\Upsilon_T - \Upsilon\right)R\right\| \leq \|R\|^2\|\Upsilon_T - \Upsilon\| \\
&\leq \|R\|^2\|I_k \otimes (\Gamma_T - \Gamma)\| = k^{1/2}\|R\|^2\|\Gamma_T - \Gamma\| \\
&= k^{1/2}\|R\|^2\left\|\frac{1}{T}\sum_{t=1}^{T} X_tX_t' - \mathsf{E}\left(X_tX_t'\right)\right\| = O_p\left(\frac{1}{T^{1/2}}\right) , \qquad (A.39)
\end{aligned}
$$

we have:

$$
\|\tilde{Q}(n_T)^{-1} - Q^{-1}\| = O_p\left(\frac{n_T}{T^{1/2}}\right) . \qquad (A.40)
$$

Finally, using the triangular inequality, we get:

$$
\|\tilde{Q}(n_T)\| \leq \|\tilde{Q}(n_T) - Q\| + \|Q\| , \qquad (A.41)
$$

$$
\begin{aligned}
\|\tilde{Q}(n_T) - Q\| &= \|\tilde{Q}(n_T)\big[\tilde{Q}(n_T)^{-1} - Q^{-1}\big]Q\| \\
&\leq \|\tilde{Q}(n_T)\|\|\tilde{Q}(n_T)^{-1} - Q^{-1}\|\|Q\| \\
&\leq \Big[\|\tilde{Q}(n_T) - Q\| + \|Q\|\Big]\|\tilde{Q}(n_T)^{-1} - Q^{-1}\|\|Q\| , \qquad (A.42)
\end{aligned}
$$

hence, for $\|\tilde{Q}(n_T)^{-1} - Q^{-1}\|\|Q\| < 1$ (an event whose probability converges to 1 as $T \to \infty$)

$$
\|\tilde{Q}(n_T) - Q\| \leq \frac{\|Q\|^2\|\tilde{Q}(n_T)^{-1} - Q^{-1}\|}{1 - \|\tilde{Q}(n_T)^{-1} - Q^{-1}\|\|Q\|} = O_p\left(\frac{n_T}{T^{1/2}}\right) . \qquad (A.43)
$$

□

PROOF OF THEOREM **4.1** Recall that $\tilde{\eta} - \eta = \tilde{Q}(n_T)\tilde{\Omega}(n_T)$. Then, we have

$$
\begin{aligned}
\|\tilde{\eta} - \eta\| &\leq \|Q\|_1\|\Omega_T\| + \|\tilde{Q}(n_T) - Q\|_1\|\Omega_T\| + \|\tilde{Q}(n_T)\|_1\|\tilde{\Omega}(n_T) - \Omega_T\| \\
&\leq \|Q\|\|\Omega_T\| + \|\tilde{Q}(n_T) - Q\|\|\Omega_T\| + \|\tilde{Q}(n_T)\|\|\tilde{\Omega}(n_T) - \Omega_T\| . \qquad (A.44)
\end{aligned}
$$

By Proposition **4.1**,

$$
\|\tilde{Q}(n_T) - Q\| = O_p\left(\frac{n_T}{T^{1/2}}\right) , \quad \|\tilde{Q}(n_T)\| = O_p(1) . \qquad (A.45)
$$

Now

$$\Omega_T = \frac{1}{T}\sum_{t=1}^{T} R'\left[I_k \otimes X_t\right] u_t = R'\mathrm{vec}\Big[\frac{1}{T}\sum_{t=1}^{T} X_t u_t'\Big]\,, \tag{A.46}$$

so that

$$\mathsf{E}\left\|\Omega_T\right\|^2 \le \left\|R\right\|^2 \mathsf{E}\|W_T\|^2 \tag{A.47}$$

where

$$W_T = \frac{1}{T}\sum_{t=1}^{T} X_t u_t'\,. \tag{A.48}$$

Then, using the fact that u_t is independent of $X_t,\, u_{t-1},\, \ldots\,,\, u_1,$

$$\begin{aligned}
\mathsf{E}\left\|W_T\right\|^2 &= \mathsf{E}[\mathrm{tr}(W_T W_T')] \\
&= \frac{1}{T^2}\Big\{\sum_{t=1}^{T}\mathsf{E}\left(\mathrm{tr}\left[X_t u_t' u_t X_t'\right]\right) + 2\sum_{t=1}^{T-1}\sum_{l=1}^{T-l}\{\mathsf{E}\left(\mathrm{tr}\left[X_t u_t' u_{t+l} X_{t+l}'\right]\right)\Big\} \\
&= \frac{1}{T^2}\Big\{\sum_{t=1}^{T}\mathsf{E}\left(\mathrm{tr}\left[u_t' u_t X_t' X_t\right]\right) + 2\sum_{t=1}^{T-1}\sum_{l=1}^{T-l}\{\mathsf{E}\left(\mathrm{tr}\left[u_{t+l} X_{t+l}' X_t u_t'\right]\right)\Big\} \\
&= \frac{1}{T^2}\Big\{\sum_{t=1}^{T}\mathrm{tr}\big[\mathsf{E}(u_t' u_t)\mathsf{E}(X_t' X_t)\big] + 2\sum_{t=1}^{T-1}\sum_{l=1}^{T-l}\{\mathsf{E}\left(\mathrm{tr}\left[\mathsf{E}(u_{t+l})\mathsf{E}(X_{t+l}' X_t u_t')\right]\right)\Big\} \\
&= \frac{1}{T^2}\Big\{\sum_{t=1}^{T}\mathrm{tr}\left[\mathsf{E}\left(u_t u_t'\right)\mathsf{E}\left(X_t' X_t\right)\right]\Big\} = \frac{1}{T}\mathrm{tr}(\Sigma_u)\mathrm{tr}(\Gamma)
\end{aligned} \tag{A.49}$$

hence

$$\|W_T\| = O_p\big(T^{-1/2}\big)\,, \quad \|\Omega_T\| = O_p\big(T^{-1/2}\big)\,. \tag{A.50}$$

Now, consider the term $\|\tilde{\Omega}(n_T) - \Omega_T\|$. We have:

$$\begin{aligned}
\tilde{\Omega}(n_T) - \Omega_T &= \frac{1}{T}R'\sum_{t=1}^{T}\big\{\big[I_k \otimes \tilde{X}_t(n_T)\big]e_t(n_T) - \big[I_k \otimes X_t\big]u_t\big\} \\
&= R'\mathrm{vec}\Big[\frac{1}{T}\sum_{t=1}^{T}\big\{\tilde{X}_t(n_T)e_t(n_T)' - X_t u_t'\big\}\Big] \\
&= R'\mathrm{vec}\big\{\tilde{\Omega}_1(n_T) + \tilde{\Omega}_2(n_T)\big\}
\end{aligned} \tag{A.51}$$

where

$$\tilde{\Omega}_1(n_T) = \frac{1}{T}\sum_{t=1}^{T} X_t\left[e_t(n_T) - u_t\right]'\,, \tag{A.52}$$

$$\tilde{\Omega}_2(n_T) = \frac{1}{T}\sum_{t=1}^{T}\big[\tilde{X}_t(n_T) - X_t\big]e_t(n_T)', \tag{A.53}$$

$$e_t\,(n_T) = \tilde{u}_t\,(n_T) + \sum_{j=0}^{\bar{p}}\Theta_j\,[u_{t-j} - \tilde{u}_{t-j}\,(n_T)]\ . \tag{A.54}$$

We can also write

$$e_t\,(n_T) - u_t = \sum_{j=0}^{\bar{p}}\bar{\Theta}_j\,[\tilde{u}_{t-j}\,(n_T) - u_{t-j}] \tag{A.55}$$

where $\bar{\Theta}_0 = I_k - \Theta_0$ and $\bar{\Theta}_j = -\Theta_j$, $j = 1, 2,\ \ldots\ , \bar{p}$, and

$$\begin{aligned}\tilde{u}_t\,(n_T) - u_t &= \big[\Pi\,(n_T) - \tilde{\Pi}\,(n_T)\big]Y_t\,(n_T) + \sum_{\tau=n_T+1}^{\infty}\Pi_\tau Y_{t-\tau} \\ &= \sum_{\tau=1}^{n_T}\big[\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big]Y_{t-\tau} + \sum_{\tau=n_T+1}^{\infty}\Pi_\tau Y_{t-\tau}\,, \end{aligned}\tag{A.56}$$

hence

$$\begin{aligned}\tilde{\Omega}_1(n_T) &= \frac{1}{T}\sum_{t=1}^{T}X_t\,[e_t(n_T) - u_t]' \\ &= \sum_{j=0}^{\bar{p}}\Big\{\frac{1}{T}\sum_{t=1}^{T}\Big\{\sum_{\tau=1}^{n_T}X_tY'_{t-j-\tau}\big[\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big]' + \sum_{\tau=n_T+1}^{\infty}X_tY'_{t-j-\tau}\Pi'_\tau\Big\}\Big\}\bar{\Theta}'_j \\ &= \sum_{j=0}^{\bar{p}}\Big\{\sum_{\tau=1}^{n_T}\Big\{\frac{1}{T}\sum_{t=1}^{T}X_tY'_{t-j-\tau}\Big\}\big[\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big]' + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}X_tY'_{t-j-\tau}\Pi'_\tau\Big\}\bar{\Theta}'_j \\ &= \tilde{\Omega}_{11}(n_T) + \tilde{\Omega}_{12}(n_T)\end{aligned}\tag{A.57}$$

where

$$\tilde{\Omega}_{11}(n_T) = \sum_{j=0}^{\bar{p}}\Big\{\sum_{\tau=1}^{n_T}\tilde{\Gamma}_{j+\tau}\,(n_T)\,\big[\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big]'\Big\}\bar{\Theta}'_j\,, \tag{A.58}$$

$$\tilde{\Gamma}_{j+\tau}\,(n_T) = \frac{1}{T}\sum_{t=1}^{T}X_tY'_{t-j-\tau}\,, \tag{A.59}$$

$$\tilde{\Omega}_{12}(n_T) = \sum_{j=0}^{\bar{p}}\Big\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}X_tY'_{t-j-\tau}\Pi'_\tau\Big\}\bar{\Theta}'_j\,. \tag{A.60}$$

Now, using the linearity and the VARMA structure of Y_t, it is easy to see that

$$\mathsf{E}\|\tilde{\Gamma}_{j+\tau}\left(n_T\right)\|^2 \leq \frac{1}{T}C_1\rho_1^{j+\tau} \tag{A.61}$$

for some constants $C_1 > 0$ and $0 < \rho_1 < 1$, hence

$$\mathsf{E}\Big[\sum_{\tau=1}^{n_T}\|\tilde{\Gamma}_{j+\tau}\left(n_T\right)\|^2\Big] \leq \frac{1}{T}C_1\sum_{\tau=1}^{n_T}\rho_1^{j+\tau} \leq \frac{1}{T}\frac{C_1}{1-\rho_1} = O_p\left(\frac{1}{T}\right). \tag{A.62}$$

Thus

$$\begin{aligned}
\|\tilde{\Omega}_{11}(n_T)\| &\leq \sum_{j=0}^{\bar{p}}\Big\{\sum_{\tau=1}^{n_T}\|\tilde{\Gamma}_{j+\tau}\left(n_T\right)\|\|\Pi_\tau-\tilde{\Pi}_\tau(n_T)\|\Big\}\,\|\bar{\Theta}_j\| \\
&\leq \sum_{j=0}^{\bar{p}}\Big\{\Big[\sum_{\tau=1}^{n_T}\|\tilde{\Gamma}_{j+\tau}\left(n_T\right)\|^2\Big]^{1/2}\Big[\sum_{\tau=1}^{n_T}\|\Pi_\tau-\tilde{\Pi}_\tau(n_T)\|^2\Big]^{1/2}\Big\}\,\|\bar{\Theta}_j\| \\
&\leq \sum_{j=0}^{\bar{p}}\Big\{\Big[\sum_{\tau=1}^{n_T}\|\tilde{\Gamma}_{j+\tau}\left(n_T\right)\|^2\Big]^{1/2}\|\tilde{\Pi}\left(n_T\right)-\Pi\left(n_T\right)\|\Big\}\,\|\bar{\Theta}_j\| \\
&= O_p\left(\frac{n_T^{1/2}}{T}\right),
\end{aligned} \tag{A.63}$$

while

$$\begin{aligned}
\mathsf{E}\|\tilde{\Omega}_{12}(n_T)\| &\leq \sum_{j=0}^{\bar{p}}\Big\{\mathsf{E}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|X_t\|\,\|Y_{t-j-\tau}\|\,\|\Pi_\tau\|\Big]\Big\}\,\|\bar{\Theta}_j\| \\
&\leq \sum_{j=0}^{\bar{p}}\Big\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\,\mathsf{E}\big[\,\|X_t\|\,\|Y_{t-j-\tau}\|\,\big]\Big\}\,\|\bar{\Theta}_j\| \\
&\leq \sum_{j=0}^{\bar{p}}\Big\{\big[\mathsf{E}(\|X_t\|^2)\mathsf{E}(\|Y_t\|^2)\big]^{1/2}\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\Big\}\,\|\bar{\Theta}_j\| \\
&= O_p(\rho^{n_T}),
\end{aligned} \tag{A.64}$$

hence $\|\tilde{\Omega}_{12}(n_T)\| = O_p(\rho^{n_T})$ and

$$\|\tilde{\Omega}_1(n_T)\| \leq \|\tilde{\Omega}_{11}(n_T)\| + \|\tilde{\Omega}_{12}(n_T)\| = O_p\left(\frac{n_T^{1/2}}{T}\right). \tag{A.65}$$

Now, using (A.55), $\tilde{\Omega}_2(n_T)$ can be decomposed as:

$$\tilde{\Omega}_2(n_T) = \tilde{\Omega}_{21}(n_T) + \tilde{\Omega}_{22}(n_T) \tag{A.66}$$

where

$$\tilde{\Omega}_{21}(n_T) = \frac{1}{T}\sum_{t=1}^{T}\big[\tilde{X}_t(n_T) - X_t\big]u_t^{'}\,, \tag{A.67}$$

$$\tilde{\Omega}_{22}(n_T) = \sum_{j=0}^{\bar{p}}\Big\{\frac{1}{T}\sum_{t=1}^{T}\big[\tilde{X}_t(n_T) - X_t\big]\left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\right]'\Big\}\bar{\Theta}_j'\,. \tag{A.68}$$

Now, in view of (A.32), consider the variables:

$$\begin{aligned} C_i(n_T) &= \frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_{t-i}\left(n_T\right) - u_{t-i}\right]u_t^{'} \\ &= \sum_{\tau=1}^{n_T}\big[\Pi_\tau - \tilde{\Pi}_\tau(n_T)\big]\Big(\frac{1}{T}\sum_{t=1}^{T}Y_{t-i-\tau}u_t^{'}\Big) + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\Pi_\tau Y_{t-i-\tau}u_t^{'}\,, \end{aligned} \tag{A.69}$$

$$C_{ij}(n_T) = \frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_{t-i}\left(n_T\right) - u_{t-i}\right]\left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\right]'\,, \tag{A.70}$$

for $i = 0,\, 1,\, \ldots\, ,\, \bar{p}$. We have:

$$\begin{aligned} \mathsf{E}\|\frac{1}{T}\sum_{t=1}^{T}Y_{t-i-\tau}u_t^{'}\|^2 &= \frac{1}{T^2}\sum_{t=1}^{T}\mathsf{E}\mathrm{tr}[Y_{t-i-\tau}u_t^{'}u_tY_{t-i-\tau}'] = \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}[\mathsf{E}(u_t^{'}u_t)\mathsf{E}(Y_{t-i-\tau}'Y_{t-i-\tau})] \\ &= \frac{1}{T}\mathrm{tr}(\Sigma_u)\mathrm{tr}[\Gamma(0)] \end{aligned} \tag{A.71}$$

where $\Gamma(0) = \mathsf{E}(Y_tY_t')$, hence

$$\sum_{\tau=1}^{n_T}\mathsf{E}\|\frac{1}{T}\sum_{t=1}^{T}Y_{t-i-\tau}u_t^{'}\|^2 = \frac{n_T}{T}\mathrm{tr}(\Sigma_u)\mathrm{tr}[\Gamma(0)]\,, \tag{A.72}$$

$$\sum_{\tau=1}^{n_T}\|\frac{1}{T}\sum_{t=1}^{T}Y_{t-i-\tau}u_t^{'}\|^2 = O_p\left(\frac{n_T}{T}\right)\,, \tag{A.73}$$

and

$$\begin{aligned} \|C_i(n_T)\| &\leq \sum_{\tau=1}^{n_T}\|\Pi_\tau - \tilde{\Pi}_\tau(n_T)\|\|\frac{1}{T}\sum_{t=1}^{T}Y_{t-i-\tau}u_t^{'}\| \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\|Y_{t-i-\tau}\|\|u_t\| \end{aligned}$$

$$
\begin{aligned}
&\leq \Big[\sum_{\tau=1}^{n_T} \|\mathit{\Pi}_\tau - \tilde{\mathit{\Pi}}_\tau(n_T)\|^2\Big]^{1/2} \Big[\sum_{\tau=1}^{n_T} \|\frac{1}{T}\sum_{t=1}^{T} Y_{t-i-\tau} u_t^{'}\|^2\Big]^{1/2} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty} \|\mathit{\Pi}_\tau\|\|Y_{t-i-\tau}\|\|u_t\| \\
&= \|\tilde{\mathit{\Pi}}\left(n_T\right) - \mathit{\Pi}\left(n_T\right)\| \Big[\sum_{\tau=1}^{n_T} \|\frac{1}{T}\sum_{t=1}^{T} Y_{t-i-\tau} u_t^{'}\|^2\Big]^{1/2} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty} \|\mathit{\Pi}_\tau\|\|Y_{t-i-\tau}\|\|u_t\| \\
&= O_p\left(\frac{n_T}{T}\right). \qquad \text{(A.74)}
\end{aligned}
$$

Further,

$$
\begin{aligned}
\|C_{ij}(n_T)\| &\leq \frac{1}{T}\sum_{t=1}^{T} \|\left[\tilde{u}_{t-i}\left(n_T\right) - u_{t-i}\right]\|\|\left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\right]'\| \\
&\leq \Big[\frac{1}{T}\sum_{t=1}^{T} \|\tilde{u}_{t-i}\left(n_T\right) - u_{t-i}\|^2\Big]^{1/2} \Big[\frac{1}{T}\sum_{t=1}^{T} \|\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\|^2\Big]^{1/2} \\
&= O_p\left(\frac{n_T^2}{T}\right). \qquad \text{(A.75)}
\end{aligned}
$$

Thus

$$
\|\tilde{\mathit{\Omega}}_{21}(n_T)\| = O_p(n_T/T), \quad \|\tilde{\mathit{\Omega}}_{22}(n_T)\| = O_p\left(\frac{n_T^2}{T}\right), \qquad \text{(A.76)}
$$

hence

$$
\|\tilde{\mathit{\Omega}}_2(n_T)\| \leq \|\tilde{\mathit{\Omega}}_{21}(n_T)\| + \|\tilde{\mathit{\Omega}}_{22}(n_T)\| = O_p\left(\frac{n_T^2}{T}\right), \qquad \text{(A.77)}
$$

$$
\begin{aligned}
\|\tilde{\mathit{\Omega}}(n_T) - \mathit{\Omega}_T\| &\leq \|R\|\big(\|\tilde{\mathit{\Omega}}_1(n_T)\| + \|\tilde{\mathit{\Omega}}_2(n_T)\|\big) \\
&= O_p\left(\frac{n_T^{1/2}}{T}\right) + O_p\left(\frac{n_T^2}{T}\right) = O_p\left(\frac{n_T^2}{T}\right). \qquad \text{(A.78)}
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
\|\tilde{\eta} - \eta\| &\leq O_p\left(\frac{1}{T^{1/2}}\right) + O_p\left(\frac{n_T}{T}\right) + O_p\left(\frac{n_T^2}{T}\right) \\
&= O_p\left(\frac{1}{T^{1/2}}\right) + O_p\left(\frac{n_T^2}{T}\right) = o_p(1). \qquad \text{(A.79)}
\end{aligned}
$$

If furthermore $n_T^4/T \longrightarrow 0$ as $T \to \infty$, the latter reduces to

$$\|\tilde{\eta} - \eta\| = O_p\left(\frac{1}{T^{1/2}}\right). \tag{A.80}$$

□

PROOF OF PROPOSITION **4.2** We have:

$$\begin{aligned} \|\tilde{S}(n_T) - S_T\| &= T^{1/2}\|\tilde{Q}(n_T)\tilde{\Omega}(n_T) - Q\Omega_T\| \\ &\le T^{1/2}\|\tilde{Q}(n_T)\|\|\tilde{\Omega}(n_T) - \Omega_T\| + T^{1/2}\|\tilde{Q}(n_T) - Q\|\|\Omega_T\|. \end{aligned} \tag{A.81}$$

By Proposition **4.1** and Theorem **4.1**, the following orders hold:

$$\|\tilde{Q}(n_T) - Q\| = O_p\left(\frac{n_T}{T^{1/2}}\right), \quad \|\tilde{Q}(n_T)\| = O_p(1), \tag{A.82}$$

$$\|\tilde{\Omega}(n_T) - \Omega_T\| = O_p\left(\frac{n_T^2}{T}\right), \quad \|\Omega_T\| = O_p\left(\frac{1}{T^{1/2}}\right). \tag{A.83}$$

Therefore,

$$\|\tilde{S}(n_T) - S_T\| = O_p\left(\frac{n_T^2}{T^{1/2}}\right). \tag{A.84}$$

□

PROOF OF THEOREM **4.3** By the standard central limit theorem for stationary processes [see Anderson (1971, section 7.7), Lewis and Reinsel (1985, section 2)] and under the assumption of independence between u_t and X_t, we have:

$$T^{1/2}\Omega_T = \frac{1}{T^{1/2}}\sum_{t=1}^{T} R'(I_k \otimes X_t)u_t = \frac{1}{T^{1/2}}\sum_{t=1}^{T} R'(u_t \otimes X_t) \underset{T\to\infty}{\longrightarrow} \mathrm{N}[0,\, \Sigma_{Xu}] \tag{A.85}$$

where

$$\begin{aligned} \Sigma_{Xu} &= \mathsf{E}\left\{R'(u_t \otimes X_t)(u_t \otimes X_t)'R\right\} = \mathsf{E}\left\{R'\left[u_t u_t' \otimes X_t X_t'\right]R\right\} \\ &= R'\left[\mathsf{E}(u_t u_t') \otimes \mathsf{E}(X_t X_t')\right]R = R'\left[\Sigma_u \otimes \Gamma\right]R. \end{aligned} \tag{A.86}$$

Then

$$S_T = T^{1/2}Q\Omega_T \underset{T\to\infty}{\longrightarrow} \mathrm{N}\left[0,\, \Sigma_\eta\right] \tag{A.87}$$

where

$$\Sigma_\eta = Q\Sigma_{Xu}Q'. \tag{A.88}$$

Finally, by Proposition **4.2**, we can conclude that

$$\sqrt{T}\,(\tilde{\eta}-\eta) = \tilde{S}(n_T) \underset{T\to\infty}{\longrightarrow} \mathrm{N}\big[0,\ \Sigma_\eta\big]\,. \tag{A.89}$$

□

References

Akaike, H. (1976), Canonical correlation analysis of stochastic processes and its application to the analysis of autoregressive moving average processes, *in* R. K. Mehra and D. G. Lainiotis, eds, 'System Identification: Advances in Case Studies', Academic Press, New York, pp. 27–96.

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley & Sons, New York.

Bartel, H. and Lütkepohl, H. (1998), 'Estimating the Kronecker indices of cointegrated echelon-form VARMA models', *Econometrics Journal* **1**, C76–C99.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992), 'Testing causality between two vectors in multivariate ARMA models', *Journal of the American Statistical Association* **87**(420), 1082–1090.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1994), 'Simplified conditions for non-causality between two vectors in multivariate ARMA models', *Journal of Econometrics* **63**, 271–287.

Cooper, D. M. and Wood, E. F. (1982), 'Identifying multivariate time series models', *Journal of Time Series Analysis* **3**(3), 153–164.

Deistler, M. and Hannan, E. J. (1981), 'Some properties of the parameterization of ARMA systems with unknown order', *Journal of Multivariate Analysis* **11**, 474–484.

Durbin, J. (1960), 'The fitting of time series models', *Revue de l'Institut International de Statistique* **28**, 233–244.

Flores de Frutos, R. and Serrano, G. R. (2002), 'A generalized least squares estimation method for VARMA models', *Statistics* **36**(4), 303–316.

Galbraith, J. W. and Zinde-Walsh, V. (1994), 'A simple, noniterative estimator for moving average models', *Biometrika* **81**(1), 143–155.

Galbraith, J. W. and Zinde-Walsh, V. (1997), 'On some simple, autoregression-based estimation and identification techniques for ARMA models', *Biometrika* **84**(3), 685–696.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Hannan, E. J. (1969), 'The identification of vector mixed autoregressive- moving average systems', *Biometrika* **57**, 223–225.

Hannan, E. J. (1970), *Multiple Time Series*, John Wiley & Sons, New York.

Hannan, E. J. (1976*a*), 'The asymptotic distribution of serial covariances', *The Annals of Statistics* **4**(2), 396–399.

Hannan, E. J. (1976*b*), 'The identification and parameterization of ARMAX and state space forms', *Econometrica* **44**(4), 713–723.

Hannan, E. J. (1979), The statistical theory of linear systems, *in* P. R. Krishnaiah, ed., 'Developments in Statistics', Vol. 2, Academic Press, New York, pp. 83–121.

Hannan, E. J. and Deistler, M. (1988), *The Statistical Theory of Linear Systems*, John Wiley & Sons, New York.

Hannan, E. J. and Kavalieris, L. (1984*a*), 'A method for autoregressive-moving average estimation', *Biometrika* **71**(2), 273–280.

Hannan, E. J. and Kavalieris, L. (1984*b*), 'Multivariate linear time series models', *Advances in Applied Probability* **16**, 492–561.

Hannan, E. J. and Kavalieris, L. (1986), 'Regression, autoregression models', *Journal of Time Series Analysis* **7**(1), 27–49.

Hannan, E. J., Kavalieris, L. and Mackisack, M. (1986), 'Recursive estimation of linear systems', *Biometrika* **73**(1), 119–133.

Hannan, E. J. and Rissanen, J. (1982), 'Recursive estimation of mixed autoregressive-moving-average order', *Biometrika* **69**(1), 81–94. Errata 70 (1983), 303.

Hillmer, S. C. and Tiao, G. C. (1979), 'Likelihood function of stationary multiple autoregressive moving average models', *Journal of the American Statistical Association* **74**(367), 652–660.

Horn, R. G. and Johnson, C. A. (1985), *Matrix Analysis*, Cambridge University Press, Cambridge, U.K.

Huang, D. and Guo, L. (1990), 'Estimation of nonstationary ARMAX models based on the Hannan-Rissanen method', *The Annals of Statistics* **18**(4), 1729–1756.

Koreisha, S. G. and Pukkila, T. M. (1989), 'Fast linear estimation methods for vector autoregressive moving-average models', *Journal of Time Series Analysis* **10**(4), 325–339.

Koreisha, S. G. and Pukkila, T. M. (1990*a*), 'A generalized least-squares approach for estimation of autoregressive-moving-average models', *Journal of Time Series Analysis* **11**(2), 139–151.

Koreisha, S. G. and Pukkila, T. M. (1990*b*), 'Linear methods for estimating ARMA and regression models with serial correlation', *Communications in Statistics, Part B -Simulation and Computation* **19**(1), 71–102.

Koreisha, S. G. and Pukkila, T. M. (1995), 'A comparison between different order-determination criteria for identification of ARIMA models', *Journal of Business and Economic Statistics* **13**(1), 127–131.

Lewis, R. and Reinsel, G. C. (1985), 'Prediction of multivariate time series by autoregressive model fitting', *Journal of Multivariate Analysis* **16**, 393–411.

Lütkepohl, H. (1987), *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Lütkepohl, H. (2001), Vector autoregressions, *in* B. Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 32, pp. 678–699.

Lütkepohl, H. and Claessen, H. (1997), 'Analysis of cointegrated VARMA processes', *Journal of Econometrics* **80**(2), 223–39.

Lütkepohl, H. and Poskitt, D. S. (1996), 'Specification of echelon-form VARMA models', *Journal of Business and Economic Statistics* **14**(1), 69–79.

Mauricio, J. A. (2002), 'An algorithm for the exact likelihood of a stationary vector autoregressive-moving average model', *Journal of Time Series Analysis* **23**(4), 473–486.

Mélard, G., Roy, R. and Saidi, A. (2002), Exact maximum likelihood estimation of structured or unit roots multivariate time series models, Technical report, Institut de Statistique, Université Libre de Bruxelles, and Départment de mathématiques et statistique, Université de Montréal.

Nsiri, S. and Roy, R. (1992), 'On the identification of ARMA echelon-form models', *Canadian Journal of Statistics* **20**(4), 369–386.

Nsiri, S. and Roy, R. (1996), 'Identification of refined ARMA echelon form models for multivariate time series', *Journal of Multivariate Analysis* **56**, 207–231.

Paparoditis, E. (1996), 'Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes', *Journal of Multivariate Analysis* **57**, 277–296.

Poskitt, D. S. (1987), 'A modified Hannan-Rissanen strategy for mixed autoregressive-moving average oder determination', *Biometrika* **74**(4), 781–790.

Poskitt, D. S. (1992), 'Identification of echelon canonical forms for vector linear processes using least squares', *The Annals of Statistics* **20**(1), 195–215.

Poskitt, D. S. and Lütkepohl, H. (1995), Consistent specification of cointegrated autoregressive moving-average systems, Technical Report 54, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.

Pukkila, T., Koreisha, S. and Kallinen, A. (1990), 'The identification of ARMA models', *Biometrika* **77**(3), 537–548.

Reinsel, G. C. (1997), *Elements of Multivariate Time Series Analysis*, second edn, Springer-Verlag, New York.

Shea, B. L. (1989), 'The exact likelihood of a vector autoregressive moving average model', *Journal of the Royal Statistical Society Series C, Applied Statistics* **38**(1), 161–184.

Tiao, G. C. and Box, G. E. P. (1981), ‘Modeling multiple time series with applications’, *Journal of the American Statistical Association* **76**(376), 802–816.

Tiao, G. C. and Tsay, R. S. (1985), A canonical correlation approach to modeling multivariate time series, *in* ‘Proceedings of the Business and Economic Statistics Section of the American Statistical Association’, Washington, D.C., pp. 112–120.

Tiao, G. C. and Tsay, R. S. (1989), ‘Model specification in multivariate time series’, *Journal of the Royal Statistical Society, Series B* **51**(2), 157–213.

Tsay, R. S. (1989*a*), ‘Identifying multivariate time series models’, *Journal of Time Series Analysis* **10**(4), 357–372.

Tsay, R. S. (1989*b*), ‘Parsimonious parameterization of vector autoregressive moving average models’, *Journal of Business and Economic Statistics* **7**(3), 327–341.

Tsay, R. S. (1991), ‘Two canonical forms for vector VARMA processes’, *Statistica Sinica* **1**, 247–269.

Zhao-Guo, C. (1985), ‘The asymptotic efficiency of a linear procedure of estimation for ARMA models’, *Journal of Time Series Analysis* **6**(1), 53–62.