Asymptotic distributions for some quasi-efficient estimators in echelon-form VARMA models [1]

Jean-Marie Dufour [2] and Tarek Jouini [3]

September 2008
Compiled: August 30, 2010

[1] We thank Vadim Marmer 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 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, and the Fonds FCAR (Government of Québec).

[2] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: 1(514) 398 8879; FAX 1(514) 398 4938; e-mail: jean-marie.dufour@mcgill.ca. Web page: http://www.jeanmariedufour.com

[3] Department of Economics, University of Windsor, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, University of Windsor, 1184 Chrysler Hall North Building, 401 Sunset avenue, Windsor, Ontario N9B 3P4, Canada. TEL: 1(519) 253-3000, Ext. 2381; FAX: 1(519) 973-7096. e-mail: tjouini@uwindsor.ca

ABSTRACT

We study two linear estimators for stationary invertible VARMA models in echelon form (for identification), with known Kronecker indices. Such linear estimators are much simpler to compute than Gaussian maximum likelihood (ML) estimators often proposed for such models, which are highly nonlinear. The first estimator is an improved two-step estimator which can be interpreted as a generalized least squares (GLS) of the two-step least-squares estimator considered in Dufour and Jouini (2005), for a more general model which allows for the presence of drift parameters. The second estimator is a new relatively simple three-step linear estimator which is asymptotically equivalent to ML, hence efficient, when the innovations of the process are Gaussian. The proposed asymptotically efficient estimator is based on using modified approximate residuals which better take into account the truncation error associated with the approximate long autoregression used in the first step of the method. We show that both estimators are consistent and asymptotically normal under the assumption that the innovations are a strong white noise, possibly non-Gaussian. Explicit formulae for the asymptotic covariance matrices are provided. The proposed estimators make it relatively easy to estimate the parameters of VARMA models in echelon form, and the distributional theory does not rely on a Gaussian assumption, like maximum likelihood or the estimators considered by Hannan and Kavalieris (1984*b*) and Reinsel, Basu and Yap (1992). We present simulation evidence which indicate that the proposed three-step estimator typically performs better in finite samples than the alternative multi-step linear estimators suggested by Hannan and Kavalieris (1984*b*), Reinsel et al. (1992), and Poskitt and Salau (1995).

Keywords: Three-Step Linear Estimation; GLS; Three-Step Linear Estimation; Stationary; Invertible; Echelon Form; Kronecker Indices; Nonlinear GLS; Simulation; ML; Asymptotically Efficient.
Journal of Economic Literature Classification: C13, C32.

Contents

1 Introduction

Modelling multivariate time series using vector autoregressive (VAR) models has received considerable attention, especially in econometrics; see Lütkepohl (1991, 2001, 2005), Hamilton (1994, Chapter 11) and Dhrymes (1998). This popularity is due to the fact that such models are easy to estimate and can account for relatively complex dynamic phenomena. However, VAR models often require very large number of parameters in order to obtain good fits. Further, the VAR specification is not invariant to many basic linear transformations. For example, instead of satisfying a VAR scheme, subvectors follow vector autoregressive moving average (VARMA) processes. Temporal and contemporaneous aggregation lead to mixed VARMA models [see Lütkepohl (1987)]. Similarly, trend and seasonal adjustment also lead to models outside the VAR class [Maravall (1993)].

The VARMA structure includes VAR models as a special case and can reproduce in a parsimonious way a much wider class of autocovariances and data generating processes (DGP). Thus, they can yield improvements in estimation and forecasting as has been pointed out in recent works [see Lütkepohl (2006), Izquierdo, Hernández and Del Hoyo (2007), Athanasopoulos and Vahid (2008*b*) and Kascha and Mertens (2008)]. VARMA modelling has been proposed years ago [see Hillmer and Tiao (1979), Tiao and Box (1981), Lütkepohl (1991), Boudjellaba, Dufour and Roy (1992, 1994), Reinsel (1993, 1997)], but it has received little attention in practice. Although building VARMA models remains similar to the procedure associated with the univariate case, the task is compounded by the multivariate nature of the data.

At the specification level, several procedures ensuring a unique parameterization have been proposed; see Hannan (1969*b*, 1970, 1971, 1976, 1979, 1980, 1981), Deistler and Hannan (1981), Deistler (1983), Hannan and Deistler (1988, Chapter 2), Lütkepohl (1991, Chapter 7) and Reinsel (1997, Chapter 3). In view of achieving parsimonious parameterization and efficiency, several methods have been considered. The main ones include: (1) techniques based on canonical analysis [Akaike (1974, 1975, 1976), Cooper and Wood (1982), Tiao and Tsay (1985, 1989), Tsay and Tiao (1985), Tsay (1989*a*), Paparoditis and Streitberg (1991) and Min and Tsay (2005)]; (2) the Kronecker index approach, which specifies an echelon-form VARMA representation [Deistler and Hannan (1981), Hannan and Kavalieris (1984*b*), Solo (1986), Tsay (1989*b*), Nsiri and Roy (1992, 1996), Poskitt (1992, 2003), Lütkepohl and Poskitt (1996) and Bartel and Lütkepohl (1998)]; (3) the scalar-component model (SCM) approach [Tiao and Tsay (1989), Tsay (1989*b*, 1991) and Athanasopoulos and Vahid (2008*a*) in a recent extension of the Tiao and Tsay (1989) method].

Once an identifiable specification has been formulated, different estimation methods have been considered. But the most widely studied estimation method is ML for independent and identically distributed (i.i.d.) Gaussian innovations; see Hannan (1969*a*), Newbold (1974), Box and Jenkins (1976), Hillmer and Tiao (1979), Nicholls and Hall (1979, 1980), Hannan, Kavalieris and Mackisack (1986), Kohn (1981), Tiao and Box (1981), Solo (1984), Shea (1989), Mélard, Roy and Saidi (2002), Mauricio (2002, 2006), Jonasson and Ferrando (2008) and recently Gallego (2009). See also Metaxoglou and Smith (2007) on ML estimation of state space VARMA models using expectation-maximization (EM) algorithm. However, maximizing the exact likelihood in stationary invertible VARMA models is computationally burdensome since for each autoregressive and moving average order (say p and q) a non-quadratic optimization with respect to inequality constraints must be performed using iterative algorithms. As noted by Tiao and Box (1981), it is much easier to maximize a conditional likelihood, although in higher dimensional systems numerical problems still occur due to the lack of suitable initial values even with known (p, q). Further, with weak white noise innovations, quasi-maximum likelihood estimates may not be consistent.

From the viewpoint of making VARMA modelling practical, one should have estimation methods that are both quick and simple to implement with standard software. 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. This suggests that simulation-based procedures—such as bootstrap techniques—should be used. However, simulation may be impractical if computing the estimator is difficult or time consuming.

In the univariate case, Hannan and Rissanen (1982) have proposed a recursive method which only requires linear regressions; see also Durbin (1960), Hannan and Kavalieris (1984*a*), Zhao-Guo (1985), Hannan et al. (1986), Poskitt (1987), Koreisha and Pukkila (1990*a*, 1990*b*, 1995), Pukkila, Koreisha and Kallinen (1990), Allende and Heiler (1992), Galbraith and Zinde-Walsh (1994, 1997) and Kavalieris, Hannan and Salau (2003). This approach is based on estimating (by least squares) the innovations of the process through a long autoregression. The resulting residuals are then used as regressors to estimate the VARMA parameters. Thereafter, new residuals are filtered and a linear regression on transformed variables is performed to achieve efficiency under Gaussian assumptions. Note that this linear estimation method (in its first two steps) has been introduced for model selection and obtaining initial values. Then using other estimation procedures, such as ML, is typically suggested.

These methods have been extended to VARMA models. For stationary processes Hannan and Kavalieris (1984*b*) have proposed four-step linear procedure for specifying and estimating ARMAX systems. The first three steps of their procedure were devoted to specifying and obtaining initial estimates using Toeplitz regressions based on Levinson-Whittle algorithm. However, it has been shown that the latter tend to deliver coefficient estimates suffering from substantial bias, especially when the ratio of the autoregression-order to the sample size is not sufficiently small [see Hannan and Deistler (1988)]. Then, they suggested in their fourth stage a GLS estimation corresponding to what should be an asymptotically efficient estimate of the system parameters. (That is to what would be true under Gaussian errors.) In line with this, Reinsel et al. (1992) analyzed the ML estimation of VARMA models from a GLS viewpoint. In particular, they considered Gaussian innovations without assuming any identification scheme. Further, after ignoring some error term in their linear regression, they have shown that the remaining error term follows a moving average process. Thus showing the equivalence between the resulting GLS and the ML estimation. However, one can see that the computational burden of this method is heavy since the inversion of a high dimensional weighting matrix is frequently involved, even in small and moderate samples dealing with big systems. Also, Poskitt and Salau (1995) have investigated the relationship between the GLS and Gaussian estimation for echelon form VARMA models by extending the three-stage linear estimation method proposed by Koreisha and Pukkila (1990*a*) for the univariate case, to the multivariate framework. Although, asymptotically equivalent to ML, these estimates suffer of substantial finite sample bias due partly to the weighting matrix used in the computation.

Furthermore, De Frutos and Serrano (2002) suggested a new GLS procedure for estimating VARMA models. It explicitly considers the stochastic structure of the approximation error that arises when the innovations are replaced with the residuals obtained from a long VAR. From a simulation study, they have shown that their method outperforms the double regression proposed by Koreisha and Pukkila (1989). However, although consistent, their procedure not only did not consider any form of identification but also found asymptotically inefficient. The same also holds for the iterative ordinary least squares (IOLS) procedure proposed by Kapetanios (2003), even though it has been shown from a simulation study that such a procedure compares favorably with ML method. More recently, Koreisha and Pukkila (2004) have proposed a three-step linear estimation procedure for specifying and estimating VARMA models without assuming any form of identification. While in the first two stages of their method they suggested a new identification approach based on the multivariate version of the residual white noise autoregressive (RWNAR) criterion through a testing procedure (with strong Gaussian innovations), their third-stage estimation procedure relies on the GLS estimation procedure suggested by Poskitt and Salau (1995). Finally, in a comparative simulation study over selected existing linear methods, based on selected criteria such as the quality of the estimates, and the accuracy of derived forecasts and impulse response estimates, Kascha (2007) highlighted the overall superiority of the fourth-stage linear estimation procedure of Hannan and Kavalieris (1984*b*), while noting situations where the investigated methods do not perform very well.

Other linear methods have been limited rather to two-stage LS estimation procedures for identifying and getting preliminary estimates of the VARMA model parameters. Thereafter, recommending the use of fully efficient methods, such as ML, to obtain efficient estimates [see; Koreisha and Pukkila (1987, 1989), Poskitt (1992) and Lütkepohl and Poskitt (1996)]. This strategy also has been extended to cointegrated VARMA systems [see; Huang and Guo (1990), Poskitt (2003), Poskitt and Lütkepohl (1995), Lütkepohl and Claessen (1997) and Bartel and Lütkepohl (1998)]. In particular, for nonstationary ARMAX models, Huang and Guo (1990) have stressed that the estimated residuals obtained from a long-autoregression are still good estimates of the true innovations. They also have shown that the VARMA orders can be consistently estimated using model selection criteria such as Schwarz criterion (SC) and Hannan-Quinn criterion (HQ).

In this paper, we propose a consistent and efficient three-step linear estimation method for stationary invertible VARMA models in echelon form, with known Kronecker indices. Our approach can easily be adapted to VARMAX models and extended to integrated and cointegrated VARMA models as well. The estimation method focuses on the echelon form as the latter tends to deliver relatively parsimonious parameterizations. Further, our estimation method is simple and more general than any other existing procedure, as it yields echelon form estimates with a general standard form much easier to obtain than other existing methods, such as in Hannan and Kavalieris (1984*b*), while remaining valid to other identifying procedures, such as final equations [see Dufour and Pelletier (2008) for new forms of identification using final equations], or any possible overidentifying restrictions that might be considered for inference purpose.

In particular, we extend the results of Dufour and Jouini (2005) to include a constant among the regressors—which is realistic in practice—and consider consistent two-stage generalized least squares (GLS) estimators. More especially, we derive a new third-stage generalized linear regression that yields fully efficient estimators that are asymptotically equivalent to Maximum Likelihood (ML) under Gaussian errors. This provides an appealing and intuitive interpretation of nonlinear estimation procedures such as Maximum Likelihood (ML) and nonlinear generalized least squares (NGLS), as we justify the implementation of the third-stage estimation without any prior knowledge of the actual distribution of the errors—unlike Hannan and Kavalieris (1984*b*), and Reinsel et al. (1992). In particular, we show that our third-stage GLS estimator is different from those proposed in the literature and corresponds exactly to a one iteration of the Gauss-Newton algorithm starting from the consistent two-stage GLS estimator. Moreover, simulation evidence shows that the finite sample properties of our third-stage estimators are comparatively better in most cases than those suggested by Hannan and Kavalieris (1984*b*), Reinsel et al. (1992), and Poskitt and Salau (1995), respectively (although asymptotically equivalent). This is mainly because we propose a new recursive method to filter the new residuals—necessary to the third-stage GLS estimation—that are function of the first-stage long-autoregression residuals and the second-stage residuals as well. Then using lagged values of these residuals as regressors in the computation of our estimates. Thus, considering different regressors compared to other alternative methods. This is with the fact that our efficient estimators and those proposed by Hannan and Kavalieris (1984*b*), Reinsel et al. (1992), and Poskitt and Salau (1995) do not use the same weighting covariance matrix. We particularly show that our weighting matrix has a rate of convergence towards the true innovation covariance matrix faster than those considered in Hannan and Kavalieris (1984*b*) and Poskitt and Salau (1995), respectively. Moreover, unlike Reinsel et al. (1992), and Poskitt and Salau (1995) procedures, our estimators are not time consuming since we don't have to deal with the inversion of high dimensional matrices. Furthermore, we provide the asymptotic distribution of the two-stage as well as the third-stage GLS estimators under the assumption of strong WN, since, to the best of our knowledge, they have not been stated anywhere, and show under general conditions that these distributions are asymptotically normal. Finally, we give their respective covariance estimators. The latter are relatively simple to use, for example, for building finite sample simulation-based inference on the echelon form model parameters, including the use of bootstrap methods.

The paper proceeds as follows. Section **2** shows how the echelon VARMA representation is used to ensure a unique parameterization. Section **3** describes the two-step GLS procedure (allowing for intercepts) and derives the estimators' properties such as convergence and asymptotic distribution. Section **4** provides a heuristic derivation of the third-stage estimators, then demonstrates its asymptotic efficiency under i.i.d. Gaussian innovations. Section **5** shows a simulation study on the finite sample performance of our proposed procedure compared to selected methods. Finally, section **6** concludes. The proofs of the lemmas, propositions and theorems are supplied in Appendix **A**.

2 Framework

We consider a k-dimensional stochastic process of the autoregressive moving-average (VARMA) type with known order (p, q). We first define the standard VARMA representation entailing identification problems. Then, among the representations ensuring parameter uniqueness in VARMA models, we proceed with the echelon form. Finally, we formulate the basic regularity assumptions we shall consider in the sequel.

2.1 Standard form

Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional random vector process with the VARMA representation

$$y_t = \mu_A + \sum_{i=1}^{p} A_i y_{t-i} + u_t + \sum_{j=1}^{q} B_j u_{t-j} \tag{2.1}$$

where $y_t = (y_{1,t}, \ldots, y_{k,t})'$, $\mu_A = A(1)\,\mu_y$, $A(1) = I_k - \sum_{i=1}^{p} A_i$, $\mu_y = \mathsf{E}(y_t)$, p and q are non-negative integers (respectively, the autoregressive and moving average orders), A_i and B_j are $k \times k$ fixed coefficient matrices, $\{u_t : t \in \mathbb{Z}\} \sim \text{WN}(0, \Sigma_u)$, i.e. u_t is a (second order) WN process, such that $\Sigma_u = \mathsf{E}(u_t u_t')$, where Σ_u is a $k \times k$ positive definite symmetric matrix. Under stationarity and invertibility conditions the coefficients A_i and B_j satisfy the constraints $\det\{A(z)\} \neq 0$ and $\det\{B(z)\} \neq 0$ for all $|z| \leq 1$, where z is a complex number, $A(z) = I_k - \sum_{i=1}^{p} A_i z^i$ and $B(z) = I_k + \sum_{j=1}^{q} B_j z^j$. Then y_t has the infinite-order autoregressive and moving average representations, respectively:

$$y_t = \mu_\Pi + \sum_{\tau=1}^{\infty} \Pi_\tau y_{t-\tau} + u_t, \quad \text{and} \quad y_t = \mu_y + u_t + \sum_{v=1}^{\infty} \Psi_v u_{t-v} \tag{2.2}$$

where $\Pi(z) = B(z)^{-1} A(z) = I_k - \sum_{\tau=1}^{\infty} \Pi_\tau z^\tau$, $\Psi(z) = A(z)^{-1} B(z) = I_k + \sum_{v=1}^{\infty} \Psi_v z^v$, with $\det\{\Pi(z)\} \neq 0$ and $\det\{\Psi(z)\} \neq 0$ for all $|z| \leq 1$, and $\mu_\Pi = \Pi(1)\,\mu_y$ where $\Pi(1) = I_k - \sum_{\tau=1}^{\infty} \Pi_\tau$. Further, there exist real constants $C > 0$ and $\rho \in (0, 1)$ such that

$$\|\Pi_\tau\| \leq C\rho^\tau, \quad \|\Psi_v\| \leq C\rho^v. \tag{2.3}$$

Hence, $\sum_{\tau=1}^{\infty} \|\Pi_\tau\| < \infty$ and $\sum_{v=1}^{\infty} \|\Psi_v\| < \infty$, where $\|.\|$ stands for Schur's norm [see Horn and Johnson (1985, Section 5.6)], $i.e.$ $\|M\|^2 = \text{tr}[M'M]$ for any matrix M.

2.2 Echelon form

The standard VARMA(p, q) representation (2.1) is not unique. The coefficient matrices A_i and B_j are not uniquely determined by the covariance structure (although Π_τ and Ψ_v typically are). To ensure a unique parameterization of (2.1) we consider the stationary invertible VARMA(p, q) process in echelon form

$$\Phi(L)\, y_t = \mu_\Phi + \Theta(L)\, u_t \tag{2.4}$$

where $\Phi(L) = \Phi_0 - \sum_{i=1}^{\bar{p}} \Phi_i L^i$, $\Theta(L) = \Theta_0 + \sum_{j=1}^{\bar{p}} \Theta_j L^j$, L denotes the lag operator, $\mu_\Phi = \Phi(1)\, \mu_y$, $\bar{p} = \max(p, q)$, $\Theta_0 = \Phi_0$, and Φ_0 is a lower-triangular matrix whose all diagonal elements are equal to one. The VARMA representation (2.4) is in echelon form if $\Phi(L) = [\phi_{lm}(L)]_{l,m=1,\dots,k}$ and $\Theta(L) = [\theta_{lm}(L)]_{l,m=1,\dots,k}$ satisfy the following conditions: given a vector of Kronecker indices $(p_1, \dots, p_k)'$, the operators $\phi_{lm}(L)$ and $\theta_{lm}(L)$ on any given row l of $\Phi(L)$ and $\Theta(L)$ have the same degree p_l and

$$\begin{aligned} \phi_{lm}(L) &= 1 - \sum_{i=1}^{p_l} \phi_{ll,i} L^i && \text{if } l = m, \\ &= -\sum_{i=p_l-p_{lm}+1}^{p_l} \phi_{lm,i} L^i && \text{if } l \neq m, \end{aligned} \tag{2.5}$$

$$\theta_{lm}(L) = \sum_{j=0}^{p_l} \theta_{lm,j} L^j, \text{ with } \Theta_0 = \Phi_0, \tag{2.6}$$

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

$$\begin{aligned} p_{lm} &= \min(p_l + 1, p_m) && \text{for } l \geq m, \\ &= \min(p_l, p_m) && \text{for } l < m. \end{aligned} \tag{2.7}$$

Note that $p_{ll} = p_l$ is the number of free varying coefficients on the l-th diagonal element of $\Phi(L)$ as well the order of the polynomials on the corresponding row of $\Theta(L)$, while p_{lm} specifies the number of free coefficients in the operator $\phi_{lm}(L)$ for $l \neq m$. $\sum_{l=1}^{k} p_l$ is the McMillan degree and $P = [p_{lm}]_{l,m=1,\dots,k}$ is the matrix formed by the Kronecker indices. 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, $\bar{p} = \max(p_1, \dots, p_k)$. Moreover, this echelon-form parameterization of VARMA(p, q) models [hereafter VARMA$(p_1, \dots, p_k)$], ensures the uniqueness of left-coprime operators $\Phi(L)$ and $\Theta(L)$. Among other identifiable parameterizations, such as the final equations form, the echelon form has been preferred for parsimony and gain efficiency criteria. For proofs of the uniqueness of the echelon form and other identification conditions, the reader should consult Hannan (1969*b*, 1970, 1976, 1979), Deistler and Hannan (1981), Hannan and Deistler (1988), and Lütkepohl (1991, Chapter 7).

The implied stationarity and invertibility conditions in (2.4) are: $\det\{\Phi(z)\} \neq 0$ and $\det\{\Theta(z)\} \neq 0$ for all $|z| \leq 1$, where $\Phi(z) = \Phi_0 - \sum_{i=1}^{\bar{p}} \Phi_i z^i$, $\Theta(z) = \Theta_0 + \sum_{j=1}^{\bar{p}} \Theta_j z^j$, with $\Pi(z) = \Theta(z)^{-1} \Phi(z)$ and $\Psi(z) = \Phi(z)^{-1} \Theta(z)$. Let also $\Theta(z)^{-1} = \sum_{\tau=0}^{\infty} \Lambda_\tau(\eta)\, z^\tau$ where by invertibility $\|\Lambda_\tau(\eta)\| \leq C\rho^\tau$, $\sum_{\tau=0}^{\infty} \|\Lambda_\tau(\eta)\| < \infty$, with η (as it will be specified further) the vector of all free varying parameters implied by the echelon form. Now, set $v_t = y_t - u_t$. We can see that

$$v_t = \Phi_0^{-1} \Big[\mu_\Phi + \sum_{i=1}^{\bar{p}} \Phi_i y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j} \Big]. \tag{2.8}$$

Obviously, v_t is uncorrelated with the error term u_t and (2.4) takes the form

$$y_t = \mu_\Phi + \big(I_k - \Phi_0\big) v_t + \sum_{i=1}^{\bar{p}} \Phi_i y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j} + u_t. \tag{2.9}$$

Set

$$\beta = \operatorname{vec}\big[\mu_\Phi, I_k - \Phi_0, \Phi_1, \ldots, \Phi_{\bar{p}}, \Theta_1, \ldots, \Theta_{\bar{p}}\big], \tag{2.10}$$

$$X_t = \big[1, v_t', y_{t-1}', \ldots, y_{t-\bar{p}}', u_{t-1}', \ldots, u_{t-\bar{p}}'\big]' \tag{2.11}$$

where β and X_t are vectors of sizes $k^2h + k$ and $kh + 1$, respectively, with $h = 2\bar{p} + 1$. Under the echelon form restrictions (2.4) – (2.7), the representation (2.9) implies a unique $\big(k^2h + k\big) \times r_{\bar{p}}$ full rank columns matrix R formed by $r_{\bar{p}}$ selected distinct vectors from the identity matrix of order $k^2h + k$ such that $R'R = I_{r_{\bar{p}}}$ and $\beta = R\eta$, where η is an $r_{\bar{p}}$-dimensional vector of free varying parameters, with $r_{\bar{p}} < \big(k^2h + k\big)$. Hence (2.9) can be restated as

$$y_t = \big[X_t' \otimes I_k\big]R\eta + u_t \tag{2.12}$$

where $\big[X_t' \otimes I_k\big]R$ is a $k \times r_{\bar{p}}$ matrix. Further, the echelon form ensures that $R'\big[X_t \otimes I_k\big]$ has a non singular covariance matrix, so that

$$\operatorname{rank}\big\{R'\big[\Gamma_X \otimes I_k\big]R\big\} = r_{\bar{p}} \tag{2.13}$$

where $\Gamma_X = \mathsf{E}\big[X_t X_t'\big]$. Now, let $y = \big[y_1', \ldots, y_T'\big]'$, $X = \big[X_1, \ldots, X_T\big]$ and $u = \big[u_1', \ldots, u_T'\big]'$. Then the stacked form of (2.12) is

$$y = \big[X' \otimes I_k\big]R\eta + u \tag{2.14}$$

where $\big[X' \otimes I_k\big]R$ is a $(kT) \times r_{\bar{p}}$ matrix. In the following, we shall assume that

$$\operatorname{rank}\big\{\big[X' \otimes I_k\big]R\big\} = r_{\bar{p}} \text{ with probability } 1. \tag{2.15}$$

Under the assumption that the process is regular with continuous distribution, the latter statement must hold.

2.3 Regularity assumptions

Assumptions on the innovation process and the truncation lag of the long autoregression are needed to establish the consistency and asymptotic distribution of the linear estimators defined below. We shall consider in the sequel the following.

Assumption 2.1 *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 *There is a finite constant m_4 such that, for all $1 \le i,\ j,\ r,\ s \le k$,*

$$\mathsf{E}\,|u_{i,t}u_{j,t}u_{r,t}u_{s,t}| \le m_4 < \infty\,, \text{ for all } t.$$

Assumption 2.3 *n_T is a function of T such that*

$$n_T \to \infty \text{ and } n_T^2/T \to 0 \text{ as } T \to \infty \tag{2.16}$$

and, for some $c > 0$ and $0 < \delta_1 < 1/2$,

$$n_T \ge cT^{\delta_1} \text{ for } T \text{ sufficiently large.} \tag{2.17}$$

Assumption 2.4 *The coefficients of the autoregressive representation* (2.2) *satisfy*

$$n_T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T, n_T \to \infty. \tag{2.18}$$

Assumption 2.5 n_T *is a function of* T *such that*

$$n_T \to \infty \text{ and } n_T^3/T \to 0 \text{ as } T \to \infty \tag{2.19}$$

and, for some $c > 0$ *and* $0 < \delta_2 < 1/3$,

$$n_T \geq cT^{\delta_2} \text{ for } T \text{ sufficiently large.} \tag{2.20}$$

Assumption 2.6 *The coefficients of the autoregressive representation* (2.2) *satisfy*

$$T^{1/2} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T, n_T \to \infty. \tag{2.21}$$

Assumption 2.7 *The coefficients of the autoregressive representation* (2.2) *satisfy*

$$T^{\delta_3} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T, n_T \to \infty \tag{2.22}$$

for some $1/2 < \delta_3 < 1$.

Assumption **2.1** implies a strong VARMA process, while Assumption **2.2** on moments of order four ensures that the empirical autocovariances of the process have finite variances. Assumption **2.3** states that n_T grows to infinity at a rate slower than $T^{1/2}$; for instance, the assumption is satisfied if $n_T = cT^\delta$ with $0 < \delta_1 \leq \delta < 1/2$. Assumption **2.4** describes the rate of decay of autoregressive coefficients relatively to n_T. While Assumptions **2.5** and **2.6** are stronger versions of Assumptions **2.3** and **2.4**, respectively. Assumption **2.7** states that for any constant $1/2 < \delta \leq \delta_3$ (with $\delta_3 < 1$) the truncated sum $T^\delta \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\|$ converges to zero as T and n_T go to infinity.

Although the above assumptions are sufficient to show consistency of the two-step linear estimator, another assumption is needed to show the asymptotic normality of its distribution.

Assumption 2.8 n_T *is a function of* T *such that*

$$n_T \to \infty \text{ and } n_T^4/T \to 0 \text{ as } T \to \infty. \tag{2.23}$$

The latter assumption means that n_T grows to infinity at a rate slower than $T^{1/4}$; for example, it is the case if $n_T = cT^\delta$ with $0 < \bar{\delta} \leq \delta < 1/4$. It is easy to see that (2.23) entails (2.19) and (2.16). Finally, it is worthwhile to note that (2.18) holds for VARMA processes whenever $n_T = cT^\delta$ with $c > 0$ and $\delta > 0$, *i.e.*

$$T^\delta \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \text{ as } T \to \infty, \quad \text{for all } \delta > 0. \tag{2.24}$$

This follows from the exponential decay of $\|\Pi_\tau\|$ for VARMA processes.

3 Generalized two-step linear estimation

We propose a two-step generalized linear regression method for obtaining consistent estimates of echelon-form VARMA models with known Kronecker indices.

Let $\{y_{-n_T+1}, \ldots, y_T\}$ be a random sample of size $n_T + T$ where n_T is a sequence function of T such that n_T grows to infinity as T goes to infinity. Now, consider the "*long*" multivariate linear autoregressive model of lag-order n_T:

$$y_t = \mu_\Pi (n_T) + \sum_{\tau=1}^{n_T} \Pi_\tau y_{t-\tau} + u_t (n_T) \tag{3.1}$$

where $\mu_\Pi (n_T) = \big(I_k - \sum_{\tau=1}^{n_T} \Pi_\tau\big)\mu_y$ and

$$u_t (n_T) = \sum_{\tau=n_T+1}^{\infty} \Pi_\tau \big(y_{t-\tau} - \mu_y\big) + u_t. \tag{3.2}$$

Setting $Y_t (n_T) = \big[1, y'_{t-1}, \ldots, y'_{t-n_T}\big]'$ and $\Pi (n_T) = [\mu_\Pi (n_T), \Pi_1, \ldots, \Pi_{n_T}]$, then the corresponding multivariate least squares (LS) estimator is:

$$\tilde{\Pi} (n_T) = \big[\tilde{\mu}_\Pi (n_T), \tilde{\Pi}_1 (n_T), \ldots, \tilde{\Pi}_{n_T} (n_T)\big] = \tilde{W}_Y (n_T) \tilde{\Gamma}_Y (n_T)^{-1} \tag{3.3}$$

where $\tilde{W}_Y (n_T) = T^{-1} \sum_{t=1}^{T} y_t Y_t (n_T)'$ and $\tilde{\Gamma}_Y (n_T) = T^{-1} \sum_{t=1}^{T} Y_t (n_T) Y_t (n_T)'$. This estimator can be obtained by running k separate univariate linear regressions, one for each component $y_{k,t}$. The Yule-Walker estimates of the theoretical coefficients Π_τ could also be considered. Set $\Gamma_Y (n_T) = \mathsf{E}\big[Y_t (n_T) Y_t (n_T)'\big]$. Also, let $\|.\|_1$ such that, for any given matrix A, $\|A\|_1$ is the largest eigenvalue of $A'A$, so that $\|A\|_1 = \sup_{x \neq 0}\big\{\frac{\|Ax\|}{\|x\|}\big\}$.

Proposition 3.1 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the Assumptions* **2.1** *to* **2.3***, we have*

$$\big\|\Gamma_Y (n_T)^{-1}\big\| = O_p\big(1\big), \tag{3.4}$$

$$\big\|\tilde{\Gamma}_Y (n_T)^{-1} - \Gamma_Y (n_T)^{-1}\big\| = \big\|\tilde{\Gamma}_Y (n_T)^{-1} - \Gamma_Y (n_T)^{-1}\big\|_1 = O_p\big(n_T/T^{1/2}\big). \tag{3.5}$$

If Assumption **2.4** is also satisfied, then the following theorem is the extension to the drift case of Theorem **1** of Lewis and Reinsel (1985) and Theorem **2.1** of Paparoditis (1996).

Theorem 3.1 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the Assumptions* **2.1** *to* **2.4***, we have:*

$$\big\|\tilde{\Pi} (n_T) - \Pi (n_T)\big\| = o_p\big(1\big). \tag{3.6}$$

If, furthermore, Assumption **2.6** *holds, then*

$$\big\|\tilde{\Pi} (n_T) - \Pi (n_T)\big\| = O_p\big(n_T^{1/2}/T^{1/2}\big). \tag{3.7}$$

Now, let l_{n_T} be a sequence of $k^2 n_T + k$-dimensional vectors such that

$$0 < M_1 \leq \|l_{n_T}\|^2 \leq M_2 < \infty \quad \text{for } n_T = 1, 2, \ldots \tag{3.8}$$

Set also

$$\tilde{S}_Y(n_T) = T^{1/2} l'_{n_T} \text{vec}\big[\tilde{\Omega}_Y(n_T)\,\tilde{\Gamma}_Y(n_T)^{-1}\big],\; S_Y(n_T) = T^{1/2} l'_{n_T} \text{vec}\big[\Omega_Y(n_T)\,\Gamma_Y(n_T)^{-1}\big], \tag{3.9}$$

with $\tilde{\Omega}_Y(n_T) = T^{-1}\sum_{t=1}^{T} u_t(n_T)\,Y_t(n_T)'$ and $\Omega_Y(n_T) = T^{-1}\sum_{t=1}^{T} u_t Y_t(n_T)'$. Then we have the following asymptotic equivalence.

Proposition 3.2 *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the Assumptions* **2.1***,* **2.2***,* **2.5** *and* **2.6***, we have:*

$$\big\|\tilde{S}_Y(n_T) - S_Y(n_T)\big\| = o_p(1). \tag{3.10}$$

If, furthermore, Assumption **2.7** *holds, then*

$$\big\|\tilde{S}_Y(n_T) - S_Y(n_T)\big\| = O_p\big(n_T^{3/2}/T^{1/2}\big). \tag{3.11}$$

The next theorem shows that asymptotic normality holds as an immediate consequence of Proposition **3.2**. This proposition and the following theorem are generalizations to the drift case of Theorems **2** and **4** of Lewis and Reinsel (1985), respectively.

Theorem 3.2 *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the Assumptions* **2.1***,* **2.2***,* **2.5***,* **2.6** *and* **2.7***, we have:*

$$\frac{T^{1/2} l'_{n_T} \text{vec}\big[\tilde{\Pi}(n_T) - \Pi(n_T)\big]}{\big\{l'_{n_T} Q_Y(n_T)\, l_{n_T}\big\}^{1/2}} \xrightarrow[T\to\infty]{d} N\big[0,1\big] \tag{3.12}$$

where

$$Q_Y(n_T) = \Gamma_Y(n_T)^{-1} \otimes \Sigma_u. \tag{3.13}$$

A possible choice for n_T to satisfy both assumptions **2.5** and **2.6** is $n_T = T^{1/\varepsilon}$ 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 fade as $T \to \infty$. Let

$$\tilde{u}_t(n_T) = y_t - \tilde{\mu}_\Pi(n_T) - \sum_{\tau=1}^{n_T} \tilde{\Pi}_\tau(n_T)\, y_{t-\tau} \tag{3.14}$$

be the LS residuals of the long autoregression (3.1), and let

$$\tilde{\Sigma}_u(n_T) = \frac{1}{T}\sum_{t=1}^{T} \tilde{u}_t(n_T)\,\tilde{u}_t(n_T)' \tag{3.15}$$

be the corresponding innovation covariance matrix estimator. Then we have the following.

Proposition 3.3 *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VAR representation* (3.1)*. Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$\big\|\tilde{\Sigma}_u(n_T) - \Sigma_u\big\| = \big\|\tilde{\Sigma}_u(n_T)^{-1} - \Sigma_u^{-1}\big\| = O_p\big(n_T/T^{1/2}\big). \tag{3.16}$$

The asymptotic equivalence stated above suggests that we may be able to estimate consistently the parameters of the VARMA model in (2.9) by replacing the unobserved lagged innovations $u_{t-1}, \ldots, u_{t-\bar{p}}$ with their corresponding first stage estimates $\tilde{u}_{t-1}(n_T), \ldots, \tilde{u}_{t-\bar{p}}(n_T)$. Hence, (2.9) can be rewritten as

$$y_t = \mu_\Phi + \big(I_k - \Phi_0\big)\tilde{v}_t\left(n_T\right) + \sum_{i=1}^{\bar{p}} \Phi_i y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j \tilde{u}_{t-j}\left(n_T\right) + e_t\left(n_T\right) \tag{3.17}$$

or equivalently,

$$y_t = \big[\tilde{X}_t\left(n_T\right)' \otimes I_k\big] R\eta + e_t\left(n_T\right) \tag{3.18}$$

where

$$\tilde{v}_t\left(n_T\right) = y_t - \tilde{u}_t\left(n_T\right), \quad e_t\left(n_T\right) = \tilde{u}_t\left(n_T\right) + \sum_{j=0}^{\bar{p}} \Theta_j \big[u_{t-j} - \tilde{u}_{t-j}\left(n_T\right)\big], \tag{3.19}$$

$$\tilde{X}_t\left(n_T\right) = \big[1, \tilde{v}_t\left(n_T\right)', y'_{t-1}, \ldots, y'_{t-\bar{p}}, \tilde{u}_{t-1}\left(n_T\right)', \ldots, \tilde{u}_{t-\bar{p}}\left(n_T\right)'\big]'. \tag{3.20}$$

Recall that running OLS on (3.17) or (3.18) corresponds to the third-stage and the second-stage estimators of Hannan and Kavalieris (1984*b*) and Dufour and Jouini (2005) methods, respectively. In the other hand, the second-stage estimator of Poskitt and Salau (1995) procedure is obtained by running OLS on a variant of (3.17), say

$$y_t - \tilde{u}_t\left(n_T\right) = \mu_\Phi + \big(I_k - \Phi_0\big)\tilde{v}_t\left(n_T\right) + \sum_{i=1}^{\bar{p}} \Phi_i y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j \tilde{u}_{t-j}\left(n_T\right) + \xi_t \tag{3.21}$$

where $\xi_t = \sum_{j=0}^{\bar{p}} \Theta_j \varepsilon_{t-j}$, with $\varepsilon_{t-j} = u_{t-j} - \tilde{u}_{t-j}\left(n_T\right)$. In this paper, we consider the explicit echelon form two-step GLS estimator of η,

$$\tilde{\eta} = \arg\min_{\eta} \sum_{t=1}^{T} e_t\left(n_T\right)' \tilde{\Sigma}_u\left(n_T\right)^{-1} e_t\left(n_T\right). \tag{3.22}$$

This estimator has the form

$$\tilde{\eta} = \tilde{Q}_X\left(n_T\right) \tilde{W}_X\left(n_T\right) \tag{3.23}$$

where

$$\tilde{Q}_X\left(n_T\right) = \Big\{R'\tilde{\Upsilon}_X\left(n_T\right) R\Big\}^{-1}, \ \tilde{\Upsilon}_X\left(n_T\right) = \tilde{\Gamma}_X\left(n_T\right) \otimes \tilde{\Sigma}_u\left(n_T\right)^{-1}, \tag{3.24}$$

$$\tilde{\Gamma}_X\left(n_T\right) = \frac{1}{T}\sum_{t=1}^{T} \tilde{X}_t\left(n_T\right) \tilde{X}_t\left(n_T\right)', \ \tilde{W}_X\left(n_T\right) = \frac{1}{T}\sum_{t=1}^{T} R'\big[\tilde{X}_t\left(n_T\right) \otimes I_k\big]\tilde{\Sigma}_u\left(n_T\right)^{-1} y_t. \tag{3.25}$$

Setting

$$\tilde{\Omega}_X\left(n_T\right) = \frac{1}{T}\sum_{t=1}^{T} R'\big[\tilde{X}_t\left(n_T\right) \otimes I_k\big]\tilde{\Sigma}_u\left(n_T\right)^{-1} e_t\left(n_T\right), \tag{3.26}$$

one can see that

$$\tilde{\eta} - \eta = \tilde{Q}_X\left(n_T\right) \tilde{\Omega}_X\left(n_T\right). \tag{3.27}$$

Using the inequality $\|AB\|^2 \leq \|A\|_1^2 \|B\|^2$, for any two conformable matrices A and B, we get

$$\|\tilde{\eta} - \eta\| \leq \big\|\tilde{Q}_X\left(n_T\right)\big\|_1 \big\|\tilde{\Omega}_X\left(n_T\right)\big\|. \tag{3.28}$$

Now, define

$$\Upsilon_X = \Gamma_X \otimes \Sigma_u^{-1}, \quad Q_X = \left\{ R' \Upsilon_X R \right\}^{-1}, \quad \Omega_X = \frac{1}{T} \sum_{t=1}^{T} R' \big[X_t \otimes I_k \big] \Sigma_u^{-1} u_t. \tag{3.29}$$

Obviously, by the regularity assumption Q_X^{-1} is positive definite, and to study the convergence and distributional properties of $(\tilde{\eta} - \eta)$ we need first to establish the following proposition.

Proposition 3.4 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7). *Then, under the assumptions* **2.1** *to* **2.4**, *we have:*

$$\big\| \tilde{Q}_X(n_T) - Q_X \big\|_1 = O_p\big(n_T / T^{1/2} \big). \tag{3.30}$$

The latter proposition shows that the regressor matrix $\tilde{X}_t(n_T)$ as well as the covariance matrix $\tilde{Q}_X(n_T)$—based on approximate innovations—are all asymptotically equivalent to their analogous based on true innovations, according to the rate $n_T / T^{1/2}$. This suggests that $\tilde{\eta}$ converges to η. The next theorem establishes the appropriate rate of such convergence.

Theorem 3.3 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7). *Then, under the assumptions* **2.1** *to* **2.8**, *we have:*

$$\|\tilde{\eta} - \eta\| = O_p\big(T^{-1/2} \big). \tag{3.31}$$

To derive the asymptotic distribution for $\tilde{\eta}$, we shall first establish the asymptotic equivalence between the following random vectors

$$\tilde{S}_X(n_T) = T^{1/2} \tilde{Q}_X(n_T) \tilde{\Omega}_X(n_T), \quad S_X = T^{1/2} Q_X \Omega_X. \tag{3.32}$$

Proposition 3.5 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7). *Then, under the assumptions* **2.1** *to* **2.8**, *we have:*

$$\big\| \tilde{S}_X(n_T) - S_X \big\| = O_p\big(n_T^2 / T^{1/2} \big). \tag{3.33}$$

The next theorem provides the asymptotic distribution of the two-step GLS estimators.

Theorem 3.4 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7). *If the assumptions* **2.1** *to* **2.8** *are satisfied, then*

$$T^{1/2}\big(\tilde{\eta} - \eta \big) \underset{T\to\infty}{\overset{d}{\longrightarrow}} N\big[0, Q_X \big], \quad T^{1/2}\big(\tilde{\beta} - \beta \big) \underset{T\to\infty}{\overset{d}{\longrightarrow}} N\big[0, V_X \big] \tag{3.34}$$

where $\tilde{\beta} = R\tilde{\eta}$ and $V_X = R Q_X R'$.

Further, $\left\{ R' \big[\sum_{t=1}^{T} \tilde{X}_t(n_T) \tilde{X}_t(n_T)' \otimes \tilde{\Sigma}_u(n_T)^{-1} \big] R \right\}^{-1}$ is a consistent estimator of its covariance matrix. It is worth noting that the GLS estimator suggested by De Frutos and Serrano (2002), although different from what we consider in (3.27), has the limiting distribution established in the above theorem for the strong WN case. A result they have mentioned in their paper for the pure vector moving-average processes but did not show for the VARMA case. Now set

$$\tilde{\Sigma}_e(n_T) = \frac{1}{T} \sum_{t=1}^{T} \tilde{e}_t(n_T) \tilde{e}_t(n_T)' \tag{3.35}$$

where

$$\tilde{e}_t(n_T) = y_t - \big[\tilde{X}_t(n_T)' \otimes I_k\big] R\tilde{\eta}. \tag{3.36}$$

Then, we have the following proposition.

Proposition 3.6 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7). *Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$\big\|\tilde{\Sigma}_e(n_T) - \Sigma_u\big\| = \big\|\tilde{\Sigma}_e(n_T)^{-1} - \Sigma_u^{-1}\big\| = O_p\big(n_T/T^{1/2}\big). \tag{3.37}$$

4 Asymptotic efficiency

The two-step linear estimator derived above is not efficient under Gaussian innovations. To allow for asymptotic efficiency [as in the fourth-stage of Hannan and Kavalieris (1984*b*)], one may perform a third-stage linear estimation that we shall describe below.

Unlike Hannan and Kavalieris (1984*b*) who assumed Gaussian errors in order to suggest their fourth-stage estimators, we show, rather, how such estimators can be derived without any prior knowledge of the actual distribution of the innovations. This will be useful to establish the asymptotic efficiency of these estimates under Gaussian assumption. In line with their procedure at the specification level which is heavy to implement even in small systems, the fourth stage estimation they suggested to achieve asymptotic efficiency does not explicitly show the echelon form zero-restrictions. In contrast, we give simple, compact and efficient echelon-form estimators that can easily be computed by running a simple linear regression. Thus, one might consider further linear regressions as they are costless. Moreover, we provide a simple estimator of its covariance matrix.

Now, recall that the main problem is to minimize an objective function that is nonlinear in the parameter vector η: we have to solve

$$\min_{\eta} \sum_{t=1}^{T} u_t' \Sigma_u^{-1} u_t \tag{4.1}$$

where $u_t = \sum_{\tau=0}^{\infty} \Lambda_\tau(\eta) \left[\Phi_0\left(y_{t-\tau} - \mu_y\right) - \sum_{i=1}^{\bar{p}} \Phi_i\left(y_{t-i-\tau} - \mu_y\right)\right]$. Setting

$$u_t(\eta) = \sum_{\tau=0}^{t-1} \Lambda_\tau(\eta) \Big[\Phi_0\big(y_{t-\tau} - \mu_y\big) - \sum_{i=1}^{\bar{p}} \Phi_i\big(y_{t-i-\tau} - \mu_y\big)\Big], \tag{4.2}$$

one can see that $\|u_t - u_t(\eta)\| = O_p\left(\rho^t\right)$, as it can be shown that

$$\mathsf{E}\,\|u_t - u_t(\eta)\| \le \sum_{\tau=t}^{\infty} \|\Lambda_\tau(\eta)\| \, \|\Phi(\bar{p})\| \, \mathsf{E}\,\big\|Y_{t-\tau}^a(\bar{p})\big\| = O\left(\rho^t\right) \tag{4.3}$$

where $\Phi(\bar{p}) = [\Phi_0, -\Phi_1, \ldots, -\Phi_{\bar{p}}]$, $Y_t^a(\bar{p}) = \left[y_t^{a\prime}, y_{t-1}^{a\prime}, \ldots, y_{t-\bar{p}}^{a\prime}\right]'$ with $y_t^a = \big(y_t - \mu_y\big)$; see the proof of Theorem **3.1**. This suggests replacing the problem (4.1) with

$$\min_{\eta} \sum_{t=1}^{T} u_t(\eta)' \, \Sigma_u^{-1} u_t(\eta)\,. \tag{4.4}$$

Also, note that (3.36) can alternatively be expressed as in (3.19), as

$$\tilde{e}_t(n_T) = \tilde{u}_t(n_T) + \sum_{j=0}^{\bar{p}} \tilde{\Theta}_j \big[\tilde{u}_{t-j} - \tilde{u}_{t-j}(n_T)\big], \tag{4.5}$$

so that, using the second-stage estimate $\tilde{\eta}$, the estimated model

$$y_t = \left[\tilde{X}_t\left(n_T\right)' \otimes I_k\right] R\tilde{\eta} + \tilde{e}_t\left(n_T\right) \tag{4.6}$$

takes the form

$$y_t = \tilde{\mu}_\Phi + \left(I_k - \tilde{\Phi}_0\right)\tilde{v}_t + \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i} + \sum_{j=1}^{\bar{p}} \tilde{\Theta}_j \tilde{u}_{t-j} + \tilde{u}_t \tag{4.7}$$

where $\tilde{v}_t = y_t - \tilde{u}_t$ with

$$\tilde{u}_t = \sum_{\tau=0}^{\infty} \Lambda_\tau\left(\tilde{\eta}\right)\left[\tilde{\Phi}_0\left(y_{t-\tau} - \tilde{\mu}_y\right) - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i\left(y_{t-i-\tau} - \tilde{\mu}_y\right)\right], \tag{4.8}$$

$\tilde{\mu}_y = \tilde{\Phi}\left(1\right)^{-1}\tilde{\mu}_\Phi$, $\tilde{\Phi}\left(1\right) = \tilde{\Phi}_0 - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i$ and $\sum_{j=0}^{\infty} \Lambda_\tau\left(\tilde{\eta}\right) z^\tau = \tilde{\Theta}\left(z\right)^{-1}$, where $\tilde{\mu}_\Phi$, $\tilde{\Phi}_i$ and $\tilde{\Theta}_j$ stand for the second-stage estimates of μ_Φ, Φ_i and Θ_j, respectively. In view of (4.7) and (4.8), it is obvious that the second-stage estimator $\tilde{\eta}$ may be used as initial value in the minimization algorithm when seeking the nonlinear GLS estimator. As for u_t and $u_t\left(\eta\right)$, we can approximate $\tilde{u}_t$ with

$$u_t\left(\tilde{\eta}\right) = \sum_{\tau=0}^{t-1} \Lambda_\tau\left(\tilde{\eta}\right)\left[\tilde{\Phi}_0 y_{t-\tau} - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i-\tau} - \tilde{\mu}_\Phi\right]. \tag{4.9}$$

This also can either be determined recursively as suggested in the literature [including Hannan and Kavalieris (1984*b*) and Reinsel et al. (1992)] using

$$u_t\left(\tilde{\eta}\right) = y_t - \tilde{\Phi}_0^{-1}\left[\tilde{\mu}_\Phi + \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i} + \sum_{j=1}^{\bar{p}} \tilde{\Theta}_j u_{t-j}\left(\tilde{\eta}\right)\right], \tag{4.10}$$

with initial values $u_t\left(\tilde{\eta}\right) = 0$, $t \leq 0$. Instead of the above recursive filtering scheme, we propose a new one. In particular, one may consider

$$u_t\left(\tilde{\eta}\right) = \tilde{\Phi}_0^{-1}\tilde{e}_t\left(n_T\right) + \left(I_k - \tilde{\Phi}_0^{-1}\right)\tilde{u}_t\left(n_T\right) + \sum_{j=1}^{\bar{p}} \tilde{\Phi}_0^{-1}\tilde{\Theta}_j\left[\tilde{u}_{t-j}\left(n_T\right) - u_{t-j}\left(\tilde{\eta}\right)\right] \tag{4.11}$$

initiating with $u_t\left(\tilde{\eta}\right) = \tilde{u}_t\left(n_T\right)$ for $0 < t \leq \bar{p}$. The latter has the feature of yielding filtered residuals that are function of the first-stage long-autoregression and second-stage residuals as well. Our argument is the following; since the error terms $e_t\left(n_T\right)$ in (3.17) or (3.18) are function of the true innovations u_t, as shown in (3.19), it follows that by simply estimating $e_t\left(n_T\right)$ one is about implicitly estimating the true innovations u_t. This is exactly described in (4.5) that we can see is satisfying (4.7). So one can obtain the new estimates of the true innovations corresponding to the second-stage echelon form parameter estimates, by solving for $u_t\left(\tilde{\eta}\right)$ in (4.5). Hence,

$$\tilde{u}_t = \tilde{u}_t\left(n_T\right) + \sum_{\tau=0}^{\infty} \Lambda_\tau\left(\tilde{\eta}\right)\left[\tilde{e}_{t-\tau}\left(n_T\right) - \tilde{u}_{t-\tau}\left(n_T\right)\right]. \tag{4.12}$$

These residuals can then be approximated with

$$u_t\left(\tilde{\eta}\right) = \tilde{u}_t\left(n_T\right) + \sum_{\tau=0}^{t-1} \Lambda_\tau\left(\tilde{\eta}\right)\left[\tilde{e}_{t-\tau}\left(n_T\right) - \tilde{u}_{t-\tau}\left(n_T\right)\right] \tag{4.13}$$

since $\tilde{u}_t\left(n_T\right)$ and $\tilde{e}_t\left(n_T\right)$ are not available for $t \leq 0$ and $t \leq \bar{p}$. Hence, setting $\tilde{u}_t\left(n_T\right) = 0$ for $t \leq 0$, and $\tilde{e}_t(n_T) = 0$ for $t \leq 0$ and $\tilde{e}_t(n_T) = \tilde{u}_t(n_T)$ for $1 \leq t \leq \bar{p}$, respectively. Finally, one can see that the above expression can be rewritten as in (4.11). However, it is worth noting that the convergence of the above two recursive schemes (4.10) and (4.11) to each other in finite sample—while remaining asymptotically equivalent—is fast only when the Kronecker indices are all equal. Let

$$\tilde{\Sigma}_u(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T} u_t(\tilde{\eta})\, u_t(\tilde{\eta})'. \tag{4.14}$$

To establish the rate of convergence of $\tilde{\Sigma}_u(\tilde{\eta})$ to Σ_u, we need the following lemma.

Lemma 4.1 *Let $\breve{\eta}$ be a $\sqrt{T}$-consistent estimator for η, i.e.*

$$T^{1/2}\|\breve{\eta} - \eta\| = O_p(1),$$

where $\|.\|$ denotes the Schur norm. Then there exists a real constant $\kappa > 0$ such that

$$T^{1/2}(1+\kappa^{-1})^{\tau}\|\Lambda_\tau(\breve{\eta}) - \Lambda_\tau(\eta)\| = O_p(1),\ \forall \tau \in \mathbb{Z}. \tag{4.15}$$

Proposition 4.1 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7)*. Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$\left\|\tilde{\Sigma}_u(\tilde{\eta}) - \Sigma_u\right\| = \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right\| = O_p\big(T^{-1/2}\big). \tag{4.16}$$

Further, consider the following lemma.

Lemma 4.2 *Let η^0 and η^1 be two distinct values of η. Then*

$$u_t(\eta^1) - u_t(\eta^0) = -Z_t^{\circ}(\eta^1,\eta^0)'\big(\eta^1 - \eta^0\big) \tag{4.17}$$

where

$$Z_t^{\circ}(\eta^1,\eta^0) = \sum_{\tau=0}^{t-1} R'\big[X_{t-\tau}(\eta^1)\otimes \Lambda_\tau(\eta^0)'\big], \tag{4.18}$$

$X_t(\eta^1) = \big[1, v_t(\eta^1)', y'_{t-1}, \ldots, y'_{t-\bar{p}}, u_{t-1}(\eta^1)', \ldots, u_{t-\bar{p}}(\eta^1)'\big]'$ *and* $v_t(\eta^1) = y_t - u_t(\eta^1)$.

Therefore, one can show that

$$u_t(\tilde{\eta}) - u_t(\eta) = -Z_t^{\circ}(\tilde{\eta},\eta)'\big(\tilde{\eta} - \eta\big) \tag{4.19}$$

where

$$Z_t^{\circ}(\tilde{\eta},\eta) = \sum_{\tau=0}^{t-1} R'\big[X_{t-\tau}(\tilde{\eta})\otimes \Lambda_\tau(\eta)'\big], \tag{4.20}$$

with $X_t(\tilde{\eta}) = \big[1, v_t(\tilde{\eta})', y'_{t-1}, \ldots, y'_{t-\bar{p}}, u_{t-1}(\tilde{\eta})', \ldots, u_{t-\bar{p}}(\tilde{\eta})'\big]'$ and $v_t(\tilde{\eta}) = y_t - u_t(\tilde{\eta})$. Hence (4.19) can be rearranged to obtain the linear regression:

$$w_t(\tilde{\eta}) = Z_t(\tilde{\eta})'\eta + \epsilon_t(\tilde{\eta},\eta) \tag{4.21}$$

where

$$w_t(\tilde{\eta}) = u_t(\tilde{\eta}) + Z_t(\tilde{\eta})'\tilde{\eta}, \quad Z_t(\tilde{\eta}) = \sum_{\tau=0}^{t-1} R'\big[X_{t-\tau}(\tilde{\eta})\otimes \Lambda_\tau(\tilde{\eta})'\big], \tag{4.22}$$

$$\epsilon_t(\tilde{\eta},\eta) = u_t(\eta) + \big[Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\big]'\big(\tilde{\eta} - \eta\big). \tag{4.23}$$

By Theorem **3.3** and Lemma **4.1**, one easily shows that $\left\|\epsilon_t(\tilde{\eta},\eta) - u_t(\tilde{\eta})\right\| = O_p\big(T^{-1/2}\big)$. This suggests obtaining a third-stage multivariate GLS estimator $\hat{\eta}$ of η by regressing $\tilde{\Sigma}_u(\tilde{\eta})^{-1/2} w_t(\tilde{\eta})$ on $\tilde{\Sigma}_u(\tilde{\eta})^{-1/2} Z_t(\tilde{\eta})'$. Hence

$$\hat{\eta} = \tilde{Q}_X(\tilde{\eta})\,\tilde{W}_X(\tilde{\eta}) \tag{4.24}$$

where

$$\tilde{Q}_X(\tilde{\eta}) = \Big\{\frac{1}{T}\sum_{t=1}^{T} Z_t(\tilde{\eta})\,\tilde{\Sigma}_u(\tilde{\eta})^{-1} Z_t(\tilde{\eta})'\Big\}^{-1}, \quad \tilde{W}_X(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T} Z_t(\tilde{\eta})\,\tilde{\Sigma}_u(\tilde{\eta})^{-1} w_t(\tilde{\eta}). \tag{4.25}$$

In view of (4.22), one can see that

$$\hat{\eta} = \tilde{\eta} + \tilde{Q}_X(\tilde{\eta})\,\tilde{\Omega}_X(\tilde{\eta}) \tag{4.26}$$

where

$$\tilde{\Omega}_X(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T} Z_t(\tilde{\eta})\,\tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t(\tilde{\eta}). \tag{4.27}$$

Obviously, our third-stage GLS estimators differ from those previously suggested in the literature [including Hannan and Kavalieris (1984*b*), Reinsel et al. (1992) and Poskitt and Salau (1995)] since we use different regressors in their computation. In particular, Hannan and Kavalieris (1984*b*) and Reinsel et al. (1992) use lagged values of the residuals filtered from (4.10) as regressors, while Poskitt and Salau (1995) use those associated with the first-stage residuals obtained from a long autoregression. Another feature making our efficient estimators different from that of Hannan and Kavalieris (1984*b*) is that, in $\tilde{Q}_X(\tilde{\eta})$ and $\tilde{\Omega}_X(\tilde{\eta})$, they used $\tilde{\Sigma}_e(n_T)$ instead of $\tilde{\Sigma}_u(\tilde{\eta})$, which corresponds to their third-stage covariance estimator of the innovations. So one can see from Propositions **3.6** and **4.1** that the weighting matrix we use has a faster convergence rate. This also holds for Poskitt and Salau (1995) as they use the fist-stage covariance estimator of the errors in the computation of their GLS estimator [see Proposition **3.3**]. Moreover, it is worth noting that, under Gaussian errors, $\hat{\eta}$ is asymptotically equivalent to ML estimator, since $\frac{\partial u_t(\eta)}{\partial \eta'}\Big|_{\eta=\tilde{\eta}} = -Z_t(\tilde{\eta})'$; see (4.19). Further, in view of (4.26), the estimator $\hat{\eta}$ corresponds to one iteration of Gauss-Newton algorithm.

Now, let

$$\tilde{Q}_X^{\circ}(\tilde{\eta}) = \Big\{\frac{1}{T}\sum_{t=1}^{T} Z_t^{\circ}(\tilde{\eta},\eta)\,\tilde{\Sigma}_u(\tilde{\eta})^{-1} Z_t^{\circ}(\tilde{\eta},\eta)'\Big\}^{-1}, \tag{4.28}$$

$$\tilde{\Omega}_X^{\circ}(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T} Z_t^{\circ}(\tilde{\eta},\eta)\,\tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t(\tilde{\eta}), \quad \tilde{\Omega}_X^{\bullet}(\tilde{\eta}) = \frac{1}{T}\sum_{t=1}^{T} Z_t^{\circ}(\tilde{\eta},\eta)\,\tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t(\eta), \tag{4.29}$$

$$Q_X(\eta) = \Big\{\mathsf{E}\big[Z_t(\eta)\,\Sigma_u^{-1} Z_t(\eta)'\big]\Big\}^{-1}, \quad \Omega_X(\eta) = \frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\,\Sigma_u^{-1} u_t, \tag{4.30}$$

$$Z_t(\eta) = \sum_{\tau=0}^{\infty} R'\big[X_{t-\tau} \otimes \Lambda_\tau(\eta)'\big]. \tag{4.31}$$

Using Lemma **4.2**, equation (4.26) can be rewritten as

$$\hat{\eta} - \eta = \tilde{Q}_X(\tilde{\eta})\,\Omega_X(\tilde{\eta}) + \tilde{Q}_X^{\circ}(\tilde{\eta})\left[\tilde{\Omega}_X^{\bullet}(\tilde{\eta}) - \tilde{\Omega}_X^{\circ}(\tilde{\eta})\right]. \tag{4.32}$$

Further, $Q_X(\eta)$ can be expressed as

$$Q_X(\eta) = \Big\{R'\Upsilon_X(\eta)\,R\Big\}^{-1} \tag{4.33}$$

where

$$\Upsilon_X(\eta) = \sum_{\tau_1=0}^{\infty}\sum_{\tau_2=0}^{\infty}\Big[\Gamma_X(\tau_1-\tau_2)\otimes\Lambda_{\tau_1}(\eta)'\,\Sigma_u^{-1}\Lambda_{\tau_2}(\eta)\Big], \tag{4.34}$$

with $\Gamma_X(\tau_1-\tau_2) = \mathsf{E}\big[X_{t-\tau_1}X'_{t-\tau_2}\big]$. By construction $Q_X(\eta)^{-1}$ is positive definite, and to study the convergence and distributional properties of $\hat{\eta}-\eta$, we first establish the following asymptotic equivalences.

Proposition 4.2 *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7)*. Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$\big\|\tilde{Q}_X(\tilde{\eta}) - \tilde{Q}^{\circ}_X(\tilde{\eta})\big\|_1 = \big\|\tilde{Q}^{\circ}_X(\tilde{\eta}) - Q_X(\eta)\big\|_1 = O_p\big(T^{-1/2}\big). \tag{4.35}$$

Then, we can give the rate of convergence of the third-stage estimator $\hat{\eta}$.

Theorem 4.1 *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7)*. Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$\|\hat{\eta}-\eta\| = O_p\big(T^{-1/2}\big). \tag{4.36}$$

Now, set

$$\tilde{S}_X(\tilde{\eta}) = T^{1/2}\Big\{\tilde{Q}_X(\tilde{\eta})\,\tilde{\Omega}_X(\tilde{\eta}) + \tilde{Q}^{\circ}_X(\tilde{\eta})\Big[\tilde{\Omega}^{\bullet}_X(\tilde{\eta}) - \tilde{\Omega}^{\circ}_X(\tilde{\eta})\Big]\Big\}, \tag{4.37}$$

$$S_X(\eta) = T^{1/2}Q_X(\eta)\,\Omega_X(\eta)\,. \tag{4.38}$$

These two vectors satisfy the following asymptotic equivalence.

Proposition 4.3 *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7)*. Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$\big\|\tilde{S}_X(\tilde{\eta}) - S_X(\eta)\big\| = O_p\big(T^{-1/2}\big)\,. \tag{4.39}$$

Finally, we establish the asymptotic normality of the third-stage GLS estimator.

Theorem 4.2 *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7)*. Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$T^{1/2}\big(\hat{\eta}-\eta\big) \underset{T\to\infty}{\overset{d}{\longrightarrow}} N\big[0, Q_X(\eta)\big], \quad T^{1/2}\big(\hat{\beta}-\beta\big) \underset{T\to\infty}{\overset{d}{\longrightarrow}} N\big[0, V_X(\eta)\big] \tag{4.40}$$

where $\hat{\beta} = R\hat{\eta}$ and $V_X(\eta) = RQ_X(\eta)\,R'$.

Its covariance matrix can then be estimated consistently with $\Big\{\sum_{t=1}^{T} Z_t(\tilde{\eta})\,\tilde{\Sigma}_u(\tilde{\eta})^{-1}Z_t(\tilde{\eta})'\Big\}^{-1}$. Further, the third-stage residuals $u_t(\hat{\eta})$ can either recursively be filtered using

$$u_t(\hat{\eta}) = \sum_{\tau=0}^{t-1}\Lambda_\tau(\hat{\eta})\Big[\hat{\Phi}_0 y_{t-\tau} - \sum_{i=1}^{\bar{p}}\hat{\Phi}_i y_{t-i-\tau} - \hat{\mu}_\Phi\Big] \tag{4.41}$$

or

$$u_t(\hat{\eta}) = y_t - \hat{\Phi}_0^{-1}\Big[\hat{\mu}_{\Phi} + \sum_{i=1}^{\bar{p}} \hat{\Phi}_i y_{t-i} + \sum_{j=1}^{\bar{p}} \hat{\Theta}_j u_{t-j}(\hat{\eta})\Big], \tag{4.42}$$

initiating with $u_t(\hat{\eta}) = 0$, $t \leq 0$, so that they satisfy

$$\hat{\Phi}(L)\, y_t = \hat{\mu}_{\Phi} + \hat{\Theta}(L)\, u_t(\hat{\eta}), \quad t = 1, \ldots, T, \tag{4.43}$$

where $\hat{\mu}_y = \hat{\Phi}(1)^{-1}\hat{\mu}_{\Phi}$ and $\hat{\Phi}(1) = \hat{\Phi}_0 - \sum_{i=1}^{\bar{p}} \hat{\Phi}_i$. Again, we would suggest filtering the residuals using

$$u_t(\hat{\eta}) = \epsilon_t(\tilde{\eta}, \hat{\eta}) - \big[Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta}, \hat{\eta})\big]'\big(\tilde{\eta} - \hat{\eta}\big), \tag{4.44}$$

initiating with $u_t(\hat{\eta}) = u_t(\tilde{\eta})$, for $0 < t \leq \bar{p}$, since the latter tends to deliver well behaved residuals in finite sample as they rely on the former [unlike (4.41) or (4.42)]. Hence, the third-stage innovation covariance matrix estimator is

$$\tilde{\Sigma}_u(\hat{\eta}) = \frac{1}{T}\sum_{t=1}^{T} u_t(\hat{\eta})\, u_t(\hat{\eta})'. \tag{4.45}$$

Its rate of convergence to Σ_u is given in the following proposition.

Proposition 4.4 *Let $\{y_t : t \in \mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in echelon form given by* (2.4)-(2.7)*. Then, under the assumptions* **2.1** *to* **2.4***, we have:*

$$\big\|\tilde{\Sigma}_u(\hat{\eta}) - \Sigma_u\big\| = O_p\big(T^{-1/2}\big). \tag{4.46}$$

5 Simulation study

In this section, we consider a Monte Carlo (MC) simulation to illustrate the finite sample performance of the proposed estimation method. We restrict our attention only to analyzing the finite sample properties of the fully efficient estimates, since the major contribution of the paper stands at that level. In particular, we consider a comparative study between our third-stage GLS estimator, described in (4.26), and those suggested by Hannan and Kavalieris (1984*b*), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. More especially, two variants of the proposed third-stage estimator were considered. The first one uses the two-stage GLS estimator, given in (3.23), as initial estimate, and the second one is based on the two-stage OLS estimator (using equationwise OLS estimation) considered in Dufour and Jouini (2005). The latter also has been used as initial estimator to obtain the GLS estimators described in Hannan and Kavalieris (1984*b*) and Reinsel et al. (1992), respectively, but, as mentioned earlier, using an alternative scheme for residual filtering. Clearly, the above two linear two-stage estimators are identical when the Kronecker indices characterizing the echelon canonical form are all equal. Further, to obtain the GLS estimator of Poskitt and Salau (1995) we have implemented their three-step procedure in full. It is worth noting that the considered GLS estimators in this simulation study are all asymptotically equivalent to ML estimates since they correspond to one iteration of Gauss-Newton algorithm starting from a $\sqrt{T}$-consistent estimator. Finally, we did not consider the ML estimation in the simulation for two reasons. First, its finite sample properties have been extensively studied in the literature and were found more or less satisfactory given the model in hand. Second, since the paper deals with efficient linear methods for estimating VARMA models, we attempted to investigate the finite sample performance of the main ones existing compared to the procedure suggested in this paper.

We simulate two bivariate stationary invertible Gaussian ARMA processes with constant terms and Kronecker indices $(1, 2)$ and $(2, 1)$, respectively, for sample sizes 100 and 200. Tables 1 to 4 report simulation results on the empirical means, the average errors and the mean squared errors (MSE) for each procedure. Moreover, these tables show the MSE ratios of the alternative efficient GLS estimators over the suggested third-stage GLS estimator TS1 [see tables' notes for more description]. These results are based on 1000 replications using GAUSS random number generator (version 3.2.37), and to avoid numerical problems that might be caused by the initialization effect, extra first 100 pseudo-data were generated then discarded. The trials associated with noninvertible processes are thrown and replaced with other ones. For all simulations, the rate of replacement did not exceed 5% in the worst case. [For how to obtain an invertible moving-average operator for echelon form VARMA model see Hannan and Kavalieris (1984*b*).] Further, the second-stage echelon form VARMA model parameters were estimated from models using, as regressors, the residuals obtained from a long-autoregression with lag-order fixed to two values; namely, the integer parts of $\ln T$ and $T^{-1/2}$ (since it has been recommended in the literature to choose the autoregression lag-order between these two extreme values). More specifically, the latter strategy has been considered to draw the effect of the choice of the long-autoregression lag-order on the finite sample quality of the echelon form estimates. In this simulation study, the error covariance matrix with $\sigma_{11} = 0.49$, $\sigma_{22} = 0.29$ and $\sigma_{11} = \sigma_{11} = -0.14$, is used for both models. Finally, the true parameter values of the simulated echelon form VARMA models as well as their related eigenvalues (describing their respective stationarity and invertibility conditions) are reported in the tables.

Simulation evidence shows for both models that, unlike TS1, TS2, and RBY methods where the estimates are characterized with small to moderate bias, HK and PS procedures provide estimates with substantial bias for sample size $T = 100$ [see upper panels of Tables 1 and 3]. These biases disappear with increasing sample size and/or lag-order of the first-stage long-autoregression [see lower panels of Tables 1 and 3, and Tables 2 and 4 for sample size $T = 200$]. It is suspected that the bias associated with PS procedure is attributed to the weighting matrix used in the computation of the estimates. Poskitt and Salau (1995) argued that the error term in the linear regression they considered follows a moving-average process of order $\bar{p}$, say $\xi_t = \sum_{j=0}^{\bar{p}} \Theta_j \varepsilon_{t-j}$ with $T^{-1}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t' = O_p\big(n_T/T\big)\Sigma_u$ [see Hannan and Kavalieris (1986) and Poskitt and Salau (1995)], but instead, they explicitly used $T^{-1}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t' = O_p\big(1\big)\Sigma_u$. The bias associated with HK procedure is due to two reasons. First, the weighting matrix used in the computation of the estimates does not correspond to what should be in the one iteration of the Gauss-Newton algorithm starting from the two-stage OLS estimates. In particular, they used the residual covariance estimator obtained at the third-stage of their procedure rather than the covariance estimator obtained from the new filtered residuals necessary for their fourth-stage estimation. Note that the latter has a convergence rate faster than the former. Second, as mentioned above, the new residuals are more or less satisfactory in finite sample given the way they are filtered. The RBY procedure uses the same filtering approach, however, compared to HK method, it delivers estimates with satisfactory finite sample properties. This is perhaps because it uses the right covariance matrix estimator (from the ML viewpoint) for the residuals in their GLS linear regression.

It is common knowledge that approximating VARMA models characterized with high persistence in their MA operators usually require long autoregressions with large number of lags, and vice versa. More especially, increasing the lag order n_T of an autoregression when approximating a VARMA model with less persistent MA operator would result in estimates with higher bias and/or MSE. This is exactly what we get with TS1 and TS2 procedures for the echelon form VARMA model with Kronecker indices $(2, 1)$ [see Tables 3 and 4]. For the same model and from the same tables, HK and PS procedures show that increasing the lag-order n_T, for a given sample size, seems to reduce the large bias and obtain parameter MSE that are decreasing for HK procedure and increasing for PS method. Further, one can see a slight increase in the bias characterizing RBY estimates, whereas the MSE of these estimates is exhibiting a mixed tendency. The same tendency characterizes all procedures when considering the echelon form VARMA$(1, 2)$ model as its largest eigenvalue, that is 0.813 (in norm), cannot be considered too high to consider the model as highly persistent in its MA part [see Tables 1 and 2]. Simulation results show that, overall, TS1, TS2 and PS procedures outperform those of HK and PS by far. To have a better idea on which procedure is providing estimates with better finite sample properties—as we may report that those of RBY procedure behave in a way similar to those associated with our suggested methods—we compute the ratios of the MSE of the parameters for each procedure with respect to those obtained with TS1 procedure. Obviously, with the exception of PS procedure all alternative methods provide estimates with MSE ratios greater than one. One should note that for the PS procedure the cases where the MSE ratios of the parameters are significantly less than unity are often matched with relative substantial bias. More precisely, these ratios are generally increasing with the sample size T and the lag-order n_T. Nevertheless, it is worth noting that TS1 has a slight advantage over TS2. So, choosing either TS1 or TS2 would have no big effect on the finite sample properties of the echelon form parameter estimates.

6 Conclusion

In this paper, we have proposed a new three-step linear estimation procedure for estimating stationary invertible echelon-form VARMA models. Our approach can easily be adapted to VARMAX models and extended to integrated and cointegrated VARMA models as well. The estimation method focuses on the echelon form as the latter tends to deliver relatively parsimonious parameterized models. Moreover, our procedure remains valid to other identifying issues such as final equations or any other restricted model for inference purposes.

Our proposed method allows for the presence of intercepts among the regressors—which, in contrast to previous works, looks more realistic—and provides a simplified general and compact standard form for the echelon-form parameter estimates that are easier to compute than those of Hannan and Kavalieris (1984*b*). This may be more advisable and tractable in practice. Further, we have extended the results of Dufour and Jouini (2005) for the two-step estimation method to derive the asymptotic distribution of the GLS estimators in the case of strong WN, since to our knowledge it has not been stated yet anywhere. Moreover, we gave its covariance estimator. In addition, we have proposed a new recursive linear method to filter the new residuals necessary to our third-stage GLS estimation. These residuals are function of the first-stage long-autoregression and second-stage residuals as well. Hence, taking into account the truncation error associated with the approximate long autoregression used in the first stage. Also, we have provided a theoretical justification for implementing a third-stage linear regression without any prior assumption of the actual distribution of the errors, unlike preceding works. We have shown that the resulting GLS estimators, for which we have derived its asymptotic distribution under strong WN and gave its covariance estimator, provide an appealing and intuitive interpretation of nonlinear estimation methods such as NGLS and ML. Thus, achieving efficiency with Gaussian errors. This shows the asymptotic equivalence between our third-stage and ML estimators. However, the finite sample properties of our estimates are not the same as those of ML estimators. Although our three-step estimation procedure is asymptotically equivalent to that of Hannan and Kavalieris (1984*b*), the estimates of the asymptotic covariances of the echelon-form parameters we have given for the second and third stages as well, are simple and easy to use for inference purposes, especially with simulation-based techniques such as bootstrap methods.

Further, simulation evidence has shown that our proposed GLS estimation methods outperform in most cases those proposed in the literature. Also, the finite sample properties of the echelon form VARMA estimates are sensitive to the lag-order of the first-stage long-autoregression when approximating the true innovations. This would suggest that more investigation should be made on this issue to provide more efficient algorithms in specifying the lag-order of the first-stage long autoregression. This lag-order may have an effect on the finite sample quality of the echelon form parameter estimates, and thus on their implied forecasts and impulse response functions subject of interest in most applied researchers.

Table 1: Estimated echelon VARMA model with Kronecker indices (1,2) and sample size T=100: A comparative simulation study on the finite sample properties of alternative fully efficient GLS estimators

		Empirical mean					Average error					Mean Squared Error					MSE ratio			
Coeff	Value	TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS2	HK	RBY	PS
$\mu_{\Phi,1}$	0.00	-0.004	-0.004	-0.007	-0.005	-0.004	0.004	0.004	0.007	0.005	0.004	0.200	0.201	0.267	0.197	0.187	1.003	1.337	0.985	0.935
$\mu_{\Phi,2}$	0.00	-0.004	-0.004	-0.005	-0.003	-0.003	0.004	0.004	0.005	0.003	0.003	0.151	0.152	0.177	0.152	0.124	1.007	1.174	1.006	0.820
$\phi_{11,1}$	1.20	1.177	1.177	1.192	1.180	1.189	0.022	0.022	0.007	0.019	0.010	0.056	0.056	0.078	0.056	0.052	1.001	1.378	1.004	0.930
$\phi_{12,1}$	0.24	0.239	0.239	0.242	0.240	0.248	0.000	0.000	0.002	0.000	0.008	0.048	0.048	0.060	0.048	0.045	1.003	1.263	1.005	0.941
$\phi_{22,1}$	0.40	0.400	0.396	0.375	0.398	0.532	0.000	0.003	0.024	0.001	**0.132**	0.111	0.104	0.127	0.116	0.121	0.943	1.144	1.049	1.096
$\phi_{21,2}$	-0.90	-0.906	-0.909	-0.916	-0.905	-0.811	0.006	0.009	0.016	0.005	**0.088**	0.078	0.075	0.088	0.079	0.085	0.960	1.133	1.019	1.086
$\phi_{22,2}$	-0.27	-0.267	-0.265	-0.256	-0.264	-0.315	0.002	0.004	0.013	0.005	0.045	0.067	0.065	0.080	0.070	0.066	0.969	1.204	1.055	0.996
$\theta_{11,1}$	0.80	0.786	0.786	0.774	0.789	0.589	0.013	0.013	0.025	0.010	**0.210**	0.094	0.095	0.104	0.097	0.058	1.008	1.109	1.036	**0.615**
$\theta_{21,1}$	0.50	0.507	0.503	0.476	0.506	0.584	0.007	0.003	0.023	0.006	**0.084**	0.090	0.090	0.103	0.096	0.081	0.990	1.135	1.060	**0.891**
$\theta_{12,1}$	0.40	0.383	0.385	0.303	0.376	0.188	0.016	0.014	**0.096**	0.023	**0.211**	0.115	0.118	0.144	0.128	0.110	1.021	1.250	1.108	**0.953**
$\theta_{22,1}$	0.40	0.374	0.380	0.358	0.383	0.238	0.025	0.019	0.041	0.016	**0.161**	0.131	0.123	0.145	0.145	0.141	0.943	1.108	1.104	1.077
$\theta_{21,2}$	0.34	0.313	0.320	0.348	0.315	0.012	0.026	0.019	0.008	0.024	**0.327**	0.162	0.152	0.168	0.170	0.172	0.937	1.034	1.049	1.062
$\theta_{22,2}$	0.85	0.774	0.781	0.653	0.770	0.444	0.075	0.068	**0.196**	0.079	**0.405**	0.143	0.140	0.154	0.147	0.105	0.979	1.072	1.023	**0.731**
$\mu_{\Phi,1}$	0.00	0.002	0.003	0.002	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.206	0.208	0.266	0.211	0.199	1.009	1.291	1.025	0.970
$\mu_{\Phi,2}$	0.00	0.005	0.005	0.002	0.004	0.003	0.005	0.005	0.002	0.004	0.003	0.169	0.168	0.210	0.169	0.155	0.994	1.241	1.003	0.919
$\phi_{11,1}$	1.20	1.175	1.174	1.175	1.176	1.117	0.024	0.025	0.024	0.023	0.022	0.056	0.057	0.076	0.060	0.055	1.008	1.337	1.055	0.981
$\phi_{12,1}$	0.24	0.239	0.239	0.236	0.240	0.240	0.000	0.000	0.003	0.000	0.000	0.046	0.047	0.073	0.050	0.046	1.007	1.562	1.068	0.988
$\phi_{22,1}$	0.40	0.393	0.396	0.403	0.403	0.420	0.006	0.003	0.003	0.003	0.020	0.105	0.102	0.110	0.111	0.105	0.969	1.047	1.057	1.001
$\phi_{21,2}$	-0.90	-0.914	-0.912	-0.905	-0.907	-0.896	0.014	0.012	0.005	0.007	0.003	0.079	0.077	0.087	0.079	0.078	0.964	1.098	0.995	0.986
$\phi_{22,2}$	-0.27	-0.266	-0.267	-0.271	-0.270	-0.277	0.003	0.002	0.001	0.000	0.007	0.064	0.064	0.075	0.070	0.063	0.997	1.175	1.091	0.985
$\theta_{11,1}$	0.80	0.787	0.790	0.775	0.799	0.746	0.012	0.009	0.024	0.000	0.053	0.099	0.100	0.099	0.103	0.091	1.010	0.998	1.039	0.926
$\theta_{21,1}$	0.50	0.499	0.498	0.505	0.506	0.509	0.000	0.001	0.005	0.006	0.009	0.090	0.090	0.096	0.095	0.093	0.999	1.064	1.048	1.031
$\theta_{12,1}$	0.40	0.388	0.390	0.374	0.396	0.359	0.011	0.009	0.025	0.003	0.040	0.121	0.124	0.159	0.135	0.127	1.020	1.310	1.113	1.046
$\theta_{22,1}$	0.40	0.371	0.371	0.367	0.373	0.343	0.028	0.028	0.032	0.026	0.056	0.122	0.119	0.122	0.129	0.113	0.972	0.997	1.052	0.925
$\theta_{21,2}$	0.34	0.320	0.317	0.300	0.313	0.273	0.019	0.022	0.039	0.026	0.066	0.159	0.154	0.156	0.170	0.177	0.972	0.981	1.072	1.117
$\theta_{22,2}$	0.85	0.799	0.798	0.773	0.816	0.759	0.050	0.051	0.076	0.033	**0.090**	0.139	0.139	0.134	0.149	0.127	1.004	0.964	1.077	**0.913**

Note – These estimates are obtained from 1000 replications. TS1, TS2 stand for the respective proposed third-stage GLS estimators based on the two-stage GLS and the two-stage OLS estimators. While HK, RBY and PS stand for the fully efficient GLS estimators suggested by Hannan and Kavalieris (1984*b*), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. The eigenvalues of the model are real 0.900, 0.400 and 0.300 for the autoregressive (AR) operator, and real 0.824 and conjugate -0.188∓0.790i (0.813 in norm) for the moving-average (MA) operator. Recall that the number of eigenvalues in each of the AR and MA operators is equal to the McMillan degree of the model. That is, the sum of the Kronecker indices. In the upper panel $n_T = [\ln T]$ was used, whereas is the lower panel a value of $n_T = [T^{1/2}]$ has been used. Finally, $[x]$ stands for the integer less or equal to x.

Table 2: Estimated echelon VARMA model with Kronecker indices (1,2) and sample size T=200: A comparative simulation study on the finite sample properties of alternative fully efficient GLS estimators

		Empirical mean					Average error					Mean Squared Error					MSE. ratio			
Coeff	Value	TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS2	HK	RBY	PS
$\mu_{\Phi,1}$	0.00	0.001	0.001	0.003	0.001	0.000	0.001	0.001	0.003	0.001	0.000	0.114	0.114	0.177	0.116	0.110	1.001	1.547	1.016	0.961
$\mu_{\Phi,2}$	0.00	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000	0.000	0.094	0.094	0.119	0.093	0.094	1.003	1.268	0.996	1.006
$\phi_{11,1}$	1.20	1.187	1.187	1.195	1.189	1.188	0.012	0.012	0.004	0.010	0.011	0.036	0.036	0.046	0.037	0.035	1.000	1.271	1.032	0.980
$\phi_{12,1}$	0.24	0.238	0.238	0.253	0.240	0.237	0.001	0.001	0.013	0.000	0.002	0.030	0.030	0.044	0.032	0.030	0.999	1.424	1.048	0.989
$\phi_{22,1}$	0.40	0.399	0.397	0.411	0.404	0.387	0.000	0.002	0.011	0.004	0.012	0.062	0.060	0.073	0.065	0.075	0.967	1.175	1.053	1.216
$\phi_{21,2}$	-0.90	-0.905	-0.906	-0.896	-0.901	-0.903	0.005	0.006	0.003	0.001	0.003	0.044	0.043	0.053	0.046	0.051	0.978	1.202	1.035	1.162
$\phi_{22,2}$	-0.27	-0.269	-0.268	-0.266	-0.271	-0.254	0.000	0.001	0.003	0.001	0.015	0.040	0.039	0.045	0.041	0.044	0.984	1.145	1.037	1.100
$\theta_{11,1}$	0.80	0.797	0.799	0.764	0.797	0.688	0.002	0.000	0.035	0.002	**0.111**	0.060	0.061	0.070	0.063	0.053	1.019	1.164	1.058	**0.880**
$\theta_{21,1}$	0.50	0.499	0.499	0.505	0.505	0.528	0.000	0.000	0.005	0.005	0.028	0.058	0.058	0.063	0.062	0.059	0.999	1.080	1.059	1.010
$\theta_{12,1}$	0.40	0.394	0.396	0.318	0.390	0.306	0.005	0.003	**0.081**	0.009	**0.093**	0.075	0.076	0.107	0.085	0.090	1.009	1.427	1.127	1.194
$\theta_{22,1}$	0.40	0.388	0.391	0.370	0.388	0.376	0.011	0.008	0.029	0.011	0.023	0.072	0.069	0.075	0.080	0.077	0.965	1.041	1.110	1.064
$\theta_{21,2}$	0.34	0.331	0.334	0.298	0.323	0.276	0.008	0.005	0.041	0.016	0.063	0.097	0.094	0.111	0.103	0.120	0.968	1.140	1.058	1.236
$\theta_{22,2}$	0.85	0.819	0.821	0.731	0.818	0.643	0.030	0.028	**0.118**	0.031	**0.206**	0.082	0.082	0.104	0.089	0.073	1.002	1.266	1.082	**0.896**
$\mu_{\Phi,1}$	0.00	0.004	0.004	0.004	0.003	0.004	0.004	0.004	0.004	0.003	0.004	0.115	0.116	0.170	0.120	0.112	1.006	1.472	1.037	0.969
$\mu_{\Phi,2}$	0.00	0.005	0.005	0.005	0.004	0.004	0.005	0.005	0.005	0.004	0.004	0.093	0.093	0.115	0.097	0.090	1.000	1.233	1.036	0.968
$\phi_{11,1}$	1.20	1.188	1.188	1.187	1.188	1.189	0.011	0.011	0.012	0.011	0.010	0.037	0.037	0.051	0.038	0.037	1.005	1.376	1.045	0.994
$\phi_{12,1}$	0.24	0.239	0.239	0.238	0.239	0.240	0.000	0.000	0.001	0.000	0.000	0.031	0.031	0.045	0.033	0.032	1.001	1.432	1.057	1.012
$\phi_{22,1}$	0.40	0.400	0.400	0.400	0.401	0.413	0.000	0.000	0.000	0.001	0.013	0.062	0.060	0.063	0.063	0.066	0.975	1.017	1.020	1.061
$\phi_{21,2}$	-0.90	-0.905	-0.905	-0.905	-0.904	-0.897	0.005	0.005	0.005	0.004	0.002	0.047	0.046	0.048	0.048	0.049	0.985	1.029	1.029	1.042
$\phi_{22,2}$	-0.27	-0.270	-0.270	-0.271	-0.270	-0.275	0.000	0.000	0.001	0.000	0.005	0.039	0.039	0.043	0.041	0.040	0.985	1.084	1.029	1.004
$\theta_{11,1}$	0.80	0.801	0.804	0.800	0.811	0.780	0.001	0.004	0.000	0.011	0.019	0.063	0.064	0.064	0.066	0.058	1.013	1.018	1.047	0.923
$\theta_{21,1}$	0.50	0.498	0.498	0.497	0.502	0.505	0.001	0.001	0.002	0.002	0.005	0.060	0.059	0.064	0.062	0.063	0.997	1.076	1.035	1.052
$\theta_{12,1}$	0.40	0.400	0.402	0.397	0.408	0.382	0.000	0.002	0.002	0.008	0.017	0.079	0.079	0.086	0.083	0.081	1.004	1.095	1.050	1.034
$\theta_{22,1}$	0.40	0.389	0.391	0.389	0.393	0.379	0.010	0.008	0.010	0.006	0.020	0.071	0.070	0.074	0.072	0.068	0.983	1.032	1.010	0.959
$\theta_{21,2}$	0.34	0.330	0.331	0.329	0.334	0.309	0.009	0.008	0.010	0.005	0.030	0.095	0.094	0.097	0.100	0.105	0.983	1.014	1.045	1.097
$\theta_{22,2}$	0.85	0.828	0.829	0.822	0.840	0.811	0.021	0.020	0.027	0.009	0.038	0.085	0.084	0.082	0.089	0.080	0.985	0.961	1.037	0.935

Note – These estimates are obtained from 1000 replications. TS1, TS2 stand for the respective proposed third-stage GLS estimators based on the two-stage GLS and the two-stage OLS estimators. While HK, RBY and PS stand for the fully efficient GLS estimators suggested by Hannan and Kavalieris (1984*b*), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. The eigenvalues of the model are real 0.900, 0.400 and 0.300 for the autoregressive (AR) operator, and real 0.824 and conjugate -0.188∓0.790i (0.813 in norm) for the moving-average (MA) operator. Recall that the number of eigenvalues in each of the AR and MA operators is equal to the McMillan degree of the model. That is, the sum of the Kronecker indices. In the upper panel $n_T = [\ln T]$ was used, whereas is the lower panel a value of $n_T = [T^{1/2}]$ has been used. Finally, $[x]$ stands for the integer less or equal to x.

Table 3: Estimated echelon VARMA model with Kronecker indices (2,1) and sample size T=100: A comparative simulation study on the finite sample properties of alternative fully efficient GLS estimators

		Empirical mean					Average error					Mean Squared Error					MSE ratio			
Coeff	Value	TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS2	HK	RBY	PS
$\mu_{\Phi,1}$	0.00	-0.001	-0.001	0.001	0.000	-0.001	0.001	0.001	0.001	0.000	0.001	0.158	0.159	0.237	0.166	0.164	1.007	1.503	1.053	1.040
$\mu_{\Phi,2}$	0.00	-0.004	-0.004	-0.006	-0.005	-0.001	0.004	0.004	0.006	0.005	0.001	0.187	0.189	0.199	0.190	0.180	1.006	1.059	1.015	0.958
$\phi_{21,0}$	0.50	0.496	0.495	0.505	0.494	0.496	0.003	0.004	0.005	0.005	0.003	0.033	0.033	0.062	0.035	0.033	1.016	1.876	1.087	1.004
$\phi_{11,1}$	1.80	1.797	1.798	1.830	1.799	1.810	0.002	0.001	0.030	0.000	0.010	0.034	0.034	0.054	0.039	0.033	1.010	1.586	1.136	0.960
$\phi_{21,1}$	-0.40	-0.362	-0.358	-0.383	-0.354	-0.361	0.037	0.041	0.016	0.045	0.038	0.089	0.091	0.178	0.102	0.087	1.021	2.002	1.147	0.973
$\phi_{22,1}$	0.80	0.735	0.730	0.761	0.724	0.732	0.064	0.069	0.038	0.075	0.067	0.129	0.131	0.249	0.148	0.125	1.017	1.931	1.143	0.966
$\phi_{11,2}$	-0.36	-0.365	-0.367	-0.517	-0.371	-0.413	0.005	0.007	**0.157**	0.011	0.053	0.111	0.111	0.188	0.129	0.103	1.006	1.699	1.166	0.927
$\phi_{12,2}$	-0.90	-0.887	-0.884	-0.648	-0.878	-0.816	0.012	0.015	**0.251**	0.021	**0.083**	0.168	0.168	0.288	0.195	0.155	1.000	1.708	1.159	**0.920**
$\theta_{11,1}$	0.33	0.274	0.274	0.305	0.286	0.211	0.055	0.055	0.024	0.043	**0.118**	0.117	0.119	0.213	0.137	0.109	1.012	1.811	1.162	**0.925**
$\theta_{21,1}$	-0.18	-0.163	-0.163	-0.265	-0.180	-0.092	0.016	0.016	**0.085**	0.000	**0.087**	0.107	0.108	0.191	0.128	0.098	1.009	1.777	1.194	**0.919**
$\theta_{12,1}$	-0.20	-0.221	-0.222	-0.266	-0.214	-0.311	0.021	0.022	0.066	0.014	**0.111**	0.139	0.142	0.170	0.153	0.128	1.018	1.222	1.098	**0.917**
$\theta_{22,1}$	-0.40	-0.327	-0.319	-0.273	-0.328	-0.211	0.072	0.080	**0.126**	0.071	**0.188**	0.161	0.165	0.312	0.189	0.136	1.028	1.936	1.176	**0.849**
$\theta_{11,2}$	-0.20	-0.261	-0.264	-0.101	-0.255	-0.270	0.061	0.064	**0.098**	0.055	0.070	0.123	0.124	0.235	0.147	0.109	1.007	1.903	1.194	0.880
$\theta_{12,2}$	0.92	0.895	0.887	0.585	0.904	0.728	0.024	0.032	**0.334**	0.015	**0.191**	0.203	0.208	0.335	0.243	0.165	1.022	1.647	1.191	**0.813**
$\mu_{\Phi,1}$	0.00	0.000	0.000	-0.005	-0.000	0.001	0.000	0.000	0.005	0.000	0.001	0.173	0.173	0.198	0.175	0.183	1.002	1.143	1.011	1.060
$\mu_{\Phi,2}$	0.00	-0.000	-0.000	0.002	0.000	-0.001	0.000	0.000	0.002	0.000	0.001	0.208	0.208	0.217	0.209	0.208	0.999	1.042	1.004	1.001
$\phi_{21,0}$	0.50	0.499	0.498	0.500	0.498	0.506	0.000	0.001	0.000	0.001	0.006	0.040	0.040	0.048	0.041	0.041	1.004	1.190	1.011	1.017
$\phi_{11,1}$	1.80	1.798	1.797	1.799	1.794	1.805	0.001	0.002	0.000	0.005	0.005	0.038	0.040	0.046	0.043	0.039	1.039	1.198	1.131	1.020
$\phi_{21,1}$	-0.40	-0.356	-0.352	-0.360	-0.352	-0.369	0.043	0.047	0.039	0.047	0.030	0.107	0.108	0.136	0.111	0.108	1.008	1.264	1.034	1.003
$\phi_{22,1}$	0.80	0.718	0.713	0.723	0.714	0.730	**0.081**	**0.086**	0.076	**0.085**	0.069	0.151	0.152	0.193	0.157	0.154	1.007	1.277	1.041	1.020
$\phi_{11,2}$	-0.36	-0.382	-0.380	-0.394	-0.371	-0.416	0.022	0.020	0.034	0.011	0.056	0.113	0.117	0.141	0.127	0.114	1.036	1.248	1.129	1.011
$\phi_{12,2}$	-0.90	-0.853	-0.856	-0.833	-0.868	-0.801	0.046	0.043	0.066	0.031	**0.098**	0.167	0.171	0.208	0.184	0.168	1.026	1.246	1.107	1.007
$\theta_{11,1}$	0.33	0.268	0.268	0.278	0.277	0.254	0.061	0.061	0.051	0.052	0.075	0.124	0.127	0.173	0.136	0.130	1.020	1.394	1.093	1.048
$\theta_{21,1}$	-0.18	-0.166	-0.166	-0.183	-0.172	-0.179	0.013	0.013	0.003	0.007	0.000	0.123	0.122	0.162	0.133	0.123	0.993	1.321	1.082	1.000
$\theta_{12,1}$	-0.20	-0.231	-0.230	-0.232	-0.225	-0.245	0.031	0.030	0.032	0.025	0.045	0.145	0.149	0.165	0.155	0.151	1.027	1.136	1.072	1.043
$\theta_{22,1}$	-0.40	-0.304	-0.299	-0.302	-0.313	-0.310	**0.095**	**0.100**	**0.097**	**0.086**	**0.089**	0.191	0.194	0.239	0.207	0.193	1.017	1.249	1.081	1.012
$\theta_{11,2}$	-0.20	-0.262	-0.263	-0.240	-0.268	-0.234	0.062	0.063	0.040	0.068	0.034	0.129	0.134	0.183	0.142	0.128	1.038	1.420	1.103	0.993
$\theta_{12,2}$	0.92	0.848	0.846	0.812	0.876	0.797	0.071	0.073	**0.107**	0.043	**0.122**	0.214	0.228	0.242	0.244	0.203	1.062	1.128	1.136	**0.946**

Note – These estimates are obtained from 1000 replications. TS1, TS2 stand for the respective proposed third-stage GLS estimators based on the two-stage GLS and the two-stage OLS estimators. While HK, RBY and PS stand for the fully efficient GLS estimators suggested by Hannan and Kavalieris (1984*b*), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. The eigenvalues of the model are real 0.800 and a double root 0.900 for the autoregressive (AR) operator, and real -0.530 and conjugate -0.350∓0.584i (0.681 in norm) for the moving-average (MA) operator. Recall that the number of eigenvalues in each of the AR and MA operators is equal to the McMillan degree of the model. That is, the sum of the Kronecker indices. In the upper panel $n_T = [\ln T]$ was used, whereas is the lower panel a value of $n_T = [T^{1/2}]$ has been used. Finally, $[x]$ stands for the integer less or equal to x.

Table 4: Estimated echelon VARMA model with Kronecker indices (2,1) and sample size T=200: A comparative simulation study on the finite sample properties of alternative fully efficient GLS estimators

		Empirical mean					Average error					Mean Squared Error					MSE ratio			
Coeff	Value	TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS2	HK	RBY	PS
$\mu_{\Phi,1}$	0.00	-0.000	-0.000	-0.002	-0.000	-0.000	0.000	0.000	0.002	0.000	0.000	0.078	0.079	0.103	0.080	0.081	1.002	1.306	1.019	1.028
$\mu_{\Phi,2}$	0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.083	0.083	0.084	0.083	0.081	1.003	1.013	0.999	0.978
$\phi_{21,0}$	0.50	0.498	0.498	0.502	0.498	0.498	0.001	0.001	0.002	0.001	0.001	0.019	0.019	0.023	0.020	0.020	1.001	1.182	1.030	1.045
$\phi_{11,1}$	1.80	1.797	1.798	1.816	1.798	1.804	0.002	0.001	0.016	0.001	0.004	0.023	0.023	0.031	0.025	0.023	1.004	1.340	1.082	0.994
$\phi_{21,1}$	-0.40	-0.383	-0.382	-0.387	-0.382	-0.383	0.016	0.017	0.012	0.017	0.016	0.056	0.057	0.065	0.059	0.060	1.006	1.158	1.052	1.064
$\phi_{22,1}$	0.80	0.771	0.769	0.772	0.769	0.772	0.028	0.030	0.027	0.030	0.027	0.083	0.083	0.095	0.087	0.087	1.006	1.147	1.051	1.057
$\phi_{11,2}$	-0.36	-0.359	-0.360	-0.436	-0.361	-0.385	0.000	0.000	0.076	0.001	0.025	0.073	0.073	0.112	0.081	0.070	1.007	1.539	1.111	0.967
$\phi_{12,2}$	-0.90	-0.897	-0.897	-0.780	-0.894	-0.859	0.002	0.002	**0.119**	0.005	0.040	0.109	0.109	0.171	0.121	0.104	1.007	1.568	1.109	0.957
$\theta_{11,1}$	0.33	0.304	0.303	0.297	0.310	0.278	0.025	0.026	0.032	0.019	0.051	0.076	0.077	0.092	0.086	0.074	1.007	1.208	1.136	0.978
$\theta_{21,1}$	-0.18	-0.170	-0.169	-0.188	-0.176	-0.127	0.009	0.010	0.008	0.003	0.052	0.069	0.069	0.098	0.077	0.064	1.009	1.429	1.116	0.933
$\theta_{12,1}$	-0.20	-0.205	-0.206	-0.253	-0.200	-0.247	0.005	0.006	0.053	0.000	0.047	0.095	0.096	0.100	0.104	0.094	1.014	1.053	1.096	0.991
$\theta_{22,1}$	-0.40	-0.374	-0.370	-0.296	-0.378	-0.310	0.025	0.029	**0.103**	0.021	**0.089**	0.104	0.105	0.124	0.118	0.103	1.013	1.198	1.133	0.994
$\theta_{11,2}$	-0.20	-0.226	-0.228	-0.200	-0.229	-0.237	0.026	0.028	0.000	0.029	0.037	0.078	0.079	0.108	0.089	0.076	1.011	1.383	1.147	0.982
$\theta_{12,2}$	0.92	0.915	0.913	0.727	0.919	0.814	0.004	0.006	**0.192**	0.000	**0.105**	0.136	0.138	0.197	0.155	0.122	1.012	1.442	1.133	**0.891**
$\mu_{\Phi,1}$	0.00	0.003	0.002	0.003	0.002	0.003	0.003	0.002	0.003	0.002	0.003	0.082	0.082	0.082	0.081	0.084	1.005	1.003	0.999	1.028
$\mu_{\Phi,2}$	0.00	-0.003	-0.003	-0.003	-0.004	-0.004	0.003	0.003	0.003	0.004	0.004	0.089	0.089	0.089	0.089	0.090	1.005	1.001	1.005	1.015
$\phi_{21,0}$	0.50	0.498	0.498	0.498	0.498	0.500	0.001	0.001	0.001	0.001	0.000	0.021	0.021	0.021	0.021	0.023	1.001	1.004	1.000	1.057
$\phi_{11,1}$	1.80	1.799	1.798	1.798	1.798	1.805	0.000	0.001	0.001	0.001	0.005	0.024	0.025	0.025	0.025	0.025	1.017	1.025	1.024	1.015
$\phi_{21,1}$	-0.40	-0.383	-0.381	-0.382	-0.381	-0.385	0.016	0.018	0.017	0.018	0.014	0.060	0.060	0.060	0.060	0.065	1.003	1.001	1.002	1.090
$\phi_{22,1}$	0.80	0.771	0.768	0.770	0.768	0.770	0.028	0.031	0.029	0.031	0.029	0.085	0.086	0.085	0.085	0.094	1.002	0.999	1.001	1.068
$\phi_{11,2}$	-0.36	-0.367	-0.366	-0.364	-0.362	-0.393	0.007	0.006	0.004	0.002	0.033	0.075	0.077	0.077	0.078	0.076	1.021	1.030	1.037	1.007
$\phi_{12,2}$	-0.90	-0.884	-0.886	-0.889	-0.891	-0.843	0.015	0.013	0.010	0.008	0.056	0.111	0.113	0.114	0.115	0.111	1.020	1.030	1.038	1.004
$\theta_{11,1}$	0.33	0.307	-0.308	0.308	0.309	0.301	0.022	0.021	0.021	0.020	0.028	0.078	0.079	0.079	0.080	0.080	1.013	1.015	1.024	1.029
$\theta_{21,1}$	-0.18	-0.172	-0.171	-0.170	-0.171	-0.183	0.007	0.008	0.009	0.008	0.003	0.072	0.072	0.073	0.073	0.076	0.998	1.017	1.016	1.049
$\theta_{12,1}$	-0.20	-0.211	-0.211	-0.212	-0.210	-0.221	0.011	0.011	0.012	0.010	0.021	0.095	0.097	0.097	0.098	0.100	1.021	1.016	1.034	1.049
$\theta_{22,1}$	-0.40	-0.366	-0.361	-0.362	-0.366	-0.354	0.033	0.038	0.037	0.033	0.045	0.112	0.114	0.113	0.115	0.118	1.018	1.006	1.028	1.048
$\theta_{11,2}$	-0.20	-0.226	-0.227	-0.228	-0.231	-0.212	0.026	0.027	0.028	0.031	0.012	0.079	0.080	0.080	0.082	0.079	1.011	1.015	1.039	1.006
$\theta_{12,2}$	0.92	0.899	0.898	0.899	0.913	0.856	0.020	0.021	0.020	0.006	0.063	0.139	0.143	0.143	0.146	0.135	1.029	1.029	1.053	0.974

Note – These estimates are obtained from 1000 replications. TS1, TS2 stand for the respective proposed third-stage GLS estimators based on the two-stage GLS and the two-stage OLS estimators. While HK, RBY and PS stand for the fully efficient GLS estimators suggested by Hannan and Kavalieris (1984*b*), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. The eigenvalues of the model are real 0.800 and a double root 0.900 for the autoregressive (AR) operator, and real -0.530 and conjugate -0.350∓0.584i (0.681 in norm) for the moving-average (MA) operator. Recall that the number of eigenvalues in each of the AR and MA operators is equal to the McMillan degree of the model. That is, the sum of the Kronecker indices. In the upper panel $n_T = [\ln T]$ was used, whereas is the lower panel a value of $n_T = [T^{1/2}]$ has been used. Finally, $[x]$ stands for the integer less or equal to x.

A Appendix: Proofs

PROOF OF PROPOSITION 3.1 Note first that

$$\mathsf{E}\left\|\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right\|_1^2 \leq \mathsf{E}\left\|\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right\|^2 \tag{A.1}$$

where $\tilde{\Gamma}_Y(n_T)=T^{-1}\sum_{t=1}^{T}Y_t(n_T)Y_t(n_T)'$ and $\Gamma_Y(n_T)=\mathsf{E}\left[Y_t(n_T)Y_t(n_T)'\right]$. Then it can be seen that

$$\mathsf{E}\left\|\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right\|^2=2\sum_{\tau=1}^{n_T}\mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left(y_{t-\tau}-\mu_y\right)\right\|^2+\sum_{\tau_1=1}^{n_T}\sum_{\tau_2=1}^{n_T}\mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left[y_{t-\tau_1}y'_{t-\tau_2}-\Gamma_y(\tau_1-\tau_2)\right]\right\|^2,$$

with $\Gamma_y(\tau_1-\tau_2)=E\left[y_{t-\tau_1}y'_{t-\tau_2}\right]$. Using (2.2), it follows from Hannan (1970, Chapter 4) that

$$\mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left(y_{t-\tau}-\mu_y\right)\right\|^2=O\left(\frac{k}{T}\right),\ \mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left[y_{t-\tau_1}y'_{t-\tau_2}-\Gamma_y(\tau_1-\tau_2)\right]\right\|^2=O\left(\frac{k^2}{T}\right), \tag{A.2}$$

hence

$$\mathsf{E}\left\|\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right\|^2=O\left(\frac{kn_T}{T}\right)+O\left(\frac{k^2n_T^2}{T}\right)=O\left(\frac{k^2n_T^2}{T}\right). \tag{A.3}$$

Further, we have

$$\begin{aligned}\left\|\tilde{\Gamma}_Y(n_T)^{-1}-\Gamma_Y(n_T)^{-1}\right\| &= \left\|\tilde{\Gamma}_Y(n_T)^{-1}\left[\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right]\Gamma_Y(n_T)^{-1}\right\| \\ &\leq \left\|\tilde{\Gamma}_Y(n_T)^{-1}\right\|\left\|\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right\|\left\|\Gamma_Y(n_T)^{-1}\right\|,\end{aligned} \tag{A.4}$$

with

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1}\right\|\leq\left\|\Gamma_Y(n_T)^{-1}\right\|+\left\|\tilde{\Gamma}_Y(n_T)^{-1}-\Gamma_Y(n_T)^{-1}\right\| \tag{A.5}$$

where as in the univariate case $\mathsf{E}\left\|\Gamma_Y(n_T)^{-1}\right\|$ is uniformly bounded by a positive constant for all n_T, hence $\left\|\Gamma_Y(n_T)^{-1}\right\|=O_p(1)$; see Berk (1974, page 491). Moreover, $\left\|\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right\|\left\|\Gamma_Y(n_T)^{-1}\right\|<1$ (an event whose probability converges to one as $T\to\infty$). Therefore

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1}-\Gamma_Y(n_T)^{-1}\right\|\leq\frac{\left\|\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right\|\left\|\Gamma_Y(n_T)^{-1}\right\|^2}{1-\left\|\tilde{\Gamma}_Y(n_T)-\Gamma_Y(n_T)\right\|\left\|\Gamma_Y(n_T)^{-1}\right\|}=O_p\left(\frac{kn_T}{T^{1/2}}\right) \tag{A.6}$$

and finally

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1}-\Gamma_Y(n_T)^{-1}\right\|_1=O_p\left(\frac{kn_T}{T^{1/2}}\right). \tag{A.7}$$

■

PROOF OF THEOREM 3.1 Recall that $\tilde{\Pi}(n_T)=\tilde{W}_Y(n_T)\tilde{\Gamma}_Y(n_T)^{-1}$, where $\tilde{W}_Y(n_T)=T^{-1}\sum_{t=1}^{T}y_tY_t(n_T)'$ and $y_t=\Pi(n_T)Y_t(n_T)+u_t(n_T)$. Set

$$U_1(n_T)=T^{-1}\sum_{t=1}^{T}\left[u_t(n_T)-u_t\right]Y_t(n_T)',\ U_2(n_T)=T^{-1}\sum_{t=1}^{T}u_tY_t(n_T)' \tag{A.8}$$

Then

$$\left\|\tilde{\Pi}(n_T)-\Pi(n_T)\right\|\leq\left\{\left\|U_1(n_T)\right\|+\left\|U_2(n_T)\right\|\right\}\left\|\tilde{\Gamma}_Y(n_T)^{-1}\right\| \tag{A.9}$$

where, by Assumption **2.3**

$$\left\|\tilde{\Gamma}_Y(n_T)^{-1}\right\|=O_p(1)+O_p\left(\frac{kn_T}{T^{1/2}}\right)=O_p(1). \tag{A.10}$$

Note that

$$\mathsf{E}\left\|U_1(n_T)\right\|=\mathsf{E}\left\|T^{-1}\sum_{t=1}^{T}\left[u_t(n_T)-u_t\right]Y_t(n_T)'\right\|\leq T^{-1}\sum_{t=1}^{T}\left\{\mathsf{E}\left\|u_t(n_T)-u_t\right\|^2\right\}^{1/2}\left\{\mathsf{E}\left\|Y_t(n_T)\right\|^2\right\}^{1/2}, \tag{A.11}$$

with

$$\mathsf{E}\left\|Y_t(n_T)\right\|^2 = 1 + n_T \mathrm{tr}\left[\Gamma_y(0)\right] \leq (1+n_T)\,\delta \tag{A.12}$$

where $\delta = \max\left\{1, \mathrm{tr}\left[\Gamma_y(0)\right]\right\}$. Further, using (2.3), one can show that

$$\begin{aligned}
\mathsf{E}\left\|u_t(n_T) - u_t\right\|^2 &= \mathsf{E}\left\|\sum_{\tau=n_T+1}^{\infty} \Pi_\tau y_{t-\tau}^a\right\|^2 = \sum_{\tau_1=n_T+1}^{\infty}\sum_{\tau_2=n_T+1}^{\infty} \mathrm{tr}\left[\Pi_{\tau_1}'\Pi_{\tau_2}\Gamma_{y^a}(\tau_2-\tau_1)\right] \\
&\leq \sum_{\tau_1=n_T+1}^{\infty}\sum_{\tau_2=n_T+1}^{\infty} \left\|\Pi_{\tau_1}\right\|\left\|\Pi_{\tau_2}\right\|\left\|\Gamma_{y^a}(\tau_2-\tau_1)\right\| \\
&\leq \frac{C^2}{1-\rho^2}\left\|\Sigma_u\right\| \sum_{\tau_1=n_T+1}^{\infty}\sum_{\tau_2=n_T+1}^{\infty} \rho^{|\tau_2-\tau_1|}\left\|\Pi_{\tau_1}\right\|\left\|\Pi_{\tau_2}\right\| \\
&\leq \frac{C^2}{1-\rho^2}\left\|\Sigma_u\right\|\Big(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\Big)^2
\end{aligned} \tag{A.13}$$

where C is a positive constant and $\Gamma_{y^a}(s-t) = \mathsf{E}\left[y_t^a y_s^{a\prime}\right]$ with $y_t^a = y_t - \mu_y = \sum_{v=0}^{\infty}\Psi_v u_{t-v}$, hence

$$\begin{aligned}
\mathsf{E}\left\|U_1(n_T)\right\| &\leq \frac{C}{(1-\rho^2)^{1/2}}\left\|\Sigma_u\right\|^{1/2}(1+n_T)^{1/2}\,\delta^{1/2}\Big(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\Big) \\
&= C_1\left(\frac{1+n_T}{n_T}\right)^{1/2}\left(n_T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right) = O(1)\left(n_T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right)
\end{aligned} \tag{A.14}$$

where C_1 is a positive constant. Then

$$\left\|U_1(n_T)\right\| = O_p(1)\left(n_T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right). \tag{A.15}$$

Since u_t and $Y_t(n_T)$ are independent, we have

$$\begin{aligned}
\mathsf{E}\left\|U_2(n_T)\right\|^2 &= \frac{1}{T^2}\sum_{t=1}^{T}\mathsf{E}\left[u_t'u_t\right]\mathsf{E}\left[Y_t(n_T)'\,Y_t(n_T)\right] \\
&= \frac{1}{T}\mathrm{tr}\left[\Sigma_u\right]\Big(1+n_T\mathrm{tr}\left[\Gamma_y(0)\right]\Big) = O\Big(\frac{k+k^2 n_T}{T}\Big) = O\Big(\frac{k^2 n_T}{T}\Big),
\end{aligned} \tag{A.16}$$

hence

$$\left\|U_2(n_T)\right\| = O_p\left(\frac{k n_T^{1/2}}{T^{1/2}}\right). \tag{A.17}$$

Then, by Assumption **2.4**, we show, using (A.9), (A.10), (A.15) and (A.17), that

$$\left\|\tilde{\Pi}(n_T) - \Pi(n_T)\right\| = o_p(1)\,. \tag{A.18}$$

Using Assumption **2.6**, we finally get

$$\left\|U_1(n_T)\right\| = O_p\left(\frac{n_T^{1/2}}{T^{1/2}}\right)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right) = o_p\left(\frac{n_T^{1/2}}{T^{1/2}}\right) \tag{A.19}$$

and

$$\left\|\tilde{\Pi}(n_T) - \Pi(n_T)\right\| = O_p\left(\frac{k n_T^{1/2}}{T^{1/2}}\right). \tag{A.20}$$

■

PROOF OF PROPOSITION 3.2 First, note that

$$\begin{aligned}
\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| &= \left\|T^{1/2} l'_{n_T} \mathrm{vec}\left[\tilde{\Omega}_Y(n_T)\tilde{\Gamma}_Y(n_T)^{-1} - \Omega_Y(n_T)\Gamma_Y(n_T)^{-1}\right]\right\| \\
&\leq T^{1/2}\left\|l_{n_T}\right\|\left\|\mathrm{vec}\left[\tilde{\Omega}_Y(n_T)\tilde{\Gamma}_Y(n_T)^{-1} - \Omega_Y(n_T)\Gamma_Y(n_T)^{-1}\right]\right\| \\
&= \left\|l_{n_T}\right\|\left\|T^{1/2}\left\{\tilde{\Omega}_Y(n_T)\tilde{\Gamma}_Y(n_T)^{-1} - \Omega_Y(n_T)\Gamma_Y(n_T)^{-1}\right\}\right\| \\
&\leq M^{1/2}\left\{\left\|q_1 + q_2\right\|\right\} \leq M^{1/2}\left\{\left\|q_1\right\| + \left\|q_2\right\|\right\} \qquad \text{(A.21)}
\end{aligned}$$

where

$$q_1 = T^{1/2}\left[U_1(n_T) + U_2(n_T)\right]\left[\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right],\; q_2 = T^{-1/2}\sum_{t=1}^{T}\left[u_t(n_T) - u_t\right]Y_t(n_T)'\,\Gamma_Y(n_T)^{-1},$$

with $U_1(n_T)$ and $U_2(n_T)$ as defined in (A.8). Then

$$\begin{aligned}
\left\|q_1\right\| &= \left\|T^{1/2}\left[U_1(n_T) + U_2(n_T)\right]\left[\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right]\right\| \\
&\leq T^{1/2}\left\{\left\|U_1(n_T)\right\| + \left\|U_2(n_T)\right\|\right\}\left\|\tilde{\Gamma}_Y(n_T)^{-1} - \Gamma_Y(n_T)^{-1}\right\|_1, \qquad \text{(A.22)}
\end{aligned}$$

$$\begin{aligned}
\mathsf{E}\left\|q_2\right\| &= \mathsf{E}\left\|T^{-1/2}\sum_{t=1}^{T}\left[u_t(n_T) - u_t\right]Y_t(n_T)'\,\Gamma_Y(n_T)^{-1}\right\| \\
&\leq T^{-1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\left\|\left[u_t(n_T) - u_t\right]Y_t(n_T)'\,\Gamma_Y(n_T)^{-1}\right\|\right\} \\
&\leq T^{-1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\left\|u_t(n_T) - u_t\right\|^2\right\}^{1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\left\|Y_t(n_T)'\,\Gamma_Y(n_T)^{-1}\right\|^2\right\}^{1/2}. \qquad \text{(A.23)}
\end{aligned}$$

By Proposition **3.1**, (A.15) and (A.17), we can see, using Assumption **2.6**, that

$$\begin{aligned}
\left\|q_1\right\| &= T^{1/2}\left\{O_p(1)\left(n_T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right) + O_p\left(\frac{n_T^{1/2}}{T^{1/2}}\right)\right\}O_p\left(\frac{n_T}{T^{1/2}}\right) \\
&= \left\{O_p(1)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right) + O_p(1)\right\}O_p\left(\frac{n_T^{3/2}}{T^{1/2}}\right) = O_p\left(\frac{n_T^{3/2}}{T^{1/2}}\right). \qquad \text{(A.24)}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\mathsf{E}\left\|Y_t(n_T)'\,\Gamma_Y(n_T)^{-1}\right\|^2 &= \mathsf{E}\left\{\mathrm{tr}\left[\Gamma_Y(n_T)^{-1}Y_t(n_T)Y_t(n_T)'\,\Gamma_Y(n_T)^{-1}\right]\right\} \\
&= \mathrm{tr}\left[\Gamma_Y(n_T)^{-1}\,\mathsf{E}\left\{Y_t(n_T)Y_t(n_T)'\right\}\Gamma_Y(n_T)^{-1}\right] = \mathrm{tr}\left[\Gamma_Y(n_T)^{-1}\right] = O(1). \qquad \text{(A.25)}
\end{aligned}$$

Then, using (A.13), we get

$$\begin{aligned}
\mathsf{E}\left\|q_2\right\| &\leq T^{-1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\left\|u_t(n_T) - u_t\right\|^2\right\}^{1/2}\left\{\sum_{t=1}^{T}\mathsf{E}\left\|Y_t(n_T)'\,\Gamma_Y(n_T)^{-1}\right\|^2\right\}^{1/2}. \\
&= T^{-1/2}\left\{O(T)\left(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right)^2\right\}^{1/2}\left\{O(T)\right\}^{1/2} = O(1)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right), \qquad \text{(A.26)}
\end{aligned}$$

$$\left\|q_2\right\| = O_p(1)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right), \qquad \text{(A.27)}$$

hence

$$\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = O_p\left(\frac{n_T^{3/2}}{T^{1/2}}\right) + O_p(1)\left(T^{1/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right). \qquad \text{(A.28)}$$

Therefore, by Assumptions **2.5** and **2.6**, we have

$$\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = o_p(1). \tag{A.29}$$

Further, we see that

$$\left(\frac{T^{1/2}}{n_T^{3/2}}\right)\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = O_p(1) + O_p(1)\left(Tn_T^{-3/2}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right) \tag{A.30}$$

where, by Assumption **2.5**, $Tn_T^{-3/2} \leq c^{-3/2}T^{1-3\delta_2/2}$, since $n_T \geq cT^{\delta_2}$. Setting also $\delta_3 = 1 - \frac{3}{2}\delta_2$, then $\frac{1}{2} < \delta_3 < 1$ since $0 < \delta_2 < 1/3$. Hence

$$\left(\frac{T^{1/2}}{n_T^{3/2}}\right)\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = O_p(1) + O_p(1)\left(T^{\delta_3}\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right). \tag{A.31}$$

If, in addition, Assumption **2.7** holds, then we get

$$\left(\frac{T^{1/2}}{n_T^{3/2}}\right)\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = O_p(1) + o_p(1) = O_p(1) \tag{A.32}$$

and finally

$$\left\|\tilde{S}_Y(n_T) - S_Y(n_T)\right\| = O_p\left(\frac{n_T^{3/2}}{T^{1/2}}\right). \tag{A.33}$$

■

PROOF OF THEOREM 3.2 By the standard central limit theorem for stationary processes [see Anderson (1971, Section 7.7), Scott (1973, Theorem 2), Berk (1974, page 491), Lewis and Reinsel (1985, Theorem 3), Chung (2001, Theorem 9.1.5)] and under the assumption of independence between u_t and $Y_t(n_T)$ we have:

$$\frac{S_Y(n_T)}{\left\{l'_{n_T}Q_Y(n_T)\,l_{n_T}\right\}^{1/2}} \xrightarrow[T\to\infty]{d} N\left[0,1\right] \tag{A.34}$$

where $Q_Y(n_T) = \Gamma_Y(n_T)^{-1}\otimes\Sigma_u$ and $\Gamma_Y(n_T) = \mathsf{E}\left[Y_t(n_T)\,Y_t(n_T)'\right]$. Therefore, by Proposition **3.2** and Assumption **2.5** we finally conclude that

$$\frac{T^{1/2}l'_{n_T}\operatorname{vec}\left[\tilde{\Pi}(n_T) - \Pi(n_T)\right]}{\left\{l'_{n_T}Q_Y(n_T)\,l_{n_T}\right\}^{1/2}} = \frac{\tilde{S}_Y(n_T)}{\left\{l'_{n_T}Q_Y(n_T)\,l_{n_T}\right\}^{1/2}} \xrightarrow[T\to\infty]{d} N\left[0,1\right]. \tag{A.35}$$

■

PROOF OF PROPOSITION 3.3 Let $\Sigma_u(T) = T^{-1}\sum_{t=1}^{T}u_tu_t'$. Then, by the triangular inequality, we have

$$\left\|\tilde{\Sigma}_u(n_T) - \Sigma_u\right\| \leq \left\|\tilde{\Sigma}_u(n_T) - \Sigma_u(T)\right\| + \left\|\Sigma_u(T) - \Sigma_u\right\| \tag{A.36}$$

where $\left\|\Sigma_u(T) - \Sigma_u\right\| = O_p\left(k/T^{1/2}\right)$ and

$$\begin{aligned}\left\|\tilde{\Sigma}_u(n_T) - \Sigma_u(T)\right\| &\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{u}_t(n_T)\,\tilde{u}_t(n_T)' - u_tu_t'\right\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\left\{\left\|\tilde{u}_t(n_T) - u_t\right\|\left\|\tilde{u}_t(n_T)\right\| + \left\|u_t\right\|\left\|\tilde{u}_t(n_T) - u_t\right\|\right\}.\end{aligned} \tag{A.37}$$

Moreover, we have

$$\begin{aligned}\left\|\tilde{u}_t(n_T) - u_t\right\|^2 &\leq 2\left\|\tilde{u}_t(n_T) - u_t(n_T)\right\|^2 + 2\left\|u_t(n_T) - u_t\right\|^2 \\ &\leq 2\left\|\tilde{\Pi}(n_T) - \Pi(n_T)\right\|^2\left\|Y_t(n_T)\right\|^2 + 2\left\|\sum_{\tau=n_T+1}^{\infty}\Pi_\tau y^a_{t-\tau}\right\|^2\end{aligned} \tag{A.38}$$

where $\left\|\tilde{\Pi}(n_T)-\Pi(n_T)\right\|^2=O_p\left(k^2 n_T/T\right)$, $\mathsf{E}\left\|Y_t(n_T)\right\|^2=O(kn_T)$ and

$$\mathsf{E}\left\|\sum_{\tau=n_T+1}^{\infty}\Pi_\tau y_{t-\tau}^a\right\|^2=O(k)\left(\sum_{\tau=n_T+1}^{\infty}\left\|\Pi_\tau\right\|\right)^2=O\Big(k\rho^{2n_T}\Big), \tag{A.39}$$

hence

$$\left\|\tilde{u}_t(n_T)-u_t\right\|^2=O_p\left(\frac{k^3n_T^2}{T}\right),\ \left\|\tilde{u}_t(n_T)-u_t\right\|=O_p\left(\frac{k^{3/2}n_T}{T^{1/2}}\right). \tag{A.40}$$

Finally, we get

$$\left\|\tilde{\Sigma}_u(n_T)-\Sigma_u(T)\right\|=O_p\left(\frac{k^2n_T}{T^{1/2}}\right),\ \left\|\tilde{\Sigma}_u(n_T)-\Sigma_u\right\|=O_p\left(\frac{k^2n_T}{T^{1/2}}\right), \tag{A.41}$$

since $\mathsf{E}\left\|u_t\right\|^2=O(k)$. Therefore, similarly, as in the proof of Proposition **3.1**, it can be seen that

$$\left\|\tilde{\Sigma}_u(n_T)^{-1}-\Sigma_u^{-1}\right\|=O_p\left(\frac{k^2n_T}{T^{1/2}}\right). \tag{A.42}$$

■

PROOF OF PROPOSITION 3.4 Let

$$\Gamma_X(T)=T^{-1}\sum_{t=1}^{T}X_tX_t',\ \Upsilon_X(T)=\Gamma_X(T)\otimes\Sigma_u^{-1},\ Q_X(T)=\Big\{R'\Upsilon_X(T)R\Big\}^{-1}. \tag{A.43}$$

Then

$$\begin{aligned}\left\|\tilde{Q}_X(n_T)^{-1}-Q_X^{-1}\right\|_1 &\le \left\|\tilde{Q}_X(n_T)^{-1}-Q_X^{-1}\right\|\\ &\le \left\|\tilde{Q}_X(n_T)^{-1}-Q_X(T)^{-1}\right\|+\left\|Q_X(T)^{-1}-Q_X^{-1}\right\|,\end{aligned} \tag{A.44}$$

with

$$\left\|Q_X(T)^{-1}-Q_X^{-1}\right\|\le\left\|R\right\|^2\left\|\Upsilon_X(T)-\Upsilon_X\right\|=\left\|R\right\|^2\left\|\Sigma_u^{-1}\right\|\left\|\Gamma_X(T)-\Gamma_X\right\| \tag{A.45}$$

where $\left\|R\right\|^2=r_{\bar{p}}$, $\left\|\Sigma_u^{-1}\right\|=O_p(k)$ and $\left\|\Gamma_X(T)-\Gamma_X\right\|=O_p\Big(hk/T^{1/2}\Big)$. Hence

$$\left\|Q_X(T)^{-1}-Q_X^{-1}\right\|=O_p\left(\frac{r_{\bar{p}}hk^2}{T^{1/2}}\right). \tag{A.46}$$

Moreover, we have

$$\left\|\tilde{Q}_X(n_T)^{-1}-Q_X(T)^{-1}\right\|\le\left\|R\right\|^2\left\|\tilde{\Upsilon}_X(n_T)-\Upsilon_X(T)\right\| \tag{A.47}$$

where

$$\left\|\tilde{\Upsilon}_X(n_T)-\Upsilon_X(T)\right\| \le \left\|\tilde{\Gamma}_X(n_T)\right\|\left\|\tilde{\Sigma}_u(n_T)^{-1}-\Sigma_u^{-1}\right\|+\left\|\tilde{\Gamma}_X(n_T)-\Gamma_X(T)\right\|\left\|\Sigma_u^{-1}\right\|, \tag{A.48}$$

$$\left\|\tilde{\Gamma}_X(n_T)-\Gamma_X(T)\right\| \le \frac{1}{T}\sum_{t=1}^{T}\Big\{\left\|\tilde{X}_t(n_T)\right\|\left\|\tilde{X}_t(n_T)-X_t\right\|+\left\|\tilde{X}_t(n_T)-X_t\right\|\left\|X_t\right\|\Big\}, \tag{A.49}$$

with

$$\begin{aligned}\left\|\tilde{X}_t(n_T)-X_t\right\| &= \left\{\sum_{j=0}^{\bar{p}}\left\|\tilde{u}_{t-j}(n_T)-u_{t-j}\right\|^2\right\}^{1/2}=\big(\bar{p}+1\big)^{1/2}\left\|\tilde{u}_t(n_T)-u_t\right\|\\ &\le h^{1/2}\left\|\tilde{u}_t(n_T)-u_t\right\|=O_p\left(\frac{h^{1/2}k^{3/2}n_T}{T^{1/2}}\right),\end{aligned} \tag{A.50}$$

using (A.40). Further, since $\mathsf{E}\left\|X_t\right\|^2=O(hk)$ and $\left\|\Sigma_u^{-1}\right\|=O_p(k)$, we have

$$\left\|\tilde{\Gamma}_X(n_T)-\Gamma_X(T)\right\|=O_p\left(\frac{hk^2n_T}{T^{1/2}}\right),\quad \left\|\tilde{\Upsilon}_X(n_T)-\Upsilon_X(T)\right\|=O_p\left(\frac{hk^3n_T}{T^{1/2}}\right), \tag{A.51}$$

then

$$\left\| \tilde{Q}_X\left(n_T\right)^{-1} - Q_X\left(T\right)^{-1}\right\| = O_p\left(\frac{r_{\bar{p}} h k^3 n_T}{T^{1/2}}\right). \tag{A.52}$$

Hence

$$\left\| \tilde{Q}_X\left(n_T\right)^{-1} - Q_X^{-1}\right\| = O_p\left(\frac{r_{\bar{p}} h k^3 n_T}{T^{1/2}}\right), \quad \left\| \tilde{Q}_X\left(n_T\right)^{-1} - Q_X^{-1}\right\|_1 = O_p\left(\frac{r_{\bar{p}} h k^3 n_T}{T^{1/2}}\right). \tag{A.53}$$

Finally, as in the proof of Proposition **3.1**, one can show that

$$\left\| \tilde{Q}_X\left(n_T\right) - Q_X\right\|_1 = O_p\left(\frac{r_{\bar{p}} h k^3 n_T}{T^{1/2}}\right). \tag{A.54}$$

■

PROOF OF THEOREM 3.3 Recall that $\tilde{\eta} - \eta = \tilde{Q}_X\left(n_T\right)\tilde{\Omega}_X\left(n_T\right)$. Then

$$\left\|\tilde{\eta} - \eta\right\| \leq \left\|\tilde{Q}_X\left(n_T\right)\right\|_1 \left\|\tilde{\Omega}_X\left(n_T\right) - \Omega_X\right\| + \left\|\tilde{Q}_X\left(n_T\right) - Q_X\right\|_1 \left\|\Omega_X\right\| + \left\|Q_X\right\|_1 \left\|\Omega_X\right\| \tag{A.55}$$

where, by Proposition **3.4**,

$$\left\| \tilde{Q}_X\left(n_T\right) - Q_X\right\|_1 = O_p\left(\frac{r_{\bar{p}} h k^3 n_T}{T^{1/2}}\right), \quad \left\| \tilde{Q}_X\left(n_T\right)\right\|_1 = O_p\left(r_{\bar{p}} h k^2\right). \tag{A.56}$$

Let also

$$W_X = \frac{1}{T}\sum_{t=1}^{T} u_t X_t{}'. \tag{A.57}$$

Then one sees that

$$\Omega_X = R'\mathrm{vec}\left[\Sigma_u^{-1} W_X\right] \tag{A.58}$$

and

$$\mathsf{E}\left\|\Omega_X\right\|^2 \leq \left\|R\right\|^2 \left\|\Sigma_u^{-1}\right\|^2 \mathsf{E}\left\|W_X\right\|^2 \tag{A.59}$$

where, by independence between u_t and X_t,

$$\mathsf{E}\left\|W_X\right\|^2 = O\left(\frac{hk^2}{T}\right). \tag{A.60}$$

Hence

$$\left\|\Omega_X\right\| = O_p\left(\frac{r_{\bar{p}}^{1/2} h^{1/2} k^2}{T^{1/2}}\right). \tag{A.61}$$

Now, consider the term $\left\|\tilde{\Omega}_X\left(n_T\right) - \Omega_X\right\|$. Then it can be shown that

$$\begin{aligned} \left\|\tilde{\Omega}_X\left(n_T\right) - \Omega_X\right\| \quad \leq \quad & \left\|R\right\|\Big[\left\|\tilde{\Sigma}_u\left(n_T\right)^{-1}\right\|\Big\{\left\|\tilde{W}_X^1\left(n_T\right)\right\| + \left\|\tilde{W}_X^2\left(n_T\right)\right\| + \left\|\tilde{W}_X^3\left(n_T\right)\right\|\Big\} \\ & + \left\|\tilde{\Sigma}_u\left(n_T\right)^{-1} - \Sigma_u^{-1}\right\|\left\|W_X\right\|\Big] \end{aligned} \tag{A.62}$$

where

$$\tilde{W}_X^1\left(n_T\right) = \frac{1}{T}\sum_{t=1}^{T}\left[e_t\left(n_T\right) - u_t\right]\left[\tilde{X}_t\left(n_T\right) - X_t\right]', \tag{A.63}$$

$$\tilde{W}_X^2\left(n_T\right) = \frac{1}{T}\sum_{t=1}^{T}\left[e_t\left(n_T\right) - u_t\right]X_t{}', \; \tilde{W}_X^3\left(n_T\right) = \frac{1}{T}\sum_{t=1}^{T} u_t\left[\tilde{X}_t\left(n_T\right) - X_t\right]'. \tag{A.64}$$

By Proposition **3.3**, we have

$$\left\|\tilde{\Sigma}_u\left(n_T\right)^{-1}\right\| = O_p\left(k\right) + O_p\left(\frac{k^2 n_T}{T^{1/2}}\right), \; \left\|\tilde{\Sigma}_u\left(n_T\right)^{-1} - \Sigma_u^{-1}\right\| = O_p\left(\frac{k^2 n_T}{T^{1/2}}\right). \tag{A.65}$$

Moreover, using (2.12), (3.18) and (A.50), one can see that

$$\left\|e_t(n_T) - u_t\right\| \leq \left\|\tilde{X}_t(n_T) - X_t\right\|\left\|R\right\|\left\|\eta\right\| = O_p\left(\frac{r_{\bar{p}} h^{1/2} k^{3/2} n_T}{T^{1/2}}\right). \tag{A.66}$$

Hence

$$\left\|\tilde{W}_X^1(n_T)\right\| \leq \left\{\frac{1}{T}\sum_{t=1}^{T}\left\|e_t(n_T) - u_t\right\|^2\right\}^{1/2}\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{X}_t(n_T) - X_t\right\|^2\right\}^{1/2} = O_p\left(\frac{r_{\bar{p}} h k^3 n_T^2}{T}\right). \tag{A.67}$$

Further, setting $F = \left[\mu_\Phi, I_k - \Phi_0, \Phi_1, \ldots, \Phi_{\bar{p}}, \Theta_1, \ldots, \Theta_{\bar{p}}\right]$, one sees that

$$\left\|\tilde{W}_X^2(n_T)\right\|^2 \leq \left\|F\right\|^2\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{X}_t(n_T) - X_t\right]X_t'\right\|^2, \tag{A.68}$$

with

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{X}_t(n_T) - X_t\right]X_t'\right\|^2 = \sum_{j=0}^{\bar{p}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_{t-j}(n_T) - u_{t-j}\right]X_t'\right\|^2 \tag{A.69}$$

and

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[u_{t-j} - \tilde{u}_{t-j}(n_T)\right]X_t'\right\|^2 \leq 2\left\|\tilde{W}_X^{21}(n_T)\right\|^2 + 2\left\|\tilde{W}_X^{22}(n_T)\right\|^2 \tag{A.70}$$

where $\tilde{W}_X^{21}(n_T) = T^{-1}\sum_{t=1}^{T}\left[\tilde{u}_{t-j}(n_T) - u_{t-j}(n_T)\right]X_t'$ and $\tilde{W}_X^{22}(n_T) = T^{-1}\sum_{t=1}^{T}\left[u_{t-j}(n_T) - u_{t-j}\right]X_t'$. In particular, we show that

$$\left\|\tilde{W}_X^{21}(n_T)\right\|^2 \leq \left\|\Pi(n_T) - \tilde{\Pi}(n_T)\right\|^2\left\|W_X^Y(n_T)\right\|^2 \tag{A.71}$$

where

$$\left\|W_X^Y(n_T)\right\|^2 = \left\|\frac{1}{T}\sum_{t=1}^{T}Y_{t-j}(n_T)X_t'\right\|^2 = \left\|\frac{1}{T}\sum_{t=1}^{T}X_t\right\|^2 + \sum_{\tau=1}^{n_T}\left\|\frac{1}{T}\sum_{t=1}^{T}y_{t-j-\tau}X_t'\right\|^2. \tag{A.72}$$

Given the VARMA structure of y_t as described above, one sees that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}X_t\right\|^2 = O_p\left(\frac{hk}{T}\right) \tag{A.73}$$

and

$$\mathsf{E}\left\|\frac{1}{T}\sum_{t=1}^{T}y_{t-j-\tau}X_t'\right\|^2 \leq \frac{hk^2 C_1 \rho_1^{\tau+j}}{T}, \tag{A.74}$$

for some constants $C_1 > 0$ and $0 < \rho < \rho_1 < 1$. Consequently, we get

$$\sum_{\tau=1}^{n_T}\mathsf{E}\left\|\frac{1}{T}\sum_{t=1}^{T}y_{t-j-\tau}X_t'\right\|^2 \leq \sum_{\tau=1}^{n_T}\frac{C_1\rho_1^{\tau+j}}{T} = \frac{hk^2 C_1 \rho_1^j}{T(1-\rho_1)} = \frac{hk^2 C_2}{T}, \tag{A.75}$$

with $C_2 = C_1\rho_1^j/(1-\rho_1)$, then

$$\left\|W_X^Y(n_T)\right\|^2 = O_p\left(\frac{hk^2}{T}\right). \tag{A.76}$$

Hence, using (A.20) and (A.76), we show that

$$\left\|\tilde{W}_X^{21}(n_T)\right\|^2 = O_p\left(\frac{hk^4 n_T}{T^2}\right). \tag{A.77}$$

In addition, we have

$$\begin{aligned}\mathsf{E}\left\|\tilde{W}_X^{22}(n_T)\right\| &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\mathsf{E}\|y_{t-j-\tau}^a X_t{}'\| \leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\left\{\mathsf{E}\|y_{t-j-\tau}^a\|^2\right\}^{1/2}\left\{\mathsf{E}\|X_t\|^2\right\}^{1/2} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left\{\operatorname{tr}\left[\Gamma_{y^a}(0)\right]\right\}^{1/2}\left\{\operatorname{tr}\left[\Gamma_X(0)\right]\right\}^{1/2}\left(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right) = O\left(\rho^{n_T}\right) \qquad \text{(A.78)}\end{aligned}$$

where $\Gamma_X(0) = \Gamma_X$. Therefore

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{u}_{t-j}(n_T) - u_{t-j}\right]X_t{}'\right\|^2 = O_p\left(\frac{hk^4 n_T}{T^2}\right), \; \left\|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{X}_t(n_T) - X_t\right]X_t{}'\right\|^2 = O_p\left(\frac{h^2k^4 n_T}{T^2}\right). \qquad \text{(A.79)}$$

Then

$$\left\|\tilde{W}_X^2(n_T)\right\| = O_p\left(\frac{hk^2 n_T^{1/2}}{T}\right). \qquad \text{(A.80)}$$

Furthermore, one can see that

$$\left\|\tilde{W}_X^3(n_T)\right\|^2 = \sum_{j=0}^{\bar{p}}\left\|\frac{1}{T}\sum_{t=1}^{T}u_t\left[\tilde{u}_{t-j}(n_T) - u_{t-j}\right]'\right\|^2 \qquad \text{(A.81)}$$

where

$$\left\|\frac{1}{T}\sum_{t=1}^{T}u_t\left[\tilde{u}_{t-j}(n_T) - u_{t-j}\right]'\right\| \leq \left\|\tilde{W}_X^{31}(n_T)\right\| + \left\|\tilde{W}_X^{32}(n_T)\right\|, \qquad \text{(A.82)}$$

with $\tilde{W}_X^{31}(n_T) = T^{-1}\sum_{t=1}^{T}u_t\left[\tilde{u}_{t-j}(n_T) - u_{t-j}(n_T)\right]'$ and $\tilde{W}_X^{32}(n_T) = T^{-1}\sum_{t=1}^{T}u_t\left[u_{t-j}(n_T) - u_{t-j}\right]'$. More especially, we have

$$\left\|\tilde{W}_X^{31}(n_T)\right\| \leq \left\|W_u^Y(n_T)\right\|\left\|\Pi(n_T) - \tilde{\Pi}(n_T)\right\|. \qquad \text{(A.83)}$$

Therefore, by independence between u_t and $Y_{t-j}(n_T)$ for $j \geq 0$,

$$\left\|W_u^Y(n_T)\right\| = \left\|\frac{1}{T}\sum_{t=1}^{T}u_t Y_{t-j}(n_T)'\right\| = O_p\left(\frac{kn_T^{1/2}}{T^{1/2}}\right). \qquad \text{(A.84)}$$

In view of (A.20), we get

$$\left\|\tilde{W}_X^{31}(n_T)\right\| = O_p\left(\frac{k^2 n_T}{T}\right). \qquad \text{(A.85)}$$

In the other hand, we have

$$\begin{aligned}\mathsf{E}\left\|\tilde{W}_X^{32}(n_T)\right\| &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\mathsf{E}\|u_t y_{t-j-\tau}^a{}'\|\|\Pi_\tau\| \leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=n_T+1}^{\infty}\left\{\mathsf{E}\|y_{t-j-\tau}^a\|^2\right\}^{1/2}\left\{\mathsf{E}\|u_t\|^2\right\}^{1/2}\|\Pi_\tau\| \\ &= \frac{1}{T}\sum_{t=1}^{T}\left\{\operatorname{tr}\left[\Gamma_{y^a}(0)\right]\right\}^{1/2}\left\{\operatorname{tr}\left[\Sigma_u\right]\right\}^{1/2}\left(\sum_{\tau=n_T+1}^{\infty}\|\Pi_\tau\|\right) = O\left(\rho^{n_T}\right). \qquad \text{(A.86)}\end{aligned}$$

Therefore, it follows that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}u_t\left[\tilde{u}_{t-j}(n_T) - u_{t-j}\right]'\right\|^2 = O_p\left(\frac{k^4 n_T^2}{T^2}\right), \; \left\|\frac{1}{T}\sum_{t=1}^{T}u_t\left[\tilde{X}_t(n_T) - X_t\right]'\right\|^2 = O_p\left(\frac{hk^4 n_T^2}{T^2}\right). \qquad \text{(A.87)}$$

Then

$$\left\|\tilde{W}_X^3(n_T)\right\| = O_p\left(\frac{h^{1/2}k^2 n_T}{T}\right). \qquad \text{(A.88)}$$

Finally, one can see that

$$\begin{aligned}\left\|\tilde{\Omega}_X(n_T)-\Omega_X\right\| &\leq \|R\|\left\|\tilde{\Sigma}_u(n_T)^{-1}\right\|\left\{\left\|\tilde{W}_X^1(n_T)\right\|+\left\|\tilde{W}_X^2(n_T)\right\|+\left\|\tilde{W}_X^3(n_T)\right\|\right\}\\ &\quad+\|R\|\left\|\tilde{\Sigma}_u(n_T)^{-1}-\Sigma_u^{-1}\right\|\|W_X\|\\ &= O_p\Big(\frac{r_{\bar{p}}hk^4n_T^2}{T}\Big). \qquad \text{(A.89)}\end{aligned}$$

As a result, we get

$$\|\tilde{\eta}-\eta\| = O_p\Big(\frac{r_{\bar{p}}^{3/2}h^{3/2}k^4}{T^{1/2}}\Big)+O_p\Big(\frac{r_{\bar{p}}^2h^2k^6n_T^2}{T}\Big) = O_p\Big(\frac{1}{T^{1/2}}\Big)+O_p\Big(\frac{n_T^2}{T}\Big). \qquad \text{(A.90)}$$

Furthermore, in view of Assumption **2.3**,

$$\|\tilde{\eta}-\eta\| = o_p(1). \qquad \text{(A.91)}$$

Moreover, one sees that

$$\left\|T^{1/2}(\tilde{\eta}-\eta)\right\| = O_p(1)+O_p\Big(\frac{n_T^2}{T^{1/2}}\Big). \qquad \text{(A.92)}$$

Hence, by Assumption **2.8**, we get

$$\|\tilde{\eta}-\eta\| = O_p\Big(\frac{r_{\bar{p}}^{3/2}h^{3/2}k^4}{T^{1/2}}\Big) = O_p\Big(\frac{1}{T^{1/2}}\Big). \qquad \text{(A.93)}$$

■

PROOF OF PROPOSITION 3.5 Note that

$$\begin{aligned}\left\|\tilde{S}_X(n_T)-S_X\right\| &= T^{1/2}\left\|\tilde{Q}_X(n_T)\tilde{\Omega}_X(n_T)-Q_X\Omega_X\right\|\\ &\leq T^{1/2}\left\|\tilde{Q}_X\right\|_1\left\|\tilde{\Omega}_X(n_T)-\Omega_X\right\|+T^{1/2}\left\|\tilde{Q}_X(n_T)-Q_X\right\|_1\|\Omega_X\| \qquad \text{(A.94)}\end{aligned}$$

where by Proposition **3.4** and Theorem **3.3**, we have:

$$\left\|\tilde{Q}_X(n_T)\right\|_1 = O_p\Big(r_{\bar{p}}hk^2\Big),\ \left\|\tilde{Q}_X(n_T)-Q_X\right\|_1 = O_p\Big(\frac{r_{\bar{p}}hk^3n_T}{T^{1/2}}\Big), \qquad \text{(A.95)}$$

$$\|\Omega_X\| = O_p\Big(\frac{r_{\bar{p}}^{1/2}h^{1/2}k^2}{T^{1/2}}\Big),\ \left\|\tilde{\Omega}_X(n_T)-\Omega_X\right\| = O_p\Big(\frac{r_{\bar{p}}hk^4n_T^2}{T}\Big) \qquad \text{(A.96)}$$

and finally

$$\left\|\tilde{S}_X(n_T)-S_X\right\| = O_p\Big(\frac{r_{\bar{p}}^2h^2k^6n_T^2}{T^{1/2}}\Big). \qquad \text{(A.97)}$$

■

PROOF OF THEOREM 3.4 By the standard central limit theorem for stationary processes [see Anderson (1971, Section 7.7), Scott (1973, Theorem 2) and Chung (2001, Theorem 9.1.5)], and under the assumption of independence between u_t and X_t, we have:

$$T^{1/2}\Omega_X \xrightarrow[T\to\infty]{d} N\big[0,\ Q_X^{-1}\big]. \qquad \text{(A.98)}$$

Then

$$S_X = T^{1/2}Q_X\Omega_X \xrightarrow[T\to\infty]{d} N\big[0,\ Q_X\big]. \qquad \text{(A.99)}$$

Further, by Proposition **3.5** and assumption **2.8** we conclude that

$$T^{1/2}(\tilde{\eta}-\eta) = \tilde{S}_X(n_T) \xrightarrow[T\to\infty]{d} N\big[0,\ Q_X\big]. \qquad \text{(A.100)}$$

■

PROOF OF PROPOSITION 3.6 By the triangular inequality we have

$$\left\|\tilde{\Sigma}_e(n_T)-\Sigma_u\right\| \leq \left\|\tilde{\Sigma}_e(n_T)-\Sigma_u(T)\right\|+\|\Sigma_u(T)-\Sigma_u\| \qquad \text{(A.101)}$$

where

$$\left\|\tilde{\Sigma}_e(n_T)-\Sigma_u(T)\right\| \leq \frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{e}_t(n_T)\tilde{e}_t(n_T)'-u_tu_t'\right\|$$
$$\leq \frac{1}{T}\sum_{t=1}^{T}\left\{\left\|\tilde{e}_t(n_T)-u_t\right\|\left\|\tilde{e}_t(n_T)\right\|+\left\|u_t\right\|\left\|\tilde{e}_t(n_T)-u_t\right\|\right\}, \quad \text{(A.102)}$$

with

$$\left\|\tilde{e}_t(n_T)-u_t\right\| = \left\|\left[X_t'\otimes I_k\right]R\eta-\left[\tilde{X}_t(n_T)'\otimes I_k\right]R\tilde{\eta}\right\| \leq \left\|I_k\right\|\left\|R\right\|\left\{\left\|X_t\right\|\left\|\eta-\tilde{\eta}\right\|+\left\|X_t-\tilde{X}_t(n_T)\right\|\left\|\tilde{\eta}\right\|\right\}$$
$$= O_p\left(\frac{r_{\bar{p}}^2h^2k^5}{T^{1/2}}\right)+O_p\left(\frac{r_{\bar{p}}h^{1/2}k^2n_T}{T^{1/2}}\right)=O_p\left(\frac{r_{\bar{p}}h^{1/2}k^2n_T}{T^{1/2}}\right). \quad \text{(A.103)}$$

Therefore, we get

$$\left\|\tilde{\Sigma}_e(n_T)-\Sigma_u(T)\right\|=O_p\left(\frac{r_{\bar{p}}h^{1/2}k^{5/2}n_T}{T^{1/2}}\right). \quad \text{(A.104)}$$

Hence

$$\left\|\tilde{\Sigma}_e(n_T)-\Sigma_u\right\|=\left\|\tilde{\Sigma}_e(n_T)^{-1}-\Sigma_u^{-1}\right\|=O_p\left(\frac{r_{\bar{p}}h^{1/2}k^{5/2}n_T}{T^{1/2}}\right). \quad \text{(A.105)}$$

■

PROOF OF LEMMA 4.1 Under the invertibility condition of the echelon form VARMA representation we have $\det\left[\Theta(z)\right]\neq 0$, $|z|\leq 1$. Then there exists a positive constant ε, such that

$$\Theta(z)^{-1}=\sum_{\tau=0}^{\infty}\Lambda_\tau(\eta)z^\tau, \quad |z|<1+\varepsilon. \quad \text{(A.106)}$$

Moreover, there exist real constants $(\varsigma,\zeta)>0$ and $\tau\geq\tau_0$ $(\tau,\tau_0\in\mathbb{Z})$, such that $\Lambda_\tau(\eta)(1+\varsigma)^\tau\to 0$ as $\tau\to\infty$, and $\Lambda_{lc,\tau}(\eta)\leq\zeta(1+\varsigma)^{-\tau}$, $|z|<1+\varsigma$, where $\Lambda_{lc,\tau}(\eta)$ is the component of $\Lambda_\tau(\eta)$ in the l-th row and c-th column $(l,c=1,...,k)$ and $0<\varsigma<\varepsilon$. This means that all components of $\Lambda_\tau(\eta)$ are geometrically bounded. Further, let $\rho=(1+\varsigma)^{-1}$ so that $\rho\in(0,1)$, then $\left\|\Lambda_\tau(\eta)\right\|\leq C\rho^\tau$, with $C=k\zeta$. In particular, there exists a positive constant κ such that $1+\kappa^{-1}<\rho^{-1}$. Hence for $|z|\leq 1+\kappa^{-1}$

$$\sum_{\tau=0}^{\infty}\left\|\Lambda_\tau(\eta)\right\||z|^\tau \leq \sum_{\tau=0}^{\infty}\left\|\Lambda_\tau(\eta)\right\|\left(1+\kappa^{-1}\right)^\tau\leq\sum_{\tau=0}^{\infty}C\rho^\tau\left(1+\kappa^{-1}\right)^\tau$$
$$= \sum_{\tau=0}^{\infty}C\left[\rho\left(1+\kappa^{-1}\right)\right]^\tau=\frac{C\kappa}{\kappa-\rho(\kappa+1)}<\infty. \quad \text{(A.107)}$$

Let also $\Lambda_{lc,\tau}(\breve{\eta})$ and $\Lambda_{lc,\tau}(\eta)$ be the components of $\Lambda_\tau(\breve{\eta})$ and $\Lambda_\tau(\eta)$, respectively. Then

$$\left|\Lambda_{lc,\tau}(\breve{\eta})-\Lambda_{lc,\tau}(\eta)\right|=(\tau!)^{-1}\left|\left[\Lambda_{lc}^{(\tau)}(\breve{\eta})(z)-\Lambda_{lc}^{(\tau)}(\eta)(z)\right]\Big|_{z=0}\right| \quad \text{(A.108)}$$

where $|.|$ stands for the euclidean norm, and $\Lambda_{lc}^{(\tau)}(\breve{\eta})$ and $\Lambda_{lc}^{(\tau)}(\eta)$ designate the τ-th derivatives of $\Lambda_{lc}(\breve{\eta})$ and $\Lambda_{lc}(\eta)$ with respect to z, respectively. Hence, applying the Cauchy inequality to the derivatives of an analytic function, here $\Lambda(\eta)(z)$ [see Ahlfors (1966, Page 122), and Churchill and Brown (1990, Page 130)], we get

$$\left|\left[\Lambda_{lc}^{(\tau)}(\breve{\eta})(z)-\Lambda_{lc}^{(\tau)}(\eta)(z)\right]\Big|_{z=0}\right|\leq(\tau!)\left(1+\kappa^{-1}\right)^{-\tau}\max_{|z|=1+\kappa^{-1}}\left|\Lambda_{lc}(\breve{\eta})(z)-\Lambda_{lc}(\eta)(z)\right|, \quad \text{(A.109)}$$

then

$$\begin{aligned}\Big|\Lambda_{lc,\tau}(\breve{\eta})-\Lambda_{lc,\tau}(\eta)\Big| &\le (1+\kappa^{-1})^{-\tau}\max_{|z|=1+\kappa^{-1}}\Big|\Lambda_{lc}(\breve{\eta})(z)-\Lambda_{lc}(\eta)(z)\Big| \\ &\le (1+\kappa^{-1})^{-\tau}\max_{|z|=1+\kappa^{-1}}\Big|\Big[\det\big\{\breve{\Theta}_{\tau}(z)\big\}\Big]^{-1}\breve{\theta}^{+}_{lc,\tau}(z)-\Big[\det\big\{\Theta_{\tau}(z)\big\}\Big]^{-1}\theta^{+}_{lc,\tau}(z)\Big| \\ &\le (1+\kappa^{-1})^{-\tau}\max_{|z|=1+\kappa^{-1}}\Big|\Big[\det\big\{\breve{\Theta}_{\tau}(z)\big\}\Big]^{-1}-\Big[\det\big\{\Theta_{\tau}(z)\big\}\Big]^{-1}\Big|\Big|\breve{\theta}^{+}_{lc,\tau}(z)\Big| \\ &\quad +(1+\kappa^{-1})^{-\tau}\max_{|z|=1+\kappa^{-1}}\Big|\Big[\det\big\{\Theta_{\tau}(z)\big\}\Big]^{-1}\Big|\Big|\breve{\theta}^{+}_{lc,\tau}(z)-\theta^{+}_{lc,\tau}(z)\Big|, \end{aligned} \tag{A.110}$$

for $\tau\in Z$ and $|z|\le 1+\kappa^{-1}$, where the polynomials $\breve{\theta}^{+}_{lc,\tau}(z)$ and $\theta^{+}_{lc,\tau}(z)$ are the (l,c)-th components of the adjoint matrices of $\breve{\Theta}(z)$ and $\Theta(z)$, respectively. By assumption $\|\breve{\eta}-\eta\|=O_p\Big(T^{-1/2}\Big)$, hence $\|\breve{\Theta}(z)-\Theta(z)\|=O_p\Big(T^{-1/2}\Big)$ for $|z|\le 1+\kappa^{-1}$. Consequently, we have

$$\Big\|\Big[\det\big\{\breve{\Theta}_{\tau}(z)\big\}\Big]^{-1}-\Big[\det\big\{\Theta_{\tau}(z)\big\}\Big]^{-1}\Big\|=O_p\Big(T^{-1/2}\Big), \tag{A.111}$$

$$\Big|\breve{\theta}^{+}_{lc,\tau}(z)-\theta^{+}_{lc,\tau}(z)\Big|=O_p\Big(T^{-1/2}\Big),\ \Big|\Lambda_{lc,\tau}(\breve{\eta})-\Lambda_{lc,\tau}(\eta)\Big|\le C\Big(1+\kappa^{-1}\Big)^{-\tau}T^{-1/2}. \tag{A.112}$$

Hence

$$\Big\|\Lambda_{\tau}(\breve{\eta})-\Lambda_{\tau}(\eta)\Big\|\le C\Big(1+\kappa^{-1}\Big)^{-\tau}T^{-1/2},\ T^{1/2}\Big(1+\kappa^{-1}\Big)^{\tau}\Big\|\Lambda_{\tau}(\breve{\eta})-\Lambda_{\tau}(\eta)\Big\|=O_p(1). \tag{A.113}$$

Finally, we get

$$\Big\|\Lambda_{\tau}(\breve{\eta})\Big\|\le C\Big[\rho^{\tau}+\Big(1+\kappa^{-1}\Big)^{-\tau}T^{-1/2}\Big]. \tag{A.114}$$

■

PROOF OF PROPOSITION 4.1 By the triangular inequality, we have

$$\begin{aligned}\Big\|\tilde{\Sigma}_u(\tilde{\eta})-\Sigma_u\Big\| &\le \frac{1}{T}\sum_{t=1}^{T}\big\|u_t(\tilde{\eta})\,u_t(\tilde{\eta})'-u_tu_t'\big\|+O_p\big(T^{-1/2}\big) \\ &\le \frac{1}{T}\sum_{t=1}^{T}\Big\{\big\|u_t(\tilde{\eta})-u_t\big\|\big\|u_t(\tilde{\eta})\big\|+\big\|u_t\big\|\big\|u_t(\tilde{\eta})-u_t\big\|\Big\}+O_p\big(T^{-1/2}\big)\end{aligned} \tag{A.115}$$

where

$$\big\|u_t(\tilde{\eta})-u_t\big\|\le\big\|u_t(\tilde{\eta})-u_t(\eta)\big\|+\big\|u_t(\eta)-u_t\big\|, \tag{A.116}$$

with $\big\|u_t(\eta)-u_t\big\|=O_p\big(\rho^t\big)$. Furthermore, let $\tilde{\Phi}(\bar{p})=\big[\tilde{\Phi}_0,-\tilde{\Phi}_1,\ldots,-\tilde{\Phi}_{\bar{p}}\big]$, $\Phi(\bar{p})=\big[\Phi_0,-\Phi_1,\ldots,-\Phi_{\bar{p}}\big]$ and $Y^a_t(\bar{p})=\big[y^{a\prime}_{t,}y^{a\prime}_{t-1},\ldots,y^{a\prime}_{t-\bar{p}}\big]'$. Then

$$\big\|u_t(\tilde{\eta})-u_t(\eta)\big\|\le\Big\|\sum_{\tau=0}^{t-1}\Big[\Lambda_{\tau}(\tilde{\eta})-\Lambda_{\tau}(\eta)\Big]\tilde{\Phi}(\bar{p})\,Y^a_{t-\tau}(\bar{p})\Big\|+\Big\|\sum_{\tau=0}^{t-1}\Lambda_{\tau}(\eta)\Big[\tilde{\Phi}(\bar{p})-\Phi(\bar{p})\Big]Y^a_{t-\tau}(\bar{p})\Big\|, \tag{A.117}$$

with

$$\begin{aligned}\Big\|\sum_{\tau=0}^{t-1}\Big[\Lambda_{\tau}(\tilde{\eta})-\Lambda_{\tau}(\eta)\Big]\tilde{\Phi}(\bar{p})\,Y^a_{t-\tau}(\bar{p})\Big\| &\le \Big\|\tilde{\Phi}(\bar{p})\Big\|\Big\|Y^a_t(\bar{p})\Big\|\Big(\sum_{\tau=0}^{t-1}\Big\|\Lambda_{\tau}(\tilde{\eta})-\Lambda_{\tau}(\eta)\Big\|\Big)\le\frac{C_1}{T^{1/2}}\Big(\sum_{\tau=0}^{t-1}\big(1+\kappa^{-1}\big)^{-\tau}\Big) \\ &= \frac{C_1}{T^{1/2}}\Big[\frac{1-\big(1+\kappa^{-1}\big)^{-t}}{1-(1+\kappa^{-1})^{-1}}\Big]=O_p\Big(T^{-1/2}\Big)\end{aligned} \tag{A.118}$$

using Lemma **4.1**. Then, by Theorem **3.3**,

$$\begin{aligned}\left\|\sum_{\tau=0}^{t-1}\Lambda_{\tau}\left(\eta\right)\left[\tilde{\Phi}\left(\bar{p}\right)-\Phi\left(\bar{p}\right)\right]Y_{t-\tau}^{a}\left(\bar{p}\right)\right\| &\leq \left\|\tilde{\Phi}\left(\bar{p}\right)-\Phi\left(\bar{p}\right)\right\|\left\|Y_{t}^{a}\left(\bar{p}\right)\right\|\left(\sum_{\tau=0}^{t-1}\left\|\Lambda_{\tau}\left(\eta\right)\right\|\right)\\ &\leq \frac{C_2}{T^{1/2}}\left(\sum_{\tau=0}^{t-1}\rho^{\tau}\right)^{2}=\frac{C_2}{T^{1/2}}\left(\frac{1-\rho^{t}}{1-\rho}\right)=O_p\left(T^{-1/2}\right), \quad \text{(A.119)}\end{aligned}$$

for some positive constants C_2 and C_2. Hence

$$\left\|u_t\left(\tilde{\eta}\right)-u_t\left(\eta\right)\right\|=O_p\left(T^{-1/2}\right),\ \left\|u_t\left(\tilde{\eta}\right)-u_t\right\|=O_p\left(T^{-1/2}\right)+O_p\left(\rho^{t}\right)=O_p\left(T^{-1/2}\right). \quad \text{(A.120)}$$

Therefore, we get

$$\left\|\tilde{\Sigma}_u\left(\tilde{\eta}\right)-\Sigma_u\right\|=O_p\left(T^{-1/2}\right),\ \left\|\tilde{\Sigma}_u\left(\tilde{\eta}\right)^{-1}-\Sigma_u^{-1}\right\|=O_p\left(T^{-1/2}\right). \quad \text{(A.121)}$$

■

PROOF OF LEMMA 4.2 Consider the two equations

$$\Phi_0^0 y_t = \mu_{\Phi^0}+\sum_{i=1}^{\bar{p}}\Phi_i^0 y_{t-i}+\Phi_0^0 u_t\left(\eta^0\right)+\sum_{j=1}^{\bar{p}}\Theta_j^0 u_{t-j}\left(\eta^0\right), \quad \text{(A.122)}$$

$$\Phi_0^1 y_t = \mu_{\Phi^1}+\sum_{i=1}^{\bar{p}}\Phi_i^1 y_{t-i}+\Phi_0^1 u_t\left(\eta^1\right)+\sum_{j=1}^{\bar{p}}\Theta_j^1 u_{t-j}\left(\eta^1\right) \quad \text{(A.123)}$$

where $\mu_{\Phi^0}=\Phi^0\left(1\right)\mu_y$, $\mu_{\Phi^1}=\Phi^1\left(1\right)\mu_y$, with $\Phi^0\left(1\right)=\Phi_0^0-\sum_{i=1}^{\bar{p}}\Phi_i^0$ and $\Phi^1\left(1\right)=\Phi_0^1-\sum_{i=1}^{\bar{p}}\Phi_i^1$. Then subtracting (A.122) from (A.123), we get

$$\begin{aligned}\left(\Phi_0^1-\Phi_0^0\right)v_t\left(\eta^1\right) &= \left(\mu_{\Phi^1}-\mu_{\Phi^0}\right)+\sum_{i=1}^{\bar{p}}\left(\Phi_i^1-\Phi_i^0\right)y_{t-i}+\sum_{j=1}^{\bar{p}}\left(\Theta_j^1-\Theta_j^0\right)u_{t-j}\left(\eta^1\right)\\ &\quad+\left(\sum_{j=0}^{\bar{p}}\Theta_j^0 L^j\right)\left[u_t\left(\eta^1\right)-u_t\left(\eta^0\right)\right] \quad \text{(A.124)}\end{aligned}$$

where $v_t\left(\eta^0\right)=y_t-u_t\left(\eta^0\right)$. Therefore

$$u_t\left(\eta^1\right)-u_t\left(\eta^0\right)=-\left(\sum_{j=0}^{\bar{p}}\Theta_j^0 L^j\right)^{-1}\left[X_t'\left(\eta^1\right)\otimes I_k\right]R(\eta^1-\eta^0)=-\sum_{\tau=0}^{t-1}\left[X_{t-\tau}'\left(\eta^1\right)\otimes\Lambda_{\tau}\left(\eta^0\right)\right]R(\eta^1-\eta^0) \quad \text{(A.125)}$$

where $X_t\left(\eta^1\right)=\left[1,v_t'\left(\eta^1\right),y_{t-1}',\ldots,y_{t-\bar{p}}',u_{t-1}'\left(\eta^1\right),\ldots,u_{t-\bar{p}}\left(\eta^1\right)\right]'$, with $v_t\left(\eta^1\right)=y_t-u_t\left(\eta^1\right)$. Finally

$$u_t\left(\eta^1\right)-u_t\left(\eta^0\right)=-Z_t^{\circ}\left(\eta^1,\eta^0\right)'\left(\eta^1-\eta^0\right),\ Z_t^{\circ}\left(\eta^1,\eta^0\right)=\sum_{\tau=0}^{t-1}R'\left[X_{t-\tau}\left(\eta^1\right)\otimes\Lambda_{\tau}\left(\eta^0\right)'\right]. \quad \text{(A.126)}$$

■

PROOF OF PROPOSITION 4.2 Set

$$\bar{Q}_X\left(\eta\right)=\left\{\frac{1}{T}\sum_{t=1}^{T}Z_t\left(\eta\right)\Sigma_u^{-1}Z_t\left(\eta\right)'\right\}^{-1}. \quad \text{(A.127)}$$

Then

$$\left\|\tilde{Q}_X^{\circ}\left(\tilde{\eta}\right)^{-1}-Q_X\left(\eta\right)^{-1}\right\|_1\leq\left\|\tilde{Q}_X^{\circ}\left(\tilde{\eta}\right)^{-1}-Q_X\left(\eta\right)^{-1}\right\|\leq\left\|\tilde{Q}_X^{\circ}\left(\tilde{\eta}\right)^{-1}-\bar{Q}_X\left(\eta\right)^{-1}\right\|+\left\|\bar{Q}_X\left(\eta\right)^{-1}-Q_X\left(\eta\right)^{-1}\right\|,$$

with

$$\begin{aligned}\left\|\bar{Q}_X(\eta)^{-1}-Q_X(\eta)^{-1}\right\| &= \left\|\frac{1}{T}\sum_{t=1}^{T}\left\{Z_t(\eta)\Sigma_u^{-1}Z_t(\eta)'-\mathsf{E}\left[Z_t(\eta)\Sigma_u^{-1}Z_t(\eta)'\right]\right\}\right\| \\ &\leq \left\|R\right\|^2\left\|\Sigma_u^{-1}\right\|\left\{\sum_{\tau=0}^{\infty}\sum_{v=0}^{\infty}\left\|\tilde{\Gamma}_X(\tau-v)-\Gamma_X(\tau-\nu)\right\|\left\|\Lambda_\tau(\eta)\right\|\left\|\Lambda_v(\eta)\right\|\right\} \quad (A.128)\end{aligned}$$

where

$$\tilde{\Gamma}_X(\tau-v)=\frac{1}{T}\sum_{t=1}^{T}X_{t-\tau}X_{t-v}',\ \Gamma_X(\tau-v)=\mathsf{E}\left[X_{t-\tau}X_{t-v}'\right]. \quad (A.129)$$

From the VARMA structure of y_t one can see that

$$\mathsf{E}\left\|\tilde{\Gamma}_X(\tau-v)-\Gamma_X(\tau-v)\right\|^2\leq\frac{\bar{C}\bar{\rho}^{|\tau-v|}}{T}, \quad (A.130)$$

for some positive constants $\bar{C}$ and $\rho<\bar{\rho}<1$. Hence

$$\left\|\bar{Q}_X(\eta)^{-1}-Q_X(\eta)^{-1}\right\|=O_p\left(T^{-1/2}\right). \quad (A.131)$$

Further, it can be seen that

$$\left\|\tilde{Q}_X^{\circ}(\tilde{\eta})^{-1}-\bar{Q}_X(\eta)^{-1}\right\|\leq\left\|Q_1\right\|+\left\|Q_2\right\|+\left\|Q_3\right\| \quad (A.132)$$

where

$$\begin{aligned}Q_1 &= \frac{1}{T}\sum_{t=1}^{T}Z_t(\eta)\Sigma_u^{-1}\left[Z_t^{\circ}(\tilde{\eta},\eta)-Z_t(\eta)\right]',\ Q_2=\frac{1}{T}\sum_{t=1}^{T}Z_t(\eta)\left[\tilde{\Sigma}_u(\tilde{\eta})^{-1}-\Sigma_u^{-1}\right]Z_t^{\circ}(\tilde{\eta},\eta)', \quad (A.133)\\ Q_3 &= \frac{1}{T}\sum_{t=1}^{T}\left[Z_t^{\circ}(\tilde{\eta},\eta)-Z_t(\eta)\right]\tilde{\Sigma}_u(\tilde{\eta})^{-1}Z_t^{\circ}(\tilde{\eta},\eta)'. \quad (A.134)\end{aligned}$$

More especially, we have

$$\left\|Q_1\right\|\leq\frac{1}{T}\sum_{t=1}^{T}\left\|Z_t(\eta)\right\|\left\|\Sigma_u^{-1}\right\|\left\|Z_t^{\circ}(\tilde{\eta},\eta)-Z_t(\eta)\right\|, \quad (A.135)$$

with

$$\begin{aligned}\mathsf{E}\left\|Z_t(\eta)\right\|^2 &= \mathsf{E}\left\|\sum_{\tau=0}^{\infty}R'\left[X_{t-\tau}\otimes\Lambda_\tau(\eta)'\right]\right\|^2\leq\left\|R\right\|^2\sum_{\tau_1=0}^{\infty}\sum_{\tau_2=0}^{\infty}\left\|\Gamma_X(\tau_1-\tau_2)\right\|\left\|\Lambda_{\tau_1}(\eta)\right\|\left\|\Lambda_{\tau_2}(\eta)\right\| \\ &\leq \bar{C}_1\left\|R\right\|^2\sum_{\tau_1=0}^{\infty}\sum_{\tau_2=0}^{\infty}\bar{\rho}_1^{|\tau_1-\tau_2|}\left\|\Lambda_{\tau_1}(\eta)\right\|\left\|\Lambda_{\tau_2}(\eta)\right\|\leq\bar{C}_2\left\|R\right\|^2\left(\sum_{\tau=0}^{\infty}\left\|\Lambda_\tau(\eta)\right\|\right)^2=O(1), \quad (A.136)\end{aligned}$$

for some constants $\bar{C}_1,\bar{C}_2>0$ and $0<\rho<\bar{\rho}<1$, and

$$\begin{aligned}\left\|Z_t^{\circ}(\tilde{\eta},\eta)-Z_t(\eta)\right\| &= \left\|\sum_{\tau=0}^{t-1}R'\left[X_{t-\tau}(\tilde{\eta})\otimes\Lambda_\tau(\eta)'\right]-\sum_{\tau=0}^{\infty}R'\left[X_{t-\tau}\otimes\Lambda_\tau(\eta)'\right]\right\| \\ &\leq \left\|R\right\|\left\{\left\|\sum_{\tau=0}^{t-1}\left[\left(X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\right)\otimes\Lambda_\tau(\eta)'\right]\right\|+\left\|\sum_{\tau=t}^{\infty}\left[X_{t-\tau}\otimes\Lambda_\tau(\eta)'\right]\right\|\right\} \quad (A.137)\end{aligned}$$

where

$$\begin{aligned}\mathsf{E}\left\|\sum_{\tau=t}^{\infty}\left[X_{t-\tau}\otimes\Lambda_\tau(\eta)'\right]\right\|^2 &\leq \left\|R\right\|^2\sum_{\tau_1=t}^{\infty}\sum_{\tau_2=t}^{\infty}\left\|\Gamma_X(\tau_1-\tau_2)\right\|\left\|\Lambda_{\tau_1}(\eta)\right\|\left\|\Lambda_{\tau_2}(\eta)\right\| \\ &\leq \bar{C}_1\left\|R\right\|^2\sum_{\tau_1=t}^{\infty}\sum_{\tau_2=t}^{\infty}\bar{\rho}_1^{|\tau_1-\tau_2|}\left\|\Lambda_{\tau_1}(\eta)\right\|\left\|\Lambda_{\tau_2}(\eta)\right\| \\ &\leq \bar{C}_2\left\|R\right\|^2\left(\sum_{\tau=t}^{\infty}\left\|\Lambda_\tau(\eta)\right\|\right)^2\leq\bar{C}_2C\left\|R\right\|^2\left(\sum_{\tau=t}^{\infty}\rho^{\tau}\right)^2=O\left(\rho^{2t}\right), \quad (A.138)\end{aligned}$$

and

$$\Big\|\sum_{\tau=0}^{t-1}\Big[\Big(X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\Big)\otimes\Lambda_{\tau}(\eta)'\Big]\Big\|\leq\sum_{\tau=0}^{t-1}\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\|\|\Lambda_{\tau}(\eta)\|, \tag{A.139}$$

with

$$\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\|^{2}=\sum_{j=0}^{\bar{p}}\|u_{t-j-\tau}(\tilde{\eta})-u_{t-j-\tau}\|^{2}=O_{p}\Big(T^{-1}\Big) \tag{A.140}$$

in view of (A.120). Therefore

$$\Big\|\sum_{\tau=0}^{t-1}\Big[\Big(X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\Big)\otimes\Lambda_{\tau}(\eta)'\Big]\Big\|=O_{p}\Big(T^{-1/2}\Big), \tag{A.141}$$

then

$$\|Z_{t}^{\circ}(\tilde{\eta},\eta)-Z_{t}(\eta)\|=O_{p}\Big(T^{-1/2}\Big)+O_{p}\left(\rho^{t}\right)=O_{p}\Big(T^{-1/2}\Big). \tag{A.142}$$

Hence

$$\|Q_{1}\|=O_{p}\Big(T^{-1/2}\Big). \tag{A.143}$$

Likewise, using (A.136), (A.142) and Proposition **4.1**, we can show that

$$\|Q_{2}\|\leq\Big\|\tilde{\Sigma}_{u}(\tilde{\eta})^{-1}-\Sigma_{u}^{-1}\Big\|\Big\{\frac{1}{T}\sum_{t=1}^{T}\|Z_{t}(\eta)\|\Big[\|Z_{t}^{\circ}(\tilde{\eta},\eta)-Z_{t}(\eta)\|+\|Z_{t}(\eta)\|\Big]\Big\}=O_{p}\Big(T^{-1/2}\Big) \tag{A.144}$$

and

$$\begin{aligned}\|Q_{3}\| &\leq \frac{1}{T}\sum_{t=1}^{T}\|Z_{t}^{\circ}(\tilde{\eta},\eta)-Z_{t}(\eta)\|\Big\{\Big\|\tilde{\Sigma}_{u}(\tilde{\eta})^{-1}-\Sigma_{u}^{-1}\Big\|\|Z_{t}^{\circ}(\tilde{\eta},\eta)-Z_{t}(\eta)\| \\ &\quad +\|\Sigma_{u}^{-1}\|\|Z_{t}^{\circ}(\tilde{\eta},\eta)-Z_{t}(\eta)\|+\Big\|\tilde{\Sigma}_{u}(\tilde{\eta})^{-1}-\Sigma_{u}^{-1}\Big\|\|Z_{t}(\eta)\|+\|\Sigma_{u}^{-1}\|\|Z_{t}(\eta)\|\Big\} \\ &= O_{p}\Big(T^{-1/2}\Big).\end{aligned} \tag{A.145}$$

Consequently, we get

$$\Big\|\tilde{Q}_{X}^{\circ}(\tilde{\eta})^{-1}-\bar{Q}_{X}(\eta)^{-1}\Big\|=O_{p}\Big(T^{-1/2}\Big),\ \Big\|\tilde{Q}_{X}^{\circ}(\tilde{\eta})^{-1}-Q_{X}(\eta)^{-1}\Big\|=O_{p}\Big(T^{-1/2}\Big), \tag{A.146}$$

then

$$\Big\|\tilde{Q}_{X}^{\circ}(\tilde{\eta})^{-1}-Q_{X}(\eta)^{-1}\Big\|_{1}=O_{p}\Big(T^{-1/2}\Big),\ \Big\|\tilde{Q}_{X}^{\circ}(\tilde{\eta})-Q_{X}(\eta)\Big\|_{1}=O_{p}\Big(T^{-1/2}\Big). \tag{A.147}$$

Further, one can show that

$$\begin{aligned}\Big\|\tilde{Q}_{X}(\tilde{\eta})^{-1}-\tilde{Q}_{X}^{\circ}(\tilde{\eta})^{-1}\Big\|_{1} &\leq \Big\|\tilde{Q}_{X}(\tilde{\eta})^{-1}-\tilde{Q}_{X}^{\circ}(\tilde{\eta})^{-1}\Big\| \\ &\leq \Big\|\tilde{\Sigma}_{u}(\tilde{\eta})^{-1}\Big\|\frac{1}{T}\sum_{t=1}^{T}\Big\{\|Z_{t}(\tilde{\eta})-Z_{t}^{\circ}(\tilde{\eta},\eta)\|\|Z_{t}(\tilde{\eta})\| \\ &\quad +\|Z_{t}^{\circ}(\tilde{\eta},\eta)\|\|Z_{t}(\tilde{\eta})-Z_{t}^{\circ}(\tilde{\eta},\eta)\|\Big\}\end{aligned} \tag{A.148}$$

where, by Proposition **4.1** and Lemma **4.1**,

$$\Big\|\tilde{\Sigma}_{u}(\tilde{\eta})^{-1}\Big\|\leq\Big\|\tilde{\Sigma}_{u}(\tilde{\eta})^{-1}-\Sigma_{u}^{-1}\Big\|+\|\Sigma_{u}^{-1}\|=O_{p}(1), \tag{A.149}$$

$$\|Z_{t}(\tilde{\eta})-Z_{t}^{\circ}(\tilde{\eta},\eta)\|\leq\|R\|\sum_{\tau=0}^{t-1}\Big\{\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\|+\|X_{t-\tau}\|\Big\}\|\Lambda_{\tau}(\tilde{\eta})-\Lambda_{\tau}(\eta)\|=O_{p}(T^{-1/2}). \tag{A.150}$$

Then, using (A.136), (A.142) and (A.150)

$$\|Z_{t}^{\circ}(\tilde{\eta},\eta)\| \leq \|Z_{t}^{\circ}(\tilde{\eta},\eta)-Z_{t}(\eta)\|+\|Z_{t}(\eta)\|=O_{p}(1), \tag{A.151}$$

$$\|Z_{t}(\tilde{\eta})\| \leq \|Z_{t}(\tilde{\eta})-Z_{t}^{\circ}(\tilde{\eta},\eta)\|+\|Z_{t}^{\circ}(\tilde{\eta},\eta)-Z_{t}(\eta)\|+\|Z_{t}(\eta)\|=O_{p}(1), \tag{A.152}$$

then

$$\left\|\tilde{Q}_X(\tilde{\eta})^{-1} - \tilde{Q}^{\circ}_X(\tilde{\eta})^{-1}\right\| = O_p\big(T^{-1/2}\big). \tag{A.153}$$

Hence

$$\left\|\tilde{Q}_X(\tilde{\eta})^{-1} - \tilde{Q}^{\circ}_X(\tilde{\eta})^{-1}\right\|_1 = O_p\big(T^{-1/2}\big),\ \left\|\tilde{Q}_X(\tilde{\eta}) - \tilde{Q}^{\circ}_X(\tilde{\eta})\right\|_1 = O_p\big(T^{-1/2}\big). \tag{A.154}$$

■

PROOF OF THEOREM 4.1 By the triangular inequality, we have

$$\begin{aligned}\|\hat{\eta}-\eta\| &\leq \left\|\tilde{Q}^{\circ}_X(\tilde{\eta})\tilde{\Omega}^{\bullet}_X(\tilde{\eta})\right\| + \left\|\tilde{Q}_X(\tilde{\eta})\tilde{\Omega}_X(\tilde{\eta}) - \tilde{Q}^{\circ}_X(\tilde{\eta})\tilde{\Omega}^{\circ}_X(\tilde{\eta})\right\| \\ &\leq \|Q_X(\eta)\|_1\|\Omega_X(\eta)\| + \left\|\tilde{Q}^{\circ}_X(\tilde{\eta}) - Q_X(\eta)\right\|_1\left\|\tilde{\Omega}^{\bullet}_X(\tilde{\eta})\right\| + \|Q_X(\eta)\|_1\left\|\tilde{\Omega}^{\bullet}_X(\tilde{\eta}) - \Omega_X(\eta)\right\| \\ &\quad + \left\|\tilde{Q}_X(\tilde{\eta}) - \tilde{Q}^{\circ}_X(\tilde{\eta})\right\|_1\left\|\tilde{\Omega}_X(\tilde{\eta})\right\| + \left\|\tilde{Q}^{\circ}_X(\tilde{\eta})\right\|_1\left\|\tilde{\Omega}_X(\tilde{\eta}) - \tilde{\Omega}^{\circ}_X(\tilde{\eta})\right\|\end{aligned} \tag{A.155}$$

where $\|Q_X(\eta)\|_1 = O_p(1)$, $\|\Omega_X(\eta)\| = O_p\left(T^{-1/2}\right)$,

$$\left\|\tilde{Q}^{\circ}_X(\tilde{\eta}) - Q_X(\eta)\right\|_1 = O_p\left(T^{-1/2}\right),\ \left\|\tilde{Q}_X(\tilde{\eta}) - \tilde{Q}^{\circ}_X(\tilde{\eta})\right\|_1 = O_p\left(T^{-1/2}\right). \tag{A.156}$$

Now, consider $\|\tilde{\Omega}^{\bullet}_X(\tilde{\eta}) - \Omega_X(\eta)\|$ and $\|\tilde{\Omega}_X(\tilde{\eta}) - \tilde{\Omega}^{\circ}_X(\tilde{\eta})\|$. For the first term, we have

$$\begin{aligned}\left\|\tilde{\Omega}^{\bullet}_X(\tilde{\eta}) - \Omega_X(\eta)\right\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\Sigma_u^{-1}\big[u_t(\eta)-u_t\big]\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right]u_t(\eta)\right\| \\ &\quad + \left\|\frac{1}{T}\sum_{t=1}^{T}\big[Z^{\circ}_t(\tilde{\eta},\eta) - Z_t(\eta)\big]\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t(\eta)\right\|\end{aligned} \tag{A.157}$$

where

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\Sigma_u^{-1}\big[u_t(\eta)-u_t\big]\right\| &= \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty} R'\big[X_{t-\tau}\otimes\Lambda_\tau(\eta)'\big]\Sigma_u^{-1}\big[u_t(\eta)-u_t\big]\right\| \\ &\leq \|R\|_1\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\operatorname{vec}\Big[\Lambda_\tau(\eta)'\Sigma_u^{-1}\big[u_t(\eta)-u_t\big]X'_{t-\tau}\Big]\right\| \\ &= \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\Lambda_\tau(\eta)'\Sigma_u^{-1}\big[u_t(\eta)-u_t\big]X'_{t-\tau}\right\|\end{aligned} \tag{A.158}$$

on using the inequality $\|AB\| \leq \|A\|_1\|B\|$, with $\|R\|_1 = 1$ by construction, and $\|\operatorname{vec}[B]\| = \|B\|$. It follows that

$$\begin{aligned}\mathsf{E}\left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\Sigma_u^{-1}\big[u_t(\eta)-u_t\big]\right\| &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\|\Sigma_u^{-1}\|\mathsf{E}\left\|\big[u_t(\eta)-u_t\big]X'_{t-\tau}\right\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\|\Sigma_u^{-1}\|\Big\{\mathsf{E}\|u_t(\eta)-u_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|X_{t-\tau}\|^2\Big\}^{1/2} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\right)\|\Sigma_u^{-1}\|\Big\{\mathsf{E}\|u_t(\eta)-u_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|X_t\|^2\Big\}^{1/2} \\ &\leq \bar{C}_3\|\Sigma_u^{-1}\|\left(\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\right)\left(\frac{1}{T}\sum_{t=1}^{T}\rho^t\right) = O\big(T^{-1}\big),\end{aligned} \tag{A.159}$$

for some positive constant $\bar{C}_3$. Moreover, we have

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right]u_t(\eta)\right\| &\leq \left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right]\big[u_t(\eta)-u_t\big]\right\| \\ &\quad + \left\|\frac{1}{T}\sum_{t=1}^{T} Z_t(\eta)\left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1}\right]u_t\right\|.\end{aligned} \tag{A.160}$$

Similarly as in (A.158) and (A.159), one sees that

$$\left\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right] \left[u_t(\eta) - u_t \right] \right\| = O_p\big(T^{-3/2}\big). \tag{A.161}$$

Manipulating as in (A.158), we also show that

$$\begin{aligned} \left\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right] u_t \right\| &\leq \left\| \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{\infty} \Lambda_\tau(\eta)' \left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right] u_t X'_{t-\tau} \right\| \\ &\leq \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right\| \left\{ \sum_{\tau=0}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \left\| \frac{1}{T} \sum_{t=1}^{T} u_t X'_{t-\tau} \right\| \right\} \end{aligned} \tag{A.162}$$

where

$$\left\{ \sum_{\tau=0}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \left\| \frac{1}{T} \sum_{t=1}^{T} u_t X'_{t-\tau} \right\| \right\} = \sum_{\tau=0}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \left\| \frac{1}{T} \sum_{t=1}^{T} u_t X'_{t-\tau} \right\|. \tag{A.163}$$

By the VARMA structure of y_t, one can see that

$$\left\| \frac{1}{T} \sum_{t=1}^{T} u_t X'_{t-\tau} \right\| = O_p\Big(T^{-1/2}\Big),\ \sum_{\tau=0}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \left\| \frac{1}{T} \sum_{t=1}^{T} u_t X'_{t-\tau} \right\| = O_p\Big(T^{-1/2}\Big). \tag{A.164}$$

Therefore, using Proposition **4.1**,

$$\left\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right] u_t \right\| = O_p\Big(T^{-1}\Big), \tag{A.165}$$

then

$$\left\| \frac{1}{T} \sum_{t=1}^{T} Z_t(\eta) \left[\tilde{\Sigma}_u(\tilde{\eta})^{-1} - \Sigma_u^{-1} \right] u_t(\eta) \right\| = O_p\Big(T^{-3/2}\Big) + O_p\Big(T^{-1}\Big) = O_p\Big(T^{-1}\Big). \tag{A.166}$$

Finally, one shows that

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t(\eta) \right\| \leq \left\| \Omega_Z^1(\tilde{\eta}) \right\| + \left\| \Omega_Z^2(\tilde{\eta}) \right\| \tag{A.167}$$

where

$$\Omega_Z^1(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right],\ \Omega_Z^2(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t. \tag{A.168}$$

More especially,

$$\left\| \Omega_Z^1(\tilde{\eta}) \right\| \leq \left\| \Omega_Z^{11}(\tilde{\eta}) \right\| + \left\| \Omega_Z^{12}(\tilde{\eta}) \right\| + \left\| \Omega_Z^{13}(\tilde{\eta}) \right\| \tag{A.169}$$

where

$$\Omega_Z^{11}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} R' \left[X_{t-\tau} \otimes \Lambda_\tau(\eta)' \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right], \tag{A.170}$$

$$\Omega_Z^{12}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} R' \Big[\big\{ X_{t-\tau}(\eta) - X_{t-\tau} \big\} \otimes \Lambda_\tau(\eta)' \Big] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right], \tag{A.171}$$

$$\Omega_Z^{13}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} R' \Big[\big\{ X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta) \big\} \otimes \Lambda_\tau(\eta)' \Big] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right], \tag{A.172}$$

with $X_t(\eta) = \left[1, v_t'(\eta), y'_{t-1}, \ldots, y'_{t-\bar{p}}, u'_{t-1}(\eta), \ldots, u'_{t-\bar{p}}(\eta) \right]'$ and $v_t(\eta) = y_t - u_t(\eta)$. Likewise, one can show that

$$\left\| \Omega_Z^{11}(\tilde{\eta}) \right\| \leq \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} \right\| \left\{ \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \left\| \Lambda_\tau(\eta) \right\| \left\| \left[u_t(\eta) - u_t \right] X'_{t-\tau} \right\| \right\} \tag{A.173}$$

where

$$\begin{aligned}
\mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\left\|\Lambda_{\tau}(\eta)\right\|\left\|\left[u_t(\eta)-u_t\right]X_{t-\tau}'\right\|\right) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\left\|\Lambda_{\tau}(\eta)\right\|\mathsf{E}\left\|\left[u_t(\eta)-u_t\right]X_{t-\tau}'\right\| \\
&\le \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\left\|\Lambda_{\tau}(\eta)\right\|\left\{\mathsf{E}\left\|u_t(\eta)-u_t\right\|^2\right\}^{1/2}\left\{\mathsf{E}\left\|X_{t-\tau}\right\|^2\right\}^{1/2} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=t}^{\infty}\left\|\Lambda_{\tau}(\eta)\right\|\right)\left\{\mathsf{E}\left\|u_t(\eta)-u_t\right\|^2\right\}^{1/2}\left\{\mathsf{E}\left\|X_t\right\|^2\right\}^{1/2} \\
&\le \frac{C}{T}\sum_{t=1}^{T}\rho^t\left(\sum_{\tau=0}^{\infty}\rho^{\tau}\right)\left\{\mathsf{E}\left\|u_t(\eta)-u_t\right\|^2\right\}^{1/2}\left\{\mathsf{E}\left\|X_t\right\|^2\right\}^{1/2} \\
&\le \frac{\bar{C}_5}{T}\left(\sum_{t=1}^{T}\rho^{2t}\right) = \left[\frac{\bar{C}_5\rho^2\left(1-\rho^{2T}\right)}{T\left(1-\rho^2\right)}\right] = O\left(T^{-1}\right). \qquad \text{(A.174)}
\end{aligned}$$

Hence

$$\left\|\Omega_Z^{11}(\tilde{\eta})\right\| = O_p\left(T^{-1}\right). \tag{A.175}$$

Further

$$\begin{aligned}
\left\|\Omega_Z^{12}(\tilde{\eta})\right\| &\le \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Lambda_{\tau}(\eta)'\tilde{\Sigma}_u(\tilde{\eta})^{-1}\left[u_t(\eta)-u_t\right]\left[X_{t-\tau}(\eta)-X_{t-\tau}\right]'\right\| \\
&\le \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left\|\Lambda_{\tau}(\eta)\right\|\left\|u_t(\eta)-u_t\right\|\left\|X_{t-\tau}(\eta)-X_{t-\tau}\right\|\right\} \\
&\le \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|u_t(\eta)-u_t\right\|\left(\sum_{\tau=0}^{t-1}\left\|\Lambda_{\tau}(\eta)\right\|^2\right)^{1/2}\left(\sum_{\tau=0}^{t-1}\left\|X_{t-\tau}(\eta)-X_{t-\tau}\right\|^2\right)^{1/2}\right\} \qquad \text{(A.176)}
\end{aligned}$$

where

$$\left\|X_{t-\tau}(\eta)-X_{t-\tau}\right\|^2 = \sum_{j=0}^{\bar{p}}\left\|u_{t-j-\tau}(\eta)-u_{t-j-\tau}\right\|^2, \tag{A.177}$$

with

$$\mathsf{E}\left\|u_{t-j-\tau}(\eta)-u_{t-j-\tau}\right\| \le \sum_{v=t-j-\tau}^{\infty}\left\|\Lambda_v(\eta)\right\|\left\|\Phi(\bar{p})\right\|\mathsf{E}\left\|Y_{t-j-\tau-v}^a(\bar{p})\right\| = O\left(\rho^{t-j-\tau}\right). \tag{A.178}$$

Hence

$$\left\|X_{t-\tau}(\eta)-X_{t-\tau}\right\|^2 = O_p\left(\rho^{2(t-\tau)}\right),\ \sum_{\tau=0}^{t-1}\left\|X_{t-\tau}(\eta)-X_{t-\tau}\right\|^2 = O_p\left(\rho^{2t}\right), \tag{A.179}$$

then

$$\left\|\Omega_Z^{12}(\tilde{\eta})\right\| = O_p\left(T^{-1}\right). \tag{A.180}$$

We also show that

$$\begin{aligned}
\left\|\Omega_Z^{13}(\tilde{\eta})\right\| &\le \left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Lambda_{\tau}(\eta)'\tilde{\Sigma}_u(\tilde{\eta})^{-1}\left[u_t(\eta)-u_t\right]\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right]'\right\| \\
&\le \left\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{t-1}\left\|\Lambda_{\tau}(\eta)\right\|\left\|u_t(\eta)-u_t\right\|\left\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right\|\right)\right\} \qquad \text{(A.181)}
\end{aligned}$$

where

$$\left\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right\|^2 = \sum_{j=0}^{\bar{p}}\left\|u_{t-j-\tau}(\tilde{\eta})-u_{t-j-\tau}(\eta)\right\|^2, \tag{A.182}$$

with

$$\left\|u_{t-j-\tau}(\tilde{\eta})-u_{t-j-\tau}(\eta)\right\| = \left\|u_t(\tilde{\eta})-u_t(\eta)\right\| = O_p\left(T^{-1/2}\right), \tag{A.183}$$

using (A.120). Hence

$$\left\|\Omega_Z^{13}(\tilde{\eta})\right\| = O_p\left(T^{-3/2}\right) \tag{A.184}$$

since $\|u_t(\eta) - u_t\| = O_p(\rho^t)$, then

$$\|\Omega_Z^1(\tilde{\eta})\| = \left\| \frac{1}{T} \sum_{t=1}^{T} \left[Z_t^{\circ}(\tilde{\eta}, \eta) - Z_t(\eta) \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} \left[u_t(\eta) - u_t \right] \right\| = O_p\Big(T^{-1}\Big) \tag{A.185}$$

In addition, one sees that

$$\|\Omega_Z^2(\tilde{\eta})\| \le \|\Omega_Z^{21}(\tilde{\eta})\| + \|\Omega_Z^{22}(\tilde{\eta})\| + \|\Omega_Z^{23}(\tilde{\eta})\| \tag{A.186}$$

where

$$\Omega_Z^{21}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} R' \left[X_{t-\tau} \otimes \Lambda_\tau(\eta)' \right] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t, \tag{A.187}$$

$$\Omega_Z^{22}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} R' \Big[\big\{ X_{t-\tau}(\eta) - X_{t-\tau} \big\} \otimes \Lambda_\tau(\eta)' \Big] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t, \tag{A.188}$$

$$\Omega_Z^{23}(\tilde{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} R' \Big[\big\{ X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta) \big\} \otimes \Lambda_\tau(\eta)' \Big] \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t, \tag{A.189}$$

Likewise, we show that

$$\|\Omega_Z^{21}(\tilde{\eta})\| \le \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} \right\| \left\{ \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \|\Lambda_\tau(\eta)\| \|u_t X_{t-\tau}'\| \right\} \tag{A.190}$$

where, by independence between u_t and X_t,

$$\begin{aligned} \mathsf{E}\left\{ \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \|\Lambda_\tau(\eta)\| \|u_t X_{t-\tau}'\| \right\} &\le \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \|\Lambda_\tau(\eta)\| \Big\{ \mathsf{E}\|u_t X_{t-\tau}'\|^2 \Big\}^{1/2} \\ &= \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=t}^{\infty} \|\Lambda_\tau(\eta)\| \Big\{ \mathsf{E}\|u_t\|^2 \mathsf{E}\|X_{t-\tau}\|^2 \Big\}^{1/2} \\ &= \frac{1}{T} \sum_{t=1}^{T} \Big(\sum_{\tau=t}^{\infty} \|\Lambda_\tau(\eta)\| \Big) \Big\{ \mathsf{E}\|u_t\|^2 \mathsf{E}\|X_t\|^2 \Big\}^{1/2} = O\Big(T^{-1}\Big), \end{aligned} \tag{A.191}$$

then

$$\|\Omega_Z^{21}(\tilde{\eta})\| = O_p\Big(T^{-1}\Big). \tag{A.192}$$

As for (A.171), using (A.179), we show that

$$\begin{aligned} \|\Omega_Z^{22}(\tilde{\eta})\| &\le \left\| \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \Lambda_\tau(\eta)' \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t \big[X_{t-\tau}(\eta) - X_{t-\tau} \big]' \right\| \\ &\le \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} \right\| \left\{ \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \|\Lambda_\tau(\eta)\| \|u_t\| \|X_{t-\tau}(\eta) - X_{t-\tau}\| \right\} \\ &\le \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} \right\| \left\{ \frac{1}{T} \sum_{t=1}^{T} \|u_t\| \Big(\sum_{\tau=0}^{t-1} \|\Lambda_\tau(\eta)\|^2 \Big)^{1/2} \Big(\sum_{\tau=0}^{t-1} \|X_{t-\tau}(\eta) - X_{t-\tau}\|^2 \Big)^{1/2} \right\} = O_p\Big(T^{-1}\Big). \end{aligned} \tag{A.193}$$

Hence

$$\begin{aligned} \|\Omega_Z^{23}(\tilde{\eta})\| &\le \left\| \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=0}^{t-1} \Lambda_\tau(\eta)' \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t \big[X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta) \big]' \right\| \\ &= \left\| \frac{1}{T} \sum_{\tau=0}^{T-1} \sum_{t=\tau+1}^{T} \Lambda_\tau(\eta)' \tilde{\Sigma}_u(\tilde{\eta})^{-1} u_t \big[X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta) \big]' \right\| \\ &\le \left\| \tilde{\Sigma}_u(\tilde{\eta})^{-1} \right\| \left\{ \sum_{\tau=0}^{T-1} \|\Lambda_\tau(\eta)\| \left\| \frac{1}{T} \sum_{t=\tau+1}^{T} u_t \big[X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}(\eta) \big]' \right\| \right\} \end{aligned} \tag{A.194}$$

where

$$\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[X_{t-\tau}\left(\tilde{\eta}\right)-X_{t-\tau}\left(\eta\right)\right]'\right\|^2=\sum_{j=0}^{\bar{p}}\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[u_{t-j-\tau}\left(\tilde{\eta}\right)-u_{t-j-\tau}\left(\eta\right)\right]'\right\|^2, \tag{A.195}$$

with

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[u_{t-\tau}\left(\tilde{\eta}\right)-u_{t-\tau}\left(\eta\right)\right]'\right\| &= \left\|\frac{1}{T}\sum_{t=\tau+1}^{T}\sum_{v=0}^{t-\tau-1}u_tY_{t-\tau-v}^{a\prime}\left[\tilde{\Phi}\left(\bar{p}\right)'\Lambda_v\left(\tilde{\eta}\right)-\Phi\left(\bar{p}\right)'\Lambda_v\left(\eta\right)\right]\right\| \\ &= \left\|\frac{1}{T}\sum_{v=0}^{T-\tau-1}\sum_{t=\tau+1+v}^{T}u_tY_{t-\tau-v}^{a\prime}\left[\tilde{\Phi}\left(\bar{p}\right)'\Lambda_v\left(\tilde{\eta}\right)-\Phi\left(\bar{p}\right)'\Lambda_v\left(\eta\right)\right]\right\| \\ &\leq \sum_{v=0}^{T-\tau-1}\left\|\frac{1}{T}\sum_{t=\tau+1+v}^{T}u_tY_{t-\tau-v}^{a\prime}\right\|\left\|\tilde{\Phi}\left(\bar{p}\right)'\Lambda_v\left(\tilde{\eta}\right)-\Phi\left(\bar{p}\right)'\Lambda_v\left(\eta\right)\right\| \end{aligned} \tag{A.196}$$

where, by independence between u_t and Y_t^a,

$$\left\|\frac{1}{T}\sum_{t=\tau+1+v}^{T}u_tY_{t-\tau-v}^{a\prime}\right\|=O_p\left(T^{-1/2}\right). \tag{A.197}$$

Further, using Theorem **3.3** and Lemma **4.1**, we have

$$\begin{aligned}\sum_{v=0}^{T-\tau-1}\left\|\tilde{\Phi}\left(\bar{p}\right)'\Lambda_v\left(\tilde{\eta}\right)-\Phi\left(\bar{p}\right)'\Lambda_v\left(\tilde{\eta}\right)\right\| &\leq \sum_{v=0}^{T-\tau-1}\left\{\left\|\tilde{\Phi}\left(\bar{p}\right)\right\|\left\|\Lambda_v\left(\tilde{\eta}\right)-\Lambda_v\left(\eta\right)\right\|+\left\|\tilde{\Phi}\left(\bar{p}\right)-\Phi\left(\bar{p}\right)\right\|\left\|\Lambda_v\left(\eta\right)\right\|\right\} \\ &= O_p\left(T^{-1/2}\right),\end{aligned} \tag{A.198}$$

$$\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[u_{t-\tau}\left(\tilde{\eta}\right)-u_{t-\tau}\left(\eta\right)\right]'\right\|=\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[u_{t-j-\tau}\left(\tilde{\eta}\right)-u_{t-j-\tau}\left(\eta\right)\right]'\right\|=O_p\left(T^{-1}\right), \tag{A.199}$$

$$\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[X_{t-\tau}\left(\tilde{\eta}\right)-X_{t-\tau}\left(\eta\right)\right]'\right\|=O_p\left(T^{-1}\right), \tag{A.200}$$

then

$$\left\|\Omega_Z^{23}\left(\tilde{\eta}\right)\right\|=O_p\left(T^{-1}\right). \tag{A.201}$$

Hence

$$\left\|\Omega_Z^{2}\left(\tilde{\eta}\right)\right\|=\left\|\frac{1}{T}\sum_{t=1}^{T}\left[Z_t^{\circ}\left(\tilde{\eta},\eta\right)-Z_t\left(\eta\right)\right]\tilde{\Sigma}_u\left(\tilde{\eta}\right)^{-1}u_t\right\|=O_p\left(T^{-1}\right), \tag{A.202}$$

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\left[Z_t^{\circ}\left(\tilde{\eta},\eta\right)-Z_t\left(\eta\right)\right]\tilde{\Sigma}_u\left(\tilde{\eta}\right)^{-1}u_t\left(\eta\right)\right\|=O_p\left(T^{-1}\right) \tag{A.203}$$

and finally

$$\left\|\tilde{\Omega}_X^{\bullet}\left(\tilde{\eta}\right)-\Omega_X\left(\eta\right)\right\|=O_p\left(T^{-1}\right). \tag{A.204}$$

Similarly, we see that

$$\left\|\tilde{\Omega}_X\left(\tilde{\eta}\right)-\tilde{\Omega}_X^{\circ}\left(\tilde{\eta}\right)\right\|\leq\left\|R\right\|\left\{\left\|\Omega_X^1\left(\tilde{\eta}\right)\right\|+\left\|\Omega_X^2\left(\tilde{\eta}\right)\right\|+\left\|\Omega_X^3\left(\tilde{\eta}\right)\right\|+\left\|\Omega_X^4\left(\tilde{\eta}\right)\right\|\right\} \tag{A.205}$$

where

$$\Omega_X^1(\tilde{\eta}) = R'\text{vec}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\big[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}\big[u_t(\tilde{\eta})-u_t\big]\big[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\big]'\Big], \quad (A.206)$$

$$\Omega_X^2(\tilde{\eta}) = R'\text{vec}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\big[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}\big[u_t(\tilde{\eta})-u_t\big]X_{t-\tau}{}'\Big], \quad (A.207)$$

$$\Omega_X^3(\tilde{\eta}) = R'\text{vec}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\big[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t\big[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\big]'\Big], \quad (A.208)$$

$$\Omega_X^4(\tilde{\eta}) = R'\text{vec}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\big[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_tX_{t-\tau}{}'\Big]. \quad (A.209)$$

Using the same arguments as before, one sees that

$$\begin{aligned}\|\Omega_X^1(\tilde{\eta})\| &\leq \Big\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\big[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}\big[u_t(\tilde{\eta})-u_t\big]\big[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\big]'\Big\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Big\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\Big\|\big\|\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big\|\Big\{\big\|u_t(\tilde{\eta})-u_t(\eta)\big\|\big\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\big\| \\ &\quad+\big\|u_t(\eta)-u_t\big\|\big\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\big\|+\big\|u_t(\tilde{\eta})-u_t(\eta)\big\|\big\|X_{t-\tau}(\eta)-X_{t-\tau}\big\| \\ &\quad+\big\|u_t(\eta)-u_t\big\|\big\|X_{t-\tau}(\eta)-X_{t-\tau}\big\|\Big\} \\ &= O_p\big(T^{-3/2}\big)+O_p\big(T^{-2}\big)+O_p\big(T^{-2}\big)+O_p\big(T^{-3/2}\big)=O_p\big(T^{-3/2}\big), \end{aligned} \quad (A.210)$$

$$\begin{aligned}\|\Omega_X^2(\tilde{\eta})\| &\leq \Big\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\big[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}\big[u_t(\tilde{\eta})-u_t\big]X_{t-\tau}{}'\Big\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Big\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\Big\|\big\|\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big\|\Big\{\big\|u_t(\tilde{\eta})-u_t(\eta)\big\|\big\|X_{t-\tau}\big\|+\big\|u_t(\eta)-u_t\big\|\big\|X_{t-\tau}\big\|\Big\} \\ &= O_p\big(T^{-1}\big)+O_p\big(T^{-3/2}\big)=O_p\big(T^{-1}\big), \end{aligned} \quad (A.211)$$

$$\begin{aligned}\|\Omega_X^3(\tilde{\eta})\| &\leq \Big\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\big[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_t\big[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\big]'\Big\| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\Big\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\Big\|\big\|\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big\|\Big\{\big\|u_t\big\|\big\|X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\big\|+\big\|u_t\big\|\big\|X_{t-\tau}(\eta)-X_{t-\tau}\big\|\Big\} \\ &= O_p\big(T^{-1}\big)+O_p\big(T^{-3/2}\big)=O_p\big(T^{-1}\big), \end{aligned} \quad (A.212)$$

$$\begin{aligned}\|\Omega_X^4(\tilde{\eta})\| &\leq \Big\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\big[\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big]'\tilde{\Sigma}_u(\tilde{\eta})^{-1}u_tX_{t-\tau}'\Big\| \\ &\leq \Big\|\tilde{\Sigma}_u(\tilde{\eta})^{-1}\Big\|\Big\{\sum_{\tau=0}^{T-1}\big\|\Lambda_\tau(\tilde{\eta})-\Lambda_\tau(\eta)\big\|\Big\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_tX_{t-\tau}'\Big\|\Big\}=O_p\big(T^{-1}\big), \end{aligned} \quad (A.213)$$

hence

$$\Big\|\tilde{\Omega}_X(\tilde{\eta})-\tilde{\Omega}_X^{\circ}(\tilde{\eta})\Big\|=O_p\big(T^{-1}\big). \quad (A.214)$$

In particular, one can see that

$$\Big\|\tilde{\Omega}_X(\tilde{\eta})\Big\|\leq\Big\|\tilde{\Omega}_X^{\circ}(\tilde{\eta})\Big\|+\Big\|\tilde{\Omega}_X(\tilde{\eta})-\tilde{\Omega}_X^{\circ}(\tilde{\eta})\Big\|\leq\Big\|\Omega_X(\eta)\Big\|+\Big\|\tilde{\Omega}_X^{\circ}(\tilde{\eta})-\Omega_X(\eta)\Big\|+\Big\|\tilde{\Omega}_X(\tilde{\eta})-\tilde{\Omega}_X^{\circ}(\tilde{\eta})\Big\| \quad (A.215)$$

Again, using the same arguments as before, it can be shown that

$$\left\|\tilde{\Omega}_X^{\circ}\left(\tilde{\eta}\right)-\Omega_X\left(\eta\right)\right\|=O_p\big(T^{-1}\big),\ \left\|\tilde{\Omega}_X\left(\tilde{\eta}\right)\right\|=O_p\big(T^{-1/2}\big),\ \left\|\hat{\eta}-\eta\right\|=O_p\big(T^{-1/2}\big). \tag{A.216}$$

■

PROOF OF PROPOSITION 4.3 Recall that

$$\begin{aligned}\left\|\tilde{S}_X\left(\tilde{\eta}\right)-S_X\left(\eta\right)\right\| \quad\leq\quad & T^{1/2}\Big\{\left\|\tilde{Q}_X^{\circ}\left(\tilde{\eta}\right)-Q_X\left(\eta\right)\right\|_1\left\|\tilde{\Omega}_X^{\bullet}\left(\tilde{\eta}\right)\right\|+\left\|Q_X\left(\eta\right)\right\|_1\left\|\tilde{\Omega}_X^{\bullet}\left(\tilde{\eta}\right)-\Omega_X\left(\eta\right)\right\| \\ & +\left\|\tilde{Q}_X\left(\tilde{\eta}\right)-\tilde{Q}_X^{\circ}\left(\tilde{\eta}\right)\right\|_1\left\|\tilde{\Omega}_X\left(\tilde{\eta}\right)\right\|+\left\|\tilde{Q}_X^{\circ}\left(\tilde{\eta}\right)\right\|_1\left\|\tilde{\Omega}_X\left(\tilde{\eta}\right)-\tilde{\Omega}_X^{\circ}\left(\tilde{\eta}\right)\right\|\Big\}.\end{aligned} \tag{A.217}$$

Then it follows, by Proposition **4.2** and Theorem **4.1**, that

$$\left\|\tilde{S}_X\left(\tilde{\eta}\right)-S_X\left(\eta\right)\right\|=O_p\big(T^{-1/2}\big). \tag{A.218}$$

■

PROOF OF THEOREM 4.2 By the central limit theorem for stationary processes [see Anderson (1971, Section 7.7), Scott (1973, Theorem 2) and Chung (2001, Theorem 9.1.5)] and under the assumption of independence between u_t and $Z_t\left(\eta\right)$, we have

$$T^{1/2}\Omega_X\left(\eta\right)\xrightarrow[T\to\infty]{d}N\Big[0,\,Q_X\left(\eta\right)^{-1}\Big]. \tag{A.219}$$

Then, by Proposition **4.3**, we get

$$T^{1/2}\big(\hat{\eta}-\eta\big)=\tilde{S}_X\left(\tilde{\eta}\right)\xrightarrow[T\to\infty]{d}N\Big[0,\,Q_X\left(\eta\right)\Big]. \tag{A.220}$$

■

References

Ahlfors, L. V. (1966), *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable*, International Series in Pure and Applied Mathematics, second edn, McGraw-Hill, New York.

Akaike, H. (1974), ‘Markovian representations of stochastic processes and its application to the analysis of autoregressive moving average processes’, *Annals of the Institute of Statistical Mathematics* **26**, 363–387.

Akaike, H. (1975), ‘Markovian representation of stochastic processes by canonical variables’, *SIAM Journal on Control and Optimization* **13**, 162–173.

Akaike, H. (1976), Canonical correlation analysis of time series and the use of an information criterion, *in* R. K. Mehra and D. G. Lainiotis, eds, ‘System Identification: Advances in Case Studies’, Academic Press, New York, pp. 27–96.

Allende, H. and Heiler, S. (1992), ‘Recursive generalized M-estimates for autoregressive moving-average models’, *Journal of Time Series Analysis* **13**(1), 1–18.

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

Athanasopoulos, G. and Vahid, F. (2008*a*), ‘A complete VARMA modelling methodology based on scalar components’, *Journal of Time Series Analysis* **29**(3), 533–554.

Athanasopoulos, G. and Vahid, F. (2008*b*), ‘VARMA versus VAR for macroeconomic forecasting’, *Journal of Business & Economic Statistics* **26**, 237–252.

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

Berk, K. N. (1974), ‘Consistent autoregressive spectral estimates’, *The Annals of Statistics* **2**(3), 489–502.

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.

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, second edn, Holden-Day, San Francisco.

Chung, K. L. (2001), *A Course in Probability Theory*, third edn, Academic Press, New York.

Churchill, R. V. and Brown, J. W. (1990), *Complex Variables and Applications*, fifth edn, McGraw-Hill, New York.

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

De Frutos, R. F. and Serrano, G. R. (2002), ‘A generalized least squares estimation method for VARMA models’, *Statistics* **36**(4), 303–316.

Deistler, M. (1983), ‘The properties of the parametrization of ARMAX systems and their relevance for structural estimation and dynamic specification.’, *Econometrica* **51**, 1187–1208.

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.

Dhrymes, P. J. (1998), *Time Series, Unit Roots, and Cointegration*, Academic Press, San Diego, California, USA.

Dufour, J.-M. and Jouini, T. (2005), Asymptotic distribution of a simple linear estimator for VARMA models in echelon form, *in* P. Duchesne and B. Rémillard, eds, 'Statistical Modeling and Analysis for Complex Data Problems', Kluwer/Springer-Verlag, Canada, chapter 11, pp. 209–240.

Dufour, J.-M. and Pelletier, D. (2008), Practical methods for modelling weak VARMA processes: Identification, estimation and specification with a macroeconomic application, Technical report, CIREQ and CIRANO.

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

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.

Gallego, J. L. (2009), 'The exact likelihood function of a vector autoregressive moving average process', *Statistics and Probability Letters* **79**(6), 711–714.

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

Hannan, E. J. (1969*a*), 'The estimation of mixed moving average autoregressive systems', *Biometrika* **56**(3), 579–593.

Hannan, E. J. (1969*b*), '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. (1971), 'The identification problem for multiple equation systems with moving average errors', *Econometrica* **39**, 751–766.

Hannan, E. J. (1976), '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. (1980), 'The estimation of the order of an ARMA process', *The Annals of Statistics* **8**(5), 1071–1081.

Hannan, E. J. (1981), 'Estimating the dimension of a linear system', *Journal of Multivariate Analysis* **11**(4), 459–473.

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.

Izquierdo, S. S., Hernández, C. and Del Hoyo, J. (2007), Forecasting VARMA processes using VAR models and subspace-based state space models, Working Paper 4235, Munich Personal RePEc Archive (MPRA), Munich University Library in Germany.

Jonasson, K. and Ferrando, S. E. (2008), ‘Evaluating exact VARMA likelihood and its gradient when data are incomplete’, *ACM Transactions on Mathematical Software* **35**(1), 16.

Kapetanios, G. (2003), ‘A note on an iterative least squares estimation method for ARMA and VARMA models’, *Economics Letters* **79**, 305–312.

Kascha, C. (2007), A comparison of estimation methods for vector autoregressive moving-average models, EUI Working Paper **ECO** 12, European University institute, Department of Economics, Via Della Piazzuola 43, 50133 Firenze, Italy.

Kascha, C. and Mertens, K. (2008), Business cycle analysis and VARMA models, Working Paper ANO 2008/5, Research Department, Norges Bank (NB), Oslo, Norway.

Kavalieris, L., Hannan, E. J. and Salau, M. (2003), ‘Generalized least squares estimation of ARMA models’, *Journal of Times Series Analysis* **24**(2), 165–172.

Kohn, R. (1981), ‘A note on an alternative derivation of the likelihood of an autoregressive moving average process’, *Economics Letters* **7**, 233–236.

Koreisha, S. G. and Pukkila, T. M. (1987), ‘Identification of nonzero elements in the polynomial matrices of mixed VARMA processes’, *Journal of the Royal Statistical Society, Series B, Methodological* **49**, 112–126. Correction 50 (1988), 155.

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 & Economic Statistics* **13**(1), 127–131.

Koreisha, S. G. and Pukkila, T. M. (2004), ‘The specification of vector autoregressive moving average models’, *Journal of Statistical Computation and Simulation* **74**(8), 547–565.

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. (2005), *New Introduction to Multiple Time Series Analysis*, Springer, Springer-Verlag, Berlin Heidelberg, Germany.

Lütkepohl, H. (2006), Forecasting with VARMA models, *in* G. Elliott, C. W. J. Granger and A. Timmermann, eds, ‘Handbook of Economic Forecasting’, Vol. 1, Elsevier, Netherland, chapter 6, pp. 287–325.

Lütkepohl, H. and Claessen, H. (1997), ‘Analysis of cointegrated VARMA processes’, *Journal of Econometrics* **80**(2), 223–239.

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

Maravall, A. (1993), ‘Stochastic linear trends: Models and estimators’, *Journal of Econometrics* **56**, 5–37.

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.

Mauricio, J. A. (2006), ‘Exact maximum likelihood estimation of partially nonstationary vector ARMA models’, *Computational Statistics & Data Analysis* **50**(12), 3644–3662.

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.

Metaxoglou, K. and Smith, A. (2007), ‘Maximum likelihood estimation of VARMA models using a state-space EM algorithm’, *Journal of Time Series Analysis* **28**(5), 666–685.

Min, W. and Tsay, R. S. (2005), ‘On canonical analysis of multivariate time series’, *Satistica Sinica* **15**, 303–323.

Newbold, P. (1974), ‘The exact likelihood function for a mixed autoregressive-moving average process’, *Biometrika* **61**, 423–426.

Nicholls, D. F. and Hall, A. D. (1979), ‘The exact maximum likelihood function of multivariate autoregressive moving average models’, *Biometrika* **66**, 259–264.

Nicholls, D. F. and Hall, A. D. (1980), ‘The evaluation of exact maximum likelihood estimates for VARMA models’, *Journal of Statistical Computation and Simulation* **3**, 251–262.

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.

Paparoditis, E. and Streitberg, B. (1991), ‘Order identification statistics in stationary autoregressive moving-average models: Vector autocorrelations and the bootstrap’, *Journal of Time Series Analysis* **13**(5), 415–434.

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. (2003), ‘On the specification of cointegrated autoregressive moving-average forecasting systems’, *International Journal of Forecasting* **19**, 503–519.

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.

Poskitt, D. S. and Salau, M. O. (1995), ‘On the relationship between generalized least squares and gaussian estimation of vector ARMA models’, *Journal of Time Series Analysis* **16**(6), 617–645.

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

Reinsel, G. C. (1993), *Multivariate Time Series*, Springer-Verlag, New York.

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

Reinsel, G. C., Basu, S. and Yap, S. F. (1992), ‘Maximum likelihood estimators in the multivariate autoregressive moving-average model from a generalized least squares viewpoint’, *Journal of Time Series Analysis* **13**(2), 133–145.

Scott, D. J. (1973), ‘Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach’, *Advances in Applied Probability* **5**, 119–137.

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.

Solo, V. (1984), ‘The exact likelihood for a multivariate ARMA model’, *Journal of Multivariate Analysis* **15**, 164–173.

Solo, V. (1986), *Topics in Advanced Time Series Analysis*, Lecture Notes in Mathematics, Springer Verlag, New York.

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 & Economic Statistics* **7**(3), 327–341.

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

Tsay, R. S. and Tiao, G. C. (1985), 'Use of canonical analysis in time series model identification', *Biometrika* **72**(2), 299–315.

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.