Asymptotic distributions for quasi-efficient estimators in echelon VARMA models

Jean-Marie Dufour [*]
McGill University

Tarek Jouini [†]
University of Windsor

December 2013

[*] Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada; e-mail: jean-marie.dufour@mcgill.ca; Web page: http://www.jeanmariedufour.com; TEL: 1(514) 398 8879; FAX 1(514) 398 4938.

[†](Corresponding Author) Department of Economics, University of Windsor, 1177 Chrysler Hall North Building, 401 Sunset avenue, Windsor, Ontario N9B 3P4, Canada; e-mail: tjouini@uwindsor.ca; TEL: 1(519) 253-3000, Ext. 2381; FAX: 1(519) 973-7096.

ABSTRACT

Two linear estimators for stationary invertible vector autoregressive moving average (VARMA) models in echelon form – to achieve parameter unicity (identification) – with known Kronecker indices are studied. It is shown that both estimators are consistent and asymptotically normal with strong innovations. The first estimator is a generalized-least-squares (GLS) version of the two-step least-squares estimator studied in Dufour and Jouini (2005). The second is an asymptotically efficient estimator which is computationally much simpler than the Gaussian maximum-likelihood (ML) estimator which requires highly nonlinear optimization, and "efficient linear estimators" proposed earlier (Hannan and Kavalieris, *Adv. App. Prob.*, 1984, Reinsel, Basu and Yap, *J. Time Series Anal.*, 1992, and Poskitt and Salau, *J. Time Series Anal.*, 1995). It stands for a new relatively simple three-step estimator based on a linear regression involving innovation estimates which take into account the truncation error of the first-stage long autoregression. The complex dynamic structure of associated residuals is then exploited to derive an efficient covariance matrix estimator of the VARMA innovations, which is of order T^{-1} more accurate than the one by the fourth-stage of Hannan and Kavalieris' procedure. Finally, finite-sample simulation evidence shows that, overall, the asymptotically efficient estimator suggested outperforms its competitors in terms of bias and mean squared errors (MSE) for the models studied.

Keywords: Stationary invertible VARMA; echelon form; Kronecker indices; truncation lag; linear estimation; simulation.
Journal of Economic Literature Classification: C13, C32.

Contents

1 Introduction

Vector autoregressive (VAR) modeling has received considerable attention, especially in time series econometrics; see Lütkepohl (2001, 2005), Hamilton (1994) 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, besides it often requires a very large number of parameters to produce a good fit, the VAR specification is not invariant to many basic linear transformations. For instance, VAR subvectors follow VARMA models. Temporal and contemporaneous aggregations lead to mixed VARMA processes [Lütkepohl (1987)]. Also, trend and seasonal adjustments lead to models outside the VAR class [Maravall (1993)]. The VARMA structure includes VAR models as a special case and can reproduce in a more parsimonious way a broader class of autocovariances and data generating processes, which can improve estimation and forecasting; see Lütkepohl (2006) and Athanasopoulos and Vahid (2008).

VARMA modeling has been proposed long ago [see Hillmer and Tiao (1979), Tiao and Box (1981), Reinsel (1997) and Lütkepohl (2005)] but has been of little use in practice. Indeed, besides fulfilling potentially complex restrictions to achieve identifiability, the task is compounded by the multivariate nature of the data. Once an identifiable specification has been formulated, different estimation methods are considered. But the most studied is ML with strong Gaussian errors; see Hannan (1969*a*), Hannan, Kavalieris and Mackisack (1986), Mauricio (2002, 2006), and Gallego (2009), among others. However, maximizing the exact likelihood in stationary invertible VARMA models is computationally burdensome. Tiao and Box (1981) stressed that it is much easier to maximize a conditional likelihood, though numerical problems still occur with high-dimensional systems in lack of suitable initial values. Recently, Metaxoglou and Smith (2007) studied the identification and ML estimation of VARMA models using EM algorithm-based state-space methods. Although this can yield improvements over earlier ML approaches, we note that recovering the echelon VARMA coefficient estimates from the state-space formulation may not necessarily lead to stationary and invertible models. Further, the Gaussian ML estimation of VARMA models still requires potentially lengthy iterative optimization over a high-dimensional parameter space. Thus, in high-dimensional systems, nonlinear estimation procedures cannot compete with linear methods from the computational cost viewpoint, especially when simulation-based inference is required.

Recursive linear regression methods, initially proposed by Hannan and Rissanen (1982) for ARMA models, have been extended to the VARMA case; see Hannan and Kavalieris (1984), Reinsel, Basu and Yap (1992) and Poskitt and Salau (1995). It consists in estimating, by least squares (LS), the innovations of the VARMA process from a long autoregression to then be used as regressors to estimate the VARMA parameters. Finally, a linear regression on transformed regressors involving newly filtered residuals is performed to achieve efficiency. Note that this multistep linear estimation was initially introduced for model selection and for obtaining consistent estimates which can be used to initialize nonlinear methods, such as ML. The seminal paper by Hannan and Kavalieris (1984) proposed a four-step linear procedure for specifying and estimating stationary ARMAX systems. The first three steps focus on model specification and on providing initial estimates, using Toeplitz regressions based on the Levinson-Whittle algorithm. However, these estimates are substantially biased especially when the ratio of the autoregression-order to the sample size is too large [Hannan and Deistler (1988)]. Finally, using a GLS regression, the fourth stage yields asymptotically efficient estimates. Reinsel et al. (1992) analyzed the ML estimation of VARMA models from a GLS viewpoint. Modulo some approximations allowing for the asymptotic equivalence between GLS and ML, they derived a linear regression with error terms following a moving average (MA) process. However, their analysis underscores the heavy computational burden of the method since it systematically requires the inversion of a high-dimensional weighting matrix. Inspired by Koreisha and Pukkila (1990), Poskitt and Salau (1995) investigated the relationship between the GLS and Gaussian estimation of echelon form VARMA models. Although asymptotically equivalent to ML, their estimates are substantially biased in finite samples. With a simulation study comparing selected linear methods on the quality of the estimates and the accuracy of implied forecasts and impulse responses, Kascha (2007) highlighted the overall superiority of the fourth-stage linear estimation procedure of Hannan and Kavalieris (1984), while noting situations where the investigated methods do not perform very well.

For making VARMA modeling practical, one needs estimation methods that are simple, quick and easy to implement with standard software. More especially as large-sample-approximation-based inference in high-dimensional dynamic models is unreliable, and that simulation-based procedures, such as bootstrap techniques, are rather recommended. However, such methods are impractical if computing the estimator is difficult or time consuming. In this paper, we study two linear estimators for stationary invertible echelon form VARMA models with known Kronecker indices. We focus on the echelon form since it often tends to deliver relatively parsimonious parameterization (involving fewer free parameters) than equivalent identification schemes, such as the final equations form; see Lütkepohl (2005). Our setup easily adapts to cointegrated VARMA and VARMAX framework and alternative identifying schemes. The first estimator is a GLS version of the two-step LS estimator studied in Dufour and Jouini (2005), using a more general setup. The second is a new relatively simple three-step linear estimator which is asymptotically equivalent to ML. Unlike predecessors, it relies on the novelty that consists on using, among the regressors, filtered residuals which take into account the truncation error of the first-stage long autoregression, based on a newly proposed recursive scheme using consistent initial values. It can also be interpreted as a one-step estimator by the scoring method, starting from a $\sqrt{T}$-consistent two-step linear estimator. The proposed estimator is computationally much simpler and more practical than the ML estimator and earlier asymptotically efficient "linear" estimators, namely those suggested by Hannan and Kavalieris (1984), Reinsel et al. (1992), and Poskitt and Salau (1995). As such, both of the estimators studied provide a handy basis for applying resampling inference methods (*e.g.*, bootstrapping).

We show that both estimators are consistent and asymptotically normal with strong innovations. Besides being computationally simpler, our efficient estimator shows distributional theory with explicit formulae of its asymptotic covariance matrix which is relatively simple and easy to estimate for inference purpose. Also, exploiting the complex dynamic structure of the third-stage regression residuals, we derive an efficient covariance estimator of the VARMA innovations, which is of order T^{-1} more accurate than the one by the fourth-stage of Hannan and Kavalieris (1984). Finally, finite-sample simulation evidence shows that two versions of our fully efficient estimator outperform the multistep linear estimators studied.

The paper proceeds as follows. Section **2** presents the echelon form VARMA setup. Section **3** derives the two-step GLS estimator and gives its properties such as convergence and asymptotic normality. Section **4** provides a heuristic derivation of the three-step GLS estimator then states its convergence and asymptotic efficiency. Section **5** shows a comparative simulation study on the finite-sample performance of competing procedures. Finally, Section **6** concludes. Proofs are given in Appendix **A**.

2 Framework

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

$$\Phi(L)\, y_t = \mu_\Phi + \Theta(L)\, u_t, \tag{2.1}$$

where $\Phi(L) = \Phi_0 - \sum_{i=1}^{\bar{p}} \Phi_i L^i$, $\Theta(L) = \Theta_0 + \sum_{j=1}^{\bar{p}} \Theta_j L^j$, $\bar{p} = \max(p_1, \ldots, p_k)$ given a vector of Kronecker indices $(p_1, \ldots, p_k)'$, L denotes the lag operator, $\Theta_0 = \Phi_0$, with Φ_0 a lower-triangular matrix whose all diagonal elements are equal to one, $\mu_\Phi = \Phi(1)\, \mu_y$, with $\mu_y = \mathsf{E}(y_t)$, and $\{u_t : t \in \mathbb{Z}\}$ is a sequence of multivariate innovations. The echelon VARMA operators $\Phi(L) = [\phi_{lm}(L)]_{l,m=1,\ldots,k}$ and $\Theta(L) = [\theta_{lm}(L)]_{l,m=1,\ldots,k}$ are left coprime and satisfy a set of restrictions such that, on any given row l

of $\Phi(L)$ and $\Theta(L)$, $\phi_{lm}(L)$ and $\theta_{lm}(L)$ have the same degree p_l with

$$\begin{aligned} \phi_{lm}(L) &= 1 - \sum_{i=1}^{p_l} \phi_{ll,i} L^i \qquad \text{if } l = m, \\ &= - \sum_{i=p_l - p_{lm}+1}^{p_l} \phi_{lm,i} L^i \quad \text{if } l \neq m, \end{aligned} \tag{2.2}$$

$$\theta_{lm}(L) = \textstyle\sum_{j=0}^{p_l} \theta_{lm,j} L^j \quad \text{with } \Theta_0 = \Phi_0, \tag{2.3}$$

$$\begin{aligned} p_{lm} &= \min(p_l + 1, p_m) \quad \text{for } l \geq m, \\ &= \min(p_l, p_m) \qquad \text{for } l < m, \end{aligned} \tag{2.4}$$

for $l, m = 1, \ldots, k$. Note that $p_l = p_{ll}$ 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 $P = [p_{lm}]_{l,m=1,\ldots,k}$ is the matrix formed by the Kronecker indices. This leads to $\sum_{l=1}^{k} \sum_{m=1}^{k} p_{lm}$ autoregressive (AR) and $k \sum_{l=1}^{k} p_l$ MA free coefficients, respectively. For proofs on the uniqueness of the echelon form and other identification conditions, one should consult Hannan (1969*b*, 1970, 1976, 1979), Deistler and Hannan (1981), Hannan and Deistler (1988), and Lütkepohl (2005, Chapter 12).

The process (2.1) is said to be stationary and invertible with respective pure infinite-order AR and MA representations:

$$\Pi(L) y_t = \mu_\Pi + u_t \qquad \text{and} \qquad y_t = \mu_y + \Psi(L) u_t, \tag{2.5}$$

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

$$\|\Pi_\tau\| \leq C\rho^\tau \qquad \text{and} \qquad \|\Psi_v\| \leq C\rho^v, \tag{2.6}$$

where $\|.\|$ is Schur's norm, $i.e.$ $\|A\|^2 = \mathrm{tr}[A'A]$ for any matrix A. Also, let $\sum_{\tau=0}^{\infty} \Lambda_\tau(\eta) z^\tau = \Theta(z)^{-1}$. Then by invertibility $\|\Lambda_\tau(\eta)\| \leq C\rho^\tau$, where η is the vector of all free varying parameters implied by the echelon form, as shall be specified below.

Now, set $v_t = y_t - u_t$. Then the latter is uncorrelated with the error term u_t since

$$v_t = \Phi_0^{-1} \Big[\mu_\Phi + \sum_{i=1}^{\bar{p}} \Phi_i y_{t-i} + \sum_{j=1}^{\bar{p}} \Theta_j u_{t-j} \Big]. \tag{2.7}$$

Also, let $X_t = \big[1, v_t', y_{t-1}', \ldots, y_{t-\bar{p}}', u_{t-1}', \ldots, u_{t-\bar{p}}'\big]'$ and $\beta = \mathrm{vec}\big[\mu_\Phi, \bar{\Phi}_0, \Phi_1, \ldots, \Phi_{\bar{p}}, \Theta_1, \ldots, \Theta_{\bar{p}}\big]$, where $\bar{\Phi}_0 = I_k - \Phi_0$, be two vectors of respective sizes $kh + 1$ and $k^2 h + k$, with $h = 2\bar{p} + 1$. Then the echelon restrictions (2.1) - (2.4) imply a unique $k^2 h + k$ by r full-rank columns matrix R formed by r selected distinct vectors from the identity matrix $I_{k^2 h + k}$ such that $R'R = I_r$ and $\beta = R\eta$, where η is an r-sized vector of free varying parameters with $r < k^2 h + k$, so that (2.1) takes the form:

$$y_t = \big[X_t' \otimes I_k\big] R\eta + u_t, \tag{2.8}$$

where $\big[X_t' \otimes I_k\big] R$ is a $k \times r$ matrix. Further, under the assumption that the process is regular [by means, with nonsingular covariance matrix of the innovations in the Wold decomposition, so that the process is not linearly predictable and has a nonsingular instantaneous covariance matrix] with continuous distribution, the echelon form ensures that $R'\left[X_t \otimes I_k\right]$ has a nonsingular covariance matrix, so that $\operatorname{rank}\left\{R'\left[\Gamma_X \otimes I_k\right]R\right\} = r$, where $\Gamma_X = \mathsf{E}\left[X_t X_t'\right]$.

3 Generalized two-step linear estimation

Let $\left\{y_{-n_T+1}, \ldots, y_T\right\}$ be a random sample of size $n_T + T$ where n_T is a sequence, function of T, such that $n_T \to \infty$ as $T \to \infty$. Further, consider the infinite-order autoregression (2.5) "*truncated*" at the lag-order n_T, precisely:

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

where $\mu_{\Pi(n_T)}$ and $u_t\left(n_T\right)$ stand respectively for a constant term and a compound innovation, such that $\mu_{\Pi(n_T)} = \left(I_k - \sum_{\tau=1}^{n_T} \Pi_\tau\right)\mu_y$ and $u_t\left(n_T\right) = \sum_{\tau=n_T+1}^{\infty} \Pi_\tau\left(y_{t-\tau} - \mu_y\right) + u_t$. The following assumptions on the VARMA innovations u_t and the truncation order n_T of the long autoregression are needed to establish the consistency and asymptotic distribution of the linear estimators studied below.

Assumption 3.1 *The vectors u_t, $t \in \mathbb{Z}$, are independent and identically distributed $(i.i.d.)$ with mean zero, positive definite (p.d.) covariance matrix $\Sigma_u = \mathsf{E}\left(u_t u_t'\right)$ and continuous distribution.*

Assumption 3.2 *There is a finite constant m_4 such that $\mathsf{E}\left|u_{i,t}u_{j,t}u_{r,t}u_{s,t}\right| \leq m_4 < \infty$, for all t and all $1 \leq i,\ j,\ r,\ s \leq k$.*

Assumption 3.3 *n_T is a function of T such that $n_T \to \infty$ and $n_T^2/T \to 0$ as $T \to \infty$, and, for some $c > 0$, $0 < \delta_1 < 1/2$ and T sufficiently large, $n_T \geq cT^{\delta_1}$.*

Assumption 3.4 *n_T is a function of T such that $n_T \to \infty$ and $n_T^4/T \to 0$ as $T \to \infty$, and, for some $c > 0$, $0 < \delta_2 < 1/4$ and T sufficiently large, $n_T \geq cT^{\delta_2}$.*

Assumption **3.1** entails a strong VARMA process, while Assumption **3.2** ensures that the empirical autocovariances of the process have finite variances. Assumptions **3.3** and **3.4** show alternative conditions on the truncation lag n_T of the first-stage long autoregression, which are required to ensure convergence and asymptotic normality of the estimators suggested. These assumptions state that n_T should grow towards infinity neither too fast nor too slowly. Further, by invertibility, $\|\Pi_\tau\|$ decays at an exponential rate [see (2.6)]. Hence, for some $\delta > 0$, whenever $n_T = cT^{\bar{\delta}}$ for some $c > 0$ and $\bar{\delta} > 0$,

$$T^{\delta} \sum_{\tau=n_T+1}^{\infty} \|\Pi_\tau\| \to 0 \quad \text{as } T \to \infty\,. \tag{3.2}$$

Let $\tilde{\Pi}\left(n_T\right) = \left[\tilde{\mu}_{\Pi(n_T)}, \tilde{\Pi}_1\left(n_T\right), \ldots, \tilde{\Pi}_{n_T}\left(n_T\right)\right] = \tilde{W}_{Y(n_T)}\tilde{\Gamma}_{Y(n_T)}^{-1}$ be the LS estimator of the coefficient matrix $\Pi\left(n_T\right) = \left[\mu_{\Pi(n_T)}, \Pi_1, \ldots, \Pi_{n_T}\right]$, where $\tilde{W}_{Y(n_T)} = T^{-1}\sum_{t=1}^{T} y_t Y_t\left(n_T\right)'$ and $\tilde{\Gamma}_{Y(n_T)} = T^{-1}\sum_{t=1}^{T} Y_t\left(n_T\right)Y_t\left(n_T\right)'$, with $Y_t\left(n_T\right) = \left[1, y_{t-1}', \ldots, y_{t-n_T}'\right]'$. Further, let

$$\tilde{u}_t\left(n_T\right) = y_t - \tilde{\mu}_{\Pi(n_T)} - \sum_{\tau=1}^{n_T} \tilde{\Pi}_\tau\left(n_T\right) y_{t-\tau}, \qquad t = 1, \ldots, T, \tag{3.3}$$

be the LS residuals of the long autoregression (3.1), and set $\tilde{\Sigma}_{u(n_T)} = T^{-1}\sum_{t=1}^{T}\tilde{u}_t(n_T)\,\tilde{u}_t(n_T)'$. Then, under Assumptions **3.1** to **3.3**, and (3.2), Dufour and Jouini (2010) showed that $\big(T^{1/2}/n_T\big)\big\|\tilde{u}_t(n_T) - u_t\big\|$ is stochastically bounded, uniformly in $t = 1, ..., T$, that is

$$\big\|\tilde{u}_t(n_T) - u_t\big\| = O_p\big(n_T/T^{1/2}\big), \quad \text{uniformly in } t = 1, ..., T. \tag{3.4}$$

Then

$$\big\|\tilde{\Sigma}_{u(n_T)} - \Sigma_u\big\|, \big\|\tilde{\Sigma}_{u(n_T)}^{-1} - \Sigma_u^{-1}\big\| = O_p\big(n_T/T^{1/2}\big). \tag{3.5}$$

The asymptotic equivalence stated above suggests that we may be able to estimate consistently the VARMA parameters in (2.8) by replacing the unobserved innovations in X_t with their respective first-stage estimates. Thus, modulo some manipulations, (2.8) can equivalently be rewritten as

$$y_t = \big[\tilde{X}_t(n_T)' \otimes I_k\big]R\eta + e_t(n_T), \tag{3.6}$$

where $\tilde{X}_t(n_T) = \big[1, \tilde{v}_t(n_T)', y_{t-1}', \ldots, y_{t-\bar{p}}', \tilde{u}_{t-1}(n_T)', \ldots, \tilde{u}_{t-\bar{p}}(n_T)'\big]'$, $\tilde{v}_t(n_T) = y_t - \tilde{u}_t(n_T)$ and

$$e_t(n_T) = \tilde{u}_t(n_T) + \sum_{j=0}^{\bar{p}}\Theta_j\big[u_{t-j} - \tilde{u}_{t-j}(n_T)\big]. \tag{3.7}$$

Noting that $\big\|e_t(n_T) - \tilde{u}_t(n_T)\big\| = O_p\big(n_T/T^{1/2}\big)$, in view of (3.7) and (3.4), an explicit two-step (feasible) GLS estimator of η is simply

$$\tilde{\eta} = \arg\min_{\eta}\sum_{t=1}^{T} e_t(n_T)'\,\tilde{\Sigma}_{u(n_T)}^{-1}e_t(n_T) = \tilde{Q}_{X(n_T)}\tilde{W}_{X(n_T)}, \tag{3.8}$$

where $\tilde{Q}_{X(n_T)} = \Big\{R'\big[\tilde{\Gamma}_{X(n_T)} \otimes \tilde{\Sigma}_{u(n_T)}^{-1}\big]R\Big\}^{-1}$ and $\tilde{W}_{X(n_T)} = T^{-1}\sum_{t=1}^{T} R'\big[\tilde{X}_t(n_T) \otimes \tilde{\Sigma}_{u(n_T)}^{-1}\big]y_t$, with $\tilde{\Gamma}_{X(n_T)} = T^{-1}\sum_{t=1}^{T}\tilde{X}_t(n_T)\,\tilde{X}_t(n_T)'$. In addition, let $Q_X = \Big\{R'\big[\Gamma_X \otimes \Sigma_u^{-1}\big]R\Big\}^{-1}$. Then under suitable conditions, Assumptions **3.1** to **3.4**, and (3.2), Dufour and Jouini (2010) have shown that

$$\|\tilde{\eta} - \eta\| = O_p\big(T^{-1/2}\big) \tag{3.9}$$

and

$$T^{1/2}\big(\tilde{\eta} - \eta\big) \underset{T\to\infty}{\overset{d}{\longrightarrow}} N\big[0,\, Q_X\big]. \tag{3.10}$$

Further, they suggested $\tilde{Q}_{X(n_T)}$ as a consistent estimator of Q_X.

Now, let $\tilde{\Sigma}_{e(n_T)} = (T - \bar{p})^{-1}\sum_{t=\bar{p}+1}^{T}\tilde{e}_t(n_T)\,\tilde{e}_t(n_T)'$, where

$$\tilde{e}_t(n_T) = y_t - \big[\tilde{X}_t(n_T)' \otimes I_k\big]R\tilde{\eta}, \qquad t = \bar{p} + 1, \ldots, T. \tag{3.11}$$

Then, using (2.8), (3.4), (3.9) and (3.11), Dufour and Jouini (2010) showed that

$$\big\|\tilde{e}_t(n_T) - u_t\big\| = O_p\big(n_T/T^{1/2}\big), \quad \text{uniformly in } t = \bar{p} + 1, \ldots, T. \tag{3.12}$$

Hence

$$\big\|\tilde{\Sigma}_{e(n_T)} - \Sigma_u\big\|, \big\|\tilde{\Sigma}_{e(n_T)}^{-1} - \Sigma_u^{-1}\big\| = O_p\big(n_T/T^{1/2}\big). \tag{3.13}$$

4 Asymptotic efficiency

The two-step linear estimator described above is not efficient. To allow for efficiency, a further linear regression is needed. As will be shown below, the latter is achieved by exploiting the nonlinear structure of the VARMA innovations in the model parameters. Unlike Hannan and Kavalieris (1984)'s procedure which is heavy to implement, even in small systems, and whose fourth-stage (efficient) estimator does not explicitly show the echelon-form restrictions, we yield a simple and compact efficient estimator with a simple estimator of its covariance matrix. However, a brief description of its competitors is required.

4.1 Competing procedures

Using our setup, we stress that running OLS on (3.6) corresponds to the third-stage and the second-stage estimation procedures of Hannan and Kavalieris (1984) and Reinsel et al. (1992), respectively. Denote by $\tilde{\eta}$ the resulting estimators and let $\tilde{\mu}_\Phi$, $\tilde{\Phi}_i$ and $\tilde{\Theta}_j$ be the implied two-step OLS estimates of μ_Φ, Φ_i and Θ_j, respectively. Further, designate by $\tilde{u}_t$ the implied "implicit" VARMA innovation estimates or residuals such that

$$\tilde{\Phi}(L)\, y_t = \tilde{\mu}_\Phi + \tilde{\Theta}(L)\, \tilde{u}_t, \tag{4.1}$$

where $\tilde{\Phi}(L) = \tilde{\Phi}_0 - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i L^i$ and $\tilde{\Theta}(L) = \sum_{j=0}^{\bar{p}} \tilde{\Theta}_j L^j$, with $\tilde{\Phi}_0 = \tilde{\Theta}_0$. Solving for $\tilde{u}_t$, one gets

$$\tilde{u}_t = \sum_{\tau=0}^{\infty} \Lambda_\tau(\tilde{\eta}) \Big[\tilde{\Phi}_0 y_{t-\tau} - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i-\tau} - \tilde{\mu}_\Phi\Big], \tag{4.2}$$

where $\sum_{\tau=0}^{\infty} \Lambda_\tau(\tilde{\eta})\, L^\tau = \tilde{\Theta}(L)^{-1}$. As suggested in the literature [see Hannan and Kavalieris (1984) and Reinsel et al. (1992)], these implicit residuals, $\tilde{u}_t$, are approximated (or filtered) with

$$\varepsilon_t(\tilde{\eta}) = \sum_{\tau=0}^{t+n_T-1} \Lambda_\tau(\tilde{\eta}) \Big[\tilde{\Phi}_0 y_{t-\tau} - \sum_{i=1}^{\bar{p}} \tilde{\Phi}_i y_{t-i-\tau} - \tilde{\mu}_\Phi\Big], \qquad t = -n_T+1, \ldots, T. \tag{4.3}$$

Hannan-Kavalieris (HK) procedure:

Let $V_t(\tilde{\eta}) = \big[1, y_t' - \varepsilon_t(\tilde{\eta})', y_{t-1}', \ldots, y_{t-\bar{p}}', \varepsilon_{t-1}(\tilde{\eta})', \ldots, \varepsilon_{t-\bar{p}}(\tilde{\eta})'\big]'$ be the regressor vector based on the two-step OLS residuals $\varepsilon_t(\tilde{\eta})$ defined above. Also, set $W_t(\tilde{\eta}) = \sum_{\tau=0}^{t+n_T-1} R'\big[V_{t-\tau}(\tilde{\eta}) \otimes \Lambda_\tau(\tilde{\eta})'\big]$. Then, the efficient estimator of Hannan and Kavalieris (1984) for η is

$$\hat{\eta}_{KH} = \tilde{\eta} + \Big\{\sum_{t=-n_T+1}^{T} W_t(\tilde{\eta})\, \tilde{\Sigma}_{e(n_T)}^{-1} W_t(\tilde{\eta})'\Big\}^{-1} \sum_{t=-n_T+1}^{T} W_t(\tilde{\eta})\, \tilde{\Sigma}_{e(n_T)}^{-1} \varepsilon_t(\tilde{\eta}) \tag{4.4}$$

where $\tilde{\eta}$ and $\tilde{\Sigma}_{e(n_T)}$ are the respective OLS estimators of η and Σ_u obtained from model (3.6). These authors have then proposed $\tilde{\Sigma}_{\varrho(\tilde{\eta},\hat{\eta}_{HK})} = (n_T + T - \bar{p})^{-1} \sum_{t=-n_T+1+\bar{p}}^{T} \varrho_t(\tilde{\eta}, \hat{\eta}_{HK})\, \varrho_t(\tilde{\eta}, \hat{\eta}_{HK})'$, where $\varrho_t(\tilde{\eta}, \hat{\eta}_{HK}) = \varepsilon_t(\tilde{\eta}) - W_t(\tilde{\eta})'(\hat{\eta}_{HK} - \tilde{\eta})$, as the fourth-stage estimator of Σ_u.

Reinsel-Basu-Yap (RBY) procedure:

Manipulating (2.1), the GLS estimator of Reinsel et al. (1992) obtains from the linear regression:

$$y_t(\tilde{\eta}) = \big[V_t(\tilde{\eta})' \otimes I_k\big] R\eta + \sum_{j=0}^{\bar{p}} \tilde{\Theta}_j u_{t-j} + D_t(\tilde{\eta}, \eta), \qquad t = -n_T+1, \ldots, T, \tag{4.5}$$

where $y_t(\tilde{\eta}) = y_t - \varepsilon_t(\tilde{\eta}) + \sum_{j=0}^{\bar{p}} \tilde{\Theta}_j \varepsilon_{t-j}(\tilde{\eta})$ and $D_t(\tilde{\eta}, \eta) = \sum_{j=0}^{\bar{p}} \left(\tilde{\Theta}_j - \Theta_j\right)\left[\varepsilon_{t-j}(\tilde{\eta}) - u_{t-j}\right]$. Dropping the compound term $D_t(\tilde{\eta}, \eta)$ – considered as being negligible – from model (4.5), then setting $y(\tilde{\eta}) = \left[y_{-n_T+1}(\tilde{\eta})', \ldots, y_T(\tilde{\eta})'\right]'$, $V(\tilde{\eta}) = \left[V_{-n_T+1}(\tilde{\eta}), \ldots, V_T(\tilde{\eta})\right]$ and $\tilde{\Theta} = \sum_{j=0}^{\bar{p}} \left[\mathbf{L}^j \otimes \tilde{\Theta}_j\right]$, where $\mathbf{L}^j$ stands for a $(n_T + T) \times (n_T + T)$ lag matrix which has ones on the j^{th} diagonal below the main diagonal and zeros elsewhere ($\mathbf{L}^0$ reduces to the identity matrix), we get the stacked form model

$$y(\tilde{\eta}) = \left[V(\tilde{\eta})' \otimes I_k\right] R\eta + \tilde{\Theta} u, \tag{4.6}$$

where $u = \left[u_{-n_T+1}', \ldots, u_T'\right]'$, with $\tilde{\Theta} u$ having a covariance matrix estimator $\tilde{\Xi}_{\varepsilon(\tilde{\eta})} = \tilde{\Theta}\left[I_{n_T+T} \otimes \tilde{\Sigma}_{\varepsilon(\tilde{\eta})}\right]\tilde{\Theta}'$, where $\tilde{\Sigma}_{\varepsilon(\tilde{\eta})} = (n_T + T)^{-1} \sum_{t=-n_T+1}^{T} \varepsilon_t(\tilde{\eta}) \varepsilon_t(\tilde{\eta})'$ and $\tilde{\Theta}$ is a $k(n_T + T) \times k(n_T + T)$ matrix based on the two-step OLS estimates. Therefore, the GLS estimator of Reinsel et al. (1992) is

$$\hat{\eta}_{RBY} = \left\{R'\left[V(\tilde{\eta}) \otimes I_k\right]\tilde{\Xi}_{\varepsilon(\tilde{\eta})}^{-1}\left[V(\tilde{\eta})' \otimes I_k\right]R\right\}^{-1} R'\left[V(\tilde{\eta}) \otimes I_k\right]\tilde{\Xi}_{\varepsilon(\tilde{\eta})}^{-1} y(\tilde{\eta}), \tag{4.7}$$

thus requiring the burdensome task, even in small samples, of inverting the $k(n_T + T) \times k(n_T + T)$ high-dimensional matrix $\tilde{\Xi}_{\varepsilon(\tilde{\eta})}$. An improved version of this estimator is obtained by deleting the first $k\bar{p}$ components of $y(\tilde{\eta})$ and $\bar{p}$ columns of $V(\tilde{\eta})$ and only retaining the $k(n_T + T - \bar{p}) \times k(n_T + T - \bar{p})$ lower right corner block matrix of $\tilde{\Xi}_{\varepsilon(\tilde{\eta})}$, but it still requires the systematic inversion of a large matrix.

Poskitt-Salau (PS) procedure:

The second-stage estimation procedure of Poskitt and Salau (1995) consists in running LS on a variant of (3.6), precisely

$$\tilde{v}_t(n_T) = \left[\tilde{X}_t(n_T)' \otimes I_k\right] R\eta + \zeta_t, \tag{4.8}$$

where $\zeta_t = \sum_{j=0}^{\bar{p}} \Theta_j \xi_{t-j}$, with $\xi_t = u_t - \tilde{u}_t(n_T)$. Further, set $\tilde{v}(n_T) = \left[\tilde{v}_1(n_T)', \ldots, \tilde{v}_T(n_T)'\right]'$, $\tilde{X}(n_T) = \left[\tilde{X}_1(n_T), \ldots, \tilde{X}_T(n_T)\right]$ and $\zeta = \left[\zeta_1', \ldots, \zeta_T'\right]'$, where $\zeta = \Theta\xi$ and $\xi = \left[\xi_1', \ldots, \xi_T'\right]'$. Then, the efficient GLS estimator of Poskitt and Salau (1995) is

$$\hat{\eta}_{PS} = \left\{R'\left[\tilde{X}(n_T) \otimes I_k\right]\tilde{\Xi}_{u(n_T)}^{-1}\left[\tilde{X}(n_T)' \otimes I_k\right]R\right\}^{-1} R'\left[\tilde{X}(n_T) \otimes I_k\right]\tilde{\Xi}_{u(n_T)}^{-1} \tilde{v}(n_T), \tag{4.9}$$

where, again, one has to invert a $kT \times kT$ high-dimensional matrix $\tilde{\Xi}_{u(n_T)} = \tilde{\Theta}\left[I_T \otimes \tilde{\Sigma}_{u(n_T)}\right]\tilde{\Theta}'$ (estimating the covariance matrix of ζ), with $\tilde{\Theta}$ now corresponding to the OLS moving-average parameter estimates from model (4.8). An improved version of $\hat{\eta}_{PS}$ is obtained in a similar way to $\hat{\eta}_{RBY}$.

4.2 Our procedure

Having shown how our setup is practical and flexible to adapt to alternative procedure, we now derive our efficient linear estimator. In view of (3.7), the two-step feasible GLS (eventually two-step OLS) residuals (3.11) are such that

$$\tilde{e}_t(n_T) = \tilde{u}_t(n_T) + \sum_{j=0}^{\bar{p}} \tilde{\Theta}_j\left[\tilde{u}_{t-j} - \tilde{u}_{t-j}(n_T)\right], \tag{4.10}$$

where, similarly, $\tilde{u}_t$ are the implicit VARMA residuals or estimates of u_t matching the two-step GLS (eventually OLS) estimator $\tilde{\eta}$ since (4.10) can be expressed as (4.1). Indeed, because the error terms $e_t(n_T)$ in (3.7) are functions of the actual innovations u_t, it follows that by estimating $e_t(n_T)$ one implicitly and simultaneously estimates u_t. More importantly, (4.10) reveals that these implicit estimates $\tilde{u}_t$ are endogenous functions not only of the two-step GLS moving average coefficient estimates $\tilde{\Theta}_j$ and the resulting residuals $\tilde{e}_t\left(n_T\right)$ as well, but also of the first-stage OLS autoregressive residuals $\tilde{u}_t\left(n_T\right)$. Hence, using the fact that $\tilde{\Theta}\left(L\right)^{-1}=\sum_{\tau=0}^{\infty}\Lambda_{\tau}\left(\tilde{\eta}\right)L^{\tau}$, one sees that

$$\tilde{u}_t=\tilde{u}_t\left(n_T\right)+\sum_{\tau=0}^{\infty}\Lambda_{\tau}\left(\tilde{\eta}\right)\left[\tilde{e}_{t-\tau}\left(n_T\right)-\tilde{u}_{t-\tau}\left(n_T\right)\right]. \tag{4.11}$$

This paper proposes a new recursive filtering scheme for approximating these implicit residuals with

$$u_t\left(\tilde{\eta}\right)=\tilde{u}_t\left(n_T\right)+\sum_{\tau=0}^{t-1}\Lambda_{\tau}\left(\tilde{\eta}\right)\left[\tilde{e}_{t-\tau}\left(n_T\right)-\tilde{u}_{t-\tau}\left(n_T\right)\right],\qquad t=1,\ldots,T, \tag{4.12}$$

initiating with $\tilde{e}_t\left(n_T\right)=\tilde{u}_t\left(n_T\right)$ [hence $u_t\left(\tilde{\eta}\right)=\tilde{u}_t\left(n_T\right)$] for $1\leq t\leq\bar{p}$. Precisely, our scheme describes the pointwise adjustment mechanism through which the approximate (or filtered) implicit VARMA residuals $u_t\left(\tilde{\eta}\right)$ are recursively computed around $\tilde{u}_t\left(n_T\right)$.

Corollary 4.1 *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in the echelon form given by (2.1)–(2.4). Let also $\tilde{u}_t$ be the implicit VARMA innovation estimates matching the two-step estimator $\tilde{\eta}$, as equivalently defined in (4.2) or (4.11) but respectively approximated with (4.3) and (4.12). Then, under Assumptions* **3.1**–**3.4**,

$$\left\|\tilde{u}_t-\varepsilon_t\left(\tilde{\eta}\right)\right\|=O_p\left(\rho^{t+n_T}\right)\quad\textit{and}\quad\left\|\tilde{u}_t-u_t\left(\tilde{\eta}\right)\right\|=O_p\Big(\rho^t\frac{n_T}{T^{1/2}}\Big). \tag{4.13}$$

Obviously, the recursive schemes (4.3) and (4.12) yield approximations with different (pointwise) convergence speeds towards the implicit VARMA residuals $\tilde{u}_t$, regardless of the persistence degree of the process and the estimation method (OLS or GLS) used for obtaining the two-step VARMA parameter estimates. However, while noting that we loose n_T observations with our recursive scheme, we stress that this is compensated with the use of better initial values, namely the first-stage autoregressive residuals that we know are consistent; see (3.4). Of course, the recursive schemes above are asymptotically equivalent only when the Kronecker indices are all equal, namely when GLS reduces to OLS.

Similarly, it is worth emphasizing that the VARMA innovation u_t can be expressed from (3.7) as

$$u_t=\tilde{u}_t\left(n_T\right)+\sum_{\tau=0}^{\infty}\Lambda_{\tau}\left(\eta\right)\left[e_{t-\tau}\left(n_T\right)-\tilde{u}_{t-\tau}\left(n_T\right)\right], \tag{4.14}$$

and then be approximated with

$$u_t\left(\eta\right)=\tilde{u}_t\left(n_T\right)+\sum_{\tau=0}^{t-1}\Lambda_{\tau}\left(\eta\right)\left[e_{t-\tau}\left(n_T\right)-\tilde{u}_{t-\tau}\left(n_T\right)\right],\qquad t=1,\ldots,T. \tag{4.15}$$

Hence, $\|u_t-u_t\left(\eta\right)\|=O_p\left(\rho^t n_T/T^{1/2}\right)$, in view of (3.4). Also, let $\tilde{\Sigma}_{u(\tilde{\eta})}=T^{-1}\sum_{t=1}^{T}u_t\left(\tilde{\eta}\right)u_t\left(\tilde{\eta}\right)'$. Then its rate of convergence to Σ_u follows.

Proposition 4.1 *Let $\{y_t : t\in\mathbb{Z}\}$ be a k-dimensional stationary invertible stochastic process with the VARMA representation in the echelon form given by (2.1)–(2.4). Then, under Assumptions* **3.1**–**3.4**,

$$\big\|\tilde{\Sigma}_{u(\tilde{\eta})}-\Sigma_u\big\|,\big\|\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}-\Sigma_u^{-1}\big\|=O_p\big(T^{-1/2}\big). \tag{4.16}$$

Now, let $X_t(\tilde{\eta}) = \left[1, v_t(\tilde{\eta})', y'_{t-1}, \ldots, y'_{t-\bar{p}}, u_{t-1}(\tilde{\eta})', \ldots, u_{t-\bar{p}}(\tilde{\eta})'\right]'$ with $v_t(\tilde{\eta}) = y_t - u_t(\tilde{\eta})$. Further, set $Z_t^{\circ}(\tilde{\eta}, \eta) = \sum_{\tau=0}^{t-1} R'\left[X_{t-\tau}(\tilde{\eta}) \otimes \Lambda_\tau(\eta)'\right]$. Then manipulating (4.15) and (4.12), one gets

$$u_t(\tilde{\eta}) - u_t(\eta) = -Z_t^{\circ}(\tilde{\eta}, \eta)'\left(\tilde{\eta} - \eta\right). \tag{4.17}$$

The latter expression can further be rearranged to obtain the linear regression model

$$\omega_t(\tilde{\eta}) = Z_t(\tilde{\eta})'\eta + \epsilon_t(\tilde{\eta}, \eta), \tag{4.18}$$

where

$$\omega_t(\tilde{\eta}) = u_t(\tilde{\eta}) + Z_t(\tilde{\eta})'\tilde{\eta} \quad \text{and} \quad \epsilon_t(\tilde{\eta}, \eta) = u_t(\eta) + \left[Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta}, \eta)\right]'\left(\tilde{\eta} - \eta\right), \tag{4.19}$$

with $Z_t(\tilde{\eta}) = \sum_{\tau=0}^{t-1} R'\left[X_{t-\tau}(\tilde{\eta}) \otimes \Lambda_\tau(\tilde{\eta})'\right]$. Note that (4.17) is an identity obtained by exploiting the nonlinear structure of the VARMA innovations in the model parameters. So it does not stand for a Taylor expansion. More importantly, the complex dynamic structure of the error terms $\epsilon_t(\tilde{\eta}, \eta)$ driving the process (4.18) – missed by Hannan and Kavalieris (1984) in their fourth stage – is completely specified up to the unknown parameter vector η; see (4.19). Hence, once estimated, these errors provide a closed form solution for computing accurately the approximate implicit VARMA residuals or innovation estimates matching the three-step efficient linear estimator that we shall define below. Such a result has not been established yet in the literature.

In view of (4.17) and (4.19) [or (4.18) and (4.19)], one sees, by Lemma **2.2** of Kreiss and Franke (1992) and (3.9), that $\left\|\epsilon_t(\tilde{\eta}, \eta) - u_t(\tilde{\eta})\right\| = O_p\left(T^{-1/2}\right)$, which suggests obtaining a third-stage GLS (fully efficient) linear estimator of η, say $\hat{\eta}$, such that

$$\hat{\eta} = \arg\min_{\eta} \sum_{t=1}^{T} \epsilon_t(\tilde{\eta}, \eta)'\, \tilde{\Sigma}_{u(\tilde{\eta})}^{-1} \epsilon_t(\tilde{\eta}, \eta) = \tilde{Q}_{X(\tilde{\eta})} \tilde{W}_{X(\tilde{\eta})}, \tag{4.20}$$

where $\tilde{Q}_{X(\tilde{\eta})} = \left\{T^{-1}\sum_{t=1}^{T} Z_t(\tilde{\eta})\, \tilde{\Sigma}_{u(\tilde{\eta})}^{-1} Z_t(\tilde{\eta})'\right\}^{-1}$ and $\tilde{W}_{X(\tilde{\eta})} = T^{-1}\sum_{t=1}^{T} Z_t(\tilde{\eta})\, \tilde{\Sigma}_{u(\tilde{\eta})}^{-1} \omega_t(\tilde{\eta})$. Further, let $\tilde{\Omega}_{X(\tilde{\eta})} = T^{-1}\sum_{t=1}^{T} Z_t(\tilde{\eta})\, \tilde{\Sigma}_{u(\tilde{\eta})}^{-1} u_t(\tilde{\eta})$. Then, in view of $\omega_t(\tilde{\eta})$ [see (4.19)],

$$\hat{\eta} = \tilde{\eta} + \tilde{Q}_{X(\tilde{\eta})} \tilde{\Omega}_{X(\tilde{\eta})}. \tag{4.21}$$

Clearly, our third-stage GLS estimators are different from their competitors since alternative regressors and weighting matrix are used in their computation. Precisely, we exploit the explicit form of the second-stage regression residuals to derive a new recursive filtering scheme for approximating the implicit VARMA residuals matching the two-step estimator [see (4.12)]. These well behaved approximate residuals stand for ”new regressors” which, unlike predecessors [see (4.3)], depend on consistent (better) initial values, and explicitly take into account the truncation error of the first-stage autoregression along with some adjustments with respect to the second-stage regression residuals. Finally, it is noteworthy that $\hat{\eta}$ is asymptotically equivalent to ML under Gaussian errors since $\frac{\partial u_t(\eta)}{\partial \eta'}|_{\eta=\tilde{\eta}} = -Z_t(\tilde{\eta})'$ [see (4.17)], and that it corresponds to an iteration of the scoring algorithm starting from $\tilde{\eta}$, in view of (4.21).

Another feature characterizing the computation of our fully efficient estimators, compared to those of Hannan and Kavalieris (1984) and Poskitt and Salau (1995), with the exception of Reinsel et al. (1992), consists in using a weighting matrix exhibiting faster rate of convergence, hence better sample properties; see (3.13) and (3.5) versus Propositions **4.1**. However, we stress that, although Reinsel et al. (1992) procedure’s relies on a refined weighting matrix, it still uses filtered residuals from an alternative scheme.

Now, let $\tilde{Q}^{\circ}_{X(\tilde{\eta})} = \Big\{ T^{-1} \sum_{t=1}^{T} Z^{\circ}_t(\tilde{\eta}, \eta)\, \tilde{\Sigma}^{-1}_{u(\tilde{\eta})} Z^{\circ}_t(\tilde{\eta}, \eta)' \Big\}^{-1}$ and $Q_{X(\eta)} = \Big\{ \mathsf{E}\big[Z_t \Sigma_u^{-1} Z_t' \big] \Big\}^{-1}$, with $Z_t = \sum_{\tau=0}^{\infty} R' \big[X_{t-\tau} \otimes \Lambda_\tau(\eta)' \big]$. Also, denote by $\|A\|_1^2$ the largest eigenvalue of $A'A$, for any matrix A.

Proposition 4.2 *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA representation in the echelon form given by* (2.1)–(2.4)*. Then, under Assumptions* **3.1–3.4**,

$$\big\| \tilde{Q}^{\circ}_{X(\tilde{\eta})} - Q_{X(\eta)} \big\|_1, \big\| \tilde{Q}_{X(\tilde{\eta})} - \tilde{Q}^{\circ}_{X(\tilde{\eta})} \big\|_1 = O_p\big(T^{-1/2}\big). \tag{4.22}$$

The next theorems establish the convergence and the asymptotic normality of our efficient estimator.

Theorem 4.1 *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA representation in the echelon form given by* (2.1)–(2.4)*. Then, under Assumptions* **3.1–3.4**,

$$\|\hat{\eta} - \eta\| = O_p\big(T^{-1/2}\big). \tag{4.23}$$

Theorem 4.2 *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA representation in the echelon form given by* (2.1)–(2.4)*. Then, under Assumptions* **3.1–3.4**,

$$T^{1/2}\big(\hat{\eta} - \eta\big) \underset{T \to \infty}{\overset{d}{\longrightarrow}} N\big[0, Q_{X(\eta)}\big]. \tag{4.24}$$

A consistent estimator of its asymptotic covariance matrix is then $\Big\{ \sum_{t=1}^{T} Z_t(\tilde{\eta})\, \tilde{\Sigma}^{-1}_{u(\tilde{\eta})} Z_t(\tilde{\eta})' \Big\}^{-1}$. As mentioned above with respect to (4.19), we suggest better filtering accurately, from the third-stage regression residuals $\epsilon_t(\tilde{\eta}, \hat{\eta})$, well-behaved VARMA innovation estimates in finite samples, say $u_t(\hat{\eta})$, such that:

$$\epsilon_t(\tilde{\eta}, \hat{\eta}) = u_t(\hat{\eta}) + \big[Z_t(\tilde{\eta}) - Z^{\circ}_t(\tilde{\eta}, \hat{\eta}) \big]' \big(\tilde{\eta} - \hat{\eta}\big), \qquad t = \bar{p} + 1, \ldots, T, \tag{4.25}$$

where $\epsilon_t(\tilde{\eta}, \hat{\eta}) = \omega_t(\tilde{\eta}) - Z_t(\tilde{\eta})' \hat{\eta}$ and $Z^{\circ}_t(\tilde{\eta}, \hat{\eta}) = \sum_{\tau=0}^{t-1} R' \big[X_{t-\tau}(\tilde{\eta}) \otimes \Lambda_\tau(\hat{\eta})' \big]$. Finally, let $\tilde{\Sigma}_{u(\hat{\eta})} = (T - \bar{p})^{-1} \sum_{t=\bar{p}+1}^{T} u_t(\hat{\eta})\, u_t(\hat{\eta})'$ be the resulting third-stage efficient estimator of the VARMA innovation covariance matrix Σ_u. Then its rate of convergence follows.

Proposition 4.3 *Let* $\{y_t : t \in \mathbb{Z}\}$ *be a* k*-dimensional stationary invertible stochastic process with the VARMA representation in the echelon form given by* (2.1)–(2.4)*. Then, under Assumptions* **3.1–3.4**,

$$\big\| \tilde{\Sigma}_{u(\hat{\eta})} - \Sigma_u \big\| = O_p\big(T^{-1/2}\big). \tag{4.26}$$

To roughly show to which extent $\tilde{\Sigma}_{\varrho(\tilde{\eta}, \hat{\eta}_{HK})}$ is less accurate than $\tilde{\Sigma}_{u(\hat{\eta})}$ in estimating Σ_u in finite samples, assume for simplicity that the HK procedure uses $u_t(\tilde{\eta})$ and $\tilde{\Sigma}_{u(\tilde{\eta})}$ instead of $\varepsilon_t(\tilde{\eta})$ and $\tilde{\Sigma}_{e(n_T)}$. Therefore, $W_t(\tilde{\eta}) = Z_t(\tilde{\eta})$, $\hat{\eta}_{HK} = \hat{\eta}$ and then $\varrho_t(\tilde{\eta}, \hat{\eta}_{HK}) = \epsilon_t(\tilde{\eta}, \hat{\eta})$. Hence, in view of (4.25), our well-behaved error covariance estimator suggested above is of order T^{-1} more accurate than the one by the fourth-stage of Hannan and Kavalieris (1984) in estimating the VARMA innovation covariance matrix Σ_u, since

$$\big\| \varrho_t(\tilde{\eta}, \hat{\eta}_{HK}) - u_t \big\| = \big\| u_t(\hat{\eta}) - u_t \big\| + O_p\big(T^{-1}\big), \qquad t = \bar{p} + 1, \ldots, T. \tag{4.27}$$

5 Simulation study

The small-sample performance of our proposed estimators is studied with Monte Carlo (MC) simulations. We only focus on the fully efficient estimates since they stand for the major contribution of the paper.

Specifically, we consider a comparative study involving those suggested by Hannan and Kavalieris (1984) (HK), Reinsel et al. (1992) (RBY) and Poskitt and Salau (1995) (PS), respectively. In these simulations, the improved versions of the last two estimators described above were used. In addition, two versions of our proposed three-step estimator, say TS1 and TS2, were considered. The first one relies on the two-step GLS estimator given in (3.8), while the second is based on the two-step OLS estimator studied in Dufour and Jouini (2005). Obviously TS1 and TS2 are identical when the Kronecker indices characterizing the echelon canonical form are all equal. While noting that a two step OLS estimation has been used for obtaining the GLS estimators of Hannan and Kavalieris (1984) and Reinsel et al. (1992), those of Poskitt and Salau (1995) were obtained by implementing their three-step procedure in full. Of course, all competing (fully efficient) estimators are asymptotically equivalent to ML estimators since they roughly correspond to one iteration of the Gauss-Newton algorithm, starting from a $\sqrt{T}$-consistent estimator. Finally, ML estimation was omitted in the simulations for the following reasons. First, its finite sample properties have been extensively studied in the literature and were found more or less satisfactory given the model at hand. Second, besides the fact that state-space formulation based ML estimation of VARMA models still requires potentially high evaluations of the EM algorithm, more especially in big or persistent systems, it also fails to handle the parsimonious echelon form parameterization since it is not guaranteed that the resulting estimated echelon VARMA models are stationary and invertible. Third, in big systems, nonlinear estimation procedures cannot compete with linear methods from the computational cost viewpoint, especially for simulation-based inference using bootstrap methods or maximized Monte Carlo (MMC) tests [see Dufour (2006) and Dufour and Jouini (2006)]. Finally, as the paper deals with efficient linear estimation methods for VARMA models, we only studied the finite-sample performance of the main procedures compared to the ones we suggested.

We simulate two bivariate stationary invertible Gaussian ARMA processes with drifts and respective Kronecker indices $(1, 2)$ and $(2, 1)$, using sample sizes 100 and 200. Simulation results on the bias (in absolute value) and MSE of the estimates for each procedure are given in Tables 1–4. These tables also show the MSE ratios of the alternative fully efficient estimators with respect to TS1. These results are based on 1000 replications using GAUSS random number generator. To avoid numerical problems due to initialization, extra first 100 pseudo-data were generated then discarded. Trials associated with estimates implying noninvertible VARMA processes are thrown then replaced. In all simulations, the rate of replacement did not exceed 5% in the worst case. The two-step echelon parameter estimates were obtained from models using, as regressors, autoregressive residuals associated with autoregression truncation set to the integer part of $\ln T$ then $T^{-1/2}$, since it has been recommended in the literature to choose the truncation order between these two values. This strategy has been considered to draw the effect of the first-stage autoregression lag-order choice on the finite sample properties of the echelon parameter estimates. The error covariance matrix with $\sigma_{11} = .49$, $\sigma_{22} = .29$ and $\sigma_{12} = \sigma_{21} = -.14$, is used for both simulated models. The parameter values of the simulated echelon VARMA models as well as the resulting eigenvalues (describing the persistence degree of the model) are given in the tables. For a better comparison with HK and RBY procedures, the latter are finally computed after discarding the first n_T values of the residuals $\varepsilon_t(\tilde{\eta})$ [namely, $\varepsilon_{-n_T+1}(\tilde{\eta})$,..., $\varepsilon_0(\tilde{\eta})$; see (4.3)] to avoid, though partially, problems due to initialization since preliminary simulations (that we omitted) showed poor HK estimates.

For both models, simulation evidence shows that, unlike TS1, TS2 and RBY methods whose respective estimates show small to moderate bias, HK and PS procedures yield estimates with substantial bias associated with relatively significant MSE for $T = 100$ [see upper panels of Tables 1 and 3]. These biases decrease with the sample size [see Tables 1–4]. It is suspected that the bias associated with PS procedure is due to the weighting matrix used in the computation of the estimates. Poskitt and Salau (1995) argued that the error term in their linear regression follows a moving-average process of order $\bar{p}$, namely $\zeta_t = \sum_{j=0}^{\bar{p}} \Theta_j \xi_{t-j}$ with $T^{-1} \sum_{t=1}^{T} \xi_t \xi_t' = O_p\big(n_T/T\big)\Sigma_u$ [see Hannan and Kavalieris (1986) and

Table 1: Echelon VARMA model with Kronecker indices (1,2): a comparative simulation study on alternative fully efficient GLS estimators

Sample Size $T = 100$																
n_T	Coefficient	Value	Bias					MSE					MSE Ratio			
			TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS2	HK	RBY	PS
4	$\mu_{\Phi,1}$	.000	.009	.010	.001	.011	.009	.200	.200	.257	.204	.190	1.000	1.286	1.020	.952
	$\mu_{\Phi,2}$	.000	.003	.004	.001	.004	.000	.145	.146	.179	.151	.123	1.006	1.231	1.039	.845
	$\phi_{11,1}$	1.200	.020	.020	.005	.018	.008	.056	.057	.079	.057	.051	1.001	1.390	1.012	.896
	$\phi_{12,1}$	.240	.000	.000	.003	.001	.009	.046	.046	.062	.047	.044	1.001	1.361	1.018	.969
	$\phi_{22,1}$	.400	.005	.000	.015	.000	.135	.111	.106	.134	.115	.181	.956	1.205	1.033	1.622
	$\phi_{21,2}$	-.900	.005	.008	.013	.006	.089	.078	.074	.094	.077	.121	.950	1.212	.987	1.558
	$\phi_{22,2}$	-.270	.002	.000	.005	.002	.048	.068	.068	.086	.073	.083	1.000	1.270	1.082	1.224
	$\theta_{11,1}$	.800	.015	.014	.025	.013	.210	.096	.097	.111	.097	.219	1.004	1.153	1.013	2.274
	$\theta_{21,1}$	.500	.007	.004	.025	.002	.081	.090	.089	.102	.095	.115	.994	1.136	1.060	1.274
	$\theta_{12,1}$	.400	.018	.017	.104	.024	.213	.117	.118	.185	.126	.239	1.006	1.583	1.081	2.043
	$\theta_{22,1}$	.400	.037	.030	.059	.033	.168	.135	.127	.160	.148	.220	.941	1.185	1.102	1.630
	$\theta_{21,2}$	.340	.035	.028	.004	.030	.334	.165	.154	.164	.166	.376	.935	.993	1.004	2.274
	$\theta_{22,2}$	.850	.073	.067	.204	.078	.406	.159	.153	.261	.159	.418	.960	1.639	.997	2.626
10	$\mu_{\Phi,1}$	.000	.002	.003	.002	.002	.002	.206	.208	.266	.211	.199	1.009	1.291	1.025	.969
	$\mu_{\Phi,2}$	.000	.005	.005	.002	.004	.003	.169	.168	.210	.169	.155	.994	1.240	1.003	.919
	$\phi_{11,1}$	1.200	.024	.025	.024	.023	.022	.062	.062	.079	.064	.060	1.009	1.284	1.038	.971
	$\phi_{12,1}$	.240	.000	.000	.003	.000	.000	.046	.047	.073	.050	.046	1.007	1.564	1.068	.988
	$\phi_{22,1}$	.400	.006	.003	.003	.003	.020	.105	.102	.110	.111	.107	.968	1.046	1.056	1.018
	$\phi_{21,2}$	-.900	.014	.012	.005	.007	.003	.081	.078	.088	.079	.078	.960	1.082	.983	.970
	$\phi_{22,2}$	-.270	.003	.002	.001	.000	.007	.064	.064	.075	.070	.064	.996	1.174	1.090	.990
	$\theta_{11,1}$	.800	.012	.009	.024	.000	.053	.100	.100	.102	.103	.106	1.006	1.020	1.031	1.062
	$\theta_{21,1}$	.500	.000	.001	.005	.006	.009	.090	.090	.096	.095	.094	.999	1.066	1.051	1.036
	$\theta_{12,1}$	.400	.011	.009	.025	.003	.040	.122	.124	.161	.135	.133	1.019	1.321	1.109	1.092
	$\theta_{22,1}$	.400	.028	.028	.032	.026	.056	.126	.122	.127	.132	.127	.974	1.006	1.047	1.007
	$\theta_{21,2}$	.340	.019	.022	.039	.026	.066	.160	.156	.161	.172	.189	.974	1.004	1.077	1.185
	$\theta_{22,2}$	.850	.050	.051	.076	.033	.090	.147	.148	.154	.153	.156	1.005	1.043	1.037	1.054

Note – TS1 and TS2 are the three-step GLS estimators based on the two-step GLS estimator and the two-step OLS estimator, respectively. While HK, RBY and PS stand for the fully efficient GLS estimators suggested by Hannan and Kavalieris (1984), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. These estimates are obtained with 1000 replications. The eigenvalues of the model are real .900, .400 and .300 for the autoregressive (AR) operator, and real .824 and conjugate -.188$\mp$.790i (.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.

Table 2: Echelon VARMA model with Kronecker indices (1,2): a comparative simulation study on alternative fully efficient GLS estimators

Sample Size $T = 200$																	
n_T	Coefficient	Value	Bias					MSE					MSE Ratio			
			TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS2	HK	RBY	PS
5	$\mu_{\Phi,1}$	.000	.001	.001	.003	.001	.000	.114	.114	.177	.116	.110	1.001	1.547	1.016	.961
	$\mu_{\Phi,2}$	.000	.000	.000	.000	.000	.000	.094	.094	.119	.093	.094	1.003	1.268	.996	1.006
	$\phi_{11,1}$	1.200	.012	.012	.004	.010	.011	.038	.038	.046	.038	.037	.999	1.208	1.016	.972
	$\phi_{12,1}$	.240	.001	.001	.013	.000	.002	.030	.030	.046	.032	.030	.999	1.488	1.048	.991
	$\phi_{22,1}$	.400	.000	.002	.011	.004	.012	.062	.060	.073	.065	.076	.968	1.189	1.056	1.232
	$\phi_{21,2}$	-.900	.005	.006	.003	.001	.003	.044	.044	.053	.046	.051	.982	1.196	1.029	1.157
	$\phi_{22,2}$	-.270	.000	.001	.003	.001	.015	.040	.039	.045	.041	.046	.984	1.148	1.038	1.164
	$\theta_{11,1}$	.800	.002	.000	.035	.002	.111	.060	.061	.078	.064	.123	1.018	1.301	1.058	2.038
	$\theta_{21,1}$	.500	.000	.000	.005	.005	.028	.058	.058	.064	.062	.065	.999	1.085	1.063	1.117
	$\theta_{12,1}$	.400	.005	.003	.081	.009	.093	.075	.076	.134	.085	.130	1.007	1.781	1.132	1.720
	$\theta_{22,1}$	.400	.011	.008	.029	.011	.023	.073	.070	.080	.081	.080	.961	1.103	1.095	1.101
	$\theta_{21,2}$	.340	.008	.005	.041	.016	.063	.097	.094	.118	.104	.136	.966	1.212	1.068	1.391
	$\theta_{22,2}$	.850	.030	.028	.118	.031	.206	.087	.087	.157	.094	.218	.993	1.794	1.074	2.490
14	$\mu_{\Phi,1}$	.000	.004	.004	.004	.003	.004	.115	.116	.170	.120	.112	1.006	1.472	1.037	.969
	$\mu_{\Phi,2}$	.000	.005	.005	.005	.004	.004	.093	.094	.115	.097	.090	1.000	1.232	1.036	.968
	$\phi_{11,1}$	1.200	.011	.011	.012	.011	.010	.038	.039	.052	.040	.038	1.006	1.351	1.038	.985
	$\phi_{12,1}$	.240	.000	.000	.001	.000	.000	.031	.031	.045	.033	.032	1.001	1.433	1.057	1.012
	$\phi_{22,1}$	.400	.000	.000	.000	.001	.013	.062	.060	.063	.063	.067	.975	1.017	1.020	1.081
	$\phi_{21,2}$	-.900	.005	.005	.005	.004	.002	.047	.046	.048	.048	.049	.985	1.028	1.025	1.035
	$\phi_{22,2}$	-.270	.000	.000	.001	.000	.005	.039	.039	.043	.041	.040	.985	1.085	1.029	1.014
	$\theta_{11,1}$	.800	.001	.004	.000	.011	.019	.063	.064	.064	.067	.061	1.015	1.017	1.062	.972
	$\theta_{21,1}$	.500	.001	.001	.002	.002	.005	.060	.059	.064	.062	.063	.997	1.077	1.035	1.056
	$\theta_{12,1}$	.400	.000	.002	.002	.008	.017	.079	.079	.086	.083	.083	1.004	1.096	1.054	1.058
	$\theta_{22,1}$	.400	.010	.008	.010	.006	.020	.072	.071	.074	.072	.071	.981	1.032	1.003	.991
	$\theta_{21,2}$	.340	.009	.008	.010	.005	.030	.096	.094	.097	.100	.109	.982	1.016	1.042	1.138
	$\theta_{22,2}$	.850	.021	.020	.027	.009	.038	.088	.087	.086	.089	.089	.984	.981	1.011	1.008

Note – TS1 and TS2 are the three-step GLS estimators based on the two-step GLS estimator and the two-step OLS estimator, respectively. While HK, RBY and PS stand for the fully efficient GLS estimators suggested by Hannan and Kavalieris (1984), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. These estimates are obtained with 1000 replications. The eigenvalues of the model are real .900, .400 and .300 for the autoregressive (AR) operator, and real .824 and conjugate $-.188\mp.790i$ (.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.

Table 3: Echelon VARMA model with Kronecker indices (2,1): a comparative simulation study on alternative fully efficient GLS estimators

			Sample Size $T = 100$													
n_T	Coefficient	Value	Bias					MSE					MSE Ratio			
			TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS2	HK	RBY	PS
4	$\mu_{\Phi,1}$	.000	.001	.001	.001	.000	.001	.158	.159	.237	.166	.164	1.007	1.503	1.053	1.040
	$\mu_{\Phi,2}$	.000	.004	.004	.006	.005	.001	.188	.189	.199	.190	.180	1.006	1.059	1.015	.958
	$\phi_{21,0}$	.500	.003	.004	.005	.005	.003	.033	.033	.062	.036	.033	1.020	1.873	1.093	1.004
	$\phi_{11,1}$	1.800	.002	.001	.030	.000	.010	.034	.034	.063	.039	.034	1.009	1.817	1.134	1.001
	$\phi_{21,1}$	-.400	.037	.041	.016	.045	.038	.096	.100	.179	.112	.095	1.033	1.852	1.155	.982
	$\phi_{22,1}$	.800	.064	.069	.038	.075	.067	.144	.149	.252	.166	.142	1.029	1.746	1.148	.982
	$\phi_{11,2}$	-.360	.005	.007	.157	.011	.053	.111	.112	.246	.130	.116	1.007	2.212	1.169	1.042
	$\phi_{12,2}$	-.900	.012	.015	.251	.021	.083	.169	.169	.382	.196	.176	1.002	2.261	1.163	1.042
	$\theta_{11,1}$	.330	.055	.055	.024	.043	.118	.130	.131	.214	.143	.160	1.012	1.651	1.105	1.237
	$\theta_{21,1}$	-.180	.016	.016	.085	.000	.087	.108	.109	.209	.128	.132	1.008	1.924	1.179	1.213
	$\theta_{12,1}$	-.200	.021	.022	.066	.014	.111	.141	.144	.183	.154	.169	1.019	1.297	1.090	1.201
	$\theta_{22,1}$	-.400	.072	.080	.126	.071	.188	.176	.184	.336	.202	.233	1.043	1.905	1.146	1.320
	$\theta_{11,2}$	-.200	.061	.064	.098	.055	.070	.138	.140	.255	.158	.129	1.014	1.849	1.145	.940
	$\theta_{12,2}$	.920	.024	.032	.334	.015	.191	.205	.210	.473	.243	.253	1.027	2.307	1.185	1.234
10	$\mu_{\Phi,1}$	.000	.000	.000	.005	.000	.001	.173	.173	.198	.175	.183	1.002	1.143	1.011	1.060
	$\mu_{\Phi,2}$	.000	.000	.000	.002	.000	.001	.208	.208	.217	.209	.208	.999	1.042	1.004	1.001
	$\phi_{21,0}$	.500	.000	.001	.000	.001	.006	.040	.040	.048	.041	.041	1.005	1.190	1.011	1.029
	$\phi_{11,1}$	1.800	.001	.002	.000	.005	.005	.038	.040	.046	.044	.039	1.040	1.197	1.139	1.028
	$\phi_{21,1}$	-.400	.043	.047	.039	.047	.030	.116	.118	.142	.121	.112	1.018	1.220	1.040	.966
	$\phi_{22,1}$	.800	.081	.086	.076	.085	.069	.172	.175	.208	.179	.169	1.017	1.208	1.043	.984
	$\phi_{11,2}$	-.360	.022	.020	.034	.011	.056	.115	.119	.145	.128	.127	1.031	1.259	1.112	1.103
	$\phi_{12,2}$	-.900	.046	.043	.066	.031	.098	.173	.176	.218	.187	.195	1.021	1.260	1.082	1.126
	$\theta_{11,1}$	.330	.061	.061	.051	.052	.075	.139	.141	.181	.146	.151	1.015	1.302	1.049	1.084
	$\theta_{21,1}$	-.180	.013	.013	.003	.007	.000	.123	.123	.162	.133	.123	.993	1.313	1.076	.994
	$\theta_{12,1}$	-.200	.031	.030	.032	.025	.045	.148	.152	.168	.157	.158	1.025	1.132	1.062	1.064
	$\theta_{22,1}$	-.400	.095	.100	.097	.086	.089	.213	.219	.258	.224	.213	1.024	1.208	1.049	.997
	$\theta_{11,2}$	-.200	.062	.063	.040	.068	.034	.143	.148	.187	.158	.132	1.035	1.311	1.103	.926
	$\theta_{12,2}$	.920	.071	.073	.107	.043	.122	.226	.239	.265	.248	.237	1.058	1.170	1.095	1.048

Note – TS1 and TS2 are the three-step GLS estimators based on the two-step GLS estimator and the two-step OLS estimator, respectively. While HK, RBY and PS stand for the fully efficient GLS estimators suggested by Hannan and Kavalieris (1984), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. These estimates are obtained with 1000 replications. The eigenvalues of the model are real .800 and a double root .900 for the autoregressive (AR) operator, and real -.530 and conjugate $-.350 \mp .584i$ (.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.

Table 4: Echelon VARMA model with Kronecker indices (2,1): a comparative simulation study on alternative fully efficient GLS estimators

			Sample Size $T = 200$													
n_T	Coefficient	Value	Bias					MSE					MSE Ratio			
			TS1	TS2	HK	RBY	PS	TS1	TS2	HK	RBY	PS	TS2	HK	RBY	PS
5	$\mu_{\Phi,1}$	.000	.000	.000	.002	.000	.000	.078	.079	.103	.080	.081	1.002	1.306	1.019	1.027
	$\mu_{\Phi,2}$	.000	.000	.000	.000	.000	.000	.083	.083	.084	.083	.081	1.003	1.013	.999	.978
	$\phi_{21,0}$	.500	.001	.001	.002	.001	.001	.019	.019	.023	.020	.020	1.004	1.185	1.031	1.046
	$\phi_{11,1}$	1.800	.002	.001	.016	.001	.004	.023	.023	.035	.025	.023	1.003	1.498	1.080	1.008
	$\phi_{21,1}$	-.400	.016	.017	.012	.017	.016	.059	.060	.067	.062	.062	1.013	1.132	1.055	1.058
	$\phi_{22,1}$	.800	.028	.030	.027	.030	.027	.087	.088	.099	.092	.092	1.012	1.132	1.054	1.050
	$\phi_{11,2}$	-.360	.000	.000	.076	.001	.025	.073	.073	.136	.081	.075	1.007	1.858	1.112	1.027
	$\phi_{12,2}$	-.900	.002	.002	.119	.005	.040	.109	.109	.208	.121	.112	1.007	1.910	1.110	1.026
	$\theta_{11,1}$	.330	.025	.026	.032	.019	.051	.080	.081	.097	.089	.090	1.009	1.214	1.103	1.123
	$\theta_{21,1}$	-.180	.009	.010	.008	.003	.052	.069	.070	.099	.077	.082	1.011	1.421	1.107	1.190
	$\theta_{12,1}$	-.200	.005	.006	.053	.000	.047	.095	.096	.113	.104	.105	1.014	1.194	1.094	1.110
	$\theta_{22,1}$	-.400	.025	.029	.103	.021	.089	.107	.109	.162	.119	.136	1.022	1.515	1.119	1.277
	$\theta_{11,2}$	-.200	.026	.028	.000	.029	.037	.082	.084	.108	.094	.085	1.015	1.308	1.140	1.032
	$\theta_{12,2}$	.920	.004	.006	.192	.000	.105	.136	.138	.275	.155	.161	1.013	2.013	1.133	1.179
14	$\mu_{\Phi,1}$	.000	.003	.002	.003	.002	.003	.082	.082	.082	.081	.084	1.005	1.003	.999	1.028
	$\mu_{\Phi,2}$	.000	.003	.003	.003	.004	.004	.089	.089	.089	.089	.090	1.005	1.001	1.005	1.015
	$\phi_{21,0}$	.500	.001	.001	.001	.001	.000	.021	.021	.021	.021	.023	1.003	1.005	1.002	1.057
	$\phi_{11,1}$	1.800	.000	.001	.001	.001	.005	.024	.025	.025	.025	.025	1.017	1.027	1.027	1.038
	$\phi_{21,1}$	-.400	.016	.018	.017	.018	.014	.062	.062	.062	.062	.067	1.011	1.004	1.010	1.079
	$\phi_{22,1}$	.800	.028	.031	.029	.031	.029	.090	.091	.090	.091	.099	1.011	1.001	1.010	1.100
	$\phi_{11,2}$	-.360	.007	.006	.004	.002	.033	.075	.077	.077	.078	.083	1.019	1.027	1.033	1.098
	$\phi_{12,2}$	-.900	.015	.013	.010	.008	.056	.112	.114	.115	.116	.125	1.018	1.025	1.031	1.113
	$\theta_{11,1}$	.330	.022	.021	.021	.020	.028	.081	.082	.082	.082	.085	1.010	1.011	1.015	1.050
	$\theta_{21,1}$	-.180	.007	.008	.009	.008	.003	.072	.072	.074	.074	.076	.998	1.020	1.017	1.045
	$\theta_{12,1}$	-.200	.011	.011	.012	.010	.021	.096	.098	.097	.099	.102	1.021	1.018	1.033	1.066
	$\theta_{22,1}$	-.400	.033	.038	.037	.033	.045	.117	.121	.119	.120	.126	1.029	1.016	1.025	1.078
	$\theta_{11,2}$	-.200	.026	.027	.028	.031	.012	.083	.084	.085	.088	.080	1.014	1.021	1.056	.966
	$\theta_{12,2}$	.920	.020	.021	.020	.006	.063	.140	.145	.144	.147	.150	1.029	1.028	1.043	1.065

Note – TS1 and TS2 are the three-step GLS estimators based on the two-step GLS estimator and the two-step OLS estimator, respectively. While HK, RBY and PS stand for the fully efficient GLS estimators suggested by Hannan and Kavalieris (1984), Reinsel et al. (1992) and Poskitt and Salau (1995), respectively. These estimates are obtained with 1000 replications. The eigenvalues of the model are real .800 and a double root .900 for the autoregressive (AR) operator, and real -.530 and conjugate $-.350\mp.584i$ (.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.

Poskitt and Salau (1995)], but instead, they used $\tilde{\Sigma}_{u(n_T)}$ which we know is $O_p\big(1\big)\Sigma_u$. The bias associated with HK procedure may be attributed to using more or less well behaved filtered residuals in finite samples, and a weighting matrix mismatching the one iteration of the scoring algorithm (starting from the two-step OLS estimates). In fact, they use the third-stage error covariance estimator of their procedure instead of the one associated with a faster rate of convergence, namely the one based on the filtered residuals necessary to their fourth-stage estimation. Although RBY procedure uses the same filtering scheme as the HK method, it relatively delivers estimates with satisfactory finite sample properties. This is due perhaps to using an error covariance estimator with better small-sample properties as a weighting matrix in their GLS linear regression.

It is well known that approximating VARMA models having highly persistent MA operators usually requires autoregressions with many lags, and vice versa. Also, approximating nonpersistent VARMA models with autoregressions using many lags would result in estimates with higher bias and/or MSE. This exactly occurs with TS1 and TS2 procedures for the echelon VARMA model with Kronecker indices $(2, 1)$ since the dominant eigenvalue associated with the MA operator, namely .681 (in norm), is not considered as persistent; see Tables 3 and 4. The same tables show that, given the sample size, increasing the lag-order n_T reduces the large bias for HK and PS procedures, and yields parameter estimates with MSE decreasing for the HK procedure but with a mixed tendency for the PS method. Further, while noticing that RBY estimates are characterized with a slight increase in the bias, they exhibit MSE with a mixed tendency. Besides noting that the bias generally decreases with n_T with all methods for the echelon VARMA$(1, 2)$, we stress that the overall tendency for the MSE is not pronounced. This is due perhaps to the fact that the largest eigenvalue associated with the MA operator, namely .824, cannot characterize the model as less or highly persistent; see Tables 1 and 2. Simulation results show that, overall, TS1, TS2 and RBY methods outperform those of HK and PS by far. For a better idea on which procedure provides estimates with better sample properties – since we note that those of RBY procedure behave in a way quite similar to ours – we compute the ratio of the MSE of each procedure's estimates relative to those associated with TS1. Obviously, with the exception of TS2 and PS procedures, those of RBY and (to a large extent) HK provide estimates with MSE ratios, overall, greater than unity. Note that the cases where the MSE ratios of PS estimates are less than unity are somehow (indirectly) attributed to relatively substantial biases characterizing some of the echelon parameter estimates. These cases also match some situations where the reduction in the standard deviation of the estimates outweighs the increase in the square of the associated bias. Further, the frequency of these below-unity ratios is generally increasing with n_T and decreasing with the sample size. Finally, it is noteworthy that, while TS2 generally dominates RBY , TS1 has a slight advantage over TS2. So, choosing either TS1 or TS2 would have no significant effect on the small-sample behavior of the resulting echelon VARMA parameter estimates for the models studied.

6 Conclusion

This paper proposes a new three-step linear estimation procedure for stationary invertible echelon VARMA models. It can be extended to VARMAX or integrated and cointegrated VARMA models as well. The estimation method focuses on the echelon form parameterization as it tends to deliver relatively parsimonious models, but may easily adapt to other parameterizations such as the final equations form.

Our setup provides simplified and practical echelon parameter estimates that are easier to obtain than those of Hannan and Kavalieris (1984), Reinsel et al. (1992), and Poskitt and Salau (1995). We extend the results of Dufour and Jouini (2005) to the two-step GLS estimator and show its consistency and asymptotic normality with strong innovations. Exploiting the explicit form of the second-stage regression residuals, we propose a new recursive filtering scheme based on consistent (hence better) initial values for obtaining well-behaved pseudo-regressors necessary to our third-stage GLS (fully efficient) estimation. These filtered residuals which approximate the implicit VARMA innovation estimates matching the two-step linear estimator, are functions of the first-stage autoregression residuals and the second-stage regression residuals as well. So, they take into account the truncation error associated with the long-autoregression used in the first-stage, along with some adjustments involving the first two-step regression residuals. Besides this novelty, our third-stage linear regression is derived by exploiting the nonlinear structure of the VARMA innovations in the model parameters without using Taylor expansion. As such, the resulting three-step GLS estimator, for which we establish its consistency and asymptotic normality with strong innovations, then show its asymptotic equivalence to ML (hence efficiency) under Gaussian innovations, provides an intuitive interpretation of nonlinear estimation methods such as nonlinear GLS and ML. Although our three-step linear estimation procedure is asymptotically equivalent to those by Hannan and Kavalieris (1984), Reinsel et al. (1992) and Poskitt and Salau (1995), it is computationally much simpler relatively. In addition, the asymptotic covariance estimators we gave for the second and third-stage echelon VARMA parameter estimators as well, are simple and more practical for inference, especially in the context of simulation-based techniques such as bootstrap methods or MMC tests. Finally, by examining the complex dynamic structure of the third-stage regression residuals, we provide an efficient estimator of the covariance matrix of the VARMA innovations, which is of order T^{-1} more accurate than the one by the fourth-stage of Hannan and Kavalieris (1984).

The small-sample performance of our efficient linear estimators is studied compared to competing ones, namely those of Hannan and Kavalieris (1984), Reinsel et al. (1992), and Poskitt and Salau (1995). Simulation evidence shows that, in most cases, our fully efficient estimators outperform their competitors in terms of bias and MSE for the models studied. It also stresses the sensitivity of the small-sample properties of the echelon VARMA parameter estimates to the truncation-order of the first-stage autoregression approximating the true innovations. This suggests that further investigation should be made in this way for developing efficient model selection procedures to estimate accurately the autoregression truncation-lag in finite samples. Indeed, such a truncation may severely affect, through the echelon VARMA parameter estimates, the finite-sample behavior of the resulting high dynamics or smooth functions of the VARMA slope parameters and innovation variances, such as impulse responses, error variance decomposition, predictability measures or long-term forecasts, usually subject of interest in most applied work.

A Appendix: Proofs

PROOF OF COROLLARY 4.1 Let $\tilde{\Phi}(\bar{p}) = \left[-\tilde{\mu}_{\Phi}, \tilde{\Phi}_0, -\tilde{\Phi}_1, \ldots, -\tilde{\Phi}_{\bar{p}}\right]$, $\Phi(\bar{p}) = \left[-\mu_{\Phi}, \Phi_0, -\Phi_1, \ldots, -\Phi_{\bar{p}}\right]$ and finally $Y_t(\bar{p}) = \left[1, y_t', y_{t-1}', \ldots, y_{t-\bar{p}}'\right]'$. Then, in view of (4.2), (4.3), (3.9) and using the multivariate version of Lemma **2.2** of Kreiss and Franke (1992), we show, for $t = -n_T + 1, ..., T$, that

$$\left\|\tilde{u}_t - \varepsilon_t(\tilde{\eta})\right\| \leq \sum_{\tau=t+n_T}^{\infty} \left\{\left\|\Lambda_\tau(\eta)\right\| + \left\|\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta)\right\|\right\}\left\{\left\|\Phi(\bar{p})\right\| + \left\|\tilde{\Phi}(\bar{p}) - \Phi(\bar{p})\right\|\right\}\left\|Y_{t-\tau}(\bar{p})\right\| = O_p\left(\rho^{t+n_T}\right). \tag{A.1}$$

On the other hand, using (4.11), (4.12), (3.9) and Lemma **2.2** of Kreiss and Franke (1992), we show, for $t = 1, ..., T$, that

$$\left\|\tilde{u}_t - u_t(\tilde{\eta})\right\| \leq \sum_{\tau=t}^{\infty} \left\{\left\|\Lambda_\tau(\eta)\right\| + \left\|\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta)\right\|\right\}\left\{\left\|\tilde{e}_{t-\tau}(n_T) - u_{t-\tau}\right\| + \left\|u_{t-\tau} - \tilde{u}_{t-\tau}(n_T)\right\|\right\} = O_p\Big(\rho^t \frac{n_T}{T^{1/2}}\Big), \tag{A.2}$$

since $\left\|u_{t-\tau} - \tilde{u}_{t-\tau}(n_T)\right\|$ and $\left\|\tilde{e}_{t-\tau}(n_T) - u_{t-\tau}\right\|$ are both $O_p\left(n_T/T^{1/2}\right)$ in view of (3.4) and (3.12), respectively. ■

PROOF OF PROPOSITION 4.1 By the triangular inequality,

$$\left\|\tilde{\Sigma}_{u(\tilde{\eta})}-\Sigma_{u}\right\| \leq \frac{1}{T}\sum_{t=1}^{T}\Big\{\left\|u_{t}\left(\tilde{\eta}\right)-u_{t}\right\|\left\|u_{t}\left(\tilde{\eta}\right)\right\|+\left\|u_{t}\right\|\left\|u_{t}\left(\tilde{\eta}\right)-u_{t}\right\|\Big\}+O_{p}\big(T^{-1/2}\big), \tag{A.3}$$

where $\left\|u_{t}\left(\tilde{\eta}\right)-u_{t}\right\| \leq \left\|u_{t}\left(\tilde{\eta}\right)-u_{t}\left(\eta\right)\right\|+\left\|u_{t}\left(\eta\right)-u_{t}\right\|$ with $\left\|u_{t}\left(\eta\right)-u_{t}\right\|=O_{p}\big(\rho^{t}n_{T}/T^{1/2}\big)$. Using (4.15) and (4.12),

$$\left\|u_{t}\left(\tilde{\eta}\right)-u_{t}\left(\eta\right)\right\| \leq \sum_{\tau=0}^{t-1}\Big\{\left\|\Lambda_{\tau}\left(\tilde{\eta}\right)\right\|\left\|\tilde{e}_{t-\tau}\left(n_{T}\right)-e_{t-\tau}\left(n_{T}\right)\right\|+\left\|\Lambda_{\tau}\left(\tilde{\eta}\right)-\Lambda_{\tau}\left(\eta\right)\right\|\left\|e_{t-\tau}\left(n_{T}\right)-\tilde{u}_{t-\tau}\left(n_{T}\right)\right\|\Big\}. \tag{A.4}$$

On substituting $e_{t-\tau}\left(n_{T}\right)$ and $\tilde{e}_{t-\tau}\left(n_{T}\right)$ with their expressions in (3.6) and (3.11), and using (3.4), (3.9) and Lemma **2.2** of Kreiss and Franke (1992), we have:

$$\sum_{\tau=0}^{t-1}\left\|\Lambda_{\tau}\left(\tilde{\eta}\right)\right\|\left\|\tilde{e}_{t-\tau}\left(n_{T}\right)-e_{t-\tau}\left(n_{T}\right)\right\| \leq k^{1/2}\sum_{\tau=0}^{t-1}\left\|\Lambda_{\tau}\left(\tilde{\eta}\right)\right\|\left\|\tilde{X}_{t-\tau}\left(n_{T}\right)\right\|\left\|R\right\|\left\|\tilde{\eta}-\eta\right\|=O_{p}\big(T^{-1/2}\big), \tag{A.5}$$

$$\sum_{\tau=0}^{t-1}\left\|\Lambda_{\tau}\left(\tilde{\eta}\right)-\Lambda_{\tau}\left(\eta\right)\right\|\left\|e_{t-\tau}\left(n_{T}\right)-\tilde{u}_{t-\tau}\left(n_{T}\right)\right\| \leq \sum_{\tau=0}^{t-1}\sum_{j=0}^{\bar{p}}\left\|\Lambda_{\tau}\left(\tilde{\eta}\right)-\Lambda_{\tau}\left(\eta\right)\right\|\left\|\Theta_{j}\right\|\left\|u_{t-\tau-j}-\tilde{u}_{t-\tau-j}\left(n_{T}\right)\right\|=O_{p}\Big(\frac{n_{T}}{T}\Big). \tag{A.6}$$

Hence, $\left\|u_{t}\left(\tilde{\eta}\right)-u_{t}\left(\eta\right)\right\|=O_{p}\big(T^{-1/2}\big)$ and then $\left\|u_{t}\left(\tilde{\eta}\right)-u_{t}\right\|=O_{p}\big(T^{-1/2}\big)+O_{p}\big(\rho^{t}n_{T}/T^{1/2}\big)$, for $t=1,...,T$. Thus,

$$\Big\{\Big\|\tilde{\Sigma}_{u(\tilde{\eta})}-\Sigma_{u}\Big\|,\Big\|\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}-\Sigma_{u}^{-1}\Big\|\Big\}=O_{p}\big(T^{-1/2}\big). \tag{A.7}$$

■

PROOF OF PROPOSITION 4.2 Note that $Q_{X(\eta)}^{-1}$ is p.d. by definition and let $\bar{Q}_{X(\eta)}=\Big\{\frac{1}{T}\sum_{t=1}^{T}Z_{t}\Sigma_{u}^{-1}Z_{t}{}'\Big\}^{-1}$. Then

$$\Big\|\tilde{Q}_{X(\tilde{\eta})}^{\circ-1}-Q_{X(\eta)}^{-1}\Big\|_{1} \leq \Big\|\tilde{Q}_{X(\tilde{\eta})}^{\circ-1}-Q_{X(\eta)}^{-1}\Big\| \leq \Big\|\tilde{Q}_{X(\tilde{\eta})}^{\circ-1}-\bar{Q}_{X(\eta)}^{-1}\Big\|+\Big\|\bar{Q}_{X(\eta)}^{-1}-Q_{X(\eta)}^{-1}\Big\|, \tag{A.8}$$

where $\left\|\bar{Q}_{X(\eta)}^{-1}-Q_{X(\eta)}^{-1}\right\|=\Big\|\frac{1}{T}\sum_{t=1}^{T}Z_{t}\Sigma_{u}^{-1}Z_{t}{}'-\mathsf{E}\big[Z_{t}\Sigma_{u}^{-1}Z_{t}{}'\big]\Big\|=O_{p}\big(T^{-1/2}\big)$. Further,

$$\Big\|\tilde{Q}_{X(\tilde{\eta})}^{\circ-1}-\bar{Q}_{X(\eta)}^{-1}\Big\| \leq \left\|Q_{1}\right\|+\left\|Q_{2}\right\|+\left\|Q_{3}\right\|, \tag{A.9}$$

where, specifically, $Q_{1}=T^{-1}\sum_{t=1}^{T}Z_{t}\Sigma_{u}^{-1}\big[Z_{t}^{\circ}\left(\tilde{\eta},\eta\right)-Z_{t}\big]'$, $Q_{2}=T^{-1}\sum_{t=1}^{T}Z_{t}\big[\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}-\Sigma_{u}^{-1}\big]Z_{t}^{\circ}\left(\tilde{\eta},\eta\right)'$ and, finally, $Q_{3}=T^{-1}\sum_{t=1}^{T}\big[Z_{t}^{\circ}\left(\tilde{\eta},\eta\right)-Z_{t}\big]\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}Z_{t}^{\circ}\left(\tilde{\eta},\eta\right)'$. In particular, $\left\|Q_{1}\right\| \leq T^{-1}\sum_{t=1}^{T}\left\|Z_{t}\right\|\Big\|\Sigma_{u}^{-1}\Big\|\left\|Z_{t}^{\circ}\left(\tilde{\eta},\eta\right)-Z_{t}\right\|$, where, by invertibility of the VARMA process, $\mathsf{E}\left\|Z_{t}\right\|^{2}=O\left(1\right)$. Further,

$$\left\|Z_{t}^{\circ}\left(\tilde{\eta},\eta\right)-Z_{t}\right\| \leq \left\|R\right\|\bigg\{\bigg\|\sum_{\tau=0}^{t-1}\Big[\Big(X_{t-\tau}\left(\tilde{\eta}\right)-X_{t-\tau}\Big)\otimes\Lambda_{\tau}\left(\eta\right)'\Big]\bigg\|+\bigg\|\sum_{\tau=t}^{\infty}\Big[X_{t-\tau}\otimes\Lambda_{\tau}\left(\eta\right)'\Big]\bigg\|\bigg\}, \tag{A.10}$$

with

$$\mathsf{E}\bigg\|\sum_{\tau=t}^{\infty}\Big[X_{t-\tau}\otimes\Lambda_{\tau}\left(\eta\right)'\Big]\bigg\|^{2} \leq \sum_{\tau_{1}=t}^{\infty}\sum_{\tau_{2}=t}^{\infty}\operatorname{tr}\big[\Gamma_{X}\left(\tau_{1}-\tau_{2}\right)\big]\left\|\Lambda_{\tau_{1}}\left(\eta\right)\right\|\left\|\Lambda_{\tau_{2}}\left(\eta\right)\right\|=O\left(\rho^{2t}\right), \tag{A.11}$$

and

$$\bigg\|\sum_{\tau=0}^{t-1}\Big[\Big(X_{t-\tau}\left(\tilde{\eta}\right)-X_{t-\tau}\Big)\otimes\Lambda_{\tau}\left(\eta\right)'\Big]\bigg\| \leq \sum_{\tau=0}^{t-1}\left\|X_{t-\tau}\left(\tilde{\eta}\right)-X_{t-\tau}\right\|\left\|\Lambda_{\tau}\left(\eta\right)\right\|, \tag{A.12}$$

with $\left\|X_{t-\tau}\left(\tilde{\eta}\right)-X_{t-\tau}\right\|^{2}=\sum_{j=0}^{\bar{p}}\left\|u_{t-j-\tau}\left(\tilde{\eta}\right)-u_{t-j-\tau}\right\|^{2}$. Hence $\left\|Z_{t}^{\circ}\left(\tilde{\eta},\eta\right)-Z_{t}\right\|=O_{p}\big(T^{-1/2}\big)+O_{p}\left(\rho^{t}\right)$ and then $\left\|Q_{1}\right\|=O_{p}\big(T^{-1/2}\big)$. Using Proposition **4.1**, we also show that $\left\|Q_{2}\right\|$ and $\left\|Q_{3}\right\|$ are both $O_{p}\big(T^{-1/2}\big)$, since $\left\|Z_{t}^{\circ}\left(\tilde{\eta},\eta\right)\right\|$ is

$O_p(1)$. Hence $\left\|\tilde{Q}^{\circ-1}_{X(\tilde{\eta})} - \bar{Q}^{-1}_{X(\eta)}\right\|$, $\left\|\tilde{Q}^{\circ-1}_{X(\tilde{\eta})} - Q^{-1}_{X(\eta)}\right\|$ and $\left\|\tilde{Q}^{\circ-1}_{X(\tilde{\eta})} - Q^{-1}_{X(\eta)}\right\|_1$ are all $O_p\big(T^{-1/2}\big)$. Finally,

$$\left\|\tilde{Q}^{\circ}_{X(\tilde{\eta})} - Q_{X(\eta)}\right\|_1 = O_p\big(T^{-1/2}\big). \tag{A.13}$$

On the other hand,

$$\left\|\tilde{Q}^{-1}_{X(\tilde{\eta})} - \tilde{Q}^{\circ-1}_{X(\tilde{\eta})}\right\|_1 \le \left\|\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\right\| \frac{1}{T}\sum_{t=1}^{T}\Big\{\|Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\|\|Z_t(\tilde{\eta})\| + \|Z_t^{\circ}(\tilde{\eta},\eta)\|\|Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\|\Big\}, \tag{A.14}$$

where, by Proposition **4.1**, $\left\|\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\right\| = O_p(1)$. Further, by Lemma **2.2** of Kreiss and Franke (1992),

$$\|Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\| \le \|R\| \sum_{\tau=0}^{t-1}\Big\{\|X_{t-\tau}(\tilde{\eta}) - X_{t-\tau}\| + \|X_{t-\tau}\|\Big\}\|\Lambda_\tau(\tilde{\eta}) - \Lambda_\tau(\eta)\| = O_p\big(T^{-1/2}\big). \tag{A.15}$$

Then, $\|Z_t(\tilde{\eta})\| \le \|Z_t(\tilde{\eta}) - Z_t^{\circ}(\tilde{\eta},\eta)\| + \|Z_t^{\circ}(\tilde{\eta},\eta)\| = O_p(1)$. Hence $\left\|\tilde{Q}^{-1}_{X(\tilde{\eta})} - \tilde{Q}^{\circ-1}_{X(\tilde{\eta})}\right\|$ and $\left\|\tilde{Q}^{-1}_{X(\tilde{\eta})} - \tilde{Q}^{\circ-1}_{X(\tilde{\eta})}\right\|_1$ are both $O_p\big(T^{-1/2}\big)$. Therefore,

$$\left\|\tilde{Q}_{X(\tilde{\eta})} - \tilde{Q}^{\circ}_{X(\tilde{\eta})}\right\|_1 = O_p\big(T^{-1/2}\big). \tag{A.16}$$

■

Proof of Theorem 4.1 Note that (4.21) can be rewritten as $\hat{\eta} - \eta = \tilde{Q}_{X(\tilde{\eta})}\tilde{\Omega}_{X(\tilde{\eta})} + \tilde{Q}^{\circ}_{X(\tilde{\eta})}\big[\tilde{\Omega}^{\bullet}_{X(\tilde{\eta})} - \tilde{\Omega}^{\circ}_{X(\tilde{\eta})}\big]$, where $\tilde{\Omega}^{\circ}_{X(\tilde{\eta})} = T^{-1}\sum_{t=1}^{T} Z_t^{\circ}(\tilde{\eta},\eta)\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}u_t(\tilde{\eta})$ and $\tilde{\Omega}^{\bullet}_{X(\tilde{\eta})} = T^{-1}\sum_{t=1}^{T} Z_t^{\circ}(\tilde{\eta},\eta)\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}u_t(\eta)$. In addition, let $\Omega_{X(\eta)} = T^{-1}\sum_{t=1}^{T} Z_t\Sigma_u^{-1}u_t$. Then, by the triangular inequality,

$$\begin{aligned}\|\hat{\eta} - \eta\| \le\;& \|Q_{X(\eta)}\|_1\|\Omega_{X(\eta)}\| + \left\|\tilde{Q}^{\circ}_{X(\tilde{\eta})} - Q_{X(\eta)}\right\|_1\left\|\tilde{\Omega}^{\bullet}_{X(\tilde{\eta})}\right\| + \|Q_{X(\eta)}\|_1\left\|\tilde{\Omega}^{\bullet}_{X(\tilde{\eta})} - \Omega_{X(\eta)}\right\| \\ &+ \left\|\tilde{Q}_{X(\tilde{\eta})} - \tilde{Q}^{\circ}_{X(\tilde{\eta})}\right\|_1\left\|\tilde{\Omega}_{X(\tilde{\eta})}\right\| + \left\|\tilde{Q}^{\circ}_{X(\tilde{\eta})}\right\|_1\left\|\tilde{\Omega}_{X(\tilde{\eta})} - \tilde{\Omega}^{\circ}_{X(\tilde{\eta})}\right\|,\end{aligned} \tag{A.17}$$

where $\|Q_{X(\eta)}\|_1 = O_p(1)$, $\|\Omega_{X(\eta)}\| = O_p\big(T^{-1/2}\big)$, while $\left\|\tilde{Q}^{\circ}_{X(\tilde{\eta})} - Q_{X(\eta)}\right\|_1$ and $\left\|\tilde{Q}_{X(\tilde{\eta})} - \tilde{Q}^{\circ}_{X(\tilde{\eta})}\right\|_1$ are both $O_p\big(T^{-1/2}\big)$. In addition, set $S_1 = T^{-1}\sum_{t=1}^{T} Z_t\Sigma_u^{-1}\big[u_t(\eta) - u_t\big]$, $S_2 = T^{-1}\sum_{t=1}^{T} Z_t\Big[\tilde{\Sigma}^{-1}_{u(\tilde{\eta})} - \Sigma_u^{-1}\Big]u_t(\eta)$ and $S_3 = T^{-1}\sum_{t=1}^{T}\big[Z_t^{\circ}(\tilde{\eta},\eta) - Z_t\big]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}u_t(\eta)$. Then

$$\left\|\tilde{\Omega}^{\bullet}_{X(\tilde{\eta})} - \Omega_{X(\eta)}\right\| \le \|S_1\| + \|S_2\| + \|S_3\|. \tag{A.18}$$

Using the fact that $\|u_t - u_t(\eta)\| = O_p\big(\rho^t n_T/T^{1/2}\big)$ and that $\|\mathrm{vec}[B]\| = \|B\|$, we show that

$$\mathsf{E}\|S_1\| \le \|R\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\|\Sigma_u^{-1}\|\Big\{\mathsf{E}\|u_t(\eta) - u_t\|^2\Big\}^{1/2}\Big\{\mathsf{E}\|X_{t-\tau}\|^2\Big\}^{1/2} = O\Big(\frac{n_T}{T^{3/2}}\Big). \tag{A.19}$$

Moreover,

$$\|S_2\| \le \left\|\frac{1}{T}\sum_{t=1}^{T} Z_t\Big[\tilde{\Sigma}^{-1}_{u(\tilde{\eta})} - \Sigma_u^{-1}\Big]\big[u_t(\eta) - u_t\big]\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T} Z_t\Big[\tilde{\Sigma}^{-1}_{u(\tilde{\eta})} - \Sigma_u^{-1}\Big]u_t\right\|, \tag{A.20}$$

where, as in (A.19), $\left\|\frac{1}{T}\sum_{t=1}^{T} Z_t\Big[\tilde{\Sigma}^{-1}_{u(\tilde{\eta})} - \Sigma_u^{-1}\Big]\big[u_t(\eta) - u_t\big]\right\| = O_p\big(n_T/T^2\big)$. Further, using Proposition **4.1**,

$$\left\|\frac{1}{T}\sum_{t=1}^{T} Z_t\Big[\tilde{\Sigma}^{-1}_{u(\tilde{\eta})} - \Sigma_u^{-1}\Big]u_t\right\| \le \left\|\tilde{\Sigma}^{-1}_{u(\tilde{\eta})} - \Sigma_u^{-1}\right\|\left\{\sum_{\tau=0}^{\infty}\|\Lambda_\tau(\eta)\|\left\|\frac{1}{T}\sum_{t=1}^{T} u_t X'_{t-\tau}\right\|\right\} = O_p\big(T^{-1}\big), \tag{A.21}$$

since $\left\|\frac{1}{T}\sum_{t=1}^{T} u_t X_{t-\tau}'\right\| = O_p\left(T^{-1/2}\right)$, by the VARMA structure of y_t. Hence, $\left\|S_2\right\| = O_p\left(T^{-1}\right)$. Finally,

$$\left\|S_3\right\| \leq \left\|\Omega^1_{Z(\tilde{\eta})}\right\| + \left\|\Omega^2_{Z(\tilde{\eta})}\right\|, \tag{A.22}$$

where $\Omega^1_{Z(\tilde{\eta})} = T^{-1}\sum_{t=1}^{T}\left[Z_t^{\circ}\left(\tilde{\eta},\eta\right) - Z_t\right]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\left[u_t\left(\eta\right) - u_t\right]$ and $\Omega^2_{Z(\tilde{\eta})} = T^{-1}\sum_{t=1}^{T}\left[Z_t^{\circ}\left(\tilde{\eta},\eta\right) - Z_t\right]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}u_t$. Also,

$$\begin{aligned}
\left\|\Omega^1_{Z(\tilde{\eta})}\right\| &\leq \left\|\Omega^{11}_{Z(\tilde{\eta})}\right\| + \left\|\Omega^{12}_{Z(\tilde{\eta})}\right\| + \left\|\Omega^{13}_{Z(\tilde{\eta})}\right\|, \qquad \text{with} && \text{(A.23)}\\
\Omega^{11}_{Z(\tilde{\eta})} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty} R'\left[X_{t-\tau}\otimes\Lambda_\tau\left(\eta\right)'\right]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\left[u_t\left(\eta\right) - u_t\right], && \text{(A.24)}\\
\Omega^{12}_{Z(\tilde{\eta})} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1} R'\left[\left\{X_{t-\tau}\left(\eta\right) - X_{t-\tau}\right\}\otimes\Lambda_\tau\left(\eta\right)'\right]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\left[u_t\left(\eta\right) - u_t\right], && \text{(A.25)}\\
\Omega^{13}_{Z(\tilde{\eta})} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1} R'\left[\left\{X_{t-\tau}\left(\tilde{\eta}\right) - X_{t-\tau}\left(\eta\right)\right\}\otimes\Lambda_\tau\left(\eta\right)'\right]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\left[u_t\left(\eta\right) - u_t\right], && \text{(A.26)}
\end{aligned}$$

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

$$\left\|\Omega^{11}_{Z(\tilde{\eta})}\right\| \leq \left\|R\right\|\left\|\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\left\|\Lambda_\tau\left(\eta\right)\right\|\left\|\left[u_t\left(\eta\right) - u_t\right]X_{t-\tau}'\right\|\right\} = O_p\left(\frac{n_T}{T^{3/2}}\right), \tag{A.27}$$

since $\mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\left\|\Lambda_\tau\left(\eta\right)\right\|\left\|\left[u_t\left(\eta\right) - u_t\right]X_{t-\tau}'\right\|\right) = O\left(n_T/T^{3/2}\right)$. Further,

$$\left\|\Omega^{12}_{Z(\tilde{\eta})}\right\| \leq \left\|R\right\|\left\|\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|u_t\left(\eta\right) - u_t\right\|\left(\sum_{\tau=0}^{t-1}\left\|\Lambda_\tau\left(\eta\right)\right\|^2\right)^{1/2}\left(\sum_{\tau=0}^{t-1}\left\|X_{t-\tau}\left(\eta\right) - X_{t-\tau}\right\|^2\right)^{1/2}\right\}, \tag{A.28}$$

where $\left\|X_{t-\tau}\left(\eta\right) - X_{t-\tau}\right\|^2 = \sum_{j=0}^{\bar{p}}\left\|u_{t-j-\tau}\left(\eta\right) - u_{t-j-\tau}\right\|^2$, with $\mathsf{E}\left\|u_{t-j-\tau}\left(\eta\right) - u_{t-j-\tau}\right\| = O\left(\rho^{t-j-\tau}n_T/T^{1/2}\right)$. Hence, $\left\|X_{t-\tau}\left(\eta\right) - X_{t-\tau}\right\|^2 = O_p\left(\rho^{2(t-\tau)}n_T^2/T\right)$ and then $\sum_{\tau=0}^{t-1}\left\|X_{t-\tau}\left(\eta\right) - X_{t-\tau}\right\|^2 = O_p\left(n_T^2/T\right)$. Therefore, $\left\|\Omega^{12}_{Z(\tilde{\eta})}\right\|$ is $O_p\left(n_T^2/T^2\right)$. On the other hand,

$$\left\|\Omega^{13}_{Z(\tilde{\eta})}\right\| \leq \left\|R\right\|\left\|\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\tau=0}^{t-1}\left\|\Lambda_\tau\left(\eta\right)\right\|\left\|u_t\left(\eta\right) - u_t\right\|\left\|X_{t-\tau}\left(\tilde{\eta}\right) - X_{t-\tau}\left(\eta\right)\right\|\right)\right\}, \tag{A.29}$$

with $\left\|X_{t-\tau}\left(\tilde{\eta}\right) - X_{t-\tau}\left(\eta\right)\right\|^2 = \sum_{j=0}^{\bar{p}}\left\|u_{t-j-\tau}\left(\tilde{\eta}\right) - u_{t-j-\tau}\left(\eta\right)\right\|^2$. Therefore, we get $\left\|\Omega^{13}_{Z(\tilde{\eta})}\right\| = O_p\left(n_T/T^2\right)$, since $\left\|u_t\left(\eta\right) - u_t\right\| = O_p\left(\rho^t n_T/T^{1/2}\right)$ and $\left\|u_t\left(\tilde{\eta}\right) - u_t\left(\eta\right)\right\| = O_p\left(T^{-1/2}\right)$. Hence, $\left\|\Omega^1_{Z(\tilde{\eta})}\right\| = O_p\left(n_T/T^{3/2}\right)$. Further,

$$\begin{aligned}
\left\|\Omega^2_{Z(\tilde{\eta})}\right\| &\leq \left\|\Omega^{21}_{Z(\tilde{\eta})}\right\| + \left\|\Omega^{22}_{Z(\tilde{\eta})}\right\| + \left\|\Omega^{23}_{Z(\tilde{\eta})}\right\|, \qquad \text{with} && \text{(A.30)}\\
\Omega^{21}_{Z(\tilde{\eta})} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty} R'\left[X_{t-\tau}\otimes\Lambda_\tau\left(\eta\right)'\right]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}u_t, && \text{(A.31)}\\
\Omega^{22}_{Z(\tilde{\eta})} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1} R'\left[\left\{X_{t-\tau}\left(\eta\right) - X_{t-\tau}\right\}\otimes\Lambda_\tau\left(\eta\right)'\right]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}u_t, && \text{(A.32)}\\
\Omega^{23}_{Z(\tilde{\eta})} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1} R'\left[\left\{X_{t-\tau}\left(\tilde{\eta}\right) - X_{t-\tau}\left(\eta\right)\right\}\otimes\Lambda_\tau\left(\eta\right)'\right]\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}u_t, && \text{(A.33)}
\end{aligned}$$

Similarly,

$$\left\|\Omega^{21}_{Z(\tilde{\eta})}\right\| \leq \left\|R\right\|\left\|\tilde{\Sigma}^{-1}_{u(\tilde{\eta})}\right\|\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\left\|\Lambda_\tau\left(\eta\right)\right\|\left\|u_t X_{t-\tau}'\right\|\right\} = O_p\left(T^{-1}\right), \tag{A.34}$$

since $\mathsf{E}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{\infty}\left\|\Lambda_{\tau}(\eta)\right\|\left\|u_t X_{t-\tau}'\right\|\right)=O\left(T^{-1}\right)$. Hence $\left\|\Omega_{Z(\tilde{\eta})}^{21}\right\|=O_p\left(T^{-1}\right)$. Further,

$$\left\|\Omega_{Z(\tilde{\eta})}^{22}\right\| \leq \|R\|\left\|\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}\right\|\left\{\frac{1}{T}\sum_{\tau=0}^{T-1}\sum_{t=1}^{T-\tau}\left\|\Lambda_{\tau}(\eta)\right\|\left\|u_{t+\tau}\left[X_t(\eta)-X_t\right]'\right\|\right\}=\|R\|\left\|\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}\right\|\Delta=O_p\left(\frac{n_T}{T^{3/2}}\right), \tag{A.35}$$

since $\mathsf{E}\left[\Delta\right] \leq \frac{1}{T}\sum_{\tau=0}^{T-1}\sum_{t=1}^{T-\tau}\left\|\Lambda_{\tau}(\eta)\right\|\left\{\mathsf{E}\left\|u_{t+\tau}\right\|^2\right\}^{1/2}\left\{\mathsf{E}\left\|X_t(\eta)-X_t\right\|^2\right\}^{1/2}=O\left(n_T/T^{3/2}\right)$. Further,

$$\left\|\Omega_{Z(\tilde{\eta})}^{23}\right\| \leq \|R\|\left\|\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}\right\|\left\{\sum_{\tau=0}^{T-1}\left\|\Lambda_{\tau}(\eta)\right\|\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right]'\right\|\right\}, \tag{A.36}$$

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

$$\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[u_{t-\tau}(\tilde{\eta})-u_{t-\tau}(\eta)\right]'\right\| \leq \Delta_1+\Delta_2, \tag{A.37}$$

where

$$\begin{aligned}
\Delta_1 &= \left\|\frac{1}{T}\sum_{t=\tau+1}^{T}\sum_{\nu=0}^{t-\tau-1}u_t\left[e_{t-\tau-\nu}(n_T)-\tilde{u}_{t-\tau-\nu}(n_T)\right]'\left[\Lambda_{\nu}(\tilde{\eta})-\Lambda_{\nu}(\eta)\right]'\right\| \leq \sum_{\nu=0}^{T-\tau-1}\Delta_{11}\left\|\Lambda_{\nu}(\tilde{\eta})-\Lambda_{\nu}(\eta)\right\|,\\
\Delta_2 &= \left\|\frac{1}{T}\sum_{t=\tau+1}^{T}\sum_{\nu=0}^{t-\tau-1}u_t\left[\tilde{e}_{t-\tau-\nu}(n_T)-e_{t-\tau-\nu}(n_T)\right]'\Lambda_{\nu}(\tilde{\eta})'\right\| \leq \sum_{\nu=0}^{T-\tau-1}\Delta_{22}\left\|\Lambda_{\nu}(\tilde{\eta})\right\|, && \text{(A.38)}\\
\Delta_{11} &= \left\|\frac{1}{T}\sum_{t=\tau+\nu+1}^{T}u_t\left[e_{t-\tau-\nu}(n_T)-\tilde{u}_{t-\tau-\nu}(n_T)\right]'\right\|=O_p\left(\frac{n_T}{T}\right), && \text{(A.39)}\\
\Delta_{22} &= \left\|\frac{1}{T}\sum_{t=\tau+\nu+1}^{T}u_t\left[\eta-\tilde{\eta}\right]'R'\left[\tilde{X}_{t-\tau-\nu}(n_T)\otimes I_k\right]\right\|=O_p\left(T^{-1}\right). && \text{(A.40)}
\end{aligned}$$

Therefore, using Lemma **2.2** of Kreiss and Franke (1992), $\Delta_1=O_p\left(n_T/T^{3/2}\right)$ and $\Delta_2=O_p\left(T^{-1}\right)$. It follows that, $\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[u_{t-\tau}(\tilde{\eta})-u_{t-\tau}(\eta)\right]'\right\|$, $\left\|\frac{1}{T}\sum_{t=\tau+1}^{T}u_t\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}(\eta)\right]'\right\|$ and $\left\|\Omega_{Z(\tilde{\eta})}^{23}\right\|$ are all $O_p\left(T^{-1}\right)$. Hence $\left\|\Omega_{Z(\tilde{\eta})}^{2}\right\|$ and $\left\|S_3\right\|$ are both $O_p\left(T^{-1}\right)$. Thus, $\left\|\tilde{\Omega}_{X(\tilde{\eta})}^{\bullet}-\Omega_{X(\eta)}\right\|=O_p\left(T^{-1}\right)$. Likewise,

$$\begin{aligned}
\left\|\tilde{\Omega}_{X(\tilde{\eta})}-\tilde{\Omega}_{X(\tilde{\eta})}^{\circ}\right\| &\leq \|R\|\left\{\left\|\Omega_{X(\tilde{\eta})}^{1}\right\|+\left\|\Omega_{X(\tilde{\eta})}^{2}\right\|+\left\|\Omega_{X(\tilde{\eta})}^{3}\right\|+\left\|\Omega_{X(\tilde{\eta})}^{4}\right\|\right\}, \qquad \text{with} && \text{(A.41)}\\
\Omega_{X(\tilde{\eta})}^{1} &= R'\mathrm{vec}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left[\Lambda_{\tau}(\tilde{\eta})-\Lambda_{\tau}(\eta)\right]'\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}\left[u_t(\tilde{\eta})-u_t\right]\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\right]'\right], && \text{(A.42)}\\
\Omega_{X(\tilde{\eta})}^{2} &= R'\mathrm{vec}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left[\Lambda_{\tau}(\tilde{\eta})-\Lambda_{\tau}(\eta)\right]'\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}\left[u_t(\tilde{\eta})-u_t\right]X_{t-\tau}'\right], && \text{(A.43)}\\
\Omega_{X(\tilde{\eta})}^{3} &= R'\mathrm{vec}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left[\Lambda_{\tau}(\tilde{\eta})-\Lambda_{\tau}(\eta)\right]'\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}u_t\left[X_{t-\tau}(\tilde{\eta})-X_{t-\tau}\right]'\right], && \text{(A.44)}\\
\Omega_{X(\tilde{\eta})}^{4} &= R'\mathrm{vec}\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=0}^{t-1}\left[\Lambda_{\tau}(\tilde{\eta})-\Lambda_{\tau}(\eta)\right]'\tilde{\Sigma}_{u(\tilde{\eta})}^{-1}u_t X_{t-\tau}'\right]. && \text{(A.45)}
\end{aligned}$$

Using the same arguments as before, we see that $\left\|\Omega_{X(\tilde{\eta})}^{1}\right\|=O_p\left(T^{-3/2}\right)$, while $\left\|\Omega_{X(\tilde{\eta})}^{2}\right\|$, $\left\|\Omega_{X(\tilde{\eta})}^{3}\right\|$ and $\left\|\Omega_{X(\tilde{\eta})}^{4}\right\|$ are all $O_p\left(T^{-1}\right)$. Thus, $\left\|\tilde{\Omega}_{X(\tilde{\eta})}-\tilde{\Omega}_{X(\tilde{\eta})}^{\circ}\right\|=O_p\left(T^{-1}\right)$. Similarly, we show that $\left\|\tilde{\Omega}_{X(\tilde{\eta})}^{\circ}-\Omega_{X(\eta)}\right\|=O_p\left(T^{-1}\right)$. Further, using the fact that

$$\left\|\tilde{\Omega}_{X(\tilde{\eta})}\right\| \leq \left\|\tilde{\Omega}^{\circ}_{X(\tilde{\eta})}\right\| + \left\|\tilde{\Omega}_{X(\tilde{\eta})} - \tilde{\Omega}^{\circ}_{X(\tilde{\eta})}\right\| \leq \left\|\Omega_{X(\eta)}\right\| + \left\|\tilde{\Omega}^{\circ}_{X(\tilde{\eta})} - \Omega_{X(\eta)}\right\| + \left\|\tilde{\Omega}_{X(\tilde{\eta})} - \tilde{\Omega}^{\circ}_{X(\tilde{\eta})}\right\|, \tag{A.46}$$

we conclude that $\left\|\tilde{\Omega}_{X(\tilde{\eta})}\right\| = O_p\big(T^{-1/2}\big)$. Finally, $\left\|\hat{\eta} - \eta\right\| = O_p\big(T^{-1/2}\big)$. ■

PROOF OF THEOREM 4.2 Let the random vectors $\tilde{S}_{X(\tilde{\eta})} = T^{1/2}\Big\{\tilde{Q}_{X(\tilde{\eta})}\tilde{\Omega}_{X(\tilde{\eta})} + \tilde{Q}^{\circ}_{X(\tilde{\eta})}\big[\tilde{\Omega}^{\bullet}_{X(\tilde{\eta})} - \tilde{\Omega}^{\circ}_{X(\tilde{\eta})}\big]\Big\}$ and $S_{X(\eta)} = T^{1/2}Q_{X(\eta)}\Omega_{X(\eta)}$. Then,

$$\begin{aligned}\left\|\tilde{S}_{X(\tilde{\eta})} - S_{X(\eta)}\right\| \quad \leq \quad & T^{1/2}\Big\{\left\|\tilde{Q}^{\circ}_{X(\tilde{\eta})} - Q_{X(\eta)}\right\|_1\left\|\tilde{\Omega}^{\bullet}_{X(\tilde{\eta})}\right\| + \left\|Q_{X(\eta)}\right\|_1\left\|\tilde{\Omega}^{\bullet}_{X(\tilde{\eta})} - \Omega_{X(\eta)}\right\| \\ & + \left\|\tilde{Q}_{X(\tilde{\eta})} - \tilde{Q}^{\circ}_{X(\tilde{\eta})}\right\|_1\left\|\tilde{\Omega}_{X(\tilde{\eta})}\right\| + \left\|\tilde{Q}^{\circ}_{X(\tilde{\eta})}\right\|_1\left\|\tilde{\Omega}_{X(\tilde{\eta})} - \tilde{\Omega}^{\circ}_{X(\tilde{\eta})}\right\|\Big\}.\end{aligned} \tag{A.47}$$

Using Proposition **4.2** and Theorem **4.1**, we show that $\left\|\tilde{S}_{X(\tilde{\eta})} - S_{X(\eta)}\right\| = O_p\big(T^{-1/2}\big)$. Therefore, 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 the assumption of independence between u_t and Z_t, we have $T^{1/2}\Omega_{X(\eta)} \xrightarrow[T\to\infty]{d} N\Big[0,\, Q^{-1}_{X(\eta)}\Big]$. Hence,

$$T^{1/2}\big(\hat{\eta} - \eta\big) = \tilde{S}_{X(\tilde{\eta})} \xrightarrow[T\to\infty]{d} N\Big[0,\, Q_{X(\eta)}\Big]. \tag{A.48}$$

■

PROOF OF PROPOSITION 4.3 Note that (4.25) reduces to

$$u_t\left(\hat{\eta}\right) = u_t\left(\tilde{\eta}\right) + Z^{\circ}_t\left(\tilde{\eta}, \hat{\eta}\right)'\left(\tilde{\eta} - \hat{\eta}\right), \tag{A.49}$$

since $\epsilon_t\left(\tilde{\eta}, \hat{\eta}\right) = \omega_t\left(\tilde{\eta}\right) - Z_t\left(\tilde{\eta}\right)'\hat{\eta}$ and $\omega_t\left(\tilde{\eta}\right) = u_t\left(\tilde{\eta}\right) + Z_t\left(\tilde{\eta}\right)'\tilde{\eta}$. Therefore, using Lemma **2.2** of Kreiss and Franke (1992), Proposition **4.1**, (3.9) and Theorem **4.1**, we get:

$$\left\|u_t\left(\hat{\eta}\right) - u_t\right\| = \left\|u_t\left(\tilde{\eta}\right) - u_t\right\| + \left\|Z^{\circ}_t\left(\tilde{\eta}, \hat{\eta}\right)\right\|\Big\{\left\|\tilde{\eta} - \eta\right\| + \left\|\hat{\eta} - \eta\right\|\Big\} = O_p\big(T^{-1/2}\big) + O_p\Big(\rho^t\frac{n_T}{T^{1/2}}\Big), \tag{A.50}$$

for $t = \bar{p} + 1, ..., T$. Then as in Proposition **4.1**, we show that $\left\|\tilde{\Sigma}_{u(\hat{\eta})} - \Sigma_u\right\| = O_p\big(T^{-1/2}\big)$. ■

B Acknowledgements

The authors thank Vadim Marmer, Peter Phillips, two anonymous referees, an anonymous Associate Editor and the Editor Stanley Azen 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).

References

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

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

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

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. (2006), ‘Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics’, *Journal of Econometrics* **133**(2), 443–477.

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 Jouini, T. (2006), ‘Finite-sample simulation-based tests in VAR models with applications to Granger causality testing’, *Journal of Econometrics* **135**(1-2), 229–254.

Dufour, J.-M. and Jouini, T. (2010), Asymptotic distributions for some quasi-efficient estimators in echelon-form VARMA models, Discussion paper, Department of Economics, McGill University.

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

Hannan, E. J. and Kavalieris, L. (1984), ‘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.

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.

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

Kreiss, J.-P. and Franke, J. (1992), ‘Bootstrapping stationary autoregressive moving-average models’, *Journal of Time Series Analysis* **13**(4), 297–317.

Lütkepohl, H. (1987), *Forecasting Aggregated Vector ARMA Processes*, 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.

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.

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.

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.

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.

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.