Online appendix for
“Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application”

Jean-Marie Dufour [*] | Denis Pelletier [†]
McGill University | North Carolina State University

January 2021

Contents

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

[†] Department of Economics, Box 8110, North Carolina State University, Raleigh, NC 27695-8110, USA. Email: denis_pelletier@ncsu.edu. Web page: http://www4.ncsu.edu/ dpellet

A. Expanded introduction

In this section we expand on some the points made in the Introduction:

- The family of VAR models is not closed under marginalization and temporal aggregation. This result is discussed in Lütkepohl (1991). If a vector satisfies a VAR model, subvectors do not typically satisfy VAR models (but VARMA models). Similarly, if the variables of a VAR process are observed at a different frequency, the resulting process is not a VAR process. In contrast, the class of VARMA models is closed under such operations.

- VARMA models can provide more accurate forecasts. There is no compelling reason for restricting macroeconomic forecasting to VAR models: VARMA models can forecast macroeconomic variables more accurately than VARs; for further discussion, see Athanasopoulos and Vahid (2008). Indeed, VARMA models can generate forecasts superior than those obtained from Bayesian VARs and factor models; see Dias and Kapetanios (2018). Models in macroeconomics often contain an MA component; for several examples, see Chen, Choi, and Escanciano (2017).

- The asymptotic efficiency of the three-step estimator presented in Hannan and Rissanen (1982) is proved by Zhao-Guo (1985). An extension of this innovation-substitution method to VARMA models was also proposed by Hannan and Kavalieris (1984a) and Koreisha and Pukkila (1989), under the assumption that the innovations constitute an m.d.s.

  In this paper we extend these results by showing that linear regression-based estimators are consistent under weaker hypotheses on the innovations and how filtering in a third step yields estimators with the same asymptotic distribution as their nonlinear counterparts (maximum likelihood when innovations are i.i.d. Gaussian, or nonlinear least squares if they are merely uncorrelated). In the non i.i.d. case, we consider strong mixing conditions [Doukhan (1995), Bosq (1998)], rather than the usual m.d.s. assumption. By using weaker assumptions on the model innovations, we broaden the class of processes to which our method can be applied. Recent work considering time series models with uncorrelated but dependent innovations include Boubacar Maïnassara and Saussereau (2018), Zhu and Li (2015), Boubacar Maïnassara and Raïssi (2015), Chen, Choi, and Escanciano (2017), Boubacar Mainassara, Carbon, and Francq (2012).

- The importance of nonlinear models has been growing in the time-series literature. Important classes of nonlinear processes admit an ARMA representation [see Francq and Zakoïan (1998), Francq, Roy, and Zakoïan (2005)]. However, the innovations in these ARMA representations do not have the usual i.i.d. or m.d.s. (martingale difference sequence) property, though they are uncorrelated. One must then be careful before applying usual results to the estimation of ARMA models because they usually rely on the above strong assumptions [*e.g.*, see Brockwell and Davis (1991) and Lütkepohl (1991)]. We refer to these as strong and semi-strong ARMA models respectively, by opposition to weak ARMA models where the innovations are only uncorrelated. The i.i.d. and m.d.s. properties are also not robust to aggregation (the i.i.d. Gaussian case being an exception); see Francq and Zakoïan (1998), Francq, Roy, and Zakoïan (2005), Boubacar Mainassara, Carbon, and Francq (2012), Palm and Nijman (1984), Nijman and Palm (1990), Drost (1993). In fact, the Wold decomposition only guarantees that the innovations are uncorrelated.

B. Existing estimation methods for VARMA models

For the estimation of VARMA models the emphasis has been on maximizing the likelihood (minimizing by nonlinear least squares) quickly. There are two ways of doing this. The first is having quick and efficient algorithm to evaluate the likelihood [*e.g.* Luceño (1994) and the reference therein, Mauricio (2002), Shea (1989), Mélard, Roy, and Saidi (2006)]. The second is to find preliminary consistent estimates that can be computed quickly to initialize the optimization algorithm. We are not the first to present a generalization to VARMA models of the Hannan and Rissanen (1982) estimation procedure for ARMA models [whose asymptotic properties are further studied in Zhao-Guo (1985) and Saikkonen (1986)]; see also Durbin (1960), Hannan and Kavalieris (1984a), Hannan, Kavalieris, and Mackisack (1986), Poskitt (1987), Koreisha and Pukkila (1990a, 1990b, 1995), Pukkila, Koreisha, and Kallinen (1990), Galbraith and Zinde-Walsh (1994, 1997), Dufour and Jouini (2005). A similar method in three steps is also presented in Hannan and Kavalieris (1984a) where the third step is presented as a correction to the second step estimates.

There are many variations around the innovation-substitution approach for the estimation of VARMA models but with the exception of Hannan and Kavalieris (1984b),[1] Dufour and Jouini (2014), [2] and us, none use a third step to get efficient estimators, surely because these procedures are often seen as a way to get initial values to start up a nonlinear optimization [*e.g.*, see Poskitt (1992), Koreisha and Pukkila (1989), Lütkepohl and Claessen (1997)]. In one of them, Koreisha and Pukkila (1989), lagged and current innovations are replaced by the corresponding residuals and a regression is performed. This is asymptotically the same as the first two steps of our method. Other variations are described in Hannan and Kavalieris (1986), Hannan and Deistler (1988), Huang and Guo (1990), Spliid (1983), Reinsel, Basu, and Yap (1992), Poskitt and Lütkepohl (1995), Lütkepohl and Poskitt (1996) and Flores de Frutos and Serrano (2002). Another approach is to use the link that exist between the VARMA parameters and the infinite VAR and VMA representations. See Galbraith, Ullah, and Zinde-Walsh (2000) for the estimation of VMA models using a VAR. VARMA models can also be estimated with subspace methods, which is based on multiple regressions and a weighted singular value decomposition [see Bauer and Wagner (2002, 2009), Bauer (2005a, 2005b)]. More recently, Dias and Kapetanios (2018) propose an Iterated OLS (IOLS) estimator where we iterate the second of our estimator until the convergence.

C. Existing methods to specify VARMA models

The identification of the orders of VARMA models depends on the representation used. Although it was one of the first representation studied, not much work has been done with the final AR equation

[1] They use a similar third step that is presented as a correction to the second step estimator but suggest that the third step should be iterated. They assume that U_t is a m.d.s.

[2] They use a similar third step for VARMA models in echelon form. They assume that U_t is i.i.d.

form. People felt that this representation gives VARMA models with too many parameters. A complete procedure to fit VARMA models under this representation is given in Lütkepohl (1993): one would first fit an ARMA$(p_i,\, q_i)$ model to every univariate time series, using maybe the procedure of Hannan and Rissanen (1982). To build the VARMA representation in final AR equation form, knowing that the VAR operator is the same for every equation we would take it to be the product of all the univariate AR polynomials. This would give a VAR operator of order $p = \sum_{i=1}^{K} p_i$. Accordingly, for the VMA part we would take $q = \max_k [q_k + \sum_{i=1, i\neq k}^{K} p_i]$. It is no wonder that people feel that the final equation form uses to many parameters.

For VARMA models in echelon form, there has been a lot more work done on the identification of Kronecker indices. The problem has been studied by, among others, Hannan and Kavalieris (1984b), Poskitt (1992) and Lütkepohl and Poskitt (1996). Non-stationary or cointegrated systems are considered by Huang and Guo (1990), Bartel and Lütkepohl (1998), and Lütkepohl and Claessen (1997). Additional references are given in Lütkepohl (1993, Chapter 8).

As for weak VARMA models estimated by QMLE, Boubacar Maïnassara (2012) propose a modified Akaike's information criteria for selecting the orders p and q.

A complementing approach to specify VARMA models, which is based on Cooper and Wood (1982), aims at finding simplifying structures via some combinations of the different series to obtain more parsimonious models. It includes Tiao and Tsay (1989), Tsay (1989a, 1989b, 1991) and Nsiri and Roy (1992, 1996).

The final stage of ARMA model specification usually involve analyzing the residuals, *i.e.* checking if they are uncorrelated. Popular tools include portmanteau tests such as Box-Pierce [Box and Pierce (1970)] and Ljung-Box [Ljung and Box (1978)] tests, and their multivariate generalization [Lütkepohl (1993, Section 5.2.9)]. Those tests are not directly applicable in our case because they are derived under strong assumptions for the innovations (independence or martingale difference). But recent developments for weak ARMA and VARMA models are applicable. They include Francq, Roy, and Zakoïan (2005), Shao (2011) and Zhu and Li (2015) (weak ARMA), Francq and Raïssi (2007) (weak VAR), Boubacar Maïnassara (2011), Katayama (2012) and Boubacar Maïnassara and Saussereau (2018) (weak VARMA).

D. Lemmas and proofs

Lemma D.1 *Let U and V be random variables measurable with respect to $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_{n}^{\infty}$, respectively where $\mathcal{F}_a^b$ is the σ-algebra of events generated by sets of the form $\{(X_{i_1}, X_{i_2}, \ldots, X_{i_n}) \in E_n\}$ with $a \leq i_1 < i_2 < \cdots < i_n \leq b$, and E_n is some n-dimensional Borel set. Let r_1, r_2, r_3 be positive numbers. Assume that $\|U\|_{r_1} < \infty$ and $\|V\|_{r_2} < \infty$ where $\|U\|_r = (E[|U|]^r)^{1/r}$. If $r_1^{-1} + r_2^{-1} + r_3^{-1} = 1$, then there exists a positive constant c_0 independent of U, V and n, such that*

$$|E[UV] - E[U]E[V]| \leq c_0 \|U\|_{r_1} \|V\|_{r_2} \alpha(n)^{1/r_3} \tag{D.1}$$

where $\alpha(n)$ is the α-mixing coefficient of order h.

Proof. See Davydov (1968).

Lemma D.2 *If the random process $\{y_t\}$ is stationary and with α-mixing coefficients $\alpha(j)$, with $E[|y_t|^{2+\zeta_1}] < \infty$ for some $\zeta_1 > 0$, and if $\sum_{j=1}^{\infty} \alpha(j)^{\zeta_1/(2+\zeta_1)} < \infty$, then*

$$\begin{aligned}\sigma^2 &\equiv \lim_{T\to\infty} Var[y_1 + \cdots + y_T] \\ &= E\left[(y_t - E[y_t])^2\right] + 2\sum_{j=1}^{\infty} E\left[(y_t - E[y_t])(y_{t+j} - E[y_{t+j}])\right].\end{aligned} \tag{D.2}$$

Moreover, if $\sigma \neq 0$ and $E[y_t] = 0$, then

$$\Pr\left[\frac{y_1 + \cdots + y_T}{\sigma\sqrt{T}} < z\right] \underset{T\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du. \tag{D.3}$$

Proof. See Ibragimov (1962).

Proof of Lemma 3.7. Clearly, $\Phi(0) = \Theta(0) = I_K$ and $\det[\Phi(0)] = \det[\Theta(0)] = 1 \neq 0$. The polynomials $\det[\Phi(z)]$ and $\det[\Theta(z)]$ are different from zero in a neighborhood of zero. So we can choose $R_0 > 0$ such that $\det[\Phi(z)] \neq 0$ and $\det[\Theta(z)] \neq 0$ for $0 \leq |z| < R_0$. It follows that the matrices $\Phi(z)$ and $\Theta(z)$ are invertible for $0 \leq |z| < R_0$. Note also that the adjoint matrices $\Phi^\star(z)$ and $\Theta^\star(z)$ are matrices of polynomials.

Let

$$C_0 = \{z \,|\, 0 \leq |z| < R_0\} \tag{D.4}$$

and

$$\Psi(z) = \Phi(z)^{-1}\Theta(z), \quad z \in C_0. \tag{D.5}$$

for $z \in C_0$. Since

$$\Phi(z)^{-1} = \frac{1}{\det[\Phi(z)]}\Phi^\star(z), \quad \Theta(z)^{-1} = \frac{1}{\det[\Theta(z)]}\Theta^\star(z), \quad \text{for } z \in C_0, \tag{D.6}$$

each element of $\Phi(z)^{-1}$ and $\Theta(z)^{-1}$ is a rational function whose denominator is different from zero on C_0. Thus, for $z \in C_0$, $\Phi(z)^{-1}$ and $\Theta(z)^{-1}$ are matrices of analytic functions, and the function

$$\Psi(z) = \Phi(z)^{-1}\Theta(z) \tag{D.7}$$

is analytic in the circle $0 \leq |z| < R_0$. Hence, it has a unique representation of the form

$$\Psi(z) = \sum_{k=0}^{\infty} \Psi_k z^k, \quad z \in C_0. \tag{D.8}$$

By assumption,

$$\Psi(z) = \Phi(z)^{-1}\Theta(z) = \bar{\Phi}(z)^{-1}\bar{\Theta}(z), \qquad z \in C_0, \tag{D.9}$$

hence, for $z \in C_0$,

$$\begin{aligned} \bar{\Phi}(z)\Phi(z)^{-1}\Theta(z) &= \bar{\Theta}(z), \\ \bar{\Phi}(z)\Phi(z)^{-1} &= \bar{\Theta}(z)\Theta(z)^{-1} \equiv \Delta(z), \end{aligned} \tag{D.10}$$

where $\Delta(z)$ is a diagonal matrix because $\Theta(z)$ and $\bar{\Theta}(z)$ are both diagonal,

$$\Delta(z) = \text{diag}\left[\delta_{ii}(z)\right], \tag{D.11}$$

where

$$\delta_{ii}(z) = \frac{\bar{\theta}_{ii}(z)}{\theta_{ii}(z)},\ \theta_{ii}(0) = 1,\ \delta_{ii}(0) = \bar{\theta}_{ii}(0),\ i = 1, \ldots , K. \tag{D.12}$$

From (D.12), it follows that each $\delta_{ii}(z)$ is rational with no pole in C_0 such that $\delta_{ii}(0) = 1$, so it can be written in the form

$$\delta_{ii}(z) = \frac{e_i(z)}{f_i(z)} \tag{D.13}$$

where $e_i(z)$ and $f_i(z)$ have no common roots, $f_i(z) \neq 0$ for $z \in C_0$ and $\delta_{ii}(0) = e_i(0) = 1$. From (D.10), it follows that for $z \in C_0$

$$\bar{\theta}_{ii}(z) = \delta_{ii}(z)\theta_{ii}(z), \quad \bar{\varphi}_{ij}(z) = \delta_{ii}(z)\varphi_{ij}(z), \quad i, j = 1, \ldots , K. \tag{D.14}$$

We first show that $\delta_{ii}(z)$ must be a polynomial. If $f_i(z) \neq 1$, then its order cannot be greater than the order $q_i \equiv \deg[\theta_{ii}(z)]$ for otherwise $\bar{\theta}_{ii}(z)$ would not be a polynomial. Similarly, if $f_i(z) \neq 1$ and is a polynomial of order less or equal to q_i, then all its roots must be roots of $\theta_{ii}(z)$ and $\varphi_{ij}(z)$, for otherwise $\bar{\theta}_{ii}(z)$ or $\bar{\varphi}_{ij}(z)$ would be a rational function. If $q_i \geq 1$, these roots are then common to $\theta_{ii}(z)$ and $\varphi_{ij}(z)$, $j = 1, \ldots , K$, which is in contradiction with Assumption **3.6**. Thus the degree of $f_i(z)$ must be zero, and $\delta_{ii}(z)$ is a polynomial.

If $\delta_{ii}(z)$ is a polynomial of degree greater than zero, this would entail that $\bar{\theta}_{ii}(z)$ and $\bar{\varphi}_{ij}(z)$ have roots in common, in contradiction with Assumption **3.6**. Thus $\delta_{ii}(z)$ must be a constant. Further, $\delta_{ii}(0) = 1$ so that for $i = 1, \ldots , K$,

$$\bar{\theta}_{ii}(z) = \theta_{ii}(z), \quad \bar{\varphi}_{ij}(z) = \varphi_{ij}(z), \quad j = 1 \ldots , K. \tag{D.15}$$

Proof of Theorem 3.8. Under the assumption that the VARMA process is invertible, we can write

$$\Theta(L)^{-1}\Phi(L)Y_t = U_t. \tag{D.16}$$

Now suppose by contradiction that there exist operators $\bar{\Phi}(L)$ and $\bar{\Theta}(L)$, with $\bar{\Theta}(L)$ diagonal and invertible, and $\bar{\Phi}(L) \neq \Phi(L)$ or $\bar{\Theta}(L) \neq \Theta(L)$, such that

$$\bar{\Theta}(L)^{-1}\bar{\Phi}(L) = \Theta(L)^{-1}\Phi(L), \tag{D.17}$$

If the above equality hold, then it must also be the case that

$$\bar{\Theta}(z)^{-1}\bar{\Phi}(z) = \Theta(z)^{-1}\Phi(z), \quad \forall z \in C_0, \tag{D.18}$$

where $C_0 = \{z \in \mathbb{C} \mid 0 \leq |z| < R_0\}$ and $R_0 > 0$. By Lemma **3.7**, it follows that

$$\bar{\Phi}(z) = \Phi(z), \quad \bar{\Theta}(z) = \Theta(z) \quad \forall z. \tag{D.19}$$

Hence, the representation is unique.

Proof of Theorem 3.10. The proof can be easily adapted from the proof of Theorem **3.8** once we replace Assumption **3.6** by Assumption **3.9**.

Lemma D.3 (Infinite VAR convergence) *If the VARMA model is invertible and if $n_T/\log(T) \to \infty$ as $T \to \infty$, then*

$$\sum_{k=1}^{K}\sum_{j=n_T+1}^{\infty} |\pi_{ik,j}| = o(T^{-1}) \quad \textit{for } i = 1, \ldots, K, \tag{D.20}$$

where $\pi_{ik,j}$ represent the parameters in $\Pi(L) = \Theta(L)^{-1}\Phi(L)$.

Proof of Lemma D.3. The matrix $\Theta(L)^{-1}$ can be seen has its adjoint matrix divided by its determinant. Since Y_t is invertible, the roots of $\det \Theta(L)$ are outside the unit circle and so the elements of $\Pi(L) = \Theta(L)^{-1}\Phi(L)$ decrease exponentially:

$$|\pi_{ik,j}| \leq c\rho^j, \ \forall i, m, \tag{D.21}$$

with $c > 0$ and $0 < \rho < 1$. From this,

$$T\sum_{k=1}^{K}\sum_{j=n_T+1}^{T} |\pi_{ik,j}| \leq T\sum_{k=1}^{K}\sum_{j=n_T+1}^{T} c\rho^j \leq c\,K\,T\,\frac{\rho^{n_T+1}}{1-\rho} \to 0 \tag{D.22}$$

as $T \to \infty$ if $n_T/\log(T) \to \infty$ because $|\rho| < 1$.

From the proof of Lemma **D.3**, we see that the condition $n_T/\log T \to \infty$ could be replaced by a weaker condition like $n_T = \kappa \log(T)$ with $\kappa > 1/\log(\rho)$ where ρ is the value given the upper bound at which the parameters $\pi_{ik,j}$ are declining to zero. A drawback if this assumption is that it would depend on the persistence of the process.

Lemma D.4 (Covariance estimation) *If the process $\{Y_t\}$ is a strictly stationary VARMA process with $\{U_t\}$ uncorrelated, $E[|u_{it}|^{4+2\zeta}] < \infty$ for some $\zeta > 0$, α-mixing with $\sum_{h=1}^{\infty} \alpha(h)^{\epsilon/(2+\epsilon)} < \infty$ for some $\epsilon > 0$, then*

$$\frac{1}{T}\sum_{t=1}^{T} y_{i,t-r}y_{i',t-s} - E[y_{i,t-r}y_{i',t-s}] = O_{ms}(T^{-1/2}) \ \ \forall i, i', r, s, \tag{D.23}$$

where ms refers to mean square convergence.

Proof of Lemma D.4. In a preliminary step, let us prove that the following result holds (assuming that $s > r$ without loss of generality):

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} Cov\left[u_{i,t-r}u_{i',t-s}\,;\,u_{i,t'-r}u_{i',t'-s}\right] = O(1/T). \tag{D.24}$$

We start by breaking this sum in the following parts:

$$\begin{aligned}
&\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} Cov\left[u_{i,t-r}u_{i',t-s}\,;\,u_{i,t'-r}u_{i',t'-s}\right] \\
= &\frac{1}{T^2}\sum_{t=1}^{T-(s-r)-1}\sum_{t'=t+1+(s-r)}^{T} Cov\left[u_{i,t-r}u_{i',t-s}\,;\,u_{i,t'-r}u_{i',t'-s}\right] \\
&+\frac{1}{T^2}\sum_{t'=1}^{T-(s-r)-1}\sum_{t=t'+1+(s-r)}^{T} Cov\left[u_{i,t-r}u_{i',t-s}\,;\,u_{i,t'-r}u_{i',t'-s}\right] \\
&+\frac{1}{T^2}\sum_{t=1+(s-r)}^{T-(s-r)}\sum_{t'=t-(s-r)}^{t+(s-r)} Cov\left[u_{i,t-r}u_{i',t-s}\,;\,u_{i,t'-r}u_{i',t'-s}\right] \\
&+\frac{1}{T^2}\sum_{t=1}^{1+(s-r)}\sum_{t'=1}^{t+(s-r)} Cov\left[u_{i,t-r}u_{i',t-s}\,;\,u_{i,t'-r}u_{i',t'-s}\right] \\
&+\frac{1}{T^2}\sum_{t'=T-(s-r)}^{T}\sum_{t=T-(s-r)-(T-t')}^{T} Cov\left[u_{i,t-r}u_{i',t-s}\,;\,u_{i,t'-r}u_{i',t'-s}\right].
\end{aligned} \tag{D.25}$$

The last three double sums are $O(1/T)$ since the covariances are finite and the number of terms is of order T. For the first two double sums, using Davydov's inequality (lemma **D.1**), the strong mixing hypothesis and the finite fourth moment we know that

$$\begin{aligned}
&\lim_{T\to\infty}\sum_{t'=t+1+(s-r)}^{T}\left|Cov\left[u_{i,t-r}u_{i',t-s}\,;\,u_{i,t'-r}u_{i',t'-s}\right]\right| \\
\leq\ &\lim_{T\to\infty}\sum_{t'=t+1+(s-r)}^{T} c_0\,\|u_{i,t-r}u_{i',t-s}\|_{2+\epsilon}\|u_{i,t'-r}u_{i',t'-s}\|_{2+\epsilon}\,\alpha(t'-t-(s-r))^{\epsilon/(2+\epsilon)} \\
<\ &\infty\,,
\end{aligned} \tag{D.26}$$

from which we conclude that the first two terms converge to zero at rate $1/T$.

Now that have the result in Equation (D.24), we first notice that by stationarity of the process,

$$E\left[\frac{1}{T}\sum_{t=1}^{T} y_{i,t-r}y_{i',t-s}\right] - E[y_{i,t-r}y_{i',t-s}] = 0. \tag{D.27}$$

Now taking the variance and writing the covariances in terms of the innovations U_t:

$$Var\left[\frac{1}{T}\sum_{t=1}^{T} y_{i,t-r}y_{i',t-s}\right] = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} Cov\left[y_{i,t-r}y_{i',t-s}\,;\, y_{i,t'-r}y_{i',t'-s}\right]$$

$$\begin{aligned} \leq \quad & \sum_{j_1=0}^{\infty}\sum_{j_1'=0}^{\infty}\sum_{j_2=0}^{\infty}\sum_{j_2'=0}^{\infty}\sum_{k_1=1}^{K}\sum_{k_1'}^{K}\sum_{k_2=1}^{K}\sum_{k_2'=1}^{K} |\psi_{ik_1,j_1}|\,|\psi_{i'k_1',j_1'}|\,|\psi_{ik_2,j_2}|\,|\psi_{i'k_2',j_2'}| \\ & \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} |Cov\left[u_{k_1,t-r-j_1}u_{k_1',t-s-j_1'}\,;\, u_{k_2,t'-r-j_2}u_{k_2',t'-s-j_2'}\right]|. \end{aligned} \tag{D.28}$$

From the assumption of stationarity we know that the ψ's are decreasing exponentially, and from Equation (D.24) we get that the right-hand side of Equation (D.28) is $O(1/T)$. Hence,

$$\frac{1}{T}\sum_{t=1}^{T} y_{i,t-r}y_{i',t-s} - E[y_{i,t-r}y_{i',t-s}] = O_{ms}(T^{-1/2}) \;\; \forall i, i', r, s. \tag{D.29}$$

Corollary D.5 (Moment estimation) *Under the assumption of Lemma* **D.4**,

$$\frac{1}{T}\sum_{t=1}^{T} y_{i,t-r}u_{i',t-s} - E[y_{i,t-r}u_{i',t-s}] = O_{ms}(T^{-1/2}) \;\; \forall i, i', r, s. \tag{D.30}$$

Proof of Lemma D.5. The proof is very similar to the proof of Lemma **D.4** where in Equation (D.28) some of the ψ's would be zero.

Proof of Theorem 4.1. We first introduce some additional matrix norms:

$$\|B\|_2^2 = \sup_{l\neq 0} \frac{l'B'Bl}{l'l}, \tag{D.31}$$

$$\|B\|_1 = \max_{i\leq j\leq n} \sum_{i=1}^{n} |b_{ij}|, \tag{D.32}$$

$$\|B\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |b_{ij}|, \tag{D.33}$$

where (D.31) is the largest eigenvalue of $B'B$. Useful inequalities relating these norms are given in

Horn and Johnson (1985, p. 313):

$$\|AB\|^2 \leq \|A\|_2^2 \|B\|^2\,, \quad \|AB\|^2 \leq \|A\|^2 \|B\|_2^2\,, \quad \|B\|_2^2 \leq \|B\|_1 \|B\|_\infty. \tag{D.34}$$

In the first step estimation, we regress

$$y_{it} = \sum_{j=1}^{n_T} \sum_{k=1}^{K} \pi_{ik,j} y_{k,t-j} + e_{it}\,, \tag{D.35}$$

when in fact

$$y_{it} = \sum_{j=1}^{\infty} \sum_{k=1}^{K} \pi_{ik,j} y_{k,t-j} + u_{it}\,. \tag{D.36}$$

On setting

$$\hat{B}(n_T) = \sum_{t=n_T+1}^{T} \frac{\mathbf{Y}'_{t-1}(n_T)\mathbf{Y}_{t-1}(n_T)}{T-n_T}, \tag{D.37}$$

OLS applied to (D.35) yields:

$$\begin{aligned}
\hat{\mathit{\Pi}}_{i\bullet}^{(n_T)} &= [\hat{\pi}_{i\bullet,1},\, \ldots\,,\, \hat{\pi}_{i\bullet,n_T}]' \\
&= \hat{B}(n_T)^{-1} \sum_{t=n_T+1}^{T} \frac{\mathbf{Y}_{t-1}^{(n_T)\prime} y_{it}}{T-n_T} \\
&= \hat{B}(n_T)^{-1} \sum_{t=n_T+1}^{T} \frac{\mathbf{Y}_{t-1}^{(n_T)\prime}}{T-n_T} \left\{ \sum_{j=1}^{\infty} \pi_{i\bullet,j} Y_{t-j} + u_{it} \right\} \\
&= \mathit{\Pi}_{i\bullet}^{(n_T)} + \hat{B}(n_T)^{-1} \sum_{t=n_T+1}^{T} \frac{\mathbf{Y}_{t-1}^{(n_T)\prime}}{T-n_T} \left\{ \sum_{j=n_T+1}^{\infty} \pi_{i\bullet,j} Y_{t-j} + u_{it} \right\}.
\end{aligned} \tag{D.38}$$

Rearranging the elements,

$$\hat{\mathit{\Pi}}_{i\bullet}^{(n_T)} - \mathit{\Pi}_{i\bullet}^{(n_T)} = \hat{B}(n_T)^{-1} \sum_{t=n_T+1}^{T} \frac{\mathbf{Y}_{t-1}^{(n_T)\prime}}{T-n_T} \left\{ \sum_{j=n_T+1}^{\infty} \pi_{i\bullet,j} Y_{t-j} \right\} + \hat{B}(n_T)^{-1} \sum_{t=n_T+1}^{T} \frac{\mathbf{Y}_{t-1}^{(n_T)\prime} u_{it}}{T-n_T}. \tag{D.39}$$

Using inequalities (D.34) and the fact that $\hat{B}(n_T)$ is symmetric,

$$\|\hat{\mathit{\Pi}}_{i\bullet}^{(n_T)} - \mathit{\Pi}_{i\bullet}^{(n_T)}\| \leq \|\hat{B}(n_T)^{-1}\|_2 \|V_{1T}\| + \|\hat{B}(n_T)^{-1}\|_2 \|V_{2T}\|\,, \tag{D.40}$$

where

$$V_{1T} = \frac{1}{T-n_T} \sum_{t=n_T+1}^{T} \mathbf{Y}_{t-1}^{(n_T)\prime} \sum_{j=n_T+1}^{\infty} \pi_{i\bullet,j} Y_{t-j}\,, \tag{D.41}$$

$$V_{2T} = \frac{1}{T-n_T} \sum_{t=n_T+1}^{T} \mathbf{Y}_{t-1}^{(n_T)\prime} u_{it} \, . \tag{D.42}$$

First, $\|V_{2T}\|^2$ can be expanded into

$$\begin{aligned} \|V_{2T}\|^2 &= tr\left(V_{2T}^{\prime} V_{2T}\right) = \sum_{k=1}^{K} \sum_{j=1}^{n_T} \left(\frac{\sum_{t=n_T+1}^{T} y_{k,t-j} u_{it}}{T-n_T} \right)^2 \\ &= \sum_{k=1}^{K} \sum_{j=1}^{n_T} \left(\underbrace{E[y_{k,t-j} u_{it}]}_{=0} + O_p(T^{-1/2}) \right)^2 . \end{aligned} \tag{D.43}$$

It follows that $\|V_{2T}\|^2 = O_p(\sqrt{n_T/T})$. Similarly, for $\|V_{1T}\|^2$

$$\begin{aligned} \|V_{1T}\|^2 &= tr\left(V_{1T}^{\prime} V_{1T}\right) = \sum_{k=1}^{K} \sum_{j=1}^{n_T} \left(\frac{\sum_{t=n_T+1}^{T} y_{k,t-j} [\sum_{j'=n_T+1}^{\infty} \sum_{k'=1}^{K} \pi_{ik',j'} y_{k',t-j'}]}{T-n_T} \right)^2 \\ &= \sum_{k=1}^{K} \sum_{j=1}^{n_T} \left(\sum_{k'=1}^{K} \sum_{j'=n_T+1}^{\infty} \pi_{ik',j'} \frac{1}{T-n_T} \sum_{t=n_T+1}^{T} y_{k,t-j} y_{k',t-j'} \right)^2 \\ &= \sum_{k=1}^{K} \sum_{j=1}^{n_T} \left(\sum_{k'=1}^{K} \sum_{j'=n_T+1}^{\infty} \pi_{ik',j'} \left[Cov[y_{k,t-j}; y_{k',t-j'}] + O_p(T^{-1/2}) \right] \right)^2 . \end{aligned} \tag{D.44}$$

From Lemma **D.3**, we know that $\sum_{j=n_T+1}^{\infty} |\pi_{ik,j}| = o(T^{-1})$ and it follows that $\sum_{j'=n_T+1}^{\infty} \pi_{ik',j'} Cov[y_{k,t-j}; y_{k',t-j'}] = o(T^{-1})$. Hence, $\|V_{1T}\|^2 = o_p(n_T T^{-1})$.

For $\|\hat{B}(n_T)^{-1}\|_1$, the existence of $\hat{B}(n_T)^{-1}$ is guaranteed by a lemma that can be found in Tiao and Tsay (1983). The argument is the following. It is clear that $\hat{B}(n_T)$ is a symmetric non-negative definite matrix. To show that it is positive definite take any arbitrary vector $c = [c_1, \dots c_{Kn_T}]'$ and consider

$$c' \hat{B}(n_T) c = \frac{1}{(T-n_T)^2} \sum_{t=n_T+1}^{T} \left(\sum_{j=1}^{n_T} \sum_{k=1}^{K} c_{(j-1)K+k} y_{k,t-j} \right)^2 . \tag{D.45}$$

If $c' \hat{B}(n_T) c = 0$, then

$$\sum_{j=1}^{n_T} \sum_{k=1}^{K} c_{(j-1)K+k} y_{k,t-j} = 0 \quad \text{for} \quad t = n_T + 1, \, \dots, \, T, \tag{D.46}$$

which, since $T > (K+1) n_T$, is a system of linear equations of $K\, n_T$ unknowns and more than $K\, n_T$ equations. Since Y_t is real-valued and non deterministic, this implies that $c = 0$ (except for a set with measure zero). This proves that $\hat{B}(n_T)$ is positive definite.

The final step is to show that $\|\hat{B}(n_T)^{-1}\|_2$ is bounded. We first see that

$$\|\hat{B}(n_T)^{-1}\|_2 \leq \|B(n_T)^{-1}\|_2 + \|\hat{B}(n_T)^{-1} - B(n_T)^{-1}\|_2 \tag{D.47}$$

where $B(n_T)$ denotes the $(K\, n_T \times K\, n_T)$ matrix of the corresponding covariances instead of the empirical covariances. As in the univariate case Berk (1974, p. 491), $\|B(n_T)^{-1}\|_2$ is uniformly bounded above by a positive constant F for all n_T since Y_t is stationary and invertible. Next, using a similar argument as in the proof of Theorem 1 in Lewis and Reinsel (1985), we show that $\|\hat{B}(n_T)^{-1} - B(n_T)^{-1}\|_2 \underset{T\to\infty}{\overset{p}{\to}} 0$. From previous results,

$$E[\|\hat{B}(n_T) - B(n_T)\|_2^2] \leq E[\|\hat{B}(n_T) - B(n_T)\|^2] \leq c_0 \, \frac{n_T^2}{T} \underset{T\to\infty}{\longrightarrow} 0 \tag{D.48}$$

for some positive constant c_0. Since

$$\begin{aligned} \|\hat{B}(n_T)^{-1} - B(n_T)^{-1}\|_2 &= \|\hat{B}(n_T)^{-1}[\hat{B}(n_T) - B(n_T)]B(n_T)^{-1}\|_2 \\ &\leq F(\|\hat{B}(n_T)^{-1} - B(n_T)^{-1}\|_2 + F)\|\hat{B}(n_T) - B(n_T)\|_2, \end{aligned} \tag{D.49}$$

we have

$$0 \leq \Xi_{n_T} = \frac{\|\hat{B}(n_T)^{-1} - B(n_T)^{-1}\|_2}{F(\|\hat{B}(n_T)^{-1} - B(n_T)^{-1}\|_1 + F)} \leq \|\hat{B}(n_T) - B(n_T)\|_2 \tag{D.50}$$

so that

$$\Xi_{n_T} \underset{T\to\infty}{\overset{p}{\to}} 0\,, \quad \|\hat{B}(n_T)^{-1} - B(n_T)^{-1}\|_2 = F^2 \Xi_{n_T}/(1 - F\Xi_{n_T}) \underset{T\to\infty}{\overset{p}{\to}} 0 \tag{D.51}$$

hence

$$\|\hat{\Pi}_{i\bullet}^{(n_T)} - \Pi_{i\bullet}^{(n_T)}\| = O_p(\sqrt{n_T/T}). \tag{D.52}$$

Proof of Theorem 4.2. If we denote by Z_{t-1} the equivalent of $\hat{Z}_{t-1}$ which contains the true innovations u_{kt} instead of the residuals $\hat{u}_{kt}$,

$$\begin{aligned} \hat{\gamma} &= \left[\sum_{t=l}^{T} \hat{Z}_{t-1}' \hat{\Sigma}_U^{-1} \hat{Z}_{t-1}\right]^{-1} \left[\sum_{t=l}^{T} \hat{Z}_{t-1} \hat{\Sigma}_U^{-1} \left(Z_{t-1}\gamma + U_t\right)\right] \\ &= \left[\sum_{t=l}^{T} \hat{Z}_{t-1}' \hat{\Sigma}_U^{-1} \hat{Z}_{t-1}\right]^{-1} \left[\sum_{t=l}^{T} \hat{Z}_{t-1}' \hat{\Sigma}_U^{-1} Z_{t-1}\right] \gamma + \\ &\quad \left[\sum_{t=l}^{T} \hat{Z}_{t-1}' \hat{\Sigma}_U^{-1} \hat{Z}_{t-1}\right]^{-1} \left[\sum_{t=l}^{T} \hat{Z}_{t-1} \hat{\Sigma}_U^{-1} U_t\right]. \end{aligned} \tag{D.53}$$

First, we show that $\hat{\Sigma}_U \xrightarrow{p} \Sigma_U$. We can write the residual $\hat{U}_t$ as

$$\begin{aligned}
\hat{U}_t &= \hat{\Pi}^{n_T}(L)Y_t = \hat{\Pi}^{n_T}(L)\Psi(L)U_t \\
&= [I_K + (\hat{\Pi}^{n_T}(L)\Psi(L) - I_K)]U_t \\
&= [I_K + (\hat{\Pi}^{n_T}(L) - \Pi(L))\Psi(L)]U_t \\
&= U_t + (\hat{\Pi}^{n_T}(L) - \Pi(L))Y_t \,. \qquad \text{(D.54)}
\end{aligned}$$

Using the results from Lemma **D.4**, Theorem **4.1** where we showed that $\sum_{l=1}^{n_T} \|\hat{\Pi}_l^{(n_T)} - \Pi_l\| = O_p(\sqrt{n_T/T})$, combined with $\sum_{l=n_T+1}^{\infty} \|\Pi_l\| = o(T^{-1})$ if $\log(T)/n_T \to 0$ as $T \to \infty$, we can conclude that

$$\hat{\Sigma}_U = \frac{1}{T-n_T}\sum_{t=n_T+1}^{T} \hat{U}_t\hat{U}_t' = \frac{1}{T-n_T}\sum_{t=n_T+1}^{T} U_tU_t' + o_p(T^{-1/2}) \xrightarrow{p} \Sigma_U \,. \qquad \text{(D.55)}$$

To show that $\frac{1}{T}\sum_{t=l}^{T} \hat{Z}_{t-1}'\hat{\Sigma}_U^{-1}\hat{Z}_{t-1}$ converge to $\tilde{J} = E[Z_{t-1}'\Sigma_U^{-1}Z_{t-1}]$ in probability, since $\hat{\Sigma}_U \xrightarrow{p} \Sigma_U$ we only have to show:

- $\frac{1}{T}\sum_{t=l}^{T} y_{i,t-j}y_{k,t-j'} \xrightarrow{p} E[y_{i,t-j}y_{k,t-j'}]$,
- $\frac{1}{T}\sum_{t=l}^{T} y_{i,t-j}\hat{u}_{k,t-j'} \xrightarrow{p} E[y_{i,t-j}u_{k,t-j'}]$,
- $\frac{1}{T}\sum_{t=l}^{T} \hat{u}_{i,t-j}\hat{u}_{k,t-j'} \xrightarrow{p} E[u_{i,t-j}u_{k,t-j'}]$.

The first is proved in Lemma **D.4**. The second can be proved in a similar manner. Start by writing

$$\begin{aligned}
\frac{1}{T}\sum_{t=l}^{T} y_{i,t-j}\hat{u}_{k,t-j'} &= \frac{1}{T}\sum_{t=l}^{T} y_{i,t-j}u_{k,t-j'} + \frac{1}{T}\sum_{t=l}^{T} y_{i,t-j}(\hat{u}_{k,t-j'} - u_{k,t-j'}) \\
&= \frac{1}{T}\sum_{t=l}^{T} y_{i,t-j}u_{k,t-j'} + \frac{1}{T}\sum_{t=l}^{T}\sum_{m=1}^{n_T}\sum_{k'=1}^{K} (\pi_{kk',m} - \hat{\pi}_{kk',m})y_{i,t-j}y_{k',t-m} \\
&\quad + \frac{1}{T}\sum_{t=l}^{T}\sum_{m=n_T+1}^{\infty}\sum_{k'=1}^{K} \pi_{kk',m}y_{i,t-j}y_{k',t-m} \qquad \text{(D.56)}
\end{aligned}$$

Proving that the first term in (D.56), $\frac{1}{T}\sum_{t=l}^{T} y_{i,t-j}u_{k,t-j'}$, converges in quadratic mean to $E[y_{i,t-j}u_{k,t-j'}]$ is very similar to the proof in Lemma **D.4** where we express $y_{i,t-j}$ as an infinite MA so we omit the details to shorten the exposition. Proving that the second and third term converge to zero in probability is also similar; combine the results of Lemma **D.4** and Theorem **4.1** for the second, Lemmas **D.3** and **D.4** for the third. Combining all these results we can conclude that $\tilde{\gamma} \xrightarrow{ms} \gamma$.

For the asymptotic distribution, since $\hat{\Sigma}_U \xrightarrow{p} \Sigma_U$, the limit distribution of $\frac{1}{\sqrt{T}}\sum_{t=l}^{T} \hat{Z}_{t-1}'\hat{\Sigma}_U^{-1}U_t$ will be the same as that of $\frac{1}{\sqrt{T}}\sum_{t=l}^{T} \hat{Z}_{t-1}'\Sigma_U^{-1}U_t$. For the latter, we can prove the asymptotic normality using an argument similar to the one used in Francq and Zakoïan (1998, Lemma 4). The argument is the following. Neglecting the constants in Σ_U^{-1}, $\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\hat{Z}'_{t-1}\Sigma_U^{-1}U_t$ contains terms such $\frac{1}{\sqrt{T}}\sum_{t=l}^{T}u_{i,t}y_{k,t-j}$ with $i,k=1,\ldots,K$ and $j=1,\ldots,\max(p,q)$. Using the MA(∞) representation of Y_t,

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=l}^{T}u_{i,t}y_{k,t-j} &= \frac{1}{\sqrt{T}}\sum_{t=l}^{T}u_{i,t}\left(\sum_{k'=1}^{K}\sum_{j'=0}^{\infty}\psi_{kk',j'}u_{k',t-j-j'}\right)\\ &= \frac{1}{\sqrt{T}}\sum_{t=l}^{T}\mathbf{A}_{r,t}^{(1)}+\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\mathbf{A}_{r,t}^{(2)} \qquad \text{(D.57)}\end{aligned}$$

where for any positive integer r,

$$\mathbf{A}_{r,t}^{(1)} = \sum_{j'=0}^{r}\sum_{k'=1}^{K}\psi_{kk',j'}u_{i,t}u_{k',t-j-j'}, \qquad \text{(D.58)}$$

$$\mathbf{A}_{r,t}^{(2)} = \sum_{j'=r+1}^{\infty}\sum_{k'=1}^{K}\psi_{kk',j'}u_{i,t}u_{k',t-j-j'}. \qquad \text{(D.59)}$$

First note that $\mathbf{A}_{r,t}^{(1)}$ is a function of a finite number of values from the process $\{U_t\}$. Therefore, the stationary process $\{\mathbf{A}_{r,t}^{(1)}\}$ satisfies a mixing property of the form (**2.3**). Lemma **D.2** implies that $\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\mathbf{A}_{r,t}^{(1)}$ has a limiting distribution $\mathcal{N}(0,\tilde{\imath}_r)$ and as $r\to\infty$, $\tilde{\imath}_r\to\tilde{\imath}$.

Now we will show that $E[\frac{1}{T}(\sum_{t=l}^{T}\mathbf{A}_{r,t}^{(2)})^2]$ converges to 0 uniformly in T as $r\to\infty$. It will follow that the limiting distribution of $\frac{1}{\sqrt{T}}\sum_{t=l}^{T}u_{i,t}y_{k,t-j}$ is the same as the limiting distribution of $\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\mathbf{A}_{r,t}^{(1)}$ from a straightforward adaptation of a result given in Anderson (1971, Corollary 7.1.1, p. 426). We have

$$\begin{aligned}&Var\left[\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\mathbf{A}_{r,t}^{(2)}\right] = Var\left[\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\sum_{j'=r+1}^{\infty}\sum_{k'=1}^{K}\psi_{kk',j'}u_{i,t}u_{k',t-j-j'}\right]\\ \leq\ & \sum_{j_1=r+1}^{\infty}\sum_{j_2=r+1}^{\infty}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}|\psi_{kk_1,j_1}||\psi_{kk_2,j_2}|\frac{1}{T}\sum_{t=l}^{T}\sum_{t'=l}^{T}|cov(u_{i,t}u_{k_1,t-j-j_1};u_{i,t'}u_{k_2,t'-j-j_2})|\\ \leq\ & \sum_{j_1=r+1}^{\infty}\sum_{j_2=r+1}^{\infty}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}|\psi_{kk_1,j_1}||\psi_{kk_2,j_2}|\frac{1}{T}\sum_{t=l}^{T}\sum_{t'=l}^{\infty}|cov(u_{i,t}u_{k_1,t-j-j_1};u_{i,t'}u_{k_2,t'-j-j_2})|\\ \leq\ & C\sum_{j_1=r+1}^{\infty}\sum_{j_2=r+1}^{\infty}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}|\psi_{kk_1,j_1}||\psi_{kk_2,j_2}| \qquad \text{(D.60)}\end{aligned}$$

for some positive constant C following a similar argument as in the proof of Lemma **D.4**. Thus,

$$\sup_T Var\left[\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\mathbf{A}_{r,t}^{(2)}\right]\to 0 \tag{D.61}$$

as $r\to\infty$.

We can extend this asymptotic normality to all the elements of $\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\hat{Z}'_{t-1}\Sigma_U^{-1}U_t$ to conclude that

$$\frac{1}{\sqrt{T}}\sum_{t=l}^{T}\hat{Z}'_{t-1}\hat{\Sigma}_U^{-1}U_t \xrightarrow{d} \mathcal{N}\left[0,\hat{I}\,\right] \tag{D.62}$$

with $\tilde{I}$ defined in Equation (4.22). From this,

$$\sqrt{T}\left(\tilde{\gamma}-\gamma\right)\xrightarrow{d}\mathcal{N}\left(0,\tilde{J}^{-1}\tilde{I}\tilde{J}^{-1}\right). \tag{D.63}$$

From the preceding results, it is obvious that $\tilde{J}$ can be consistently estimated by $\tilde{J}_T$ as defined in Equation (4.24) and using Theorem 2 of Newey and West (1987) or more explicit results from Francq and Zakoïan (2000) for weak ARMA models, we know that $\tilde{I}_T\xrightarrow{p}\tilde{I}$ if we take $m_T^4/T\to 0$ with $m_T\to\infty$ as $T\to\infty$.

Proof of Theorem 4.3. First we can rewrite X_t, W_t and $\tilde{V}_t$ as

$$X_t=\hat{\Theta}(L)^{-1}Y_t\,,\quad W_t=\hat{\Theta}(L)^{-1}\tilde{U}_t\,,\quad \tilde{V}_t=\hat{\Theta}(L)^{-1}\tilde{Z}_t. \tag{D.64}$$

We can also rewrite $\tilde{U}_t+X_t-W_t$ as

$$\begin{aligned}\tilde{U}_t+X_t-W_t &= \hat{\Theta}(L)^{-1}Y_t+\tilde{U}_t-\hat{\Theta}(L)^{-1}\tilde{U}_t\\ &= \hat{\Theta}(L)^{-1}\left[Z_{t-1}\gamma+U_t\right]+\tilde{U}_t-\hat{\Theta}(L)^{-1}\tilde{U}_t\\ &= \hat{\Theta}(L)^{-1}Z_{t-1}\gamma+\hat{\Theta}(L)^{-1}U_t+\tilde{U}_t-\hat{\Theta}(L)^{-1}\tilde{U}_t\\ &= V_{t-1}\gamma+U_t+[(\tilde{U}_t-U_t)-\hat{\Theta}(L)^{-1}(\tilde{U}_t-U_t)]\\ &= V_{t-1}\gamma+U_t+O_p(T^{-1/2}).\end{aligned} \tag{D.65}$$

With this, the regression becomes

$$\begin{aligned}\hat{\gamma} &= \left[\sum_{t=l'}^{T}\tilde{V}'_{t-1}\tilde{\Sigma}_U^{-1}\tilde{V}_{t-1}\right]^{-1}\left[\sum_{t=l'}^{T}\tilde{V}'_{t-1}\tilde{\Sigma}_U^{-1}\left(\tilde{U}_t+X_t-W_t\right)\right]\\ &= \left[\sum_{t=l'}^{T}\tilde{V}'_{t-1}\tilde{\Sigma}_U^{-1}\tilde{V}_{t-1}\right]^{-1}\left[\sum_{t=l'}^{T}\tilde{V}'_{t-1}\tilde{\Sigma}_U^{-1}V_{t-1}\right]\gamma+\\ &\quad\left[\sum_{t=l'}^{T}\tilde{V}'_{t-1}\tilde{\Sigma}_U^{-1}\tilde{V}_{t-1}\right]^{-1}\left[\sum_{t=l'}^{T}\tilde{V}'_{t-1}\tilde{\Sigma}_U^{-1}U_t\right]+o_p(T^{-1/2}).\end{aligned} \tag{D.66}$$

Just like in the proof of theorem **4.2** we see that $\hat{\gamma} - \gamma = O_p(T^{-1/2})$. With a similar application of Ibragimov's central limit theorem as in the proof of Theorem **4.2**, we conclude that

$$\sqrt{T}(\hat{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}\left(0, \hat{J}^{-1}\hat{I}\hat{J}\right)$$

where $\hat{I}$ and $\hat{J}$ are defined in Equation (4.26). As in the proof of theorem **4.2** the matrices $\hat{I}$ and $\hat{J}$ can be consistently estimated respectively by $\hat{I}_T$ and $\hat{J}_T$ as defined in Equations (4.27) and (4.28).

Proof of Theorem 4.4. The variance of MLE for i.i.d. Gaussian innovations is given in Lütkepohl (1993):

$$I = \operatorname{plim}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial U_t'}{\partial \gamma}\Sigma^{-1}\frac{\partial U_t}{\partial \gamma'}\right]^{-1}. \tag{D.67}$$

We can transform this expression so as to obtain an equation more closely related to our previous results. First, we split γ in the same two vectors γ_1 (the AR parameters) and γ_2 (the MA parameters), then we compute $\partial U_t/\partial\gamma_1'$ and $\partial U_t/\partial\gamma_2'$. We know that

$$U_t = Y_t - \Phi_1 Y_{t-1} - \cdots - \Phi_p Y_{t-p} + \Theta_1 U_{t-1} + \cdots + \Theta_q U_{t-q}. \tag{D.68}$$

So taking the derivative with respect to γ_1':

$$\frac{\partial U_t}{\partial \gamma_1'} = -Z_{\bullet 1:\dim(\gamma_1),t-1} + \Theta_1\frac{\partial U_{t-1}}{\partial \gamma_1'} + \cdots + \Theta_q\frac{\partial U_{t-q}}{\partial \gamma_1'}\ , \tag{D.69}$$

$$\Theta(L)\frac{\partial U_t}{\partial \gamma_1'} = -Z_{\bullet 1:\dim(\gamma_1),t-1}\ , \tag{D.70}$$

$$\frac{\partial U_t}{\partial \gamma_1'} = -\Theta(L)^{-1}Z_{\bullet 1:\dim(\gamma_1),t-1}\ , \tag{D.71}$$

where $Z_{\bullet 1:\dim(\gamma_1),t-1}$ is the first $\dim(\gamma_1)$ columns of Z_{t-1}. Similarly, the derivative with respect to γ_2' is

$$\begin{aligned}\frac{\partial U_t}{\partial \gamma_2'} &= -Z_{\bullet\,\dim(\gamma_1)+1:\dim(\gamma),t-1} + \Theta_1\frac{\partial U_{t-1}}{\partial \gamma_2'} + \cdots + \Theta_q\frac{\partial U_{t-q}}{\partial \gamma_2'} \\ &= -\Theta(L)^{-1}Z_{\bullet\,\dim(\gamma_1)+1:\dim(\gamma),t-1}\end{aligned} \tag{D.72}$$

Combining the two expressions we see that

$$\frac{\partial U_t}{\partial \gamma'} = -V_{t-1} \tag{D.73}$$

so the variance matrix for the maximum likelihood estimator I is equal to the matrix $J_{(3)}^{-1}$ from the third step estimation. Moreover if U_t is i.i.d. we see that we have the equality $J_{(3)} = I_{(3)}$ so that the asymptotic variance matrix that we get in the third step of our method is the same as one would get by doing maximum likelihood.

Proof of Theorem 4.5.

For the weak VARMA case, from the results in Boubacar Maïnassara and Francq (2011) we know that the asymptotic covariance matrix of the QMLE estimator of γ is equal to $J^{-1}IJ^{-1}$ with

$$I = 4\sum_{k=-\infty}^{\infty} Cov\left[U_t\ \Sigma^{-1}\ \frac{\partial U_t}{\partial \gamma'}\ ;\ U_{t-k}\ \Sigma^{-1}\ \frac{\partial U_{t-k}}{\partial \gamma'}\right]\ ,\quad J = 2E\left[\frac{\partial U_t'}{\partial \gamma}\ \Sigma^{-1}\ \frac{\partial U_t}{\partial \gamma'}\right]\ . \qquad \text{(D.74)}$$

In the proof of Theorem **4.4** we established that $\partial U_t/\partial\gamma' = V_{t-1}$. From this we see that $J = 2J_{(3)}$, $I = 4I_{(3)}$ and our third-step estimator has the same asymptotic variance-covariance matrix as the non-linear least squares estimator if the innovations are only uncorrelated.

Proof of Theorem 5.1.

Let us denote by $\tilde{\Sigma}_U(p,q)$ the value taken by $\tilde{\Sigma}_U$ for given orders p and q. The true value of p and q is denoted by p_0 and q_0. The difference between the information criterion for given values of the orders p and q, and the true values p_0, q_0 is equal to

$$\log\left(\det\tilde{\Sigma}_U(p,q)\right) - \log\left(\det\tilde{\Sigma}_U(p_0,q_0)\right) + [\dim\gamma(p,q) - \dim\gamma(p_0,q_0)]\,\frac{(\log T)^{1+\delta}}{T}. \qquad \text{(D.75)}$$

First, consider the case where $p < p_0$ or $q < q_0$. In this case, as T grows to infinity, uniformly across the orders (p,q), eventually $\det\tilde{\Sigma}_U(p,q) > \det\tilde{\Sigma}_U(p_0,q_0)$ because of the left-coprime property. As argued for example in Hannan and Rissanen (1982, p.90) if this result does not hold, then it would mean that a model with smaller orders would be giving the minimum prediction error. So while the penalty is increasing with the sample size, (D.75) would become positive as $T\to\infty$. So eventually we must have $\hat{p} \geq p_0$ and $\hat{q} \geq q_0$.

Next, to discuss the case where $p \geq p_0$ and $q \geq q_0$, we can start by writing the residuals of the second step estimation as

$$\begin{aligned}
\tilde{U}_t &= \tilde{\Phi}(L)Y_t - \left(\tilde{\Theta}(L) - I_K\right)\hat{U}_t \\
&= \tilde{\Phi}(L)Y_t - \left(\tilde{\Theta}(L) - I_K\right)\hat{\Pi}^{(n_T)}(L)Y_t \\
&= \left[\tilde{\Phi}(L) - \left(\tilde{\Theta}(L) - I_K\right)\hat{\Pi}^{(n_T)}(L)\right]Y_t \\
&= \left[\tilde{\Phi}(L) - \left(\tilde{\Theta}(L) - I_K\right)\hat{\Pi}^{(n_T)}(L)\right]\Psi_0(L)U_t \\
&= \left[\tilde{\Phi}(L) - \tilde{\Theta}(L)\hat{\Pi}^{(n_T)}(L) + \hat{\Pi}^{(n_T)}(L)\right]\Psi_0(L)U_t \\
&= \left[\left(\tilde{\Phi}(L) - \Phi_0(L)\right) + \Phi_0(L) - \tilde{\Theta}(L)\hat{\Pi}^{(n_T)}(L) + \left(\hat{\Pi}^{(n_T)}(L) - \Pi_0(L)\right) + \Pi_0(L)\right]\Psi_0(L)U_t \\
&= \left[\left(\tilde{\Phi}(L) - \Phi_0(L)\right) + \left(\Theta_0(L) - \tilde{\Theta}(L)\right)\Pi_0(L) - \tilde{\Theta}(L)\left(\hat{\Pi}^{(n_T)}(L) - \Pi_0(L)\right) + \right. \\
&\qquad \left.\left(\hat{\Pi}^{(n_T)}(L) - \Pi_0(L)\right) + \Pi_0(L)\right]\Psi_0(L)U_t
\end{aligned}$$

$$
\begin{aligned}
&= \Big[\Big(\tilde{\Phi}(L) - \Phi_0(L)\Big)\Psi_0(L) + \Big(\Theta_0(L) - \tilde{\Theta}(L)\Big) - \tilde{\Theta}(L)\Big(\hat{\Pi}^{(n_T)}(L) - \Pi_0(L)\Big)\Psi_0(L) + \\
&\qquad \Big(\hat{\Pi}^{(n_T)}(L) - \Pi_0(L)\Big)\Psi_0(L) + I_K\Big] U_t \\
&= \Big[\Big(\tilde{\Phi}(L) - \Phi_0(L)\Big)\Psi_0(L) + \Big(\Theta_0(L) - \tilde{\Theta}(L)\Big) - \tilde{\chi}(L) + C(L) + I_K\Big] U_t \,, \qquad \text{(D.76)}
\end{aligned}
$$

where

$$
\tilde{\chi}(L) = \tilde{\Theta}(L)\Big(\hat{\Pi}^{(n_T)}(L) - \Pi_0(L)\Big)\Psi_0(L)\,, \qquad \text{(D.77)}
$$

$$
C(L) = \Big(\hat{\Pi}^{(n_T)}(L) - \Pi_0(L)\Big)\Psi_0(L). \qquad \text{(D.78)}
$$

For the case where $p = p_0$ and $q = q_0$, from the results of Theorems **4.1** and **4.2**, it follows that with an obvious abuse of notation[3]

$$
\|(\tilde{\Phi}(L) - \Phi_0(L))\Psi_0(L)\| = O_p(T^{-1/2})\,, \qquad \text{(D.79)}
$$

$$
\|\tilde{\chi}(L)\| = O_p(\sqrt{n_T/T})\,, \quad \|C(L)\| = O_p(\sqrt{n_T/T})\,. \qquad \text{(D.80)}
$$

Using the above representation of the residuals $\tilde{U}_t$, we get

$$
\tilde{\Sigma}_U(p_0, q_0) = \frac{1}{T}\sum_{t=n_T+1}^{T} U_t U_t' + O_p(n_T T^{-1}). \qquad \text{(D.81)}
$$

Next, we discuss the case where $p \geq p_0$ and $q \geq q_0$ with $p > p_0$ or $q > q_0$. To put a lower bound on $\det(\tilde{\Sigma}_U(p,q))$ across (p,q), we re-arrange equation D.76 as

$$
\tilde{U}_t = \big[\Upsilon(L)\Theta_0(L)^{-1}\Psi_0(L) - \chi(L) + C(L) + I_K\big] U_t\,, \qquad \text{(D.82)}
$$

where

$$
\Upsilon(L) = \Phi(L)\Theta_0(L) - \Theta(L)\Phi_0(L)\,, \qquad \text{(D.83)}
$$

$$
\chi(L) = \Theta(L)\Big(\hat{\Pi}^{(n_T)}(L) - \Pi_0(L)\Big)\Psi_0(L)\,, \qquad \text{(D.84)}
$$

with $\Phi(L)$ and $\Theta(L)$ of orders p and q respectively that would satisfy Assumption **2.1**. We can first point out that the term $C(L)$ does not vary with the orders (p,q). Also, uniformly across the orders $\|\chi(L)\| = O_p(\sqrt{n_T/T})$. Now, We can minimize $\det\tilde{\Sigma}_U(p,q)$ with respect to Υ freely and then with respect to Θ. This will give a minimum no greater than the true minimum obtained from the minimization over Φ and Θ. Through the minimization with respect to Υ we see that uniformly across the orders

$$
\det(\tilde{\Sigma}_U(p,q)) \geq \det\left(\frac{1}{T}\sum_{t=1}^{T} U_t U_t'\right) + O_p(n_T/T)\,. \qquad \text{(D.85)}
$$

[3]For example, by $\|(\Theta(L) - \tilde{\Theta}(L))\|^2$ we mean $\sum_{i=1}^{K}\sum_{k=1}^{K}\sum_{j=1}^{q}(\theta_{ik,j} - \tilde{\theta}_{ik,j})^2$.

Thus eventually

$$\log[\det \tilde{\Sigma}_U(p,q)] - \log[\det \tilde{\Sigma}_U(p_0,q_0)] + [\dim\gamma(p,q) - \dim\gamma(p_0,q_0)]\,\frac{(\log T)^{1+\delta}}{T} \geq [\dim\gamma(p,q) - \dim\gamma(p_0,q_0)]\,\frac{(\log T)^{1+\delta}}{T} + O_p(n_T/T)\,. \tag{D.86}$$

If we choose $n_T = O((\log T)^{1+\delta_1})$ with $\delta_1 > 0$ and $\delta_1 < \delta$, the above lower bound is positive unless $p = p_0$ and $q = q_0$. and the probability that $\hat{p} = p_0$ and $\hat{q} = q_0$ converges to one as $T \to \infty$.

Proof of Theorem 5.2. The proof is similar to the proof of Theorem **5.1**.

References

ANDERSON, T. (1971): *The Statistical Analysis of Time Series*. Wiley, New York.

ATHANASOPOULOS, G., AND F. VAHID (2008): “VARMA versus VAR for macroeconomic forecasting,” *Journal of Business and Economic Statistics*, 26(2), 237–252.

BARTEL, H., AND LÜTKEPOHL (1998): “Estimating the Kronecker indices of cointegrated echelon-form VARMA models,” *Econometrics Journal*, 1, C76–C99.

BAUER, D. (2005a): “Comparing the CCA Subspace Method to Pseudo Maximum Likelihood Methods in the Case of No Exogenous Inputs,” *Journal of Time Series Analysis*, 26(5), 631–668.

——— (2005b): “Estimating Linear Dynamical Systems Using Subspace Methods,” *Econometric Theory*, 21, 181–211.

BAUER, D., AND M. WAGNER (2002): “Estimating cointegrated systems using subspace algorithms,” *Journal of Econometrics*, 111, 47–84.

——— (2009): “Using subspace algorithm cointegration analysis: Simulation performance and application to the term structure,” *Computational Statistics and Data Analysis*, 53(5), 1954–1973.

BERK, K. N. (1974): “Consistent autoregressive spectral estimates,” *The Annals of Statistics*, 2(3), 489–502.

BOSQ, D. (1998): *Nonparametric Statistics for Stochastic Processes - Estimation and Prediction*, no. 110 in Lecture Notes in statistics. Springer-Verlag, Berlin, second edn.

BOUBACAR MAÏNASSARA, Y. (2011): “Multivariate portmanteau test for structural VARMA models with uncorrelated but non-independent error terms,” *Journal of Statistical Planning and Inference*, 141, 2961–2975.

——— (2012): “Selection of weak VARMA models by modified Akaike’s information criteria,” *Journal of Time Series Analysis*, 33, 121–130.

BOUBACAR MAINASSARA, Y., M. CARBON, AND C. FRANCQ (2012): “Computing and estimating informaiton matrices of weak ARMA models,” *Computational Statistics and Data Analysis*, 56, 345–361.

BOUBACAR MAÏNASSARA, Y., AND C. FRANCQ (2011): “Estimating structural VARMA models with uncorrelated but non-independent error terms,” *Journal of Multivariate Analysis*, 102, 496–505.

BOUBACAR MAÏNASSARA, Y., AND H. RAÏSSI (2015): “Semi-strong linearity testing in linear models with dependent but uncorrelated errors,” *Statistics and Probability Letters*, 103, 110–115.

BOUBACAR MAÏNASSARA, Y., AND B. SAUSSEREAU (2018): “Diagnostic checking in multivariate ARMA models with dependent errors using normalized residual autocorrelations,” *Journal of the American Statistical Association*, 113(524), 1813–1827.

BOX, G. E. P., AND D. A. PIERCE (1970): “Distribution of residual autocorrelations in autoregressive-integrated moving average time series models,” *Journal of the American Statistical Association*, 65, 1509–1526.

BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time series: theory and methods*, Springer Series in Statistics. Springer-Verlag, New York, second edn.

CHEN, B., J. CHOI, AND J. C. ESCANCIANO (2017): “Testing for fundamental vector moving average representations,” *Quantitative Economics*, 8(1), 149–180.

COOPER, D. M., AND E. F. WOOD (1982): “Identifying Multivariate Time Series Models,” *Journal of Time Series Analysis*, 3, 153–164.

DAVYDOV, Y. A. (1968): “Convergence of Distributions Generated by Stationary Stochastic Processes,” *Theory of Probability and its Applications*, pp. 691–696.

DIAS, G. F., AND G. KAPETANIOS (2018): “Estimation and forecasting in vector autoregressive moving average models for rich datasets,” *Journal of Econometrics*, 202(1), 75 – 91.

DOUKHAN, P. (1995): *Mixing - Properties and Examples*, no. 85 in Lecture Notes in Statistics. Springer-Verlag.

DROST, F. C. (1993): “Temporal Aggregation of Time-Series,” in *Econometric Analysis of Financial Markets*, ed. by J. Kaehler, and P. Kugler, pp. 11–21. Physica-Verlag, New York.

DUFOUR, J.-M., AND T. JOUINI (2005): “Asymptotic Distribution of a Simple Linear Estimator for Varma Models in Echelon Form,” in *Statistical Modeling and Analysis for Complex Data Problems*, ed. by P. Duchesne, and B. Rémillard, chap. 11, pp. 209–240. Springer US, Boston, MA.

——— (2014): “Asymptotic distributions for quasi-efficient estimators in echelon VARMA models,” *Computational Statistics & Data Analysis*, 73, 69 – 86.

DURBIN, J. (1960): “The Fitting of Time Series Models,” *Revue de l’Institut International de Statistique*, 28, 233.

FLORES DE FRUTOS, R., AND G. R. SERRANO (2002): “A Generalized Least Squares Estimation Method for VARMA Models,” *Statistics*, 36(4), 303–316.

FRANCQ, C., AND H. RAÏSSI (2007): “Multivariate Portmenteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors,” *Journal of Time Series Analysis*, 28(3), 454–470.

FRANCQ, C., R. ROY, AND J.-M. ZAKOÏAN (2005): “Diagnostic checking in ARMA models with uncorrelated errors,” *Journal of the American Statistical Association*, 100, 532–544.

FRANCQ, C., AND J.-M. ZAKOÏAN (1998): “Estimating Linear Representations of Nonlinear Processes,” *Journal of Statistical Planning and Inference*, 68, 145–165.

——— (2000): “Covariance matrix estimation for estimators of mixing weak ARMA models,” *Journal of Statistical Planning and Inference*, 83, 369–394.

GALBRAITH, J. W., A. ULLAH, AND V. ZINDE-WALSH (2000): "Estimation of the Vector Moving Average Model by Vector Autoregression," *Econometric Reviews*, 21(2), 205–219.

GALBRAITH, J. W., AND V. ZINDE-WALSH (1994): "A Simple Noniterative Estimator for Moving Average Models," *Biometrika*, 81(1), 143–155.

——— (1997): "On Simple, Autoregression-Based Estimation and Identification Techniques for ARMA Models," *Biometrika*, 84(3), 685–696.

HANNAN, E. J., AND M. DEISTLER (1988): *The Statistical Theory of Linear Systems*. John Wiley & Sons, New York.

HANNAN, E. J., AND L. KAVALIERIS (1984a): "A Method for Autoregressive-Moving Average Estimation," *Biometrika*, 71(2), 273–280.

——— (1984b): "Multivariate Linear Time Series Models," *Advances in Applied Probability*, 16, 492–561.

——— (1986): "Regression, Autoregression Models," *Journal of Time Series Analysis*, 7(1), 27–49.

HANNAN, E. J., L. KAVALIERIS, AND M. MACKISACK (1986): "Recursive Estimation of Linear Systems," *Biometrika*, 73(1), 119–133.

HANNAN, E. J., AND J. RISSANEN (1982): "Recursive Estimation of Mixed Autoregressive-Moving-Average Order," *Biometrika*, 69, 81–94, Errata 70 (1982), 303.

HORN, R. G., AND C. A. JOHNSON (1985): *Matrix Analysis*. Cambridge University Press, Cambridge, U.K.

HUANG, D., AND L. GUO (1990): "Estimation of Nonstationary ARMAX Models Based on the Hannan-Rissanen Method," *The Annals of Statistics*, 18(4), 1729–1756.

IBRAGIMOV, I. A. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and its Applications*, 7, 349–382.

KATAYAMA, N. (2012): "Chi-squared portmanteau tests for structural VARMA models with uncorrelated errors," *Journal of Time Series Analysis*, 33, 863–872.

KOREISHA, S. G., AND T. M. PUKKILA (1989): "Fast Linear Estimation Methods for Vector Autoregressive Moving-Average Models," *Journal of Time Series Analysis*, 10(4), 325–339.

——— (1990a): "A Generalized Least-Squares Approach for Estimation of Autoregressive-Moving-Average Models," *Journal of Time Series Analysis*, 11(2), 139–151.

——— (1990b): "Linear Methods for Estimating ARMA and Regression Models with Serial Correlation," *Communications in Statistics, Part B -Simulation and Computation*, 19, 71–102.

——— (1995): "A Comparison Between Different Order-Determination Criteria for Identification of ARIMA Models," *Journal of Business and Economic Statistics*, 13, 127–131.

LEWIS, R., AND G. C. REINSEL (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16, 393–411.

LJUNG, G. M., AND G. E. P. BOX (1978): "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65(2), 297–303.

LUCEÑO, A. (1994): "A Fast Algorithm For The Exact Likelihood Of Stationary And Partially Nonstationary Vector Autoregressive-Moving Average Processes," *Biometrika*, 81(3), 555–565.

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

——— (1993): *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, second edn.

LÜTKEPOHL, H., AND H. CLAESSEN (1997): "Analysis of Cointegrated VARMA Processes," *Journal of Econometrics*, 80, 223–239.

LÜTKEPOHL, H., AND D. S. POSKITT (1996): "Specification of Echelon-Form VARMA Models," *Journal of Business and Economic Statistics*, 14, 69–80.

MAURICIO, J. A. (2002): "An Algorithm for the Exact Likelihood of a Stationary Vector Autoregressive-Moving Average Model," *Journal of Time Series Analysis*, 23(4), 473–486.

MÉLARD, G., R. ROY, AND A. SAIDI (2006): "Exact maximum likelihood estimation of structured or unit root multivariate time series models," *Computational Statistics & Data Analysis*, 50(11), 2958 – 2986.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

NIJMAN, T. H., AND F. PALM (1990): "Parameter Identification in ARMA Processes in the Presence of Regular But Incomplete Sampling," *Journal of Time Series Analysis*, 11(3), 239–248.

NSIRI, S., AND R. ROY (1992): "On the Identification of ARMA Echelon-Form Models," *Canadian Journal of Statistics*, 20(4), 369–386.

NSIRI, S., AND R. ROY (1996): "Identification of Refined ARMA Echelon Form Models for Multivariate Time Series," *Journal of Multivariate Analysis*, 56, 207–231.

PALM, F., AND T. H. NIJMAN (1984): "Missing Observations in the Dynamic Regression Model," *Econometrica*, 52(6), 1415–1436.

POSKITT, D. S. (1987): "A Modified Hannan-Rissanen Strategy for Mixed Autoregressive-Moving Average Oder Determination," *Biometrika*, 74(4), 781–790.

——— (1992): "Identification of Echelon Canonical Forms for Vector Linear Processes Using Least Squares," *The Annals of Statistics*, 20, 195–215.

POSKITT, D. S., AND H. LÜTKEPOHL (1995): "Consistent Specification of Cointegrated Autoregressive Moving-Average Systems," Discussion Paper 54, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.

PUKKILA, T., S. KOREISHA, AND A. KALLINEN (1990): "The Identification of ARMA Models," *Biometrika*, 77(3), 537–548.

REINSEL, G. C., S. BASU, AND S. F. YAP (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.

SAIKKONEN, P. (1986): "Asymptotic Properties of some Preliminary Estimators for Autoregressive Moving Average Time Series Models," *Journal of Time Series Analysis*, 7, 133–155.

SHAO, X. (2011): "Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models," *Econometric Theory*, 27, 312–343.

SHEA, B. L. (1989): "The Exact Likelihood of a Vector Autoregressive Moving Average Model," *Applied Statistics*, 38(1), 161–184.

SPLIID, H. (1983): "A Fast Estimation Method for the Vector Autoregressive Moving Average Model with Exogenous Variables," *Journal of the American Statistical Association*, 78(384), 843–849.

TIAO, G. C., AND R. S. TSAY (1983): "Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models," *The Annals of Statistics*, 11, 856–871.

——— (1989): "Model Specification in Multivariate Time Series," *Journal of the Royal Statistical Society, Series B*, 51(2), 157–213.

TSAY, R. S. (1989a): "Identifying Multivariate Time Series Models," *Journal of Time Series Analysis*, 10, 357–372.

TSAY, R. S. (1989b): "Parsimonious Parameterization of Vector Autoregressive Moving Average Models," *Journal of Business and Economic Statistics*, 7(3), 327–341.

TSAY, R. S. (1991): "Two Canonical Forms for Vector VARMA Processes," *Statistica Sinica*, 1, 247–269.

ZHAO-GUO, C. (1985): "The Asymptotic Efficiency of a Linear Procedure of Estimation for ARMA Models," *Journal of Time Series Analysis*, 6(1), 53–62.

ZHU, K., AND W. K. LI (2015): "A boostrapped spectral test for adequacy in weak ARMA models," *Journal of Econometrics*, 187, 113–130.