**ADVANCED REVIEW** OPEN ACCESS

# Vector AutoRegressive Moving Average Models: A Review

Marie-Christine Düker[1] | David S. Matteson[2] | Ruey S. Tsay[3] | Ines Wilms[4] 🆔

[1]Department of Statistics and Data Science, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany | [2]Department of Statistics and Data Science, Cornell University, Ithaca, New York, USA | [3]Booth School of Business, University of Chicago, Chicago, Illinois, USA | [4]School of Business and Economics, Maastricht University, Maastricht, The Netherlands

**Correspondence:** Ines Wilms (i.wilms@maastrichtuniversity.nl)

## ABSTRACT

Vector AutoRegressive Moving Average (VARMA) models form a powerful and general model class for analyzing dynamics among multiple time series. While VARMA models encompass the Vector AutoRegressive (VAR) models, their popularity in empirical applications is dominated by the latter. Can this phenomenon be explained fully by the simplicity of VAR models? Perhaps many users of VAR models have not fully appreciated what VARMA models can provide. The goal of this review is to provide a comprehensive resource for researchers and practitioners seeking insights into the advantages and capabilities of VARMA models. We start by reviewing the identification challenges inherent to VARMA models thereby encompassing classical and modern identification schemes and we continue along the same lines regarding estimation, specification, and diagnosis of VARMA models. We then highlight the practical utility of VARMA models in terms of Granger Causality analysis, forecasting and structural analysis as well as recent advances and extensions of VARMA models to further facilitate their adoption in practice. Finally, we discuss some interesting future research directions where VARMA models can fulfill their potentials in applications as compared to their subclass of VAR models.

## 1 | Introduction

Vector AutoRegressive Moving Averages (VARMAs) have long been considered a fundamental model class for multivariate time series. VARMA models extend the popular ARMA framework (Box and Jenkins 1976) to vector time series thereby permitting practitioners to learn dynamic interrelationships between the component series and to explore the cross-dependence to add prediction of each individual series.

Several strong reasons exist for modeling multivariate time series in a VARMA framework: (i) VARMA models typically permit more parsimonious representations of the data generating process than pure Vector Autoregressive (VAR) models; which may lead, in turn, to estimation and forecast accuracy gains; see, for instance, Tiao and Box (1981). (ii) The class of VARMA models is closed under many basic linear transformations, marginalization and temporal aggregation, whereas the class of VAR models is not; see Lütkepohl (2005, Chapter 11), and Amendola, Niglio, and Vitale (2010) for textbook introductions. (iii) VARMA models are closely linked to other widely used econometric models such as linear simultaneous equation models (e.g., Wallis 1977; Zellner and Palm 1974 for the link with VARMA in final equation form) or dynamic models such as Dynamic Stochastic General Equilibrium (DSGE) models or rational expectation (RE) models in economics: linearized DSGEs imply that the

variables of interest are generated by a VARMA process, not a VAR one (e.g., Fernández-Villaverde et al. 2007; Komunjer and Ng 2011) and also RE models have VARMAs as their reduced form (e.g., Vinod 1996). Already more than 25 years ago, Cooley and Dwyer (1998) highlighted "While VARMA models involve additional estimation and identification issues, these complications do not justify systematically ignoring these moving average components, as in the SVAR approach."

In theory, VARMA models ought to be preferable over VAR models. Nonetheless, VARMA models are scarcely used in practice as their adoption is complicated by identification, estimation, and specification difficulties, which arise primarily due to the flexibility of the model structure leading easily to over-parametrization if care is not exercised. In contrast, VARs dominate empirical work in multivariate time series analysis since they are direct generalizations of multivariate linear regression. Over the years, various proposals have been made to make VARMA more accessible to applied researchers. In this paper, we review the literature on VARMA models to further ease their adoption by practitioners and applied researchers.

The remainder of this article is structured as follows. Section 2 starts by briefly reviewing the VARMA model. Section 3 highlights the identification problem in VARMA models and reviews some commonly used identification schemes available in the literature. Section 4 addresses estimation, order specification and diagnosis of VARMA models. Section 5 reviews the main usage of VARMA while Section 6 presents the most commonly adopted extensions. Section 7 concludes with a perspective on the need for future research directions.

## 2 | The VARMA Model

In this section, we present a compact review of the VARMA model and refer the reader to Hannan and Deistler (1988), Reinsel (1993), Brockwell and Davis (1991), Lütkepohl (2005), or Tsay (2013), among many others, for more extensive introductions to VARMA models.

Let $y_t$ be a stationary $d$-dimensional mean-zero vector time series. It follows a $\text{VARMA}_d(p, q)$ model if

$$y_t = \sum_{\ell=1}^{p} \Phi_\ell y_{t-\ell} + \sum_{m=1}^{q} \Theta_m a_{t-m} + a_t, \qquad (1)$$

where $\left\{\Phi_\ell \in \mathbb{R}^{d\times d}\right\}_{\ell=1}^{p}$ are the autoregressive (AR) parameter matrices, $\left\{\Theta_m \in \mathbb{R}^{d\times d}\right\}_{m=1}^{q}$ the moving-average (MA) parameter matrices, and $\{a_t\}$ denotes a $d$-dimensional mean-zero white noise vector time series with $d \times d$ nonsingular contemporaneous covariance matrix $\Sigma_a$. The VARMA model states that $y_t$ is a function of its own $p$ past values and $q$ lagged error terms. Model (1) can be re-written as

$$\Phi(L)y_t = \Theta(L)a_t,$$

using the compact AR and MA matrix polynomials in lag operator given by

$$\Phi(L) = I - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$$
$$\text{and} \quad \Theta(L) = I + \Theta_1 L + \Theta_2 L^2 + \cdots + \Theta_q L^q,$$

where the lag operator $L^\ell$ is defined as $L^\ell y_t = y_{t-\ell}$ and $I$ denotes the $d \times d$ identity matrix.

The VARMA model is stable if $\det\{\Phi(z)\} \neq 0$ for all $|z| \leq 1$ ($z \in \mathbb{C}$) and is invertible if $\det\{\Theta(z)\} \neq 0$ for all $|z| \leq 1$ ($z \in \mathbb{C}$). It is unit-root nonstationary if $\det\{\Phi(z)\} = 0$ for some $|z| = 1$. If the VARMA model is invertible, it has a pure VAR representation given by

$$\Pi(L)y_t = a_t,$$

where $\Pi(L) = \Theta^{-1}(L)\Phi(L) = I - \Pi_1 L - \Pi_2 L^2 - \cdots$. The $\Pi$-matrices in the infinite-order VAR representation can be obtained recursively from the AR matrices $\{\Phi_\ell\}$ and MA matrices $\{\Theta_m\}$:

$$\Pi_i = \Phi_i + \Theta_i - \sum_{j=1}^{i-1} \Theta_{i-j}\Pi_j, \quad i = 1, 2, \ldots$$

with $\Theta_0 = I$, $\Theta_i = 0$, for $i > q$ and $\Phi_i = 0$, for $i > p$. The VARMA model is uniquely defined in terms of the operator $\Pi(L)$, but not in terms of the AR and MA operators $\Phi(L)$ and $\Theta(L)$, in general. See some specific examples in Tsay (2013), among others. In fact, this identification problem for VARMA models is well known in the literature, early discussions on this date back to, among others, Hannan (1969, 1971); Akaike (1974, 1976). Some identification conditions of VARMA models are also available in the literature. See, for instance, the block identifiability conditions in Dunsmuir and Hannan (1976). Next, we discuss this identification problem in further detail and review two approaches available in the literature to overcome this identifiability problem.

## 3 | Identification

Consider the $\text{VARMA}_d(p, q)$ of equation (1) with fixed AR order $p$ and MA order $q$. For a given $\Pi(L)$, $p$, and $q$, one can define an equivalence class of AR and MA matrix polynomial pairs,

$$\mathscr{E}_{p,q}(\Pi(L)) = \{(\Phi, \Theta): \Phi(L) = \Theta(L)\Pi(L)\}, \qquad (2)$$

where $\Phi = \begin{bmatrix} \Phi_1 \cdots \Phi_p \end{bmatrix}$ and $\Theta = \begin{bmatrix} \Theta_1 \cdots \Theta_q \end{bmatrix}$. This class can easily consist of multiple (or even infinitely many) pairs, implying that further identification restrictions on the AR and MA matrices are needed for meaningful model specification and estimation.

Simply put, for a given $d$-dimensional time series $y_t$ with $d > 1$, the identifiability problem arises because the two integers $p$ and $q$ are not sufficient to describe its dynamic structure. To illustrate, suppose that $d = 2$ and $p = q = 1$. In this case, we have $y_t = (y_{1,t}, y_{2,t})^\top$ and $\Phi(L) = I - \Phi_1 L$ and $\Theta(L) = I + \Theta_1 L$. Here $p$ and $q$ do not provide any information on the structures of $\Phi_1$ and $\Theta_1$, yet these structures provide the dynamic dependence of $y_t$. Suppose further that $y_{1,t}$ is in fact a white noise series while $y_{2,t}$ depends on $\{y_{1,t-1}, y_{2,t-1}$ and $a_{2,t-1}\}$. In this particular instance, the $\text{VARMA}_2(1, 1)$ model for $y_t$ should assume the form

$$y_t - \begin{bmatrix} 0 & 0 \\ \Phi_{21,1} & \Phi_{22,1} \end{bmatrix} y_{t-1} = a_t - \begin{bmatrix} 0 & 0 \\ 0 & \Theta_{22,1} \end{bmatrix} a_{t-1}, \quad (3)$$

for the model to be estimable. One would encounter problems in estimation if an unrestricted $VARMA_2(1, 1)$ model were used in estimation, because the likelihood function is not well defined then. Of particular interest in the specification of the model in equation (3) is that we set $\Theta_{21,1} = 0$. This is so because, being white noise, $y_{1,t-1} = a_{1,t-1}$ so that only one of the two parameters $\{\Phi_{21,1}, \Theta_{21,1}\}$ can be used in the model. The identifiability problem becomes more complicated for higher values of $d$, $p$ and $q$.

We review here the most commonly used identification schemes. They are the Echelon methodology (Section 3.1) and the scalar component methodology (Section 3.2). We conclude this section with some more recent approaches on model identification (Section 3.3).

## 3.1 | Echelon-Form

Arguably the most popular identification procedure is the *Echelon form identification*. The Echelon form methodology was developed in the engineering literature under the linear dynamic system; see Hannan and Deistler (1988), Hannan and Kavalieris (1984), Poskitt (1992), and the references therein. The basic idea amounts to the use of a set of $d$ Kronecker indices, which are integers and invariant with respect to the ordering of the components of $y_t$, to determine the lag structure and the number of free parameters in the model. These indices effectively capture the depth of each variable's influence in the system, allowing for a more efficient representation by excluding unnecessary parameters. The end result of using Echelon form is to obtain an identifiable $VARMA_d(k, k)$ model for $y_t$, with $k = \max\{p, q\}$.

We discuss the basic concept here. A simple way to understand the Echelon form of the time series $y_t$ is from the prediction point of view. For simplicity, we assume that $y_t$ is stationary with mean-zero and let $\Gamma_k = \text{Cov}(y_t, y_{t-k})$ be the lag-$k$ autocovariance matrix of $y_t$. If $y_t$ is unit-root nonstationary, then one can replace $\Gamma_k$ by the $\Pi_k$ matrix of the VAR representation of $y_t$ in the following discussion.

### 3.1.1 | Kronecker Index

Consider a time index $t$. Let $F_t = (y_t^\top, y_{t+1}^\top, y_{t+2}^\top, \ldots)^\top$ and $P_{t-1} = (y_{t-1}^\top, y_{t-2}^\top, \ldots)^\top$ be, respectively, the *future* and *past* vectors of $y_t$. Define an infinite-dimensional *Hankel matrix* of $y_t$ as

$$H_\infty = \text{Cov}(F_t, P_{t-1}) = E(F_t P_{t-1}^\top) = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 & \cdots \\ \Gamma_2 & \Gamma_3 & \Gamma_4 & \cdots \\ \Gamma_3 & \Gamma_4 & \Gamma_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (4)$$

Clearly, $H_\infty$ is a *Toeplitz matrix* in which the 2nd $d$-block row is a subset of the first $d$-block row, and so on.

Let $m = \text{rank}(H_\infty)$. It is easy to show, via the moment equations, that $y_t$ follows a $VARMA(p, q)$ model if and only if the rank $m$ is finite; see Lemma 4.1 of Tsay (2013). Let $h(i, j)$ denote the $[(i - 1)d + j]$th row of $H_\infty$, where $j = 1, \ldots, d$ and $i = 1, 2, \ldots$. From the definition in equation (4), we see that $h(i, j) = E(y_{j,t+i-1} P_{t-1}^\top)$, which represents the linear dependence of $y_{j,t+i-1}$ on the past vector $P_{t-1}$ of $y_t$. Next, we say that $h(i, j)$ is a *predecessor* of $h(u, v)$ if $(i - 1)d + j < (u - 1)d + v$. Using the Toeplitz property of $H_\infty$, one can easily see that if $h(i, j)$ is a linear combination of its predecessors $\{h(i_1, j_1), h(i_2, j_2), \cdots, h(i_s, j_s)\}$, then $h(i + 1, j)$ is a linear combination of its predecessors $\{h(i_1 + 1, j_1), h(i_2 + 1, j_2), \cdots, h(i_s + 1, j_s)\}$; see Lemma 4.2 of Tsay (2013).

**Definition:** For the $j$th component $y_{j,t}$ of $y_t$, the *Kronecker index* $k_j$ is the smallest non-negative integer $i$ such that $h(i + 1, j)$ of $H_\infty$ is linearly dependent of its predecessors.

To illustrate, consider the bi-variate VARMA(1,1) model in equation (3). Since $y_{1,t}$ is white noise, which does not depend on $P_{t-1}$, we have $h(1, 1) = 0$ and the Kronecker index $k_1 = 0$. Next, for the VARMA(1,1) model, the moment equations are $\Gamma_k - \Phi_1 \Gamma_{K-1} = 0$, for $k > 1$, implying that the 2nd $d$-block row of $H_\infty$ is a linear combination of the 1st $d$-block row. Consequently, we have $k_2 = 1$ for $y_{2,t}$.

The collection of Kronecker indices $\{k_1, \ldots, k_d\}$ of $y_t$ forms the Kronecker index set of the series, and they provide a clear description of the dynamic dependence of $y_t$. For the particular $VARMA_2(1, 1)$ process $y_t$ in equation (3), the Kronecker index set is $\{0, 1\}$. Note that the Kronecker index $k_j$ is for the component $y_{j,t}$ so that the index depends on the ordering of the components of $y_t$, but the Kronecker index set is invariant with respect to the ordering of the components of $y_t$. Furthermore, it is also easy to see that, for a VARMA model, $\sum_{j=1}^{d} k_j = m$, which is the rank of $H_\infty$. Again, for the model in equation (3), it is easily seen that $\sum_{j=1}^{2} k_j = 1 = m$, which is the rank of $H_\infty$ of $y_t$.

### 3.1.2 | Model Specification via Kronecker Indices

In this section, we show that the Kronecker index set $\{k_1, \ldots, k_d\}$ provides a concrete structural specification of the $VARMA_d(p, q)$ model for $y_t$. The notation used in this section is a bit complicated as we try to give a detailed description of the dynamic dependence of each component of $y_t$. We refer the reader to Tsay (2013) for further details.

To facilitate a better understanding of the dynamic structure implied by Kronecker indices, it is helpful to think of the Hankel matrix of $y_t$ as follows:

| Block | Future component | $P_{t-1}^\top = (y_{t-1}^\top, y_{t-2}^\top, \ldots)$ |
|---|---|---|
| 1 | $y_{1,t}$ | $h(1, 1)$ |
| | $y_{2,t}$ | $h(1, 2)$ |
| | $\vdots$ | $\vdots$ |
| | $y_{d,t}$ | $h(1, d)$ |

| Block | Future component | $P_{t-1}^{\top} = (y_{t-1}^{\top}, y_{t-2}^{\top}, \dots)$ |
|---|---|---|
| 2 | $y_{1,t+1}$ | $h(2,1)$ |
| | $y_{2,t+1}$ | $h(2,2)$ |
| | $\vdots$ | $\vdots$ |
| | $y_{d,t+1}$ | $h(2,d)$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $k_j$ | $y_{1,t+k_j-1}$ | $h(k_j,1)$ |
| | $y_{2,t+k_j-1}$ | $h(k_j,2)$ |
| | $\vdots$ | $\vdots$ |
| | $y_{d,t+k_j-1}$ | $h(k_j,d)$ |
| $k_j+1$ | $y_{1,t+k_j}$ | $h(k_j+1,1)$ |
| | $y_{2,t+k_j}$ | $h(k_j+1,2)$ |
| | $\vdots$ | $\vdots$ |
| | $y_{d,t+k_j}$ | $h(k_j+1,d)$ |

Let $\{k_1, \cdots, k_d\}$ be the set of Kronecker indices of $y_t$. Consider the first component $y_{1,t}$. By the definition, $h(i,1)$ is not a linear combination of its predecessors, for $i = 1, \dots, k_1$, but $h(k_1+1,1)$ is a linear combination of its predecessors. Therefore, from the aforementioned $H_\infty$ structure, we have

$$h(k_1+1,1) = \sum_{u=1}^{k_1} \sum_{i=1}^{d} \alpha_{u,i,1} h(u,i),$$

where $\alpha_{u,i,1}$ is a real number and the summation is zero if its upper limit is smaller than its lower limit. In general, for the $j$th component $y_{j,t}$ with Kronecker index $k_j$, we have

$$h(k_j+1,j) = \sum_{i=1}^{j-1} \alpha_{k_j+1,i,j} h(k_j+1,i) + \sum_{u=1}^{k_j} \sum_{i=1}^{d} \alpha_{u,i,j} h(u,i), \quad (5)$$

where, again, $\alpha_{u,i,j}$ denotes a real number and the first subscript $k_j+1$ of $\alpha_{k_j+1,i,j}$ signifies a concurrent time index. By rearranging the summation according to the second argument of $h(u,i)$, we can rewrite equation (5) as

$$h(k_j+1,j) = \sum_{i=1}^{j-1} \sum_{u=1}^{k_j+1} \alpha_{u,i,j} h(u,i) + \sum_{i=j}^{d} \sum_{u=1}^{k_j} \alpha_{u,i,j} h(u,i). \quad (6)$$

Next, consider jointly all Kronecker indices. That is, consider equation (6) simultaneously for $j = 1, \dots, d$. For each $i$, $h(u,i)$ is a linear combination of its predecessors if $u > k_i$. Therefore, equation (6) can be simplified as

$$h(k_j+1,j) = \sum_{i=1}^{j-1} \sum_{u=1}^{k_j+1 \wedge k_i} \beta_{u,i,j} h(u,i) + \sum_{i=j}^{d} \sum_{u=1}^{k_i \wedge k_j} \beta_{u,i,j} h(u,i), \quad j = 1, \dots, d,$$

$$(7)$$

where $u \wedge v = \min(u,v)$ and coefficients $\beta_{u,i,j}$ are linear combinations of the coefficients $\alpha_{u,i,j}$ in equation (6). The $d$ equation in (7) jointly specify a detailed structure of VARMA model for $y_t$. In particular, the number of coefficients of the $j$th equation in (7) is

$$\delta_j = \sum_{i=1}^{j-1} \min(k_j+1, k_i) + k_j + \sum_{i=j+1}^{d} \min(k_j, k_i), \quad (8)$$

which turns out to be the number of AR parameters needed for $y_{j,t}$ in the specified VARMA model for $y_t$.

To make it more precisely, we define an infinite dimensional vector $\psi_j$ based on the $j$th equation of (7) below. Denote the $[(u-1)d+i]$th element of $\psi_j$ by $\psi_{u,i,j}$. Then,

1. let $\psi_{k_j+1,j,j} = 1$, that is, the $[k_j \times d + j]$th element of $\psi_j$ is 1;

2. for each $\beta_{u,i,j}$ coefficient on the right hand side of equation (7), let $\psi_{u,i,j} = -\beta_{u,i,j}$;

3. let all other elements of $\psi_j$ be zero.

By equation (7), we have

$$\psi_j^{\top} H_\infty = 0. \quad (9)$$

Let $w_{j,t+k_j} = \psi_j^{\top} F_t$, where $F_t$ is the future vector of $y_t$ at time index $t$ and the last non-zero element of $w_{j,t+k_j}$ is $y_{j,t+k_j}$. Then, equation (9) implies, from the definition of $H_\infty$, that $w_{j,t+k_j}$ is uncorrelated with the past vector $P_{t-1}$ of $y_t$. Consequently, $w_{j,t+k_j}$ must be a linear combination of $\{a_{t+k_j}, a_{t+k_j-1}, \dots, a_t\}$. As a matter of fact, we have

$$w_{j,t+k_j} = \sum_{i=0}^{k_j} u_{i,j}^{\top} a_{t+k_j-i}, \quad (10)$$

where $u_{i,j}$s are $d$-dimensional row vectors such that

$$u_{0,j} = \left(\psi_{k_j+1,1,j}, \dots, \psi_{k_j+1,j-1,j}, 1, 0, \dots, 0\right),$$

with 1 being in the $j$th position and it is understood that $\psi_{k+j+1,i,j} = 0$ if $k_i < k_j + 1$ and $i < j$. Equation (10) says that $w_{j,t+k_j}$ is an $\mathrm{MA}(k_j)$ time series.

Finally, from the definition of $\psi_j$ and equation (7), we also have

$$w_{j,t+k_j} = y_{j,t+k_j} + \sum_{i=1}^{j-1} \sum_{u=1}^{k_j+1 \wedge k_i} \psi_{u,i,j} y_{i,t+u-1} + \sum_{i=j}^{d} \sum_{u=1}^{k_j \wedge k_i} \psi_{u,i,j} y_{i,t+u-1}. \quad (11)$$

Combining equations (10) and (11) and noting that $\psi_{k_j+1,i,j} = 0$ if $k_i < k_j + 1$ and $i < j$, we have specified an equation for $y_{j,t}$ as

$$y_{j,t+k_j} + \sum_{i=1}^{j-1} \sum_{u=1}^{k_j+1 \wedge k_i} \psi_{u,i,j} y_{i,t+u-1} + \sum_{i=j}^{d} \sum_{u=1}^{k_j \wedge k_i} \psi_{u,i,j} y_{i,t+u-1}$$

$$= a_{j,t+k_j} + \sum_{i<j, k_i<k_j+1} \psi_{k_j+1,i,j} a_{i,t+k} + \sum_{i=1}^{k_j} u_{i,j}^{\top} a_{t+k_j-i}. \quad (12)$$

By stationarity of $y_t$, we can change the time index from $t + k_j$ to $t$ throughout the above equation and it continues to hold. Putting equation (12) together, for $j = 1, \ldots, d$, we see that the Kronecker index set $\{k_1, \ldots, k_d\}$ specifies a well-defined $\text{VARMA}_d(p^*, p^*)$ model for $y_t$, where $p^* = \max\{k_1, \ldots, k_d\}$. Also, from equation (12), the number of parameters for $y_{j,t}$ is $\delta_j + k_j \times d$, where $\delta_j$ is defined in equation (8) and $k_j \times d$ is the number of parameters in the MA part. Consequently, the number of coefficient parameters of the specified $\text{VARMA}_d(p^*, p^*)$ model via the Kronecker indices is $N = \sum_{j=1}^{d} \delta_j + d \sum_{j=1}^{d} k_j$. The resulting $\text{VARMA}_d(p^*, p^*)$ model for $y_t$ is said to be in the *Echelon Form*.

Based on the specification in equation (12), we see that the Echelon form puts coefficients in the MA part when AR and MA parameters are exchangeable. For the particular $\text{VARMA}_2(1, 1)$ example in equation (3), Echelon form would estimate $\Theta_{21,1}$ instead of $\Phi_{21,1}$. Theoretically speaking, this is not a problem as only one of $\{\Phi_{21,1}, \Theta_{21,1}\}$ is allowed in the VARMA model.

### 3.1.3 | Echelon VARMA Models

Given the set of Kronecker indices $\{k_1, \ldots, k_d\}$ of a $d$-dimensional time series $y_t$, we can obtain the structural specification of a $\text{VARMA}_d(p^*, p^*)$ model for $y_t$ by considering jointly the $d$ equation in (12), where $p^* = \max\{k_1, \ldots, k_d\}$. The specified Echelon form contains further information of the dynamic dependence of $y_t$ than an overall model. To see this, we summarize the specified VARMA model below: The model assumes the form

$$\Phi(L)y_t = \Theta(L)a_t \quad \text{with} \quad \Phi(L) = \Phi_0 - \sum_{i=1}^{p^*} \Phi_i L^i \quad \text{and} \quad \Theta(L) = \Theta_0 + \sum_{j=1}^{p^*} \Theta_j L^j,$$

where $\Phi_0 = \Theta_0$ is a lower triangular matrix with diagonal elements being 1. Denote further the $(r, s)$th elements of the $i$th matrices $\Phi_i$ and $\Theta_i$ by $\Phi_{rs,i}$ and $\Theta_{rs,i}$, respectively, and write $\Phi(L) = [\Phi_{rs}(L)]$ and $\Theta(L) = [\Theta_{rs}(L)]$, for $r, s = 1, \ldots, d$. Let $n_{rs}$ be the number of coefficients in the polynomial $\Phi_{rs}(L)$ and $m_{rs}$ be the number of coefficients in the polynomial $\Theta_{rs}(L)$. Here both $n_{rs}$ and $m_{rs}$ include the unknown coefficients in $\Phi_0$, if any. From the equation in (12), we have

$$n_{rs} = \begin{cases} \min\{k_r, k_s\} & \text{if } r \leq s, \\ \min\{k_r + 1, k_s\} & \text{if } r > s, \end{cases}$$

$$m_{rs} = \begin{cases} k_r & \text{if } r \leq s \text{ or } (r > s \text{ and } k_r \geq k_s), \\ k_r + 1 & \text{if } r > s \text{ and } k_r < k_s. \end{cases}$$

The equation in (12) also imply that

$$\Phi_{rs}(L) = \begin{cases} 1 - \sum_{i=1}^{k_r} \Phi_{rr,i} L^i & \text{if } r = s, \\ -\sum_{i=k_r+1-n_{rs}}^{k_r} \Phi_{rs,i} L^i & \text{if } r \neq s, \end{cases}$$

for $r, s = 1, \ldots, d$ and

$$\Theta_{rs}(L) = \begin{cases} 1 + \sum_{i=1}^{k_r} \Theta_{rr,i} L^i & \text{if } r = s, \\ \sum_{i=k_r+1-m_{rs}}^{k_r} \Theta_{rs,i} L^i & \text{if } r \neq s, \end{cases}$$

for $r, s = 1, \ldots, d$.

### 3.1.4 | Discussion

The Echelon form offers multiple benefits in identifying VARMA representations. Firstly, its definition is solely reliant on the Kronecker indices, eliminating the need for additional constraints on the coefficients to distinctly determine the VARMA structure. As a matter of fact, it specifies an equation for each component $y_{i,t}$ in a matrix framework. Secondly, it gives positions of estimable coefficients of the VARMA models. Thirdly, its inherent simplicity alleviates computational challenges associated with likelihood maximization. Lastly, there are established methods for accurately estimating the Kronecker indices in finite-dimensional vector processes.

While offering a reliable and well-studied identification procedure, the Echelon form has also some drawbacks. In particular, in the high-dimensional setting, when the dimension $d$ and orders $p, q$ might be large, the Echelon form suffers from selecting *Kronecker orders* from a $O((p+q)^d)$-dimensional set, by comparing an equally large number of models. Data-driven strategies, involving a series of canonical correlation tests, or regressions based on model selection criteria (e.g., AIC, BIC, information theoretic criterion) were proposed (Akaike 1976; Tsay 1989; Poskitt 1992). However, all of these methods are computationally intensive and require a large sample size to work well. Assuming $d$ is fixed, Poskitt (1992) proves asymptotic theory for the specification step. Then, assuming Kronecker orders are known, consistency of parameter estimation follows via maximum likelihood methods under the multivariate Gaussian assumption. This procedure has been tested only on very small $d$, and finite sample performances deserve a further investigation; see Section 3.4 in Lütkepohl (2006) and Chapter 4 of Tsay (2013).

### 3.1.5 | Finding Kronecker Indices

The Echelon form is the most commonly used identification scheme and has ever since its development been an active research area by either benefiting from its advantages or attempting to make the identification scheme more tractable. We review some of those efforts here.

To identify the Echelon form, Tsay (1989) and Nsiri and Roy (1992, 1996) present procedures based on the examination of the linear dependence among rows of the Hankel matrix that either summarizes autocorrelation or employ the smallest canonical correlation between the past and future vectors of $y_t$. To be more precise, they define test statistics for the null hypothesis of linear dependence between correlation vectors; see Section 4.4 of Tsay (2013) for details and examples.

Note that it is not necessary to estimate the AR or MA parameters to specify the Kronecker indices in data analysis. We briefly outline the procedure in Section 4.4 of Tsay (2013). For a given data set $\{y_t\}_{t=1}^{T}$, where $T$ is the sample size. Let $P_t^o = \left( y_{t-1}^{\top}, \ldots, y_{t-p_o}^{\top} \right)^{\top}$ be a proxy of the past vector of $y_t$ at the time index $t$. In practice, the order $p_o$ can be selected either using the AIC or examining the sequential chi-square test of zero partial autoregressive matrices of Tiao and Box (1981). Then, one can test the linear dependence of elements of the future vector $F_t$ on $P_t^o$ one by one starting with the first row $y_{1t}$. Specifically, we divide the future vector $F_t$ into two disjoint vectors $F_t^*$ and $F_t^c$, where $F_t^*$ denote the future vector to be tested and $F_t^c$ consists of the remaining candidates. One starts with $F_t^* = \{y_{1t}\}$ and $F_t^c = F_t - F_t^*$ and tests the null hypothesis that the smallest canonical correlation between $F_t^*$ and $P_t^o$ is zero. To this end, Tsay (2013) uses a chi-square test. If the null hypothesis is rejected, then move the first element of $F_t^c$ into $F_t^*$ and repeat the testing procedure. If the null hypothesis cannot be rejected, the last row of $F_t^*$ is linearly dependent on $P_t^o$. In this case, the Kronecker index for the last element of $F_t^*$ is found. One then deletes the last element of $F_t^*$ and removes all of its future elements from $F_t^c$ for any further consideration. If $F_t^c$ becomes an empty set, all Kronecker indices have been found and the search is stopped. If $F_t^c$ is not empty, one moves its first element into $F_t^*$ and repeats the testing procedure. Obviously, $F_t^c$ will be empty when $d$ kronecker indices are found for a $d$-dimensional series $y_t$.

Ratsimalahelo (2001) proposed an algorithm which selects a maximal set of linearly independent rows of the Hankel matrix of the estimated covariances. This set is obtained by sequentially testing the smallest singular value of the Hankel matrix and yields estimates of Kronecker indices which characterize the Echelon form. Using the matrix perturbation theory, the asymptotic distribution of the test statistic is seen to be chi-squared.

Poskitt (2016) develops a new methodology for identifying the structure of VARMA time series models. The analysis proceeds by examining the Echelon canonical form and presents a fully automatic data driven approach to model specification using a new technique to determine the Kronecker invariants. In a more recent work, Bhansali (2020) identifies three major difficulties with an established Echelon form approach in identifying a model from observed data: A lack of choice, overparameterization and structural rigidity. Their approach to address those issues is to specify a range of different multistep Echelon forms.

## 3.2 | Scalar Component Methodology

Another popular identification and specification method is the *Scalar Component Model* (SCM) which was first introduced in Tiao and Tsay (1989) and further developed in Athanasopoulos and Vahid (2008); Athanasopoulos, Poskitt, and Vahid (2012). We refer to Tsay (1991) for a comparison of the Echelon and the SC methodologies.

The SCM approach decomposes a multivariate series into scalar components, which are linear combinations of $y_{i,t}$s. This decomposition simplifies the model identification process by

allowing the researcher (1) to seek linear transformations of $y_t$ to reveal its dynamic structure and (2) to focus on specification of each SCM within a VARMA framework. Consequently, the SCM approach is considerably easier to handle than the full VARMA structure. For a $d$-dimensional series $y_t$, once $d$ linearly independent SCMs are given, one can specify a $\text{VARMA}_d(p,q)$ model for $y_t$ in which all estimable coefficients are identified. In contrast to the Kronecker index approach, the SCM approach specifies a $\text{VARMA}_d(p,q)$ model for $y_t$ without any constraints on $p$ and $q$ so long as they are finite. This refinement over the Kronecker index approach comes with the price of requiring more intensive computation in searching for the SCMs.

### 3.2.1 | Scalar Components

One of the motivations for developing SCM is that in many empirical applications some linear combinations of $y_t$ become a white noise series, even when some components $y_{i,t}$ are unit-root nonstationary; see, for instance, Box and Tiao (1977). The $j$th component $y_{j,t}$ can be written as $y_{j,t} = e_{0,j}^{\top} y_t$, where $e_{0,j}$ is the $j$th unit vector. That is, $e_{0,j} = (0, \ldots, 0, 1, 0, \ldots, 0)^{\top}$ with 1 being at the $j$th position. SCM is simply to employ a general non-zero $d$-dimensional vector $v_0$.

**Definition**: $w_t = v_0^{\top} y_t$ is a scalar component of order $(r,s)$ of $y_t$, where $v_0$ is a non-zero $d$-dimensional vector, if there exist $r$ vectors $v_1, \cdots, v_r$, with $v_r \neq 0$, such that $z_t = w_t + \sum_{i=1}^{r} v_i^{\top} y_{t-i}$ satisfies (a) $\text{E}(a_{t-h} z_t) = 0$, for $h > s$, and (b) $\text{E}(a_{t-s} z_t) \neq 0$.

We denote the $w_t$ of the above definition as a $\text{SCM}(r,s)$ component. Recall that $P_{t-h} = (y_{t-h}^{\top}, y_{t-h-1}^{\top}, \ldots)^{\top}$, for $h > 0$. From the definition, we see that $\text{E}(z_t P_{t-h}) = 0$, for $h > s$, but $\text{E}(z_t P_{t-s}) \neq 0$. Thus, if $w_t$ is a $\text{SCM}(r,s)$ of $y_t$, then $w_t$ depends on $y_{t-r}$ and $a_{t-s}$. It may or may not depend on $y_{t-1}, \ldots, y_{t-r+1}$ or $a_{t-1}, \ldots, a_{t-s+1}$. In fact, if $w_t = v_0^{\top} y_t$ is a $\text{SCM}(r,s)$ of $y_t$, then there exist vectors $v_1, \cdots, v_r$ and $u_1, \cdots, u_s$ such that

$$v_0^{\top} y_t + \sum_{i=1}^{r} v_i^{\top} y_{t-i} = v_0^{\top} a_t + \sum_{i=1}^{s} u_i^{\top} a_{t-i}, \quad (13)$$

where $v_r$ and $u_s$ are non-zero. The MA part of the above equation follows from that the left hand side of equation (13) $z_t = \sum_{i=0}^{r} v_i^{\top} y_{t-i}$ is uncorrelated with $a_{t-h}$ for $h > s$. The SCM approach to VARMA model specification is to make use of equation (13) jointly for $d$ linearly independent SCMs. Details are in the next subsection.

Three properties of SCM are relevant to our discussion below. First, if $w_t$ is a $\text{SCM}(r,s)$ of $y_t$, then $cw_t$ is also a $\text{SCM}(r,s)$ of $y_t$ if $c \neq 0$. This implies that SCMs are scale invariant. Second, if $w_{1,t}$ is a $\text{SCM}(r_1, s_1)$ and $w_{2,t}$ is a $\text{SCM}(r_2, s_2)$ of $y_t$, then $\alpha_1 w_{1,t} + \alpha_2 w_{2,t}$ is a $\text{SCM}(r^*, s^*)$ of $y_t$, provided that $(\alpha_1, \alpha_2) \neq 0$, where $r^* = \max\{r_1, r_2\}$ and $s^* = \max\{s_1, s_2\}$. This property is easily seen from the definition of SCM. Third, suppose $w_{1,t}$ and $w_{2,t}$ are SCMs of $y_t$ with orders $(r_1, s_1)$ and $(r_2, s_2)$, respectively. If $r_1 < r_2$ and $s_1 < s_2$, then one can embed $w_{1,t}$ in $w_{2,t}$ so that $\min\{r_2 - r_1, s_2 - s_1\}$ coefficients in $w_{2,t}$ can be set to zero. Tiao and Tsay (1989) refer to those parameters as *redundant parameters*. The simple example in equation (3)

serves as an illustration, for which $y_{1,t}$ is a SCM(0,0) and $y_{2,t}$ is a SCM(1,1) of $y_t$. Therefore, there is a redundant parameter in the equation of $y_{2,t}$. As another example, suppose that $w_{i,t}$ is a SCM$(r_i, s_i)$ of $y_t$, for $i = 1$ and 2, with $r_1 = s_1 = 1$ and $r_2 = s_2 = 2$. In this case, by the definition, we have

$$v_{0,1}^\top y_t + v_{1,1}^\top y_{t-1} = v_{0,1}^\top a_t + u_{1,1}^\top a_{t-1}, \tag{14}$$

where all three vectors $v_{0,1}, v_{1,1}$ and $u_{1,1}$ are non-zero and it is understood that $w_{1,t} = v_{0,1}^\top y_t$. Similarly, we have

$$v_{0,2}^\top y_t + v_{1,2}^\top y_{t-1} + v_{2,2}^\top y_{t-2} = v_{0,2}^\top a_t + u_{1,2}^\top a_{t-1} + u_{2,2}^\top a_{t-2}, \tag{15}$$

where $v_{0,2}, v_{2,2}$ and $u_{2,2}$ are non-zero vectors and $w_{2,t} = v_{0,2}^\top y_t$. Let $v_{1,1,2}$ and $u_{1,1,2}$ be the first elements of $v_{1,2}$ and $u_{1,2}$, respectively. Then, we can see that only one of $\{v_{1,1,2}, u_{1,1,2}\}$ is needed in equation (15). This is so because, from equation (14), we have

$$y_{1,t-1} + \sum_{j=2}^{d} v_{j,0,1} y_{j,t-1} + v_{1,1}^\top y_{t-2} = a_{1,t-1} + \sum_{j=2}^{d} v_{j,0,1} a_{t-1} + u_{1,1}^\top a_{t-2}, \tag{16}$$

where $v_{j,0,1}$ is the $j$th element of $v_{0,1}$, for $j = 1, \ldots, d$, and we assume $v_{0,0,1} = 1$ for simplicity as SCM is scale invariant. Multiplying (16) by $-v_{1,1,2}$ and adding the resulting equation to equation (15), we see that the coefficient of $y_{1,t-1}$ becomes zero while we maintain $w_{2,t}$ as a SCM(2, 2) of $y_t$. Consequently, we can set either $v_{1,1,2}$ or $u_{1,1,2}$ to zero. In general, for any two SCMs $w_{i,t}$ of order $(r_i, s_i)$, for $i = 1$ and 2, the total number of redundant parameters in the first equation is $\eta_1 = \max\{0, \min(r_1 - r_2, s_1 - s_2)\}$ and that of the second equation is $\eta_2 = \max\{0, \min(r_2 - r_1, s_2 - s_1)\}$.

### 3.2.2 | Model Specification by SCM

For a $d$-dimensional time series $y_t$, suppose that we have $d$ scalar components of orders $(r_i, s_i)$, for $i = 1, \ldots, d$. That is, we have $w_{j,t} = v_{0,j}^\top y_t$ is SCM$(r_j, s_j)$. We say that the $d$ SCMs are linearly independent if the matrix $T$ is non-singular, where $T$ is a $d \times d$ matrix with $j$th row being $v_{0,j}^\top$. In practice, we want the orders $(r_j, s_j)$ to be as small as possible in the sense that $r_j + s_j$ is minimized. This requirement is achieved in the searching procedure in finding SCM, which we discussed in the next subsection.

Let $p^* = \max\{r_1, \ldots, r_d\}$ and $q^* = \max\{s_1, \ldots, s_d\}$. Then, the $d$ SCMs specify a VARMA$_d(p^*, q^*)$ for $y_t$. This specification is achieved by putting together the equation (13) for each $w_{j,t}$. More precisely, we have

$$Ty_t + \sum_{i=1}^{p^*} \Xi_i y_{t-i} = Ta_t + \sum_{i=1}^{q^*} \Omega_i a_{t-i}, \tag{17}$$

where $T$, as before, is the matrix consisting of $v_{0,j}$ and $\Xi_i$ and $\Omega_i$ are coefficient matrices whose rows are given as follows. Let the $j$th row of $\Xi_i$ and $\Omega_i$ be $\Xi_{j,.i}$ and $\Omega_{j,.i}$, respectively. Then,

$$\Xi_{j,.i} = \begin{cases} v_{i,j}^\top & \text{if } i \leq r_j \\ 0 & \text{if } i > r_j \end{cases} \quad \text{and} \quad \Omega_{j,.i} = \begin{cases} u_{i,j}^\top & \text{if } i \leq s_j \\ 0 & \text{if } i > s_j, \end{cases}$$

where $v_{i,j}$s are the vectors associated with the SCM $w_{j,t}$ and $u_{i,j}$ is a $d$-dimensional vector.

The VARMA model in equation (17) is not complete because there may exist some redundant parameters. The positions of those redundant parameters can be identified by using the method discussed in Section 3.2.1. For the joint VARMA$_d(p^*, q^*)$ model in equation (17), the total number of redundant parameters is

$$\tau = \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \text{IND}\left[\min(r_j - r_i, s_j - s_i) > 0\right],$$

where IND$(\cdot)$ denotes the indicator function.

Note that if we define $w_t = Ty_t$, then we can rewrite equation (17) in terms of the transformed series $w_t$. The model structure remains unchanged because $\Xi_i y_{t-i} = \Xi_i T^{-1} T y_{t-i} \equiv \Xi_i^* w_{t-i}$, where $\Xi_i$ and $\Xi_i^*$ have the same zero row structure, as a zero row vector multiplied by a matrix remains a zero row vector.

### 3.2.3 | Finding SCM

Tiao and Tsay (1989) propose a procedure to find SCMs. The procedure performs sequentially eigen-analysis of certain expanded covariance matrices of $y_t$ and applies a chi-square test to detect the number of SCMs. Similarly to the search of Kronecker indices, it is unnecessary to estimate AR or MA parameters in specifying SCMs in data analysis. What is needed is the sample covariance matrices of some expanded vectors of $y_t$. We briefly review the procedure in this section.

For a $d$-dimensional time series $y_t$, define an expanded vector $Y_{m,t} = (y_t^\top, y_{t-1}^\top, \ldots, y_{t-m}^\top)^\top$, which is of dimension $d(m + 1)$, where $m \geq 0$. For $m \geq 0$ and $j \geq 0$, consider the covariance matrix

$$\Gamma(m, j) = \text{Cov}(Y_{m,t}, Y_{m,t-j-1}) = \begin{bmatrix} \Gamma_{j+1} & \Gamma_{j+2} & \Gamma_{j+3} & \cdots & \Gamma_{j+1+m} \\ \Gamma_j & \Gamma_{j+1} & \Gamma_{j+2} & \cdots & \Gamma_{j+m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Gamma_{j+1-m} & \Gamma_{j+2-m} & \Gamma_{j+3-m} & \cdots & \Gamma_{j+1} \end{bmatrix}.$$

Tiao and Tsay (1989) consider a two-way table of $\Gamma(m, j)$, for $m, j = 0, 1, \ldots$. From the moment equations of $y_t$, the existence of a SCM$(r, s)$ implies that there is a zero eigenvalue in $\Gamma(r, s)$. In fact, if $w_t$ is a SCM$(r, s)$ of $y_t$, then there exist $d$-dimensional vectors $\{v_0, v_1, \ldots, v_r\}$, with $v_0 \neq 0$ and $v_r \neq 0$, such that $z_t = \sum_{i=0}^{r} v_i^\top y_{t-i}$ satisfies $\text{E}(z_t a_{t-h}) = 0$, for $h > s$. Thus, by counting the number of zero eigenvalues in the two-way table formed by $\Gamma(m, j)$, one can gain ideas on the SCMs. A complication arises, however. For the above SCM$(r, s)$ component $w_t$, there exist two zero eigenvalues in $\Gamma(r + 1, s + 1)$. This is so because both $\{0, v_0, v_1, \ldots, v_r\}$ and $\{v_0, v_1, \ldots, v_r, 0\}$ would give rise to the same SCM, where 0 denotes a $d$-dimensional zero vector. More precisely, the two SCMs are $w_t$ and $w_{t-1}$. They are identical under stationarity. This issue is referred to as a *double counting problem* in Tiao and Tsay (1989). To overcome this issue, the authors consider a *diagonal difference* of the number of zero eigenvalues. Specifically, let $n(m, j)$

denote the number of zero eigenvalues of $\Gamma(m, j)$. The diagonal difference is defined as $d(m, j) = n(m + 1, j + 1) - n(m, j)$. Then, the overall VARMA order for $y_t$ is the position of the upper-left corner of a two-way table of $d(m, j)$ formed by a lower-right square consisting of entries $d$.

Finally, Tiao and Tsay (1989) propose to search for SCM sequentially starting with $\Gamma(0, 0)$, then along the sequence given by $m + j = 1, 2, \ldots$ until $d$ linearly independent SCMs are found. In this way, the procedure ensures that the selected orders $(r_i, s_i)$ are as small as possible. For $m + j = c$, one can start with $\Gamma(c, 0), \Gamma(c - 1, 1)$, and so forth. We refer the reader to Tiao and Tsay (1989) and Chapter 4 of Tsay (2013) for more details.

### 3.2.4 | Discussion

One of the main advantages of the SCM is its ability to simplify the otherwise daunting task of parameter estimation in VARMA models. By breaking down the model into more manageable parts, SCM reduces the computational burden and potential estimation errors associated with high-dimensional parameter spaces. Additionally, this method enhances the interpretability of the model by revealing the hidden structures of the observed time series as its seeks linear transformations to simplify the dynamic structure of the observed series $y_t$.

Critically, the effectiveness of SCM hinges on the initial decomposition of the time series, which must preserve the essential dynamics among the variables. Incorrect or suboptimal decomposition can lead to misleading conclusions and poor model performance. Therefore, careful consideration and robust testing of the decomposition strategy are imperative.

In practice, the application of SCM has been demonstrated in various studies, showing improved accuracy and efficiency in model estimation compared to traditional methods. This is particularly evident in cases where the time series data exhibit complex interdependencies and when the dimensionality of the dataset is high.

Based on the currently available methods for finding SCMs and Kronecker indices of $y_t$, both methods can be carried out by canonical correlation analysis of certain expanded vectors of $y_t$ and by asymptotic chi-square tests for checking the number of zero correlations. The method for finding Kronecker indices is faster to compute and requires fewer numbers of hypothesis testings. The method for finding SCMs is more computational intensive and requires more hypothesis testing, especially in sorting out the double counting problem. On the other hand, Kronecker indices specify a $\text{VARMA}_d(p, p)$ model for $y_t$ whereas SCMs identify a general $\text{VARMA}_d(p, q)$ model for $y_t$. Part of the intensive computation of the approach is devoted to the separation of the AR and MA orders.

We remark that the methods for finding Kronecker indices and SCMs are available in the R package MTS of Tsay and Wood (2022). In addition, the structural specification of the VARMA model given a set of Kronecker indices or a set of SCMs is also available there.

## 3.3 | Recent Advances

Due to the limitations of the Echelon- and SCM-form, in particular in high-dimensional VARMA modeling, recent advances suggest new approaches for model identification.

Dufour and Pelletier (2022) propose new identified VARMA representations, the *diagonal MA equation form* and the *final MA equation form*, where the MA operators are respectively diagonal and scalar elements. These two formulations simply extend the traditional VAR model class by incorporating a basic MA operator, which may be either diagonal or scalar. Adding an MA component can lead to more parsimonious representations while maintaining simplicity and avoiding unnecessary complexity.

Wilms et al. (2023) address the identifiability issue for high-dimensional VARMA models by proposing an automatic identification of parsimonious VARMA models. The idea is to find a "simple" element in the equivalence set $\mathscr{E}_{p,q}$ in (2) of all AR-MA matrices by identifying such a parsimonious element in an intuitive yet objective fashion—using a suitable convex penalty—that results in an optimization-based identification procedure. Earlier work on parameter reduction in VARMA models (i.e., identification of non-zero elements in the AR and MA parameter matrices) dates back to Koreisha and Pukkila (1987).

## 4 | Estimation, Specification, and Diagnosis

In Section 4.1 we review popular estimation methods for an identified VARMA model with fixed AR and MA order. We thus assume the identifiability conditions associated to the VARMA to hold. In particular, the identification procedure (e.g., Kronecker index approach or SCM) results in a specified maximum order of the AR and MA polynomials together with an identified set of parameters that are constrained to zero and parameters that are non-zero, hence estimable. One can then proceed with standard estimation procedures, discussed below, that apply these zero constraints during estimation to ensure a well-defined likelihood function and to avoid overparametrization. Section 4.2 subsequently considers integral approaches towards estimation and specification of VARMA models, thereby focusing on the problem of determining the AR and MA orders. Section 4.3 reviews diagnosis tests to investigate the adequacy of estimated VARMAs.

### 4.1 | Estimation

#### 4.1.1 | Maximum Likelihood-Based Estimation

In early works, the most commonly used estimation method for identifiable VARMA models (with fixed AR and MA orders) is maximum likelihood. The Gaussian log-likelihood of the VARMA (1) takes on the form

$$\ell\left(\Phi, \Theta, \Sigma_a\right) = \ell_p - \frac{(T - p)}{2} \ln |\Sigma_a| - \frac{1}{2} \text{trace} \sum_{t=p+1}^{T} \left(\left(\Sigma_a\right)^{-1} a_t a_t^\top\right),$$

where $\ell_p = \ell\left(y_1, y_2, \ldots, y_p, a_{p-q+1}, a_{p-q+2}, \ldots, a_p\right)$ captures the contribution to the log-likelihood of the starting values of the

response and the error term; see, for instance, Reinsel (1993) for a textbook discussion. Over the years, different proposals have been made regarding exact, approximate, and conditional maximum likelihood estimation.

For a general class of linear multivariate models including VARMA, Dunsmuir and Hannan (1976) consider approximate likelihood procedures and establishes the strong law of large numbers and the central limit theorem for estimators of the parameters in such models; see Deistler, Dunsmuir, and Hannan (1978) for a generalization and corrections. Kohn (1979) consider Gaussian likelihood procedures for general linear multivariate time series models and establish the strong consistency and asymptotic normality of the parameter estimates.

For (stable) VARMA models specifically, early work on maximum likelihood estimation dates back to Akaike (1973). Tunnicliffe Wilson (1973) starts from the Gaussian likelihood of the VARMA and consecutively alternates between estimating the AR and MA parameters on the one hand and the error covariance matrix on the other hand.

Nicholls (1976) proposes spectral techniques to estimate VARMA models with exogenous variables, and Nicholls (1977) subsequently shows that the obtained estimates are identical to those obtained by applying Newton–Raphson to the approximate likelihood function; thereby establishing the consistency, asymptotic normality and efficiency of the estimator. Nicholls and Hall (1979) derive the exact likelihood function of a stationary vector process generated by a VARMA by writing it as a function of the observed data and backcasted values of pre-sample innovations. Hall and Nicholls (1980) then propose an algorithm for the evaluation of the derived exact likelihood whereas Gallego (2009) offers an improved version of it oriented towards nonlinear least squares estimation.

Rissanen and Caines (1979) consider multivariate Gaussian stationary vector time series following a VARMA and establish the strong consistency of the parameter estimates obtained with maximum likelihood. Reinsel (1979) considers full information maximum likelihood estimation for dynamic simultaneous equation models with VARMA errors. Hillmer and Tiao (1979) propose Gaussian approximate likelihood procedures for VARMA without relying on the invertibility condition (as commonly maintained in earlier work). Hannan, Dunsmuir, and Deistler (1980) derive the asymptotic properties of maximum likelihood estimates in VARMA models with exogenous variables under general conditions.

Ansley (1980); Kohn and Ansley (1982) provide expressions for the theoretical autocovariances of VARMA processes. Mittnik (1990) proposes an efficient procedure for computing autocovariance sequences of VARMA models in order to reduce the computational burden of exact maximum likelihood estimation; see Mittnik (1993) for an computational extension particularly suited for models with high order AR components and/ or a large number of variables and McElroy (2017) for a detailed discussion on the algorithmic implementation.

Mauricio (1995) focuses on computational techniques for maximizing the exact likelihood of VARMA models, as opposed to earlier studies that focus on evaluating the likelihood but oftentimes resort to standard optimization algorithms to maximize it. Mauricio (1997, 2002) provides details on the corresponding algorithmic implementation and Jonasson and Ferrando (2008); Jonasson (2008) handles the extension to VARMA models with missing data based on a Cholesky decomposition method and Gallego (2009) provides a simplified version of the Mauricio (1995) algorithm oriented towards maximum likelihood estimation. Kharrati-Kopaei, Nematollahi, and Shishebor (2009) consider a likelihood-based approach to find an approximate sufficient statistics for the VARMA model in terms of the periodogram. Finally, heavy-tailed VARMA models are studied in She, Mi, and Ling (2022) who offer Whittle estimation for VARMA models with heavy-tailed noises.

Note that many of the proposals above consider the exact likelihood of VARMA models, but maximizing it is computationally burdensome. Tiao and Box (1981) stress that the maximization of a conditional likelihood is much easier, other alternatives are discussed in detail below.

### 4.1.2 | State-Space Representations and Kalman Filter

Consider the VARMA model (1) in state-space form, thereby following the notation in Metaxoglou and Smith (2007),

$$
\begin{aligned}
y_t &= \Phi x_t + Z w_t + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_\epsilon) \\
w_t &= S w_{t-1} + \eta_t, \quad\quad\quad \eta_t \sim N(0, \Sigma_\eta),
\end{aligned}
\tag{18}
$$

where $x_t^\top = \left[ y_{t-1}^\top, y_{t-2}^\top, \ldots, y_{t-p}^\top \right]^\top$, $w_t^\top = \left[ v_t^\top, \ldots, v_{t-q}^\top \right]^\top$, $\eta_t^\top = \left[ v_t^\top, 0, \ldots, 0 \right]^\top$ for $\Theta(L) a_t = \Gamma(L) v_t + \epsilon_t$ with $v_t$ and $\epsilon_t$ white noise processes such that $v_t$ and its lags can be treated as observable in the complete-data log-likelihood. Furthermore, $\Phi = \left[ I \ \Phi_1 \ \cdots \ \Phi_p \right]$ and $Z = \left[ I \ \Gamma_1 \ \cdots \ \Gamma_q \right]$ and

$$
S = \begin{bmatrix} 0 & 0 \\ I_{dq} & 0 \end{bmatrix} \quad \text{and} \quad \Sigma_\eta = \begin{bmatrix} \Sigma_v & 0 \\ 0 & 0 \end{bmatrix}.
$$

Early work on VARMA models in state space form dates back to Ansley and Kohn (1983), Solo (1984), Deistler (1985) where the usage of the Kalman filter is proposed to compute its exact Gaussian likelihood thereby allowing for missing data. Shea (1989) offers details on the algorithmic implementation of the former and Shea (1987, 1988) provides a detailed assessment on the choice of initial estimates. Zadrozny (1989, 1992) presents algorithms to compute the exact Gaussian likelihood of discrete time, linear dynamic models in state space form that also encompass VARMA. Metaxoglou and Smith (2007) focuses on likelihood maximization and proposes maximum likelihood estimation of VARMA models in state space representation (18) using the EM algorithm; the Kalman filter also facilitates back-casting to account for the pre-sample values of the AR component which are treated as missing data.

Klein and Mélard (2014) derive the exact Fisher Information Matrix (FIM)—crucial for describing the covariance structure of the maximum likelihood estimator—of multivariate

Gaussian time series models in state space form, thereby giving a detailed treatment for VARMA models; the properties of the FIM are further investigated in Klein and Spreij (2006); Klein, Mélard, and Saidi (2008); Mélard (2022), an algorithm for Mathematica is discussed in Klein and Mélard (2023). Bao and Hua (2014) propose a compact representation of the asymptotic Fisher information matrix that does not involve any integral.

### 4.1.3 | Least Squares-Based Estimation

Recursive linear regression methods have also been extensively considered for VARMA processes as they form an appealing alternative to maximum likelihood estimation from a computational point of view, especially since the latter is expensive to apply for large time series models. The general idea is to estimate, by least squares, the errors of the VARMA process from a high-order (i.e., $\widetilde{p}$ large) VAR given by

$$y_t = \sum_{\tau=1}^{\widetilde{p}} \Pi_\tau y_{t-\tau} + \varepsilon_t,$$

and to subsequently use these residuals $\widehat{\varepsilon}_t$ as regressors when estimating the (approximated) VARMA model

$$y_t = \sum_{\ell=1}^{p} \Phi_\ell y_{t-\ell} + \sum_{m=1}^{q} \Theta_m \widehat{\varepsilon}_{t-m} + u_t.$$

Spliid (1983) offers an early proposal of such a two-stage least squares based procedure, a similar proposal was made by Koreisha and Pukkila (1989). Poskitt (1992) considers VARMA models in Echelon canonical form and proposes a method for identification and estimation based on a sequence of least squares regressions. Poskitt and Salau (1994, 1995) subsequently discuss the relationship between the least squares and Gaussian estimation schemes and the asymptotic (in)efficiency of using least squares relative to Gaussian maximum likelihood to estimate the parameters of Echelon-form VARMA models, numerical methods for computing the asymptotic covariance matrix of the conditional maximum likelihood estimator and the least squares estimator are discussed in Salau (1997, 1999). Kascha (2012) provides a Monte Carlo comparison of maximum likelihood and least squares based estimation methods for VARMA models.

Reinsel, Basu, and Yap (1992) discusses a Gauss-Newton iterative procedure to obtain the maximum likelihood estimate of the VARMA parameters, which has a computational form in terms of generalized least squares estimation. De Frutos and Serrano (2002) also propose a generalized least squares estimation procedure for VARMA models that explicitly accounts for the stochastic nature of the approximation errors when the lagged errors are replaced by the lagged residuals of the high-order VAR model.

Dufour and Jouini (2005) consider a two-step least squares based estimator for the VARMA, in their follow-up work (Dufour and Jouini 2014), a generalized least squares version of the former and

a three-step linear estimator that is asymptotically equivalent, yet computationally more efficient, to the Gaussian maximum likelihood are introduced. Jouini (2015) develops practical and asymptotically valid methods for bootstrapping VARMA models using the simple linear estimation methods developed in Dufour and Jouini (2014). Dufour and Pelletier (2022) then consider a three-stage procedure where in addition to the commonly used two steps in the linear regression based approach, a third step is added where the data from the VARMA with approximated errors is filtered to obtain estimates with the same asymptotic covariance matrix as their nonlinear counterparts (i.e., the Gaussian maximum likelihood estimator).

Dias and Kapetanios (2018) propose an iterative, instead of two-step, least squares estimator for VARMA models in the spirit of Kapetanios (2003) and establish its consistency and asymptotic distribution. Finally, Wilms et al. (2023); Zheng (2024) consider high-dimensional time series and offer penalized regression based approaches to sparsely estimate high-dimensional VARMA and infinite-order VAR models respectively.

### 4.1.4 | Bayesian Estimation

Bayesian estimation contributions to VARMA modeling remain rather scarce. Shaarawy (1989) initiated the proposal of Bayesian solutions to the problems of estimation of and forecasting with VARMA models. Albassam, Soliman, and Ali (2023) conduct a wide simulation study to investigate the effectiveness of this proposal.

Ravishanker and Ray (1997) consider Bayesian estimation of VARMA models using Metropolis Hastings to obtain samples from the joint posterior density of the VARMA parameters based on the exact Gaussian likelihood, the VARMA model is identified using Bayesian variable selection techniques.

Li and Tsay (1998) offer a Bayesian procedure for simultaneous identification (via Kronecker indices) and estimation of VARMA models; their proposal uses stochastic search variable selection priors and can handle cointegrated as well as noninvertible systems. Chan, Eisenstat, and Koop (2016) also offer a unified approach to identification and estimation of Echelon form VARMA models through the usage of a hierarchical prior that permits joint selection of identification restrictions and shrinkage in the resulting model to accommodate high-dimensional settings; they offer an efficient Markov chain Monte Carlo algorithm to this end.

Roy, McElroy, and Linton (2019) consider a reparametrized VARMA model to permit parameter estimation under the constraints of causality and invertibility which facilitates the computation of Bayesian estimates via a prior specification on the constrained space (as well as maximum likelihood estimation). Lastly, Shaarawy (2023) provide a Bayesian methodology based on the conditional likelihood to unify the four stages of model identification, estimation, diagnostic checking, and forecasting.

## 4.2 | Specification

While the approaches discussed in the previous section mainly consider parameter estimation for a given VARMA, we now review the problem of specifying the VARMA model with respect to its AR and MA order. Related problems of such model building steps for VARMA models date back to the early work of Akaike (1976); Chan and Wallis (1978); Jenkins and Alavi (1981).

In the seminal paper by Hannan and Kavalieris (1984) on VARMA models, a regression-based approach is used for estimating the VARMA parameters in the context of determining the AR and MA orders. The first three steps of the procedure focus on specifying the VARMA model by choosing the AR and MA orders through an information criterion and providing initial estimates. The final stage uses generalized least squares regression to obtain asymptotically efficient estimates.

Tiao and Box (1981), on the other hand, offer an iterative procedure for building VARMA models which consists of three stages (i) specification, (ii) estimation, and (iii) diagnostic checking where the use of cross correlations and partial autoregressions is advocated to tentatively specify the VARMA orders in the first stage. Tiao and Tsay (1983) provide a subsequent discussion on the first model specification stage where an extended sample cross-correlation procedure, that extends the proposal of Tsay and Tiao (1984) for univariate ARMA models, is advocated. Tiao and Tsay (1989) then turn to model specification for VARMA using SCM where canonical correlation analysis is used to determine the orders of the AR and MA polynomials. The value of canonical correlation analysis for time series analysis in general and VARMA model specification in particular has been discussed by, among others, Akaike (1976); Box and Tiao (1977); Cooper and Wood (1982); Tsay and Tiao (1985); Peña and Box (1987); Tsay (1989); Toscano and Reisen (2000).

To make VARMA models more accessible for practitioners and to promote their use over VARs, Lütkepohl and Poskitt (1996) offer a general strategy for specifying VARMA models in Echelon form which consists of choosing a set of Kronecker indices. Koreisha and Pukkila (2004) propose to select the VARMA orders based on the residual white noise autoregressive criterion of Pukkila, Koreisha, and Kallinen (1990). Boubacar Maïnassara (2012) consider model specification based on a modified Akaike information criterion for weak VARMA models where the errors are uncorrelated but not necessarily independent. Dufour and Pelletier (2022) also consider weak VARMA processes and develop practical methods for identifying, specifying, and estimating such processes in diagonal MA equation form. To specify the VARMA orders, an information criterion is used that yields consistent estimates of the AR and MA orders. Kathari and Tangirala (2020) use a pre-estimation approach based on scalar (inverse) autocorrelation functions to specify the orders across a variety of multivariate time series models including VARMA. Wilms et al. (2023) use penalized regression methods to simultaneously identify, specify and estimate the VARMA model.

## 4.3 | Diagnosis

After specifying and estimating the VARMA model, it is good practice to continue with various diagnostic checks to evaluate the adequacy of the fitted VARMA. In practice, the specification and estimation can be carried out via the methods discussed in the previous sections resulting in different representations. The choice of methods often depends on the goal of data analysis. The suitability of the chosen representation can always be evaluated by the diagnostic procedures discussed below. In this section, we review some of the adequacy tests that are specifically proposed for VARMA models.

Hosking (1980) proposes a Portmanteau goodness-of-fit for the VARMA and subsequently shows that it can be obtained as a Lagrange-multiplier test (Hosking 1981). Li and McLeod (1981) obtain the large-sample distribution of the multivariate residual autocorrelations in VARMA models and offer a Portmanteau test based on it. Hallin and Liu (2023) recently revisited the tests of Hosking (1980) and Li and McLeod (1981) and propose a class of rank- and sign-based Portmanteau tests for a broad family of error distributions. Mahdi and Ian McLeod (2012) extend the univariate Portmanteau test of Peña and Rodriguez (2002) to VARMA models.

Arbués (2008) considers a Portmanteau test for constrained VARMA models where the whole system (including the error covariance matrix) is constrained to a certain class of models, whereas Boubacar Maïnassara and Francq (2011); Katayama (2012) offer Portmanteau tests for structural VARMA models (see Section 5.3), Boubacar Maïnassara and Saussereau (2018) for VARMA models with uncorrelated but nonindependent errors and Ilmi Amir and Boubacar Maïnassara (2020) for seasonal VARMA models (see Section 6.2).

Hallin and Paindaveine (2004) derive an optimal rank-based test for verifying the adequacy of elliptical VARMA models, Hallin and Paindaveine (2005) consider optimal rank-based procedures for affine-invariant linear hypothesis testing in multivariate general linear models with elliptical VARMA errors.

Paparoditis (2005) offers a goodness-of-fit-test for VARMA models that can be applied when no a priori information exists on expected departures from the null that the observed process follows a VARMA with fixed AR and MA orders, this in contrast to earlier work by Kohn (1979), Hosking (1981) and Poskitt and Tremayne (1982) who consider testing a VARMA model against a higher order VARMA alternative. Velilla and Thu (2018); Velilla and Nguyen (2019) offer techniques for testing the adequacy of VARMA models where the goodness-of-fit process is shown to converge to the Brownian bridge.

## 5 | Usage of VARMA Models

We review the main usage of VARMA models to test Granger causality relations (Section 5.1), to conduct forecasting tasks (Section 5.2) and to perform structural analysis (Section 5.3).

## 5.1 | Granger Causality

Granger causality captures, intuitively speaking, the incremental predictability of one variable for another given a particular information set Granger (1969, 1980). While for pure VAR and VMA models, sufficient and necessary conditions for the hypothesis that "x does not cause z" can be directly related to the nullity of the corresponding (respectively) AR and MA parameters in the model, the same is not true for VARMA models.

To this end, consider two multivariate stationary stochastic processes $z_t$ and $x_t$ and let their joint VARMA representation be given by

$$
\begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \Theta_{11}(L) & \Theta_{12}(L) \\ \Theta_{21}(L) & \Theta_{22}(L) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.
$$

While the joint nullity of $\Phi_{12}(L)$ and $\Theta_{12}(L)$ is a sufficient condition for Granger non-causality from $x$ to $z$ it is not a necessary condition. Indeed, to this end, consider the pure VMA representation of the VARMA as given by

$$
\begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.
$$

Then $x$ is not Granger causal for $z$ if and only if $\Psi_{12}(L) = 0$; see Lütkepohl (2005) for a textbook introduction. Typically a set of non-linear restrictions—as opposed to linear restrictions for VAR and VMA representations—is required to characterize Granger causality in VARMA models.

Granger causality in the context of bivariate VARMA models dates back to Kang (1981); Eberts and Steece (1984); Newbold and Hotopp (1986); Taylor (1989), whereas Osborn (1984); Boudjellaba, Dufour, and Roy (1991, 1994) consider Granger causality in VARMA models beyond the bivariate case; see also James, Koreisha, and Partch (1985), Hundley and Koreisha (1987) and Das (2003) for economic applications on Granger causal relations using VARMA models. Dufour and Renault (1998); Dufour and Taamouti (2010) consider a wide class of dynamic models including VARMA and derive general parametric and nonparametric characterizations of non-causality at various horizons. Himdi and Roy (1997); Hallin and Saidi (2005) generalize the procedure by Haugh (1976) for univariate time series to test the hypothesis of non-correlation between two multivariate stationary ARMA processes and discuss how their test can be adapted to determine the direction of Granger causality.

## 5.2 | Forecasting

VARMA models are powerful tools for jointly forecasting a set of time series variables. We review both theoretical and practical work that focuses on forecasting with VARMA models.

Theoretical work on forecasting dates back to Yamamoto (1980, 1981) who derive the optimal prediction scheme for multiperiod predictions with VARMA models, while Hung and Alt (1994) offer an approximation of the one-step ahead forecast error covariance of VARMA models.

Aksu and Narayan (1991); Grillenzoni (1991) offer a theoretical and practical perspective on forecasting with VARMA models, the former thereby adopt the MTS software package, nowadays available via the package MTS (Tsay and Wood 2022) for the software environment R (R Core Team 2024). Reinsel (1995) considers the traditional estimation procedure based on the exact likelihood function and establishes general results on exact finite sample forecasts and their mean squared errors.

Oke and Öller (1999) offer a short-memory test for VARMA models, to help distinguish whether a series cannot be predicted from the past (i.e., "no" memory), is partially predictable in the future (i.e., "short" memory) or can be predicted far or indefinitely into the future (i.e., "long" memory).

Lütkepohl (2006) provides a general exposition on forecasting with VARMA models in Echelon form in the presence of stationary and cointegrated variables, thereby paying special attention to forecasting issues related to VARMA processes under contemporaneous and temporal aggregation.

Peña and Sánchez (2007) provide insight into the advantages of using a dynamic multivariate forecast models, such as a VARMA, over univariate ones, thereby offering an a priori measure for the increase in precision to be attained by the multivariate approach over the univariate one. Anthanasopoulos and Vahid (2008) compare VARMA to VAR models for macroeconomic forecasting and conclude that there is no compelling reason for restricting the model class to VARs since VARMAs forecast more accurately on the various macroeconomic data sets they considered.

More recently in the field of statistics and computer science, forecasting with VARMA models attracted attention: Guo, Liu, and Sun (2016) propose a hybrid combination of VARMA models and Bayesian networks to improve the forecasting performance of multivariate time series, Yang et al. (2018) present an online time series prediction framework for VARMA models and Isufi et al. (2019) offer VAR and VARMA models for forecasting the temporal evolution of time series on graphs, Shi and Sheng (2024) propose uncertain vector autoregressive smoothly moving average models to consider forecasting under imprecise observations.

## 5.3 | Structural Analysis

Structural VARMA, in short SVARMA, models extend the VARMA framework by incorporating structural information, which allows for the identification of causal relationships among variables. The "structural" aspect refers to the imposition of theoretically informed restrictions on the model, which are often based on economic theory or prior empirical findings. These restrictions enable the disentanglement of shock transmission mechanisms within the system, offering insights into how exogenous shocks to one variable can propagate through and impact other variables in the model.

To be more precise, consider the representation

$$\Phi(L)y_t = \Theta(L)\varepsilon_t,$$

where the error terms $\varepsilon_t$ represent structural shocks, which are unobservable innovations that have a direct interpretation within the context of, for instance, economic theory being studied. These shocks are assumed to be uncorrelated with each other and often have a direct economic meaning, such as supply shocks, demand shocks, policy shocks.

The concept of SVARMA was first introduced in Angulo (1999) to better understand money supply processes. Ever since it has been a popular tool in the macroeconomic literature. Focusing here on the theoretical aspects of SVARMA, Boubacar Maïnassara and Francq (2011) study the consistency and the asymptotic normality of the quasi MLE for a structural model.

Boubacar Maïnassara (2011); Katayama (2012) subsequently introduce hypothesis tests for SVARMA models to discover their adequacy. In Gouriéroux, Monfort, and Renne (2020), non-Gaussian strong SVARMA models are identified. Strong SVARMA refers to cross-sectional correlation in the $\varepsilon_t$. Gouriéroux, Monfort, and Renne (2020) further propose parametric and semi-parametric estimation methods to consistently estimate possibly non-fundamental representation in the moving average dynamics.

For SVARMA models driven by independent and non-Gaussian shocks, Funovits (2024) discusses parameterization, identifiability, and maximum likelihood (ML) estimation. More recently, Velasco (2023) suggests a frequency domain criterion for identification based on a new representation of the higher order spectral density arrays of vector linear processes.

# 6 | Extensions of VARMA Models

VARs nowadays still dominate VARMAs especially so in the development of flexible extensions of the basic VAR. Nonetheless, also for VARMA, a wide variety of useful extensions have been proposed over the years. We review a collection of most widely adopted extensions in this section.

## 6.1 | Cointegrated VARMA

Cointegrated VARMA models extend traditional VARMA models by incorporating cointegration, a statistical property indicating that a linear combination of nonstationary variables is stationary. This integration allows the models to capture both short-term dynamics and long-term relationships among variables, making them particularly suitable for analyzing economic and financial time series that exhibit long-run equilibrium relationships.

The first work that extended the basic ideas of cointegration from VAR (see Sections 8.1–8.2. in Lütkepohl 2005) to VARMA models goes back to Yap and Reinsel (1995). Yap and Reinsel (1995) introduce a vector error correction form (VEC) for VARMA models, given by

$$\Phi^*(I_d - L)y_t = Cy_{t-1} + \Theta(L)a_t,$$

where $C$ has reduced rank. The VEC concentrates the nonstationarity of the AR operator in the behavior of the coefficient matrix $C$. Yap and Reinsel (1995) derive the asymptotic properties of the full-rank and reduced-rank Gaussian estimators. These results are utilized to derive the asymptotic distribution of the likelihood ratio statistic and for testing the number of unit roots.

Estimating cointegrated VARMA models involves several steps, including determining the rank of cointegration, identifying the cointegration space, and estimating the parameters of the model. The Echelon form and other identification constraints play a crucial role in simplifying these processes, ensuring the model is both identifiable and estimable.

One of the main challenges in cointegrated VARMA modeling is the computational complexity and the difficulty in model specification and selection. Recent advances involve developing more efficient estimation techniques and software implementations, as well as extending the models to handle issues like structural breaks and nonlinearities.

Later Lütkepohl and Claessen (1997) combine the general VEC model for VARMA models with the Echolon form. In a subsequent work, Bartel and Lütkepohl (1998) discuss the estimation of the corresponding Kronecker indices to derive the Echelon form for VEC. Other extensions of the Echelon methodology for cointegrated VARMA can be found in Poskitt (2003, 2006); see also Lütkepohl (2005, Chapter 14). More recently, Mélard, Roy, and Saidi (2006) evaluate the exact likelihood function of Gaussian, nonstationary VARMA models in VEC form. Cubadda, Hecq, and Palm (2009) study some implications of cointegration on the univariate time series. Finally, Guo and Ling (2024) propose full and reduced rank least squares estimators of heavy-tailed and partially nonstationary VARMA models.

While the Echelon methodology has been extended to cointegrated VARMA models, similar extensions of the scalar-components methodology are not currently available.

## 6.2 | Seasonal VARMA

Many time series contain a seasonal component that repeats itself after a regular period of time. To capture the seasonal component, one can resort to seasonal VARMA models. The seasonal VARMA model is given by

$$\Phi(L)\widetilde{\Phi}(L^s)y_t = \Theta(L)\widetilde{\Theta}(L^s)a_t,$$

where $s$ and the seasonal matrix polynomials are given by

$$\begin{aligned}\widetilde{\Phi}(L^s) &= I - \widetilde{\Phi}_1 L^s - \widetilde{\Phi}_2 L^{2s} - \dots - \widetilde{\Phi}_P L^{Ps} \\ \text{and} \quad \widetilde{\Theta}(L^s) &= I + \widetilde{\Theta}_1 L^s + \widetilde{\Theta}_2 L^{2s} + \dots + \widetilde{\Theta}_Q L^{Qs}.\end{aligned} \tag{19}$$

The seasonal period $s$ is typically known a priori, for instance 4 for quarterly data or 12 for monthly data. Note that unlike

for seasonal ARMA representations, seasonal VARMA representations are not unique. Indeed, a different representation is obtained when swapping the standard and seasonal lag polynomials in equation (19) due to the non-uniqueness of the matrix polynomials; see Yozgatligil and Wei (2009) for such different representations of seasonal VARMA models.

McElroy (2022) offers a frequency domain-based approach to compute the autocovariances from the parameters in the SVARMA which may then be used to estimate the SVARMA via maximum likelihood or to forecast from a VARMA model.

## 6.3 | FAVARMA

Dufour and Stevanović (2013) study the relationship between VARMA and factor representations of a vector stochastic process and find that multivariate times series and their factors cannot—in general—both follow finite order VAR processes. In fact, VAR factor dynamics induce a VARMA process, while a VAR process entails VARMA factors. The authors therefore propose to combine factor and VARMA modeling using a parsimonious Factor Augmented VARMA (FAVARMA) representation to represent dynamic interactions between a large collection of time series.

The FAVARMA for the $d$-dimensional stationary stochastic process $y_t$ and $r$ factors is given by

$$y_{it} = \lambda_i(L)f_t + u_{it}$$
$$u_{it} = \delta_i(L)u_{i,t-1} + v_{it}$$
$$\Phi(L)f_t = \Theta(L)\eta_t$$
$$i = 1, \ldots, d \quad t = 1, \ldots, T,$$

where $\lambda_i(L)$ is an $r$-dimensional vector of lag polynomials $\lambda_i(L) = \big(\lambda_{i1}(L), \ldots, \lambda_{ir}(L)\big)$ with $\lambda_{ij}(L) = \sum_{k=0}^{p_{i,j}} \lambda_{i,j,k}L^k$, $\delta_i(L)$ is a $p_{y,i}$-degree lag polynomial, $\Phi(L)$ and $\Theta(L)$ are the usual AR and MA polynomials in a VARMA representation and $v_{it}$ is $d$-dimensional white noise that is uncorrelated with the $r$-dimensional white noise process $\eta_t$. Dufour and Stevanović (2013) and Zadrozny and Chen (2019) illustrate the good forecast performance of FAVARMA for macroeconomic forecasting.

## 6.4 | VARMA-GARCH

The VARMA-GARCH model is designed to capture the dynamics of multivariate time series data, specifically addressing both mean and volatility fluctuations. This model integrates VARMA approach with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process, which effectively models time-varying volatility. The model was introduced in Ling and McAleer (2003) and has been employed extensively in financial econometrics.

Following the representation in Ling and McAleer (2003), the model can be written as

$$\Phi(L)y_t = \Theta(L)a_t, \quad \text{with} \quad a_t = D_t\eta_t,$$
$$H_t = W + \sum_{l=1}^{r} A_l\zeta_{t-l} + \sum_{l=1}^{s} B_lH_{t-l}, \tag{20}$$

where $H_t = \big(h_{1,t}, \ldots, h_{d,t}\big)$, $D_t^2 = \text{diag}\big(h_{1,t}, \ldots, h_{d,t}\big)$, $\zeta_t = \big(\varepsilon_{1,t}^2, \ldots, \varepsilon_{d,t}^2\big)^\top$. Ling and McAleer (2003) establish the structural and statistical properties, including the sufficient conditions for the existence of moments and the sufficient conditions for consistency and asymptotic normality of the QMLE for model (20).

McAleer et al. (2008) generalize model (20) towards letting the standardized residuals follow a random coefficient VAR process to allow for dynamic conditional correlations. McAleer, Hoti, and Chan (2009) develop structural and statistical properties of the model. Wang and Tsay (2013) consider diagnostic checking of VARMA-GARCH models with Gaussian and Student-$t$ innovations.

The VARMA-GARCH model has found its way into numerous fields and has been used in financial econometrics to study velocity and variability of money growth (Serletis and Shahmoradi 2006), to analyze the oil market (Rahman and Serletis 2012; Serletis and Xu 2018), water quality (Wu, Kuo, and Liu 2012), dynamic spillovers between stock and money markets (Salisu, Isah, and Assandri 2019), examine investment opportunities (Do, Bhatti, and Shahbaz 2020), relationships among air pollutants and how their concentration changed (Wu and Kuo 2020).

## 6.5 | Nonstationary VARMA

Throughout the literature, one can find several attempts to lift the assumption of stationarity in VARMA models. For instance by introducing a *thresholded* VARMA model, allowing for *change-points*, *time varying* coefficient matrices or *Markov Switching* models.

### 6.5.1 | Threshold VARMA

Introduced in Niglio and Vitale (2015), the threshold VARMA (TVARMA) model is a type of time series model that incorporates regime-switching based on the value of an observable variable, typically a lagged value of the time series itself. The regimes switch when this variable crosses certain thresholds, that is,

$$y_t = \Phi^{(j)}(L)y_t + \Theta^{(j)}(L)a_t + \varepsilon_t^{(j)}, \quad \text{if} \quad \tau_{j-1} < z_t \leq \tau_j.$$

### 6.5.2 | Change-Point Detection

Galeano and Peña (2007) studies step changes in the variance and in the correlation structure modeled through

$$\Phi(L)y_t = \Theta(L)e_t \quad \text{with} \quad e_t = a_t + WS_t^{(h)}a_t,$$

where $S_t^{(h)} = 1_{\{t \geq h\}}$ is a step function creating a change at $t = h$ from $e_t$ having covariance $\Sigma$ pre break and $\Omega = (I + W)\Sigma(I + W)^\top$ post break under suitable assumptions on $W$. Galeano and Peña (2007) introduce two approaches using a likelihood ratio approach and a CUSUM type approach.

Later, Steland ([2020]) attempts to address similar questions in a high-dimensional regime, letting the dimension grow with the sample size. Their approach uses bilinear forms of the centered or non-centered sample variance–covariance matrices. Change-point testing and estimation are based on maximally selected weighted CUSUM statistics. Large sample approximations under a change-point regime are provided including a multivariate CUSUM transform of increasing dimension.

Golosnoy and Seifert ([2021]) introduce a framework for sequentially (online) monitoring changes in the mean vector of high-dimensional persistent VARMA time series by using multivariate control charts.

### 6.5.3 | Time Varying VARMA

One of the first works to consider time varying VARMA models, meaning that the coefficient matrices are allowed to vary over time, is Hallin ([1978]). Hallin ([1978]) derives conditions for time varying VARMA models to be purely nondeterministic and invertible. Shelton Peiris and Thavaneswaran ([2001]) generalize the results in Hallin ([1978]) by allowing the innovations to a general class of stable distributions instead of imposing Gaussianity. In another early work, Zadrozny and Mittnik ([1994]) consider a recursive Kalman-filtering method for computing exact sample and asymptotic information matrices for time-invariant, periodic, or time-varying Gaussian VARMA models.

In an empirical work, Chan and Eisenstat ([2017]) study different types of time varying VARMA models and address computational challenges associated with VARMA estimation through a Bayesian approach developing a Gibbs sampler. Their considered extensions of the classical VARMA model, allow for time-varying vector moving average coefficients and stochastic volatility.

More recently, maximum likelihood estimation for time varying VARMA models has been studied, see Mélard ([2024]) for a recent discussion. Alj, Jónasson, and Mélard ([2016]) proposes an algorithm for the evaluation of the exact Gaussian likelihood including a time dependent innovation covariance matrix. Subsequently, the author study quasi-maximum likelihood estimators in Alj et al. ([2017]); Alj, Azrak, and Mélard ([2025]). Mélard ([2022]) prove strong consistency and asymptotic normality of a Gaussian quasi-maximum likelihood estimator for the parameters of a causal, invertible, and identifiable vector autoregressive moving average.

### 6.5.4 | Markovian VARMA

Multivariate Markov-switching autoregressive moving-average (MS-ARMA) models incorporate regime-switching elements into multivariate ARMA models, allowing the model parameters to change depending on the state of a Markov process. These models are especially useful in capturing the behavior of time series that exhibit changes in regime or state, such as shifts in economic conditions.

To be more precise, one typically writes

$$y_t = \mu_{S_t} + \Phi_{S_t}(L)y_t + \Theta_{S_t}(L)a_t$$

such that the model parameters depend on the state of an unobserved Markov chain $(S_t)$ with finite state-space.

A natural idea when estimating these models is to impose local stationarity conditions, that is, stationarity within each regime. Francq and Zakoïan ([2001]) show that local stationarity of the observed process is neither sufficient nor necessary to obtain global stationarity. Another observation, first made in Francq and Zakoïan ([2001]) is that the autocovariance structure coincides with that of a standard ARMA. Later, Zhang and Stine ([2001]) show that the autocovariance structure of a model belonging to a general class of second order stationary Markov regime switching processes coincides with that of a VARMA whose orders are bounded above by functions of the number of Markov regimes. Cavicchioli ([2016]) improve their bound on the VARMA orders. Stelzer ([2009]) introduce stationarity and ergodicity conditions as well as an easy-to-check sufficient stationarity condition based on a tailor-made norm. Cavicchioli ([2017b]) propose conditions for higher-order stationarity.

In another line of research that also aims to find stationarity conditions, several authors use a spectral domain perspective. Pataracchia ([2011]) propose a method to derive the spectral density function of Markov switching ARMA model by applying the Riesz–Fischer theorem which defines the spectral representation as the Fourier transform of the autocovariance functions. Cavicchioli ([2013]) derive a formula in closed form for the spectral density of MS-VARMA models and describe some of its properties.

In economics and finance, in particular, shocks are often regarded as being heavily tailed and a straightforward way to include this feature into MS-ARMA models is to use a regularly varying and thus heavy-tailed noise sequence. Stelzer ([2008]) show that heavy tailed noise implies that under appropriate summability conditions, the MS-ARMA process is again heavy tailed as a sequence.

Cavicchioli studied how to determine the number of regimes in a MS-VARMA model. Cavicchioli ([2014]) propose a stable finite-order VARMA representations for M-state Markov switching second-order stationary time series under suitable conditions on the autocovariances.

In a different line of research, Cavicchioli contributed to understanding the asymptotic and exact Fisher information matrices of MS-VARMA models; see Cavicchioli ([2017a], [2020]). In particular, the explicit representation to derive the asymptotic covariance matrix of the Gaussian maximum likelihood estimator of the parameters in the MS-VARMA model.

## 7 | Conclusion and Outlook

We explored various aspects of VARMA models, highlighting their key role in multivariate time series analysis. The application of VARMA models spans numerous fields, such as economics, finance, environmental studies, and more, offering a robust framework for understanding system dynamics through the interdependencies among multiple time series. Throughout the review, we have dissected the methodological advancements

that have enhanced the accuracy and efficiency of these models. Despite these advancements, the literature still presents a number of unresolved issues and challenges.

One of the notable gaps in the current VARMA literature is the complexity involved in model specification, particularly in selecting appropriate order parameters $(p, q)$. The model selection process is crucial as it significantly affects the model's performance, but it remains largely heuristic and computationally intensive. Future research could focus on developing more automated and data-driven techniques for determining the model parameters, potentially leveraging advancements in machine learning. Additionally, there is a need for more robust methods to handle model estimation in the presence of missing data, outliers, and high-dimensional data. Furthermore, the integration of VARMA models with other data types and sources remains an underexplored area. As data becomes increasingly multidimensional and heterogeneous, integrating diverse data types such as high-frequency time series, spatial data, network data, or tensor-valued data into the VARMA framework could open new avenues for multidisciplinary research and application.

In terms of software implementations, VARMA models are notably less represented compared to VAR models, particularly in widely used statistical programming environments such as R. While R offers comprehensive packages for VAR modeling, such as the "vars" package (Pfaff 2008) which provides extensive functionalities for estimation, diagnostics, forecasting, and causality analysis, the resources for VARMA are comparatively limited. The MTS package (Tsay and Wood 2022) is one of the few that supports VARMA and related models, yet it does not provide as rich a feature set as those available for VAR, especially in areas like model diagnostics and interactive model selection tools. The bigtime package (Wilms et al. 2021) is a more recent attempt to make VARMA models more accessible addressing questions of model selection for high-dimensional time series. This disparity in software tools reflects the broader challenges associated with the computational complexity and parameter estimation difficulties inherent in VARMA models. Enhancing the software support for VARMA in R and other software languages, could significantly increase their accessibility and usability, encouraging more widespread adoption and innovation in the analysis of multivariate time series data.

### Author Contributions

**Marie-Christine Düker:** conceptualization (equal), methodology (equal), writing – original draft (equal), writing – review and editing (equal). **David S. Matteson:** conceptualization (equal), methodology (equal). **Ruey S. Tsay:** conceptualization (equal), methodology (equal), writing – original draft (equal), writing – review and editing (equal). **Ines Wilms:** conceptualization (equal), methodology (equal), writing – original draft (equal), writing – review and editing (equal).

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

### Related Wires Articles

[Autoregressive processes](#)

### References

Akaike, H. 1973. "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models." *Biometrika* 60, no. 2: 255–265.

Akaike, H. 1974. "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes." *Annals of the Institute of Statistical Mathematics* 26, no. 3: 363–387.

Akaike, H. 1976. "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion." In *Mathematics in Science and Engineering*, vol. 126, 27–96. Amsterdam, the Netherlands: Elsevier.

Aksu, C., and J. Narayan. 1991. "Forecasting With Vector ARMA and State Space Methods." *International Journal of Forecasting* 7, no. 1: 17–30.

Albassam, M., E. Soliman, and S. Ali. 2023. "An Effectiveness Study of the Bayesian Inference With Multivariate Autoregressive Moving Average Processes." *Communications in Statistics: Simulation and Computation* 52, no. 10: 4773–4788.

Alj, A., R. Azrak, C. Ley, and G. Mélard. 2017. "Asymptotic Properties of QML Estimators for VARMA Models With Time-Dependent Coefficients." *Scandinavian Journal of Statistics* 44, no. 3: 617–635.

Alj, A., R. Azrak, and G. Mélard. 2025. "General Estimation Results for tdVARMA Array Models." *Journal of Time Series Analysis* 46, no. 1: 137–151.

Alj, A., K. Jónasson, and G. Mélard. 2016. "The Exact Gaussian Likelihood Estimation of Time-Dependent VARMA Models." *Computational Statistics & Data Analysis* 100: 633–644.

Amendola, A., M. Niglio, and C. Vitale. 2010. "Temporal Aggregation and Closure of VARMA Models: Some New Results." In *Data Analysis and Classification: Proceedings of the 6th Conference of the Classification and Data Analysis Group of the SocietàItaliana di Statistica*, 435–443. New York, NY: Springer.

Angulo, V. 1999. "A Structural VARMA Approach to Modelling the Money Supply Process." *Applied Economics* 31, no. 2: 195–206.

Ansley, C. 1980. "Computation of the Theoretical Autocovariance Function for a Vector ARMA Process." *Journal of Statistical Computation and Simulation* 12, no. 1: 15–24.

Ansley, C., and R. Kohn. 1983. "Exact Likelihood of Vector Autoregressive-Moving Average Process With Missing or Aggregated Data." *Biometrika* 70, no. 1: 275–278.

Anthanasopoulos, G., and F. Vahid. 2008. "VARMA Versus VAR Macroeconomic Forecasting." *Journal of Business & Economic Statistics* 26, no. 2: 237–252.

Arbués, I. 2008. "An Extended Portmanteau Test for VARMA Models With Mixing Nonlinear Constraints." *Journal of Time Series Analysis* 29, no. 5: 741–761.

Athanasopoulos, G., D. Poskitt, and F. Vahid. 2012. "Two Canonical VARMA Forms: Scalar Component Models Vis-à-Vis the Echelon Form." *Econometric Reviews* 31, no. 1: 60–83.

Athanasopoulos, G., and F. Vahid. 2008. "A Complete VARMA Modelling Methodology Based on Scalar Components." *Journal of Time Series Analysis* 29, no. 3: 533–554.

Bao, Y., and Y. Hua. 2014. "On the Fisher Information Matrix of a Vector ARMA Process." *Economics Letters* 123, no. 1: 14–16.

Bartel, H., and H. Lütkepohl. 1998. "Estimating the Kronecker Indices of Cointegrated Echelon-Form VARMA Models." *Econometrics Journal* 1, no. 1: 76–99.

Bhansali, R. 2020. "Model Specification and Selection for Multivariate Time Series." *Journal of Multivariate Analysis* 175: 104539.

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, no. 8: 2961–2975.

Boubacar Maïnassara, Y. 2012. "Selection of Weak VARMA Models by Modified Akaike's Information Criteria." *Journal of Time Series Analysis* 33, no. 1: 121–130.

Boubacar Maïnassara, Y., and C. Francq. 2011. "Estimating Structural VARMA Models With Uncorrelated but Non-independent Error Terms." *Journal of Multivariate Analysis* 102, no. 3: 496–505.

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, no. 524: 1813–1827.

Boudjellaba, B., J.-M. Dufour, and R. Roy. 1991. *Testing Causality Between Two Vectors in Multivariate ARMA Models. Technical Report.* Montreal, Canada: Universite de Montreal, Departement de sciences economiques.

Boudjellaba, H., J.-M. Dufour, and R. Roy. 1994. "Simplified Conditions for Noncausality Between Vectors in Multivariate ARMA Models." *Journal of Econometrics* 63, no. 1: 271–287.

Box, G., and G. Jenkins. 1976. *Time Series Analysis Forecasting and Control.* San Francisco: Holkden-Day.

Box, G., and G. Tiao. 1977. "A Canonical Analysis of Multiple Time Series." *Biometrika* 64, no. 2: 355–365.

Brockwell, P., and R. Davis. 1991. *Time Series: Theory and Methods.* New York, NY: Springer Series in Statistics.

Cavicchioli, M. 2013. "Spectral Density of Markov-Switching VARMA Models." *Economics Letters* 121, no. 2: 218–220.

Cavicchioli, M. 2014. "Determining the Number of Regimes in Markov Switching VAR and VMA Models." *Journal of Time Series Analysis* 35, no. 2: 173–186.

Cavicchioli, M. 2016. "Weak VARMA Representations of Regime-Switching State-Space Models." *Statistical Papers* 57, no. 3: 705–720.

Cavicchioli, M. 2017a. "Asymptotic Fisher Information Matrix of Markov Switching VARMA Models." *Journal of Multivariate Analysis* 157: 124–135.

Cavicchioli, M. 2017b. "Higher Order Moments of Markov Switching VARMA Models." *Econometric Theory* 33, no. 6: 1502–1515.

Cavicchioli, M. 2020. "A Note on the Asymptotic and Exact Fisher Information Matrices of a Markov Switching VARMA Process." *Statistical Methods & Applications* 29, no. 1: 129–139.

Chan, J., and E. Eisenstat. 2017. "Efficient Estimation of Bayesian VARMAs With Time-Varying Coefficients." *Journal of Applied Econometrics* 32, no. 7: 1277–1297.

Chan, J., E. Eisenstat, and G. Koop. 2016. "Large Bayesian VARMAs." *Journal of Econometrics* 192, no. 2: 374–390.

Chan, W., and K. Wallis. 1978. "Multiple Time Series Modelling: Another Look at the Mink-Muskrat Interaction." *Journal of the Royal Statistical Society: Series C: Applied Statistics* 27, no. 2: 168–175.

Cooley, T., and M. Dwyer. 1998. "Business Cycle Analysis Without Much Theory a Look at Structural VARs." *Journal of Econometrics* 83, no. 1–2: 57–88.

Cooper, D., and E. Wood. 1982. "Identifying Multivariate Time Series Models." *Journal of Time Series Analysis* 3, no. 3: 153–164.

Cubadda, G., A. Hecq, and F. C. Palm. 2009. "Studying Co-Movements in Large Multivariate Data Prior to Multivariate Modelling." *Journal of Econometrics* 148, no. 1: 25–35.

Das, S. 2003. "Modelling Money, Price and Output in India: A Vector Autoregressive and Moving Average (VARMA) Approach." *Applied Economics* 35, no. 10: 1219–1225.

De Frutos, R., and G. Serrano. 2002. "A Generalized Least Squares Estimation Method for VARMA Models." *Statistics: A Journal of Theoretical and Applied Statistics* 36, no. 4: 303–316.

Deistler, M. 1985. "General Structure and Parametrization of ARMA and State-Space Systems and Its Relation to Statistical Problems." In *Handbook of Statistics*, vol. 5, 257–277. Amsterdam, the Netherlands: Elsevier.

Deistler, M., W. Dunsmuir, and E. Hannan. 1978. "Vector Linear Time Series Models: Corrections and Extensions." *Advances in Applied Probability* 10, no. 2: 360–372.

Dias, G., and G. Kapetanios. 2018. "Estimation and Forecasting in Vector Autoregressive Moving Average Models for Rich Datasets." *Journal of Econometrics* 202, no. 1: 75–91.

Do, H. Q., M. I. Bhatti, and M. Shahbaz. 2020. "Is 'Oil and Gas' Industry of ASEAN5 Countries Integrated With the US Counterpart?" *Applied Economics* 52, no. 37: 4112–4134.

Dufour, J., and T. Jouini. 2014. "Asymptotic Distributions for Quasi-Efficient Estimators in Echelon VARMA Models." *Computational Statistics & Data Analysis* 73: 69–86.

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*, 209–240. New York, NY: Springer.

Dufour, J.-M., and D. Pelletier. 2022. "Practical Methods for Modeling Weak VARMA Processes: Identification, Estimation and Specification With a Macroeconomic Application." *Journal of Business & Economic Statistics* 40, no. 3: 1140–1152.

Dufour, J.-M., and E. Renault. 1998. "Short Run and Long Run Causality in Time Series: Theory." *Econometrica* 66, no. 5: 1099–1125.

Dufour, J.-M., and D. Stevanović. 2013. "Factor-Augmented VARMA Models With Macroeconomic Applications." *Journal of Business & Economic Statistics* 31, no. 4: 491–506.

Dufour, J.-M., and A. Taamouti. 2010. "Short and Long Run Causality Measures: Theory and Inference." *Journal of Econometrics* 154, no. 1: 42–58.

Dunsmuir, W., and E. Hannan. 1976. "Vector Linear Time Series Models." *Advances in Applied Probability* 8, no. 2: 339–364.

Eberts, R., and B. Steece. 1984. "A Test for Granger-Sausality in a Multivariate ARMA Model." *Empirical Economics* 9, no. 1: 51–58.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. Sargent, and M. Watson. 2007. "ABCs (and Ds) of Understanding VARs." *American Economic Review* 97, no. 3: 1021–1026.

Francq, C., and J.-M. Zakoïan. 2001. "Stationarity of Multivariate Markov–Switching ARMA Models." *Journal of Econometrics* 102, no. 2: 339–364.

Funovits, B. 2024. "Identifiability and Estimation of Possibly Non-Invertible SVARMA Models: A New Parametrisation." *Journal of Econometrics* 241, no. 2: 105766.

Galeano, P., and D. Peña. 2007. "Covariance Changes Detection in Multivariate Time Series." *Journal of Statistical Planning and Inference* 137, no. 1: 194–211.

Gallego, J. 2009. "The Exact Likelihood Function of a Vector Autoregressive Moving Average Process." *Statistics & Probability Letters* 79, no. 6: 711–714.

Golosnoy, V., and M. I. Seifert. 2021. "Monitoring Mean Changes in Persistent Multivariate Time Series." *Statistics* 55, no. 3: 475–488.

Gouriéroux, C., A. Monfort, and J.-P. Renne. 2020. "Identification and Estimation in Non-Fundamental Structural VARMA Models." *Review of Economic Studies* 87, no. 4: 1915–1953.

Granger, C. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37, no. 3: 424–438.

Granger, C. 1980. "Testing for Causality: A Personal Viewpoint." *Journal of Economic Dynamics and Control* 2, no. 1: 329–352.

Grillenzoni, C. 1991. "Simultaneous Transfer Functions Versus Vector ARMA Models." *Journal of Forecasting* 10, no. 5: 477–499.

Guo, F., and S. Ling. 2024. "Statistical Inference for Heavy-Tailed and Partially Nonstationary Vector ARMA Models." *Statistica Sinica*.

Guo, H., X. Liu, and Z. Sun. 2016. "Multivariate Time Series Prediction Using a Hybridization of VARMA Models and Bayesian Networks." *Journal of Applied Statistics* 43, no. 16: 2897–2909.

Hall, A., and D. Nicholls. 1980. "The Evaluation of Exact Maximum Likelihood Estimates for VARMA Models." *Journal of Statistical Computation and Simulation* 10, no. 3–4: 251–262.

Hallin, M. 1978. "Mixed Autoregressive-Moving Average Multivariate Processes With Time-Dependent Coefficients." *Journal of Multivariate Analysis* 8, no. 4: 567–572.

Hallin, M., and H. Liu. 2023. "Center-Outward Rank- and Sign-Based VARMA Portmanteau Tests: Chitturi, Hosking, and Li–McLeod Revisited." *Econometrics and Statistics*.

Hallin, M., and D. Paindaveine. 2004. "Rank-Based Optimal Tests of the Adequacy of an Elliptic VARMA Model." *Annals of Statistics* 32, no. 6: 2642–2678.

Hallin, M., and D. Paindaveine. 2005. "Affine-Invariant Aligned Rank Tests for the Multivariate General Linear Model With VARMA Errors." *Journal of Multivariate Analysis* 93, no. 1: 122–163.

Hallin, M., and A. Saidi. 2005. "Testing Non-Correlation and Non-Causality Between Multivariate ARMA Time Series." *Journal of Time Series Analysis* 26, no. 1: 83–105.

Hannan, E. 1969. "The Identification of Vector Mixed Autoregressive-Moving Average System." *Biometrika* 56, no. 1: 223–225.

Hannan, E. 1971. "The Identification Problem for Multiple Equation Systems With Moving Average Errors." *Econometrica* 39, no. 5: 751–765.

Hannan, E., W. Dunsmuir, and M. Deistler. 1980. "Estimation of Vector ARMAX Models." *Journal of Multivariate Analysis* 10, no. 3: 275–295.

Hannan, E., and L. Kavalieris. 1984. "Multivariate Linear Time Series Models." *Advances in Applied Probability* 16, no. 3: 492–561.

Hannan, E. J., and M. Deistler. 1988. *The Statistical Theory of Linear Systems*. Vol. 68. Hoboken, NJ: Wiley.

Haugh, L. 1976. "Checking the Independence of Two Covariance-Stationary Time Series: A Univariate Residual Cross-Correlation Approach." *Journal of the American Statistical Association* 71, no. 354: 378–385.

Hillmer, S. C., and G. Tiao. 1979. "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models." *Journal of the American Statistical Association* 74, no. 367: 652–660.

Himdi, K., and R. Roy. 1997. "Tests for Noncorrelation of Two Multivariate ARMA Time Series." *Canadian Journal of Statistics* 25, no. 2: 233–256.

Hosking, J. 1980. "The Multivariate Portmanteau Statistic." *Journal of the American Statistical Association* 75, no. 371: 602–608.

Hosking, J. 1981. "Lagrange-Multiplier Tests of Multivariate Time-Series Models." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 43, no. 2: 219–230.

Hundley, G., and S. Koreisha. 1987. "The Specification of Econometric Strike Models: A VARMA Approach." *Applied Economics* 19, no. 4: 511–530.

Hung, K., and F. Alt. 1994. "The Approximation of the One-Step Ahead Forecast Error Covariance for Vector ARMA Models." *International Journal of Forecasting* 10, no. 1: 59–64.

Ilmi Amir, A., and Y. Boubacar Maïnassara. 2020. "Multivariate Portmanteau Tests for Weak Multiplicative Seasonal VARMA Models." *Statistical Papers* 61, no. 6: 2529–2560.

Isufi, E., A. Loukas, N. Perraudin, and G. Leus. 2019. "Forecasting Time Series With VARMA Recursions on Graphs." *IEEE Transactions on Signal Processing* 67, no. 18: 4870–4885.

James, C., S. Koreisha, and M. Partch. 1985. "A VARMA Analysis of the Causal Relations Among Stock Returns, Real Output, and Nominal Interest Rates." *Journal of Finance* 40, no. 5: 1375–1384.

Jenkins, G., and A. Alavi. 1981. "Some Aspects of Modelling and Forecasting Multivariate Time Series." *Journal of Time Series Analysis* 2, no. 1: 1–47.

Jonasson, K. 2008. "Algorithm 878: Exact VARMA Likelihood and Its Gradient for Complete and Incomplete Data With Matlab." *ACM Transactions on Mathematical Software (TOMS)* 35, no. 1: 1–11.

Jonasson, K., and S. Ferrando. 2008. "Evaluating Exact VARMA Likelihood and Its Gradient When Data Are Incomplete." *ACM Transactions on Mathematical Software (TOMS)* 35, no. 1: 1–16.

Jouini, T. 2015. "Linear Bootstrap Methods for Vector Autoregressive Moving-Average Models." *Journal of Statistical Computation and Simulation* 85, no. 11: 2214–2257.

Kang, H. 1981. "Necessary and Sufficient Conditions for Causality Testing in Multivariate ARMA Models." *Journal of Time Series Analysis* 2, no. 2: 95–101.

Kapetanios, G. 2003. "A Note on an Iterative Least-Squares Estimation Method for ARMA and VARMA Models." *Economics Letters* 79, no. 3: 305–312.

Kascha, C. 2012. "A Comparison of Estimation Methods for Vector Autoregressive Moving-Average Models." *Econometric Reviews* 31, no. 3: 297–324.

Katayama, N. 2012. "Chi-Squared Portmanteau Tests for Structural VARMA Models With Uncorrelated Errors." *Journal of Time Series Analysis* 33, no. 6: 863–872.

Kathari, S., and A. Tangirala. 2020. "Scalar Correlation Functions for Model Structure Selection in High-Dimensional Time-Series Modelling." *ISA Transactions* 100: 275–288.

Kharrati-Kopaei, M., A. Nematollahi, and Z. Shishebor. 2009. "On the Sufficient Statistics for Multivariate ARMA Models: Approximate Approach." *Statistical Papers* 50: 261–276.

Klein, A., and G. Mélard. 2014. "An Algorithm for the Exact Fisher Information Matrix of Vector ARMAX Time Series." *Linear Algebra and Its Applications* 446: 1–24.

Klein, A., and G. Mélard. 2023. "An Algorithm for the Fisher Information Matrix of a VARMAX Process." *Algorithms* 16, no. 8: 364.

Klein, A., G. Mélard, and A. Saidi. 2008. "The Asymptotic and Exact Fisher Information Matrices of a Vector ARMA Process." *Statistics & Probability Letters* 78, no. 12: 1430–1433.

Klein, A., and P. Spreij. 2006. "An Explicit Expression for the Fisher Information Matrix of a Multiple Time Series Process." *Linear Algebra and Its Applications* 417, no. 1: 140–149.

Kohn, R. 1979. "Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models." *Econometrica* 47, no. 4: 1005–1030.

Kohn, R., and C. Ansley. 1982. "A Note on Obtaining the Theoretical Autocovariances of an ARMA Process." *Journal of Statistical Computation and Simulation* 15, no. 4: 273–283.

Komunjer, I., and S. Ng. 2011. "Dynamic Identification of Dynamic Stochastic General Equilibrium Models." *Econometrica* 79, no. 6: 1995–2032.

Koreisha, S., and T. Pukkila. 1987. "Identification of Nonzero Elements in the Polynomial Matrices of Mixed VARMA Processes." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 49, no. 1: 112–126.

Koreisha, S., and T. Pukkila. 1989. "Fast Linear Estimation Methods for Vector Autoregressive Moving-Average Models." *Journal of Time Series Analysis* 10, no. 4: 325–339.

Koreisha, S., and T. Pukkila. 2004. "The Specification of Vector Autoregressive Moving Average Models." *Journal of Statistical Computation and Simulation* 74, no. 8: 547–565.

Li, H., and R. Tsay. 1998. "A Unified Approach to Identifying Multivariate Time Series Models." *Journal of the American Statistical Association* 93, no. 442: 770–782.

Li, W., and A. McLeod. 1981. "Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 43, no. 2: 231–239.

Ling, S., and M. McAleer. 2003. "Asymptotic Theory for a Vector ARMA-GARCH Model." *Econometric Theory* 19, no. 2: 280–310.

Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin-Germany: Springer-Verlag.

Lütkepohl, H. 2006. "Forecasting With VARMA Models." In *Handbook of Economic Forecasting*, vol. 1, 287–325. Amsterdam, the Netherlands: Elsevier.

Lütkepohl, H., and H. Claessen. 1997. "Analysis of Cointegrated VARMA Processes." *Journal of Econometrics* 80, no. 2: 223–239.

Lütkepohl, H., and D. Poskitt. 1996. "Specification of Echelon-Form VARMA Models." *Journal of Business & Economic Statistics* 14, no. 1: 69–79.

Mahdi, E., and A. Ian McLeod. 2012. "Improved Multivariate Portmanteau Test." *Journal of Time Series Analysis* 33, no. 2: 211–222.

Mauricio, J. A. 1995. "Exact Maximum Likelihood Estimation of Stationary Vector ARMA Models." *Journal of the American Statistical Association* 90, no. 429: 282–291.

Mauricio, J. A. 1997. "The Exact Likelihood Function of a Vector Autoregressive Moving Average Model." *Journal of the Royal Statistical Society: Series C: Applied Statistics* 46, no. 1: 157–171.

Mauricio, J. A. 2002. "An Algorithm for the Exact Likelihood of a Stationary Vector Autoregressive-Moving Average Model." *Journal of Time Series Analysis* 23, no. 4: 473–486.

McAleer, M., F. Chan, S. Hoti, and O. Lieberman. 2008. "Generalized Autoregressive Conditional Correlation." *Econometric Theory* 24, no. 6: 1554–1583.

McAleer, M., S. Hoti, and F. Chan. 2009. "Structure and Asymptotic Theory for Multivariate Asymmetric Conditional Volatility." *Econometric Reviews* 28, no. 5: 422–440.

McElroy, T. 2017. "Computation of Vector ARMA Autocovariances." *Statistics & Probability Letters* 124: 92–96.

McElroy, T. 2022. "Frequency Domain Calculation of Seasonal VARMA Autocovariances." *Journal of Computational and Graphical Statistics* 31, no. 1: 301–303.

Mélard, G. 2022. "An Indirect Proof for the Asymptotic Properties of VARMA Model Estimators." *Econometrics and Statistics* 21: 96–111.

Mélard, G. 2024. *Time-Dependent Time Series Models: Comments on Marc Hallin's Early Contributions and a Pragmatic View on Estimation*, 429–446. Cham: Springer Nature Switzerland.

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, no. 11: 2958–2986.

Metaxoglou, K., and A. Smith. 2007. "Maximum Likelihood Estimation of VARMA Models Using a State-Space EM Algorithm." *Journal of Time Series Analysis* 28, no. 5: 666–685.

Mittnik, S. 1990. "Computation of Theoretical Autocovariance Matrices of Multivariate Autoregressive Moving Average Time Series." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 52, no. 1: 151–155.

Mittnik, S. 1993. "Computing Theoretical Autocovariances of Multivariate Autoregressive Moving Average Models by Using a Block Levinson Method." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 55, no. 2: 435–440.

Newbold, P., and S. Hotopp. 1986. "Testing Causality Using Efficiently Parametrized Vector ARMA Models." *Applied Mathematics and Computation* 20, no. 3–4: 329–348.

Nicholls, D. 1976. "The Efficient Estimation of Vector Linear Time Series Models." *Biometrika* 63, no. 2: 381–390.

Nicholls, D. 1977. "A Comparison of Estimation Methods for Vector Linear Time Series Models." *Biometrika* 64, no. 1: 85–90.

Nicholls, D., and A. Hall. 1979. "The Exact Likelihood Function of Multivariate Autoregressive-Moving Average Models." *Biometrika* 66, no. 2: 259–264.

Niglio, M., and C. D. Vitale. 2015. "Threshold Vector ARMA Models." *Communications in Statistics - Theory and Methods* 44, no. 14: 2911–2923.

Nsiri, S., and R. Roy. 1992. "On the Identification of ARMA Echelon-Form Models." *Canadian Journal of Statistics* 20, no. 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, no. 2: 207–231.

Oke, T., and L.-E. Öller. 1999. "Testing for Short Memory in a VARMA Process." *Journal of Forecasting* 18, no. 7: 477–487.

Osborn, D. R. 1984. "Causality Testing and Its Implications for Dynamic Econometric Models." *Economic Journal* 94, no. Supplement: 82–96.

Paparoditis, E. 2005. "Testing the Fit of a Vector Autoregressive Moving Average Model." *Journal of Time Series Analysis* 26, no. 4: 543–568.

Pataracchia, B. 2011. "The Spectral Representation of Markov Switching ARMA Models." *Economics Letters* 112, no. 1: 11–15.

Peña, D., and G. E. Box. 1987. "Identifying a Simplifying Structure in Time Series." *Journal of the American Statistical Association* 82, no. 399: 836–843.

Peña, D., and J. Rodriguez. 2002. "A Powerful Portmanteau Test of Lack of Fit for Time Series." *Journal of the American Statistical Association* 97, no. 458: 601–610.

Peña, D., and I. Sánchez. 2007. "Measuring the Advantages of Multivariate vs. Univariate Forecasts." *Journal of Time Series Analysis* 28, no. 6: 886–909.

Pfaff, B. 2008. "VAR, SVAR and SVEC Models: Implementation Within R Package vars." *Journal of Statistical Software* 27, no. 4: 1–32.

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

Poskitt, D. 2003. "On the Specification of Cointegrated Autoregressive Moving-Average Forecasting Systems." *International Journal of Forecasting* 19, no. 3: 503–519.

Poskitt, D. 2006. "On the Identification and Estimation of Nonstationary and Cointegrated ARMAX Systems." *Econometric Theory* 22, no. 6: 1138–1175.

Poskitt, D. 2016. "Vector Autoregressive Moving Average Identification for Macroeconomic Modeling: A New Methodology." *Journal of Econometrics* 192, no. 2: 468–484.

Poskitt, D., and M. Salau. 1994. "On the Asymptotic Relative Efficiency of Gaussian and Least Squares Estimators for Vector ARMA Models." *Journal of Multivariate Analysis* 51, no. 2: 294–317.

Poskitt, D., and M. Salau. 1995. "On the Relationship Between Generalized Least Squares and Gaussian Estimation of Vector ARMA Models." *Journal of Time Series Analysis* 16, no. 6: 617–645.

Poskitt, D. S., and A. Tremayne. 1982. "Diagnostic Tests for Multiple Time Series Models." *Annals of Statistics* 10, no. 1: 114–120.

Pukkila, T., S. Koreisha, and A. Kallinen. 1990. "The Identification of Arma Models." *Biometrika* 77, no. 3: 537–548.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rahman, S., and A. Serletis. 2012. "Oil Price Uncertainty and the Canadian Economy: Evidence From a VARMA, GARCH-in-Mean, Asymmetric BEKK Model." *Energy Economics* 34, no. 2: 603–610.

Ratsimalahelo, Z. 2001. "Specification of Varma Echelon Form Models." *IFAC Proceedings Volumes* 34, no. 20: 129–136.

Ravishanker, N., and B. Ray. 1997. "Bayesian Analysis of Vector ARMA Models Using Gibbs Sampling." *Journal of Forecasting* 16, no. 3: 177–194.

Reinsel, G. 1979. "FIML Estimation of the Dynamic Simultaneous Equations Model With ARMA Disturbances." *Journal of Econometrics* 9, no. 3: 263–281.

Reinsel, G. 1993. *Elements of Multivariate Time Series Analysis.* New York: Springer.

Reinsel, G. 1995. "Finite Sample Forecast Results for Vector Autoregressive Moving Average Models." *Journal of Forecasting* 14, no. 4: 405–412.

Reinsel, G., S. Basu, and S. Yap. 1992. "Maximum Likelihood Estimators in the Multivariate Autoregressive Moving-Average Model From a Generalized Least Squares Viewpoint." *Journal of Time Series Analysis* 13, no. 2: 133–145.

Rissanen, J., and P. Caines. 1979. "The Strong Consistency of Maximum Likelihood Estimators for ARMA Processes." *Annals of Statistics* 7, no. 1: 297–315.

Roy, A., T. McElroy, and P. Linton. 2019. "Constrained Estimation of Causal Invertible VARMA." *Statistica Sinica* 29: 455–478.

Salau, M. 1997. "Numerical Computation of Asymptotic Covariance Matrix of the Gaussian Estimators for Vector Arma Models." *Communications in Statistics: Simulation and Computation* 26, no. 1: 173–192.

Salau, M. O. 1999. "On the Numerical Evaluation of the Theoretical Variance-Covariance Matrix of Least Squares Estimators for Echelon-Form VARMA Models." *Communications in Statistics: Simulation and Computation* 28, no. 3: 743–766.

Salisu, A., K. Isah, and A. Assandri. 2019. "Dynamic Spillovers Between Stock and Money Markets in Nigeria: A VARMA-Garch Approach." *Review of Economic Analysis* 11, no. 2: 255–283.

Serletis, A., and A. Shahmoradi. 2006. "Velocity and the Variability of Money Growth: Evidence From a VARMA, GARCH-M Model." *Macroeconomic Dynamics* 10, no. 5: 652–666.

Serletis, A., and L. Xu. 2018. "The Zero Lower Bound and Crude Oil and Financial Markets Spillovers." *Macroeconomic Dynamics* 22, no. 3: 654–665.

Shaarawy, S. 1989. "Bayesian Inferences and Forecasts With Multiple Autoregressive Moving Average Models." *Communications in Statistics: Simulation and Computation* 18, no. 4: 1481–1509.

Shaarawy, S. 2023. "Bayesian Modeling and Forecasting of Vector Autoregressive Moving Average Processes." *Communications in Statistics - Theory and Methods* 52, no. 11: 3795–3815.

She, R., Z. Mi, and S. Ling. 2022. "Whittle Parameter Estimation for Vector ARMA Models With Heavy-Tailed Noises." *Journal of Statistical Planning and Inference* 219: 216–230.

Shea, B. 1987. "Estimation of Multivariate Time Series." *Journal of Time Series Analysis* 8, no. 1: 95–109.

Shea, B. 1988. "A Note on the Generation of Independent Realizations of a Vector Autoregressive Moving-Average Process." *Journal of Time Series Analysis* 9, no. 4: 403–410.

Shea, B. 1989. "Algorithm as 242: The Exact Likelihood of a Vector Autoregressive Moving Average Model." *Applied Statistics* 38: 161–184.

Shelton Peiris, M., and A. Thavaneswaran. 2001. "Multivariate Stable ARMA Processes With Time Dependent Coefficients." *Metrika* 54, no. 2: 131–138.

Shi, Y., and Y. Sheng. 2024. "Uncertain Vector Autoregressive Smoothly Moving Average Model." *Communications in Statistics: Simulation and Computation* 53, no. 12: 6038–6049.

Solo, V. 1984. "The Exact Likelihood for a Multivariate ARMA Model." *Journal of Multivariate Analysis* 15, no. 2: 164–173.

Spliid, H. 1983. "A Fast Estimation Method for the Vector Autoregressive Moving Average Model With Exogenous Variables." *Journal of the American Statistical Association* 78, no. 384: 843–849.

Steland, A. 2020. "Testing and Estimating Change-Points in the Covariance Matrix of a High-Dimensional Time Series." *Journal of Multivariate Analysis* 177: 104582.

Stelzer, R. 2008. "Multivariate Markov-Switching ARMA Processes With Regularly Varying Noise." *Journal of Multivariate Analysis* 99, no. 6: 1177–1190.

Stelzer, R. 2009. "On Markov-Switching ARMA Processes–Stationarity, Existence of Moments, and Geometric Ergodicity." *Econometric Theory* 25, no. 1: 43–62.

Taylor, S. 1989. "A Comparison of Classical Tests of Criterion When Determining Granger Causality With a Bivariate ARMA Model." *Empirical Economics* 14: 257–271.

Tiao, G., and G. Box. 1981. "Modeling Multiple Time Series With Applications." *Journal of the American Statistical Association* 76, no. 376: 802–816.

Tiao, G., and R. Tsay. 1983. "Multiple Time Series Modeling and Extended Sample Cross-Correlations." *Journal of Business & Economic Statistics* 1, no. 1: 43–56.

Tiao, G., and R. Tsay. 1989. "Model Specification in Multivariate Time Series." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 51, no. 2: 157–195.

Toscano, E., and V. A. Reisen. 2000. "The Use of Canonical Correlation Analysis to Identify the Order of Multivariate ARMA Models: Simulation and Application." *Journal of Forecasting* 19, no. 5: 441–455.

Tsay, R. 1989. "Identifying Multivariate Time Series Models." *Journal of Time Series Analysis* 10, no. 4: 357–372.

Tsay, R. 1991. "Two Canonical Forms for Vector ARMA Processes." *Statistica Sinica* 1: 247–269.

Tsay, R. 2013. *Multivariate Time Series Analysis: With R and Financial Applications*. Hoboken, NJ: John Wiley & Sons.

Tsay, R., and G. Tiao. 1984. "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models." *Journal of the American Statistical Association* 79, no. 385: 184–196.

Tsay, R., and G. Tiao. 1985. "Use of Canonical Analysis in Time Series Model Identification." *Biometrika* 72, no. 2: 299–315.

Tsay, R., and D. Wood. 2022. "MTS: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models." R Package Version 1.2.1.

Tunnicliffe Wilson, G. 1973. "The Estimation of Parameters in Multivariate Time Series Models." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 35, no. 1: 76–85.

Velasco, C. 2023. "Identification and Estimation of Structural VARMA Models Using Higher Order Dynamics." *Journal of Business & Economic Statistics* 41, no. 3: 819–832.

Velilla, S., and H. Nguyen. 2019. "A New Diagnostic Tool for VARMA($p,q$) Models." *Statistics* 53, no. 4: 866–884.

Velilla, S., and H. Thu. 2018. "A Goodness-of-Fit Test for VARMA($p,q$) Models." *Journal of Statistical Planning and Inference* 197: 126–140.

Vinod, H. 1996. "Empirically Feasible Solutions and Explicit Dynamics for Rational Expectation Models." *Journal of Statistical Planning and Inference* 49, no. 2: 203–222.

Wallis, K. 1977. "Multiple Time Series Analysis and the Final Form of Econometric Models." *Econometrica* 45, no. 6: 1481–1497.

Wang, Y., and R. S. Tsay. 2013. "On Diagnostic Checking of Vector ARMA-GARCH Models With Gaussian and Student-$t$ Innovations." *Econometrics* 1: 1–31.

Wilms, I., S. Basu, J. Bien, and D. Matteson. 2023. "Sparse Identification and Estimation of Large-Scale Vector Autoregressive Moving Averages." *Journal of the American Statistical Association* 118, no. 541: 571–582.

Wilms, I., S. Basu, D. Matteson, J. Bien, W. Nicholson, and E. Wegner. 2021. "bigtime: Sparse Estimation of Large Time Series Models."

Wu, E. M.-Y., and S.-L. Kuo. 2020. "VARMA-EGARCH Model for Air-Quality Analyses and Application in Southern Taiwan." *Atmosphere* 11, no. 10: 1096.

Wu, E. M.-Y., S.-L. Kuo, and W.-C. Liu. 2012. "Generalized Autoregressive Conditional Heteroskedastic Model for Water Quality Analyses and Time Series Investigation in Reservoir Watersheds." *Environmental Engineering Science* 29, no. 4: 227–237.

Yamamoto, T. 1980. "On the Treatment of Autocorrelated Errors in the Multiperiod Prediction of Dynamic Simultaneous Equation Models." *International Economic Review* 21, no. 3: 735–748.

Yamamoto, T. 1981. "Predictions of Multivariate Autoregressive-Moving Average Models." *Biometrika* 68, no. 2: 485–492.

Yang, H., Z. Pan, Q. Tao, and J. Qiu. 2018. "Online Learning for Vector Autoregressive Moving-Average Time Series Prediction." *Neurocomputing* 315: 9–17.

Yap, S., and G. Reinsel. 1995. "Estimation and Testing for Unit Roots in a Partially Nonstationary Vector Autoregressive Moving Average Model." *Journal of the American Statistical Association* 90, no. 429: 253–267.

Yozgatligil, C., and W. W. Wei. 2009. "Representation of Multiplicative Seasonal Vector Autoregressive Moving Average Models." *American Statistician* 63, no. 4: 328–334.

Zadrozny, P. 1989. "Analytic Derivatives for Estimation of Linear Dynamic Models." *Computers & Mathematics With Applications* 18, no. 6–7: 539–553.

Zadrozny, P. 1992. "Errata to "Analytic Derivatives for Estimation of Linear Dynamic Models": Computers and Mathematics With Applications, vol. 18, no. 6/7, pp. 539–553, 1989." *Computers & Mathematics With Applications* 24, no. 8–9: 289–290.

Zadrozny, P., and B. Chen. 2019. "Weighted-Covariance Factor Decomposition of VARMA Models Applied to Forecasting Quarterly Us Real GDP at Monthly Intervals." *Journal of Time Series Analysis* 40, no. 6: 968–986.

Zadrozny, P., and S. Mittnik. 1994. "Kalman-Filtering Methods for Computing Information Matrices for Time-Invariant, Periodic, and Generally Time-Varying VARMA Models and Samples." *Computers & Mathematics With Applications* 28, no. 4: 107–119.

Zellner, A., and F. Palm. 1974. "Time Series Analysis and Simultaneous Equation Econometric Models." *Journal of Econometrics* 2, no. 1: 17–54.

Zhang, J., and R. Stine. 2001. "Autocovariance Structure of Markov Regime Switching Models and Model Selection." *Journal of Time Series Analysis* 22, no. 1: 107–124.

Zheng, Y. 2024. "An Interpretable and Efficient Infinite-Order Vector Autoregressive Model for High-Dimensional Time Series." *Journal of the American Statistical Association*: 1–26.