Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application*

Jean-Marie Dufour† McGill University

Denis Pelletier‡ North Carolina State University

January 2021

* The authors thank Marine Carasco, John Galbraith, Nour Meddahi and Rui Castro, an Associate Editor and two anonymous referees for several useful comments. The second author gratefully acknowledges financial assistance from the Social Sciences and Humanities Research Council of Canada, the Government of Québec (fonds FCAR), the CRDE and CIRANO. Earlier versions of the paper circulated under the title *Linear Estimation of Weak VARMA Models With a Macroeconomic Application*. This work was supported by the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the CIREQ, the CIRANO, and the Fonds FCAR (Government of Québec).

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

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

ABSTRACT

We consider the problem of developing practical methods for modelling weak VARMA processes. We first propose new identified VARMA representations, the *diagonal MA equation form* and the *final MA equation form*, where the MA operator is either diagonal or scalar. Both these representations have the important feature that they constitute relatively simple modifications of a VAR model (in contrast with the echelon representation). Second, for estimating VARMA models, we develop computationally simple methods which only require linear regressions, through an extension of the regression-based approach proposed by Hannan and Rissanen (1982). The asymptotic properties of the estimator are derived under weak hypotheses on the innovations (uncorrelated and strong mixing), in order to broaden the class of models to which it can be applied. Third, we present a modified information criterion which yields consistent estimates of the orders under the proposed representations. The estimation methods are studied by simulation. To demonstrate the importance of using VARMA models to study multivariate time series, we compare the impulse-response functions and the out-of-sample forecasts generated by VARMA and VAR models. The proposed methodology is applied to a six-variable macroeconomic model of monetary policy, based on U.S. monthly data over the period 1962-1996. The results demonstrate the advantages of using the VARMA methodology for impulse response estimation and forecasting, in contrast with standard VAR models.

Key words: linear regression; VARMA; final equation form; information criterion; weak representation; strong mixing condition; impulse-response function.

Journal of Economic Literature Classification: C13, C32, C51, E0.

1. Introduction

In time series analysis and econometrics, VARMA models are scarcely used to represent multivariate time series. VAR models are much more widely employed because they are easier to implement. The latter models can be estimated by least-squares methods. Specification is also easier for VAR models because only one lag order need be chosen. VAR models, however, have important drawbacks. First, they are typically less parsimonious than VARMA models [*e.g.*, see Lütkepohl and Poskitt (1996)]. Second, the family of VAR models is not closed under marginalization and temporal aggregation [see Lütkepohl (1991)].

It follows that VARMA models appear to be preferable from a theoretical viewpoint, but their adoption is complicated by identification and estimation difficulties. The unrestricted VARMA representation is not identified [see Lütkepohl (1991, Section 7.1.1)], and we need to decide on a set of constraints to impose so as to achieve identification. The dominant representation, *i.e.* the echelon form, consist in choosing a set of Kronecker indices which is not as intuitive as choosing the order of a VAR. Standard estimation methods for VARMA models (such as maximum likelihood) require nonlinear optimization which may not be feasible as soon as the model involves a few time series because the number of parameters can increase quickly.

The importance of nonlinear models has been growing in the time-series literature. Important classes of nonlinear processes admit an ARMA representation [see Francq and Zakoïan (1998), Francq, Roy, and Zakoïan (2005)], but involve innovations which may not satisfy the usual i.i.d. (independent and identically distributed) or m.d.s. (martingale difference sequence) property, though they are uncorrelated. We refer to these as strong and semi-strong ARMA models respectively, by opposition to weak ARMA models where the innovations are only uncorrelated. In fact, the Wold decomposition only guarantees that the innovations are uncorrelated.

In this paper, we consider the problem of modeling weak VARMA processes. Our goal is to develop a procedure which will ease the use of these models. It will cover three basic modeling operations: identification, estimation and specification.

First, in order to avoid identification problems and to make VARMA models easier to use, we introduce three new identified VARMA representations: the *diagonal MA equation form*, the *final MA equation form*, and the *diagonal AR equation form*. Under the diagonal MA (or AR) representation, the MA (or AR) operator is diagonal, and the lag operators for different variables (in the diagonal form) may have different orders. Under the final MA equation form representation the MA operator is scalar, *i.e.* the operators are equal across equations.

Second, we consider the problem of estimating VARMA models by relatively simple methods which only require linear regressions. For this purpose, we consider a multivariate generalization of the regression-based estimation method proposed by Hannan and Rissanen (1982) for univariate ARMA models. The method involves in three steps. First, a long autoregression is fitted to the data. Second, the lagged innovations in the ARMA model are replaced by the corresponding residuals from the long autoregression, and a regression is performed. Third, the data from the second step are filtered, which yields estimates with the same asymptotic covariance matrix as their nonlinear counterparts (maximum likelihood when innovations are i.i.d. Gaussian, or nonlinear least squares if they are merely uncorrelated).

Third, we suggest a modified information criterion to choose the orders of VARMA models under these representations. This criterion is minimized in the second step of the estimation method over the orders of the AR and MA operators and yields consistent estimates of these orders.

Our results do not require an i.i.d. or m.d.s. assumption on the innovations. We consider strong mixing conditions [Doukhan (1995), Bosq (1998)], rather than the usual m.d.s. assumption thus broadening the class of models to which our results apply.

We apply the proposed methodology to U.S. macroeconomic data previously studied by Bernanke and Mihov (1998) and McMillin (2001). To illustrate the impact of using VARMA models instead of VAR models to study multivariate time series, we compare the impulse-response functions generated by each model. We show that we can obtain much more precise estimates of the impulse-response function by using VARMA models instead of VAR models.

The rest of the paper is organized as follows. Our framework and notation are described in section 2. The new identified representations are presented in section 3. In section 4, we present the estimation method. In section 5, we describe the information criterion for choosing the orders of VARMA models under the proposed representations. Section 6 contains results of Monte Carlo simulations which illustrate the properties of our method. Section 7 presents the macroeconomic application where we compare the impulse-response functions from a VAR model and VARMA models. Section 8 contains a few concluding remarks. Finally, an expanded version of this introduction, proofs and lemmas are in an online appendix.

2. Framework

Consider the following K-variate VARMA(p, q) model in standard representation for real-valued series:

$$Y_t = \sum_{i=1}^{p} \Phi_i Y_{t-i} + U_t - \sum_{j=1}^{q} \Theta_j U_{t-j} \tag{2.1}$$

where U_t is a sequence of uncorrelated random variables with zero mean, defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$, and $t \in \mathbb{Z}$. The vectors Y_t and U_t contain the K univariate time series: $Y_t = [y_{1t}, \ldots, y_{Kt}]'$ and $U_t = [u_{1t}, \ldots, u_{Kt}]'$. We can also write the previous equation using matrix lag operators:

$$\Phi(L) Y_t = \Theta(L) U_t \tag{2.2}$$

where

$$\Phi(L) = I_K - \Phi_1 L - \cdots - \Phi_p L^p, \quad \Theta(L) = I_K - \Theta_1 L - \cdots - \Theta_q L^q. \tag{2.3}$$

Let H_t be the Hilbert space generated by $(Y_s,\ s < t)$. We consider the following assumptions on the Y_t process.

Assumption 2.1 *The process Y_t is stable and invertible,* i.e. $\det[\Phi(z)] \neq 0$ *and* $\det[\Theta(z)] \neq 0$ *for all* $|z| \leq 1$.

Assumption 2.2 *Y_t is a strictly stationary and ergodic sequence and that the process U_t has common variance $\mathrm{Var}[U_t] = \Sigma_U$ and finite fourth moment $E[|u_{it}|^{4+2\zeta}] < \infty$, for all i and t, for some $\zeta > 0$.*

Assumption 2.3 *The innovations U_t constitute a strictly stationary process and satisfy the mixing assumption $\sum_{h=1}^{\infty} \alpha(h)^{\epsilon/(2+\epsilon)} < \infty$ for some $\epsilon > 0$, where $\alpha(h)$ is the α-mixing coefficient of order h.*

Under Assumption **2.1**, where $\det[\Phi(L)]$ represents the determinant of the matrix lag operator $\Phi(L)$, U_t can be interpreted as the linear innovation of Y_t:

$$U_t = Y_t - E_L[Y_t|H_t]\,. \tag{2.4}$$

The case of $I(1)$ and cointegrated variables is left to future work. We make the zero mean-mean hypothesis for Y_t only to simplify notation. Being stable and invertible, the process Y_t has an infinite vector autoregressive (VAR) representation,

$$\Pi(L)Y_t = U_t\,, \tag{2.5}$$

$$\Pi(L) = \Theta(L)^{-1}\Phi(L) = I_K - \sum_{i=1}^{\infty} \Pi_i L^i\,, \tag{2.6}$$

and an infinite vector moving average (VMA) representation,

$$Y_t = \Psi(L)U_t\,, \tag{2.7}$$

$$\Psi(L) = \Phi(L)^{-1}\Theta(L) = I_K - \sum_{j=1}^{\infty} \Psi_j L^j\,. \tag{2.8}$$

We denote by $\varphi_{ik}(L)$ the polynomial in row i and column k of $\Phi(L)$, and the row i or column k of $\Phi(L)$ by:

$$\Phi_{i\bullet}(L) = [\varphi_{i1}(L), \ldots, \varphi_{iK}(L)]\,, \tag{2.9}$$

$$\Phi_{\bullet k}(L) = [\varphi_{1k}(L), \ldots, \varphi_{Kk}(L)]'. \tag{2.10}$$

The diag operator creates a diagonal matrix,

$$\mathrm{diag}[\varphi_{ii}(L)] = \mathrm{diag}[\varphi_{11}(L), \ldots, \varphi_{KK}(L)] = \begin{bmatrix} \varphi_{11}(L) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varphi_{KK}(L) \end{bmatrix}, \tag{2.11}$$

$$\varphi_{ii}(L) = 1 - \varphi_{ii,1}L - \cdots - \varphi_{ii,p}L^p. \tag{2.12}$$

The function $\deg[\varphi(L)]$ returns the degree of the polynomial $\varphi(L)$ and the function $\dim(\gamma)$ gives the length of the vector γ.

We need to impose some structure on the U_t process. The typical assumption made in the time series literature is that the U_t's are either independent and identically distributed (i.i.d.) or a martingale difference sequence (m.d.s.). In this work, we do not impose such strong assumptions because we wish to broaden the class of models to which it can be applied. We only assume that it satisfies the mixing condition in Assumption **2.3**; see *e.g.* Doukhan (1995) or Bosq (1998). This is a minimal condition that will be satisfied by many processes of interest.[1]

3. Identification and diagonal VARMA representations

It is important to note that we cannot work with the standard representation (2.1) because it is not identified. To help us gain intuition on the identification of VARMA models we can consider a more general representation where Φ_0 and Θ_0 are not identity matrices:

$$\Phi_0 Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \Theta_0 U_t - \Theta_1 U_{t-1} + \cdots + \Theta_q U_{t-q}\,. \tag{3.1}$$

By this specification, we mean the well-defined process

$$Y_t = (\Phi_0 - \Phi_1 L - \cdots - \Phi_p L^p)^{-1}(\Theta_0 + \Theta_1 L + \cdots + \Theta_q L^q)\, U_t\,. \tag{3.2}$$

But we can see that this process has a standard representation if Φ_0 and Θ_0 are nonsingular. To see this, we premultiply (3.1) by Φ_0^{-1} and define $\bar{U}_t = \Phi_0^{-1}\Theta_0 U_t$:

$$\begin{aligned} Y_t \;=\; & \Phi_0^{-1}\Phi_1 Y_{t-1} + \cdots + \Phi_0^{-1}\Phi_p Y_{t-p} \\ & + \bar{U}_t - \Phi_0^{-1}\Theta_1\Theta_0^{-1}\Phi_0\bar{U}_{t-1} - \cdots - \Phi_0^{-1}\Theta_q\Theta_0^{-1}\Phi_0\bar{U}_{t-q}\,. \end{aligned} \tag{3.3}$$

Redefining the matrices, we get a representation of type (2.1). As long as Φ_0 and Θ_0 are nonsingular, we can transform a non-standard VARMA model into a standard one.

We say that two VARMA representations are equivalent if the MA operator $\Psi(L) = \Phi(L)^{-1}\Theta(L)$ are the same. To ensure uniqueness of a VARMA representation, we must impose restrictions on the AR and MA operators so that, for given $\Psi(L)$, there is one and only one set of operators $\Phi(L)$ and $\Theta(L)$ which can generate this infinite MA representation. This is typically done by setting a set of parameters equal to zero.

A first restriction that we impose is a multivariate equivalent of the coprime property in the univariate case. We do not want factors of $\Phi(L)$ and $\Theta(L)$ to "cancel out" when $\Phi(L)^{-1}\Theta(L)$ is computed. This feature is called the *left-coprime* property [see Hannan (1969), Hannan and Deistler (1988), Lütkepohl (1993)]. There exist more than one representation which guarantee the uniqueness of the left-coprime operators. In the econometric literature, the predominant representation is the *echelon form* [see Deistler and Hannan (1981), Hannan and Kavalieris (1984), Lütkepohl (1993), Lütkepohl and Poskitt (1998)]. It requires the selection of Kronecker indices, which conceptually is not as easy as selecting the orders p and q of an ARMA model.[2] This may be a reason why

[1] See Andrews (1984) for some examples of first-order autoregressive processes that do not satisfy this mixing assumption.

[2] Specification of VARMA models in echelon form is discussed for example in Hannan and Kavalieris (1984), Lütke-

practitioners are reluctant to employ VARMA models.

Here, to make the use of VARMA models easier, we propose new VARMA representations which can be viewed as simple extensions of the VAR model. To introduce these, we first review another identified representation, the *final equation form*, which will refer to as the *final AR equation form*, under which the AR operator is scalar; see Zellner and Palm (1974), Hannan (1976), Wallis (1977), and Lütkepohl (1993).

Definition 3.1 (Final AR equation form) *We say that the VARMA model* (2.1) *is in final AR equation form if* $\Phi(L) = \varphi(L)I_K$*, where* $\varphi(L) = 1 - \varphi_1 L - \cdots - \varphi_p L^p$ *is a scalar polynomial with* $\varphi_p \neq 0$*.*

The Final AR equation form stands on its own, just like the other three representations we present below, in that we can directly assume that the series Y_t follows one of these representations. That being said, the intuition behind this formulation proceeds as follows. Take the standard representation in (2.2). By standard linear algebra, we have

$$\Phi^\star(L)\Phi(L) = \Phi(L)\Phi^\star(L) = \det\left[\Phi(L)\right] I_K \tag{3.4}$$

where $\Phi^\star(L)$ is the adjoint matrix of $\Phi(L)$. On multiplying both sides of (2.2) by $\Phi^\star(L)$, we get:

$$\det\left[\Phi(L)\right] Y_t = \Phi(L)^\star \Theta(L) U_t. \tag{3.5}$$

This representation may not be attractive for several reasons. First, it is quite far from usual VAR models by excluding lagged values of other variables in each equation; *e.g.*, the AR part of the first equation includes lagged values of y_{1t} but no lagged values of $y_{2t}, \ldots, y_{Kt}$. Further, the AR coefficients are the same in all the equations, which will require a polynomial with higher order $p\,K$. Second, the interaction between the different variables is modeled through the MA part of the model, which may be quite difficult to estimate.

Indeed, we propose here more convenient alternative representations. On multiplying both sides of (2.2) by $\Theta^\star(L)$, we get:

$$\Theta(L)^\star \Phi(L) Y_t = \det\left[\Theta(L)\right] U_t \tag{3.6}$$

where $\Theta(L)^\star$ is the adjoint matrix of $\Theta(L)$. We say VARMA models with the form (3.6) are in *final MA equation form*.

Definition 3.2 (Final MA equation form) *The VARMA representation* (2.1) *is said to be in final MA equation form if* $\Theta(L) = \theta(L)I_K$*, where* $\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$ *is a scalar operator with* $\theta_q \neq 0$*.*

The above form also raises a parsimony problem similar to the one associated with the final AR equation form. It is however possible to get a more parsimonious representation by looking

pohl and Claessen (1997), Poskitt (1992), Nsiri and Roy (1992, 1996), Lütkepohl and Poskitt (1996), Bartel and Lütkepohl (1998). A more general and in-depth discussion of identification of VARMA models can be found in Hannan and Deistler (1988, Chapter 2).

at common structures across equations. Suppose there are common roots across rows for some columns of $\Theta(L)$, so that starting from (2.2) we can write:

$$\Phi(L)Y_t = \bar{\Theta}(L)D(L)U_t\,, \tag{3.7}$$
$$\bar{\Theta}^*(L)\Phi(L)Y_t = \det\left[\bar{\Theta}(L)\right]D(L)U_t\,, \tag{3.8}$$

where $D(L) = \mathrm{diag}[d_1(L), \ldots, d_K(L)]$ and $d_j(L)$ is a polynomial common to $\theta_{ij}(L)$, $\forall i = 1, \ldots, K$. We see that allowing non-equal diagonal polynomials in the moving average as in equation (3.8) may yield a more parsimonious representation than in (3.6). We call the representation (3.8) the *diagonal MA equation form* representation.

Definition 3.3 (Diagonal MA equation form) *The VARMA representation* (2.1) *is said to be in diagonal MA equation form if*

$$\Theta(L) = \mathrm{diag}[\theta_{ii}(L)] = I_K - \Theta_1 L - \cdots - \Theta_q L^q \tag{3.9}$$

where $\theta_{ii}(L) = 1 - \theta_{ii,1}L - \cdots - \theta_{ii,q_i}L^{q_i}$, $\theta_{ii,q_i} \neq 0$, *and* $q = \max_{1\leq i\leq K}(q_i)$.

This representation is interesting because contrary to the echelon form it is easy to specify. One does not need rules for the orders of the off-diagonal elements in the AR and MA operators. The fact that it can be seen as a simple extension of the VAR model is also appealing. Practitioners are comfortable using VAR models, so simply adding lags of u_{it} to equation i is a natural extension of the VAR model which could give a more parsimonious representation. It also has the advantage of imposing a relatively simple structure on the MA polynomials, the component which complicates the estimation, rather than the AR component as in the final AR equation form. Notice that in VARMA models, it is not necessary to include lags of all the innovations $u_{1t}, \cdots, u_{Kt}$ in every equation. This could lead practitioners to consider VARMA models if it is combined with a simple regression-based estimation method. For this representation to be useful, it needs to be identified. This is demonstrated in Theorem **3.8** below under the following assumptions and using Lemma **3.7** below.

Assumption 3.4 *The matrices* $\Phi(z)$ *and* $\Theta(z)$ *have the following form:*

$$\Phi(z) = I_K - \Phi_1 z - \cdots - \Phi_p z^p\,, \quad \Theta(z) = I_K - \Theta_1 z - \cdots - \Theta_q z^q\,.$$

Assumption 3.5 $\Theta(z)$ *is diagonal:*

$$\Theta(z) = \mathrm{diag}\left[\theta_{ii}(z)\right]$$

where $\theta_{ii}(z) = 1 - \theta_{ii,1}z - \cdots - \theta_{ii,q_i}z^{q_i}$ *and* $\theta_{ii,q_i} \neq 0$.

Assumption 3.6 *For each* $i = 1, \ldots, K$, *there are no roots common to* $\Phi_{i\bullet}(z)$ *and* $\theta_{ii}(z)$, i.e. *there is no value* $z^\star$ *such that* $\Phi_{i\bullet}(z^\star) = 0$ *and* $\theta_{ii}(z^\star) = 0$.

Lemma 3.7 *Let* $[\Phi(z), \Theta(z)]$ *and* $\left[\bar{\Phi}(z), \bar{\Theta}(z)\right]$ *be two pairs of polynomial matrices which satisfy Assumptions* **3.4** *to* **3.6**. *If* R_0 *is a positive constant such that* $\Phi(z)^{-1}\Theta(z) = \bar{\Phi}(z)^{-1}\bar{\Theta}(z)$ *for* $0 \leq |z| < R_0$, *then*

$$\Phi(z) = \bar{\Phi}(z) \text{ and } \Theta(z) = \bar{\Theta}(z), \ \forall z\,.$$

The proofs of this lemma and other propositions are available in the online appendix. Assumptions **3.4** to **3.6** and the conditions in Lemma **3.7** allow $\det[\Phi(z)]$ and $\det[\Theta(z)]$ to have roots on or inside the unit circle $|z| = 1$. It should be noted that Assumption **3.6** is weaker than the hypothesis that $\det[\Phi(L)]$ and $\det[\Theta(L)]$ have no common roots, which would be a generalization of the usual identification condition for ARMA models. If invertibility is assumed, condition

$$\Phi(z)^{-1}\Theta(z) = \bar{\Phi}(z)^{-1}\bar{\Theta}(z) \tag{3.10}$$

in Lemma **3.7** can be replaced by

$$\Theta(z)^{-1}\Phi(z) = \bar{\Theta}(z)^{-1}\bar{\Phi}(z)\,. \tag{3.11}$$

We will now show that the diagonal MA representation is unique.

Theorem 3.8 (Identification of diagonal MA equation form representation) *Suppose* Y_t *satisfies the VARMA model in* (2.1) *along with Assumptions* **2.1**, **3.4 -3.6**. *If the VARMA model is in diagonal MA equation form, then it is identified.*

Similarly, we can demonstrate that the final MA equation form representation is identified under the following assumption.

Assumption 3.9 *There are no roots common to* $\Phi(z)$ *and* $\theta(z)$, i.e. *there is no value* $z^\star$ *such that* $\Phi(z^\star) = 0$ *and* $\theta(z^\star) = 0$.

Theorem 3.10 (Identification of final MA equation form representation) *Suppose* Y_t *satisfies the VARMA model in* (2.1) *along with Assumptions* **2.1**, **3.4** *and* **3.9**. *If the VARMA model is in final MA equation form, then it is identified.*

A strong appeal of the diagonal and final MA equation form representations is that it is easy to get the equivalent (in term of autocovariances) invertible MA representation of a non-invertible representation. With ARMA models, we simply have to invert the roots of the MA polynomial which are inside the unit circle and adjust the standard deviation of the innovations (divide it by the square of these roots): see Hamilton (1994, Section 3.7). The same procedure could be applied to VARMA models in diagonal or final MA equation form.

For VARMA representations where no particular simple structure is imposed on the MA part, at the moment we are not aware of an algorithm to go from the non-invertible to the invertible representation though theoretically this invertible representation exist and is unique as long as $\det[\Theta(z)] \neq 0$ for $|z| = 1$; see Hannan and Deistler (1988, chapter 1, section 3). So it might be troublesome to use a nonlinear optimization with these VARMA representations since it is not clear how to go from the non-invertible to an invertible representation.

We can also consider the following natural generalization of the final AR equation form, where we simply replace the scalar AR operator by a diagonal operator.

Definition 3.11 (Diagonal AR equation form) *The VARMA representation* (2.1) *is said to be in diagonal AR equation form if*

$$\Phi(L) = \mathrm{diag}[\varphi_{ii}(L)] = I_K - \Phi_1 L - \cdots - \Phi_p L^p \tag{3.12}$$

where $\varphi_{ii}(L) = 1 - \varphi_{ii,1}L - \cdots - \varphi_{ii,p_i}L^{p_i}$ *and* $p = \max_{1\leq i\leq K}(p_i)$.

Assumption 3.12 *For each* $i = 1, \ldots, K$*, there are no roots common to* $\varphi_{ii}(z)$ *and* $\Theta_{i\bullet}(z)$, i.e. *there is no value* $z^\star$ *such that* $\varphi_{ii}(z^\star) = 0$ *and* $\Theta_{i\bullet}(z^\star) = 0$.

Theorem 3.13 (Identification of diagonal AR equation form representation) *Let the VARMA model be defined by equation* (2.1) *and let Assumptions* **2.1**, **3.4** *and* **3.12** *hold. If the VARMA model is in diagonal AR equation form, then it is identified.*

From Theorem **3.8**, we can see that one way to ensure identification consists in imposing constraints on the MA operator. This can be viewed as an alternative to the approach developed by Hannan (1971, 1976), where identification is obtained by restricting the autoregressive part to be lower triangular with $deg[\varphi_{ik}(L)] \leq deg[\varphi_{ii}(L)]$ for $k > i$, or to the final AR equation form where $\Phi(L)$ is scalar. It may be more interesting to impose constraints on the moving average part instead because it is this part which causes problems in the estimation of VARMA models. Other identified representations which do not have a simple MA operator include the reversed echelon canonical form [see Poskitt (1992)] where the rows of the VARMA model in echelon form are permuted so that the Kronecker indices are ordered from smallest to largest, and the scalar component model [see Tiao and Tsay (1989)] where contemporaneous linear transformations of the vector process are considered. A general treatment of algebraic and topological structure underlying VARMA models is given in Hannan and Kavalieris (1984). For the maximum likelihood estimation of linear state space models, data driven local coordinates are often used; see *e.g.* Ribarits, Deistler, and McKelvey (2004) and McKelvey, Helmersson, and Ribarits (2004). Theorem 2.7.1 in Hannan and Deistler (1988) provides general conditions for a class of ARMAX models to be identifiable. These conditions are satisfied by the proposed representations.

4. Estimation

We next introduce elements of notation for the parameters of our model. First, irrespective of the VARMA representation employed, we split the whole vector of parameters γ in two parts γ_1 (the parameters for the AR part) and γ_2 (MA part):

$$\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}. \tag{4.1}$$

For a VARMA model in diagonal MA equation form,

$$\gamma_1 = \left[\varphi_{1\bullet,1}, \ldots, \varphi_{1\bullet,p}, \ldots, \varphi_{K\bullet,1}, \ldots, \varphi_{K\bullet,p}\right]', \tag{4.2}$$

$$\gamma_2 = \left[\theta_{11,1}, \ldots, \theta_{11,q_1}, \ldots, \theta_{KK,1}, \ldots, \theta_{KK,q_K}\right]', \tag{4.3}$$

while for a VARMA model in final MA equation form,

$$\gamma_2 = \left[\theta_1, \ldots, \theta_q\right]'.$$

For VARMA models in diagonal AR equation form, we simply invert γ_1 and γ_2:

$$\gamma_1 = \left[\varphi_{11,1}, \ldots, \varphi_{11,p_1}, \ldots, \varphi_{KK,1}, \ldots, \varphi_{KK,p_K}\right]', \tag{4.4}$$

$$\gamma_2 = \left[\theta_{1\bullet,1}, \ldots, \theta_{1\bullet,q}, \ldots, \theta_{K\bullet,1}, \ldots, \theta_{K\bullet,q}\right]', \tag{4.5}$$

while for a VARMA model in final AR equation form,

$$\gamma_1 = \left[\varphi_1, \ldots, \varphi_p\right]'. \tag{4.6}$$

The estimation method involves three steps. The observations go from $t = 1, \ldots, T$.

Step 1. Estimate a VAR(n_T) model to approximate the VARMA(p, q) model, and compute the residuals

$$\hat{U}_t = Y_t - \sum_{j=1}^{n_T} \hat{\Pi}_j^{(n_T)} Y_{t-j} \quad \text{for} \quad t = n_T + 1, \ldots, T. \tag{4.7}$$

Step 2. Using the residuals $\hat{U}_t$, compute an estimate of the covariance matrix of U_t, $\hat{\Sigma}_U = \frac{1}{T}\sum_{t=n_T+1}^{T} \hat{U}_t \hat{U}_t'$ and estimate by GLS the multivariate regression

$$\Phi(L) Y_t = [\Theta(L) - I_K]\hat{U}_t + e_t \tag{4.8}$$

to get estimates $\tilde{\Phi}(L)$ and $\tilde{\Theta}(L)$ of $\Phi(L)$ and $\Theta(L)$. The estimator is

$$\tilde{\gamma} = \left[\sum_{t=l}^{T} \hat{Z}_{t-1}' \hat{\Sigma}_U^{-1} \hat{Z}_{t-1}\right]^{-1} \left[\sum_{t=l}^{T} \hat{Z}_{t-1}' \hat{\Sigma}_U^{-1} Y_t\right] \tag{4.9}$$

where $l = n_T + \max(p, q) + 1$. Setting

$$\mathbf{Y}_{t-1}^{(p)} = \left[y_{1,t-1}, \ldots, y_{K,t-1}, \ldots, y_{1,t-p}, \ldots, y_{K,t-p}\right], \tag{4.10}$$

$$\hat{\mathbf{U}}_{t-1}^{(q)} = \left[\hat{u}_{1,t-1}, \ldots, \hat{u}_{K,t-1}, \ldots, \hat{u}_{1,t-q}, \ldots, \hat{u}_{K,t-q}\right], \tag{4.11}$$

$$\mathbf{y}_{t-1}^{(k)} = \left[y_{k,t-1}, \ldots, y_{k,t-p_k}\right], \quad \hat{\mathbf{u}}_{t-1}^{(k)} = \left[\hat{u}_{k,t-1}, \ldots, \hat{u}_{k,t-q_k}\right], \tag{4.12}$$

the matrix $\hat{Z}_{t-1}$ for the various representations is:

$$\hat{Z}_{t-1}^{DMA} = \begin{bmatrix} \mathbf{Y}_{t-1}^{(p)} & \cdots & 0 & -\hat{\mathbf{u}}_{t-1}^{(1)} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Y}_{t-1}^{(p)} & 0 & \cdots & -\hat{\mathbf{u}}_{t-1}^{(K)} \end{bmatrix}, \tag{4.13}$$

$$\hat{Z}_{t-1}^{FMA} = \begin{bmatrix} \mathbf{Y}_{t-1}^{(p)} & \cdots & 0 & -\hat{\mathbf{u}}_{t-1}^{(1)} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathbf{Y}_{t-1}^{(p)} & -\hat{\mathbf{u}}_{t-1}^{(K)} \end{bmatrix}, \tag{4.14}$$

$$\hat{Z}_{t-1}^{DAR} = \begin{bmatrix} \mathbf{y}_{t-1}^{(1)} & \cdots & 0 & -\hat{\mathbf{U}}_{t-1}^{(q)} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{y}_{t-1}^{(K)} & 0 & 0 & -\hat{\mathbf{U}}_{t-1}^{(q)} \end{bmatrix}, \tag{4.15}$$

$$\hat{Z}_{t-1}^{FAR} = \begin{bmatrix} \mathbf{y}_{t-1}^{(1)} & -\hat{\mathbf{U}}_{t-1}^{(q)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t-1}^{(K)} & 0 & 0 & -\hat{\mathbf{U}}_{t-1}^{(q)} \end{bmatrix}, \tag{4.16}$$

where DMA, FMA, DAR and FAR respectively stand for Diagonal MA, Final MA, Diagonal AR and Final AR equation form.

Step 3. Using the second step estimates, we first form new residuals

$$\tilde{U}_t = Y_t - \sum_{i=1}^{p} \tilde{\Phi}_i Y_{t-i} + \sum_{j=1}^{q} \tilde{\Theta}_j \tilde{U}_{t-j} \tag{4.17}$$

initiating with $\tilde{U}_t = 0$ for $t \leq \max(p, q)$, and we define

$$X_t = \sum_{j=1}^{q} \tilde{\Theta}_j X_{t-j} + Y_t\,, \quad W_t = \sum_{j=1}^{q} \tilde{\Theta}_j W_{t-j} + \tilde{U}_t\,, \tag{4.18}$$

initiating with $X_t = W_t = 0$ for $t \leq \max(p, q)$. We also compute a new estimate of Σ_U, $\tilde{\Sigma}_U = \frac{1}{T}\sum_{t=l'}^{T} \tilde{U}_t \tilde{U}_t'$, with $l' = \max(p, q) + 1$. Then we regress $\tilde{U}_t + X_t - W_t$ on $\tilde{V}_{t-1}$ by GLS, with

$$\tilde{V}_t = \sum_{j=1}^{q} \tilde{\Theta}_j \tilde{V}_{t-j} + \tilde{Z}_t \tag{4.19}$$

where $\tilde{Z}_t$ is computed like $\hat{Z}_t$ from step 2 with $\tilde{U}_t$ replaced by $\hat{U}_t$. This yields

$$\hat{\gamma} = \left[\sum_{t=l'}^{T} \tilde{V}_{t-1}' \tilde{\Sigma}_U^{-1} \tilde{V}_{t-1}\right]^{-1} \left[\sum_{t=l'}^{T} \tilde{V}_{t-1}' \tilde{\Sigma}_U^{-1} [\tilde{U}_t + X_t - W_t]\right] \tag{4.20}$$

from which we extract the final estimates $\hat{\Phi}_i$ and $\hat{\Theta}_j$.

The properties of the above estimators are summarized in the following three theorems. Theorem **4.1** is a generalization of results from Lewis and Reinsel (1985) where convergence is demonstrated for mixing rather than i.i.d. innovations. We denote the Euclidean norm by $\|B\|^2 = \text{tr}(B'B)$.

Theorem 4.1 (VARMA first-step estimates) *Suppose Y_t satisfies the VARMA model in (2.1) along with Assumptions* **2.1**, **2.2** *and* **2.3**. *If $n_T/\log(T) \to \infty$ and ${n_T}^2/T \to 0$ as $T \to \infty$, then the estimators $\hat{\Pi}_j^{(n_T)}$ in (4.7) satisfy*

$$\sum_{j=1}^{n_T} \|\hat{\Pi}_j^{(n_T)} - \Pi_j\| = O_p(\sqrt{n_T/T}\,). \tag{4.21}$$

Theorem 4.2 (VARMA second-step estimates) *Under the assumptions of Theorem* **4.1**, *suppose the parameters of the model are identified. Then the second-step estimator $\tilde{\gamma}$ in (4.9) converge in probability to the true value γ, and*

$$\sqrt{T}\,(\tilde{\gamma} - \gamma) \xrightarrow[T\to\infty]{\text{d}} \mathcal{N}\big[0,\, J_{(2)}^{-1} I_{(2)} J_{(2)}^{-1}\big]$$

where

$$I_{(2)} = \sum_{j=-\infty}^{\infty} E\{\big[Z'_{t-1}\Sigma_U^{-1}U_t\big]\,\big[Z'_{t-1-j}\Sigma_U^{-1}U_{t-j}\big]'\}\,, \quad J_{(2)} = E\,\big[Z'_{t-1}\Sigma_U^{-1}Z_{t-1}\big]\,, \tag{4.22}$$

and Z_{t-1} is equal to the matrix $\hat{Z}_{t-1}$ where $\hat{U}_t$ is replaced by U_t. Further, if $\omega(j, m_T) = 1 - |j|/(m_T+1)$ with $m_T^4/T \to 0$ and $m_T \to \infty$ as $T \to \infty$, we have:

$$\hat{I}_{(2)} := \frac{1}{T}\sum_{j=-m_T}^{m_T} \omega(j, m_T) \sum_{t=l+|j|}^{T} \big[\hat{Z}'_{t-1}\hat{\Sigma}_U^{-1}\tilde{U}_t\big]\,\big[\hat{Z}'_{t-1-j}\hat{\Sigma}_U^{-1}\tilde{U}_{t-j}\big]' \xrightarrow[T\to\infty]{\text{p}} I_{(2)}\,, \tag{4.23}$$

$$\hat{J}_{(2)} := \frac{1}{T}\sum_{t=l}^{T} \hat{Z}'_{t-1}\hat{\Sigma}_U^{-1}\hat{Z}_{t-1} \xrightarrow[T\to\infty]{\text{p}} J_{(2)}\,. \tag{4.24}$$

Theorem 4.3 (VARMA third-step estimates) *Under the assumptions of Theorem* **4.2**, *the third-step estimator $\hat{\gamma}$ in (4.20) converges in probability to the true value γ, and*

$$\sqrt{T}\,(\hat{\gamma} - \gamma) \xrightarrow[T\to\infty]{\text{d}} \mathcal{N}[0,\, J_{(3)}^{-1} I_{(3)} J_{(3)}^{-1}] \tag{4.25}$$

with

$$I_{(3)} = \sum_{j=-\infty}^{\infty} E\{\big[V'_{t-1}\Sigma_U^{-1}U_t\big]\,\big[V'_{t-1-j}\Sigma_U^{-1}U_{t-j}\big]'\}\,, \quad J_{(3)} = E\,\big[V'_{t-1}\Sigma_U^{-1}V_{t-1}\big] \tag{4.26}$$

and V_{t-1} is equal to the matrix $\tilde{V}_{t-1}$ where $\tilde{U}_t$ is replaced by U_t. Further, if $m_T^4/T \to 0$ and $m_T \to \infty$ as $T \to \infty$, we have:

$$\hat{I}_{(3)} = \frac{1}{T} \sum_{j=-m_T}^{m_T} \omega(j, m_T) \sum_{t=l'+|j|}^{T} \left\{ \tilde{V}'_{t-1} \tilde{\Sigma}_U^{-1} \bar{U}_t \right\} \left\{ \tilde{V}'_{t-1-j} \tilde{\Sigma}_U^{-1} \bar{U}_{t-j} \right\}' \xrightarrow[T\to\infty]{\text{p}} I_{(3)} \tag{4.27}$$

$$\hat{J}_{(3)} = \frac{1}{T} \sum_{t=l'}^{T} \tilde{V}'_{t-1} \tilde{\Sigma}_U^{-1} \tilde{V}_{t-1} \xrightarrow[T\to\infty]{\text{p}} J_{(3)}\,, \tag{4.28}$$

where $\omega(j, m_T) = 1 - |j|/(m_T + 1)$ and $\bar{U}_t$ are the filtered residuals computed with $\hat{\gamma}$.

Proofs of these theorems can be found in the online appendix (part D). In Step 1, if the order n_T of the VAR grows at the appropriate rate, the approximation error from estimating a finite order VAR when the truth is VAR(∞) will shrink to zero with the sample size.[3] In Step 2, the measurement error problem in using residuals instead of the true innovations as regressors asymptotically disappears which makes the regression coefficients consistent estimates. Replacing the innovation U_t with a residual induces a MA structure in the measurement error. Step 3 builds an MA filter to correct it and improve the efficiency of the estimator.

Notice the simplicity of this estimation method. Only three regressions are needed so we can avoid all the caveats associated with nonlinear optimizations. This is a central problem with VARMA models where one typically deals with a high number of parameters and numerical convergence may be hard to obtain. This is especially important when we consider the fact that the asymptotic distribution of our estimators, on which we would base our inference, may be a bad approximation to the finite-sample distribution in high-dimensional dynamic models. Because of this, an estimation procedure which only requires linear methods is interesting since it suggests that simulation-based procedures – bootstrap techniques for example – should be used, something that would be impractical if the estimation is based on non-linear optimizations.

It is also important to mention that this procedure is not specific to the representations considered in this work. The expressions can be easily adapted to other identified representation. Since our estimation method is only based on linear regressions, we can afford to use a less parsimonious representation, whereas it is critical to keep the number of parameters to a minimum when a nonlinear estimation method is used. An advantage of the proposed Diagonal MA and Final MA representations is that if the second step estimates do not correspond to an invertible MA representation (roots inside the unit circle), it is easy to get the corresponding invertible representation to be able to perform Step 3.[4]

A review of existing estimation methods for VARMA models is included in the online appendix (part B). With the exception of Dias and Kapetanios (2018) who propose an Iterated OLS (IOLS) estimator where we iterate the second step of our estimator until the convergence, all the articles reviewed consider data generating processes with innovations that are either i.i.d. or at a minimum

[3] Our experience is that the specific value of n_T does not matter much as long as it is large enough to give residuals that appear white noise.

[4] See the comments after Theorem **3.10**.

form a m.d.s. Here, however, we supply a distributional theory which holds under much weaker assumptions. This allow us to study a broader class of models, *e.g.* temporally aggregated processes, marginalized processes, weak representation of nonlinear models. For example, the test of Chen, Choi, and Escanciano (2017) for fundamental VMA representations is specifically based on allowing the innovations U_t to not be a m.d.s. Our results are based on the assumption that the models are identified. We leave for future work the analysis of our estimator in the context of weak identification as in Andrews and Cheng (2012).

We can ask ourselves what is the cost of not doing the nonlinear estimation. For a given sample size we will certainly lose some efficiency because of the first step estimation. We can nonetheless compare the asymptotic variance matrix of our estimator with the corresponding nonlinear estimator. We have the following results.

Theorem 4.4 (Efficiency of third step estimates under Gaussianity) *Under the assumptions of Theorem* **4.3***, suppose that the innovations U_t are i.i.d. Gaussian. Then, the asymptotic variance of the third stage estimator is equal to the asymptotic variance of the maximum likelihood estimator.*

Theorem 4.5 (Efficiency of third step estimates under weak innovations) *Under the assumptions of Theorem* **4.3***, the asymptotic variance of the third stage estimator is equal to the asymptotic variance of the quasi-maximum likelihood estimator.*

We can see that the asymptotic variance of our three-step linear estimator is the same as nonlinear least squares if the innovations are only uncorrelated or maximum likelihood if they are i.i.d. Gaussian. To get a feel for the loss of efficiency in finite samples due to replacing the true innovations by residuals from a long VAR we performed Monte Carlo simulations and report the results in section 6. We can interpret the third step as a Newton-Raphson step in the minimization of the sum of squared residuals $\sum_{t=1}^{T} U_t' U_t$.[5] It can also be viewed as a GLS correction induced by the MA structure in the error we make, when the true error term U_t is replaced by the first step residual $\hat{U}_t$ [see Reinsel, Basu, and Yap (1992)].

Our proposed estimator of the variance matrices in the asymptotic distribution of the second and third-step estimators are based on the results of Newey and West (1987), which require the choice of a bandwidth m_T. Beyond the rate requirement to guarantee the consistency of the estimators, practitioners can use recent results in Lazarus, Lewis, Stock, and Watson (2018) to choose a value for m_T. They favor using a larger bandwidth than typically recommended, $m_T = 1.3T^{1/2}$ is one of their recommendation. Different kernel could also be used; see Lazarus, Lewis, and Stock (2019).

5. Order selection

We still have unknowns in our model, the orders of the AR and MA operators. If no theory specifies these parameters, we have to use a statistical procedure to choose them. We propose the following information criterion method to choose the orders for VARMA models in the different identified

[5] See the proof of Theorem **4.5** in the online Appendix. We thank a referee for pointing this out to us.

representations proposed in Section 3. In the second step of the estimation, we compute for all $p \leq P$ and $q \leq Q$ the following information criterion:

$$DP(p,\, q,\, T) := \log(\det \tilde{\Sigma}_U) + \dim(\gamma)\frac{(\log T)^{1+\delta}}{T} \tag{5.1}$$

for some $\delta > 0$. We then choose $\hat{p}$ and $\hat{q}$ as the set which minimizes the information criterion. We assume that the upper bound P and Q on the orders of the AR and MA part are bigger than the true values of p and q (or that they slowly grow with the sample size). The properties of $\hat{p}$ and $\hat{q}$ are summarized in the following theorem.[6]

Theorem 5.1 (Estimation of the orders p and q in VARMA models) *Under the assumptions of Theorem* **4.2**, *suppose that* $0 \leq p \leq P$ *and* $0 \leq q \leq Q$. *If* $n_T = O\big((\log T)^{1+\delta_1}\big)$ *from some* $\delta_1 > 0$ *with* $\delta_1 < \delta$, *and* $(\hat{p}, \hat{q})$ *minimize* $DP(p,\, q,\, T)$, *then* $\hat{p}$ *and* $\hat{q}$ *converge in probability to their true values.*

In practice, this procedure can lead to a search over too many models for the diagonal representations. A valid alternative is to search for the true orders by proceeding equation by equation. In the second step of the estimation, rather than a simultaneous estimation, just perform univariate regressions. For a VARMA model in diagonal MA equation form, we estimate the regressions

$$y_{it} = \sum_{j=1}^{p_i}\sum_{k=1}^{K}\varphi_{ik,j}y_{k,t-j} - \sum_{j=1}^{q_i}\theta_{ii,j}\hat{u}_{i,t-j} + e_{it} \tag{5.2}$$

for $i = 1,\, \ldots\, ,\, K$, while for a VARMA models in diagonal AR equation form, we consider

$$y_{it} = \sum_{j=1}^{p_i}\varphi_{ii,j}y_{i,t-j} - \sum_{j=1}^{q_i}\sum_{k=1}^{K}\theta_{ik,j}\hat{u}_{k,t-j} + e_{it}\,. \tag{5.3}$$

We then chose $\hat{p}_i$ and $\hat{q}_i$ as the orders which minimize the following information criterion:

$$\log(\hat{\sigma}_i^2) + g(p_i, q_i)\frac{(\log T)^{1+\delta}}{T} \tag{5.4}$$

where $\delta > 0$ and $g(p_i, q_i) = p_i\, K + q_i$ or $g(p_i, q_i) = p_i + q_i\, K$ for the diagonal MA or AR equation form representation respectively. The global order for the autoregressive operator is then $\hat{p} = \max(\hat{p}_1,\, \ldots\, ,\, \hat{p}_K)$ for the diagonal MA representation, and similarly $\hat{q} = \max(\hat{q}_1,\, \ldots\, ,\, \hat{q}_K)$ for the diagonal AR representation. We see that this equation by equation selection procedure is not only easier to apply, but can lead to more parsimonious representations by identifying rows of zeros coefficients in Φ_j or Θ_j.

Theorem 5.2 (Estimation of the order p and q in diagonal VARMA models) *Under the assumption of Theorem* **5.1**, *if the VARMA model is in either the diagonal MA or AR equation form*

[6] Convergence is pointwise to the assumed true value of the orders p and q.

and the orders are chosen by minimizing $DP(p,\, q,\, T)$ *in* (5.4)*, then* $\hat{p}_i$ *and* $\hat{q}_i$*,* $i = 1, \ldots, K$*, converge in probability to their true value.*

The criterion $DP(p,\, q,\, T)$ is a generalization of the information criterion proposed by Hannan and Rissanen (1982). However, these authors later recognized that this criterion should be modified to yield consistent estimates of the orders p and q. The original criterion was

$$\log \tilde{\sigma}^2 + (p+q)\frac{(\log T)^\delta}{T} \tag{5.5}$$

with $\delta > 0$. Hannan and Rissanen (1983) showed that $\tilde{\sigma}^2 - \frac{1}{T}\sum_{t=1}^{T} u_t^2$ is $O_p(n_T T^{-1})$, rather than $O_p(T^{-1})$, so that the penalty $(\log T)^\delta/T$ is not strong enough. Two possible modifications are then considered. The first one is relatively simple and consists in taking $(\log T)^{1+\delta}$ instead of $(\log T)^\delta$ in the information criterion, so that the penalty on $p+q$ dominates $\log \hat{\sigma}^2$ in the criterion. The second one – which they favor and use in later work [see Hannan and Kavalieris (1984)] – considers a modification of the first step of the procedure: instead of taking $n_T = O(\log T)$, another information criterion is used to choose the order of the long autoregression, and the whole procedure is iterated picking potentially different values p and q at every iteration. A similar approach is also proposed by Poskitt (1987). In this work, we prefer the first solution in order to keep the procedure as simple as possible.

The online appendix (part C) contains a review of existing methods for specifying VARMA models under different identified representations, as well as testing if the innovations are uncorrelated by analyzing the residuals.

6. Monte Carlo simulations

To illustrate the performance of our estimation method we ran simulations where the innovations are not independent nor a m.d.s. but merely uncorrelated. We simulate weak VARMA processes by directly simulating weak innovations, from which we build the simulated series. Different approaches can be taken to simulate weak VARMA processes. Temporal aggregation of a strong VARMA process with innovations that have a skewed marginal distribution [see Francq and Zakoïan (1998, Section 2.2.1)] or of a strong GARCH process [see Drost and Nijman (1993) and Francq and Zakoïan (2000)] can generate weak innovations. For our simulations we instead employ a specification presented in Boubacar Maïnassara and Saussereau (2018, equation 27):

$$U_t = \begin{pmatrix} \varepsilon_{1,t}^2 \varepsilon_{2,t-1}\varepsilon_{1,t-2} \\ \varepsilon_{2,t}^2 \varepsilon_{1,t-1}\varepsilon_{2,t-2} \end{pmatrix} \tag{6.1}$$

where ε_t is i.i.d. $N(0, I_2)$. The process U_t is uncorrelated over time but is not a m.d.s. All the simulated models are bivariate so the results are easier to analyze. We performed 10,000 simulations for each model.

In these examples, because the innovations are not a m.d.s., we cannot do maximum likelihood. We instead employ nonlinear generalized least-squares (GLS), *i.e.* we minimize the nonlinear least

squares, compute an estimate of the variance matrix of the innovations and then do nonlinear GLS. We did not iterate this procedure, partly to reduce the estimation time in our Monte Carlo study, partly because there is no asymptotic gain in iterating.

In these simulations the sample size is 250 observations, which represent about 20 years of monthly data, a reasonable sample size for macroeconomic data. Table 1 gives results for a VARMA model in final MA equation form [VARMA$(1,1)$], while results for VARMA models in diagonal MA equation form are given in Tables 2 and 3 [VARMA$(1,1)$ with $q = (1,1)$ and VARMA(2,1) with $q = (1,1)$ respectively]. We present the results (mean, standard deviations, root mean square error, 5% quantile, 95% quantile and median) for the second (when the number of parameters does not exceed five) and third-step estimates, and the nonlinear GLS estimates (using the true value of the parameters as initial values).

From looking at the RMSE, a first thing to notice is that there can be sizable improvement in doing the third step. Some of the third-step RMSEs in Tables 1 and 2 are more than 50% smaller than for the second step. This is an interesting observation considering that the third step basically involve only one extra regression. Comparing the third-step RMSEs and the RMSEs for the nonlinear GLS estimates, we see that the former are usually no more than 15% bigger. This is also an interesting observation. The cost of avoiding a numerical optimization, which can become quite challenging as the number of time series or the orders of the operators increase, appears to be small.

In the top part of these tables we also present the results for the selection of the order of the operators using our proposed information criterion. For models in final MA equation form, we have to select the orders p and q, and for models in diagonal MA equation, the selection is over p, q_1 and q_2. In Table 1, we see that for VARMA models in final MA equation form the most frequently chosen orders are the true ones, and the criterion will tend to pick a higher value for q than for p. This result might partially be skewed by the fact that the simulated models have a highly persistent moving average ($\theta_1 = 0.9$). For VARMA models in diagonal equation form (Tables 2 and 3), we get similar results, the orders selected with the highest frequency are the true ones. In these simulations, we set $\delta = 0.5$ for the information criterion. Our experience is that this value, or something smaller like $\delta = 0.25$ works well.

7. Application to a macroeconomics model of the U.S. monetary policy

Our application is based on McMillin (2001) who compare numerous identification restrictions for the structural effects of monetary policy shocks using the same dataset as Bernanke and Mihov (1998).[7] The series are plotted in Figure 1. One of the model studied is a VAR applied to the first difference of the series, in order, *gdpm*, (*psscom-pgdpm*), *fyff*, *nbrec1*, *tr1*, *psscom*. With an

[7]The dataset consist of the log of the real GDP (*gdpm*), total bank reserves (*tr1*), nonborrowed reserves (*nbrec1*), federal funds rate (*fyff*), log of the GDP deflator (*pgdpm*), log of the Dow-Jones index of spot commodity prices (*psccom*). These are monthly data and cover the period January 1962 to December 1996. The monthly data for real GDP and the GDP deflator were constructed by state space methods, using a list of monthly interpolator variables and assuming that the interpolation error is describable as an AR(1) process. Both total reserves and nonborrowed reserves are normalized by a 36-month moving average of total reserves.

argument based on Keating (2002), the author state that using this ordering of the variables the Cholesky decomposition, based on long-run macroeconomic restrictions, which are described in an appendix, of the variance matrix of the innovations will identify the structural effects of the policy variable *nbrec1* without imposing any contemporaneous restrictions among the variables. Since the model is in first difference, the impulse-response at a given order is the cumulative shocks up to that order.

By fitting a VAR(12) to these series we get basically the same impulse-response functions and confidence bands as in McMillin (2001) They are plotted in Figure 3(a). The impulse-response function for the output and federal funds rate tends to zero as the order increases which is consistent with the notion that a monetary variable does not have a long term impact on real variables. The impulse response of the price level increases as we let the order grow and does not revert to zero.

We next estimate VARMA models. Of the four representations presented in Section 3, we present results for the Final MA representation, for which the information criterion picked orders $(\hat{p} = 3, \hat{q} = 10)$. The impulse-response functions for this model are plotted in Figure 3(b). The behavior of the impulse-response function for GDP, the federal funds rate and the price level from the VARMA models are similar to what we obtained with a VAR. The most notable differences are that the initial decrease in the federal funds rate is smaller (0.20 versus 0.32 percentage point) and the GDP is peaking earlier with Final MA. Furthermore, in Figure 3(c) we report the impulse-response functions for a VARMA model in echelon form.[8] They are somewhat similar to the ones from VAR and Final MA except for two important aspects. For output the IRF is decreasing instead of having a hump shape. We usually believe that it takes some time for monetary shocks to affect the real side of the economy. For the price level the IRF has pretty much leveled off after 24 months instead of still increasing.

It is not surprising that VAR and VARMA models are giving similar impulse-response functions since they are all ways of getting an infinite MA representation. What is more interesting is the comparison of the width of the confidence bands for the VAR and VARMA's impulse-response functions.[9] For GDP and the federal funds rate, we see that the bands are much smaller for the VARMA model in Final MA equation form, and they shrink more quickly as the horizon increases. The confidence bands for these two variables should be collapsing around their IRF since there should be no long-term effect of the policy variable so the uncertainty should decrease as the horizon increases. The situation is different for the price level. For this variable the confidence band grows with the order. Again this is not surprising, since we expect that a change in the non-borrowed reserves should have a long-term impact on the price level. With a non-dying impact it is natural that the uncertainty about this impact can grow as time passes.

The result that the confidence bands around IRFs can be shorter with a VARMA, in particular with our proposed representations, than with a VAR could be expected since these models are simple extensions of the VAR approach. The introduction of a simple MA operator allows the reduction of the required AR order so we can get more precise estimates, which translate into more precise

[8]The Kronecker indices are initially selected using the method in Poskitt (1992), which gives $[1, 1, 1, 1, 1, 1]$ using the BIC information criterion. From these, we looked at IRFs for Kronecker indices slightly different and found through visual inspection that Kronecker indices equal to $[1, 1, 1, 1, 2, 1]$ give the closest match to the ones from a VAR(12).

[9]For all models the confidence bands are computed by performing a parametric bootstrap using Gaussian innovations.

impulse-response functions.

As mentioned at the beginning of this section, another way of comparing the performance of VAR and VARMA models is to look at out-of-sample forecasts. For this exercise, using the same dataset as for the IRF above, we compared the RMSE for VAR models, VARMA models in Final and Diagonal MA equation form and VARMA models in echelon form. We use the first 264 observations to select the particular specification of each model. This leaves 120 observations for the out-of-sample evaluation of the forecasts. For the VAR model, we include a VAR(6) and a VAR(12). The former is the order selected by AIC, the latter is the order used by McMillin (2001) to model this dataset. For the VARMA in the Diagonal MA form, using our equation-by-equation information criterion (with $\delta = 0.5$) we get $\hat{p} = 1$ and $\hat{q} = [2, 1, 1, 1, 1, 1]$. For the Final MA form, our information criterion (with $\delta = 0.5$) selects $\hat{p} = 0$ and $\hat{q} = 1$. For the echelon form, using the method in Poskitt (1992) we select Kronecker indices $[1, 1, 1, 1, 1, 1]$ using BIC and $[2, 1, 1, 1, 1, 2]$ using AIC.[10] As we move through the out-of-sample period, the parameters of all the models are re-estimated and we compute forecasts up to 12 periods ahead.[11] The results are presented in Table 4. We can see that for all forecasting horizons, the Diagonal and Final MA representations achieve a lower RMSE than the VAR models and the VARMA models in echelon forms. For one step ahead forecasts, the Final MA representation has a lower RMSE than the Diagonal MA representation. For horizons 2 to 7 periods ahead, it's the Diagonal MA that Dominates Final MA. For horizons greater than 7 periods ahead, Final MA and Diagonal MA have similar RMSEs. Comparing the RMSEs for the VAR models, and for the VARMA models in echelon form, we see that the more parsimonious representations perform better.

8. Conclusion

In this paper, we have developed a modeling and estimation method which will make the use of VARMA models easier and more practical. We first introduced new identified VARMA representations, the final MA equation form and the diagonal MA equation form. These two representations are simple extensions of the class of VAR models where we add a simple MA operator, either a scalar or a diagonal operator. The addition of a MA part can give more parsimonious representations, yet the simple form of the MA operators does not introduce undue complications.

To simplify the estimation, we proposed relatively simple methods which only require linear regressions. For that purpose, we considered a generalization of the regression-based estimation method proposed by Hannan and Rissanen (1982) for univariate ARMA models. Our method has three steps. First, a long VAR is fitted to the data. Second, the lagged innovations in the VARMA model are replaced by the corresponding lagged residuals from the first step, and a regression is performed. Third, the data from the second step are filtered, and another regression is performed. We showed that the third-step estimators have the same asymptotic variance as their nonlinear counterpart (Gaussian maximum likelihood if the innovations are i.i.d., or generalized nonlinear least

[10] The number of AR and MA parameters is 92 for the representation selected by AIC and 72 for BIC.

[11] We estimate the VARs by OLS, the Diagonal and Final MA forms using our three-step methods. We use the SSM-MATLAB toolbox, see Gómez (2015), to estimate the echelon form by conditional likelihood. An alternative would be the method in Poskitt (2016).

squares if they are merely uncorrelated). In the non i.i.d. case, we consider strong mixing conditions, rather than the usual martingale difference sequence assumption. These minimal assumptions on the innovations broaden the class of models to which the method can be applied.

We also proposed a modified information criterion that gives consistent estimates of the orders of the AR and MA operators of the proposed VARMA representations. This criterion is to be minimized in the second step of the estimation method over a set of possible values for the different orders.

Monte Carlo simulation results indicates that the estimation method works well for small sample sizes, and the information criterion picks the true value of the order p and q most of the time. These results holds for sample sizes commonly used in macroeconomics, *i.e.* 20 years of monthly data or 250 sample points. To demonstrate the importance of using VARMA models to study multivariate time series, we compared the impulse-response functions and the out-of-sample forecasts generated by VARMA and VAR models when these models are applied to the dataset of macroeconomic time series used by Bernanke and Mihov (1998).

References

ANDREWS, D. W. K. (1984): "Non-Strong Mixing Autoregressive Processes," *Journal of Applied Probability*, 21(4), 930–934.

ANDREWS, D. W. K., AND X. CHENG (2012): "Estimation and inference with weak, semi-strong, and strong identification," *Econometrica*, 80(5), 2153–2211.

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

BERNANKE, B. S., AND I. MIHOV (1998): "Measuring Monetary Policy," *The Quarterly Journal of Economics*, 113(3), 869–902.

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

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

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

DEISTLER, M., AND E. J. HANNAN (1981): "Some Properties of the Parametrization of ARMA Systems with Unknown Order," *Journal of Multivariate Analysis*, 11, 474–484.

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

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

DROST, F. C., AND T. E. NIJMAN (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61(4), 909–927.

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

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

——— (2000): "Estimating Weak GARCH Representations," *Econometric Theory*, 16, 692–728.

GÓMEZ, V. (2015): "SSMMATLAB: A Set of MATLAB Programs for the Statistical Analysis of State Space Models," *Journal of Statistical Software, Articles*, 66(9), 1–37.

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

HANNAN, E. J. (1969): "The Identification of Vector Mixed Autoregressive- Moving Average Systems," *Biometrika*, 57, 223–225.

——— (1971): "The Identification Problem for Multiple Equation Sysytem with Moving Average Errors," *Econometrica*, 39, 751–766.

——— (1976): "The Identification and Parameterization of ARMAX and State Space Forms," *Econometrica*, 44(4), 713–723.

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

HANNAN, E. J., AND L. KAVALIERIS (1984): "Multivariate Linear Time Series Models," *Advances in Applied Probability*, 16, 492–561.

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

——— (1983): "Errata: "Recursive Estimation of Mixed Autoregressive-Moving Average Order"," *Biometrika*, 70(1), 303.

KEATING, J. W. (2002): "Structural Inference with Long-Run Recursive Empirical Models," *Macroeconomic Dynamics*, 6(2), 266–283.

LAZARUS, E., D. J. LEWIS, AND J. H. STOCK (2019): "The size-power tradeoff in HAR inference," working paper.

LAZARUS, E., D. J. LEWIS, J. H. STOCK, AND M. W. WATSON (2018): "HAR Inference: Recommendations for Practice," *Journal of Business & Economic Statistics*, 36(4), 541–559.

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

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

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

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

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

LÜTKEPOHL, H., AND D. S. POSKITT (1998): "Consistent Estimation of the Number of Cointegration Relations in a Vector Autoregressive Model," in *Econometrics in Theory and Practice: Festschrift for Hans Schneeweiß*, ed. by R. Galata, and H. Küchenhoff, pp. 87–100. Physica-Verlag HD, Heidelberg.

MCKELVEY, T., A. HELMERSSON, AND T. RIBARITS (2004): "Data driven local coordinates for multivariable linear systems and their application to system identification," *Automatica*, 40, 1629–1635.

MCMILLIN, W. D. (2001): "The Effect of Monetary Policy Shocks: Comparing Contemporaneous versus Long-Run Identifying Restrictions," *Southern Economic Journal*, 67(3), 618–636.

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

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

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

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

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

——— (2016): "Vector autoregressive moving average identification for macroeconomic modeling: A new methodology," *Journal of Econometrics*, 192, 468–484.

REINSEL, G. C., S. BASU, AND S. F. YAP (1992): "Maximum likelihood estimators in the multivariate autoregressive moving-average model from a generalized least squares viewpoint," *Journal of Time Series Analysis*, 13(2), 133–145.

RIBARITS, T., M. DEISTLER, AND T. MCKELVEY (2004): "An analysis of the parametrization by data driven local coordinates for multivariable linear systems," *Automatica*, 40, 789–803.

TIAO, G. C., AND R. S. TSAY (1989): "Model Specification in Multivariate Time Series," *Journal of the Royal Statistical Society, Series B*, 51(2), 157–213.

WALLIS, K. F. (1977): "Multiple time series analysis and the final form of econometric models," *Econometrica*, 45(6), 1481–1497.

ZELLNER, A., AND F. PALM (1974): "Time series analysis and simultaneous equation econometric model," *Journal of Econometrics*, 2(1), 17–54.

Table 1. Estimation of a weak final MA equation form VARMA$(1,1)$

The simulated model is a weak VARMA$(1,1)$ in final MA equation form with $\varphi_{11,1} = 0.5$, $\varphi_{21,1} = 0.7$, $\varphi_{12,1} = -0.6$, $\varphi_{22,1} = 0.3$ and $\theta_1 = 0.9$. The innovations are simulated according to equation (6.1). The sample size is 250, the length of the long VAR is $n_T = 15$, the number of repetition is 10,000. The parameter in the criterion is $\delta = 0.5$.

Frequencies of selection of $(\hat{p}, \hat{q})$ using the information criteria

$p \setminus q$	0	1	2	3	4	5
0	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.722	0.074	0.024	0.015	0.012
2	0.000	0.067	0.066	0.005	0.005	0.007
3	0.000	0.002	0.001	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000

	Value	Average	Std. dev.	RMSE	5%	95%	Median
Second step							
$\varphi_{11,1}$	0.5	0.423	0.064	0.101	0.312	0.518	0.427
$\varphi_{21,1}$	0.7	0.699	0.121	0.121	0.493	0.877	0.706
$\varphi_{12,1}$	-0.6	-0.587	0.112	0.113	-0.751	-0.392	-0.596
$\varphi_{22,1}$	0.3	0.233	0.070	0.097	0.112	0.339	0.238
θ_1	0.9	0.817	0.062	0.104	0.712	0.916	0.820
Third step							
$\varphi_{11,1}$	0.5	0.498	0.048	0.048	0.420	0.574	0.499
$\varphi_{21,1}$	0.7	0.694	0.095	0.095	0.534	0.834	0.699
$\varphi_{12,1}$	-0.6	-0.591	0.088	0.088	-0.717	-0.438	-0.598
$\varphi_{22,1}$	0.3	0.298	0.050	0.050	0.220	0.375	0.298
θ_1	0.9	0.892	0.048	0.049	0.807	0.963	0.897
NLLS							
$\varphi_{11,1}$	0.5	0.494	0.046	0.047	0.416	0.565	0.496
$\varphi_{21,1}$	0.7	0.700	0.093	0.093	0.541	0.842	0.705
$\varphi_{12,1}$	-0.6	-0.597	0.086	0.086	-0.724	-0.446	-0.603
$\varphi_{22,1}$	0.3	0.294	0.045	0.045	0.223	0.363	0.296
θ_1	0.9	0.892	0.042	0.043	0.832	0.937	0.898

Table 2. Estimation of a weak diagonal MA equation form VARMA$(1,1)$

The simulated model is a weak VARMA$(1,1)$ in diagonal MA equation form with $\varphi_{11,1} = 0.5$, $\varphi_{12,1} = -0.6$, $\varphi_{21,1} = 0.7$, $\varphi_{22,1} = 0.3$, $\theta_{1,1} = 0.9$ and $\theta_{1,1} = 0.7$. The innovations are simulated according to equation (6.1). The sample size is 250, the length of the long VAR is $n_T = 15$, the number of repetition is 10,000. The parameter in the criterion is $\delta = 0.5$.

Frequencies of selection of $(\hat{p}, \hat{q})$ using the information criteria.

(p, q_1, q_2)	Frequency
1,1,1	0.619
2,1,1	0.126
2,1,0	0.060
1,2,1	0.038
1,1,2	0.029
2,2,1	0.024
1,3,1	0.010

	Value	Average	Std. dev.	RMSE	5%	95%	Median
Second step							
$\varphi_{11,1}$	0.5	0.414	0.078	0.116	0.276	0.528	0.422
$\varphi_{21,1}$	0.7	0.695	0.122	0.122	0.490	0.875	0.704
$\varphi_{12,1}$	-0.6	-0.589	0.111	0.112	-0.754	-0.393	-0.599
$\varphi_{22,1}$	0.3	0.242	0.077	0.096	0.107	0.359	0.249
$\theta_{1,1}$	0.9	0.803	0.093	0.134	0.640	0.942	0.811
$\theta_{2,1}$	0.7	0.639	0.096	0.114	0.473	0.787	0.644
Third step							
$\varphi_{11,1}$	0.5	0.503	0.083	0.083	0.414	0.589	0.501
$\varphi_{21,1}$	0.7	0.692	0.107	0.107	0.516	0.843	0.701
$\varphi_{12,1}$	-0.6	-0.589	0.098	0.098	-0.729	-0.425	-0.598
$\varphi_{22,1}$	0.3	0.300	0.075	0.075	0.202	0.392	0.298
$\theta_{1,1}$	0.9	0.889	0.069	0.069	0.765	0.980	0.899
$\theta_{2,1}$	0.7	0.695	0.086	0.086	0.548	0.828	0.698
NLLS							
$\varphi_{11,1}$	0.5	0.495	0.046	0.046	0.418	0.564	0.497
$\varphi_{12,1}$	0.7	0.694	0.100	0.101	0.524	-0.840	0.702
$\varphi_{21,1}$	-0.6	-0.599	0.088	0.088	-0.734	0.447	-0.605
$\varphi_{22,1}$	0.3	0.296	0.050	0.050	0.214	0.373	0.298
$\theta_{1,1}$	0.9	0.897	0.054	0.054	0.823	0.959	0.902
$\theta_{2,1}$	0.7	0.699	0.072	0.072	0.579	0.811	0.702

Table 3. Estimation of a weak diagonal MA equation form VARMA$(2,1)$

The simulated model is a weak VARMA(2,1) in diagonal MA equation form with $\varphi_{11,1} = 0.9$, $\varphi_{12,1} = -0.5$, $\varphi_{21,1} = 0.3$, $\varphi_{22,1} = 0.1$, $\varphi_{11,2} = -0.1$, $\varphi_{12,2} = -0.2$, $\varphi_{21,2} = 0.1$, $\varphi_{22,2} = -0.15$, $\varphi_{1,1} = 0.9$, and $\varphi_{2,1} = 0.7$. The innovations are simulated according to equation (6.1). The sample size is 250, the length of the long VAR is $n_T = 15$, the number of repetition is 10,000. The parameter in the criterion is $\delta = 0.5$.

Frequencies of selection of $(\hat{p}, \hat{q})$ using the information criteria.

(p, q_1, q_2)	Frequency
2,1,1	0.677
2,1,0	0.080
2,1,2	0.066
2,2,1	0.030
3,1,1	0.029
3,1,0	0.022
2,1,3	0.012

	Value	Average	Std. dev.	RMSE	5%	95%	Median
Third step							
$\varphi_{11,1}$	0.9	0.894	0.057	0.057	0.800	0.980	0.896
$\varphi_{21,1}$	0.3	0.298	0.118	0.118	0.119	0.484	0.296
$\varphi_{12,1}$	-0.5	-0.501	0.202	0.202	-0.821	-0.179	-0.501
$\varphi_{22,1}$	0.1	0.096	0.175	0.175	-0.162	0.353	0.094
$\varphi_{11,2}$	-0.5	-0.492	0.081	0.081	-0.613	-0.361	-0.494
$\varphi_{21,2}$	0.1	0.093	0.173	0.173	-0.184	0.320	0.108
$\varphi_{12,2}$	-0.2	-0.204	0.105	0.105	-0.369	-0.038	-0.206
$\varphi_{22,2}$	-0.15	-0.137	0.176	0.176	-0.388	0.142	-0.151
$\theta_{1,1}$	0.9	0.889	0.059	0.060	0.785	0.972	0.896
$\theta_{2,1}$	0.7	0.686	0.142	0.143	0.434	0.907	0.699
NLLS							
$\varphi_{11,1}$	0.9	0.892	0.054	0.054	0.803	0.973	0.894
$\varphi_{21,1}$	0.3	0.298	0.118	0.118	0.117	0.490	0.295
$\varphi_{12,1}$	-0.5	-0.501	0.196	0.196	-0.809	-0.189	-0.501
$\varphi_{22,1}$	0.1	0.081	0.118	0.120	-0.126	0.253	0.092
$\varphi_{11,2}$	-0.5	-0.493	0.076	0.076	-0.608	-0.372	-0.495
$\varphi_{21,2}$	0.1	0.105	0.129	0.129	-0.111	0.304	0.110
$\varphi_{12,2}$	-0.2	-0.208	0.099	0.100	-0.367	-0.047	-0.209
$\varphi_{22,2}$	-0.15	-0.147	0.159	0.160	-0.389	0.118	-0.157
$\theta_{1,1}$	0.9	0.894	0.058	0.058	0.815	0.961	0.899
$\theta_{2,1}$	0.7	0.686	0.113	0.114	0.490	0.837	0.700

Table 4. RMSE for VAR and VARMA models

RMSE for out-of-sample forecasts for the dataset used in the computation of the impulse-response functions. The models are recursively estimated starting at observation 264 so the out-of-sample period consists of 100 observations. The specification of the models is determined at observation 264. VAR(6) is the number of lags chosen by AIC. VAR(12) is the order used to compute the impulse-response functions. Using our proposed information criterion with $\delta = 0.5$ the orders of the Diagonal MA (one equation at a time) and Final MA models are respectively $p = 1$, $q = (2, 1, 1, 1, 1, 1)$ and $p = 0$, $q = 1$. The Kronecker indices of the echelon form are determined according to the method in Poskitt (1992): $(1, 1, 1, 1, 1, 1)$ and $(2, 1, 1, 1, 1, 2)$ with respectively BIC or AIC.

Steps ahead	VAR(6)	VAR(12)	Diagonal MA	Final MA	$\text{Echelon}_{\text{BIC}}$	$\text{Echelon}_{\text{AIC}}$
1	0.1124	0.1181	0.1004	0.0866	0.1074	0.1034
2	0.1260	0.1275	0.0904	0.0917	0.1172	0.1357
3	0.1230	0.1299	0.0869	0.0922	0.1216	0.1524
4	0.1227	0.1263	0.0891	0.0924	0.1209	0.1454
5	0.1127	0.1195	0.0901	0.0911	0.1109	0.1748
6	0.1064	0.1115	0.0907	0.0914	0.1025	0.1558
7	0.0969	0.1030	0.0916	0.0918	0.1086	0.1446
8	0.0964	0.1049	0.0923	0.0923	0.1098	0.1357
9	0.0936	0.1026	0.0909	0.0910	0.1078	0.1162
10	0.0927	0.1057	0.0914	0.0915	0.1033	0.1178
11	0.0911	0.0998	0.0888	0.0888	0.0992	0.1151
12	0.0919	0.1074	0.0892	0.0893	0.0967	0.1217

Figure 1. Macroeconomic series

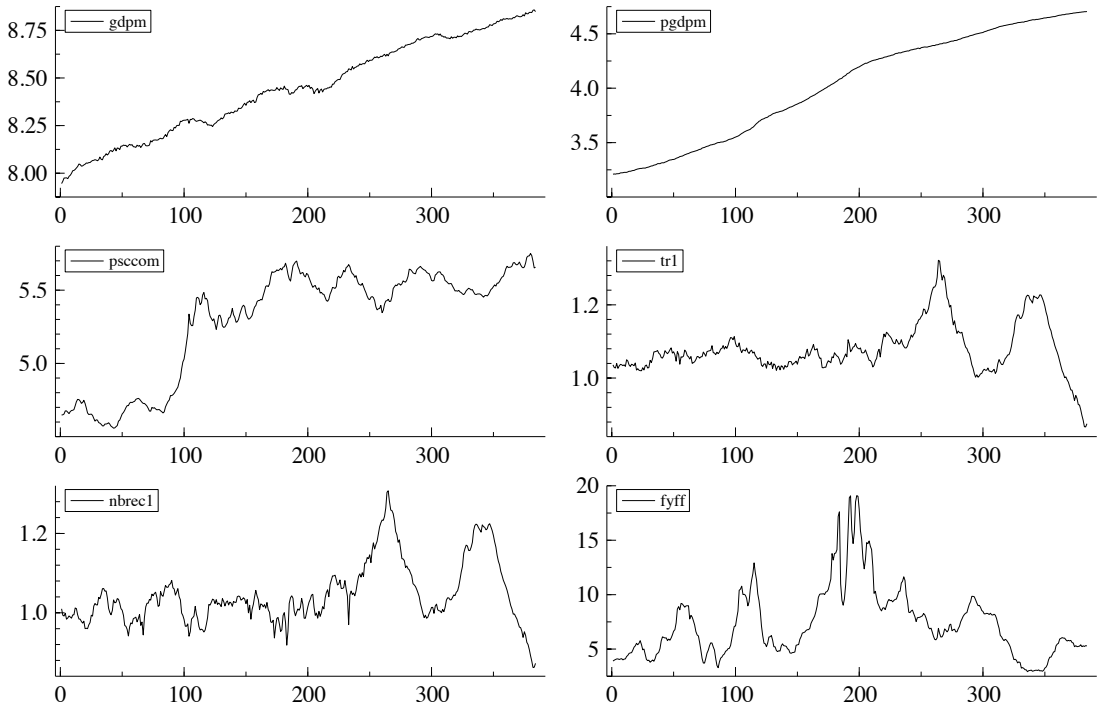
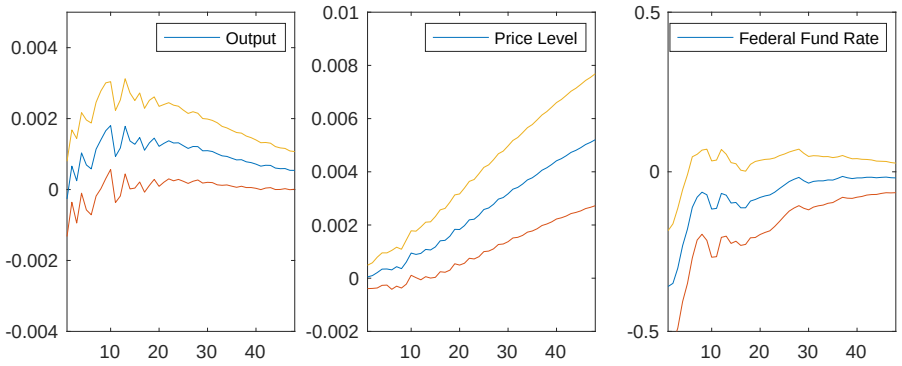
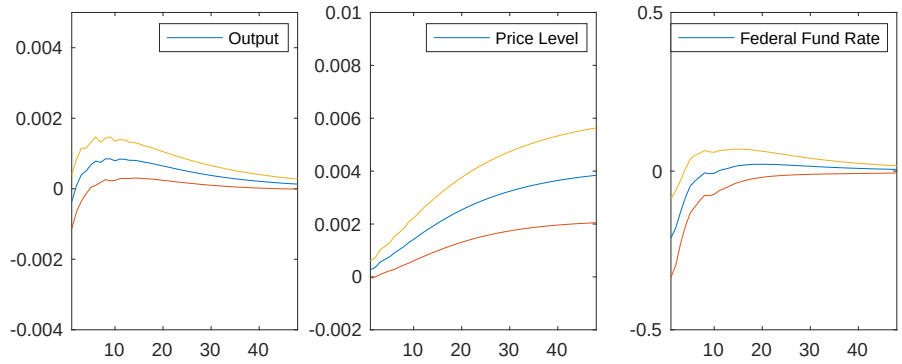


Figure 2. Impulse-response functions for VAR model
and VARMA model in final MA equation form.

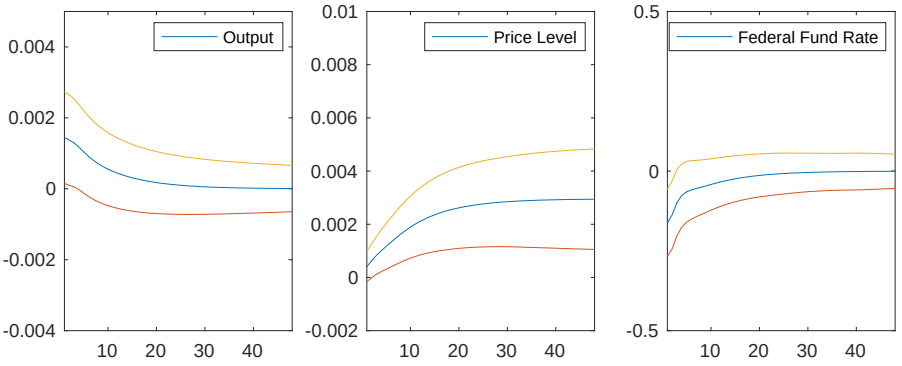
VAR(12), VARMA(3,10) in final MA equation form and VARMA in echelon form with Kronecker indices $[1, 1, 1, 1, 2, 1]$ are fitted to the first difference of the six time series. The confidence band represent a one standard deviation. The standard deviations are derived from a parametric bootstrap using Gaussian innovations.



(a) VAR(12)



(b) VARMA Final MA



(c) VARMA Echelon