# LINEAR METHODS FOR ESTIMATING VARMA MODELS WITH A MACROECONOMIC APPLICATION

Jean-Marie Dufour and Denis Pelletier
Département de sciences économiques, Université de Montréal,
C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7.
E-mail: jean.marie.dufour@umontreal.ca, denis.pelletier@umontreal.ca.

**Key Words:** Linear regression, VARMA, final equation form, information criterion, impulse-response function.

## 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, while VARMA models typically require nonlinear methods (such as maximum likelihood).

VAR models, however, have important drawbacks. First, they are typically much less parsimonious than VARMA models. Second, the family of VAR models is not closed under marginalization and temporal aggregation. If a vector satisfies a VAR model, subvectors do not typically satisfy VAR models (but VARMA models). Similarly, if the variables of a VAR process are observed at a different frequency, the resulting process is not a VAR process. In contrast, the class of (weak) VARMA models is closed under such operations. We say that a VARMA model is strong if the innovations are independent, and it is weak if they are merely uncorrelated.

It follows that VARMA models appear to be preferable from a theoretical viewpoint, but their adoption is complicated by estimation difficulties. Standard estimation methods for VARMA models (maximum likelihood, nonlinear least squares) 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.

In this paper, we consider the problem of estimating VARMA models by relatively simple methods which only require linear regressions. For that purpose, we consider a generalization of the regression-based estimation method proposed by Hannan and Rissanen (1982) for univariate ARMA models. In this method, lagged innovations are replaced by the corresponding residuals from a long autoregression and a regression is performed on the resulting equation. Extension of this method to VARMA models was also proposed by Koreisha and Pukkila (1989), but these authors did not provide a detailed asymptotic theory for their proposed extension.

Here, we first provide such a theory by showing that the linear-regression-based estimators are consistent and have the same asymptotic distribution as their nonlinear counterparts (maximum likelihood if the innovations are i.i.d., or nonlinear least squares if they are merely uncorrelated). In the non i.i.d. case, we consider strong mixing conditions [Doukhan (1995), Bosq (1998)], rather than the usual martingale difference sequence (m.d.s) assumption.

Second, in order to avoid identification problems, we study how the method can be applied to VARMA models in *final equation form*.

Thirdly, we suggest a modified information criterion to choose the orders $p$ and $q$ of the VARMA models in this representation.

Fourth, the method is applied to U.S. macroeconomic data previously studied by Bernanke and Mihov (1998).

We only summarize our results here. Detailed results, assumptions and proofs are presented in a discussion paper [Dufour and Pelletier (2002)].

The rest of the paper is organized as follow. Our notation and the VARMA representation are introduced in section 2. In section 3, we present the estimation method. In section 4, we describe the information criterion used for choosing the orders $p$ and $q$ of the VARMA models in final equation form. Section 5 contains results of Monte Carlo simulations which illustrate the properties of our method. Section 6 presents the macroeconomic application. Section 7 contains a few concluding remarks.

---

The authors thank Marine Carasco, John Galbraith, Nour Meddahi and Rui Castro 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. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds FCAR (Government of Québec).

## 2. VARMA representations

Consider the following $K$-variate zero mean VARMA$(p,\, q)$ model in standard representation:

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + U_t - \sum_{j=1}^{q} B_j U_{t-j} \qquad (1)$$

where $U_t$ is a sequence of uncorrelated random variables defined on some probability space $(\Omega, \mathcal{A}, \mathcal{P})$, $Y_t = [y_t(1),\, y_t(2),\, \ldots\, ,\, y_t(K)]'$ and $U_t = [u_t(1), u_t(2),\, \ldots\, ,\, u_t(K)]'$. We can also write the above equation using lag operators:

$$A(L)Y_t = B(L)U_t$$

where

$$\begin{aligned} A(L) &= I_K - A_1 L - \cdots - A_p L^p\,, &(2)\\ B(L) &= I_K - B_1 L - \cdots - B_q L^q\,. &(3)\end{aligned}$$

We also denote by $A_l(i,j)$ the element on row $i$ and column $j$ of the matrix $A_l$.

Let $H_t$ be the Hilbert space generated by $(Y_s,\ s < t)$. The process $(U_t)$ can be interpreted as the linear innovation of $Y_t$:

$$U_t = Y_t - E_L[Y_t|H_t].$$

We also assume that $Y_t$ is a strictly stationary and ergodic process, while the innovations $U_t$ have common variance $[Var[U_t] = \Sigma_U]$ and finite fourth moment ( $E[|u_t(i)|^{4+2\delta}] < \infty$ for some $\delta > 0$ ). The zero mean hypothesis is only used to simplify notation.

Assuming the process $Y_t$ is stable ( $\det[A(z)] \neq 0$ for $|z| \leq 1$ ) and invertible ( $\det[B(z)] \neq 0$ for $|z| \leq 1$ ), it can be represented as an infinite VAR

$$\Pi(L)Y_t = U_t,$$

where

$$\Pi(L) = B(L)^{-1}A(L) = I_K - \sum_{i=1}^{\infty} \Pi_i L^i\,,$$

or an infinite VMA

$$Y_t = \Psi(L)U_t,$$

where

$$\Psi(L) = A(L)^{-1}B(L) = I_K - \sum_{j=1}^{\infty} \Psi_j L^j.$$

The matrices $\Pi_i$ and $\Psi_j$ could be zero past a finite order if $\det[B(L)]$ or $\det[A(L)]$ respectively is a non-zero constant.

We need to impose a minimum of structure on the innovation process $U_t$ because stating that it is uncorrelated is not sufficient to get useful results. The typical assumption 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). Here, we shall consider weaker assumptions because we wish to broaden the class of models considered. We only assume that it satisfies a strong mixing condition [Doukhan (1995), Bosq (1998)]. Let $U_t$ be a strictly stationary process, then its $\alpha$-mixing coefficient of order $h$, $h \geq 1$, is defined as

$$\alpha(h) = \sup_{\substack{B\in\sigma(U_s, s\leq t)\\ C\in\sigma(U_s, s\geq t+h)}} |P(B\cap C) - P(B)P(C)|\,.$$

The strong mixing condition that we impose is

$$\sum_{h=1}^{\infty} \alpha(h)^{\delta/(2+\delta)} < \infty \quad \text{for some} \quad \delta > 0\,.$$

This is a fairly minimal condition that will be satisfied by many processes of interest.

It is important to note that we cannot work with the standard representation (1) because it is not identified [see Lütkepohl (1993)]. The identified representation that we use in this work is the final equation form.

**Definition 1 (Final Equation Form)** *The VARMA representation (1) is said to be in final equation form if $A(L) = a(L)I_K$, where $a(L) = 1 - a_1 L - \cdots - a_p L^p$ is a scalar operator with $a_p \neq 0$.*

This representation is quite similar to a univariate ARMA representation. An alternative representation that we could use is the echelon form, but we prefer the former for the sake of simplicity.

## 3. Estimation method

We now introduce elements of notation for the parameters of our model. Assuming a VARMA model in final equation form, we split the parameter vector $\gamma$ into two subvectors $\gamma_1$ (the parameters for the VAR operator) and $\gamma_2$ (the VMA operator):

$$\begin{aligned} \gamma &= [\gamma_1',\, \gamma_2']'\,,\\ \gamma_1 &= [a_1,\, \ldots\, ,\, a_p]'\,,\\ \gamma_2 &= vec\left[B^1,\, \ldots\, ,\, B^K\right]\,,\\ B^i &= [B_1(i,1),\, \ldots\, ,\, B_1(1,K),\, \ldots\, ,\\ &\qquad B_q(i,1),\, \ldots\, ,\, B_q(i,K)]'\,,\end{aligned}$$

where $vec$ is the operator that stack the columns of a matrix.

The estimation method considered involves three steps.

**Step 1.** *Estimate a VAR($n_T$) to approximate the VARMA(p, q) and recuperate the residuals*

$$\hat{U}_t = Y_t - \sum_{l=1}^{n_T} \hat{\Pi}_l^{n_T} Y_{t-l}$$

*with $T > 2 \times K \times n_T$.*

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

$$a(L)Y_t = (B(L) - I_K)\hat{U}_t + e_t$$

*to get estimates $\tilde{a}(L)$ and $\tilde{B}(L)$ of $a(L)$ and $B(L)$. This yields the estimator*

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

*with*

$$\begin{aligned} l &= n_T + \max(p,q) + 1\,, \\ \hat{Z}_{t-1} &= \left[Y_{t-1}, \ldots, Y_{t-p}, I_K \otimes \hat{\mathbf{U}}_{t-1}'\right], \\ \hat{\mathbf{U}}_{t-1} &= [\hat{u}_{t-1}(1), \ldots, \hat{u}_{t-1}(K), \ldots, \\ &\qquad \hat{u}_{t-q}(1), \ldots, \hat{u}_{t-q}(K)]' \end{aligned}$$

*where $\otimes$ is the Kronecker product operator.*

**Step 3.** *Using the estimates obtained in step 2, we form*

$$\tilde{U}_t = Y_t - \sum_{i=1}^{p} \tilde{a}_i Y_{t-i} + \sum_{j=1}^{q} \tilde{B}_j \tilde{U}_{t-j}$$

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

$$\begin{aligned} X_t &= \sum_{j=1}^{q} \tilde{B}_j X_{t-j} + Y_t\,, \\ W_t &= \sum_{j=1}^{q} \tilde{B}_j W_{t-j} + \tilde{U}_t\,, \end{aligned}$$

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

$$\tilde{V}_t = \sum_{j=1}^{q} \tilde{B}_j \tilde{V}_{t-j} + \tilde{Z}_t\,,$$

*and $\tilde{Z}_t$ is just like $\hat{Z}_t$ from step 2 except that it is computed with $\tilde{U}_t$ instead of $\hat{U}_t$ to obtain regression coefficients that we call $\hat{a}_i$ and $\hat{B}_j$. This yields the estimator*

$$\begin{aligned} \hat{\gamma} &= \left[\sum_{t=\max(p,q)+1}^{T} \tilde{V}_{t-1}' \tilde{\Sigma}_U^{-1} \tilde{V}_{t-1}\right]^{-1} \times \\ &\quad \left[\sum_{t=\max(p,q)+1}^{T} \tilde{V}_{t-1}' \tilde{\Sigma}_U^{-1} [\tilde{U}_t + X_t - W_t]\right]. \end{aligned}$$

The properties of this estimation method are summarized in the following theorem.

**Theorem 1 (Third step estimates)** *If $n_T$ grows at a rate faster than $\log T$ with ${n_T}^4/T \to 0$, then the third stage estimates converge in quadratic mean to their true value. The asymptotic distribution of $\hat{\gamma}$ is given by*

$$\sqrt{T}\,(\hat{\gamma} - \gamma) \xrightarrow{d} N\big[0,\ \hat{J}^{-1}\hat{I}\hat{J}^{-1}\big]$$

*where*

$$\begin{aligned} \hat{I} &= \sum_{j=-\infty}^{\infty} E\left[\left\{V_{t-1}'\Sigma_U^{-1}U_t\right\}\left\{V_{t-1-j}'\Sigma^{-1}U_{t-j}\right\}'\right], \\ \hat{J} &= E\left[V_{t-1}'\Sigma_U^{-1}V_{t-1}\right] \end{aligned}$$

*and $V_{t-1}$ is equal to the matrix $\tilde{V}_{t-1}$ where $\tilde{U}_t$ is replaced by $U_t$. If furthermore $m_T^4/T \to 0$ with $m_T \to \infty$, then the matrices $\hat{I}$ and $\hat{J}$ can be consistently estimated by*

$$\begin{aligned} \hat{I}_T &= \sum_{j=-m_T}^{m_T} \frac{\omega(j, m_T)}{T} \sum_{t}^{T} \left\{\tilde{V}_{t-1}'\tilde{\Sigma}_U^{-1}\bar{U}_t\right\} \times \\ &\qquad \left\{\tilde{V}_{t-1-j}'\tilde{\Sigma}_U^{-1}\bar{U}_{t-j}\right\}', \\ \hat{J}_T &= \frac{1}{T} \sum_{t=\max(p,q)+1}^{T} \tilde{V}_{t-1}'\tilde{\Sigma}_U^{-1}\tilde{V}_{t-1} \end{aligned}$$

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

The simplicity of this estimation method is remarkable. Only three linear multivariate regressions are needed, and we meet none of the problems associated with the maximization of the likelihood (local optima, optimization over many parameters, non-uniqueness of the VMA operator).

As for the asymptotic variance of our estimators, we can show that the asymptotic covariance matrix of the third step estimator is the same as the one of: (i) the Gaussian maximum likelihood estimator, if the innovations are i.i.d or follow a m.d.s.; (ii) nonlinear least squares, if the innovations are merely uncorrelated. In a later section, Monte Carlo simulations will illustrate the good finite sample properties of our method for relevant sample sizes.

## 4. Estimation of the orders $p$ and $q$ for VARMA models in final equation form

We still have two unknowns in our model, the orders $p$ and $q$. If no theory specifies these parameters, we have to use a statistical procedure to choose them. We propose the following method to choose the order $\hat{p}$ and $\hat{q}$ of the VARMA($p$,$q$) model. In the second step of the estimation we compute for all $p \leq P$ and $q \leq Q$ the following information criterion:

$$\log|\det \tilde{\Sigma}_U| + (p + q\,K)\frac{(\log T)^{1+\delta}}{T}$$

with $\delta > 0$ and we choose ($\hat{p}$, $\hat{q}$) as the pair which minimizes the information criterion. We assume that the upper bound $P$ and $Q$ on the order of the VAR and VMA operator are bigger than the true value of $p$ and $q$ (or that they slowly grow with the sample size). The property of $\hat{p}$ and $\hat{q}$ are summarized in the following theorem.

**Theorem 2 (Estimation of $p$ and $q$)** *If $n_T$ grows at a rate faster than $\log T$ with ${n_T}^4/T \to 0$ then $\hat{p}$ and $\hat{q}$ converge in probability to their true value.*

This criterion is a generalization of the information criterion proposed by Hannan and Rissanen (1982) and later modified in Hannan and Rissanen (1983).

The earlier method for choosing $p$ and $q$ of VARMA model in final equation form was first to fit an ARMA($p_i$, $q_i$) for each univariate series $y_i$. Because the VAR operator is the same for each equation, we would take $p = \sum_{i=1}^{K} p_i$ for the order of the VAR operator (the product of each univariate AR operator). Accordingly we would get \$q = \max_k\{q_k + \sum_{i=1, i\neq k}^{K} p_i\}$. This procedure will obviously give bloated values for $p$ and $q$ and this is certainly a reason why the final equation form is rarely used.

## 5. Monte Carlo simulations

To evaluate the performance of our estimation method, we ran two sets of Monte Carlo simulations. In the first set, strong VARMA models (Gaussian i.i.d. innovations) are simulated and we compare our method to maximum likelihood. In the second set, weak VARMA models (uncorrelated innovations) are simulated and our method is compared to generalized nonlinear least squares. The weak VARMA models are simulated by time-aggregating strong VARMA processes. We also use our information criterion to choose the order $p$ and $q$ for both set of simulations. The results for only one weak VARMA experiment are reported here.

The simulated model is a weak VARMA(1, 1) in final equation form with $a_1 = 0.729$, $b(1,\,1) = 0.0593618$, $b(1,\,2) = -0.14134$, $b(2,\,1) = 0.20598$, $b(2,\,2) = 0.296472$. The variance of the innovations is 2.64155 and 1.70611 and the covariance is 0.650962. The sample size is 200 which would correspond to a bit less than 17 years of monthly data, a reasonable sample size in macroeconomics. The length of the long AR in the first step is $n_T = 15$ and the number of repetitions is 1000. The parameter in the information criterion is $\delta = 0.5$. The results for the information criterion appear in Table 1 and the results for the parameter estimation in Table 2.

Table 1: Estimation of $p$ and $q$

| $p \setminus q$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 1 | 0.135 | 0.791 | 0.000 | 0.000 | 0.000 |
| 2 | 0.033 | 0.033 | 0.000 | 0.000 | 0.000 |
| 3 | 0.002 | 0.004 | 0.000 | 0.000 | 0.000 |
| 4 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |
| 5 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |

From the results of Table 1, we see that the performance of the information criterion is pretty good. The criterion chooses the true order almost 80 % of the time. When the criterion does not choose the true order it will more often choose a lower order for the VMA operator.

The performance of our regression-based estimation method when compared to its corresponding nonlinear counterpart (nonlinear least squares) is excellent (see Table 2). The root mean squared errors over the 1000 simulations of the third step estimates is basically the same as those of NLLS. In this simulation the generalized NLLS are computed in two steps. In a first step, the variance of the innovations is taken to be an identity matrix and in the second step the identity matrix is replaced by the variance of the first step residuals. It is very interesting that we get the same efficiency with a regression-based estimation method that with a nonlinear method which come with all its caveats enumerated previously.

## 6. Application to macroeconomic time series

To illustrate the impact of using a VARMA model instead of a VAR model, we consider the following macroeconomic application. One common analysis of macroeconomic time series is to study the impulse-response functions, i.e. the effect of a change in a variable on the future values of another. The usual procedure to obtain the impulse responses consists in fitting a VAR model to the time series, from which the implied infinite VMA representation is obtained. When one fits a VAR model to macroeconomic time series, a high value for the order $p$ is often needed to get uncorrelated residuals. For ex-

Table 2: Estimation of the VAR and MA operators

| | Value | Average | Std. dev. | RMSE | Min | Max | Median |
|---|---|---|---|---|---|---|---|
| Third step | | | | | | | |
| $a_1$ | 0.7290 | 0.7229 | 0.0547 | 0.0550 | 0.5347 | 0.8534 | 0.7266 |
| $b_1(1,1)$ | 0.0594 | 0.0607 | 0.0831 | 0.0831 | -0.2312 | 0.3392 | 0.0613 |
| $b_1(1,2)$ | -0.1413 | -0.1462 | 0.0909 | 0.0910 | -0.4173 | 0.1763 | -0.1454 |
| $b_1(2,1)$ | 0.2060 | 0.2061 | 0.0599 | 0.0599 | 0.0092 | 0.4079 | 0.2069 |
| $b_1(2,2)$ | 0.2965 | 0.2914 | 0.0954 | 0.0955 | -0.0300 | 0.5718 | 0.2943 |
| NLLS | | | | | | | |
| $a_1$ | 0.7290 | 0.7242 | 0.0543 | 0.0545 | 0.5352 | 0.8518 | 0.7281 |
| $b_1(1,1)$ | 0.0594 | 0.0642 | 0.0835 | 0.0836 | -0.2250 | 0.3316 | 0.0640 |
| $b_1(1,2)$ | -0.1413 | -0.1481 | 0.0909 | 0.0912 | -0.4252 | 0.1729 | -0.1484 |
| $b_1(2,1)$ | 0.2060 | 0.2076 | 0.0603 | 0.0603 | 0.0072 | 0.4073 | 0.2088 |
| $b_1(2,2)$ | 0.2965 | 0.2953 | 0.0953 | 0.0953 | -0.0284 | 0.5666 | 0.2937 |

ample, Bernanke and Mihov (1998) uses a VAR(13) to model six monthly macroeconomic time series (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*)) when about 30 years of data are available (January 1962 to December 1996), hardly a parsimonious model.

A typical impulse-response analysis can be found in McMillin (2001). In this work the author uses the six time series of Bernanke and Mihov (1998) to study the impact of a change in the nonborrowed reserves on the variables *gdpm*, *pgdpm*, *fyff*. The do this he fits a VAR to the first difference of the series, in order, *gdpm*, (*psscom-pgdpm*),*fyff*, *nbrec1*, *tr1*, *psscom*. With an argument based on Keating (1999), the author states that using this ordering, the Choleski decomposition, based on long-run macroeconomic restrictions, of the variance matrix of the innovations will identify the structural effect of the policy variable *nbrec1*.

Fitting a VAR(12) to the same dataset we can replicate McMillin's experiment. The resulting impulse-responses are plotted in Figure 1 with one standard deviation confidence band (derived from Monte Carlo simulations with 1000 draws). These figures are typical in the macroeconomic literature. An increase of the nonborrowed reserves raises the output in the short run, increases permanently the price level, and decreases the level of the federal funds rate in the short run. A frequent result in this kind of analysis is that the confidence bands, i.e. the degree of uncertainty around these impulse-responses is quite large. We can wonder how much of this uncertainty is due to the use of such non-parsimonious models. Note that a VAR(12) operator with six time series contains 432 parameters.

To answer this question, we performed the same exercise with a VARMA model instead of a VAR to get the infinite VMA representation. Using our information criterion, with a VAR(15) in the first step estimation and with $\delta = 0.5$ in the information criterion, we identified a VARMA$(9,\ 3)$ model for these series. This particular model contains 119 parameters in the VAR and VMA operators. The impulse-response functions from this model and their one standard deviation confidence band (Monte Carlo simulations with 1000 draws) are plotted in Figure 2. The shape of the impulse-responses for the three series are similar. What is very different is the width of the confidence bands, which are now much smaller. An interesting result is that for the two variables which should not be affected in the long run by a change in the nonborrowed reserves according to the long-run restrictions, the confidence band shrinks as we move forward in time. It is natural that we get this result since the level of uncertainty should diminish if the restrictions are true. In contrast, for the price level, a variable which can be affected in the long-run by a change in the nonborrowed reserves, the level of uncertainty increases with the horizon. Note that we are not getting this result when the analysis is performed with a VAR, presumably because the uncertainty due the non-parsimonious representation is masking everything else.

## 7. Conclusion

In this paper, we have presented a regression-based estimation method in three steps for weak VARMA models. The estimates from the third step have the same asymptotic variance than the corresponding (i) Gaussian maximum likelihood estimates – if we assume that the innovations are i.i.d. or follow a m.d.s. – or (ii) generalized nonlinear least squares – if we merely assume that the innovations are uncorrelated. This estimation

Figure 1: Impulse-response functions for VAR model

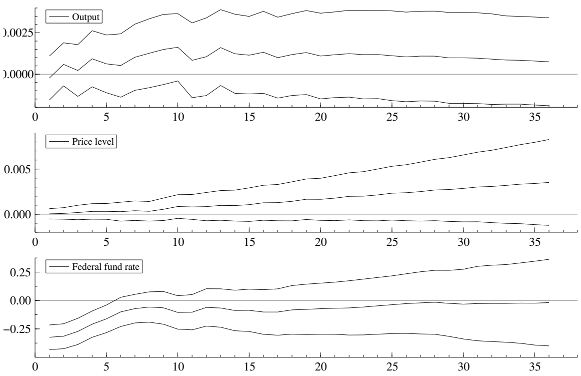
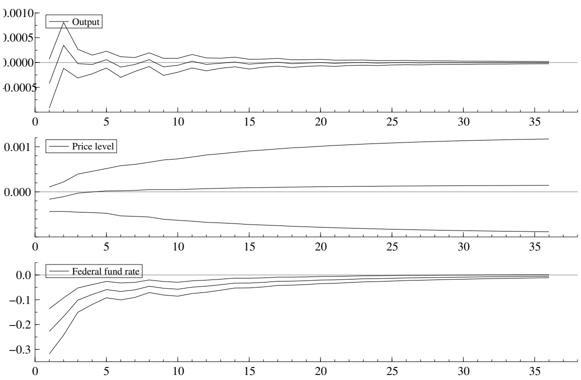



Figure 2: Impulse-response functions for VARMA model




method is a generalization of the method presented in Hannan and Rissanen (1982) for univariate ARMA models. Our results are derived under weaker assumptions than those usually considered in this literature (an alpha mixing condition instead of i.i.d. or m.d.s. conditions).

To make the implementation of the VARMA models easier, we chose the final equation form representation as the identified VARMA representation. An information criterion for choosing the order $p$ and $q$ of VARMA models under this representation was also presented.

A small Monte Carlo simulation showed the very good performance of our estimation method for small sample sizes when compared to their nonlinear counterparts. The performance of the information criterion also appears to be good.

Finally, to demonstrate the importance of using VARMA models instead of only relying on VAR models, we compared the impulse-responses generated by both models when applied to U.S. macroeconomic data. The results indicate that the impulse-responses obtained from the more parsimonious VARMA representation are more precisely estimated.

# References

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, second edn.

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

DUFOUR, J.-M., AND D. PELLETIER (2002): "Linear Estimation of Weak VARMA Models with Macroeconomic Applications," Discussion paper, Département de sciences économiques and CIREQ, Université de Montréal, Montréal, Canada.

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

——— (1983): "Errata: "Recursive estimation of mixed autoregressive-moving average order"," *Biometrika*, 70(1), 303.

KEATING, J. W. (1999): "Structural Inference With Long-run Recursive Empirical Models," Unpublished Paper, University of Kansas.

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

LÜTKEPOHL, H. (1993): "Testing for Causation Between Two Variables in Higher Dimensional VAR Models," in *Studies in Applied Econometrics*, ed. by H. Schneeweiss, and K. Zimmermann. Springer-Verlag, Heidelberg.

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