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

November 15, 2013
November 15, 2013, 13:34

[*] This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (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 4400 ext. 09156; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

1. Introduction

Motivation

- VAR models are widely used to do forecasting and policy analysis.
  - Estimation is easy (least squares).
  - VAR models are relatively easy to specify: choose the order.
  - Forecasting and (Granger) causality analysis are straightforward.

- Shortcomings of VAR modelling

  1. The number of parameters tend to go up quickly,
     leading to imprecision in estimation and low power in testing.
     Dimension reduction methods often used:
     Bayesian, methods, shrinking, factor models
  2. VAR modelling is logically incoherent, because the VAR class of models is not closed under marginalization and temporal aggregation:
     – if a vector time series Y_t follows a finite-order VAR,
       its subvectors do not typically follow a VAR;
       instead they follow VARMA models;
     – if a VAR process is temporally aggregated,
       the aggregated process is not typically a VAR;
       the aggregated process is a VARMA.

- This suggests to replace VAR models by VARMA models.
- Advantages of VARMA models:

1. considerably more parsimonious than VAR models;
   a VMA(1) is a VAR(∞) model, so a long VAR model may be required to approximate it reasonably well;
   alternative (or complement) to other dimension reduction methods;
2. closed under marginalization and temporal aggregation;
3. regularly follow from structural macroeconomic models (DSGE models),
   not VAR models. [Komunjer and Ng (2011, *Econometrica*)].

- Difficulties associated with VARMA models
  1. Raise identification problems not present in VAR models:
     - restrictions must be imposed to make sure a VARMA model has a unique representation;
     - the echelon form (Hannan, Deistler) is the most well known: not intuitive, difficult to specify.
  2. VARMA models are difficult to estimate and require nonlinear methods (maximum likelihood assuming Gaussian errors).

Contributions

The general goal of this paper is to develop a practical VARMA modelling methodology.

1. **Identification** – We propose new identified representations which are more intuitive and easier to specify and use than earlier ones [e.g., the echelon form]: **diagonal MA equation form** especially attractive.
2. **Estimation** – We propose linear estimation methods which can be applied by using least squares (involving estimated innovations).
   (a) Weak VARMA – Consistency and asymptotic normality under relatively weak assumptions on the model innovations (uncorrelated, strong mixing).
   (b) Asymptotic efficiency in the case of Gaussian innovations (asymptotic equivalence with ML when innovations are Gaussian).

3. **Model selection** – We present an information criterion for choosing the orders of the different operators in the proposed VARMA models.
4. **Simulation results** showing that the VARMA modelling does improve estimation efficiency (*e.g.*, for impulse response coefficients).
5. **Application** to a small macroeconomic model of monetary policy [Bernanke and Mihov (1998), McMillin (2001)]. Results show improvements in the estimation of impulse responses.

Contents

2. VARMA representations

K-variate VARMA(p,q) model in standard representation:

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

where

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

- The standard representation is not identified. We can have $A(L) \neq \tilde{A}(L)$, $B(L) \neq \tilde{B}(L)$ while $A(L)^{-1}B(L) = \tilde{A}(L)^{-1}\tilde{B}(L)$.
- Two identified representations:
  - Echelon form: restrictions on the order of the elements of $A(L)$ and $B(L)$.
  - Final equation form: $A(L)$ is scalar.

Echelon form: all operators $a_{ij}(L)$ and $b_{ij}(L)$ in the i-th row of $A(L)$ and $B(L)$ have the same degree p_i and have the form

$$
\begin{aligned}
a_{ii}(L) &= 1 - \sum_{m=1}^{p_i} a_{ii,m} L^m, \quad \text{for } i = 1, \ldots, K \\
a_{ij}(L) &= - \sum_{m=p_i-p_{ij}+1}^{p_i} a_{ij,m} L^m, \quad \text{for } j \neq i \\
b_{ij}(L) &= \sum_{m=0}^{p_i} b_{ij,m} L^m \quad \text{for } i, j = 1, \ldots, K, \quad \text{with } B_0 = A_0.
\end{aligned}
$$

Further, in the VAR operator $a_{ij}(L)$,

$$
p_{ij} = \begin{cases} \min(p_i + 1, p_j) & \text{for } i \geq j \\ \min(p_i, p_j) & \text{for } i < j \end{cases} \qquad i, j = 1, \ldots, K.
$$

i.e., p_{ij} specifies the number of free coefficients in the operator $a_{ij}(L)$ for $j \neq i$. The row orders $(p_1, \ldots, p_K)$ are the *Kronecker indices*.

- The Echelon form is hard to use, not intuitive.
- The final equation form puts the restrictions on the VAR operator.
- Focus is on restricting the VAR operator.
- We propose new identified VARMA representations
  - **Final MA equation form**: $B(L)$ is scalar, $B(L) = b(L)I_K$.
  - **Diagonal MA equation form**: $B(L)$ is diagonal.
  - **Diagonal AR equation form**: $A(L)$ is diagonal.

3. Identification

$$\Phi(L)Y_t = \Theta(L)U_t$$

where

$$\begin{aligned}\Phi(L) &= I_K - \Phi_1 L - \cdots - \Phi_p L^p,\\ \Theta(L) &= I_K - \Theta_1 L - \cdots - \Theta_q L^q.\end{aligned}$$

Then, under stationarity,

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

$$\Psi(L) = \Phi(L)^{-1}\Theta(L)$$

The impulse responses $\Psi(L)$ are identifiable. We must impose restrictions on $\Phi(L)$ and $B(L)$ to ensure a unique factorization.

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

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

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

$$\Theta(z) = \operatorname{diag}[\theta_{ii}(z)]$$

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

3.3 Assumption *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 $\theta_{ii}(z^\star) = 0$.*

3.1 Lemma *Let* $[\Phi(z), \Theta(z)]$ *and* $[\bar{\Phi}(z), \bar{\Theta}(z)]$ *be two pairs of polynomial matrices which satisfy the Assumptions 3.1 to 3.3. 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*

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

3.1 Definition (Diagonal MA equation form) *The VARMA model is said to be in diagonal MA equation form if* $\Theta(L) = \text{diag}[\theta_{ii}(L)] = I_K - \Theta_1 L - \cdots - \Theta_q L^q$ *where* $\theta_{ii}(L) = 1 - \theta_{ii,1}L - \cdots - \theta_{ii,q_i}L^{q_i}$, $\theta_{ii,q_i} \neq 0$, *and* $q = \max_{1\leq i\leq K}(q_i)$.

3.2 Theorem (Identification of diagonal MA equation form) *Suppose the VARMA model satisfies the Assumptions 3.1-3.3 hold. If the VARMA model is in diagonal MA equation form, then it is identified.*

Invertibility of the model is not required.
If invertibility is imposed, a similar result holds for the diagonal AR form (more difficult to use).

4. Regression-based estimation method

Generalization of Hannan and Rissanen (1982), presented in Hannan and Kavalieris (1984).

Estimation in three steps.

Step 1: Estimate a long VAR and keep the residuals

$$\hat{U}_t = Y_t - \sum_{i=i}^{n_T} \hat{\Pi}_i^{n_T} Y_{t-i}$$

with n_t growing at a rate faster than $\log T$ and $n_t^2/T \to 0$.

Step 2: Replace the lagged innovations by the lagged residuals and estimate by GLS

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

Step 3: Using $\tilde{U}_t$, the residuals from the second step, define

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

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

where $\tilde{Z}_t$ is the matrix of regressors from step 2, except that we replace $\hat{U}_t$

by $\tilde{U}_t$.

Regress by GLS $\tilde{U}_t + X_t - W_t$ on $\tilde{V}_{t-1}$.

$\Longrightarrow$ Same asymptotic distribution as MLE/NLLS.

- We do not assume that the innovation process is i.i.d. or m.d.s.
- We only assume that it is uncorrelated and strongly mixing.
- Let $\{U_t\}$ be a strictly stationary process, then its α-mixing coefficient of order h is defined as

$$\alpha(h) = \sup_{\substack{B\in\sigma(U_s, s\leq t)\\ C\in\sigma(U_s, s\geq t+h)}} |\Pr(B\cap C) - \Pr(B)\Pr(C)| \ , \ \ h \geq 1.$$

- The strong mixing condition that we impose is

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

- Why? We can then study linear representations of nonlinear processes.
- Looking at quadratic mean convergence
- Need to study elements like $Var\left[\frac{1}{T}\sum_{t=1}^{T} y_{t-r}(k) y_{t-s}(k')\right]$
- i.e., $Cov\left[u_{t-r}(k)u_{t-s}(k')\,;\,u_{t'-r}(k)u_{t'-s}(k')\right]$

- Davydov (1968): Let U and V be random variables measurable with respect to $\mathscr{F}_{-\infty}^{0}$ and $\mathscr{F}_{n}^{\infty}$, respectively. Let r_1, r_2, r_3 be positive numbers. Assume that $\|U\|_{r_1} < \infty$ and $\|V\|_{r_2} < \infty$ where $\|U\|_r = (E[|U|]^r)^{1/r}$. If $r_1^{-1} + r_2^{-1} + r_3^{-1} = 1$, then there exists a positive constant C independent of U, V and n, such that

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

- Ibragimov (1962): Central Limit Theorem for alpha-mixing processes

5. Information criterion

- We propose the following information criterion for choosing p_i and q_i:

$$\log(\det\tilde{\Sigma}) + c_0 \dim(\gamma) \frac{(\log T)^{1+\delta}}{T}, \quad c_0 > 0, \delta > 0$$

- We minimize it over $p_i \leq P$ and $q_i \leq Q$ in the second step.
- For the diagonal representations, the criterion can be minimized equation by equation [(p_i, q_i) for $i = 1, 2, \ldots, K$].

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

- Diagnostic: check for uncorrelated residuals using the results in Francq, Roy and Zakoïan (2005).

6. Monte Carlo results

- We simulate strong VARMA models with i.i.d. Gaussian innovations.
- We simulate weak VARMA models where the innovations are uncorrelated but not a m.d.s. [Francq and Zakoïan (1998), time-aggregation of GARCH models].
- Results:
  - RMSEs of our three-step estimator is usually less than 15 % higher than of nonlinear estimator.
  - Information criterion: in most cases, the most frequently chosen orders are the true ones.

Weak diagonal MA equation form VARMA(1,1). The simulated model is a weak VARMA(1,1) in diagonal MA equation form with $a(1,1) = 0.5$, $a(1,2) = -0.6$, $a(2,1) = 0.7$, $a(2,2) = 0.3$, $b_1(1) = 0.9$ and $b_1(1) = 0.7$. The variance of the innovations is 1.3 and the covariance is 0.91. Sample size is 250, the length of the long AR is $n_T = 20$, the number of repetition is 1000. The parameters in the criterion are $\delta = 0.3$ and $c_0 = 1$.

(p, q_1, q_2)	Frequency	(p, q_1, q_2)	Frequency
1,1,1	0.588	1,1,3	0.026
1,2,1	0.123	2,1,1	0.014
1,1,2	0.062	1,4,1	0.014
1,3,1	0.045	1,5,1	0.012
2,2,2	0.043	1,1,5	0.010

	Value	Average	Std. dev.	RMSE	5%	95%	Median
Second step							
$a_1(1,1)$	0.5	0.4277	0.0601	0.0940	0.3284	0.5233	0.4303
$a_1(1,2)$	-0.6	-0.6439	0.0507	0.0671	-0.7291	-0.5594	-0.6444
$a_1(2,1)$	0.7	0.6732	0.0514	0.0579	0.5863	0.7550	0.6729
$a_1(2,2)$	0.3	0.2314	0.0526	0.0865	0.1446	0.3193	0.2309
$b_1(1)$	0.9	0.8130	0.0707	0.1122	0.6976	0.9266	0.8150
$b_1(2)$	0.7	0.6364	0.0708	0.0952	0.5185	0.7476	0.6393
Third step							
$a_1(1,1)$	0.5	0.5064	0.0469	0.0473	0.4324	0.5845	0.5062
$a_1(1,2)$	-0.6	-0.5960	0.0552	0.0554	-0.6762	-0.5183	-0.5969
$a_1(2,1)$	0.7	0.6988	0.0418	0.0418	0.6314	0.7659	0.6997
$a_1(2,2)$	0.3	0.3021	0.0469	0.0469	0.2272	0.3830	0.3032
$b_1(1)$	0.9	0.8885	0.0442	0.0456	0.8100	0.9531	0.8910
$b_1(2)$	0.7	0.6967	0.0522	0.0523	0.6092	0.7843	0.6969
NLLS							
$a_1(1,1)$	0.5	0.4973	0.0453	0.0453	0.4222	0.5703	0.4972
$a_1(1,2)$	-0.6	-0.6116	0.0443	0.0458	-0.6864	-0.5371	-0.6114
$a_1(2,1)$	0.7	0.7009	0.0411	0.0411	0.6334	0.7683	0.7006
$a_1(2,2)$	0.3	0.2897	0.0441	0.0453	0.2185	0.3645	0.2893
$b_1(1)$	0.9	0.8874	0.0349	0.0371	0.8260	0.9385	0.8894
$b_1(2)$	0.7	0.6950	0.0446	0.0449	0.6198	0.7673	0.6955

7. Macroeconomic application

- One use of VAR and VARMA models is the computation of impulse-response functions (IRFs).
- Macroeconomists typically use VARs to get the infinite moving average representation.
- The confidence bands around the IRFs are big! Quite often, zero is within the bands at every horizon.
- Some of this must be due to the high number of parameters estimated when using a VAR.
- Could a VARMA model do better?

- Our example is based on McMillin (2001).
  - Take the first difference of the series, in order, log of real GDP, log of spot commodity price minus log of the GDP deflator, federal funds rate, nonborrowed reserves, total bank reserves, log of spot commodity price.
  - Using the ordering of the variables the Choleski decomposition, based on long-run macroeconomic restrictions, will identify the structural effects.
  - Use a VAR(12).
- We fit VARMA models.
- Impact of a one standard deviation shock to nonborrowed reserves.

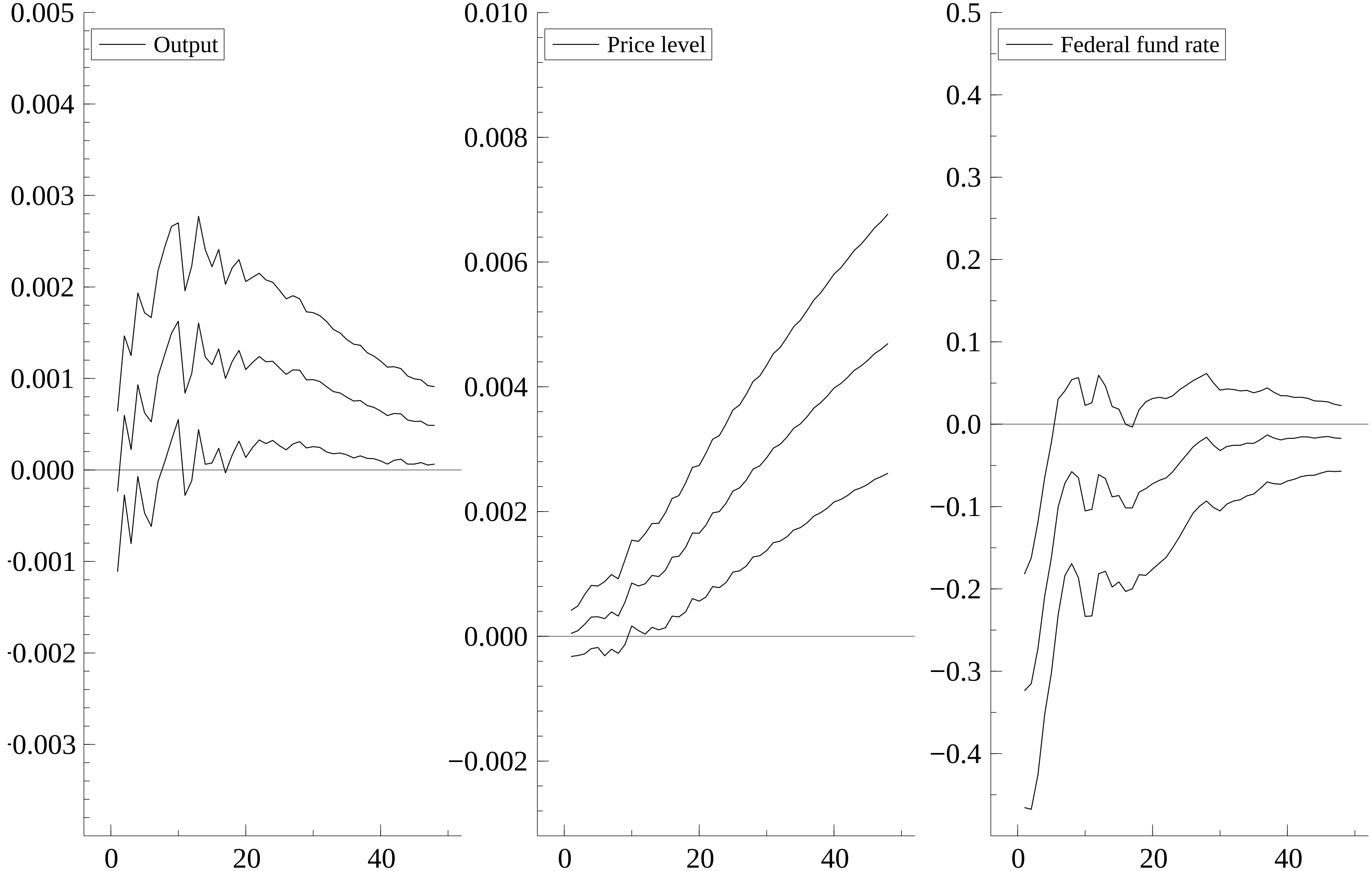

Figure 1. IRFs for VAR(12) model

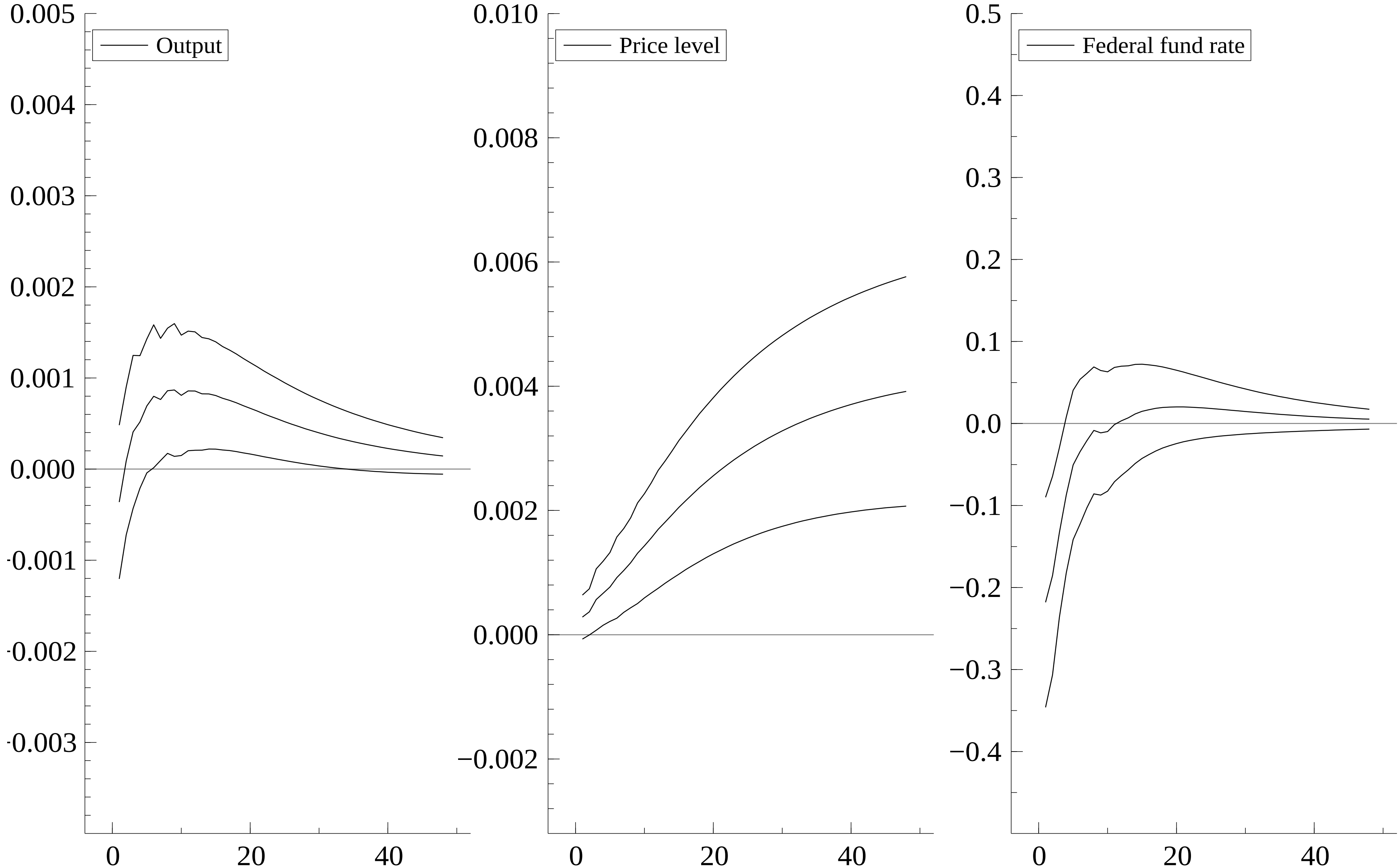

Figure 2. IRFs for VARMA(3,10) model in final MA equation form

RMSE of out-of-sample forecasts with VAR and VARMA models

Step ahead	VAR	VARMA diag. MA	VARMA final MA
1	0.0834	0.0764	0.0743
	$p=1$	$p=0$	$p=0$
		$q=(1,2,2,1,1,1)$	$q=12$
3	0.0799	0.0788	0.0744
	$p=1$	$p=1$	$p=1$
		$q=(1,1,1,1,0,1)$	$q=12$
6	0.0826	0.0767	0.0790
	$p=7$	$p=3$	$p=1$
		$q=(4,4,1,4,0,4)$	$q=12$
9	0.0871	0.0774	0.0829
	$p=2$	$p=4$	$p=0$
		$q=(5,5,3,4,5,5)$	$q=12$
12	0.0819	0.0728	0.0803
	$p=4$	$p=4$	$p=1$
		$q=(3,5,3,4,5,5)$	$q=12$

8. Conclusion

We present a modelling method for VARMA models

- We present new VARMA identified representations which are easy to use.
- We present a regression-based estimation method in three steps for VARMA models which has the same asymptotic efficiency as MLE or NLLS.
- We present an information criterion for choosing the orders for the VARMA representation we proposed.
- So as to broaden the class of processes to which this method can be applied, we only assume that innovations are uncorrelated.