Estimation of the Vector Moving Average model by Vector Autoregression

John W. Galbraith, Aman Ullah and Victoria Zinde-Walsh[1]

ABSTRACT

We examine a simple estimator for the multivariate moving average model based on vector autoregressive approximation. In finite samples the estimator has a bias which is low where roots of the determinantal equation are well away from the unit circle, and more substantial where one or more roots have modulus near unity. We show that the representation estimated by this multivariate technique is consistent and asymptotically invertible. This estimator has significant computational advantages over Maximum Likelihood, and more importantly may be more robust than ML to mis-specification of the vector moving average model. The estimation method is applied to a VMA model of wholesale and retail inventories, using Canadian data on overall aggregate, non-durable and durable inventory investment, and allows us to examine the propagation of shocks between the two classes of inventory.

[1]Galbraith and Zinde-Walsh: Department of Economics, McGill University, 855 Sherbrooke St. West, Montreal, QC, Canada H3A 2T7; Ullah: Department of Economics, University of California, Riverside, CA 92521, USA.

1. Introduction.

Estimation techniques for vector moving average (VMA) models are in general extensions of the techniques used for the scalar case. Typically, algorithms are based on the method of maximum likelihood (ML), approximate or exact (see Osborn, 1977), or on iterative procedures based on long autoregressions; the latter may be used with ML estimation to provide starting values (see, for example, Lütkepohl, 1993 for a discussion). The main advantage of ML is asymptotic efficiency in a correctly specified model. However, the computational complexity and potential for non-convergence of ML may lead to difficulties in employing these estimates in Monte Carlo studies; moreover, non-robustness to mis-specification may undermine its reliability. In the present study we examine an estimator designed to provide better performance in these respects.

ARMA or vector ARMA processes often arise as nuisance elements in a model, where simplicity, robustness and good finite-sample performance are issues of greater concern than asymptotic efficiency. For the scalar case, a simple estimator based on an autoregressive approximation to the MA process was examined in Galbraith and Zinde-Walsh (1994); for an ARMA process a method is given in Galbraith and Zinde-Walsh (1997). In the MA case, in addition to ease of calculation such methods exhibited small-sample performance quite similar to that of ML estimator, unless one or more roots was close to the limits of the invertibility region, and could improve upon the performance of ML in over-specified and, especially, in under-specified models. Lütkepohl (1988) and Lütkepohl and Poskitt (1991) use an analogous procedure to compute impulse response (MA) representations of VAR processes, and obtain the asymptotic distribution. Alternative non-ML algorithms for the scalar MA case include those of Durbin (1959) and Koreisha and Pukkila (1990).

This paper examines the simple estimator in the multivariate case, concentrating on finite-sample performance of the estimator when used to obtain coefficients of an underlying vector moving average process. We compare it with exact ML, in both correctly-specified and mis-specified cases. We then use the estimator in an application in which we investigate the propagation of shocks between wholesale and retail inventories, both overall aggregate and divided into durable and non-durable goods, with Canadian data.

Section 2 provides the necessary definitions, introduces the estimator, and considers

the asymptotic properties of the estimates. Simulation results, including the comparison with ML, are presented in section 3, and section 4 presents the analysis of the inventory investment data. Section 5 concludes.

2. The vector moving average estimator
2.1 Estimation algorithm

We consider an n-dimensional discrete-time multiple time series $\{X_t\}$, with $X_t = (X_{1t}, X_{2t}, \ldots, X_{nt})'$ consisting of zero-mean, finite-variance, stationary stochastic processes with finite fourth moments. The multiple time series follows a vector moving average process of order q, that is,

$$X_t = A_0^* u_t + A_1^* u_{t-1} + \cdots + A_q^* u_{t-q} \quad (t = 1, 2, \ldots, T), \tag{1}$$

where u_t is an n-dimensional white-noise process with covariance matrix Σ, and where $A_0^*, A_1^*, \ldots A_q^*$ are coefficient matrices of dimension $n \times n$, with A_0^* of full rank; no roots of the determinantal equation

$$|A_0^* m^q + A_1^* m^{q-1} + \cdots + A_q^*| = 0 \tag{2}$$

are on the unit circle.

A moving average process has the property that its covariance structure is invariant with respect to two classes of transformation:

(i) transformation by a non-singular $n \times n$ matrix W, so that $v_t = W u_t$, and $C_i = A_i^* W^{-1}$, leading to the representation

$$X_t = C_0 v_t + C_1 v_{t-1} + \ldots + C_q v_{t-q},$$

with $E(v_t v_t') = W \Sigma W'$. The roots of $|C_0 m^q + C_1 m^{q-1} + \cdots + C_q| = 0$ coincide with those of (2);

(ii) the unique transformation of the representation (1) that replaces every root of the determinantal equation (2) that is outside the unit circle by its inverse, yielding a determinantal equation with all roots inside the unit circle (see, e.g., Hannan and Deistler 1988).

3

We therefore consider, without further loss of generality, the process (1) to be such that all roots of (2) are inside the unit circle.

Here we will choose a transformation matrix as in (i) above, denoted S, to yield a set of orthogonalized innovations $\varepsilon = Su$, where S is the Cholesky decomposition of Σ^{-1} (that is, $S'S = \Sigma^{-1}$), so that $E(\varepsilon_t'\varepsilon_t) = I$. We have

$$X_t = A_0^* S^{-1} \varepsilon_t + A_1^* S^{-1} \varepsilon_{t-1} + \cdots + A_q^* S^{-1} \varepsilon_{t-q}, \quad (t = 1, 2, \ldots, T). \tag{3}$$

Next define $A_i = A_0^* S^{-1}$, $i = 1, 2$, and we obtain

$$X_t = A_0 \varepsilon_t + A_1 \varepsilon_{t-1} + \cdots + A_q \varepsilon_{t-q}, \quad (t = 1, 2, \ldots, T), \tag{4}$$

Note that if $\Sigma = I$, then $A_i = A_i^*$.

In general, because of the invariance property, model (1) can be estimated only up to transformations of types (i) and (ii). From this point on, we therefore consider models of the type given in (4), and such that all roots of the corresponding determinantal equation are inside the unit circle.

It follows from the determinantal condition (2), with all roots inside the unit circle, that the multiple time series (1) or its transformation (4) can be expressed as an infinite-order VAR process. We examine the transformed process, which can be represented as (see, for example, Fuller, 1976)

$$X_t = \sum_{j=1}^{\infty} B_j X_{t-j} + B_0 \varepsilon_t, \tag{5}$$

with coefficient matrices given by

$$B_0 = A_0$$

$$B_1 = A_1$$

$$B_2 = -B_1 A_1 + A_2$$

$$B_3 = -B_2 A_1 - B_1 A_2 + A_3$$

$$\vdots \tag{6}$$

$$B_q = -B_{q-1} A_1 - B_{q-2} A_2 - \cdots - B_1 A_{q-1} + A_q;$$

$$B_j = \sum_{i=1}^{q} -B_{j-i} A_i, \quad \text{for } j = q+1, \ldots \quad .$$

4

By a method analogous to that of Galbraith and Zinde-Walsh (1994), relations of this type can be used to yield estimates of the moving average coefficient matrices. By estimating a VAR approximation to the system, we may deduce the VMA coefficient matrices from those of the VAR and the set of relations above.

The following algorithm is one method of using the information embodied in the VAR approximation, and is similar to the method used by Lütkepohl (1988) and Lütkepohl and Poskitt (1991) in impulse response estimation.[2] Fit a VAR(p) model with $p > q$ to the sample by least squares; from the p estimated coefficient matrices of dimension $n \times n$ of the VAR representation $X_t = B_1 X_{t-1} + B_2 X_{t-2} + \cdots + B_p X_{t-p} + B_0 \varepsilon_t$, estimate the q moving average coefficient matrices of dimension $n \times n$ by the following relations based on (6):

$$A_1 = B_1$$
$$A_2 = B_1 A_1 + B_2$$
$$A_3 = B_1 A_2 + B_2 A_1 + B_3 \tag{7}$$
$$\vdots$$
$$A_q = B_1 A_{q-1} + B_2 A_{q-2} + \cdots + B_{q-1} A_1 + B_q.$$

Finally, the estimated error covariance matrix provides an estimate of $B_0 B_0'$, from which an estimate of A_0 can be obtained as the Cholesky decomposition.

Before examining the performance of this vector analogue of the simple estimator of Galbraith and Zinde-Walsh (1994), we will summarize briefly the results obtained there for the scalar case. First, the simple estimator is biased; in the MA(1) case with parameter θ, for example, the estimator has an asymptotic bias toward zero which equals $\frac{-\theta^{2p+1}(1-\theta^2)}{1-\theta^{2p+2}}$, and thus declines monotonically with p. Second, the asymptotic mean squared error to $O_p(T^{-1/2})$ is provided, and for the MA(1) case an explicit expression for it is given as a function of θ, T and p. Minimization of this part of the mean squared error over choices of p leads to the conclusion that small values of p (around 1 to 3) perform best, by a small margin, for moderate values of θ (around 0.5 and less) even for quite large sample sizes.

---

[2]Lütkepohl (1988) does not orthogonalize the innovations, being concerned with asymptotic properties of the impulse responses, which are unaffected.

However, for θ near the unit circle, high values of p are needed to minimize the mean squared error. For example, the optimal value of p is about 24 for $\theta = .95$ and $T = 200$; small values of p result in quite large mean squared errors in this case.

In the next sections we comment on consistency and invertibility of the estimated representations. We show that for fixed p the simple estimator has an asymptotic bias, which declines exponentially if p is allowed to grow with T at an appropriate rate.

2.2 Consistency of the estimates

In order to discuss consistency we need notation to distinguish the coefficients B_i of the infinite order system (5) from the coefficients of the population analogue of the p-th order least-squares VAR approximation. We will now refer to the former as $B_{i(\infty)}$ and to the latter as $B_{i(p)}$. Define $\Gamma_i = E(X_t X_{t-i})$, and set $B_i = 0$ if $i > p$; then the $B_{i(p)}$ are the solutions to the following system of Yule-Walker equations:

$$
\begin{aligned}
B_1\Gamma_0 &+ B_2\Gamma_1' &+ \cdots &+ B_{q+1}\Gamma_q' & &= \Gamma_1 \\
B_1\Gamma_1 &+ B_2\Gamma_0 &+ \cdots &+ B_{q+1}\Gamma_{q-1}' &+ B_{q+2}\Gamma_q' &= \Gamma_2 \\
&\vdots & & & \vdots & \\
B_1\Gamma_q &+ B_2\Gamma_{q-1} &+ \cdots + & B_{2q+1}\Gamma_q' & &= \Gamma_q;
\end{aligned}
\tag{8}
$$

$$
B_i\Gamma_q + B_{i+1}\Gamma_{q-1} + \cdots + B_{q+i}\Gamma_0 + \cdots + B_{2q+i}\Gamma_q' = 0, \quad i = 2, \ldots, p.
$$

As $T \to \infty$ and for p of order $o(T^{\frac{1}{3}})$, it follows from Berk (1974) that the OLS estimates $(\hat{B}_1, \ldots, \hat{B}_p)$ of the VAR(p) coefficients converge to the solutions $B_{i(p)}$, $i = 1, \ldots, p$ of (8); that is, $B_{i(p)} = plim(\hat{B}_{i(p)})$. Define the asymptotic bias for order p as $\Delta B_i = B_{i(p)} - B_{i(\infty)}$, and for estimated values, let $\Delta\hat{B}_i = \hat{B}_{i(p)} - B_{i(\infty)}$. Then $\Delta B = [\Delta B_1 : \Delta B_2 : \ldots : \Delta B_p]$ solves the system

$$
\Delta B\Gamma = \Omega,
\tag{9}
$$

where Ω is an $np \times n$ matrix of the form $\begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \vdots \\ \Omega_p \end{bmatrix}$ with $n \times n$ sub-matrices $\Omega_1 = \Omega_2 = $

6

$\cdots = \Omega_{p-2q} = 0$ and $\Omega_{p-2q+1} = -B_{p+1}\Gamma'_q$, and so on. This representation follows from comparing the Yule-Walker equations for the infinite-order VAR in (5) with the equations for the truncated VAR(p) in (8).

Each of the Ω_i can be expressed as a linear combination of at most $2q - 1$ matrices B_j, $j \geq p + 1$. From the fact that the B_j are the coefficient matrices of the infinite autoregressive representation of an invertible MA(q) process it follows that the elements of $B_j, j \geq p + 1$ decline in absolute value at an exponential rate as $p \to \infty$, and therefore also that the asymptotic bias $\Delta B = \Gamma^{-1}\Omega$ declines to zero as p increases. Therefore, the estimate $\hat{A}_1$ differs from A_1 by $\Delta\hat{B}_1$ and so shares the distribution and asymptotic bias of $\hat{B}_1$; similarly the estimate $\hat{A}_2$ differs from A_2 by $B_1\Delta\hat{B}_1 + \Delta\hat{B}_1A_1 + (\Delta\hat{B}_1)^2 + \Delta\hat{B}_2$ and so has the same order of asymptotic bias as $\Delta\hat{B}_2$, and so on. Since $\Delta\hat{B}_i \to 0$ $\forall i$ as $T \to \infty$ with $p \to \infty$ and $p = o(T^{\frac{1}{3}})$, consistency of the estimator of A_i follows.

2.3 Asymptotic invertibility of the estimated representations

The following proposition demonstrates that the estimated representation in also invertible as $T, p \to \infty$, whether or not the original process is invertible, as long as there are no roots on the unit circle.

Proposition. Let the VMA(q) process $X_t = A(L)\varepsilon_t$ ($t = 1, 2, \ldots T$) be such that the determinantal equation $|A_0 m^q + A_1 m^{q-1} + \cdots + A_q| = 0$ has no roots on the unit circle. Let $\hat{A}(L)$ be a set of estimated lag polynomial matrices produced by the simple estimator of a VMA(q) model, using a VAR approximation of order p. Then for sufficiently large p, T, the VAR(p) representation of the process is stationary and the determinantal equation $|\hat{A}_0 m^q + \hat{A}_1 m^{q-1} + \cdots + \hat{A}_q| = 0$ has all roots within the unit circle (the estimated representation is invertible) with probability arbitrarily close to 1.

Proof: If X_t satisfies the given determinantal condition, then there exists an invertible VMA(q) process $Y_t = C_{(q)}(L)\nu_t$ with an identical autocovariance function ($C_{(q)}(L) = A(L)$, $Y_t = X_t$ if the original process is invertible.) Being invertible, Y_t has an infinite VAR representation of the form (5)-(6) based on $(C_{(q)}(L))^{-1}$, which is well defined as a mean square limit (see, e.g., Fuller, 1976; Lütkepohl, 1993) and which, since Y_t must be stationary, is stationary as well. Let this VAR representation be $B_\infty(L)Y_t = u_t$. Consider now LS estimation of the VAR(p) model $B_{(p)}(L)X_t = u_t$. The estimated lag polynomial

matrices $\hat{B}_{i(p)}$ will be identical to those estimated from a VAR(p) model of Y_t, since LS estimation uses only the autocovariances. By the consistency of the LS estimator, $\hat{B}_{(p)}(L) \xrightarrow{p} B_\infty(L) = [C_{(q)}(L)]^{-1}$ where $p, T \to \infty$ such that, for example, $p = o(T^{\frac{1}{3}})$.[3] Hence the estimated process $\hat{B}_{(p)}(L)X_t = u_t$ will also be stationary for sufficiently large p, T. Finally consider estimates of the coefficient matrices $\hat{A}_i$ of an order–q model of X_t, based on (7) and using the estimated matrices $\hat{B}_{i(p)}$. Since the latter coefficient matrices approach those of $B_\infty(L)$ in probability as $p, T \to \infty$, the corresponding solutions to (7) converge as well, to those of $C_{(q)}(L)$. Since the process $C_{(q)}(L)$ is invertible, this estimated process will also be invertible, with probability arbitrarily close to 1, for sufficiently large p, T. ∎

As examples, consider the two non-invertible VMA(1) processes (subscripted a and b) described by the lag polynomial matrices $A_{1,a} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ and $A_{1,b} = \begin{bmatrix} -1 & 1.75 \\ 1 & -4 \end{bmatrix}$ in (4). The roots of the determinantal equation (2) corresponding to the two processes are $(-1/2, -3/2)$ and $(1/2, 9/2)$ respectively. In simulations of 5000 replications with $T = 200$ and $p = 8$, we obtained average values of the estimated coefficient matrices (corresponding to invertible counterparts of the processes) of $\hat{C}_{1,a} = \begin{bmatrix} 0.577 & -0.082 \\ -0.082 & 0.578 \end{bmatrix}$ and $\hat{C}_{1,b} = \begin{bmatrix} -0.137 & 0.040 \\ -0.715 & -0.576 \end{bmatrix}$. The roots of the determinantal equations corresponding to these estimated processes are $(-0.496, -0.660)$ and $(0.488, 0.225)$ respectively, and are consistent estimates of the roots of the invertible counterparts, respectively $(-1/2, -2/3)$ and $(1/2, 2/9)$.

Sufficiently large p ensures stationarity of the VAR(p) representation of the VMA(q) process, by the consistency of the estimator. Again, a sufficient condition is that the order p of the VAR approximation increase at the rate $o(T^{\frac{1}{3}})$ as $T \to \infty$. A faster rate, close to $T^{\frac{1}{3}}$, entails a loss of precision in the estimation of parameter matrices of the VAR(p) process through the loss of degrees of freedom, but is offset by the faster decline in asymptotic bias (for given p) of the VMA estimates. A non-stationary estimated VAR(p) representation is a signal that the process is not in fact VMA, or that p is too low.

---

[3] See Lütkepohl (1988).

For practical purposes one needs some indication of appropriate small-sample orders p; the following section provides some indication, by simulation, of appropriate p for moderate sample size and order of VMA process.

3. Simulation results

In order to examine the performance of the simple VAR-based estimator we choose a selection of processes yielding real and complex roots of the determinantal equation (2). The processes are described in Table I, and the simulation results for VMA(1) and VMA(2) cases appear in Tables II and III respectively; sample size is $T = 200$ in each case. In Table IV we report results for mis-specified models in which VMA(1) models are used for the VMA(2) processes of Table I (cases v-viii). Note that all of the processes selected below are invertible, and for each, $A_0 = I$ and $\Sigma = I$.

In each case we report both the root mean squared error for the estimates based on (7), and the ratio of this RMSE to the corresponding RMSE from a Maximum Likelihood estimator. We use an exact ML algorithm in which the Kalman Filter is used to generate the likelihood function for the model case in state-space form (see for example Lütkepohl, 1993), followed by optimization using the Powell algorithm (Press et al., 1986). Because we perform Monte Carlo simulation with ML, we need an optimization algorithm which is very reliable in the sense of very rarely failing to converge to a solution; discarding non-convergent cases could lead to systematic bias in the simulation results. The Powell algorithm meets this requirement, and we were able to obtain 1000 (VMA(2) models) or 2000 (VMA(1) models) replications without the necessity of discarding any experiments. In ML estimation, we treat Σ as known to be diagonal in order to optimize over only two values in the covariance matrix.[4]

Consider the root mean squared errors (that is, $\mathrm{MSE}^{\frac{1}{2}}$) of estimation for the VMA(1) cases, in Table II. In parallel with the univariate case (Galbraith and Zinde-Walsh 1994), we see that the "optimal" value of p tends to be larger where roots are nearer the unit circle (although biases again do decline monotonically in p); of course, the value of p which

---

[4]This represents a modest advantage to ML in that a valid restriction is imposed, saving degrees of freedom.

9

gives lowest RMSE may differ for different elements of the coefficient matrix. A low value of p performs relatively well in Case i, which has the smallest-modulus roots; in each of cases ii–iv there are some parameters for which there is a substantial improvement in moving from $p = 4$ to $p = 8$. It is also noteworthy that RMSE's of estimates from (7) show little variation across parameters, and little sensitivity to p within this range.[5] This reflects in part the fact that with $p = 8$ biases are low for any of these models, so that the RMSE reflects mainly the estimator's variance. Estimates of A_0 (not reported in the tables) produce similar root mean squared errors.

Relative to ML, we note that ML is generally superior (the indicated RMSE ratio exceeds unity), but that this is not always the case; in case ii, for example, ML gives superior estimates of the (1,1) parameter, but (7) performs much better for the (2,2) parameter. Note also the relatively high variability across parameters of the ratios, reflecting relatively high variability of Maximum Likelihood RMSE's across parameters.

Similar patterns arise in the VMA(2) cases, Table III a/b. For the first coefficient matrix, where roots are complex, $p = 8$ almost invariably produces gains relative to $p = 4$; for the second coefficient matrix the two values produce similar results. Maximum Likelihood produces lower RMSE in a majority (37/64) of cases, but ratios are generally near unity, and there are cases of substantially better performance by (7) as well as by ML. In general, RMSE's are higher than for the VMA(1) cases for given p; case viii, in particular, indicates that for moderately large roots we may need more than eight lags to minimize RMSE.

In Table IV we consider mis-specified cases. Note that for the estimator based in (7), the RMSE's are precisely the same as those for the first parameter matrix in a VMA(2) model (Table IIIa), since this estimator has the property that estimates of parameters to order q are unaffected by whether or not estimates at $q + 1$ are also calculated. We therefore report only the RMSE ratio, as defined above. Here we see, as did Galbraith and Zinde-Walsh (1994), that mis-specification can affect ML badly; the RMSE for (7) is

---

[5]Results are very similar for $p = 12$ (not reported in these tables), with the exception of Table IIIa, case viii, where a substantial improvement in RMSE is available in moving from $p = 8$ to $p = 12$.

superior in a large majority of cases, sometimes by a large margin.[6]

Finally, it is interesting to compare the results in case iv of Table II (in which the two processes are in fact univariate) with the mean squared errors of comparable univariate models (see Table 3a of Galbraith and Zinde-Walsh, 1994). In the univariate case the simple estimator gives root mean squared errors of approximately 0.07 for MA parameters of 0.1 and 0.5, and approximately 0.08 for a parameter of 0.9 ($T = 200$, using $p = 12$). These are very similar to the Table II-case iv ($p = 8$) results, suggesting that the sacrifice which follows from specifying a vector MA model for what are in fact independent univariate processes may be negligible.[7]

---

[6] Comparison of the estimates from the first parameter matrix remains the correct comparison in these mis-specified cases; second-order MA terms do not project onto first-order terms, and so capturing first-order autocorrelation correctly is the best achievable with a mis-specified model of this type.

[7] As noted above, results for $p = 12$ are not reported in tables given here, but are in this case very close to those for $p = 8$.

Table I
Coefficient matrices and roots of (2) for cases considered in Tables II and III

VMA(1) processes

Case	A_1	μ_1	μ_2[8]
i	$\begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.6 \end{bmatrix}$	-0.176	-0.624
ii	$\begin{bmatrix} 0.8 & 0.2 \\ -0.6 & 0.4 \end{bmatrix}$	$-0.6 \pm 0.283i$	mod=0.663
iii	$\begin{bmatrix} -0.5 & 1.2 \\ -0.8 & 0.4 \end{bmatrix}$	$0.05 \pm 0.870i$	mod=0.872
iv	$\begin{bmatrix} 0.8 & 0.0 \\ 0.0 & 0.4 \end{bmatrix}$	-0.4	-0.8

VMA(2) processes

Case	A_1	A_2	μ_1, μ_2	$\mathrm{mod}(\mu_i)$	μ_3	μ_4
v	$\begin{bmatrix} 1.20 & -0.443 \\ 0.20 & 0.80 \end{bmatrix}$	$\begin{bmatrix} 0.161 & 0.00 \\ 0.00 & 0.30 \end{bmatrix}$	$-0.565 \pm 0.298i$	mod=0.639	-0.701	-0.169
vi	$\begin{bmatrix} 1.20 & 0.00 \\ 0.00 & 0.80 \end{bmatrix}$	$\begin{bmatrix} 0.440 & 0.20 \\ 0.00 & 0.11 \end{bmatrix}$	$-0.6 \pm 0.283i$	mod=0.663	-0.624	-0.176
vii	$\begin{bmatrix} 0.80 & 0.00 \\ 1.21 & 0.30 \end{bmatrix}$	$\begin{bmatrix} 0.36 & -0.40 \\ 0.284 & 0.360 \end{bmatrix}$	$0.05 \pm 0.870i$	mod=0.872	-0.8	-0.4
viii[9]	$\begin{bmatrix} 1.10 & 0.00 \\ 0.115 & 1.30 \end{bmatrix}$	$\begin{bmatrix} 0.90 & 1.00 \\ -0.258 & -0.13 \end{bmatrix}$	$-0.6 \pm 0.283i$	mod=0.663	-0.8	-0.4

---

[8]For complex-root processes this column contains the modulus.

[9]Some entries in the matrices of case VIII are rounded relative to the values necessary to generate the indicated roots.

Case i

Quantity	p	parameter: $A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
RMSE: (7)	4	0.072	0.073	0.073	0.073
(7)/ML	4	1.00	1.03	1.26	1.14
RMSE: (7)	8	0.075	0.076	0.075	0.075
(7)/ML	8	1.04	1.07	1.29	1.17

Case ii

Quantity	p	parameter: $A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
RMSE: (7)	4	0.073	0.071	0.076	0.084
(7)/ML	4	1.66	1.04	1.17	0.46
RMSE: (7)	8	0.075	0.076	0.075	0.076
(7)/ML	8	1.70	1.12	1.15	0.42

Case iii

Quantity	p	parameter: $A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
RMSE: (7)	4	0.078	0.115	0.106	0.074
(7)/ML	4	1.70	1.00	0.91	1.28
RMSE: (7)	8	0.076	0.080	0.078	0.075
(7)/ML	8	1.65	0.70	0.67	1.29

Case iv

Quantity	p	parameter: $A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
RMSE: (7)	4	0.094	0.076	0.070	0.072
(7)/ML	4	1.11	1.65	1.00	1.06
RMSE: (7)	8	0.076	0.076	0.075	0.075
(7)/ML	8	0.89	1.65	1.07	1.09

---

[10]The notation "RMSE: (7)" indicates the square root of the mean squared error of the estimator based on equation (7); the notation "(7)/ML" indicates the ratio of this root mean squared error to that of the exact ML estimator described in the text.

Table IIIa
Estimated Root-mean-squared errors of estimates of VMA(2) coefficients
First coefficient matrix

$T = 200$; 10000 replications; 1000 replications for ML

Case v

Quantity	p	parameter:			
		$A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
RMSE: (7)	4	0.220	0.107	0.129	0.092
(7)/ML	4	0.80	1.27	1.28	1.15
RMSE: (7)	8	0.077	0.076	0.075	0.076
(7)/ML	8	0.28	0.90	0.74	0.95

Case vi

Quantity	p	parameter:			
		$A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
RMSE: (7)	4	0.134	0.086	0.071	0.075
(7)/ML	4	0.45	1.18	0.85	0.96
RMSE: (7)	8	0.077	0.076	0.075	0.076
(7)/ML	8	0.26	1.04	0.89	0.97

Case vii

Quantity	p	parameter:			
		$A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
RMSE: (7)	4	0.091	0.071	0.126	0.076
(7)/ML	4	1.34	1.04	2.14	1.36
RMSE: (7)	8	0.076	0.076	0.080	0.075
(7)/ML	8	1.12	1.12	1.36	1.34

Case viii

Quantity	p	parameter:			
		$A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
RMSE: (7)	4	0.461	0.453	0.332	0.361
(7)/ML	4	3.29	2.03	2.70	1.66
RMSE: (7)	8	0.160	0.117	0.110	0.100
(7)/ML	8	1.14	0.52	0.89	0.46

Table IIIb

Estimated Root-mean-squared errors of estimates of VMA(2) coefficients

Second coefficient matrix

$T = 200$; 10000 replications; 1000 replications for ML

Case v

Quantity	p	parameter:			
		$A_2[1,1]$	$A_2[1,2]$	$A_2[2,1]$	$A_2[2,2]$
RMSE: (7)	4	0.115	0.121	0.085	0.096
(7)/ML	4	0.46	1.13	0.99	1.68
RMSE: (7)	8	0.118	0.123	0.096	0.096
(7)/ML	8	0.47	1.15	1.12	1.68

Case vi

Quantity	p	parameter:			
		$A_2[1,1]$	$A_2[1,2]$	$A_2[2,1]$	$A_2[2,2]$
RMSE: (7)	4	0.110	0.116	0.087	0.091
(7)/ML	4	0.57	1.61	1.05	1.30
RMSE: (7)	8	0.114	0.118	0.095	0.096
(7)/ML	8	0.59	1.64	1.14	1.37

Case vii

Quantity	p	parameter:			
		$A_2[1,1]$	$A_2[1,2]$	$A_2[2,1]$	$A_2[2,2]$
RMSE: (7)	4	0.087	0.117	0.112	0.129
(7)/ML	4	1.89	2.34	1.19	1.34
RMSE: (7)	8	0.094	0.098	0.119	0.120
(7)/ML	8	2.04	1.96	1.27	1.25

Case viii

Quantity	p	parameter:			
		$A_2[1,1]$	$A_2[1,2]$	$A_2[2,1]$	$A_2[2,2]$
RMSE: (7)	4	0.102	0.190	0.120	0.132
(7)/ML	4	0.55	0.57	0.83	0.52
RMSE: (7)	8	0.103	0.112	0.116	0.121
(7)/ML	8	0.56	0.34	0.81	0.47

Table IV

Estimated Root-mean-squared errors in mis-specified cases:
VMA(1) model parameters where processes are VMA(2)

$T = 200$; 10000 replications; 2000 replications for ML

Case v

Quantity	p	parameter: $A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
(7)/ML	4	0.62	0.96	0.86	0.27
(7)/ML	8	0.64	1.00	0.88	0.28

Case vi

Quantity	p	parameter: $A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
(7)/ML	4	0.24	0.39	0.81	1.17
(7)/ML	8	0.24	0.42	0.80	1.06

Case vii

Quantity	p	parameter: $A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
(7)/ML	4	0.27	0.57	0.30	1.04
(7)/ML	8	0.27	0.40	0.22	1.06

Case viii

Quantity	p	parameter: $A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
(7)/ML	4	0.13	0.17	0.28	0.24
(7)/ML	8	0.11	0.17	0.30	0.25

4. An application to wholesale and retail inventories

In this section we apply the methods just described to estimate VMA models of the bivariate system composed of aggregate Canadian inventory investments by wholesalers and retailers respectively. The data are available in real terms, quarterly from 1981 through 1997, and are used in seasonally-adjusted form.[11] Note that the available span of data comprises only 68 observations. As Galbraith and Zinde-Walsh (1994) noted in univariate MA cases, the asymptotic efficiency of ML may not be realized in small samples, and lower RMSE of estimation may be achieved with methods of the type used here.

Inventories are often considered to be well described by moving average processes because the periodic resetting of inventories to optimal levels implies a bound on the number of periods that shocks can affect inventory levels, and this point beyond which autocorrelations are nil is characteristic of the moving average process. Moreover, by a theorem of Granger and Morris (1976), the aggregation of a set of $MA(q_i)$ processes will in general lead to an aggregate process which is $MA(\max_i(q_i))$, so that we expect the same type of pattern to be visible in data which aggregate different firms and sectors. We therefore expect MA models to provide adequate characterizations of each of the data series just described (used below in de-meaned form).

The primary question that we investigate here is the propagation of shocks from wholesalers' inventories to retailers' inventories, and *vice versa*. To do so, we begin with a choice of order for the approximating autoregressions, followed by estimation of VMA models of (i) aggregate investment in inventories, (ii) investment in inventories of durable goods, (iii) investment in inventories of non-durable goods, by wholesalers and retailers respectively. We therefore deal with three bivariate systems, using six data series in all.

For each of the three systems, we chose an autoregressive order of four. Only one of the twelve F-transformed (for better finite-sample properties) LM residual autocorrelation

---

[11]All series are available from CANSIM, and are given in constant 1992 dollars, seasonally adjusted at annual rates; all refer to business investment in inventories. The series used are: D15576, retail inventories, total; D15577, retail inventories, durable goods; D15579, retail inventories, non-durable goods; D15581, wholesale inventories, total; D15582, wholesale inventories, durable goods; D15583, wholesale inventories, non-durable goods. Between the two, durable and non-durable goods make up the largest part, but not all, of the aggregate figure.

tests in these AR models, against autocorrelation in lags 1–4 and 5–8, was significant at the 5% level;[12] the SIC criterion chose an AR order of four or less for all series except D15576, for which the order chosen was five (with a value close to that for four). A still lower order was not used because a number of the AR models showed substantial coefficients at lag 4, possibly an annual-inventory effect. The choice of a relatively modest order is also consistent with the modest roots of the estimated moving-average representations, as we see below, and is in keeping with the restricted sample size as well.

Table V reports the first two VMA coefficient matrices derived from this fourth-order bivariate VAR; recall that, by the estimation method above, estimates of VMA(1) parameters are not affected by the presence or absence of VMA(2) or higher terms, and so on. There are several interesting points and regularities.

First, the roots of the determinantal equations for the VMA(1) and VMA(2) representations are quite small; these are not highly persistent series. Both retail and wholesale inventory investments show similar degrees of persistence. However, the immediate transmission of shocks is asymmetric: in the aggregate, the effect of a shock to retail inventory investment is transmitted to wholesale inventories with a coefficient of 0.47, whereas a shock to wholesale inventories shows only an effect of 0.06 (with standard error of 0.13, and so insignificantly different from zero on this sample).[13] That is, retailers appear able to offset inventory shocks to a substantial extent by changing orders from wholesalers; wholesalers do not have reciprocal short-term ability to cause retailers to adjust to their own shocks. This is compatible with the fact that the purchasing choice is made by retailers in response to demand; nonetheless, there must be eventual adjustment to supply factors communicated to retailers by wholesalers through price adjustments. We see this at the two-quarter (six-month) lag, where in aggregate data the responses of retailers to

---

[12]These statistics ranged from a low of 0.71 (lags 5–8, aggregate data, retail, distributed $F_{(4,43)}$) to a high of 3.16 (lags 5–8, durables data, wholesale, also distributed $F_{(4,43)}$). The second-highest value was 2.41. The 5% critical value is approximately 2.6.

[13]For the first VMA parameter matrix, A_1, standard errors are those of B_1 and are obtained from the VAR model variances; for the second matrix, A_2, the coefficients are transformations of those of B_1 and B_2, and approximate standard errors for each element are obtained from the Jacobian of the transformation and the VAR variances and covariances, restricting cross-equation covariances to zero.

wholesalers' shocks, and *vice versa,* are now quite close at 0.24 and 0.28 respectively.

A similar pattern appears in durable goods inventory investment; the cross-effect from wholesalers to retailers is again the only coefficient insignificantly different from zero among the VMA(1) parameters. At the two-quarter lag, the two are much closer, at 0.20 and 0.35 respectively (in VMA(3) and VMA(4) representations, not reported in Table V, the cross-effects are 0.33, 0.31 at three quarters and 0.12, 0.13 at four quarters). For non-durable goods, by contrast, the cross-effects are similar at the one-quarter lag at 0.25 and 0.20; at the two- and three- quarter lags, however, the effect of wholesalers' shocks on retailers is very close to zero. Note however that the coefficients in the models of non-durables are smaller overall, indicating less persistence in non-durables inventories in general and less scope for substantial cross-effects at longer lags. This observation would be consistent with lower costs of adjustment of non-durables inventories.

Table V
Estimated coefficient matrices, inventory investment models

X_1 : Retail; X_2 : Wholesale

Aggregate inventories:

VMA(1)	$A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
coefficient	0.39	0.06	0.47	0.37
(std error)	0.11	0.13	0.09	0.12

VMA(2)	$A_2[1,1]$	$A_2[1,2]$	$A_2[2,1]$	$A_2[2,2]$
coefficient	0.25	0.24	0.28	0.16
(std error)	0.13	0.15	0.12	0.14

Durable goods inventories:

VMA(1)	$A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
coefficient	0.45	-0.04	0.39	0.24
(std error)	0.12	0.12	0.11	0.11

VMA(2)	$A_2[1,1]$	$A_2[1,2]$	$A_2[2,1]$	$A_2[2,2]$
coefficient	0.19	0.20	0.35	0.21
(std error)	0.14	0.14	0.13	0.13

Non-durable goods inventories:

VMA(1)	$A_1[1,1]$	$A_1[1,2]$	$A_1[2,1]$	$A_1[2,2]$
coefficient	0.12	0.25	0.20	0.09
(std error)	0.10	0.13	0.09	0.12

VMA(2)	$A_2[1,1]$	$A_2[1,2]$	$A_2[2,1]$	$A_2[2,2]$
coefficient	0.19	-0.02	0.13	0.23
(std error)	0.10	0.13	0.10	0.12

5. Concluding remarks

A generalization of the results for the simple moving average estimator, obtained in Galbraith and Zinde-Walsh (1994), shows that it is convenient for use with vector moving average models, and exhibits very similar properties. The estimator is biased in finite samples; nonetheless, in cases where the moduli of roots are not too close to the unit circle, the biases become small, and mean squared errors approach their minimum values at moderate orders of approximating VAR. For sample sizes in the neighbourhood of 200, a VAR of order 4 proves to be a reasonable choice in typical cases for estimation of VMA(1) models, and VAR of order 8 for VMA(2) models. In choosing the order, there is a trade-off between a moderate loss of degrees of freedom (therefore efficiency) if order is chosen too high, versus poorer approximation (greater bias), particularly in models with roots near the unit circle, if order is chosen too low. The algorithm comes close to achieving the overall performance of ML in correctly-specified models, and appears to be substantially more robust to mis-specification of the VMA model.

We apply the method to estimation of VMA models of quarterly time series of aggregate inventory investment. The results suggest that, because the moving average roots are far from the unit circle, low-order models are sufficient to characterize the data. Although the time series is short, it is nonetheless possible to detect some asymmetry in short-term responses of wholesalers' and retailers' inventories to shocks in the other sector.

References

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.

Durbin, J. (1959) "Efficient estimation of parameters in moving-average models." *Biometrika* 46, 306–16.

Fuller, W.A. (1976) *Introduction to Statistical Time Series*. Wiley, New York.

Galbraith, J.W. and V. Zinde-Walsh (1994) "A simple, non-iterative estimator for moving average models." *Biometrika* 81, 143–155.

Galbraith, J.W. and V. Zinde-Walsh (1997) "On some simple, autoregression-based estimation and identification techniques for ARMA models." *Biometrika* 84, 685–696.

Granger, C.W.J. and R.J. Morris (1976) "Time series modelling and interpretation." *Journal of the Royal Statistical Society* A, 139, 246–257.

Hannan, E.J. and M. Deistler (1988) *The Statistical Theory of Linear Systems*. Wiley, New York.

Koreisha, S. and T. Pukkila (1990) "A generalized least squares approach for estimation of autoregressive moving average models." *Journal of Time Series Analysis* 11, 139-151.

Lütkepohl, H. (1988) "Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process." *Econometric Theory* 4, 77-85.

Lütkepohl, H. and D.S. Poskitt (1991) "Estimating orthogonal impulse responses via vector autoregressive models." *Econometric Theory* 7, 487-496.

Lütkepohl, H. (1993) *Introduction to Multiple Time Series Analysis*. (Revised ed.) Springer-Verlag, Berlin.

Osborn, D.R. (1977) "Exact and approximate maximum likelihood estimators for vector moving average processes." *Journal of the Royal Statistical Society*, B, 39, 114–18.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1986) *Numerical Recipes: the Art of Scientific Computing*. Cambridge University Press, Cambridge.