INFERENCE IN DYNAMIC MODELS CONTAINING 'SURPRISE' VARIABLES*

Richard T. BAILLIE

Michigan State University, East Lansing, MI 48824, USA

Some new results on calculating moving average representation (MAR) coefficients and their limiting distribution from estimated vector ARMA processes are presented. The technique is applied to the problem of estimating the coefficients of unanticipated or 'surprise' variables in a single equation for a multi-period expectations horizon. The method naturally conditions the expectations on all past values of the process and avoids the necessity of using two-step regression procedures and adjusting the resulting standard errors.

1. Introduction

Many recent macroeconomic studies have considered situations where unanticipated or 'surprise' variables enter the formulation of a model. Some of the initial empirical work in this area was performed by Barro (1977, 1978), Sheffrin (1979) and Barro and Rush (1980) where unanticipated money growth was used as an explanatory variable to account for deviations from the natural rate of output or employment. Other examples include unanticipated money supply differentials and unanticipated real output differentials to explain unanticipated changes in the nominal exchange rate, e.g., Edwards (1982a, 1982b, 1983), Copeland (1984) and MacDonald (1984).

Many of these studies use a two-step regression procedure to estimate the coefficient of the unanticipated variable. First the residuals from a separate auxiliary regression are taken as a proxy for the unanticipated variable and are then used as an explanatory variable in the equation of interest. Pagan (1984) has considered some of the econometric problems that may arise in these situations and has shown how the standard errors of coefficient estimates can be adjusted to avoid a downward bias that sometimes occur when residuals from an auxiliary regression are used as explanatory variables.

This paper shows that when the data-generating process can be interpreted as including an equation with an unanticipated or 'surprise' variable, then that

*This paper was written while I was visiting the University of California, San Diego. I am very grateful to Tim Bollerslev, Rob Engle, to two anonymous referees, and particularly to Peter Phillips who all made extremely helpful comments on an earlier version of the paper. Thanks also go to Leigh Henry and Maxine Shafer for their skillful and efficient typing.

equation should be implicit in the unrestricted moving average representation (MAR) of the variables in the system.

When an equation contains an explanatory variable which is the rational expectation of the dependent variable \( l \) periods in the future, then the resulting disturbance will be a moving average process of order \( l - 1 \). Section 2 of the paper shows how the disturbance can be regarded as the sum of separate component moving average processes with the interpretation of being 'surprise' or 'news' on different variables. Furthermore, their coefficients in the equation of interest are reparameterizations of the MAR coefficients.

Section 3 shows how estimates of vector MAR coefficient matrices can be derived from the estimated finite-order vector autoregression (VAR) and vector autoregressive moving average (VARMA) models. The technique is then applied in section 4 to obtain parameter estimates and associated standard errors of 'surprise' variables in the equation of interest.

The suggested procedure appears to have several desirable features:

(a) It naturally conditions expectations on current and past values of the variables in the process and hence avoids the problem of having to artificially generate expectations or news from a separate regression.

(b) The technique allows news to be considered more than one period ahead and avoids the need for generalizing Pagan's (1984) technique to this situation.

(c) The estimated parameters of interest and their standard errors can be found without two-step procedures and complicated adjustments of their respective standard errors.

At the same time it should be acknowledged that a potential disadvantage with the technique is that it does require the specification of a complete multivariate time-series model, rather than just being based on single-equation estimation as in previous studies. Also the matrix algebra needed to compute the required quantities is formidable, though by no means overwhelming. Indeed a numerical example of applying the method to modeling surprises of the interest rate differential variable in a model relating the forward rate to the future spot rate is also presented in section 4.

2. Properties of component moving average processes

Consider a \( g \)-component vector \( y \), whose second element is the observed expectation of the first element at some future time period. Hence

\[
y_{2t} = E_t y_{1,t+1},
\]

where \( E_t \) represents the expectations operator conditioned on current and past
values of the process \( y_t \) at time \( t \). Brown and Maital (1981) and Hayashi and Sims (1983), among others, have considered the testing of relationships such as (1) when \( l > 1 \). The error associated with (1) is given by

\[
u_{t+1} = y_{1,t+1} - y_{2t},
\]

and \( \text{E}(u_t, u_{1-t}) = 0 \) for \( j \geq l \); so that \( u_t \) could be a moving average process of order \( l - 1 \), i.e., MA\((l-1)\), although non-linear processes such as bi-linear models would also satisfy this property. Also assume that \( y_t \) is a linear, non-deterministic jointly covariance stationary process with a Wold decomposition and unique infinite-order MA process given by

\[
y_t = \sum_{j=0}^{\infty} B_j e_{t-j},
\]

where

\[
B_0 = I,
\]

\[
\text{E}(e_t e'_s) = \Omega, \quad s = t,
\]

\[
= 0, \quad s \neq t.
\]

Since

\[
y_{1,t+1} - y_{2t} = e'_t \sum_{j=0}^{\infty} B_j e_{t+1-j} - e'_2 \sum_{j=0}^{\infty} B_j e_{t-j},
\]

where \( e'_t \) is a \( g \)-dimensional row vector, full of zeros except for unity in its \( k \)th element; then from (2)

\[
u_{t+1} = e'_t \sum_{j=0}^{l-1} B_j e_{t+1-j} + \eta_t,
\]

where

\[
\eta_t = \sum_{j=0}^{\infty} (e'_t B_{1+j} - e'_2 B_j) e_{t-j},
\]

and under the assumption of rationality contained in (1) it follows that \( \eta_t = 0 \), so that all relevant and worthwhile information is contained in \( y_{2t} \), and the past history of \( \eta_t \) contains no important extra information. Given (1) and (5) it follows that \( u_{t+1} \) in (4) will be a moving average process of order \( l - 1 \), i.e.,
MA(\(l-1\)), as in eq. (2). Violation of the property of rationality will require \(\eta_t\) and \(u_t\) to be infinite-order MA processes. It is of interest to focus on the interpretation of \(u_{t+l}\) under the assumption of rational expectations and straightforward to show that with \(\eta_t = 0\), \(u_{t+l}\) in (4) can be expressed as

\[
u_{t+l} = \sum_{j=0}^{l-1} (e_t B_j e_1) \varepsilon_{1,t+l-j} + \sum_{k=2}^{g} \sum_{j=1}^{l-1} (e_t B_j e_k) \varepsilon_{k,t+l-j}.
\] (6)

Hence \(u_t\) can be represented as the sum of \(g\)-component moving average processes. The properties of such processes have been studied previously by Granger and Morris (1976), Ansley, Spivey and Wrobleski (1977), MacDonald and Darroch (1983), and others. Most of these studies have assumed contemporaneous uncorrelated increments between the different component processes. In the above formulation \(u_{t+l}\) can be expressed as

\[
u_{t+l} = \sum_{k=1}^{g} \nu_{k,t+l},
\]

where \(\nu_{k,t+l}\) is the unanticipated component associated with the \(k\)th variable and can be represented as

\[
u_{k,t+l} = y_{k,t+l} - E_t y_{k,t+l} = \sum_{j=1}^{l-1} (e_t B_j e_k) \varepsilon_{k,t+l-j}.
\] (7)

Then the 'news' or 'surprise' coefficient associated with the \(k\)th innovation in period \(t+l-j\) is given by

\[\gamma_{jk} = e_t B_j e_k,\] (8)

and the total surprise coefficient is then

\[\gamma_k = \sum_{j=1}^{l-1} (e_t B_j e_k).
\]

Since the hypothesis of interest (1) concerns the first and second elements of \(y_t\), it follows that in a regression of the form

\[y_{1,t+l} = \psi y_{2,t} + \gamma_k u_{k,t+l} + v_{t+l},\]

the hypothesis of rationality embodied in eq. (1) restricts \(\psi = 1\) and \(v_{t+l}\) to be
an MA($l - 1$) process which is the sum of $(g - 1)$ separate component processes. It is then of interest to determine the coefficient $\gamma_k$, also defined in (7), which is the coefficient associated with $u_{k,t+1}$, the 'surprise' on variable $k$, where $k$ is any integer between 3 and $g$ inclusive.

3. The distribution of estimated MAR coefficient matrices

This section develops a simple procedure for the estimation of the news coefficients and their asymptotic standard errors. To facilitate this, we first show how infinite-order moving average representation (MAR) coefficients can be calculated from a finite-parameter multivariate time-series model and then derive the MAR coefficient estimators limiting distribution.

Suppose $y_t$ can be represented by a model containing a finite number of parameters $\theta$ and from a sample of $n$ observations; the limiting distribution of the maximum likelihood estimator $\hat{\theta}$ is given by

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, V),$$

and from (3) the row vectorized coefficients of the $j$th MA coefficient matrix $B_j$ are given by

$$\beta_j' = \text{vec}(B_j),$$

and $\beta_j$ is clearly a non-linear function of the original models parameters $\theta$ so that

$$\beta_j' = f(\theta).$$

Then to a first-order approximation

$$\sqrt{n} (\hat{\beta}_j - \beta_j) \sim N(0, D_j'V D_j),$$

where

$$D_j = \partial \beta_j / \partial \theta|_{\theta = \hat{\theta}}.$$

Clearly (10) is a very general relationship that is true for any vector linear time-series process that depends on some basic parameters $\theta$. In order to make (10) of direct use, it is necessary to consider some specific cases:
Case I: The VAR(p) model

Suppose \( y_t \) can be represented as a vector autoregression of order \( p \), i.e., VAR(\( p \)), so that

\[
y_t = \sum_{j=1}^{p} A_j y_{t-j} + \epsilon_t, \tag{11}
\]

and the \( \epsilon_t \) are defined as in (3). Then in first-order system form (11) can be expressed as

\[
\begin{bmatrix}
  y_t \\
  y_{t-1} \\
  \vdots \\
  y_{t-p+1}
\end{bmatrix} = \begin{bmatrix}
  A_1 & A_2 & \cdots & A_{p-1} & A_p \\
  I & 0 & & & \\
  \vdots & \vdots & \ddots & \ddots & \\
  0 & \cdots & \cdots & I & 0
\end{bmatrix} \begin{bmatrix}
  y_{t-1} \\
  y_{t-2} \\
  \vdots \\
  y_{t-p}
\end{bmatrix} + \begin{bmatrix}
  \epsilon_t \\
  0 \\
  \vdots \\
  0
\end{bmatrix}.
\]

Or,

\[
y_t = A Y_{t-1} + \eta_t. \tag{12}
\]

On defining the \((g \times gp)\)-dimensional matrix \( N \) as

\[
N' = [I : 0],
\]

and premultiplying through (12) by \( N' \),

\[
y_t = N' A Y_{t-1} + \epsilon_t,
\]

and the MAR coefficient matrices \( B_j \) in (3) can be derived from

\[
B_j = N' A^j N, \tag{13}
\]

so that successive MAR coefficient matrices can be obtained as the north-west blockmatrix of \( A \) raised to different powers.

From (9),

\[
\beta'_j = \text{vec}(N'A^j N) = \text{vec}(N'A^j)(I \otimes N).
\]

If \( y_t \) in (11) is a purely unrestricted VAR(\( p \)) model containing \( g^2 p \) parameters, then

\[
\theta' = \text{vec}(N'A),
\]
and
\[ \sqrt{n} (\hat{\theta} - \theta) \sim N(0, \Omega \otimes \Gamma^{-1}), \]  
(14)
where
\[ \Gamma = \text{plim} \frac{1}{n} \sum Y_{t-1} Y_{t-1}'. \]
Then \( D_j \) in eq. (10) is given by
\[ D_j = \frac{\partial \text{vec}(N^t A^j) (I \otimes N)}{\partial \theta}, \]
and on invoking a standard result originally due to Schmidt (1973),
\[ D_j = N^t \left( \sum_{k=0}^{j-1} A^k \otimes A^{j-1-k} \right) (I \otimes N), \]
which in this case reduces to the simpler expression of
\[ D_j = \left( \sum_{k=0}^{j-1} B_k' \otimes A^{j-1-k} \right) (I \otimes N), \]
(15)
and
\[ \sqrt{n} \left( \hat{\beta}_j - \beta_j \right) \sim N(0, D_j' V D_j). \]
(16)

It should be noted that although the above provides a proper limiting distribution for the MAR coefficient estimates, the covariance matrix will be singular for \( j > p \), so that for higher-order MAR coefficients it will not be possible to produce confidence intervals.

While the above formula applies to unrestricted VAR models it is generally desirable to have a parsimonious parameterization. For example, Doan, Litterman and Sims (1984) have recently shown that a very small number of time-varying parameters derived from a Bayesian prior distribution can be quite successful in parameterizing high-order, high-dimensional VAR systems.

**Case II: The VARMA(p, q) model**

In order to treat the situation where \( y_t \) is generated by the vector autoregressive moving average process of order \( p \) and \( q \) so that
\[ y_t = \sum_{j=1}^{p} A_j y_{t-j} + \epsilon_t + \sum_{j=1}^{q} C_j \epsilon_{t-j}, \]
(17)
it is necessary to use a form of the state space representation as given by

\[
\begin{bmatrix}
y_t \\
y_{t-1} \\
\vdots \\
y_{t-r+1} \\
\epsilon_t \\
\vdots \\
\epsilon_{t-r+1}
\end{bmatrix} = \begin{bmatrix} A_1 & \cdots & A_{r-1} & A_r & C_1 & \cdots & C_{r-1} & C_r \\
I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \begin{bmatrix} y_{t-1} \\
y_{t-2} \\
\vdots \\
y_{t-r} \\
\epsilon_{t-1} \\
\epsilon_{t-2} \\
\vdots \\
\epsilon_{t-r}
\end{bmatrix} + \begin{bmatrix} \epsilon_t \\
\vdots \\
\epsilon_t \\
\epsilon_t \\
\vdots \\
\epsilon_t \\
\epsilon_t \\
\epsilon_t 
\end{bmatrix}.
\]

Or,

\[ Y_t = A Y_{t-1} + \eta_t, \quad (18) \]

and

\[ r = \max(p, q). \]

From this formulation it can be seen that the MAR coefficient matrices \( B_j \) are given by

\[ B_j = N'A/M, \]

where \( M \) is a \((g \times 2gr)\)-dimensional matrix defined as

\[ M' = [I:0:I:0], \]

where both the null matrices are of order \( g \times g(r - 1) \). Again we define the parameter vector \( \theta \), which now contains \( g^2(p + q) \) parameters as

\[ \theta' = \text{vec}(N'A), \quad \sqrt{n}(\hat{\theta} - \theta) \sim N(0, V). \]

Now

\[ D_j = \frac{\partial \text{vec}(N'A^j)(I \otimes M)}{\partial \theta} = N' \left( \sum_{i=0}^{j-1} A'^i \otimes A^{j-1-i} \right)(I \otimes M), \quad (19) \]

and substitution for \( D_j \) and \( V \) realizes a parametric expression for eq. (10). The exact details of \( V \) can be found in Hannah (1970).
Case III: The VAR(p) model with standardized error covariance matrix

Following Sims (1980), it has become popular in much applied macroeconometric analysis to give MAR plots once the error covariance matrix has been diagonalized, or reduced to an identity matrix. This can be readily done by premultiplying through (11) by a lower triangular matrix $R$ with diagonal elements greater than zero.

Then

$$Ry_t = \sum_{j=1}^{p} RA_j y_{t-j} + \eta_t,$$

(20)

where

$$\eta_t = Re_t \quad \text{and} \quad C_j = B_j R^{-1}.$$

Then

$$E(\eta_t \eta_t') = R\Omega R' = I, \quad s = t,$$

$$= 0 \quad \text{if} \quad s \neq t.$$

The standard MAR corresponding to (20) is given by

$$y_t = \sum_{j=0}^{\infty} C_j \eta_{t-j},$$

(21)

where

$$C_0 = R^{-1},$$

and

$$C_j = N' A^j NR^{-1}, \quad j \geq 1,$$

(22)

so that a typical MAR coefficient matrix is now a function of $\theta$, the VAR parameters, and also of the elements of $R^{-1}$ which is a function of the error covariance matrix parameters. On defining the $g(g+1)/2$ unique parameters of $\Omega$ as $\omega$, and given a $(g^2 \times g(g+1)/2)$-dimensional selection matrix $J$ of zeros and ones, then

$$\text{vec} \ \Omega = J\omega \quad \text{and} \quad \theta' = \text{vec}(N' A).$$
It is shown in the appendix that
\[
\sqrt{n} \left[ (\hat{\theta} - \theta)' : (\hat{\omega} - \omega)' \right] \sim N(0, V),
\]
where
\[
V = \begin{bmatrix}
\Omega \otimes \Gamma^{-1} & 0 \\
0 & 2 \left[ J'(\Omega^{-1} \otimes \Omega^{-1}) J \right]^{-1}
\end{bmatrix}.
\]
(23)

Thus the MLE of \( \theta \) and \( \omega \) are asymptotically uncorrelated with a well-defined limiting distribution. The vectorized \( j \)th standardized MAR coefficient matrix is given by
\[
\beta_j' = \text{vec}(C_j) = \text{vec}(N' A^j N R^{-1}).
\]

Then to first-order approximation
\[
(\beta_j - \beta_j)' = \left[ (\theta - \theta)' : (\omega - \omega)' \right] \begin{bmatrix}
D_1' \\
D_2'
\end{bmatrix},
\]
and from (23) it follows that
\[
\sqrt{n} \left( \hat{\beta}_j - \beta_j \right) \sim N\left( 0, D_1' (\Omega \otimes \Gamma^{-1}) D_1 \right.
\]
\[
+ 2 D_2' \left[ J'(\Omega^{-1} \otimes \Omega^{-1}) J \right]^{-1} D_2 \},
\]
(24)

where from (15)
\[
D_1 = \frac{\partial (N' A^j N R^{-1})}{\partial \theta} = \frac{\partial \text{vec}(N' A^j) (I \otimes NR^{-1})}{\partial \theta}
\]
\[
= N' \left( \sum_{i=0}^{j-1} B_i \otimes A^{j-1-i} \right) (I \otimes NR^{-1}),
\]
and
\[
D_2 = \frac{\partial (N' A^j N R^{-1})}{\partial \omega} = \frac{\partial \text{vec}(R^{-1}) (N' A^j N \otimes I)}{\partial \omega}.
\]
Unfortunately it does not seem possible to write down a simple closed form expression for \( \partial \text{vec}(R^{-1}) / \partial \omega \); but for modest sized systems, the expression is easy to derive analytically.

4. Modeling surprise variables

In order to demonstrate a specific use of the results developed in the previous section consider the model

\[
y_{1,t+1} = \psi y_{2t} + \gamma_k u_{k,t+1} + \nu_{t+1},
\]

where \( u_{k,t+1} \) is MA(\( l - 1 \)) as defined in eq. (7), while \( \nu_{t+1} \) is also MA(\( l - 1 \)) so that \( \nu_{t+1} = u_{t+1} - u_{k,t+1} \), where \( u_{t+1} \) is the aggregate MA(\( l - 1 \)) process defined in eq. (4). Under the null hypothesis \( y_{2t} = E_t y_{1,t+1} \), which implies that \( \psi = 1 \) and \( u_{k,t+1} \) is the 'news' or 'surprise' on the \( k \)th variable in the information set \( Y_r \).

An example of the above is given by the literature on whether or not the logarithm of the forward exchange rate \( f_t \) is an efficient predictor, or expectation of the logarithm of the future spot exchange rate \( s_t \). So that under the null hypothesis

\[
f_t = E_t s_{t+1}.
\]

Mussa (1979) noted that during the 1970's the variance of the actual change of the spot rate was some twenty times the variance of the expected change. Mussa attributed some 90\% of exchange rate fluctuations to the presence of unanticipated information. These excessive fluctuations in the exchange rate presumably occur in periods of great uncertainty where expectations of future fundamentals are disturbed by news, announcements and the general acquisition of new information. Similarly, Frenkel (1981) noted that the 'key factor affecting exchange rates has been news' and led him to estimate an equation of the form

\[
s_{t+1} = \psi f_t + \gamma (x_{t+1} - E_t x_{t+1}) + u_{t+1},
\]

where \( x_t \) is the interest rate differential between a domestic and foreign country. Apart from testing the 'efficiency' of the logarithm of the forward rate, i.e., whether or not \( \psi = 1 \), Frenkel was also interested in the sign of \( \gamma \) and whether or not it was significant. Given an exchange rate expressed in terms of the number of U.S. dollars to a currency, an unanticipated rise in domestic U.S. interest rates may attract foreign capital and lead to a capital account surplus, or reduce domestic expenditure and cause a surplus in the current account, or through the interest rate parity condition lead to a higher
forward premium, or a combination of any two, or all three. In any event, these mechanisms will lead to an appreciation of the domestic U.S. currency and hence \( \gamma < 0 \). Alternatively, a period dominated by inflationary expectations will see a rise in the domestic rate of interest and may lead to a domestic currency depreciation and hence to a positive \( \gamma \) news coefficient. This mechanism works through the route of reducing the demand for real balances and producing a situation where asset market equilibrium requires a price level higher than prevailing prices so that the spot rate depreciates to bring purchasing power parity into line.

In order to estimate eq. (25), Frenkel (1981) generated the news variable \( \{x_{t+1} - E_t x_{t+1}\} \) by regressing the interest rate differential on its last two lags and the lagged forward rate and using the residuals as a genuine explanatory variable in (25). This running of auxiliary regression equations has been examined by Pagan (1984), who has shown that in certain cases it will lead to downward biased standard errors of parameter estimates. Extremely similar methodology to Frenkel (1981) has also been adopted by Turnovsky and Ball (1983) who consider the interest rate differential, and Edwards (1982a, b), Copeland (1984) and MacDonald (1984) who consider money supply differentials and real output differentials. All of these studies use extremely restrictive information sets to generate \( E_t x_{t+1} \), e.g., Frenkel (1981) and Turnovsky and Ball (1983) used the last two lags on the interest rate differential and the lagged forward rate, Edwards (1982a) uses residuals from a third-order scalar autoregression to generate news on money supply differentials and real output differentials.

In order to generalize the above approaches and to illustrate the methodology developed earlier in the previous section of this paper, we consider estimation of the model

\[
(s_{t+1} - s_t) = \psi(f_t - s_t) + \gamma(x_{t+1} - E_t x_{t+1}) + u_{t+1},
\]

so that the actual change in the spot exchange rate is considered a function of the expected change \( (f_t - s_t) \), and news on the change in the interest rate differential \( x_t \) and \( u_{t+1} \) will be an MA(1) error process.

The approach outlined previously can be used to generate the parameter associated with the multi-step news variable and will automatically be conditioned on current and past values of the process. The data set consisted of 328 weekly observations taken from the New York foreign exchange market between September 1973 and April 1980. The spot rate data were recorded on a Tuesday and the forward rate data on a Monday to ensure exactly 90 days difference between the variables.

A VAR model of order two (i.e., \( p = 2 \)) was estimated for the West German 90-day forward rate, spot rate, and appropriate Eurocurrency bond rates of
90-day maturity times. Likelihood ratio test statistics indicated that the choice of \( p = 2 \) could not be rejected against an alternative higher-order model with \( p = 8 \). The exact variables used in the estimated VAR model were

\[
y_t' = \left[ (s_t - s_{t-12}) (f_t - s_t) (i_t - i_{t-12}) \right],
\]

where \( i_t \) is the interest rate differential. A test of the hypothesis given by eq. (1) is of the form

\[
(s_{t+12} - s_t) = E_t (f_t - s_t),
\]

(27)

and is basically similar to tests performed on spot rates and 30-day forward rates by Baillie, Lippens and McMahon (1983). General properties of tests of models with forward expectations variables and estimation of models containing cointegrated variables subject to rational expectations restrictions are given by Baillie (1986). In fact Wald and Likelihood Ratio statistics of (27) were 15.72 and 15.29, respectively, which were at the 0.20 significance level for a chi-squared distribution with six degrees of freedom.

The associated \( \gamma_{jk} \) coefficients in eqs. (7) and (8), and hence in the regression

\[
y_{1,t+12} = \psi y_{2t} + \sum_{j=1}^{11} \gamma_{j3} e_{3,t+12-j} + v_{t+1}.
\]

(28)

were obtained from (8) as

\[
\hat{\gamma}_{j3} = e_j \hat{\beta}_j e_3 = e_j \hat{\beta}_j.
\]

Then

\[
\text{var}(\hat{\gamma}_{j3}) \approx \frac{1}{n} e_j' D_j' V D_j e_3 \quad \text{and} \quad \text{cov}(\hat{\gamma}_{j3} \hat{\gamma}_{i3}) \approx \frac{1}{n} e_j' D_j' V D_i e_3,
\]

and the total news coefficient \( \gamma_3 \) is then estimated by

\[
\hat{\gamma}_3 = \sum_{j=1}^{11} \hat{\gamma}_{j3}.
\]

The estimates were obtained from unrestricted MLE so that some autocorrelation was still present in \( \eta_t \) in (4). A more complicated but feasible alternative would be to obtained restricted MLE of the VAR parameters so that condition
R.T. Baillie, Models containing 'surprise' variables

Table 1
Estimated parameters of 'news' on interest rate differential between West Germany and USA.

<table>
<thead>
<tr>
<th>j</th>
<th>( \hat{\gamma}_j )</th>
<th>s.e. (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.427</td>
<td>0.032</td>
</tr>
<tr>
<td>2</td>
<td>0.277</td>
<td>0.071</td>
</tr>
<tr>
<td>3</td>
<td>0.098</td>
<td>0.089</td>
</tr>
<tr>
<td>4</td>
<td>0.115</td>
<td>0.072</td>
</tr>
<tr>
<td>5</td>
<td>-0.189</td>
<td>0.084</td>
</tr>
<tr>
<td>6</td>
<td>-0.133</td>
<td>0.087</td>
</tr>
<tr>
<td>7</td>
<td>0.119</td>
<td>0.069</td>
</tr>
<tr>
<td>8</td>
<td>0.004</td>
<td>0.081</td>
</tr>
<tr>
<td>9</td>
<td>-0.084</td>
<td>0.083</td>
</tr>
<tr>
<td>10</td>
<td>-0.098</td>
<td>0.082</td>
</tr>
<tr>
<td>11</td>
<td>0.092</td>
<td>0.097</td>
</tr>
<tr>
<td>Total</td>
<td>0.627</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) s.e. = standard error.

(27) was imposed, which is equivalent to restricting \( \psi \) to unity in (28). Full details of such restricted estimates in these and more complicated models are to be found in Baillie (1986).

The news coefficients were estimated from the VAR(2) model with unrestricted error covariance matrix so that no diagonalization was imposed; so that (16) gave the limiting distribution for the MAR coefficients estimates. The estimated \( \gamma_{3} \), the 'news' or 'surprise' parameters, are presented in table 1. It should be noted that most of them are positive which lends support to the 'inflationary expectations' theory. The overall sum is also positive with a significant standard error.

5. Conclusion

This paper has derived fairly simple methods to compute and also to find the limiting distribution of MAR coefficients from various finite parameter vector linear time-series models. The methods allow the parameter values and asymptotic standard errors to be calculated for parameters associated with news or surprise variables. Unlike previous studies, the methods allow multi-step expectations models to be considered; where MA errors are present and hence avoid auxiliary regressions and the adjustment problems discussed by Pagan (1984). The methods also automatically generate news as the difference between a variables actual outcome and an expectation conditioned on all past and current values of a variable, which seems more consistent with the notion of rational expectations than previous studies which rather arbitrarily choose information sets to use in auxiliary regressions. In certain situations, the approach appears a reasonably attractive method to generate news coefficients.
Appendix

On assuming that $\varepsilon_t$ is NID(0, $\Omega$) in eqs. (3) and (12), and given a sample of $n$ observations, it is necessary to maximize the log likelihood function

$$\ln l = c - \frac{n}{2} \ln |\Omega| - \frac{1}{2} \sum_{t=1}^{n} (y_t - A Y_{t-1})' \Omega^{-1} (y_t - A Y_{t-1}).$$

In the case where $\theta = \text{vec}(A)$ is unrestricted so that no elements are set to zero, OLS provides asymptotically efficient estimates of $\theta$, and $\Omega$ is estimated as

$$\hat{\Omega} = n^{-1} \sum_{t=1}^{n} (y_t - \hat{A} Y_{t-1}) (y_t - \hat{A} Y_{t-1})' = n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_t'.$$

It can be shown that

$$\frac{\partial \ln l}{\partial \text{vec} \Omega} = -\frac{n}{2} \text{vec} \Omega^{-1} + \frac{1}{2} (\Omega^{-1} \otimes \Omega^{-1})$$

$$\times \text{vec} \left\{ \sum_{t=1}^{n} (y_t - A Y_{t-1}) (y_t - A Y_{t-1})' \right\}. \quad (A.1)$$

Since

$$\frac{\partial \text{vec} \Omega^{-1}}{\partial \text{vec} \Omega} = -(\Omega^{-1} \otimes \Omega^{-1}),$$

it follows that

$$\frac{\partial^2 \ln l}{\partial \text{vec} \Omega (\partial \text{vec} \Omega)'} = \frac{n}{2} (\Omega^{-1} \otimes \Omega^{-1}) - \frac{1}{2} \Omega^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_t' (\Omega^{-1} \otimes \Omega^{-1})$$

$$- \frac{1}{2} (\Omega^{-1} \otimes \Omega^{-1}) \sum_{t=1}^{n} (\hat{\varepsilon}_t \hat{\varepsilon}_t' \Omega^{-1}). \quad (A.2)$$

If we follow the route of taking the expectation of minus the second derivatives of the log likelihood to form the information matrix in the usual way, it is easily shown that the asymptotic covariance matrix of (vec $\hat{\Omega}$) is given by

$$(2/n)(\Omega \otimes \Omega),$$

a result which has been previously derived by an alternative method by Rothenberg (1973, pp. 87–90). However, the result fails to recognize the duplication of elements in vec $\Omega$. For inferential procedures, it is more
helpful to define $\omega$ as the vector containing $g(g+1)/2$ distinct parameters. Then

$$\text{vec } \Omega = J\omega,$$

and $J$ is a selection matrix consisting of zeros or ones and is of dimension $g^2 \times g(g+1)/2$. For example, in the case of $g = 3$,

$$J = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}.$$  

Since

$$\frac{\partial \ln l}{\partial \omega} = \frac{\partial \text{vec } \omega}{\partial \omega} \frac{\partial \ln l}{\partial \text{vec } \Omega} = J' \frac{\partial \ln l}{\partial \text{vec } \Omega},$$

and appropriately modifying eqs. (A.1) and (A.2), we obtain

$$\frac{\partial \ln l}{\partial \omega \partial \omega'} = -\frac{n}{2} J'(\Omega^{-1} \otimes \Omega^{-1})J,$$

so that result (23) easily follows.

References


