86-F-2

On Incoherency of Testing Rational Expectation Hypotheses by Vector Autoregressive Models

by

Naoto Kunitomo
University of Tokyo

and

Taku Yamamoto
Yokohama National University

April 1986

PREPARED UNDER
THE PROJECT ON ECONOMETRICS
RESEARCH INSTITUTE FOR THE JAPANESE ECONOMY

Abstract

We show two methodological difficulties for testing rational expectation (RE) hypotheses based upon fitting vector autoregressive (VAR) time series models. The methods used in a number of econometric studies to test RE hypotheses for the term structure of interest rates and the forward foreign exchange markets are shown to be incoherent with the RE hypotheses in their theoretical consideration. We found that in most cases random walk processes are not consistent with RE hypotheses. Incidentally, Shiller's assertion on the RE hypothesis for the term structure of interest rates (1981) can be viewed as a special case of our Corollary 1 in Section 2. We explore the relationship between the RE hypotheses and the cross-equation restrictions imposed by those in the vector autoregressive moving-average (VARMA) time series models. Our results suggest that in most cases the conventional use of VAR modelling in macroeconometric applications is not consistent with RE hypotheses.

1. Introduction

Econometric analyses with rational expectations (RE) have been voluminous in the last decade. The type of rational expectation (RE) hypotheses of interest in the present paper is expressed as

$$(1.1)\qquad \sum_{i=1}^{m_1} \sum_{j=0}^{n_i-1} w_{ij} E(y_{i\,t+j} \mid I_t) = 0 \quad ,$$

for $t=0, \pm 1, \pm 2, \ldots,$ where $w_{ij} (i=1,\ldots,m_1;\ j=0,1,\ldots,n_i-1)$ are some constants, m_1 is the number of relevant variables included in RE hypotheses, and $n_i\ (i=1,\ldots,m_1)$ are some fixed integers; a vector $(y_{1t},\ldots,y_{m_1 t})$ is a subset of an $m\times 1$ stochastic process $\{y_t\}$ we are considering; I_t is the information set available at period t and $E(.\mid I_t)$ is the conditional expectation operator given I_t.

Several methods have been proposed to test the RE hypotheses (1.1). Among them, one method commonly used in empirical studies is to fit vector autoregressive (VAR) time series models and construct statistical test procedures on the nonlinear cross-equation restrictions imposed by the RE hypotheses. Originally, Sargent (1979) proposed this method in connection with a RE hypothesis in the term structure of interest rates. In his study, the hypothesis of interest is (1.1) when $m_1 = 2$ and

$$(1.2)\qquad w_{1j} = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 \end{cases} \quad , \quad w_{2j} = -\frac{1}{n_2} \quad (j=0,1,\ldots,n_2-1),$$

where y_{1t} is the long-term interest rate and y_{2t} is the short-term interest rate. Later, Hakkio (1981a, 1981b), Baillie et.al. (1983), and Ito (1985) applied some variants of this method for testing the RE hypothesis in the foreign exchange rate market. In their studies the hypothesis of interest is (1.1) when $m_1 = 2$ and

$$(1.3) \qquad w_{1j} = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 \end{cases}, \quad w_{2j} = \begin{cases} -1 & \text{if } j = h \ (h \geq 1) \\ 0 & \text{if } j \neq h \end{cases},$$

where h stands for the prediction horizon of economic agents, y_{1t} is the forward exchange rate, and y_{2t} is the spot exchange rate.

This paper exhibits two kinds of serious problems inherent in this method of which econometricians have been often unaware. Our results imply that most previous studies using this approach, some published in this journal, are logically misleading with questionable empirical results. More specifically, Section 2 shows that if we fit VAR models to filtered time series, the RE hypotheses of the type (1.1) never hold for the original stochastic processes. The difference filter commonly used in practice is an example of our general formulation. Incidentally, Shiller's assertion on the RE hypothesis for the term structure of interest rates (1981) is a special case of Corollary 1 in Section 2. In Section 3 we discuss a problem in treating information sets when we construct statistical tests based on the VAR models. It will be seen that the conventional procedure using the statistical tests based upon a limited information set is likely to lead to a model misspecification. It is not necessarily appropriate for testing RE hypotheses of the type (1.1). We shall point out that a number of previous studies suffer from incoherency problems with respect to their RE hypotheses in the theoretical consideration. Section 4 summarizes our results in this paper. We shall point out that our results have implication for VAR modelling in recent macroeconometric applications. In most cases, the method frequently used in empirical studies is not consistent with the RE hypotheses.

2. Testing RE hypotheses by VAR Models

In this section we first present a general method of testing the RE hypotheses by VAR models, and then show a serious incoherency when VAR models are fitted to filtered time series. Suppose that an m-variate time series $\{y_t\}$ is generated by the following vector autoregressive (VAR) process with order p, denoted by $VAR_m(p)$,

$$(2.1) \qquad y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t ,$$

where $y_t = (y_{1t}, \dots, y_{mt})$; $u_t = (u_{1t}, \dots u_{mt})$ is the disturbance vector with $E(u_t) = 0$, $E(u_t u_t') = \Omega$ (positive definite), and $E(u_t u_s') = 0$ for $t \neq s$; $A_1, \dots, A_p$ are $m \times m$ coefficient matrices.[1/] The process $\{y_t\}$ can be either stationary or nonstationary at this stage. However, if some of the absolute values of characteristic roots of the associated equation $| z^p I_m - \sum_{j=1}^{p} A_j z^{p-j} | = 0$ are equal or greater than one, we assume that the initial values y_0, $y_{-1}, \dots, y_{-(p-1)}$ are fixed. The process (2.1) can be expressed in a Markovian form:

$$(2.2) \qquad Y_t = A\, Y_{t-1} + U_t ,$$

where $Y_t = (y_t', y_{t-1}', \dots, y_{t-p+1}')'$ is an $mp \times 1$ vector, $U_t = (u_t', 0, \dots, 0)'$ is an $mp \times 1$ vector, and

$$A = \begin{pmatrix} A_1 \; A_2 \; \dots \; A_{p-1} & A_p \\ I_{m(p-1)} & 0 \end{pmatrix} \; (mp \times mp).$$

Using the above representation, it is straightforward to derive the optimal predictor of Y_{t+h} ($h \geq 1$) given the information set I_t, where $I_t=\{y_t, y_{t-1}, \ldots\}$. By repeating the insertion of (2.2), we express Y_{t+h} as

$$(2.3) \qquad Y_{t+h} = A^h Y_t + \sum_{i=0}^{h-1} A^i U_{t+h-i} .$$

The second term on the right-hand side of (2.3) consists solely of future disturbances while the first term is in the set I_t. Thus in this case the optimal predictor of Y_{t+h} given I_t is the least squares prediction of Y_{t+h} on I_t:

$$(2.4) \qquad E(Y_{t+h} | I_t) = A^h Y_t .$$

From (2.4) we obtain the following lemma.

<u>Lemma 1</u>: Let $\{y_t\}$ be generated from a VAR process (2.1). Then (1.1) holds if and only if

$$(2.5) \qquad \sum_{i=1}^{m_1} e_i'(mp)\Big(\sum_{j=0}^{n_i-1} w_{ij} A^j \Big) = 0 ,$$

where $e_i(n)$ is the $n \times 1$ vector with one in the i-th element and zeros in all others.

<u>Proof</u>: Let c' be the left-hand side vector of (2.5). Then (1.1) is equivalent to the condition $c'Y_t = 0$. Using (2.2), this condition is rewritten as

$$(2.5) \qquad \sum_{i=1}^{p} c'A^{i-1} J_1(p)\, u_{t-i+1} + c'A^p Y_{t-p} = 0 ,$$

where $J_i(p) = e_i(p) \otimes I_m = (0,\ldots,0,I_m,0,\ldots,0)'$ is an mp×m matrix with identity matrix in the i-th block and zero matrices in others. From (2.2) and (2.6), $c'AJ_1(p) = c'\{J_1(p)A_1+J_2(p)\} = 0$. Since $c'J_1(p) = 0$ from (2.6), we have $c'J_2(p) = 0$. Similarly we obtain $c'J_i(p) = 0$ $(i=1,\ldots,p)$ and thus $c' = 0'$. The other direction is obvious. **Q.E.D.**

This lemma is a slight generalization of previous studies which considered some special cases of the hypotheses (1.1).

We now consider the effects of linear filters on the testing procedure of RE hypotheses by VAR models. Since most observed data of economic time series exhibit considerable nonstationarities including trends and seasonality, many econometricians have applied the difference filter and the seasonal adjustment procedure to remove the observed non-stationarities. These transformations of data are generally expressed by the linear filter

$$(2.7) \qquad \Delta = c_0F^r - c_1F^{r-1} - \ldots - c_r - c_{r+1}L - \ldots - c_{r+s}L^s,$$

where c_j $(j=0,1,\ldots,r+s)$ are fixed scalars and F and L are forward and backward shift operators such that $y_{t+k} = F^k y_t$ and $y_{t-k} = L^k y_t$. The linear filter (2.7) includes the d-th difference operator $\Delta = (1-L)^d$ and the moving average operators on which most seasonal adjustment procedures are based. The filtered data is denoted by $y_t^* = \Delta y_t$ in what follows.

Sargent's strategy (1979) of testing RE hypotheses (1.1) can be summarized as follows. The assumption of RE hypotheses gives a set of nonlinear cross-equation restrictions on the underlying stochastic process $\{y_t\}$. Then by the law of iterated projection the same restrictions should be imposed on the stochastic process $\{y_t^*\}$. If the restrictions on the VAR model of $\{y_t^*\}$ cannot be rejected by any statistical standard (say, 1 %

significance level), it has been usually interpreted that RE hypotheses on $\{y_t\}$ cannot be rejected. We argue that this procedure sometimes leads to a false conclusion and thus is not valid for testing the RE hypotheses (1.1).

Suppose that a filtered series $\{y_t^*\}$ is generated by the $VAR_m(p^*)$ process[2/]

$$(2.8) \qquad y_t^* = A_1^* y_{t-1}^* + \dots + A_{p*}^* y_{t-p*}^* + u_t^* ,$$

or equivalently, $\{\Delta y_t\}$ is generated by

$$(2.9) \qquad \Delta y_t = A_1^* \Delta y_{t-1} + \dots + A_{p*}^* \Delta y_{t-p*} + u_{t+r} ,$$

where Δ is defined by (2.7). Let $\Delta' = L^r \Delta$, and noting that $\Delta' y_t = \{I_m - (I_m - \Delta')\} y_t$, we can write (2.8) in terms of y_t as follows:

$$(2.10) \qquad y_t = (I_m - \Delta') y_t + A_1^* \Delta' y_{t-1} + \dots + A_{p*}^* \Delta' y_{t-p*} + u_t$$

$$= \sum_{i=1}^{r+s} c_i y_{t-i} + A_1^* (y_{t-1} - \sum_{i=1}^{r+s} c_i y_{t-1-i})$$

$$+ \dots + A_{p*}^* (y_{t-p*} - \sum_{i=1}^{r+s} c_i y_{t-p*-i}) + u_t .$$

Collecting terms, we rewrite (2.10) as

$$(2.11) \qquad y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t ,$$

where $p = p^* + r + s$ and

$$A_1 = c_1 I_m + A_1^* ,$$
$$A_2 = c_2 I_m - c_1 A_1^* + A_2^* ,$$
$$A_3 = c_3 I_m - c_2 A_1^* - c_1 A_2^* + A_3^*,$$
$$\vdots$$
$$A_{p-1} = -c_{r+s} A_{p*-1}^* - c_{r+s-1} A_{p*}^* ,$$
$$A_p = -c_{r+s} A_{p*}^* .$$

Let λ_j ($j = 1,\ldots,\ell$; $\ell \leqq r+s$) be non-zero solutions of

$$(2.12) \qquad c_0 \lambda^{r+s} - c_1 \lambda^{r+s-1} - \ldots - c_{r+s} = 0 .$$

Now we make an additional assumption:

(A1) There exists at least one solution of (2.12) such that for some i

$$\sum_{j=0}^{n_i - 1} w_{ij} \lambda^j \neq 0 .$$

Then we have the following result:

<u>Theorem 1</u>: Let a filtered series $\{y_t^*\}$ be generated by the $VAR_m(p^*)$ process (2.8), or equivalently $\{y_t\}$ from (2.11). Then under (A1) the restriction (1.1) does not hold.

<u>Proof</u>: We prove this proposition by showing that the necessary and sufficient condition (2.5) cannot be satisfied. Let D be given by

$$(2.13) \qquad D = d \otimes I_m ,$$

where $d = (d_1,d_2,\ldots,d_p)'$, and $d_i = \lambda^{p-i} (i=1,\ldots,p)$ and we take some λ_j in (2.12) as λ here. Then by the structure of A_i in (2.11), we have

$$(2.14)\quad \sum_{i=1}^{p} d_i A_i = \sum_{i=1}^{p} \lambda^{p-i} A_i$$

$$= \lambda^{p-1}(c_1 I_m + A_1^*) + \lambda^{p-2}(c_2 I_m - c_1 A_1^* + A_2^*) + \dots$$

$$+ \lambda(-c_{r+s} A^*_{p*-1} - c_{r+s-1} A^*_{p*}) - c_{r+s} A^*_{p*}$$

$$= I_m(\sum_{i=1}^{r+s} \lambda^{p-i} c_i) + A_1^*(\lambda^{p-1} - \sum_{i=1}^{r+s} \lambda^{p-1-i} c_i) + \dots + A^*_{p*-1} (\lambda^{p-(p^*-1)}$$

$$- \sum_{i=1}^{r+s} \lambda^{p-(p^*-1)-i} c_i) + A^*_{p*}(\lambda^{p-p^*} - \sum_{i=1}^{r+s} \lambda^{p-p^*-i} c_i).$$

In view of (2.12), we have

$$(2.15)\quad \sum_{i=1}^{r+s} \lambda^{p-i} c_i = \lambda^p ,$$

$$\lambda^{p-k} - \sum_{i=1}^{r+s} \lambda^{p-k-i} c_i = 0 \quad (k=1,2,\dots,p) .$$

Thus we obtain the relation:

$$(2.16)\quad \sum_{i=1}^{p} d_i A_i = \lambda^p I_m .$$

Then by the structure of A and A_i defined by (2.2) and (2.11), we have

$$(2.17)\quad A\, D = \lambda\, D .$$

Thus $\lambda_j (j=1,\dots,r+s)$ and row vectors are the characteristic roots and vectors of matrix A, respectively. Accordingly, for any integer k,

$$(2.18)\quad A^k\, D = \lambda^k\, D .$$

Multiplying D from the right to the left hand side of (2.5) and using (2.18), we obtain

$$(2.19)\qquad \sum_{i=1}^{m_1} e_i'(mp)\, D\, \Big(\sum_{j=0}^{n_i-1} w_{ij} \lambda^j \Big) = \sum_{i=1}^{m_1} e_i'(m) \Big(\sum_{j=0}^{n_i-1} w_{ij} \lambda^{j+p-1} \Big) .$$

By assumption (A1), (2.19) cannot be equal to zero. This contradicts the equation given by (2.5). **Q.E.D.**

The above proposition states that, under (A1), when VAR models are fitted to filtered series, the original (or non-filtered) stochastic process does not satisfy the cross-equation restrictions imposed by the RE hypotheses in (2.5). Therefore, the statistical tests based upon the VAR models fitted to the filtered time series are meaningless, if we are interested in the RE hypotheses for the original stochastic process. Actually, as soon as fitting a VAR model to the filtered series is judged appropriate, we must automatically reject RE hypotheses for the original series under (A1).

It should be stressed that the assumption (A1) is not restrictive in most applications. Violation of (A1) seems to require very rare specifications of the RE hypotheses and linear filters. Actually, it is easily seen that (A1) is always satisfied for the RE hypotheses (1.2) and (1.3). Thus, the immediate but important consequence of the above result is given as follows:

<u>Corollary 1</u>: Let $\{\Delta y_t\}$ be generated by a VAR process (2.9) and

$$(2.20)\qquad \Delta = (1 - L)^d$$

for any integer d, then the RE hypotheses (1.2) and (1.3) do not hold.

Shiller (1981) has suggested a special case of this result for $d = 1$, $m = m_1 = 2$, and the RE hypothesis (1.2). The above corollary is particularly interesting because in many empirical studies VAR models have been fitted to the differenced time series and statistical tests based upon them have been conducted. The difference filter of the type (2.20) is widely used since many observed economic time series exhibit nonstationarities. It is sometimes asserted that they are well characterized by the existence of unit roots in the autoregressive parts of the time series models. In other words, random walk processes are appropriate for describing macroeconomic time series. (See, for example, Meese and Singleton (1982), and Nelson and Plosser (1982).) However, our result suggests that the usual practice of applying the difference filter to each series in VAR models is not coherent with the RE hypotheses for the original stochastic process. Thus, the studies by Sargent (1979), Hakkio (1981a,1981b), and Baillie et.at. (1983) are subject to this incoherency. Further, a part of the results by Shiller (1979) where his discussion is based upon the differenced model may be questionable.

3. Roles of Information Set in VAR Models

In this section we re-examine the misspecification problem which arises when VAR models are fitted to the smaller rather than the full information set. As a solution to this problem, we present a proposition later, which generalizes Lemma 1 to the vector autoregressive moving-average (VARMA) models.

First we consider the following example. Suppose that $y_t = (y_{1t}, y_{2t}, y_{3t})' = (y_t^{*\prime}, y_{3t})'$ is generated by the $VAR_3(1)$ with

$$A_1 = \begin{pmatrix} A^* & \begin{matrix} -c \\ c \end{matrix} \\ 0' & c \end{pmatrix} \ (3\times3), \quad A^* = \begin{pmatrix} c & 2c \\ -c^2 & c \end{pmatrix} \ (2\times2),$$

where $c = .5$, y_{1t} is the forward exchange rate, y_{2t} is the spot exchange rate, and y_{3t} is the interest differential. By construction, $e_1'(3)A_1^2 = e_2'(3)$ and the original stochastic process $\{y_t\}$ satisfies (1.3) with h=2. Suppose that we ignore the third variable y_{3t}. Then we can write

$$(3.1) \qquad y_t^* = A^* y_{t-1}^* + u_t^* ,$$

where

$$u_t^* = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} + \begin{pmatrix} -c \\ c \end{pmatrix} u_{3t}/(1-cL) .$$

Thus multiplying (1-cL) to (3.1), we obtain the representation

$$(3.2) \qquad y_t^* = (cI_2 + A^*)y_{t-1}^* - cA^* y_{t-2}^* + v_t ,$$

where

$$v_t = (1-cL)\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} + \begin{pmatrix} -c \\ c \end{pmatrix} u_{3t} .$$

If we treat v_t as if it were a white noise process despite a $VMA_2(1)$ process and use the condition (2.5), it is easy to see that $e_1'(4)\tilde{A}^2 \neq e_2'(4)$, where

$$\tilde{A} = \begin{pmatrix} cI_2 + A^* & -cA^* \\ I_2 & 0 \end{pmatrix} .$$

Hence in this case the cross-equation restrictions of the type (2.5) on the smaller information set $I_t^* = \{y_t^*, y_{t-1}^*, \ldots\}$ are not necessary conditions of the RE hypothesis for the original stochastic process $\{y_t\}$.

We now generalize the above example. Let $\{y_t\}$ in (2.1) be the true process and suppose that y_t is decomposed as $y_t = (y_t^{*\prime}, y_t^{**\prime})'$ where $y_t^* = (y_{1t}, \ldots, y_{m^*t})'$, $y_t^{**} = (y_{m^*+1}, \ldots, y_{mt})$, $m^* \geq 2$, $m^{**} \geq 1$, and $m = m^* + m^{**}$. The law of iterated projection implies that, if (1.1) holds for $I_t = \{y_t, y_{t-1}, \ldots\}$, then we also have

$$(3.3) \qquad \sum_{i=1}^{m_1} \sum_{j=0}^{n_i-1} w_{ij} \, E(y_{i\,t+j} \mid I_t^*) = 0$$

where $I_t = \{y_t^*, y_{t-1}^*, \ldots\}$. Needless to say, (3.3) is a necessary condition of (1.1). Utilizing (3.3), some previous studies such as Sargent (1979) and Hakkio (1981a, 1981b) have fitted VAR models to the smaller information set I^*, and conducted statistical tests based upon these fitted models by following the procedure described in section 2. It has been claimed that those tests were justified as tests of a necessary condition of (1.1). However, we argue that the justification often made is not warranted, since fitting VAR models to the smaller information set involves a high possibility of model misspecification as shown below.

Using the above decomposition of $\{y_t\}$, (2.1) is expressed as

$$(3.4) \qquad A(L)\, y_t = u_t ,$$

where

$$A(L) = \begin{pmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{pmatrix} = I_m - A_1 L - \ldots - A_p L^p ,$$

and $u_t = (u_t^{*\prime}, u_t^{**\prime})$ with

$$\Omega = E(u_t u_t') = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

To make our discussion simpler in this section, we further assume the following:

(A2) The vector y_t^{**} does not cause the vector y_t^* in the sense of Granger (1969), that is $A_{12}(L) \neq 0$.

Also let H(L) be an m×m matrix with lag polynomials which is decomposed as A(L):

$$H(L) = \begin{pmatrix} H_{11}(L) & H_{12}(L) \\ 0 & I_{m^{**}} \end{pmatrix} .$$

Multiply this matrix to (3.4) from the left, and the resulting first m^* equations are given by

$$C_{11}(L)\, y_t^* + C_{12}(L)\, y_t^{**} = H_{11}(L)\, u_t^* + H_{12}(L)\, u_t^{**} , \tag{3.5}$$

where $C_{11}(L) = H_{11}(L)A_{11}(L) + H_{12}(L)A_{21}(L)$, and $C_{12}(L) = H_{11}(L)A_{12}(L) + H_{12}(L)A_{22}(L)$. In order to erase y_t^{**} from (3.5), we have to choose $H_{11}(L)$ and $H_{12}(L)$ such that

$$C_{12}(L) = H_{11}(L)\, A_{12}(L) + H_{12}(L)\, A_{22}(L) = 0 . \tag{3.6}$$

Let $\tilde{A}_{22}(L)$ be the adjoint matrix of $A_{22}(L)$. Multipying $\tilde{A}_{22}(L)$ from the right to (3.6), we have

$$H_{11}(L)\, A_{12}(L)\, \tilde{A}_{22}(L) + H_{12}(L)\, |\, A_{22}(L)\, | = 0 , \tag{3.7}$$

where $|A_{22}(L)|$ is the determinant of $A_{22}(L)$. Thus if we choose

$$H_{11}(L) = |\, A_{22}(L)\, |\, I_{m^*} \quad \text{and} \quad H_{12}(L) = -\, A_{12}(L)\, \tilde{A}_{22}(L), \tag{3.8}$$

then (3.6) is satisfied. Thus (3.5) can be rewritten as

$$A^*(L)\, y_t^* = F(L)\, u_t^* + G(L)\, u_t^{**} , \tag{3.9}$$

where $A^*(L) = |A_{22}(L)|A_{11}(L) - A_{12}(L)\, \tilde{A}_{22}(L)\, A_{21}(L)$,

$$F(L) = |A_{22}(L)| = I_{m*} + F_1 L + \dots + F_{q_1} L^{q_1} ,$$

$$G(L) = - A_{12}(L) \tilde{A}_{22}(L) = G_1 L + \ldots + G_{q_2} L^{q_2} ,$$

and $q_1 \leq pm^{**}$ and $q_2 \leq p^2(m^{**}-1)$.

By the Wold' Decomposition Theorem, we can find the appropriate moving average process such that

$$(3.10) \qquad x_t = D(L)\, v_t = F(L)\, u_t^{*} + G(L)\, u_t^{**}$$

where v_t is the $m^{*}\times 1$ disturbance vector with $E(v_t) = 0$ and $E(v_t v_s') = 0$ for $t \neq s$, and $D(L) = D_0 + D_1 L + \ldots + D_{q^*} L^{q^*}$ ($D_0 = I_{m^*}$) is appropriately defined $m^{*}\times m^{*}$ lag polynomial matrix with order $q^{*} \leq \max\{q_1, q_2\}$. Then we have the next lemma.

<u>Lemma 2</u>: Suppose the moving average process is generated by (3.10). Assume that (A2) holds. The necessary and sufficient condition for $\{x_t\}$ to be a white noise process (i.e. $q^{*}= 0$) is given by

$$(3.11) \qquad \Gamma_x(k) = 0 \quad \text{for } k = \pm 1, \pm 2, \ldots, \pm q_2 ,$$

and $q_1 \leq q_2$ when $\Omega_{12} \neq 0$, and $q_1 + 1 \leq q_2$ when $\Omega_{12} = 0$, where $\Gamma_x(k) = \Gamma_x'(-k) = E(x_t x_{t-k}')$ are defined by (3.13) in detail.

<u>Proof</u>: It is well known that the necessary and sufficient condition of $q^{*}= 0$ is generally given by

$$(3.12) \qquad \Gamma_x(k) = 0 \quad \text{for } k = \pm 1, \pm 2, \ldots, \pm \max\{q_1, q_2\} .$$

We qualify the above condition in what follows. First in case of $\Omega_{12} \neq 0$, $\Gamma_x(k)$ is given by

$$(3.13) \qquad \Gamma_x(k) = \sum_{j=0}^{q_1-k} F_{j+k}\Omega_{11}F_j' + \sum_{j=1}^{q_2-k} G_{j+k}\Omega_{22}G_j'$$

$$+ \sum_{j=0}^{\min(q_1-k,q_2-k)} G_{j+k}\Omega_{21}F_j' + \sum_{j=1}^{\min(q_1-k,q_2-k)} F_{j+k}\Omega_{12}G_j' \ , \ k = 0,1,2,\ldots.$$

If $q_1 > q_2$, then we have $\Gamma_x(q_1) = F_{q_1}\Omega_{11} \neq 0$ since Ω_{11} is nonsingular. Thus we must have $q_2 \geqq q_1$ so that (3.12) holds. Second, in case of $\Omega_{12} = 0$, $\Gamma_x(k)$ is the same as (3.13) except the third and fourth terms are dropped. If $q_1+1 > q_2$, we have $\Gamma_x(q_1) = F_{q_1}\Omega_{11} \neq 0$. Thus we must have $q_1 + 1 \leqq q_2$.
Q.E.D.

The above lemma indicates that in order for $\{x_t\}$ to be a white noise process, strong restrictions (3.11) must be imposed on A(L) and Ω. Since these conditions are hardly satisfied, it is natural to assume that $\{x_t\}$ is a VMA process with positive order in general. Although there is a possibility of cancelling out the lag polynomials in AR and MA parts, it is natural to assume that (3.9) is a VARMA process with positive orders p^* and q^*

$$(3.14) \qquad A^*(L)\ y_t^* = D(L)\ v_t \quad ,$$

where $p^* \leqq \max\{p^2m^{**},\ p^3(m^{**}-1)\}$, $q^* \leqq \max\{q_1,\ q_2\}$, rather a VAR process. The above inequalities for p^* and q^* are derived after considering cancelling out effects. (See Granger and Morris (1976).) It may be remarked that, if the assumption (A2) is violated, that is $A_{12}(L) = 0$, (3.14) is always reduced to a VAR process $A_{11}(L)y_t^* = u_t^*$.

The argument here implies that fitting lower order VAR models to the smaller information set generally involves a high possibility of model misspecification. Apparently, previous studies by Sargent (1979) and Hakkio (1981a, 1981b) are likely to be subject to this possibility since the orders of their VAR models are 4. Of course, their results are valid if we reinterpret that their models are based upon the full rather than smaller information set.

To avoid this misspecification problem, include all relevant variables in estimating VAR models. However, this may require a very large model, and the problem of multicollinearity or the degree of freedom shortage for estimation may arise. An alternative solution is to fit VARMA rather than VAR models to avoid possible misspecifications. However, it appears that for VARMA models, (2.5) is no longer necessary or sufficient for the RE hypotheses (1.1). We present the next result on this problem.

<u>Theorem 2</u>: Assume that $\{y_t\}$ is generated by the vector autoregressive moving average (VARMA) process with orders p and q

$$y_t = A_1 y_t + \dots + A_p y_{t-p} + u_t + B_1 u_{t-1} + \dots + B_q u_{t-q} , \tag{3.15}$$

where $B_1, \dots, B_q$ are m×m coefficient matrices and $\{u_t\}$ are defined as in (2.1). If the absolute values of some roots of the associated equation $|z^p I_m - \sum_{i=1}^{p} A_i z^{p-i}| = 0$ are not less than one, we assume that $y_{-1}, y_{-2}, \dots, y_{-(p+1)}$ are fixed and $u_{-1} = u_{-2} = \dots = u_{-(q+1)} = 0$. Then the necessary and sufficient condition of (1.1) is given by

(3.16) $$a'(A^{+})^{k}\left[J_1(p+q) + J_{p+1}(p+q) \right] = 0 \quad (k = 0, 1, \ldots, p^{*}-1)$$

where

(3.17) $$A^{+} = \begin{bmatrix} A & J_1(p)B_1 \; \cdots & J_1(p)B_q \\ 0 & \begin{matrix} 0 \; 0 \; \cdots \\ I_{m(q-1)} \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} \end{bmatrix} \quad (m(p+q)\times m(p+q)) ,$$

(3.18) $$a' = \sum_{i=1}^{m_1} e_i'(m(p+q)) \sum_{j=0}^{n_i-1} w_{ij}(A^{+})^{j} ,$$

and p^{*} ($\leq$ mp) is the order of minimal polynomial of matrix A and A is defined by (2.2)

<u>Proof</u>: The process (3.15) can be expressed as

(3.19) $$Y_t^{+} = A^{+} Y_{t-1}^{+} + U_t^{+} ,$$

where $Y_t^{+} = (Y_t', U_t')'$ and $U_t^{+} = [J_1(p+q)+J_{p+1}(p+q)]u_t$ are $m(p+q)\times 1$ stacked vectors. Using the same argument as in (2.3), we obtain the optimal predictor of $Y_{t+h}^{+}(h \geq 1)$ given the information set $I_t = \{y_t, y_{t-1},\ldots\}$. The resulting formula is (2.4) where Y_{t+h} and A should be replaced by Y_{t+h}^{+} and A^{+}, respectively. Thus (1.1) is equivalent to the condition

(3.20) $$a'Y_t^{+} = 0 .$$

Using (3.19) and repeating its substitution, we get

(3.21) $$a'\sum_{j=0}^{k-1}(A^{+})^{j}[J_1(p+q)+ J_{p+1}(p+q)]u_{t-j} + a'(A^{+})^{k}Y_{t-k-1}^{+} = 0 .$$

Therefore (3.16) is a necessary condition. Now we consider the characteristic equation of A^{+}:

$$(3.22) \qquad 0 = | A^{+} - \lambda I_{m(p+q)} | = (-\lambda)^{mq} | A - \lambda I_{mp} |$$

$$= (-1)^{p}(-\lambda)^{mq} \prod_{k=1}^{mp} (\lambda - \lambda_k) ,$$

where $\lambda_k (k=1,\ldots,mp)$ are the characteristic roots of matrix A. Then by the Cayley-Hamilton Theorem

$$(3.23) \qquad (A^{+})^{p^*} = b_1 (A^{+})^{p^*-1} + \ldots + b_{p*-1} A^{+} ,$$

where $b_j (j=1,\ldots,p^*-1)$ are some constants depending upon matrix A. In using (3.23), the condition (3.16) for $k=0,1,\ldots,p^*-1$ implies (3.16) for $k=p^*$. **Q.E.D.**

The above result is a generalization of Lemma 1 to the VARMA models. When there exist multiple characteristic roots of the associated equation, the order of mimimal polynomial of the AR part p^* is less than mp. This is always the case if we use the linear filter (2.12) before fitting VARMA models. We note that when q=0, (3.16) is reduced to the condition $a' = 0$. But when $q \geq 1$ we cannot necessarily reduce (3.16) to $a' = 0$. In this case $a' = 0$ (and hence the condition (2.5) on the AR part of (3.15) holds) is merely a sufficient but not necessary condition for (3.1). For example, consider the VARMA(1,1) model. In the present case the conditions given by (3.16) for k=0 and 1 become

$$(3.24) \qquad a'J_1(2)[A_1 + B_1] = a'J_2(2)[A_1 + B_1] = 0' .$$

If we take A_1 and B_1 such that $\mathrm{rank}(A_1 + B_1) < m$, and $\mathrm{rank}(A_1) = \mathrm{rank}(B_1) = m$ for identification (for instance, see Hannan (1969)), it is clear that (3.24) does not mean $a' = 0'$. Thus, the above proposition implies that even if we reject the condition (2.5), we should not reject the RE hypotheses (1.1).

Although the condition (3.16) is far more complicated than (2.5), it is possible for this problem to derive some test statistics and the asymptotic test procedures based on them by following the method developed by Kunitomo and Yamamoto (1985). Alternatively, one may first test several restrictions of (3.16) which are necessary conditions for (1.1) in VARMA models. The condition (3.6) can be simplified further if we have additional information on the parameters of the VARMA models. A fuller investigation on these problems with applications will be reported in another occasion.

One may argue that the alternative test procedures advocated by Geweke and Feige (1979) and Hansen and Hodrick (1980) are superior to those discussed in the present paper, since their methods remain valid for the smaller information set. However, it may be fair to say that their methods have not been justified for nonstationary stochastic processes due to their asymptotic theories.[3/]

4. Conclusion

The present paper exhibited two methodological difficulties for testing the rational expectation (RE) hypotheses based upon fitting vector autoregressive (VAR) time series models. The first one discussed in Section 2 is crucial. We have shown that a widely accepted practice of differencing time series is not coherent with certain RE hypotheses and should not be used. The second one discussed in Section 3 indicated the possibility of model misspecification when VAR models are fitted to the smaller information set. We examined this problem and derived instead a new formula based upon vector autoregressive moving-average (VARMA) models to avoid such misspecification.

Once revealed, these two points may seem trivial to some statisticians. However, since some econometric studies have already been trapped into these troubles, it is worthwhile to examine the problem and state propositions in formal fashion as we have done here.

Finally, our results have some implications not only for the problem of testing RE hypotheses but also for more general econometric modelling with RE hypotheses. In recent macroeconometric studies VAR models are often fitted to filtered time series (for instance, Sims (1980)). Our result (Theorem 1) indicates that this procedure automatically excludes RE hypotheses for the original stochastic processes in most cases. It always does so if the original stochastic processes are random walks. We hope that this finding will correct a misunderstanding commonly held by time series econometricians.

Footnotes

1/ The law of iterated projection states that $E(y|z) = E[E(y|x,z)|z]$ where x, y, and z are random variables. See Shiller (1978) for details and their relevance for the RE hypotheses.

2/ In some cases it is necessary to include a constant term and trend terms when the process is expressed for the original series as in (3.6). It is easy to incorporate these terms into matrix A as in Fuller and Hasza (1981). The results of the paper are essentially unchanged by such treatment of the problem. Thus we ignore those terms for simplicity.

3/ When the absolute value of the characteristic roots of the determinantal equation $| z^p I_m - \sum_{i=1}^{m} A_i z^{p-i} | = 0$ is not smaller than one, the usual asymptotic theory for stationary stochastic processes cannot be used. For instance, the order of convergence is not $\sqrt{T}$ but c(T), where c(T) is some function of the sample size T. See Anderson (1959), and Dickey and Fuller (1979) for instance.

References

Anderson, T.W. (1959), "On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations," The Annals of Mathematical Statistics, 30, September, 676-687.

Baillie, Richard, Robert E. Lippens, and Patrick C. McMahon.(1983)," Testing Rational Expectations and Efficiency in the Foreign Exchange Market," Econometrica, 51, May, 553-563.

Dickey, David A. and Wayne A. Fuller. (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," Journal of the American Statistical Association, 74, 427-431.

Fuller, Wayne A. and David P. Hasza. (1981), "Properties of Predictors for Autoregressive Time Series, Journal of the American Statistical Association, 74, 155-161.

Geweke, John and Edger Feige, (1979). "Some Joint Tests of the Efficiency of Markets for Forward Foreign Exchange," Review of Economics and Statistics, 61, August, 334-341.

Granger, C.W. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," Econometrica, 37, 424-38.

Granger, C.W.J. and M.J. Morris. (1976), "Time Series Modelling and Interpretation," Journal of the Royal Statistical Society, Ser. A., 139, 246-257.

Hakkio, Craig S. (1981a), "The Term Structure of the Forward Premium," Journal of Monetary Economics, 8, July, 41-58.

Hakkio, Craig S. (1981b), "Expectations and the Forward Exchange Rate," International Economic Review, 22, October, 663-678.

Hannan E,J. (1969), "The Identification of Vector Mixed Autoregressive Moving Average Systems," Biometrika, 56, 223-225.

Hansen, Lars Peter and Robert J. Hodrick. (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," Journal of Political Economy, 88, 829-853.

Ito Takatoshi. (1985), "Use of Time-Domain Vector Autoregressions to Test Uncovered Interest Parity," Mimeographed, University of Minnesota.

Kunitomo, N. and T. Yamamoto (1985), "Testing Hypotheses and Causal Ordering in Multivariate Time Series Models with an Analysis of Japanese Macroeconomy, Unpublished Mimeograph, University of Tokyo.

Meese, Richard A. and Kenneth J. Singleton. (1982), "On Unit Roots and the Empirical Modelling of Exchange Rates," Journal of Finance, 37, September, 1029-1035.

Nelson, Charles R. and Charles I. Plosser. (1982), "Trends and Random Walks in Macroeconomic Time Series," Journal of Monetary Economics, 10, September, 139-162.

Sargent, Thomas J. (1979), "A Note on Maximum Likelihood Estimation of the Rational Expectation Model of the Term Structure, Journal of Monetary Economics, January, 133-143.

Shiller, Robert J. (1978), "Rational Expectations and the Dynamic Structure of Macroeconomic Models: A Critical Review, Journal of Monetary Economics, 4, 1-44.

Shiller, Robert J. (1979), "The Volatility of Long-Term Interest Rates and Expectation Models of Term Structure," Journal of Political Economy, 87, January, 1190-1219.

Shiller, Robert J. (1981), "Alternative Tests of Rational Expectation Models: The Case of Term Structure," Journal of Econometrics, Vol.16, 71-87.

Sims, Christopher.(1980), "Macroeconomics and Reality," Econometrica, 48, 1-48.