88-F-6

Conditions on Consistency for Testing Rational Expectation Hypotheses by Vector Autoregressive Models and Cointegration*

by

Naoto Kunitomo
University of Tokyo

and

Taku Yamamoto
Tsukuba University

April 1988

*This paper is a revised version of the present authors' Discussion Paper No. 86-F-2, "On Incoherency of Testing Rational Expectation Hypotheses by Vector Autoregressive Models," April 1986, Research Institute for the Japanese Economy, University of Tokyo. The research was supported by Grant-in-Aid for Scientific Research No. 60301081 and 62301071 of the Ministry of Education, Science and Culture. The earlier version was presented at the annual meeting of Japan Association of Economics and Econometrics held at Nagoya University, November 2, 1986. We thank Michio Hatanaka and Makoto Yano for their comments.

1. Introduction

Econometric analyses with rational expectations (RE) have been expanding in the last decade. The type of cross-equational constraints under the rational expectation (RE) hypothesis of interest in the present paper is expressed as

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

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 the cross-equational constraints under the RE hypothesis, and n_i $(i = 1,\ldots,m_1)$ are positive integers, which can be infinity; a vector $(y_{1t},\ldots,y_{m_1 t})$ is a subset of an $m\times 1$ vector y_t of a stochastic process $\{y_t\}$ we are considering; $I_t = \{y_s; s \leq t\}$ is the information set available at period t and $E(. \ I_t)$ is the conditional expectation operator given I_t.[1/]

Several methods have been proposed to test the cross-equational constraints (1.1) under the RE hypothesis. 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-equational restrictions imposed by the RE hypotheses. Originally, Sargent (1979) proposed this method in connection with a cross-equational restriction under the RE hypothesis in the term structure of interest rates. Later, Hakkio (1981a), (1981b), Baillie et.al. (1983), and Ito (1985) applied some variants of this method in order to test the RE hypotheses in the foreign exchange rate market. In the following, we give only three important examples of the cross-equational constraints under the RE hypothesis, which are special cases of (1.1).

__Example 1__: The cross-equational constraint under the RE hypothesis in Sargent (1979) is that the long-term interest rate is a weighted average of expected short-term interest rates in future. This hypothesis can be expressed as (1.1) when

$m_1 = 2$ and

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

where y_{1t} is the long-term interest rate and y_{2t} is the short-term interest rate.

__Example 2__: In Hakkio (1981a), (1981b), Baillie et. al. (1983), and Ito (1985), the cross-equational constraint under the RE hypothesis of interest is that the forward exchange rate is equal to the expected spot exchange rate in future. This hypothesis can be expressed as (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.

__Example 3__: The present-value relation under the RE hypothesis written in the form:

$$(1.4) \qquad y_{1t} = \sum_{i=2}^{m_1} \sum_{j=0}^{\infty} \beta(i)^j E(y_{i\,t+j} | I_t)$$

is often of our interest, where $\beta(i)$ is the coefficient of the i-th explanatory variable y_{it}. A typical example of this type is the case when

y_{1t} is the stock price and y_{it} are some measures of earnings. It is clear that this relation is a special case of (1.1).

The purpose of this paper is to point out two serious problems which are inherent in testing these cross-equational constraints under the RE hypothesis by fitting VAR models and which econometricians have been often unaware of. Our results imply that many previous studies using this approach are logically misleading. More specifically, first, in Section 2 we show that if we fit VAR models to pre-filtered time series, the cross-equational constraints of the type (1.1) under the RE hypothesis are not necessarily realized with respect to the original stochastic processes in many cases. The common difference filter widely used in practice is an example of our general formulation. Incidentally, Shiller's similar assertion on the cross-equational constraints under RE hypothesis for the term structure of interest rates (1981) is a special case of Corollary 2.3 in Section 2. We also relate our results to the cointegrated processes , which has been recently proposed by Engle and Granger (1987) to describe the non-stationary stochastic processes with unit roots in thier autoregressive representation. In order to avoid the logical inconsistency problem, new sufficient conditions on the cointegrated filter will be given. Second, in Section 3 we discuss a problem which originates from an improper treatment of information sets when we construct statistical tests based on the VAR models. It will be seen that the conventional test procedure based upon a limited information set, which is seemingly justified by law of iterated projection, is likely to lead to a model misspecification. In such a case, we show that VARMA models, rather than VAR models, are more appropriate for testing the cross-equational constraints (1.1) under the RE hypothesis. We shall point out that a number of previous studies suffer from the above two inconsistency problems with respect to their cross-equational constraints under the RE hypothesis in their theoretical frameworks. Section 4 summarizes our results in this paper. We further discuss the implication of our results to more general VAR modelling in recent macroeconometric applications by using pre-filtered time series data. It appears that, in many cases, the method frequently used in macroeconometric studies is not consistent with the cross-equational restrictions under the RE hypothesis.

2. Testing Cross-Equational Constraints by VAR Models

In this section we first present a general method of testing the cross-equational constraints under RE hypothesis 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×m coefficient matrices.[2/] 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 order p is either known or unknown, and can be infinite. However, when p is infinite, some care should be taken to use the following discussion in this section. The derivations of Lemma 1 can always be stated in terms of finite size matrices by choosing components from vectors and matrices with infinite dimension. (See Appendix for this further discussion of this case.) When p is finite, the process (2.1) can be expressed in a Markovian form:

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

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

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

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.

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

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

where $e_i(n)$ is the $n\times1$ vector with one in the i-th element and zeros in all others.

Proof: 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

$$(2.6) \qquad \sum_{i=1}^{p} c'A^{i-1}J_1(p)\, u_{t-i+1} + c'A^p\, Y_{t-p} = 0 ,$$

where $J_i(p) = e_i(p)\otimes I_m = (0,\ldots,0,I_m,0,\ldots,0)'$ is an $mp\times m$ 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 hypothesis (1.1). (Sargent (1979), Baillie et. al. (1983), for instance.)

We now consider the effects of linear filters on the testing procedure of the cross-equational constraints under the RE hypothesis 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

$$\Delta = C_0F^r - C_1F^{r-1} - \ldots - C_r - C_{r+1}L - \ldots - C_{r+s}L^s, \tag{2.7}$$

where $C_k = (c_{ij}(k))$ $(i,j = 1,\ldots,m\ ;\ k = 0,1,\ldots,r+s)$ are m×m matrices with $|C_0| \neq 0$, and F and L are forward and backward shift operators such that $y_{t+k} = F^k y_t$ and $y_{t-k} = L^k y_t$. When $\{C_k\}$ matrices are diagonal, Δ is called diagonal filter. Also when $c_{ii}(k) = c(k)$ for all k, Δ is called common filter. The linear filter (2.7) includes the d-th common difference operator $\Delta_1 = (1-L)^d I_m$, which is a diagonal filter, 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.

Suppose that a filtered series $\{y_t^*\} = \{(y_{it}^*)\}$ is generated by the $VAR_m(p^*)$ process

$$y_t^* = A_1^* y_{t-1}^* + \ldots + A_{p*}^* y_{t-p*}^* + u_t^*, \tag{2.8}$$

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

$$\Delta y_t = A_1^* \Delta y_{t-1} + \ldots + A_{p*}^* \Delta y_{t-p*} + u_t^*, \tag{2.9}$$

where Δ is defined by (2.7).[3/]

Testing the cross-equational constraints under the RE hypothesis, which was originally proposed by Sargent (1979) and has been adopted by many subsequent empirical studies, can be summarized as follows. The assumption of RE hypothesis gives a set of nonlinear cross-equation restrictions on the coefficient matrices $\{A_i,\ i\geq 1\}$ of the underlying stochastic process $\{y_t\}$ given by (2.5). We transform the original process $\{y_t\}$ into a filtered process $\{y_t^*\}$ by a <u>common filter</u> such as Δ_1 and use the law of iterated projection. Then the same restrictions $\{A_i^*,\ i\geq 1\}$ 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 the cross-equational constraints imposed by the RE hypothesis on $\{y_t\}$ cannot be rejected. We first argue that this procedure sometimes leads to a false conclusion and thus is not necessarily valid for testing the cross-equational constraints under the RE hypothesis given by (1.1). We then shall give a sufficient condition for justifying the above procedure.

Let $\Delta' = L^r\Delta$, and noting that $\Delta' y_t = \{C_0 - (C_0 - \Delta')\}y_t$, we can write (2.8) in terms of y_t as follows:

$$(2.10)\qquad C_0 y_t = (C_0 - \Delta')y_t + A_1^*\Delta' y_{t-1} + \cdots + A_{p*}^*\Delta' y_{t-p*} + u_{t-r}^*$$

$$= \sum_{i=1}^{r+s} C_i y_{t-i} + A_1^*(C_0 y_{t-1} - \sum_{i=1}^{r+s} C_i y_{t-1-i})$$

$$+ \cdots + A_{p*}^*(C_0 y_{t-p*} - \sum_{i=1}^{r+s} C_i y_{t-p*-i}) + u_{t-r}^*.$$

Collecting terms, we rewrite (2.10) as

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

where $p = p^* + r + s$, $C_0 u_t = u_{t-r}^*$, and

$$C_0A_1 = C_1 + A_1^*C_0 ,$$
$$C_0A_2 = C_2 - A_1^*C_1 + A_2^*C_0 ,$$
$$C_0A_3 = C_3 - A_1^*C_2 - A_2^*C_1 + A_3^*C_0 ,$$
$$\vdots$$
$$C_0A_{p-1} = -A_{p*-1}^*C_{r+s} - A_{p*}^*C_{r+s-1},$$
$$C_0A_p = -A_{p*}^*C_{r+s} .$$

Let also $\{\lambda_i\}$ and $\{d_1^{(i)}\}$ be the roots and m×1 vectors satisfying

$$(2.12) \qquad (C_0 \lambda^{r+s} - C_1\lambda^{r+s-1} - \dots - C_{r+s})d_1 = 0 .$$

We now present the fundamental result for our arguments on the inconsistency of testing the cross-equational restrictions imposed by the RE hypothesis in this section.

<u>Theorem 2.1</u>: Let a filtered process $\{y_t^*\}$ be generated by a VAR(p^*) process with a finite p^* given by (2.8). Then the restrictions on the original process $\{y_t\}$ given by (1.1) implies $g(\lambda) = 0$ for any root $\{\lambda_k\}$ of (2.12), where

$$(2.13) \qquad g(\lambda) = \sum_{i=1}^{m_1} \{ \sum_{j=0}^{n_i-1} w_{ij} \lambda^j\} e_i(m)' d_1 .$$

<u>Proof of Theorem 2.1</u>: If a root λ of (2.12) is zero, it is trivial to have $g(\lambda) = 0$. Let λ_k be a non-zero root of (2.12) and we construct a pm×1 vector $d^{(k)} = (d_1^{'(k)}, d_2^{'(k)}, \dots, d_p^{'(k)})'$ by defining

$$(2.14) \qquad d_i^{(k)} = \lambda_k^{1-i} d_1^{(k)} \quad (i = 2,\dots,p).$$

Then by the structure of A_i in (2.11), we have

$$(2.15)\qquad C_0\sum_{i=1}^{p} A_i d_i^{(k)} = \sum_{i=1}^{p} \lambda_k^{1-i} C_0 A_i d_1^{(k)}$$

$$= \Big(\sum_{i=1}^{r+s} \lambda_k^{1-i} C_i\Big) d_1^{(k)} + A_1^*\Big(C_0 - \sum_{i=1}^{r+s} \lambda_k^{-i} C_i\Big) d_1^{(k)} + \cdots$$

$$+ A_{p*-1}^*\Big(\lambda_k^{-(p^*-2)} C_0 - \sum_{i=1}^{r+s} \lambda_k^{-(p^*-2)-i} C_i\Big) d_1^{(k)}$$

$$+ A_{p*}^*\Big(\lambda_k^{-(p^*-1)} C_0 - \sum_{i=1}^{r+s} \lambda_k^{-(p^*-1)-i} C_i\Big) d_1^{(k)}.$$

$$= \lambda_k C_0 d_1^{(k)} - \lambda_k^{-(r+s+p^*-1)}\Big(\lambda_k^{p^*} I_m - A_1^*\lambda_k^{p^*-1} - \cdots - A_{p*}\Big)$$

$$\times\Big(\lambda_k^{r+s} C_0 - \sum_{i=1}^{r+s} \lambda_k^{r+s-i} C_i\Big) d_1^{(k)}.$$

In view of (2.12), we have

$$(2.16)\qquad \Big(\lambda_k^{p-\ell} C_0 - \sum_{i=1}^{r+s} \lambda_k^{p-\ell-i} C_i\Big) d_1^{(k)} = 0 \quad (\ell = 0,1,2,\ldots,p)\ .$$

Thus we obtain the relation:

$$(2.17)\qquad \sum_{i=1}^{p} A_i d_i^{(k)} = \lambda_k d_1^{(k)}\ .$$

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

$$A\, d^{(k)} = \lambda_k d^{(k)}.$$

Thus λ_k and $d^{(k)}$ are the characteristic root and vector of matrix A, respectively. Accordingly, for any integer ℓ, (2.18) $A^{\ell} d^{(k)} = \lambda_k^{\ell} d^{(k)}$.

Multiplying $d^{(k)}$ from the right to the left-hand side of (2.5) and using (2.18), we obtain $g(\lambda_k) = 0$ as a necessary condition for (2.5). **Q.E.D.**

This proposition gives a necessary condition on the filter matrices $\{C_i, i \geq 0\}$ for the cross-equational restrictions imposed by the RE hypothesis given by (2.5). An important inconsistency result on testing the hypothesis (1.1) by the VAR modelling, which is a direct consequence of Theorem 2.1, is stated in the following corollary.

<u>Corollary 2.1</u>: Let a filtered series $\{y_t^*\}$ be generated by the $VAR_m(p^*)$ process (2.8), or equivalently $\{y_t\}$ from (2.11) with a finite p. If there exists at least a non-zero scalor λ and a non-zero m×1 vector d_1 satisfying $g(\lambda) \neq 0$, the cross-equation restrictions on $\{y_t\}$ imposed by (1.1) do not hold.

Let $w_j = (w_{1j},\ldots,w_{m_1 j},0,\ldots,0)'$ be m×1 vectors for $j \geq 0$ and n = $\max\{n_i-1\}$ for $i = 1,\ldots,m_1$. Then the cross-equational constraint given by (1.1) is written as

(2.19) $E[\ w(F)'y_t \mid I_t\ ] = 0$,

where

$$w(F) = \sum_{j=0}^{n} w_j F^j .$$

Because $y_t^* = \Delta y_t$, (2.19) is rewritten as

(2.20) $E[\ w^*(F)'y_t^* \mid I_t\ ] = 0$,

where $w^*(F)$ is assumed to be properly defined by $w(F)' = w^*(F)'\Delta$, $w_j^* = (w_{1j}^*,\ldots,w_{mj}^*)'$, and for a finite integer ℓ

$$w^*(F) = \sum_{j=\ell}^{n-r} w_j^* F^j .$$

Since $I_{t-r}^* \subset I_t$ for $I_{t-r}^* = \{y_s^* ; s \leq t-r\}$ and $I_t = \{y_s ; s \leq t\}$, we obtain the condition

$$\text{(2.21)} \qquad \sum_{i=1}^{m} \sum_{j=\ell}^{n-r} w_{ij}^* E(y_{it+j}^* \mid I_{t-r}^*) = 0 \quad ,$$

which is the cross-equation restriction on the filtered process $\{y_t^*\}$ imposed by the RE hypothesis. Then the condition we obtained in Theorem 2.1 is also a sufficient condition under additional assumptions, which are stated formally in the following proposition.

<u>Theorem 2.2</u>: Suppose that (i) we have (2.21) with $\ell \geq -r$, (ii) the coefficient matrices $\{A_i^*\}$ satisfy

$$\text{(2.22)} \qquad \sum_{i=1}^{m} e_i'(mp^*)\left(\sum_{j=\ell}^{n-r} w_{ij}^* A^{*j+r} \right) = 0',$$

where

$$A^* = \begin{pmatrix} A_1^* \; A_2^* \; \cdots & A_{p^*-1}^* & A_{p^*}^* \\ I_{m(p^*-1)} & & 0 \end{pmatrix} (mp^* \times mp^*) ,$$

and (iii) the roots of

$$\text{(2.23)} \qquad |\lambda^{p^*} I_m - A_1^* \lambda^{p^*-1} - \ldots - A_{p*}^*| = 0$$

and the roots of (2.12) are all distinct. Then the condition $g(\lambda_k) = 0$ for all roots $\{\lambda_k\}$ of (2.12) implies the restrictions on the original stochastic process $\{y_t\}$ given by (1.1).

Proof of Theorem 2.2: Let $\lambda^{(k)}$ and $f_1^{(k)}$ be the non-zero roots and vectors of (2.23). Then $\lambda^{(k)}$ and

$$d_1^{(k)} = [\lambda_k^{r+s}C_0 - \sum_{i=1}^{r+s} \lambda_k^{r+s-i}C_i]^{-1}f_1^{(k)}$$

satisfy (2.17) since the first term on the right-hand side of $d_1^{(k)}$ is non-singular under the assumptions in Theorem 2.2. Let $J_p' = e_1'(p)\otimes I_m$. Then

$$(2.24)\quad \sum_{j=0}^{n} w_j' J_p' A^j d^{(k)} = (\lambda_k)^{\ell-s}\{\sum_{j=0}^{n-r-\ell} (\lambda_k)^j w_{\ell+j}^{*\,'} \ [\lambda_k^{r+s}C_0 - \sum_{i=1}^{r+s} \lambda_k^{r+s-i}C_i]d_1^{(k)}\}$$

$$= (\lambda_k)^{-s-r}\{\sum_{j=\ell}^{n-r} (\lambda_k)^{j+r} w_j^{*\,'}\}f_1^{(k)}$$

$$= (\lambda_k)^{-s-r}\{\sum_{j=\ell}^{n-r} w_j^{*\,'} J_{p*}' A^{*j+r}\}f^{(k)},$$

which is zero-vector from (2.22), where $f^{(k)}$ is constructed from $f_1^{(k)}$ as $d^{(k)}$ in Theorem 2.1. If $\lambda_k = 0$, $d_1^{(k)} = 0$ because $Ad^{(k)} = 0$ and $d^{(k)} \neq 0$. In this case, the left-hand side of (2.24) becomes $w_0'd_1^{(k)} = 0$. Then we can construct a $pm\times pm$ matrix D such that the first $m\times p^*$ ($p = p^*+r+s$) columns are $[d^{(1)}, \ldots, d^{(m\times p^*)}]$ and the remaining $m\times(p-p^*)$ columns are the characteristic vectors, which are corresponding to the characteristic roots of (2.12). Since all roots of (2.12) and (2.23) are distinct under the assumptions in Theorem 2.2, the matrix D is non-singular. Thus the conditions of (2.22) and $g(\lambda_k) = 0$ imply that

$$(2.25)\quad \sum_{i=1}^{m_1} e_i(mp)'\{\sum_{j=0}^{n_i-1} w_{ij}\ A^j\}D = 0'.$$

Multiplying D^{-1} to (2.25) from the left, we obtain (2.5). Then we use Lemma 1 to establish the result. Q.E.D.

When the filter matrices $\{C_j, 0 \leq j \leq r+s\}$ are diagonal in (2.7), we have stronger results than Theorems 2.1 and 2.2. Let $\{\lambda_i\}$ be the roots satisfying

$$c_{ii}(o)\lambda^{r+s} - c_{ii}(1)\lambda^{r+s-1} - \ldots - c_{ii}(r+s) = 0 . \tag{2.26}$$

<u>Theorem 2.3</u>: Let a filtered process $\{y_t^*\}$ be generated by a VAR(p^*) process with a finite p^* given by (2.8). Then the restrictions on the original stochastic process $\{y_t\}$ given by (1.1) implies $g_k(\lambda_k) = 0$ for any root $\{\lambda_k\}$ of (2.26), where

$$g_k(\lambda_k) = \sum_{j=0}^{n_i-1} w_{kj} \lambda_k^{j}. \tag{2.27}$$

<u>Theorem 2.4</u>: Suppose that (i) we have (2.21) with $\ell \geq -r$, (ii) the coefficient matrices $\{A_i^*\}$ satisfy (2.22), and (iii) the roots of (2.23) and the roots of (2.26) are all distinct. Then if there are at least m_1 roots of (2.26) satisfying the condition $g_k(\lambda_k) = 0$, then the restrictions on the original stochastic process $\{y_t\}$ given by (1.1) hold.

<u>Proofs of Theorems 2.3 and 2.4</u>: Because the matrices $\{C_i, i \geq 0\}$ are diagonal, we can take $d_1^{(k)} = e_k(m)$ in the proof of Theorems 2.1 and 2.2. Then from (2.13) and (2.27) $g(\lambda_k) = g_k(\lambda_k)$ for $k = 1,\ldots,m_1$. We note that $g_k(\lambda_k) = 0$ for $k = m_1+1,\ldots, m$ since $d_1^{(k)} = e_k(m)$. Thus the conditions in Theorems 2.1 and 2.2 become the conditions in Theorems 2.3 and 2.4. **Q.E.D.**

Theorem 2.3 gives a necessary condition on the diagonal filter matrices $\{C_i, i \geq 0\}$ for the cross-equational restrictions imposed by the RE hypothesis given by (2.5). Thorem 2.4 shows that the condition we obtained is also a sufficient condition under additional assumptions. An important

inconsistency result on testing the hypothesis (1.1) by the VAR modelling, which is a direct consequence of Theorem 2.3, is stated in the following corollary.

<u>Corollary 2.2</u>: Let a filtered series $\{y_t^*\}$ be generated by the $VAR_m(p^*)$ process (2.8) and the filter matrices $\{C_i,\ i \geq 0\}$ are diagonal. If there exists at least a non-zero scalor λ_k satisfying $g_k(\lambda_k) \neq 0$, the cross-equation restrictions on $\{y_t\}$ imposed by (1.1) do not hold.

Corollaries 2.1 and 2.2 state that 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 hypothesis in (2.5) unless $g(\lambda_k) = 0$ or $g_k(\lambda_k) = 0$. Therefore, the statistical tests based upon the VAR models fitted to the filtered time series are meaningless, if we are interested in the cross-equation restrictions imposed by the RE hypothesis for the original stochastic process unless $g(\lambda_k) = 0$ or $g_k(\lambda_k) = 0$. Actually, as soon as fitting a VAR model to the filtered series is judged appropriate, we must automatically reject the cross-equation restrictions imposed by the RE hypothesis for the original series if these conditions are not satisfied.

It should be stressed that the conditions $g(\lambda_k) \neq 0$ or $g_k(\lambda_k) \neq 0$ are not restrictive in many applications when we use linear diagonal filters. Violation of these conditions seems to require very rare specifications of the cross-equation restrictions imposed by the RE hypothesis and linear diagonal filters. Actually, it is easily seen that the condition $g_k(\lambda_k) \neq 0$ is always satisfied for the cross-equation restrictions imposed by the RE hypotheses (1.2) and (1.3). Thus, an immediate but important consequence of the above result is given as follows:

Corollary 2.3: Let $\{\Delta_1 y_t\}$ be generated by a VAR process (2.9) and a common filter

(2.28) $$\Delta_1 = (1 - L)^d I_m$$

for any integer d, then the cross-equation restrictions (1.2) and (1.3) imposed by the RE hypothesis 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 have been conducted based upon them. The common difference filter of the type Δ_1 has been 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 common difference filter Δ_1 to time series in VAR models is often inconsistent with the cross-equation restrictions imposed by the RE hypothesis for the original stochastic process. Thus, the studies by Sargent (1979), Hakkio (1981a,1981b), and Baillie et.at. (1983) are subject to this inconsistenccy. Further, a part of the results by Shiller (1979) where his discussion is based upon the differenced model should be reconsidered.

It may be natural to ask if we obtain the corresponding result for the vector autoregressive moving average (VARMA) models. So far, we have obtained only the necessary part to answer this question, which is an extension of Theorems 2.1 and 2.3. The proof, which is given in Appendix, is similar to those of Theorems 2.1 and 2.3.

<u>Theorem 2.5</u>: Let a filtered series $\{y_t^*\}$ be generated by the $VARMA_m(p^*,q)$ process with a finite q

$$(2.29) \qquad y_t^* = A_1^* y_{t-1}^* + \dots + A_{p*}^* y_{t-p*}^* + u_t + B_1 u_{t-1} + \dots + B_q u_{t-q} ,$$

where $B_1,\dots,B_q$ are m×m coefficient matrices and $\{u_t\}$ are defined as in (2.1). Let ρ be the maximum absolute value of the roots of the associated MA equation $|z^q I_m - \sum_{i=1}^{q} z^{q-i} B_i| = 0$ and λ_k be the roots of (2.12) or (2.26). We assume that $|\lambda_k| > \rho$ and $\rho < 1$. Suppose the restrictions on $\{y_t^*\}$ imposed by (1.1) hold. Then the restrictions on $\{y_t\}$ imposed by (1.1) implies that (i) for any root $\{\lambda_k\}$ of (2.12) $g(\lambda_k) = 0$ in the general case and (ii) for any root $\{\lambda_k\}$ of (2.26) $g(\lambda_k) = 0$ when $\{C_i,\ i \geq 0\}$ are diagonal.

This proposition shows that if $g(\lambda_k) \neq 0$ or $g_k(\lambda_k) \neq 0$ for some k, then the restrictions (1.1) imposed by the RE hypothesis on $\{y_t\}$ can not be true. It may be interesting to note that the conditions in Theorem 2.5 are the same as in the VAR models even when the true stochastic process is a VARMA process.

From Theorems 2.1-2.5 and Corollaries 2.1-2.3 in this section, the integrated vector autoregressive (IAR) processes and the vector autoregressive integrated moving average (ARIMA) processes are often inconsistent with the cross-equation restrictions imposed by RE hypothesis. Recently, Engle and Granger (1987) has proposed the cointegrated processes for dealing with unit roots in the autoregressive part of VARMA stochastic processes. The cointegrated process can be defined as the nonstatinary process having (i) autoregressive (AR) representation, (ii) the number of unit roots in the AR part is less than the dimension of variables $\{y_t\}$, and (iii) the absolute values of other roots in the AR part are less than 1. From Theorems 2.1-2.5, the cross-equation constraints under the RE hypothesis in the form of (1.1) often imply that the stochastic process of $\{y_t\}$ is cointegrated if it is a nonstationary VARMA process with unit roots in the autoregressive part.

Levy and Nobay (1986) has used the cointegrated filter

$$(2.30) \qquad \Delta_2 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} L$$

for testing the cross-equation restriction in Example 2. They rewrite the cross-equation restriction for the filtered process and test the restriction, which is given by (1.1) with

$$(2.31) \qquad w^*_{1j} = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 \end{cases}, \quad w^*_{2j} = \begin{cases} 1 & \text{if } j = 1,\dots,h \\ 0 & \text{otherwise} \end{cases} .$$

This procedure is valid if (i) the original stochastic process is cointegrated, (ii) the linear filter Δ_2 remove its non-stationarity, and (iii) the coefficient matrix A^* for the filtered process does not have roots of 0 and 1. This is because the roots of (2.12) are 1 and 0. If we take $d^{(1)} = (1,\ 1)'$ and $d^{(2)} = (0,1)'$, $g(1) = g(0) = 0$ in (2.13). Thus the necessary condition in Theorem 2.1 for (1.3) is automatically satisfied in this case.

Similarly, Campbell and Shiller (1987) used the cointegrated filter

$$(2.32) \qquad \Delta_3 = \begin{pmatrix} 1 & -\theta \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} L$$

for testing the cross-equation restriction (1.4) imposed by the RE hypothesis in Example 3, where $\beta(i) = \theta(1-\delta)\delta^i$ and θ and δ are parameters in their notation. They rewrite the cross-equation restriction for the filtered process and test the restriction, which is given by (1.1) with

$$(2.33) \qquad w^*_{1j} = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 \end{cases}, \quad w^*_{2j} = -\theta\delta^j \quad j \geq 1 .$$

This procedure is valid if (i) the original stochastic process is cointegrated, (ii) the linear filter Δ_3 remove its non-stationarity, (iii) the coefficient matrix A^* for the filtered process does not have roots of 0 and 1, and (iv) the parameter θ is <u>a priori</u> known. This is because the roots of (2.12) are also 1 and 0. If take $d^{(1)} = [\theta,1]'$ and $d^{(2)} = [\theta(1-\delta),1]'$, $g(1) = g(0) = 0$ in (2.13). Thus again the necessary condition in Theorem 2.1 is automatically satisfied in this case. We summarize the consistency results for cointegrated processes with filters Δ_2 and Δ_3 as the following corollary, whose proof is straightforward.

<u>Corollary 2.4</u>: Let $\{\Delta y_t\}$ be generated by a VAR process (2.9). We assume that the roots of (2.18) are all distinct and they are not 1 and 0. Then the cointegrated filter Δ_2 is consistent with the cross-equation restriction (1.3) imposed by the RE hypothesis and the cointegrated filter Δ_3 is consistent with the cross-equation restriction (1.4) imposed by the RE hypothesis.

It is important to notice that a cointegrated stochastic process is not necessarily consistent with a set of cross-equation restrictions imposed by the RE hypothesis. For instance, the cointegrated filters used by Levy and Nobay (1986) and Campbell and Shiller (1987) are valid if the VAR processes for the filtered time series do not have the characteristic roots of 1 and 0 in their autoregressive parts. Thus it is important to check if a set of sufficient conditions for (1.1) is satisfied with the linear filter as well as the VAR process for the filtered time series. For this purpose, our results in Theorems 2.2 and 2.4 may be useful for empirical studies.

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 in a particular manner.

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

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

where $c = .5$, y_{1t} is the spot exchange rate, y_{2t} is the forward exchange rate, and y_{3t} is the money supply. 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

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

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

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

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

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

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

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

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 a smaller information set I_t^*, and conducted statistical tests of similar cross-equational restrictions based upon these 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 a 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

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

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

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

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

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

To make our discussion meaningful, we further assume the following:

(A) The vector y_t^{**} causes the vector y_t^{*} in the sense of Granger (1969), that is, $A_{12}(L) \neq 0$.

Multiplying the matrix H(L) to (3.4) from the left, and the resulting first m^* equations are given by

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

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

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

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

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

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

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

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

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

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

$$F(L) = |A_{22}(L)| I_{m*} = I_{m*} + \text{diag}\{f_1\}L + \ldots + \text{diag}\{f_{q_1}\}L^{q_1},$$

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

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

By the Wold's 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 $q^* \leq \max\{q_1, q_2\}$.

It is well known that the necessary and sufficient condition for $\{x_t\}$ to be a multivariate white noise process (i.e., $q^* = 0$) is given by

$$(3.11) \qquad \Gamma_x(k) = 0 \quad \text{for } k = \pm 1, \pm 2, \ldots, \pm \max\{q_1, q_2\}.$$

where

$$(3.12) \qquad \Gamma_x(k) = \sum_{j=0}^{q_1-k} F_{j+k}\Omega_{11}F_j' + \sum_{j=1}^{q_2-k} G_{j+k}\Omega_{22}G_j'$$

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

When $\Omega_{12} = 0$, the second and third terms of (3.12) are dropped. The above condition indicates that, in order for $\{x_t\}$ to be a multivariate white noise process, strong restrictions must be imposed upon A(L) and Ω.

Since the above condition (3.11) is too general and complicated to use in practice, we present the next theorem which gives a simple necessary condition with respect to the order of lags in G(L).

<u>Theorem 3.1</u>: Suppose the moving average process is generated by (3.10). Assume that (A) holds. Let $g_i(L)$ be the i-th row vector of G(L), and q_3 be the lowest lag order in G(L), i.e.,

$$(3.13) \qquad G_k = 0 \quad \text{for } k = 0,1,2,\cdots,q_3-1, \quad \text{and } G_{q_3} \neq 0.$$

Then, a necessary condition for $\{x_t\}$ to be a multivariate white noise process is given as follows:

(i) When $\Omega_{12} \neq 0$, $g_i(L)u_t^{**}$ is a moving average process with the q_1-th order of lags for $i=1,2,\cdots,m^*$.

(ii) When $\Omega_{12} = 0$, $g_i(L)u_t^{**}$ is a moving average process with the (q_1+q_3)th order of lags, and its spectral density has the same shape for $i=1,2,\cdots,m^*$.

Proof: Note that the lag polynomial matrix F(L) is diagonal with a common element $|A_{22}(L)|$. Thus, each element of $F(L)u_t^*$ has the q_1-th order of lags and its spectral density has the same shape. In order for $\{x_t\}$ to be a multivariate white noise process, every moving average process must cancel out the autocorrelations derived from the common lag polynomial $|A_{22}(L)|$. Thus, it is easy to see that statement (i) holds.

When $\Omega_{12} = 0$, there is no cross autocorrelations between $F(L)u_t^*$ and $G(L)u_t^{**}$. Then, in order to cancel out the common lag structure, $g_i(L)u_t^{**}$ must have the same spectral shape for all i. Since lags between u_t^* and u_{t-k}^{**} ($k \neq 0$) can be ignored in this case, it is necessary for $G(L)u_t^{**}$ to have the q_1-th autocorrelation. It means that the order of lags of $g_i(L)u_t^{**}$ must be q_1+q_3 for all i, when $G_k = 0$ for $k \leq q_3-1$. This completes the proof of statement (ii). **Q.E.D.**

Since (3.11) or conditions in Theorem 2 are hardly satisfied, it is natural to regard 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) $$A^*(L)\, y_t^* = D(L)\, v_t ,$$

where $p^* \leq \max\{p^2 m^{**},\ p^3(m^{**}-1)\}$, $q^* \leq \max\{q_1, q_2\}$, rather than 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 (A) is violated, that is $A_{12}(L) = 0$, (3.14) is always reduced to a VAR process $A_{11}(L)y_t^* = u_t^*$. That is, when y_t^{**} does not cause y_t^*

in Granger's sense, it is correct to test RE hyhothses by VAR models based upon a smaller information set.

The argument here implies that the correct model for a smaller information set is a VARMA process, and fitting lower order VAR models to a 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. However, their results are valid if we reinterpret that their models are based upon the full rather than a smaller information set.

To avoid this misspecification problem, we should include all relevant variables in estimating VAR models. However, this may require a very large model, and problems of multicollinearity or the degree of freedom shortage in 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 theorem which gives the necessary and sufficient condition of RE hypothses (1.1) for VARMA models.

<u>Theorem 3.2</u>: Assume that $\{y_t\}$ is generated by the vector auto-regressive moving average (VARMA) process with orders p and q

$$(3.15) \qquad y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t + B_1 u_{t-1} + \dots + B_q u_{t-q} .$$

If the absolute values of some roots of the associated AR 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

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

where $J_i(p+q) = e_i(p+q)\otimes I_m$ is an $m(p+q)\times m$ matrix,

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

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

and p^{*} ($\leqq$ 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

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

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 $Y_{t+h}^{+}(h \geqq 1)$ 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

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

Using (3.19) and repeating its substitution, we get

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

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 minimal polynomial of the AR part p^* is less than mp. This is always the case if we use the linear filter (2.13) 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

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

If we take A_1 and B_1 such that $rank(A_1 + B_1) < m$, and $rank(A_1) = rank(B_1) = m$ 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 straight-forward to derive some test statistics and the asymptotic test procedures based on them by following the method developed by Kunitomo and Yamamoto (1986) if the stochastic process is a stationary vector ARMA process. ALso Levy and Nobay (1986) has developed a similar method for a special case in Example 1. Alternatively, one may test several restrictions of (3.16) which are necessary conditions for (1.1) in VARMA models. The condition (3.16) can be simplified further if we have additional information on the parameters of the VARMA models.

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 a smaller information set. However, their methods are applicable only to a very special type of RE hypothsis given in Example 1.1, and their methods could not be justified if their asymptotic theories were not valid for nonstationary stochastic processes.[4/]

4. Conclusion

The present paper pointed out two methodological difficulties in the test of the cross-equation restrictions under the rational expectation (RE) hypothesis, which is based upon fitted vector autoregressive (VAR) time series models. The first one, discussed in Section 2, is crucial. We have shown that a widely accepted practice of pre-filtering or common differencing of time series is often inconsistent with the cross-equational constraints under the RE hypothesis and should not be used. We also pointed out that the cointegration filter can be a solution to this inconsistency, but should be used carefully. In order to avoid the inconsistency problem, we have obtained new sufficient conditions for the logical consistency, which may be useful for empirical studies. The second one, discussed in Section 3, showed that a high possibility of model misspecification exists when VAR models are fitted to a smaller information set. We then derived the necessary and sufficient condition of cross-equational constraints imposed by the RE hypothsis for vector autoregressive moving-average (VARMA) models in order to avoid such a misspecification. Since many econometric studies have been trapped into these troubles, it is worthwhile to examine the problems and state propositions in formal fashion as we have done here.

Finally, our results have some implications with respect not only to the testing of cross-equational constraints imposed by RE hypothesis but also to more general econometric modelling that adopts the RE hypothesis. In recent macroeconometric studies VAR models are often fitted to pre-filtered time series. For instance, it seems that Sims (1980) used seasonally adjusted data and estimated VAR models by the pre-filtered time series data in most cases. Our results (Theorems 2.1-2.5) indicate that this procedure automatically excludes the cross-equational constraints imposed by the RE hypothesis for the original stochastic processes in many cases. It always does so if the original stochastic process is an integrated VAR or an integrated VARMA process for the restrictions such as (1.2) and (1.3) as is seen in Corollary 2.3. We hope that our finding will alarm time series econometricians of the fact.

5. Appendix: A Sketch of Proof of Theorem 2.5

To prove Theorem 2.5, we use a similar argument as in the proofs of Theorems 2.1 and 2.3. When $\rho<1$ for the general VARMA process (2.29), we obtain the AR representation

$$y_t = \sum_{j\geq 1} A_j(1)\ y_{t-j} + u_t. \tag{A.1}$$

By the successive substitution, we have

$$y_t = \sum_{j\geq 1} A_j(k) y_{t-(k-1)-j} + u_t + \sum_{j\geq 1}^{k-1} A_1(j) u_{t-j} \tag{A.2}$$

where $A_j(k)$ are defined by the recursive formula

$$A_j(k) = A_1(k-1)A_j(1) + A_{j+1}(k-1) \quad (k = 2,3,\ldots), \tag{A.3}$$

and $A_j(0) = \delta(j,1)I_m$, where $\delta(j,1)$ is the delta function. Then Lemma 1 should be modified as the conditions

$$c_k' = \sum_{i=1}^{m_1} \sum_{j=0}^{n_i-1} w_{ij} e_i'(m) A_k(j) = 0 \quad (k=1,2,\ldots). \tag{A.4}$$

Let λ_k and d_1 be a solution of (2.12) or (2.26). For (2.26), we take $d_1 = e_k(m)$. Then we construct $\{d_j;\ 1 \leq j\}$ as in the proofs of Theorems 2.1 and 2.3. By the use of (A.3), we have

$$\sum_{j \geq 1} A_j(\ell) d_j = \lambda_k^{\ell} d_1 , \tag{A.5}$$

for $\ell \geq 0$. Since $A_j(1) = O(\rho)$ when $\rho < 1$ (Yamamoto [1981], for instance), we note that for a sufficiently large M,

$$\left|\left| \sum_{j \geq 1}^{\infty} A_j(1) d_j - \sum_{j \geq 1}^{M} A_j(1) d_j \right|\right|^2 \leq c\left(\frac{\rho}{|\lambda_k|} \right)^{2M} , \tag{A.6}$$

where c is a constant and $||a||^2 = \sum_i a_i^2$ for a vector $a = (a_i)$. Then we obtain $g(\lambda_k) = 0$ as a necessary condition under the present assumptions when $p = \infty$. Q.E.D.

Footnotes

1/ There are some interesting examples when a finite number of lagged values of y_{it} and a contemporary disturbance term appear in equation (1.1) in addition to the expected future values of y_{it}. Since it is straightforward to extend our discussion to this case, we deal with only (1.1) in this paper for simplicity.

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 (2.1). 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.

Also if we consider a complete system of endogenous and exogenous variables in simultaneous equations models in the sense of Koopmans (1950), a large number of conventional econometric models may be expressed as (2.1).

3/ The original system is a VAR process, then the filtered system is a vector ARMA process in general. If the filter has a unit root in this formulation, then the MA part is noninvertible. The statistical inference of vector ARMA processes with unit roots in the MA part may be an interesting research topic. This problem has been pointed out by a referee. However, many econometric studies have been made under the assumption that the filtered system of $\{y_t^*\}$ is a VAR (or VARMA) process. This formulation implies i) that the system for the original process $\{y_t\}$ derived from the system of the filtered process $\{y_t^*\}$ is a VAR (or VARMA), and ii) that the VAR (or VARMA) system for $\{y_t\}$ contains constraints among its coefficient matrices. We adopted the latter formulation, which is crucial for the following results.

4/ 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 White (1958) and Dickey and Fuller (1979) for univariate AR models with unit roots and Anderson (1959) for an AR model with an explosive root. For the unit root case, the asymptotic distribution of the ordinary least squares estimator is a function of the Brownian motion B(t). Phillip (1987) has recently elaborated this finding.

References

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

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

Campbell, J.Y. and Shiller, Robert J. (1987), "Cointegration and Tests of Present Value Models," Journal of Political Economy, 95, 1062-1088.

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.

Engle, F. Robert and C. W. J. Granger (1987), "Co-Integration and Error Correction: Representastion, Estimation, and Testing," Econometrica, 55, 251-276.

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, 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, 41-58.

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

Koopmans, Tjalling C. (1950), "When is an Equation System Complete for Statistical Purposes ?," in Statistical Inference in Dynamic Economic Models, Wiley, 393-409.

Kunitomo, N. and T. Yamamoto (1986), "Testing Hypotheses and Causal Ordering in Multivariate Time Series Models with an Analysis of Japanese Macroeconomy I and II", The Journal of Economics (in Japanese), 52, University of Tokyo.

Levy, E. and Nobay, A.R. (1986) ,"The Speculative Efficiency Hypothesis: A Bivariate Analysis," The Economic Journal, 100-112.

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

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

Phillips, P.C.B. (1987), "Time Series Regression with a Unit root," Econometrica, 55, 277-301.

Sargent, Thomas J. (1979), "A Note on Maximum Likelihood Estimation of the Rational Expectation Model of the Term Structure, Journal of Monetary Economics, 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, 1190-1219.

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

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

Yamamoto, Taku. (1981), "Predictions of Multivariate Autoregressive-Moving Average Models," Biometrika, 68, 485-92.

White, J.S. (1958), "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," Annals of Mathematical Statistics, 57, 477-480.