No.1

Tests for Long-Run Granger Non-Causality in Cointegrated Systems

Taku Yamamoto and Eiji Kurozumi

December 2003

Hitotsubashi University Research Unit
for Statistical Analysis in Social Sciences

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

TESTS FOR LONG-RUN GRANGER NON-CAUSALITY IN COINTEGRATED SYSTEMS *

TAKU YAMAMOTO† EIJI KUROZUMI

Department of Economics
Hitotsubashi University

Key Words and Phrases : Vector autoregression; Cointegration; Long-run causality; Hypothesis testing.

JEL classifications : C32, C12

*Kurozumi's earlier research was partially supported by JSPS Research Fellowships for Young Scientists and by Ministry of Education, Science and Culture under JSPS Fellows (No. 7582), respectively. We are grateful to S. Bates and J. G. Cragg for valuable comments and to Hiroaki Chigira for able research assistance. We thank Professor E. Jondeau for making available his data for the present research. An earlier version of the paper was presented at International Congress on Modelling and Simulation, Canberra, Australia, 2001.

†Corresponding author: Taku Yamamoto. Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, JAPAN. E-mail: yamamoto@econ.hit-u.ac.jp

ABSTRACT

In this paper, we propose a new approach to test the hypothesis of long-run Granger non-causality in cointegrated systems. We circumvent the problem of singularity of the variance-covariance matrix associated with the usual Wald type test by proposing a generalized inverse procedure, and an alternative simple procedure which can be approximated by a suitable chi-square distribution. A test for the ranks of submatrices of the cointegration matrix and its orthogonal matrix plays a vital role in the former. The relevant small sample experiments indicate that the proposed method performs reasonably well in finite samples. As empirical applications, we examine long-run causal relations among long-term interest rates of three and five nations.

1. Introduction

The Granger non-causality has been one of major concepts in time series analysis of economic data for past three decades. In stationary vector autoregressive (VAR) processes, it is based upon the least squares prediction of finite period ahead, usually of the first period ahead. We may call it the "short-run Granger non-causality." See Dufour and Renault (1998) for classification of the Granger non-causality for different prediction horizons. Tests for the short-run Granger non-causality are straightforward in a stationary framework.

In cointegrated systems, such tests become more complex, since the existence of unit roots gives various complications in statistical inference. See, for example, Sims, Stock, and Watson (1990), Park and Phillips (1989), Toda and Yamamoto (1995), and in particular Toda and Phillips (1993, 1994). Further, in cointegrated systems, the least squares prediction of infinte horizon becomes meaningful in the sense it converges to finite values, contrary to stationary systems where the infinite horizon prediction converges to zero (or sample mean of the process). Then, in cointegrated system, the "long-run Granger non-causality" can be defined in addition to the usual "short-run Granger non-causality." See, for example, Bruneau and Jondeau (1999).

As a closely related concept, the long-run neutrality has also been discussed. Contrary to the long-run causality, various definitions of the long-run neutrality have been proposed. See, for example, Geweke (1986), Stock and Watson (1989), Fisher and Seater (1993), Weber (1994), and Boschen and Mills (1995) among others, in addition to Bruneau and Jondeau (1999).

In this paper, we generalize the definitions of the long-run causality and the long-run neutrality given in Bruneau and Jondeau (1999), which are based upon the infinite horizon least squares prediction derived from the vector error correction (VEC) representation of cointegrated systems. Here, the term "generalization" means that we consider "block causality", that is, causal relation from a set of variables to a set of variables, while they are concerned with "single variable causality", that is, causal relation from one variable to one variable.

Inference on the long-run prediction in cointegrated system suffers the same complictions due to unit roots discussed above, if T-asymptotics are considered where T

is the sample size. In order to circumvent the difficulty, we confine our analysis to $\sqrt{T}$ -asymptotics in this paper. Then, we instead encounter degeneracy of the variance covariance matrix of the estimator, which is vital in the derivation of the usual Wald test statistic. This degeneracy problem has been noted or discussed in the context of the long-run impact matrix, for example, in Johansen (1995) and specifically Paruolo (1997). This problem is more likely to occur in the block causality, but it can happen even in the single variable causality as empirical applications in section 5 show.

In this paper, we propose two procedures to escape the degeneracy problem for testing the long-run *block* Granger non-causality in cointegrated systems. Needless to say, the generalized inverse procedure is a standard way to circumvent such situations. However, in practice, its success crucially depends upon how we detect the true (degenerated) rank of a matrix concerned. We show that it depends upon the ranks of submatrices of the cointegrating matrix and its orthogonal matrix. In order to get the necessary rank information, we resort to a newly developed testing procedure by Kurozumi (2003) for testing those ranks. We also propose an alternative simple test statistic which is practically free from such a rank information.

The remainder of the paper is organized as follows. In section 2 we introduce the model and give the definitions of long-run Granger non-causality and long-run neutrality, and testable conditions for them. In section 3 we first derive the asymptotic distribution of the coefficient matrix of the infinite horizon prediction, and explain why the usual Wald test statistic may fail. Then, we propose two test procedures, one based upon the generalized inverse method and an alternative simple one, to circumvent the degeneracy problem. In section 4, we examine finite sample properties of two proposed test procedures. In section 5 we apply the test procedures to examine causal relations among long-term interest rates in five nations; the U.S., Germany, France, the Great Britain, and Japan. Finally, in section 6, we give a brief concluding remarks.

2. Model, Assumptions, and Long-Run Non-Causality

We first define the block long-run non-causality, i.e. the non-causality from a set of variables to a set of variables. Let $\{x = [x_i]\}$ be the m-element process, integrated of order one. Without loss of generality, we consider the case where the last p_2 ($p_2 \geq 1$) variables $R_R^* x$ do not cause the first p_1 ($p_1 \geq 1$) variables $R_L x$, where R_R^* and R_L are the choice matrices such that $R_R^* = [0, I_{p_2}]$, $R_L = [I_{p_1}, 0]$ and I_k is the identity matrix of rank k. Let $\underline{x}_t$ be a set of past variables x_{t-k} ($k \geq 0$), and $\underline{x}_t^*$ be $\underline{x}_t$ but without $R_R^* x_{t-k}$ ($k \geq 0$). Then, the long-run non-causality is defined interms of the best (in the sense of mean square error) linear predictions $EL(R_L x_{t+h}|\underline{x}_t)$ and $EL(R_L x_{t+h}|\underline{x}_t^*)$ where h is the prediction horizon.

Definition 1 (Long-Run Non-Causality)

$R_R^* x$ does not Granger cause $R_L x$ in the long-run if

$$\lim_{h \to +\infty} EL(R_L x_{t+h} \,|\, \underline{x}_t) = \lim_{h \to +\infty} EL(R_L x_{t+h} \,|\, \underline{x}_t^*), \tag{1}$$

that is, the knowledge of the lagged variables $R_R^* x_{t-k}$ ($k \geq 0$) does not improve the best linear prediction of $R_L x_{t+h}$.

Needless to say, the above definition is the straightforward generalization of Bruneau and Jondeau (1999) where the long-run non-causality is defined as a causal relation from one variable to one variable.

We now derive a testable condition of long-run non-causality. Consider m-vector process $\{x = [x_i]\}$ generated by vector autoregressive (VAR) model of order p,

$$A(L) x_t = d + \Theta D_t + \varepsilon_t\,, \tag{2}$$

where $x_t = [x_{it}]$, $A(L) = I_m - A_1 L - \cdots - A_p L^p$, L is the lag operator, d is the $m \times 1$ constant vector, $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and nonsingular covariance matrix $\Sigma_{\varepsilon\varepsilon}$. The deterministic terms D_t can contain a linear time, seasonal dummies, intervention dummies, or other regressors that we consider fixed and non-stochastic. Suppose that we know the true lag length p. Following Johansen (1988, 1991), we assume the following:

Assumption (Cointegration): *System* (2) *satisfies*

(i) $|A(z)| = 0$ *has its all roots outside the unit circle or equal to* 1.

(ii) $\Pi = \alpha\beta'$, *where* $\Pi = -A(1)$, α *and* β *are* $m \times r$ *matrices of rank* r, $0 < r < m$, *and rank*$\{\Pi\} = r$. *Without loss of generality, it will be assumed that* β *is orthonormal.*

(iii) *rank*$\{\alpha_\perp'\Gamma\beta_\perp\} = m - r$, *where* $\alpha_\perp$ *and* $\beta_\perp$ *are* $m \times (m-r)$ *matrices such that* $\alpha_\perp'\alpha = 0$, $\beta_\perp'\beta = 0$, *and* $\Gamma = -(\partial A(z)/\partial z)_{z=1} - \Pi$.

These assumptions imply that each component of x_t is I(1), and linear combinations of $\beta' x_t$ are stationary. The components of x_t are cointegrated with the cointegrating matrix β and the cointegration rank r. Subtracting x_{t-1} from both sides of (2) and rearranging the variables, we get Johansen's (1991) vector error correction (VEC) form of the process,

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + d + \Theta D_t + \varepsilon_t \, , \tag{3}$$

where $\Gamma_j = -\sum_{i=j+1}^{p} A_i \quad (j = 1, \cdots, p-1)$. The differenced process has representation

$$\Delta x_t = C(L)(d + \Theta D_t + \varepsilon_t) \, ,$$

where $C(L) = \sum_{i=0}^{\infty} C_i L^i$ with $C_0 = I_m$. Further, the vector moving average (VMA) representation of $\{x_t\}$ can be explicitly expressed as

$$x_t = C \sum_{i=1}^{t} \varepsilon_i + C_1(L)\varepsilon_t + \tau t + C(L)\Phi \sum_{i=1}^{t} D_i + x_0 - s_0 \, , \tag{4}$$

where $C = [c_{ij}] = C(1) = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'$, $C_1(L) = (C(L) - C(1))/(1-L)$, $\tau = Cd$, and $s_0 = C_1(L)\varepsilon_0$ such that $\beta' x_0 = \beta' s_0$.

In the above representation (4), C is often called the long-run impact matrix.

Next, we derive the least squares prediction of the process. Consider the companion form of the system (2) in order to express the prediction of h-period ahead explicitly.

$$X_t = \bar{A} X_{t-1} + \Xi_t \, , \tag{5}$$

where
$$\begin{aligned} X_t' &= [x_t', x_{t-1}', \cdots, x_{t-p+1}'] \, , \\ \Xi_t' &= [\varepsilon_t', 0, \cdots, 0] \, , \end{aligned}$$

$$
\begin{aligned}
\bar{A} &= \begin{bmatrix} A \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ I_{(p-1)m} \;\vdots\; 0 \end{bmatrix} \\
&= \begin{bmatrix} A_1 & A_2 & \cdot & \cdot & \cdot & A_{p-1} & A_p \\ I_m & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & I_m & & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & & \cdot & \cdot & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & I_m & 0 \end{bmatrix},
\end{aligned}
$$

and $A_1 = I_m + \alpha\beta' + \Gamma_1$, $A_i = \Gamma_i - \Gamma_{i-1}$ $(i = 2, \cdots, p-1)$, $A_p = -\Gamma_{p-1}$. The h-th period ahead best linear prediction of x_{t+h} given X_t is given by

$$x_{t+h\,|\,t} = M'\bar{A}^h X_t \equiv B_h X_t\,,$$

where $B_h = M'\bar{A}^h$, and $M' = [I_m, 0, \cdots, 0]$. The long-run prediction is defined as the least squares prediction of infinite horizon, that is, when h goes to infinity. It is known that B_h converges to a non-zero finite matrix as h goes to infinity. (See, for example, Phillips (1998).) The coefficient matrix of the long-run prediction is defined as

$$\bar{B} = [\bar{B}_1, \bar{B}_2, \cdots, \bar{B}_p] = \lim_{h\to\infty} B_h\,. \tag{6}$$

Then, the hypothesis of long-run Granger non-causality is given in the following proposition.

Proposition 1: *Let x_t be the stochastic process generated by the VAR model (2). Then, $R_R^* x$ does not Granger cause $R_L x$ in the long-run, if and only if*

$$
\begin{aligned}
R_L \bar{B} R_R' &= 0\,, \quad \textit{or equivalently} \\
R_L \bar{B}_i R_R^{*\prime} &= 0\,, \quad (i = 1, 2, \cdots, p),
\end{aligned} \tag{7}
$$

where $R_R = I_p \otimes R_R^*$.

In what follows we take the condition (7) as the null hypothesis H_0 for testing the long-run Granger non-causality. Proposition 1 of Bruneau and Jondeau (1999) gives a similar result for the case of $p_1 = p_2 = 1$. Our result gives an alternative expression of testable restrictions for the case where p_1 and/or p_2 are greater than unity. Since expressions of testable restrictions in (7) and in Bruneau and Jondeau (1999) are quite different, their equivalence is shown in Appendix A for completeness. It is easily seen that we have

$$\bar{B} = \left[\bar{B}_1, \bar{B}_2, \cdots, \bar{B}_p\right] = C\left[I_m, -\Gamma_1, \cdots, -\Gamma_{p-1}\right], \tag{8}$$

where C is the long-run impact matrix defined in (4). (See, for example, Chigira (2003).)

While there are various definition of long-run neutrality in the literature, we here adopt that of Bruneau and Jondeau (1999), which is defined in terms of the long-run impact matrix as follows:

Definition 2 (Long-Run Neutrality)

Let x_t be the stochastic process generated by the VAR model (2). Then, $R_R^* x$ is neutral to $R_L x$ in the long-run if

$$\begin{aligned} R_L \bar{B} R'_{R,N} &= 0, \quad \textit{or equivalently} \\ R_L C R_R^{*\prime} &= 0, \end{aligned} \tag{9}$$

where $R_{R,N} = e'_p \otimes R_R^*$, and e_p is the $p \times 1$ vector such that $e_p = [1, 0, \cdots, 0]'$.

In what follows, we take the condition (9) as the null hypothesis H_{0N} for testing the long-run neutrality. Needless to say, the long-run neutrality is a necessary condition of the long-run Granger causality.

3. Tests for Long-Run Non-Causality

3.1. Asymptotic Distribution and Wald-Type Test Statistics

In this subsection, we first derive the asymptotic distribution of coefficient matrix of the best linear prediction, and then we show that the usual Wald-type test is generally not feasible for the test of long-run non-causality. In order to test the hypothesis (7), we first estimate the VEC form (3) of the process by the ML method. See, for example, Johansen (1988,1991) for ML estimation. It is important to note that the model should be estimated in the VEC form (3) by ML rather than the levels VAR form (2), since, as Phillips (1998) points out, the latter cannot give the consistent estimate of the coefficients for the long-run prediction. Here, the coefficients of the levels VAR form

(2) are derived from the VEC estimates. The asymptotic distributions of coefficient matrices of the h-period ahead prediction $\hat{B}_h$ and the long-run prediction $\hat{\bar{B}}$ are given in the following Proposition.

Proposition 2: *Let Assumption holds and let $\hat{B}_h$ be estimates of the least squares prediction matrix B_h obtained from the ML estimates on the VEC representation* (3).

(i) *For fixed h, we have*

(a) $\hat{B}_h \xrightarrow{p} B_h$, *and*

(b) $\sqrt{T}\, vec(\hat{B}_h - B_h) \xrightarrow{d} N(0, \Sigma_h)$,

where $vec(\,\cdot\,)$ is the row-stacking operator, $\Sigma_h = F_h \Sigma_{vec} F_h'$, $\Sigma_{vec} = \Sigma_{\varepsilon\varepsilon} \otimes \Sigma_{\xi\xi}^{-1}$, $\Sigma_{\xi\xi} = E[\xi_t \xi_t']$, $\xi_t = [(\beta' x_{t-1})', \Delta x_{t-1}', \cdots, \Delta x_{t-p+1}']'$, $F_h = \sum_{i=0}^{h-1} C_i \otimes \bar{A}'^{h-1-i} K'^{-1} G_\xi$, $C_i = M' \bar{A}^i M$ *is the i-th impulse response matrix,*

$$G_\xi = \begin{bmatrix} \beta & 0 \\ 0 & I_{(p-1)m} \end{bmatrix}, \quad \textit{and}$$

$$K^{-1} = \begin{bmatrix} I_m & & & & \\ I_m & -I_m & & 0 & \\ & & \cdot & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & 0 & \cdot & \cdot & \\ & & & I_m & -I_m \end{bmatrix}.$$

(ii) *If $h \to \infty$ as $T \to \infty$ with either $h = fT$ or $h/T \to 0$ where $f > 0$ is a fixed fraction of the sample, we have*

(a) $\hat{\bar{B}}_h \xrightarrow{p} \bar{B}$, *and*

(b) $\sqrt{T}\, vec(\hat{\bar{B}} - \bar{B}) \xrightarrow{d} N(0, \Sigma)$,

where $\Sigma = F \Sigma_{vec} F'$, $F = C \otimes P = C \otimes K'^{-1} G L (I_{(p-1)m+r} - E_{22}')^{-1} \begin{bmatrix} I_r & 0 \\ 0 & I_{p-1} \otimes H' \end{bmatrix}$, $G = I_p \otimes H$, $H = [\beta_\perp, \beta]$, $L' = [0, I_{(p-1)m+r}]$, *and E_{22} is defined in Appendix* B.

Proof: See Appendix B.

Before proceeding, it should be noted that, in closely related results of Phillips (1998, Ths. 2.3 and 2.9), there is an important misprint in the expression of the crucial asymptotic distribution. In his notation, $N_i = \sum_{j=0}^{i-1} \Theta_{i-1-j} \otimes M' C'^j K^{-1}$ in Theorems 2.3 and 2.9 should be $N_i = \sum_{j=0}^{i-1} \Theta_{i-1-j} \otimes M'C'^j K'^{-1}$. That is, K^{-1} should be transposed. Actually, it is correctly derived in the 14th line from the bottom of p.50 in his article, but is misprinted in the theorems.

From (ii)(b) above, we have, under H_0,

$$
\begin{aligned}
(10) \qquad \sqrt{T}\, R\, vec\{\hat{b} - b\} &= \sqrt{T}\, R\, vec(\hat{b}) \\
&\xrightarrow{d} N(0, R\Sigma R'),
\end{aligned}
$$

where $b = vec(\bar{B})$, $\hat{b} = vec(\hat{\bar{B}})$, and $R = R_L \otimes R_R$. It should be noted that the usual Wald type test statistic, under H_0,

$$
(11) \qquad W = T\{R\hat{b}\}'(R\Sigma R')^{-1}\{R\hat{b}\},
$$

is generally infeasible, because $R\Sigma R'$ is degenerate. The degeneracy of $R\Sigma R'$ comes from that of Σ

$$
\begin{aligned}
(12) \qquad \Sigma &= F\Sigma_{vec}F' \\
&= C\Sigma_{\varepsilon\varepsilon}C' \otimes P\Sigma_{\xi\xi}^{-1}P'.
\end{aligned}
$$

We may note that both $C\Sigma_{\varepsilon\varepsilon}C'$ $(m \times m)$ and $P\Sigma_{\xi\xi}^{-1}P'$ $(mp \times mp)$ are degenerate, because $C = \beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}\alpha'_\perp$ with rank $m - r$, and P is the $mp \times \{(p-1)m + r\}$ matrix. For example, if $p = 1$ and $r = 1$, then $\Sigma_{\xi\xi}$ is a scalar and $rank(P\Sigma_{\xi\xi}^{-1}P') = 1$.

3.2. Generalized Inverse Procedure

It is a usual practice to resort a generalized inverse procedure when we have invert a degenerate matrix. That is, we have, under H_0,

$$
(13) \qquad W^- = T(R\hat{b})'(R\Sigma R')^- R\hat{b} \sim \chi_s^2,
$$

where $(R\Sigma R')^-$ is the generalized inverse of $R\Sigma R'$, χ_s^2 is the chi-square distribution with s degrees of freedom, and $s = rank(R\Sigma R')$. See, for example, Rao and Mitra (1971, Th. 9.2.2).

As a special case, it is easy to obtain the test statistic, say W_N^-, for the null hypothesis of long-run neutrality H_{0N}, since (9) is a subset of (7).

$$
(14) \qquad W_N^- = T(R_N\hat{b})'(R_N\Sigma R_N')^- R_N\hat{b},
$$

where $R_N = R_L \otimes R_{R,N}$, and $R_{R,N}$ is defined in (9). Obviously, W_N^- is asymptotically distributed as χ^2_s where $s = rank(R_N \Sigma R_N')$.

In practice, it is important to obtain the information on the rank of $R\Sigma R'$ (or $R_N \Sigma R_N'$). We have the following result.

Proposition 3: *The rank of* $R\Sigma R'$ *in* (10) *is given by*

$$rank(R\Sigma R') = rank(R_L\beta_\perp) \times \{rank(R_R^*\beta) + (p-1)p_2\} \tag{15}$$

Proof: See Appendix C.

Remark 1: Since $R\Sigma R'$ is the $pp_1p_2 \times pp_1p_2$ matrix, it is easily seen that the necessary and sufficient condition for $R\Sigma R'$ to be of full rank is

$$rank(R_L\beta_\perp) = p_1 \qquad \text{and} \qquad rank(R_R^*\beta) = p_2\,.$$

Remark 2: When $rank(R_L\beta_\perp) = 0$, we have that $rank(R\Sigma R') = 0$. In this case, we also have

$$\begin{aligned} R_L\bar{B}R_R &= R_L C[I_m, -\Gamma_1, \cdots, -\Gamma_{p-1}]R_R \\ &= 0\ [I_m, -\Gamma_1, \cdots, -\Gamma_{p-1}]R_R \\ &= 0\,. \end{aligned}$$

The second equality in the above comes from the fact that $rank(R_L\beta_\perp) = 0$ means that $R_L\beta_\perp = 0$ and thus $R_L C = R_L\beta_\perp(\alpha_\perp\Gamma\beta_\perp)^{-1}\alpha_\perp' = 0$. In sum, when $rank(R_L\beta_\perp) = 0$, it automatically indicates that $R_R^* x_t$ does not Granger cause $R_L x_t$ in the long run. (See, for example, Chigira (2003).)

When $p_1 > m - r$ or $p_2 > r$, we immediately notice that $R\Sigma R'$ is degenerate by order condition. When $p_1 \le m - r$ or $p_2 \le r$, we have to detect $rank(R_L\beta_\perp)$ or $rank(R_R^*\beta)$, respectively. For that purpose, we resort to a newly proposed testing procedure by Kurozumi (2003). He develops the test procedures for

$$\begin{aligned} &H_{0r} \,:\, rank(\beta_1) = f \qquad \text{v.s.} \quad H_{1r} \,:\, rank(\beta_1) > f\,, \quad \text{and} \\ &H_{0r\perp} \,:\, rank(\beta_{\perp,1}^*) = g \quad \text{v.s.} \quad H_{1r\perp} \,:\, rank(\beta_{\perp,1}^*) > g\,, \end{aligned}$$

where $0 \le f < \min(p_2, r)$, $0 \le g < \min(p_1, m-r)$, $\beta_1 = R_R^*\beta$, $\beta_{\perp,1} = R_R^*\beta_\perp$, $\beta_1^* = R_L\beta$, and $\beta_{\perp,1}^* = R_L\beta_\perp$. Then, we have

Theorem: *Suppose that there is no trend but $d \neq 0$ in the model (3). Let $\hat{\mu}_1 \geq \hat{\mu}_2 \geq \cdots \geq \hat{\mu}_{p_2}$ and $\hat{\mu}_1^* \geq \hat{\mu}_2^* \geq \cdots \geq \hat{\mu}_{p_1}^*$ be the ordered characteristic roots of*

$$\left|\hat{\beta}_1 \hat{\Psi} \hat{\beta}_1' - \hat{\mu}\hat{\Phi}\right| = 0\,, \quad \textit{and}$$

$$\left|\hat{\beta}_{\perp,1} \hat{\bar{\Psi}} \hat{\beta}_{\perp,1}' - \hat{\mu}^*\hat{\bar{\Phi}}\right| = 0\,,$$

where $\hat{\Psi} = \hat{\alpha}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\alpha}$, $\hat{\bar{\Psi}} = \{L'(\Upsilon_T' S_{11}^+ \Upsilon_T)^{-1}L\}^{-1}$, $\hat{\bar{\beta}}_\perp = \hat{\beta}_\perp(\hat{\beta}_\perp'\hat{\beta}_\perp)^{-1}$, $S_{11}^+ = T^{-1}\sum_{t=1}^T R_{1t}R_{1t}'$, *$R_{1t}$ being the regression residual of x_{t-1}^+ on $\Delta x_{t-1}, \cdots, \Delta x_{t-p+1}$, $x_{t-1}^+ = [x_{t-1}', 1]'$, " ^ " indicates the maximum likelihood estimate of the corresponding parameter, L and Υ_T are $(m-r+1)\times(m-r)$ and $(m+1)\times(m-r+1)$ matrices defined by*

$$\begin{aligned} L &= \begin{bmatrix} I_{m-r} \\ 0 \end{bmatrix}, \qquad \Upsilon_T = \begin{bmatrix} T^{-1/2}\hat{\bar{\beta}}_\perp & 0 \\ 0 & 1 \end{bmatrix}, \\ \hat{\Phi} &= \hat{\beta}_1(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}_1' + \hat{\beta}_{\perp,1}(\hat{\beta}_\perp'\hat{\beta}_\perp)^{-1}L'(\Upsilon_T' S_{11}^+ \Upsilon_T)^{-1}L(\hat{\beta}_\perp'\hat{\beta}_\perp)^{-1}\hat{\beta}_{\perp,1}'\,, \end{aligned}$$

and

$$\hat{\bar{\Phi}} = \hat{\beta}_{\perp,1}^*(\hat{\beta}_\perp'\hat{\beta}_\perp)^{-1}\hat{\beta}_{\perp,1}^{*\prime} + \hat{\beta}_1^*(\hat{\beta}'\hat{\beta})^{-1}(\hat{\alpha}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\alpha})^{-1}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}_1^{*\prime}\,.$$

Then, under H_{0r} and $H_{0r\perp}$, we have

$$\begin{aligned} \mathcal{L} = T^2\sum_{i=f+1}^{p_2}\hat{\mu}_i &\xrightarrow{d} \chi^2_{(p_2-f)(r-f)}\,, \quad \textit{and} \\ \mathcal{L}_\perp = T^2\sum_{i=g+1}^{p_1}\hat{\mu}_i^* &\xrightarrow{d} \chi^2_{(p_1-g)(m-r-g)}\,, \end{aligned}$$

respectively.

Proof: See Theorems 3 and 4 in Kurozumi (2003).

The above theorem specifically concerns with the case where the constant term d in (3) is such that $d = \alpha\rho_0$ where ρ_0 is the $r \times 1$ vector, and the model (3) can be rewritten as

$$\Delta x_t = \alpha\beta^{+\prime}x_{t-1}^+ + \sum_{j=1}^{p-1}\Gamma_j\Delta x_{t-j} + \Theta D_t + \varepsilon_t\,, \tag{16}$$

where $\beta^+ = [\beta', \rho_0]'$. This specification of d corresponds to empirical applications discussed in section 5. For different specifications of d, the test statistics should be slightly modified. See Kurozumi (2003) for detail.

We conduct the above test sequentially. For example, we first test $H_{0r} : f = 0$ against $H_{1r} : f > 0$. If it is accepted, we conclude that $f = 0$. If it is rejected, we proceed to test $H_{0r}: f = 1$ against $H_{1r}: f > 1$, and continues the process until H_{0r} is accepted. If $H_{0r}: f = \min(p_2, r) - 1$ is rejected, it is judged that $R_R^{*}\beta$ is of full rank. A similar sequential procedure is used for testing $H_{0r\perp}$.

Remark 3: When the cointegration rank is found to be one, i.e. $r = 1$, H_{0r} : $rank(R_R^{*}\beta) = 0$ is equivalent to the exclusion hypothesis $H'_{0r} : R_R^{*}\beta = 0$, and the testing procedure by Johansen (1991) or Johansen and Juselius (1990) may be used, instead of Kurozumi's test.

3.3. An Alternative Test Statistic and Its Approximate Distribution

In this subsection, we propose an alternative test statistic

$$W^{+} = T(R\hat{b})'(R\hat{b})\,, \tag{17}$$

that is, the sum of squares of restricted coefficient estimates, $R\hat{b}$. It will be shown that its asymptotic distribution is approximated by a suitable chi-square distribution. The following approximation was applied, for example, in Kunitomo and Yamamoto (1986) in a different context, namely the development of a test statistic for the variance decomposition, but it is given here for completeness. First, we need the following lemma,

Lemma: Suppose that U is the $m \times 1$ vector such that $U \sim N(0, G)$, where $\mathrm{rank}(G) = s \le m$. *Let $\lambda_j > 0$ $(j = 1, 2, \cdots, s)$ be distinct characteristic roots of G. Then,*

$$U'U \sim \sum_{j=1}^{s} \lambda_j X_j^2\,, \tag{18}$$

where $\{X_j\}_{j=1}^{s}$ are i.i.d. $N(0, 1)$.

Proof: Let $t_j\,(j = 1, \cdots, s)$ be characteristic vectors corresponding to $\lambda_j\,(j = 1, 2, \cdots, s)$, that is,

$$GT_1 = T_1\Lambda_1\,,$$

where $\Lambda_1 = diag\{\lambda_j\}$, $T_1 = [t_j]$ and $T_1'T_1 = I_s$. Since $\mathrm{rank}(G) = s$, there exists $T_2 = [t_1^{*}, \cdots, t_{m-s}^{*}]$ such that

$$GT_2 = 0\,, \quad T_2'T_2 = I_{m-s}\,, \quad \text{and} \quad T_1'T_2 = 0\,.$$

Then, we can define the $m \times 1$ vector $X = [X_i]$ as

$$X = \begin{bmatrix} X_I \\ X_{II} \end{bmatrix} = \begin{bmatrix} X_I \\ 0 \end{bmatrix} = \Lambda^{-\frac{1}{2}} T' U \sim N\left(0, \begin{bmatrix} I_s & 0 \\ 0 & 0 \end{bmatrix}\right),$$

where $T = [T_1, T_2]$ and

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \Lambda^{-1/2} = \begin{bmatrix} \Lambda_1^{-1/2} & 0 \\ 0 & 0 \end{bmatrix}.$$

Then, we have

$$\begin{aligned} U'U = (X'\Lambda^{\frac{1}{2}} T')(T\Lambda^{\frac{1}{2}} X) &= X'\Lambda X \\ &= X_I' \Lambda_1 X_I \qquad \text{Q.E.D.} \end{aligned}$$

Next, we consider the distribution of

$$Y = \sum_{j=1}^{s} \lambda_j X_j^2 , \tag{19}$$

where $\{X_j\}_{j=1}^s$ are i.i.d. $N(0,1)$ and $\lambda_j > 0$ for all j. In general, the exact distribution of Y depends on the nuisance parameter λ_j and it may be tedious to derive it for a practitioner. Instead, we approximate the distribution of Y by $a\chi_f^2$, as discussed in Chapter 29 of Johnson and Kotz (1970) and Satterthwaite (1941), where a and f chosen to make the first two moments in agreement with those of Y. These moments can be easily calculated and given by

$$\begin{aligned} E[Y] &= \sum_{j=1}^{s} \lambda_j , & Var[Y] &= 2\sum_{j=1}^{s} \lambda_j^2 , \\ E[a\chi_f^2] &= af , & Var[a\chi_f^2] &= 2a^2 f . \end{aligned}$$

Then, we have

$$a = \frac{\sum_{j=1}^{s} \lambda_j^2}{\sum_{j=1}^{s} \lambda_j}, \quad f = \frac{\left(\sum_{j=1}^{s} \lambda_j\right)^2}{\sum_{j=1}^{s} \lambda_j^2} . \tag{20}$$

If we regard $\sqrt{T}R\hat{b}$ in (13) as U and the characteristic roots of $R\Sigma R'$ as λ_j ($j = 1, 2, \cdots, s$) in the above lemma, we have, under H_0,

$$W^+ = T(R\hat{b})'R\hat{b} \underset{\text{approx.}}{\sim} a\chi_f^2 . \tag{21}$$

We also derive the test statistic for the long-run neutrality as a special case of W^+ as follows:

$$W_N^+ = T(R_N\hat{b})'(R_N\hat{b})\,, \tag{22}$$

where R_N is defined in (14). It is easily seen that, under H_{0N}, W_N^+ can be approximated by $a\chi_f^2$, where a and f are calculated from (20) with λ_i's being the characteristic roots of $R_N\Sigma R_N'$.

Note that, in general, the degrees of freedom, f, is fractional. Significance points for χ^2 with degrees of freedom differing by 0.2 are given in Pearson and Hartley (1976). Further, the computer package GAUSS has a convenient built-in function "cdfchinc" which returns a p-value for a chi-square distribution with the fractional degrees of freedom. We will use it in the experiments and applications later.

Finally, it should be noted that, contrary to the generalized inverse procedure in the previous subsection, the choice of s for $rank(R\Sigma R')$ is not so crucial in the present procedure as long as we take s to be large enough. Because, adding redundant λ_i's does not increase (19) so much, since they should be negligibly small by definition.

3.4. Proposed Test Procedures

Obviously, we should use W in (11) when $R\Sigma R'$ is of full rank, whereas we should use W^- in (13) or W^+ in (21) when $R\Sigma R'$ is degenerate. Thus, we propose the following test procedures which consist of thee steps.

Step 1 : Determine the cointegration rank r by the Johansen procedure (1991), estimating the VEC model by the maximum likelihood method.

Step 2 : Given the cointegration rank r, determine the rank of $R\Sigma R'$, s, by testing $rank(R_R^*\beta)$ and $rank(R_L\beta_\perp)$ with the Kurozumi procedure (2003).

Step 3 : Test the long-run Granger no-causality with W when $R\Sigma R'$ is found to be of full rank, and with W^- or W^+ with an appropriate rank s when $R\Sigma R'$ is degenerate. The combination of W and W^-, which is denoted here as com^-, and that of W and W^+ as com^+, are the ultimate test statistics proposed in this paper.

4. Finite Sample Experiments

In this section, we examine and compare the finite sample properties, namely, empirical size and (size corrected) empirical power of test statistics, com^+ and com^- proposed in Section 3.4. See also our earlier study (Yamamoto and Kurozumi (2001)) for preliminary finite sample experiments on W^+ .

Model and Design of Experiment

We examine a simple model with $m = 4$, $p = 2$, and $r = 2$, which can be described in the following VEC form,

$$\Delta x_t = \alpha\beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_t \,,$$

where $\{\varepsilon_t\}$ is i.i.d. $N(0, I_4)$.

We are concerned with the hypothesis that x_3 and x_4 do not cause x_1 and x_2 in the long-run. Namely, we test the hypothesis H_0 in (7) with

$$R_L = [I_2, 0], \quad \text{and} \quad R_R^* = [0, I_2]\,.$$

We consider two particular cases for the above model.

Case 1:

$$\alpha = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -0.5 & 0.2 \\ -0.5 & 0.5 \end{bmatrix}, \quad \beta = \begin{bmatrix} 0.5 & -1 \\ -0.5 & 0 \\ 1 & 1 \\ 1 & 0.3 \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} 0.3 & -0.5 & 0 & 0 \\ 0.5 & -0.5 & \delta & 0 \\ -0.1 & 0.1 & -0.3 & 0.3 \\ -0.1 & 0.1 & -0.3 & 0.6 \end{bmatrix}$$

$$\alpha_\perp = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad \beta_\perp = \begin{bmatrix} 1 & 0.3 \\ 3 & 2.3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Case 2:

$$\alpha = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -0.5 & 0.2 \\ -0.5 & 0.5 \end{bmatrix}, \quad \beta = \begin{bmatrix} 0.5 & -1 \\ -0.5 & 0 \\ 1 & 1 \\ 0.5 & 0.5 \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} 0.3 & -0.5 & 0 & 0.1 \\ 0.5 & -0.5 & \delta & 0 \\ -0.1 & 0.1 & -0.3 & 0.3 \\ -0.1 & 0.1 & -0.3 & 0.6 \end{bmatrix}$$

$$\alpha_\perp = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad \beta_\perp = \begin{bmatrix} 1 & 0.5 \\ 3 & 1.5 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

In Case 1, $R_R^*\beta$ and $R_L\beta_\perp$ are both of full rank, whereas in Case 2, they are both degenerate, i.e., $rank(R_R^*\beta) = rank(R_L\beta_\perp) = 1$. In both cases, we set $\delta = 0.0, 0.1$, and 0.2. The case of $\delta = 0.0$ corresponds to the experiment for empirical size and those of $\delta = 0.1$ and 0.2 to empirical power. The sample size T is taken to be 100, 200, and 400, and the number of replication is 5,000 throughout the experiment. All computations are done on GAUSS.

Notation for Tables from 1a to 2b

We first explain the notation in Tables from 1a to 2b. The column "r" indicates a possible cointegration rank to be selected by the trace test in Johansen (1988) at 1% significance level. The column "%" next to it shows an empirical distribution of the selected cointegration rank. The critical value is drawn from Table 0 of Osterwald-Lenum (1992). Note that the row for $r = 0$ is omitted from the table, since there are virtually no occurrence. The row for $r = 4$ is added for completeness. While $r = 4$ is selected in Tables 2a and 2b, there are no entries. When $r = 4$, the system is purely stationary and there should be no long-run relations in the system.

The column "rank" indicates the rank of $R\Sigma R'$ selected by the Kurozumi procedure (2003) at 1% significance level: "full" means that $R\Sigma R'$ is of full rank, i.e., $rank(R\Sigma R') = pp_1p_2$, and "deg" means that $R\Sigma R'$ is degenerate, i.e., $0 < rank(R\Sigma R') < pp_1p_2$. Further, "null" means that $rank(R\Sigma R') = 0$, which corresponds to the case of no causality as described in Remark 2 in Section 3.2. Thus, there should be no entries in the row "null". The column "%" next to "rank" shows an empirical distribution of $rank(R\Sigma R')$ for a given r.

The columns "W", "W^+" and "W^-" show rejection percentages for testing H_0 in (7) at 5% significance level for a given $rank(R\Sigma R')$. We employ the usual W statistic when $R\Sigma R'$ is of full rank, and W^+ or W^- when it is degenerate.

The column "com$^+$" shows a weighted sum of the corresponding rejection percentages in columns "W" and "W^+". The column "com$^-$" is a similar weighted sum of the corresponding columns "W" and "W^-". As explained earlier, com$^+$ and com$^-$ represent the proposed procedures for testing the long-run Granger non-causality in the present paper.

Finally, the row "total" in each sample size shows an appropriate weighted average of rejection percentages for each test statistic.

Results of Experiment: Case 1

Table 1a shows the empirical size for Case 1, where the true $R\Sigma R'$ is of full rank. When, the correct cointegration rank, 2, is selected, the rank of $R\Sigma R'$ is correctly detected for all sample sizes. In this case, the usual Wald statistic W is employed. However, it appears that the empirical size is much greater than the nominal size of 5% when $T = 100$, although it decreases to a reasonable level of 7.3% when $T = 400$.

When $r = 1$ or $r = 3$, that is, when an incorrect cointegration rank is selected, W^+ or W^- is exclusively selected. Their absolute size distortion generally smaller than that of W, and they are conservative when $T = 200$ and 400. The combined statistics com^+ and com^- show essentially similar results as W, but slightly less size distorted than W, because of the contribution of conservative W^+ and W^-.

Table 1b shows the empirical power for Case 1 when $\delta = 1$ and 2. It appears that the empirical powers of com^+ and com^- increases smoothly as δ or T increases. But it should be noted that their seemingly good power performance actually come from a large weight on W when $r = 2$.

Results of Experiment: Case 2

Table 2a shows the empirical size for Case 2, where the true $R\Sigma R'$ is degenerate. There are a few disturbing results. The usual Wald statistic W shows 100% rejection when $r = 2$ and $R\Sigma R'$ is of full rank, and W^- shows 40% or more rejection when $r = 3$ and $R\Sigma R'$ is degenerate, for all sample sizes. However, fortunately these disturbing results do not contribute to severe size distortion in combined statistics com^- and com^+, because their weights are relatively small. Other entries show relatively conservative results. Overall size performance of com^+ and com^- are relatively liberal, while the size distortion is slightly smaller for com^+. It may be noted that we now have a positive percentage in "null" case described in Remark 2 in Section 3.2.

We can examine the empirical size property of the Kurozumi procedure (2002) in this particular specification. Given that the correct cointegration rank, 2, is selected, we expect that the selection of full rank to be 1%. We find that they are 6%, 2.9%, and 1.6% when T = 100, 200, and 400, respectively. Thus, while the size distortion is relatively large when the sample size is small, say $T = 100$, it quickly diminishes as

T increases.

Table 2b shows the empirical power of Case 2 for $\delta = 1$ and 2. It appears that the empirical power of com^+ does not increases in comparison with that of com^-, when δ or T increases. This is because the power of W^+ does not increase smoothly and its weight is high in calculating com^+.

Summary of the Experiments

From the above results, we can see that both com^+ and com^- show similar and reasonable size performance when the sample size is large, say, $T = 400$. In terms of empirical power, com^- appears to be more powerful than com^+ as shown in Case 2. Thus, we recommend the use of com^- in practice in samples with about $T = 400$ or more. If we use it in smaller samples, we should be reminded that the test is rather liberal.

5. Empirical Applications

We examine the long-run Granger causality among long-term interest rates among several countries.

5.1. Three Country Case

We first examine a three country model studied by Bruneau and Jondeau (1999) with the same dataset. The dataset consists of 10-year benchmark interest rates for the US dollar (USD), the Deutschmark (DEM), and the French franc (FRF). The sample covers weekly data from January 5, 1990 to June 27, 1997 with the sample size $T = 391$. Following Bruneau and Jondeau (1999), dummy variables are used for 92:09:04, 94:06:17, 94:07:29, 94:09:30, and 97:01:17.

Main estimation and test results are given in Table 3. In Tables 3 and 4, superscripts a, b, and c indicate that statistics are statistically significant at 1%, 5% and 10% level, respectively. Panel (A) of Table 3 gives the results of the ADF test for a unit root and the Leybourne and McCabe (1994) test for stationarity. They both strongly suggest the existence of a unit root in every series. The VEC model is fitted by Johansen's (1991) maximum likelihood method. The optimal lag length is selected as 4 by the Hannan and Quinn (1979) criterion (Panel (B)). Panel (C) gives the results of the Johansen (1991) likelihood ratio statistic of testing for a trend in the system. It indicates that it is accepted that there is no trend in the system. Given this result, the estimates based upon the VEC model (16) is adopted. Panel (D) gives the results of the Johansen (1991) tests for the cointegration rank, where "Eig" denotes the ordered eigen values, "trace" the trace test statistic, and "l-max" the maximum eigen value test statistic. We conclude that the cointegration rank is one at 5% significance level. Here, the critical value for the test is drawn from Table 1 in Osterwald-Lenum (1992). The above results are all conformable with those of Bruneau and Jondeau (1999). Panels (E) and (F) give estimates of the loading vector α and the cointegrating vector β, respectively where the last element in β is an estimate of a constant term in the cointegrating vector.

Panel (G) gives the results of the test for the long-run Granger non-causality. Here, we resort to the com^{-} procedure because it was shown to be more powerful than com^{+} in the previous section. Figure 1 depicts the long-run Granger causality which is statistically significant at 5% significance level. In the top figure, the single variable causality is depicted. We may note that $H_{0r} : rank(R_R^* \beta) = 0$ is not rejected for USD by Kurozumi's test at 5% significance level, although Bruneau and Jondeau (1999) found that USD is not excluded from the cointegrating vector. Thus, we use the test statistic W^{-} in testing causality from USD to DEM or to FRF. It is interesting to note that there is no causal relation between USD and FRF, but there are feedbacks between USD and DEM and between DEM and FRF. These results are generally conformable with those of Bruneau and Jondeau (1999), except two relatively minor differences. Namely, they found causality from USD to FRF at 10% significance level but we find no such causality, and they found causality from FRF to DEM at 10% significance level but we find it at 1% siginificance level. These differences may come from the fact that we explicitly take into account the degeneracy problem. In the bottom figure, FRF and DEM are grouped. In this case, we find feedback between USD and a group of FRF and DEM. We may note that, since the cointegration rank is one, the test statistic W^{-} must be used for testing causality from a group of FRF and DEM to USD.

5.2. Five Country Case

We next examine a five country case by adding interest rates of the Great Britain pound (GBP) and Japanese yen (JPY) to those examined above. The sample covers weekly data from January 5, 1990 to October 2, 1998 with the sample size $T = 457$, which is slightly longer than the three country case.

Main estimation and test results are given in Table 4. Panel (A) of Table 4 shows again that the results of the ADF test for a unit root and the Leybourne and McCabe (1994) test both strongly suggest the existence of a unit root in every series. The optimal lag length of a VEC model is selected as 3 by the Hannan and Quinn (1979) criterion (Panel(B)). Panel (C) gives the results of the Johansen (1991) likelihood ratio statistic of testing for a trend in the system. It again indicates that it is accepted that there is no trend in the system. Panel (D) gives the results of the Johansen (1991) tests for the cointegration rank. We conclude that the cointegration rank is one at 1% significance level. Panels (E) and (F) give estimates of the loading vector α and the cointegration vector β, respectively, where the last element in β is an estimate of a constant term in the cointegration vector.

Panel (G) gives the results of the test for the long-run Granger non-causality. "D-F-G" denotes a group of Germany, France and the Great Britain. Figure 2 depicts the long-run Granger causality which is statistically significant at 5% significance level. In the top figure, the single variable causality is depicted. We may note that $H_{0r} : rank(R_R^* \beta) = 0$ is not rejected for USD and for GBP by Kurozumi's test at 5% significance level. Again, we use the test statistic W^- in testing causality from USD or GBP to others even in the single variable causality. It is interesting to note that GBP causes all other nations but not caused by them . On the other hand, USD causes only DEM, but caused by the other countries. The feedbacks are only between USD and DEM and between JPY and FRF. The rest are unidirectional causalities. In the middle and the bottom figures, countries are grouped according to their regions. It is interesting to note that the long-run feedbacks are more evident between sets of nations rather than the unidirectional causalities observed between individual countries. As in the previous three country case, since the cointegration rank is one, we have to use the test statistic W^- when testing for causality from a set of variables to others in the middle and the bottom figures.

6. Conclusion

In this paper, we proposed two procedures to test the hypothesis of long-run Granger non-causality between sets of variables in cointegrated systems; one based on the generalized inverse procedure and the other on the direct sum of squares of restricted coefficient estimates. They circumvent the problem of possible degeneracy of the variance-covariance matrix associated with the usual Wald type test statistic. In order to detect the degeneracy, the testing procedure by Kurozumi (2003) plays an important role. The relevant finite sample experiments suggested that the former test procedure, denoted here as com^{-} is preferable, because it turned out to be more powerful in finite samples. In empirical applications, we examined long-run causal relations among long-term interest rates of three and of five nations. We found that there are many cases where the degeneracy happens, even in the single variable causality, and the proposed procedure appears to be useful.

References

Arai, Y., and T. Yamamoto (2000): "Alternative Representation for Asymptotic Distributions of Impulse Responses in Cointegrated VAR Systems," *Economics Letters*, 67, 261-271.

Boschen, J. F., and L. O. Mills (1995): "Tests of Long-Run Neutrality Using Permanent Monetary and Real Shocks," *Journal of Monetary Economics*, 35, 25-44.

Bruneau, C. and E.Jondeau (1999): "Long-Run causality, with Application to International Links between Long-Term Interest Rates," *Oxford Bulletin of Economics and Statistics*, 61, 545-568.

Chigira, H. (2003): "Test for the Long-Run Causality,"(in Japanese) M.A. Thesis, Department of Economics, Hitotsubashi University.

Fisher, M. E. and J. J. Seater (1993): "Long-Run Neutrality and Superneutrality in an ARMA," *American Economic Review*, 83, 402-415.

Dufour, J-. M. and E. Renault (1998): "Short Run and Long Run Causality in Time Series: Theory," *Econometrica*, 66, 1099-1125.

Geweke, J. (1986): "The Superneutrality of Money in the United States: An Interpretation of the Evidence," *Econometrica*, 54, 1-21.

Hannan, E. J. and B. G. Quinn (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, Series B*, 41, 190-195.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551-1580.

Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Johansen, S. and K. Juselius (1990): "Maximum Likelihood Estimation and Inference on Cointegration - with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169-240.

Johnson, N. L. and S. Kotz (1970): *Continuous Univariate Distributions-2*, Houghton Mifflin Company, Boston.

Kunitomo, N. and T. Yamamoto (1986): "Causal Ordering and Hypothesis Testing in Multivariate Time Series with Macro Econometric Applications I," (in Japanese), *Keizaigaku Ronshu*, 52, 15-27.

Kurozumi, E. (2003): "The Rank of a Sub-Matrix of Cointegration," Discussion Paper 2002-15, Department of Economics, Hitotsubashi University.

Leybourne, S. J. and B. P. M. McCabe (1994): "A Consistent Test for Unit Root," *Journal of Business and Economic Statistics*, 12, 157-166.

Osterwald-Lenum, M. (1992): "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54, 461-472.

Park, J. Y. and P. C. B. Phillips (1989): "Statistical Inference in Regressions with Integrated Processes, Part 2," *Econometric Theory*, 5, 95-131.

Paruolo, P. (1997): "Asymptotic Inference on Moving Average Impact Matrix in Cointegrated I(1) VAR Systems," *Econometric Theory*, 13, 79-118.

Pearson, E. S. and H. O. Hartley (1976): *Biometrika Tables for Statisticians Vol. II*, Cambridge University Press.

Phillips, P. C. B. (1998): "Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs," *Journal of Econometrics*, 83, 21-56.

Rao, C. R. and T. Mitra (1971): *Generalized Inverse of Matrices and its Applications*, John Wiley and Sons, New York.

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113-144.

Satterthwaite, F. E. (1941): "Synthesis of Variance," *Psychometrika*, 6, 309-316.

Stock, J. H. and M. W. Watson (1989): "Interpreting the Evidence on Money Income Causality," *Journal of Econometrics*, 40, 161-181.

Toda, H. Y. and P. C. B. Phillips (1993): "Vector Autoregression and Causality," *Econometrica*, 59, 229-255.

Toda, H. Y. and P. C. B. Phillips (1994): "Vector Autoregression and Causality: A Theoretical Overview and Simulation Study," *Econometric Reviews*, 13, 259-285.

Toda, H. Y. and T. Yamamoto (1995): "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes," *Journal of Econometrics*, 66, 225-250.

Weber, A. A, (1994): "Testing Long-Run Neutrality: Empirical Evidence for G7-Countries with Special Emphasis on Germany," *Carnegie-Rochester Conference Series on Public Policy*, 41, 67-117.

Yamamoto, T. and E. Kurozumi (2001): "Finite Sample Properties of the Test for Long-Run Granger Non-Causality in Cointegrated Systems," Proceedings of International Congress on Modelling and Simulation 2001, Modelling and Simulation Society of Australia and New Zealand Inc., 1243-1248.

Appendix A

In this appendix, we prove the equivalence of definitions of long-run non-causality in Bruneau and Jondeau (1999) and in this paper. Their Proposition 1 concerns only with a variable to a variable non-causality, but it is easily generalized to non-causality between two sets of variables and is written, in terms of our notation, as follows:

$$R_L \bar{B}_1 A(L) R_R = 0\,. \tag{A.1}$$

This is specifically rewritten as

$$R_L \bar{B}_1 R_R^* = 0 \quad \text{and} \quad R_L \bar{B}_1 A_i R_R^* = 0 \quad (i = 1, 2, \cdots, p)\,. \tag{A.2}$$

In what follows, we show that (A.2) is equivalent to (7).

Proposition A1: *Condition* (7) *is necessary and sufficient for* (A.2).

Proof: We first note that, since $\bar{B}$ is the limit of $B_h = M'\bar{A}^h$ as defined in (6), we have

$$\bar{B} = \bar{B}\bar{A}\,. \tag{A.3}$$

By the structure of $\bar{A}$, it implies the following relations.

$$\begin{aligned} &\bar{B}_i = \bar{B}_1 A_i + \bar{B}_{i+1} \quad (i = 1, 2, \cdots, p-1)\,, \quad \text{and} \\ &\bar{B}_p = \bar{B}_1 A_p\,. \end{aligned} \tag{A.4}$$

(Necessity) Suppose that (A.2) holds. We proceed $R_L \bar{B}_i R_R^*$ backward from $i = p$ to $i = 1$. When $i = p$, $R_L \bar{B}_p R_R^* = 0$ is immediate from the last relation in (A.4). For $i = p - 1$, it is immediate from $R_L \bar{B}_p R_R^* = 0$ and (A.2).

$$R_L \bar{B}_{p-1} R_R^* = R_L \bar{B}_1 A_{p-1} R_R^* + R_L \bar{B}_p R_R^* = 0 + 0 = 0\,.$$

A similar argument continues to hold until $i = 1$.

(Sufficiency) Suppose that (7) holds. The relations in (A.3) can be rearranged as follows:

$$\begin{aligned} &\bar{B}_1 A_i = \bar{B}_{i+1} - \bar{B}_i \quad (i = 1, 2, \cdots, p-1)\,, \quad \text{and} \\ &\bar{B}_1 A_p = \bar{B}_p\,. \end{aligned} \tag{A.5}$$

It is immediately seen that (7) implies (A.2).

Appendix B Proof of Proposition 2

Proof of (i)(a): Since the ML estimates $[\hat{\alpha}, \hat{\Gamma}_1, \cdots, \hat{\Gamma}_{p-1}]$ is consistent, the result immediately follows.

Proof of (i)(b): The following result is a straightforward generalization of Lütkepohl and Reimers (1992) and Phillips (1998) who deal with the asymptotic distribution of the estimate of impulse response matrix. We first note that the asymptotic distribution of vec$\{[\hat{\alpha}, \hat{\Gamma}_1, \cdots, \hat{\Gamma}_{p-1}]\}$ is given, for example in Johansen (1995, Th.13.2),

$$\text{(B.1)} \qquad \begin{aligned} &\sqrt{T} vec \left\{ [\hat{\alpha}, \hat{\Gamma}_1, \cdots, \hat{\Gamma}_{p-1}] - [\alpha, \Gamma_1, \cdots, \Gamma_{p-1}] \right\} \\ &\qquad \xrightarrow{d} N(0, \Sigma_{vec}) \,. \end{aligned}$$

The coefficients in a levels VAR model are related to those in a VEC model as follows:

$$\begin{aligned} A &= [I_m + \alpha\beta' + \Gamma_1, \Gamma_2 - \Gamma_1, \cdots, \Gamma_{p-1} - \Gamma_{p-2}, -\Gamma_{p-1}] \\ &= [\alpha\beta, \Gamma_1, \cdots, \Gamma_{p-1}]K^{-1} + [I_m, 0, \cdots, 0] \\ &= [\alpha, \Gamma_1, \cdots, \Gamma_{p-1}]G'_\xi K^{-1} + [I_m, 0, \cdots, 0] \,. \end{aligned}$$

Then, we have

$$\text{(B.2)} \qquad \frac{\partial\, vec[A]}{\partial \{vec[\alpha, \Gamma_1, \cdots, \Gamma_{p-1}]\}'} = I_m \otimes K'^{-1} G_\xi \,,$$

and

$$\text{(B.3)} \qquad \sqrt{T} vec[\hat{A} - A] \xrightarrow{d} N(0, \Sigma_A) \,,$$

where $\Sigma_A = (I_m \otimes K'^{-1} G_\xi) \Sigma_{vec} (I_m \otimes K'^{-1} G_\xi)' = \Sigma_{\varepsilon\varepsilon} \otimes K'^{-1} G_\xi \Sigma_{\xi\xi}^{-1} G'_\xi K^{-1}$.

Finally, we note that

$$\text{(B.4)} \qquad \begin{aligned} \frac{\partial\, vec[B_h]}{\partial \{vec[A]\}'} &= \frac{\partial\, vec[M' \bar{A}^h]}{\partial \{vec[A]\}'} \\ &= \sum_{i=0}^{h-1} M' \bar{A}^i \otimes \bar{A}'^{h-1-i} \left[\frac{\partial\, vec[\bar{A}]}{\partial \{vec[A]\}'} \right] \\ &= \sum_{i=0}^{h-1} M' \bar{A}' \otimes \bar{A}'^{h-1-i} [M \otimes I_{mp}] \\ &= \sum_{i=0}^{h-1} C_i \otimes \bar{A}'^{h-1-i} \,. \end{aligned}$$

Thus, we have

$$\text{(B.5)} \qquad \begin{aligned} \Sigma_h &= \left(\sum_{i=0}^{h-1} C_i \otimes \bar{A}'^{h-1-i} \right) \Sigma_A \left(\sum_{i=0}^{h-1} C_i \otimes \bar{A}'^{h-1-i} \right)' \\ &= F_h \Sigma_{vec} F_h' \,, \end{aligned}$$

where $F_h = \sum_{i=0}^{h-1} C_i \otimes \bar{A}'^{h-1-i} K'^{-1} G_\xi$. This completes the proof of (i)(b).

Proof of (ii): The following proof is a simple generalization of Arai and Yamamoto (2000) which originally heavily draws upon results in Phillips (1998, Appendix). Here, we are concerned with $B_h = M'\bar{A}^h$, whereas Arai and Yamamoto and Phillips are concerned with the first m columns of B_h, namely $C_h = M'\bar{A}^h M$ which is the h-th impulse response matrix.

Proof of (ii)a: By estimating the VEC representation (2) by the ML method, we can construct the estimate of B_h as in (A.4) of Phillips (1998, Appendix), namely

$$\text{(B.6)} \qquad \hat{B}_h = M' K \hat{D}^h K^{-1}$$

where $D = K^{-1}\bar{A}K$, is the companion matrix associated with an error correction form (2), $\hat{D}$ is its estimate

$$D = \begin{bmatrix} I_m + \alpha\beta' & \Gamma_1 & & \Gamma_{p-1} \\ \alpha\beta' & \Gamma_1 & & \Gamma_{p-1} \\ 0 & I_m & & 0 \\ \cdot & & \cdot & \\ 0 & & I_m & 0 \end{bmatrix}, \qquad \text{and}$$

$$K = \begin{bmatrix} I_m & 0 & \cdots & 0 \\ I_m & -I_m & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ I_m & -I_m & \cdots & -I_m \end{bmatrix}.$$

We further express $\hat{B}_h$ in terms of $\hat{E}$ that is the estimated companion matrix associated with the I(1)/I(0) VAR representation – see Phillips (1998, Appendix A.1) for the I(1)/(0) VAR representation.

$$\text{(B.7)} \qquad \hat{B}_h = M' K \hat{D}^h K^{-1} = M' K \hat{G} \hat{E}^h \hat{G}' K^{-1} \,,$$

where

$$
\text{(B.8)} \qquad E = G'DG = \begin{bmatrix} I_s & \beta_\perp'\alpha & \bar\Gamma_1 & \cdots & & \bar\Gamma_{p-1} \\ 0 & I_r+\beta'\alpha & & & & \\ 0 & \beta_\perp'\alpha & \bar\Gamma_1 & \cdots & & \bar\Gamma_{p-1} \\ 0 & \beta'\alpha & & & & \\ 0 & 0 & I_m & & & 0 \\ \cdot & \cdot & \cdot & \ddots & & \\ 0 & \cdot & 0 & & I_m & 0 \end{bmatrix},
$$

$$
= \begin{bmatrix} I_s & E_{12} \\ 0 & E_{22} \end{bmatrix} \quad \text{(say)},
$$

where

$$
E_{12} = \left[\beta_\perp'\alpha, \ddot\Gamma_1, \cdots, \dddot\Gamma_{p-1}\right],
$$

$$
E_{22} = \begin{bmatrix} I_r+\beta'\alpha & \dddot\Gamma_1 & & & \dddot\Gamma_{p-1} \\ \beta_\perp'\alpha & \bar\Gamma_1 & & & \bar\Gamma_{p-1} \\ \beta'\alpha & & & & \\ I_m & 0 & \cdots & & 0 \\ 0 & \ddots & & & \\ \cdot & & \ddots & & \\ 0 & 0 & & I_m & 0 \end{bmatrix},
$$

and $\bar\Gamma_i = H'\Gamma_i H = [\ddot\Gamma_i', \dddot\Gamma_i']'$, $\ddot\Gamma_i = \beta_\perp'\Gamma_i H$, and $\dddot\Gamma_i = \beta'\Gamma_i H$ $(i = 1, 2, \cdots, p-1)$. It is known (e.g. Phillips (1998, Appendix A.2)) that E_{22} corresponds to the stationary part of the system and has only stable roots. Note further that we assume here that H is orthonormalized without loss of generality, namely, $H'H = I_m$ and $G'G = I_{mp}$.

Now, we are in the position to consider the case where $h \to \infty$ as $n \to \infty$ with either $h = fn$ or h/n. Noting that $\hat\alpha$, $\hat\beta_\perp$ and $\hat\Gamma_i$ $(i = 1, 2, \cdots, p-1)$ are consistent estimates, we have

$$
\text{(B.9)} \qquad \begin{aligned} \hat E^h &= \begin{bmatrix} I_s & \hat E_{12}(I_{m(p-1)+r} + \hat E_{22} + \hat E_{22}^2 + \cdots + \hat E_{22}^{h-1}) \\ 0 & \hat E_{22}^h \end{bmatrix} \\ &\xrightarrow{p} \begin{bmatrix} I_s & E_{12}(I - E_{22})^{-1} \\ 0 & 0 \end{bmatrix}. \end{aligned}
$$

Thus we have

$$
\hat B_h = M'K\hat G\hat E^i\hat G'K^{-1} \xrightarrow{p} M'KG \begin{bmatrix} I_s & E_{12}(I - E_{22})^{-1} \\ 0 & 0 \end{bmatrix} G'K^{-1}
$$

$$
\begin{aligned}
\text{(B.10)} \qquad &= \beta_\perp \beta_\perp' M' + \beta_\perp E_{12}(I - E_{22})^{-1} L' G' K^{-1} \\
&= \bar{B} \quad (\text{say}),
\end{aligned}
$$

where $L' = [0, I_{(p-1)m+r}]$. It gives the required results of (ii)(a).

Proof of (ii)(b): Since $\bar{A} = KDK^{-1} = KGEG'K^{-1}$, we have $\bar{A}^k = KGE^k G' K^{-1}$. Then, F_h in (B.5) is alternatively given as

$$
F_h = \sum_{k=0}^{h-1} C_{h-1-k} \otimes K'^{-1} G E'^k G' G_\xi \,. \tag{B.11}
$$

Note that, by partitioning G as

$$
G = \begin{bmatrix} \beta_\perp & \beta & 0 \\ 0 & 0 & I_{p-1} \otimes H \end{bmatrix} = \begin{bmatrix} \beta_\perp & G_{12} \\ 0 & G_{22} \end{bmatrix} \quad (\text{say}),
$$

we have

$$
\begin{aligned}
G_\xi' G E^k G' K^{-1} &= \begin{bmatrix} \beta' & 0 \\ 0 & I_{m(p-1)} \end{bmatrix} \begin{bmatrix} \beta_\perp & G_{12} \\ 0 & G_{22} \end{bmatrix} \begin{bmatrix} I_s & E_{12}(I + E_{22} + \cdots + E_{22}^{k-1}) \\ 0 & E_{22}^k \end{bmatrix} \\
&\qquad \times G' K^{-1} \\
\text{(B.12)} \qquad &= \begin{bmatrix} 0 & \beta' G_{12} \\ 0 & G_{22} \end{bmatrix} \begin{bmatrix} I_s & E_{12}(I + E_{22} + \cdots + E_{22}^{k-1}) \\ 0 & E_{22}^k \end{bmatrix} G' K^{-1} \\
&= \begin{bmatrix} 0 & \begin{pmatrix} \beta' G_{12} \\ G_{22} \end{pmatrix} E_{22}^k \\ 0 & \end{bmatrix} G' K^{-1} \\
&= \begin{bmatrix} \beta' G_{12} \\ G_{22} \end{bmatrix} E_{22}^k L' G' K^{-1} \\
&= \begin{bmatrix} I_r & 0 \\ 0 & I_{p-1} \otimes H \end{bmatrix} E_{22}^k L' G' K^{-1} \,.
\end{aligned}
$$

Thus F_h can be written in terms of E_{22}

$$
F_h = \sum_{k=0}^{h-1} C_{h-1-k} \otimes K'^{-1} G L E_{22}^{k'} \begin{bmatrix} I_r & 0 \\ 0 & I_{p-1} \otimes H' \end{bmatrix} . \tag{B.13}
$$

Since E_{22} corresponds to the coefficient matrix for the stationary components, this representation implies the convergent property of F_h.

With regard to deriving the asymptotic distributions of $\hat{\bar{B}}$, it is enough to show that $F_h \to F$ as $h \to \infty$. From the equation (B.13),

$$
\begin{aligned}
F_h &= \sum_{k=0}^{h-1} C_{h-1-k} \otimes K'^{-1} G L E_{22}^{k'} \begin{bmatrix} I_r & 0 \\ 0 & I_{p-1} \otimes H' \end{bmatrix} \\
\text{(B.14)} \qquad & \\
&\to C \otimes K'^{-1} G L (I - E_{22}')^{-1} \begin{bmatrix} I_r & 0 \\ 0 & I_{p-1} \otimes H' \end{bmatrix} .
\end{aligned}
$$

This completes the proof of (ii)(b) of the Proposition.

Appendix C Proof of Proposition 3

We first note that Σ in (12) can be conveniently decomposed as follows:

$$
\text{(C.1)} \qquad \begin{aligned}
\Sigma &= C\Sigma_{\varepsilon\varepsilon}C' \otimes P\Sigma_{\xi\xi}^{-1}P' \\
&= \beta_\perp(\alpha_\perp\Gamma\beta_\perp)^{-1}\alpha_\perp'\Sigma_{\varepsilon\varepsilon}\alpha_\perp(\alpha_\perp\Gamma\beta_\perp)^{-1\prime}\beta_\perp' \otimes K'^{-1}GLUL'G'K^{-1} \\
&= \beta_\perp V\beta_\perp' \otimes \bar{G}U\bar{G}'\,,
\end{aligned}
$$

where $\bar{G} = K'GL,\;\; V = (\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'\Sigma_{\varepsilon\varepsilon}\alpha_\perp(\alpha_\perp\Gamma\beta_\perp)^{-1\prime}$, and

$$
U = (I_{(p-1)m+r} - E_{22}')^{-1}\begin{bmatrix} I_r & 0 \\ 0 & I_p\otimes H' \end{bmatrix}\Sigma_{\xi\xi}^{-1}\begin{bmatrix} I_r & 0 \\ 0 & I_p\otimes H \end{bmatrix}(I_{(p-1)m+r} - E_{22}')^{-1\prime}\,.
$$

Obviously, V and U are symmetric matrices and they are both full rank. Then we can decompose $R\Sigma R'$ as

$$
\text{(C.2)} \qquad R\Sigma R' = (R_L\beta_\perp V\beta_\perp' R_L') \otimes (R_R\bar{G}U\bar{G}'R_R')\,.
$$

We have

$$
\text{(C.3)} \qquad \begin{aligned}
rank(R\Sigma R') &= rank(R_L\beta_\perp V\beta_\perp' R_L') \times rank(R_R\bar{G}U\bar{G}'R_R') \\
&= rank(R_L\beta_\perp) \times rank(R_R\bar{G})
\end{aligned}
$$

The second equality in the above comes from the fact that V and U are both full rank. We further note that

$$
\text{(C.4)} \qquad R_R\bar{G} = \left[\begin{array}{c|cccc}
R_R^*\beta & R_R^* & 0 & \dots & 0 \\ \hline
0 & -R_R^* & R_R^* & & 0 \\
\vdots & & -R_R^* & \ddots & \\
\vdots & & & \ddots & R_R^* \\
0 & 0 & & & -R_R^*
\end{array}\right] \equiv \begin{bmatrix} \bar{G}_{11}^* & \bar{G}_{12}^* \\ \bar{G}_{21}^* & \bar{G}_{22}^* \end{bmatrix} \quad \text{(say)}.
$$

Then, we have

$$
\text{(C.5)} \qquad \begin{aligned}
rank(R_R\bar{G}) &= rank(\bar{G}_{11}^*) + rank(\bar{G}_{22}^*) \\
&= rank(R_R^*\beta) + (p-1)p_2\,.
\end{aligned}
$$

Inserting it into (C.3), we have the desired result.

Table 1a   Case 1: Empirical Size

T	r	%	rank	%	W	W^+	W^-	com$^+$	com$^-$
100	1	4.0	full	0.0	.	.	.	5.4	7.4
			deg	100	.	5.4	7.4		
	2	94.4	full	100	13.1	.	.	13.1	13.1
			deg	0.0	.	.	.		
	3	1.5	full	0.0	.	.	.	1.3	6.7
			deg	100	.	1.3	6.7		
	4	0.1		.	.	.	.	.	.
	total				13.1	4.3	7.2	12.6	12.8
200	1	0.0	full	.	.	.	.	.	.
			deg	.	.	.	.		
	2	98.4	full	100	9.9	.	.	9.9	9.9
			deg	0.0	.	.	.		
	3	1.5	full	0.0	.	.	.	0.0	0.0
			deg	100	.	0.0	0.0		
	4	0.1		.	.	.	.	.	.
	total				9.9	0.0	0.0	9.7	9.7
400	1	0.0	full	.	.	.	.	.	.
			deg	.	.	.	.		
	2	98.8	full	100	7.2	.	.	7.2	7.2
			deg	0.0	.	.	.		
	3	1.1	full	0.0	.	.	.	1.8	1.8
			deg	100	.	1.8	1.8		
	4	0.1		.	.	.	.	.	.
	total				7.2	1.8	1.8	7.2	7.2

For explanation of the notation, see subsection *Notation for Tables from 1a to 2b* in section 4.

Table 1b   Case 1: Empirical Power

δ	T	r	%	rank	%	W	W^+	W^-	com$^+$	com$^-$
0.1	100	1	3.5	full	0.0	.	.	.	12.6	18.3
				deg	100	.	12.6	18.3		
		2	95.0	full	100.0	17.1	.	.	17.1	17.1
				deg	0.0	.	.	.		
		3	1.4	full	0.0	.	.	.	5.4	6.8
				deg	100.0	.	5.4	6.8		
		4	0.1		.	.	.	.	.	.
		total				17.1	10.4	14.9	16.8	17.0
	200	1	0.0	full	.	.	.	.	.	.
				deg	.	.	.	.		
		2	98.4	full	100.0	36.7	.	.	36.7	36.7
				deg	0.0	.	.	.		
		3	1.5	full	0.0	.	.	.	8.1	20.3
				deg	100.0	.	8.1	20.3		
		4	0.1		.	.	.	.	.	.
		total				36.7	8.1	20.3	36.2	36.4
	400	1	0.0	full	.	.	.	.	.	.
				deg	.	.	.	.		
		2	98.8	full	100.0	75.6	.	.	75.6	75.6
				deg	0.0	.	.	.		
		3	1.1	full	0.0	.	.	.	8.6	31.0
				deg	100.0	.	8.6	31.0		
		4	0.1		.	.	.	.	.	.
		total				75.6	8.6	31.0	74.9	75.1

Table 1b (continued)

δ	T	r	%	rank	%	W	W^+	W^-	com$^+$	com$^-$
0.2	100	1	3.1	full	0.0	.	.	.	43.3	71.3
				deg	100	.	43.3	71.3		
		2	95.3	full	100.0	65.1	.	.	65.1	65.1
				deg	0.0	.	.	.		
		3	1.5	full	0.0	.	.	.	5.4	17.6
				deg	100.0	.	5.4	17.6		
		4	0.1		.	.	.	.	.	.
		total				65.1	31.2	54.1	63.5	64.6
	200	1	0.0	full	.	.	.	.	.	.
				deg	.	.	.	.		
		2	98.5	full	100.0	96.9	.	.	96.9	96.9
				deg	0.0	.	.	.		
		3	1.4	full	0.0	.	.	.	17.4	33.3
				deg	100.0	.	17.4	33.3		
		4	0.1		.	.	.	.	.	.
		total				96.9	17.4	33.3	95.8	96.0
	400	1	0.0	full	.	.	.	.	.	.
				deg	.	.	.	.		
		2	98.8	full	100.0	100.0	.	.	100.0	100.0
				deg	0.0	.	.	.		
		3	1.1	full	0.0	.	.	.	28.8	44.1
				deg	100.0	.	28.8	44.1		
		4	0.1		.	.	.	.	.	.
		total				100.0	28.8	44.1	99.2	99.3

Table 2a Case2: Empirical Size

T	r	%	rank	%	W	W^+	W^-	com$^+$	com$^-$
100	1	0.5	full	0.0	.	.	.	100.0	100.0
			deg	100	.	100.0	100.0		
	2	96.9	full	6.2	100.0	.	.	9.9	10.1
			deg	93.7	.	3.9	4.1		
	3	2.5	full	0.0	.	.	.	1.9	48.1
			deg	84.6	.	1.9	48.1		
			(null)	15.4	.	.	.		
	4	0.1		.	.	.	.	.	.
	total				100.0	4.3	5.6	10.2	11.4
200	1	0.0	full	.	.	.	.	.	.
			deg	.	.	.	.		
	2	98.0	full	2.9	100.0	.	.	7.5	7.0
			deg	97.1	.	4.8	4.2		
	3	1.9	full	0.0	.	.	.	2.5	43.8
			deg	85.1	.	2.5	43.8		
			(null)	14.9	.	.	.		
	4	0.2		.	.	.	.	.	.
	total				100.0	4.7	4.9	7.5	7.6
400	1	0.0	full	.	.	.	.	.	.
			deg	.	.	.	.		
	2	98.5	full	1.6	100.0	.	.	6.9	6.5
			deg	98.4	.	5.4	4.9		
	3	1.4	full	0.0	.	.	.	3.1	41.5
			deg	94.2	.	3.1	41.5		
			(null)	5.8	.	.	.		
	4	0.1		.	.	.	.	.	.
	total				100.0	5.4	5.4	6.9	6.9

Table 2b Case 2: Empirical Power

δ	T	r	%	rank	%	W	W^+	W^-	com$^+$	com$^-$
0.1	100	1	0.6	full	0.0	.	.	.	35.7	10.7
				deg	100	.	35.7	10.7		
		2	97.3	full	6.7	7.7	.	.	7.4	54.3
				deg	93.3	.	7.4	57.7		
		3	2.0	full	0.0	.	.	.	3.4	11.2
				deg	88.1	.	3.4	11.2		
				(null)	11.9	.	.	.		
		4	0.1		.	.	.	.	.	.
		total				7.7	7.5	56.5	7.5	53.3
	200	1	0.0	full	.	.	.	.	.	.
				deg	.	.	.	.		
		2	98.3	full	3.2	7.7	.	.	7.4	85.8
				deg	96.8	.	7.4	88.3		
		3	1.5	full	0.0	.	.	.	4.9	21.3
				deg	80.3	.	4.9	21.3		
				(null)	19.7	.	.	.		
		4	0.2		.	.	.	.	.	.
		total				7.7	7.4	87.5	7.4	85.0
	400	1	0.0	full	.	.	.	.	.	.
				deg	.	.	.	.		
		2	98.7	full	1.5	5.4	.	.	9.4	97.9
				deg	98.5	.	9.4	99.3		
		3	1.2	full	0.0	.	.	.	7.8	9.8
				deg	85.0	.	7.8	9.8		
				(null)	15.0	.	.	.		
		4	0.1		.	.	.	.	.	.
		total				5.4	9.4	98.3	9.4	97.0

Table 2b (continued)

δ	T	r	%	rank	%	W	W^+	W^-	com$^+$	com$^-$
0.2	100	1	0.5	full	0.0	.	.	.	73.9	60.9
				deg	100	.	73.9	60.9		
		2	97.6	full	7.0	16.1	.	.	11.0	90.0
				deg	93.0	.	10.6	95.6		
		3	1.8	full	0.0	.	.	.	6.8	16.4
				deg	81.1	.	6.8	16.4		
				(null)	18.9	.	.	.		
		4	0.1		.	.	.	.	.	.
		total				16.1	10.9	94.2	11.3	88.8
	200	1	0.0	full	.	.	.	.	.	.
				deg	.	.	.	.		
		2	98.5	full	3.3	16.0	.	.	12.1	97.1
				deg	96.7	.	12.0	99.9		
		3	1.4	full	0.0	.	.	.	10.0	31.7
				deg	83.3	.	10.0	31.7		
				(null)	16.7	.	.	.		
		4	0.1		.	.	.	.	.	.
		total				16.0	12.0	99.0	12.1	96.3
	400	1	0.0	full	.	.	.	.		
				deg	.	.	.	.		
		2	98.7	full	1.5	22.4	.	.	22.4	98.8
				deg	98.5	.	22.4	100.0		
		3	1.2	full	0.0	.	.	.	19.2	44.2
				deg	89.7	.	19.2	44.2		
				(null)	10.3	.	.	.		
		4	0.1		.	.	.	.	.	.
		total				22.4	22.3	99.4	22.3	98.2

Table 3 Long-Run Causality Between Long-Term Interest Rates: Three Country Case

(A) Test for Non-Stationarity of Interest Rates

	ADF test	L-M test
USD	−1.630	13.125^a
DEM	−0.584	29.757^a
FRF	−0.771	10.759^a

(B) Estimated lag length of VAR 4

(C) Test statistics for $\alpha'_{\perp}\mu = 0$ 2.423

(D) Test for the number of cointegrating vectors

Eig.	0.058	0.030	0.008
H_0	$r = 0$	$r \leq 1$	$r \leq 2$
trace	38.503^b	15.178	3.290
lmax	23.325^b	11.888	3.290

(E) Standardized adjustment coefficients α'

−0.092	0.062	−0.130

(F) Standardized cointegrating vectors β'

0.487	0.154	−0.719	0.471

(G) Test statistics for long-run Granger non-causality

from:	to:	USD	DEM	FRF
USD		.	15.578^a	4.983
DEM		11.553^b	.	90.612^a
FRF		4.097	11.239^b	.

from: DEM FRF to: USD 16.920^b

from: USD to: DEM FRF 8.328^b

Table 4 Long-Run Causality Between Long-Term Interest Rates: Five Country Case

(A) Test for Non-Stationarity of Interest Rates

	ADF test	L-M test
USD	−0.951	16.919[a]
DEM	0.699	6.670[a]
FRF	−0.096	16.206[a]
GBP	−1.081	32.298[a]
JPY	0.081	36.415[a]

(B) Estimated lag length of VAR 3

(C) Test statistics for $\alpha'_{\perp}\mu = 0$ 5.636

(D) Test for the number of cointegrating vectors

Eig.	0.109	0.058	0.034	0.014	0.007
H_0	$r = 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$	$r \leq 4$
trace	104.839[a]	52.606	25.256	9.733	3.373
lmax	52.234[a]	27.350	15.522	6.360	3.373

(E) Standardized adjustment coefficients α'

−0.193	0.107	−0.204	0.063	−0.146

(F) Standardized cointegrating vectors β'

0.711	0.177	−0.568	0.337	−0.114	0.118

Table 4 (continued)

(G) Test statistics for long-run Granger non-causality

from:	to:	USD	DEM	FRF	GBP	JPY
USD		.	17.152^a	4.679^c	4.149	3.944
DEM		16.875^a	.	85.080^a	6.684^c	23.024^a
FRF		14.020^a	2.135	.	4.71739	12.932^a
GBP		14.895^a	11.843^a	14.752^a	.	13.067^a
JPY		18.697^a	7.149^c	13.388^a	4.116	.

from:	to:	USD	D-F-G	JPY
USD		.	9.154^c	3.944
D-F-G		21.140^a	.	25.483^a
JPY		18.697^a	28.179^a	.

from: DEM FRF GBP to: USD JPY 20.726^a
from: USD JPY to: DEM FRF GBP 43.615^a

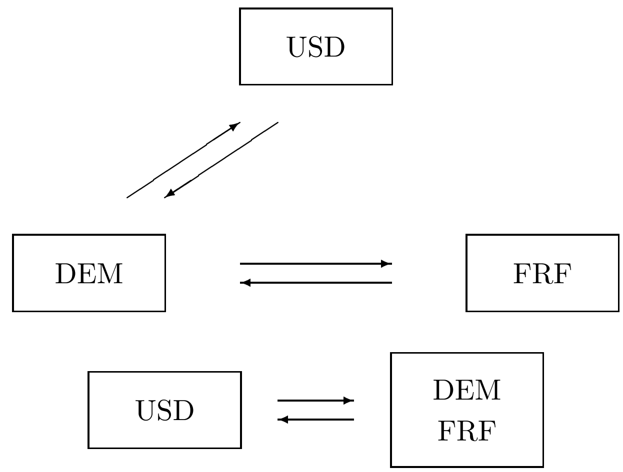

Figure 1 The long-run Granger causality at 5% significance level: Three Country Case

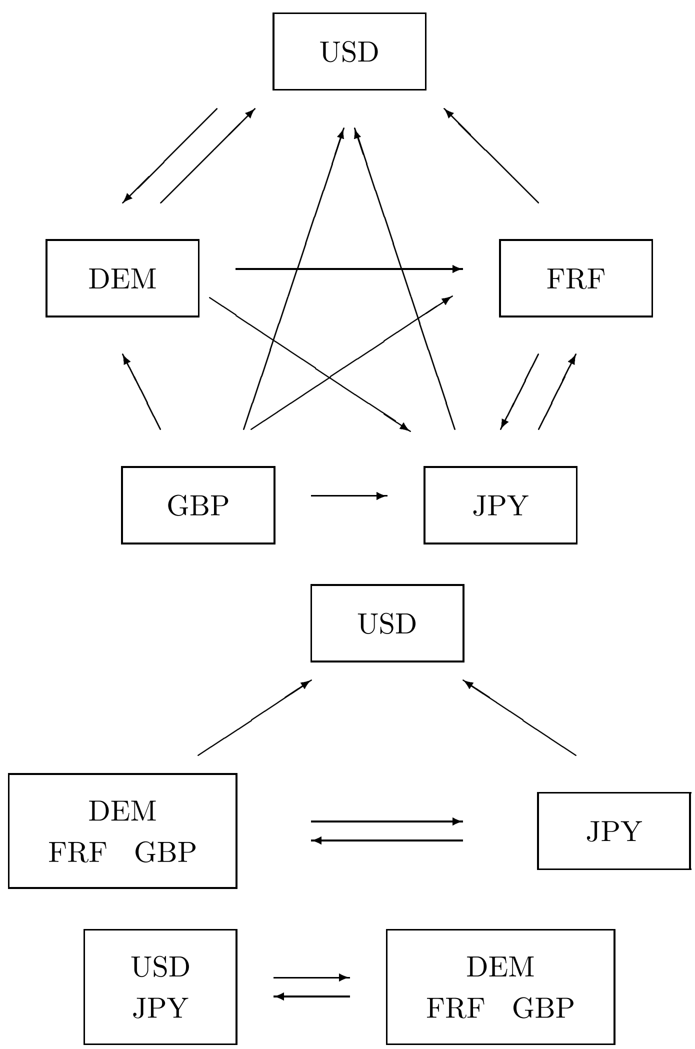

Figure 2 The long-run Granger non-causality at 5% significance level: Five Country Case