J. Japan Statist. Soc.
Vol. 36 No. 2 2006 149–171

INFERENCE ON THE COINTEGRATION RANK AND A PROCEDURE FOR VARMA ROOT-MODIFICATION

Taro Takimoto* and Yuzo Hosoya**

The paper presents a feasible numerical procedure for evaluating the maximum Whittle likelihood estimates and the likelihood-ratio statistics, where to obtain the maximum Whittle likelihood estimates under specific cointegration ranks, we introduce an iterative method in which the set of the ARMA coefficient estimates is adjusted so as to guarantee that in each step they satisfy the root conditions imposed by respective cointegration rank hypotheses. The method is incorporated in the Whittle likelihood maximization.

Key words and phrases: Cointegration rank test, invertibility, Jordan canonical form, stationarity, Whittle estimator.

1. Introduction

Since the multivariate time-series analysis enables us to scrutinize not only temporal dependence structures but also interactive relations among the component variables, it has a wide range of applications to such fields as the portfolio analysis and the term-structure analysis of interest rates in finance in addition to traditional macroeconomic time-series analysis. In this paper, we make two major improvements to the Whittle-likelihood-based computational procedure proposed previously by Takimoto and Hosoya (2004). Namely, on the basis of the multivariate autoregressive-movingaverage (ARMA) processes, we propose in this paper an eigenvalue control algorithm for estimating the maximum Whittle likelihood. Our algorithm consists of

(1) eigenvalue contraction method, and

(2) penalty-imposed likelihood function maximization.

The objective of (1) is to locate the initial values of coefficient estimates for the optimization iteration in the admissible set (where we call a set of ARMA coefficients admissible if it satisfies the root conditions given in Section 2) and (2) is to prevent the maximizer of the likelihood to depart from the admissible set. Secondly, we propose testing a null-hypothesis rank not against a constant-mean stationary VARMA process but against the trend-stationary full rank alternative.

Cointegration analysis initiated by Granger (1981) has been prevalent in the literature of long-run relationships among economic time series. Particularly for vector autoregressive (VAR) models, a number of estimation and testing methods have been proposed aiming at numerical tractability or mitigating computational

Received May 24, 2004. Revised August 9, 2004. Accepted June 8, 2006.

*Faculty of Economics, Kyushu University, Hakozaki 6-19-1, Higashi-ku, Fukuoka 812-8581, Japan. Email: takimoto@en.kyushu-u.ac.jp

**Department of Economics, Meisei University, Hodokubo 2-1-1, Hino 191-8506, Japan.

requirement. (See, for example, Engle and Granger (1987) for a single equation approach and Johansen (1988, 1991) for a system estimation approach.) On the other hand, it is known that the VARMA model can capture more complex features of data fluctuation by parsimonious use of parameters compared with the VAR model. Recently procedures to deal with the VARMA model have been provided by Yap and Reinsel (1995) (the Gaussian (conditional) likelihood approach), Lütkepohl and Claessen (1997), Poskitt (2003) (the echelon form combined with the error correction form) and Takimoto and Hosoya (2004) (the Whittle likelihood approach) among others. One reason we prefer the frequency-domain approach to the time-domain counterpart is its mathematical and computational tractability when the framework is extended to a wider class of time series models such as a fractional cointegrating system. (See Hosoya (2004) and Shimotsu and Phillips (2005).) For another reason, simulation and empirical results given by Takimoto (2001) and Takimoto and Hosoya (2004) suggest that the frequency-domain method is superior to the time-domain counterpart in locating the maximum of the likelihood.

Suppose that a p-vector MA process $\{\xi(t)\}$ is generated by $\xi(t) = \sum_{l=0}^{b} \Theta(l)\varepsilon(t-l)$, where $\Theta(0) = I_p$, the $\varepsilon(t)$'s are independent random vectors with mean 0, covariance matrix Ω and $E[\{\varepsilon_i(t)\}^4] < \infty$, $(i = 1, \ldots, p)$, and a p-vector observation process $\{Z(t)\}$ is generated by

$$\triangle Z(t) = \Pi Z(t-1) + \sum_{k=1}^{a-1} \Gamma(k)\triangle Z(t-k) + \mu + 1\{\mathrm{rank}(\Pi) = p\}\nu t + \xi(t), \tag{1.1}$$

where $\triangle$ indicates $(1-L)I_p$ for the lag operator L, and $1\{\cdot\}$ is the indicator function. The model having the term $1\{\mathrm{rank}(\Pi) = p\}\nu t$ has the merit that it generates a trend-stationary process when the Π is full-rank, whereas, if the term is missing as in Johansen (1995), the full-rank stationary process does not generate trend effects. As the null hypothesis generates a unit-root process with a linear time trend, it would be more appropriate to allow a trend stationary component in the full-rank stationary hypothesis, for it to be a comparable alternative to the null hypothesis. On the basis of the model (1.1) and the asymptotic theory of Hosoya and Takimoto (2003), the paper presents a feasible numerical procedure for evaluating the maximum Whittle likelihood (MWL) estimator and the likelihood-ratio statistics.

The stationarity and invertibility conditions of VARMA coefficients are usually presumed to hold in the literature on numerical estimation methods for stationary ARMA models, although computational outcomes may sometimes violate those conditions unless certain restrictions are imposed on it. Neither Hannan and Rissanen (1982)'s three step estimation method for stationary ARMA models nor Johansen's likelihood ratio method based on the concentrated likelihoods is equipped with a built-in mechanism automatically producing estimates satisfying those conditions. Johansen's derivation of the concentrated likelihood, for example, conceals that the estimated AR coefficients implicitly obtained by the projections of $\triangle Z(t)$ and $Z(t-1)$ on $\triangle Z(t-1), \ldots, \triangle Z(t-a+1)$ may not satisfy those conditions. In contrast, our method uses an iterative method, where in each step the root conditions of the ARMA coefficient estimate are explicitly examined; that is, it examines whether all values of z satisfying $\det A(z) = 0$ and $\det B(z) = 0$ lie outside or on the unit circle exactly as specified by the respective hypotheses, where

$$A(z) = \triangle(z) - \Pi z - \sum_{j=1}^{a-1} \Gamma(j) \triangle(z) z^j \quad \text{and} \quad B(z) = \sum_{l=0}^{b} \Theta(l) z^l.$$

By means of the Jordan canonical form, the eigenvalues of companion matrices of the AR and the MA parts for the stationary component of the estimated cointegration model are evaluated and they are adjusted, if necessary, so as to fulfill the stationarity and invertibility conditions in the reduced-rank set-up of the cointegration model. Thanks to this eigenvalue contraction procedure, we can test a statistical hypothesis against others, in which the coefficient estimates belong to the respective proper parameter spaces.

Based on Monte Carlo simulations, Section 5 reports rejection probabilities by the rank test of the hypothesis $\text{rank}(\Pi) = 0$ against a full-rank model, where our root-modification procedure is observed useful to guarantee the valid root requirements automatically. For numerical illustration, our statistical method is applied in Section 6 to a trivariate U.S. short-term interest-rates series data and compared to the previous results in Takimoto and Hosoya (2004) which assumed a constant-mean stationary process when $\text{rank}(\Pi) = 3$. The empirical analysis concludes that the trend-stationary process is the appropriate modelling in contrast to the previous result which claimed that the series had at least one cointegration rank. (See also Reinsel and Ahn (1992) and Yap and Reinsel (1995).)

The paper is organized as follows: Section 2 presents our root modification procedure. Section 3 provides a modified three-step procedure in optimizing the Whittle likelihood. In Section 4, we present a cointegrating rank testing procedure which is essentially based on the Whittle log-likelihood ratio and give a computational algorithm for evaluating the p-value for this cointegration rank test. Section 5 conducts Monte Carlo simulations in order to examine the small-sample performance of the proposed procedure for a simple case. Section 6 analyses the trivariate U.S. interest-rate series by our test procedure.

As for the notations used in the paper: A' denotes the ordinary transpose of a matrix A, whereas A^* is the conjugate transpose; all vectors are assumed to be column vectors; the identity matrix of order p is denoted by I_p.

2. The root conditions and a root-contraction procedure

In this section, we propose an eigenvalue contraction algorithm which guarantees a set of ARMA coefficient estimates to satisfy the stationarity and invertibility conditions. Suppose that a p-vector process $\{Z(t), t = 1, \ldots, T\}$ is generated by (1.1), where the initial values $Z(-a+1), \ldots, Z(0)$ are assumed to be observed. Let β and $(\beta, \beta_\perp)$ be $p \times r$ and $p \times p$ matrices such that $\beta \perp \beta_\perp$ and $\mathrm{rank}(\beta, \beta_\perp) = p$. We focus on the model of cointegration rank r. As Hosoya (2003, pp. 38–39) showed, we have the relationship:

$$A(z)[\beta, \beta_\perp] = \left[A(z)\beta, \beta_\perp - \sum_{k=1}^{a-1} \Gamma(k) z^k \beta_\perp\right] \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & (1-z)I_{p-r} \end{array}\right]$$

$$\equiv C(z) \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & (1-z)I_{p-r} \end{array}\right]. \tag{2.1}$$

Therefore, all the roots of $\det C(z) = 0$ are outside the unit circle if and only if $(p - r)$ roots of $\det A(z) = 0$ are ones and the rest of them are outside the unit circle.

In practical estimation circumstances, it is difficult to specify the characteristic roots of $\det A(z) = 0$ derived from an estimate of $A(z)$ so that r of them are unities and $(p - r)$ are outside of the unit circle. For computational tractability, we propose a root modification method based on $C(z)$, since under the rank r hypothesis, it is equivalent to that all of the roots exceed unity in modulus. So we define the root conditions as follows:

Definition (the root conditions). Under the cointegration rank r hypothesis, the characteristic polynomial satisfies the condition that $\det C(z) = 0$, only if $|z| > 1$. Moreover all the roots of $\det B(z) = 0$ are assumed to be outside the unit circle and do not share any common zero with $\det A(z)$.

In the sequel, a set of ARMA coefficients is said to be admissible if it satisfies the root conditions.

The remainder of this section exhibits a root-contraction procedure based on the Jordan representation, which is implemented in Steps 2 and 3 to generate an admissible set of estimators in our three-step algorithm. To be explicit, the matrix $C(z)$ in (2.1) is represented as

$$C(z) = \begin{cases} [\beta, \beta_\perp] - [(I_p + \Pi)\beta, 0]z, & \text{if} \quad a = 1, \\ [\beta, \beta_\perp] - [(I_p + \Pi + \Gamma(1))\beta, \Gamma(1)\beta_\perp]z & \\ \quad - 1\{a \geq 3\} \sum_{k=2}^{a-1} [(\Gamma(k) - \Gamma(k-1))\beta, \Gamma(k)\beta_\perp] z^k & \\ \quad + [\Gamma(a-1)\beta, 0]z^a \quad \text{if} \quad a \geq 2 & \end{cases}$$

$$\equiv \sum_{k=0}^{a} C_k z^k, \tag{2.2}$$

where the bracketed matrix pairs consist of $p \times r$ and $p \times (p-r)$ matrices. Suppose at first that $a \geq 2$, and set $D_a \equiv \Gamma(a-1)\beta$ and define by J_r the $r \times p$ matrix whose (i, j) component is 1 if $i = j$ and 0 otherwise. The companion matrix for the generating mechanism $C(z)$ is defined as the following square matrix of size $(a-1)p+r \equiv q$

$$D \equiv \begin{bmatrix} -C_0^{-1}C_1 & -C_0^{-1}C_2 & \cdots & -C_0^{-1}C_{a-1} & -C_0^{-1}D_a \\ I_p & 0 & \cdots & \cdots & 0 \\ 0 & I_p & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_r & 0 \end{bmatrix}. \tag{2.3}$$

Let the v_i and ω_i, $i = 1, \ldots, q$, be the eigenvectors and eigenvalues of D respectively, and let $v_i^{(j)}$ be the partitions of the eigenvector v_i such that the $v_i^{(j)}$ are p-vectors for $1 \le j \le a-1$ and $v_i^{(a)}$ is a r-vector, and hence $v_i = (v_i^{(1)\prime}, \ldots, v_i^{(a)\prime})'$. In view of the form of D, the characteristic equation

$$Dv_i = \omega_i v_i \tag{2.4}$$

implies that for each i we have, as long as $|\omega_i| \neq 0$,

$$\left(I + C_0^{-1} \sum_{k=1}^{a-1} C_j \omega_i^{-k} + C_0^{-1} D_a J_r \omega_i^{-a} \right) v_i^{(1)} = 0,$$
$$v_i^{(j)} = \omega_i^{-j+1} v_i^{(1)}, \qquad j = 2, \ldots, a-1,$$
$$v_i^{(a)} = \omega_i^{-a+1} J_r v_i^{(1)}.$$

It also follows from (2.4) that

$$D = [v_1, \ldots, v_q] \mathrm{diag}(\omega_1, \ldots, \omega_q) [v_1, \ldots, v_q]^{-1}.$$

(In this paper, we assume without much loss of generality that all the non-zero eigenvalues $v_1, \ldots, v_q$ are distinct.)

The characteristic roots of $\det C(z) = 0$ are nothing but the inverse of the eigenvalues of the matrix D, so that the zero condition of $C(z)$ is equivalent to that all ω_i's are inside the unit circle. (See Miller (1968), p. 37.) We would sometimes encounter such cases as some ω_j's for D in conventional estimation procedures of (1.1) fall on or outside the unit circle, violating the assumption of the posited hypothesis whose rank of Π is equal to r. This would be the case in particular if a data set is generated by a process whose cointegration rank is less than r. For example, the C_j's induced from the unrestrictive ML estimate might quite possibly produce such an eigenvalue set. For testing cointegration rank r, $r = 0, 1, \ldots, p-1$ against rank p based on LR statistics, we have to prepare all of the likelihoods, where each likelihood should be estimated with a specified rank of Π. Our root-modification procedure aims at keeping the respective hypothesis in concern valid. A way of modification of the model parameters to secure the condition of stationarity is given as follows. If the ω_i, $i = 1 \ldots, s$, are eigenvalues such that $|\omega_i| > 1 - \varepsilon_1$ for a suitably chosen small number ε_1 and $|\omega_1| > |\omega_2| > \cdots > |\omega_s|$, contract them to $\omega_1^\dagger, \ldots, \omega_s^\dagger$ so that $|\omega_1^\dagger| > |\omega_2^\dagger| > \cdots > |\omega_s^\dagger|$, $|\omega_{s+1}| \le$

$|\omega_i^\dagger| \le 1 - \varepsilon_1$ and $\arg(\omega_i^\dagger) = \arg(\omega_i)$, $i = 1, \ldots, s$, where ω_{s+1} is the eigenvalue such that $|\omega_{s+1}| = \max_{s+1 \le j \le q} |\omega_j|$. A way to do this is to set

$$\omega_i^\dagger \equiv \frac{\omega_i}{|\omega_i|} \left\{ |\omega_{s+1}| + \frac{(s-i+1)(1-\varepsilon_1 - |\omega_{s+1}|)}{s} \right\}. \tag{2.5}$$

We contract only the roots on or outside the $(1-\varepsilon_1)$ circle for the correction to be minimal. Using those modified $\omega_i^\dagger$'s, define $D^\dagger$ by

$$\begin{aligned} D^\dagger \equiv{}& [v_1^\dagger, \ldots, v_s^\dagger, v_{s+1}, \ldots, v_q] \operatorname{diag}(\omega_1^\dagger, \ldots, \omega_s^\dagger, \omega_{s+1}, \ldots, \omega_q) \\ &\cdot [v_1^\dagger, \ldots, v_s^\dagger, v_{s+1}, \ldots, v_q]^{-1}, \end{aligned}$$

where the new eigenvectors $v_i^\dagger$, $i = 1, \ldots, s$, are given by

$$\begin{aligned} &v_i^{(1)\dagger} = v_i^{(1)}; \qquad v_i^{(j)\dagger} = (\omega_i^\dagger)^{-j+1} v_i^{(1)}, \qquad j = 2, \ldots, a-1; \\ &v_i^{(a)\dagger} = (\omega_i^\dagger)^{-a+1} J_r v_i^{(1)}. \end{aligned}$$

By virtue of this construction, $D^\dagger$ is given as

$$D^\dagger = \begin{bmatrix} -C_0^{-1}C_1^\dagger & -C_0^{-1}C_2^\dagger & \cdots & -C_0^{-1}C_{a-1}^\dagger & -C_0^{-1}D_a^\dagger \\ I_p & 0 & \cdots & 0 & 0 \\ 0 & I_p & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & J_r & 0 \end{bmatrix}, \tag{2.6}$$

whose eigenvectors and eigenvalues are given by $v_1^\dagger, \ldots, v_s^\dagger, v_{s+1}, \ldots, v_q$ and $\omega_1^\dagger, \ldots, \omega_s^\dagger, \omega_{s+1}, \ldots, \omega_q$ respectively. Based on the elements of $D^\dagger$, we can produce the desirable AR coefficients whose characteristic roots are outside the unit circle.

In the case of $a = 1$, the companion matrix D is defined by $D = C_0^{-1}C_1$. If we denote the $r \times r$ upper left block of $C_0^{-1}C_1$ by $(C_0^{-1}C_1)^{(1,1)}$, the nonzero eigenvalues of D is given by the roots of $\det\{\omega I_r - (C_0^{-1}C_1)^{(1,1)}\} = 0$. Suppose that $\omega_1, \ldots, \omega_r$ are those non-zero eigenvalues (hence $\omega_{r+1} = \cdots = \omega_p = 0$) and that these are distinct. If the absolute values of the eigenvalues ω_i, $i = 1, \ldots, s$, are greater than $(1-\varepsilon_1)$, then define the contracted $\omega_i^\dagger$'s as in (2.5), so that the modified $D^\dagger$ is given by

$$D^\dagger = (v_1, \ldots, v_p) \operatorname{diag}(\omega_1^\dagger, \ldots, \omega_s^\dagger, \omega_{s+1}, \ldots, \omega_p)(v_1, \ldots, v_p)^{-1}.$$

The change of the coefficients C_i to $C_i^\dagger$ produces the new parameters $\alpha^\dagger$, $\Gamma(k)^\dagger$, $k = 1, \ldots, a-1$. Namely, setting $C_i^\dagger = [C_i^{(1)\dagger}, C_i^{(2)\dagger}]$ where $C_i^{(1)\dagger}$ and $C_i^{(2)\dagger}$ are $p \times r$ and $p \times (p-r)$ matrices, we have for given β and $\beta_\perp$,

$$\begin{aligned} &\Gamma(a-1)^\dagger = [C_a^{(1)\dagger}, C_{a-1}^{(2)\dagger}](\beta, -\beta_\perp)^{-1}; \\ &\Gamma(k-1)^\dagger = [C_k^{(1)\dagger} + \Gamma(k)^\dagger \beta, C_{k-1}^{(2)\dagger}][\beta, -\beta_\perp]^{-1} \qquad \text{for} \quad 2 \le k \le a-1; \\ &\alpha^\dagger = -(C_1^{(1)\dagger} - \beta - \Gamma(1)^\dagger \beta)(\beta'\beta)^{-1}. \end{aligned} \tag{2.7}$$

Thus we have a new set of AR coefficients whose Jordan-form eigenvalues are inside the unit circle.

The validity of the root condition for the MA coefficient estimate $B(z)$ is also examined and if some eigenvalues of the Jordan form are outside the unit circle, we modify them in a parallel way so that all of them are inside the unit circle. For MA coefficients, there is another modification approach. See Remark 2.2.

Remark 2.1. Johansen's algorithm for LR is based on solving an eigen-equation derived from the concentrated likelihood in which the parameters $\Gamma(j)$'s are eliminated by maximization. The point is that the procedure has no built-in mechanism for the estimated $\Gamma(j)$'s to satisfy the root conditions. In contrast, our numerical method can keep number of unit roots exactly the same as the one specified by a hypothesis in concern. The case may be illustrated by the Dickey-Fuller test for a scalar-value process

$$\triangle x(t) = ax(t-1) + \varepsilon(t), \qquad t = 1, \ldots, T,$$

where $\varepsilon(t)$ is a Gaussian white noise. For testing $a = 0$ against $-2 < a < 0$. Suppose the test is conducted as is commonly done by the t-statistic $\sqrt{T}\hat{a}/\{\sum_{t=1}^{T} x(t-1)^2\}^{1/2}$, where $\hat{a}$ is "uncontrolled" OLS estimate of a. The test statistic is not proper for $\hat{a} > 0$ if we test the random-walkness against stationarity. But the case is automatically ignored in one-side test in which the null hypothesis is rejected if $\hat{a} < -c$ for some positive c. But if we apply F test (or the LR test) which corresponds to Johansen's test, we face difficulty since the F statistic in general is not "signed" as the t-statistic. A large F may be due to large positive $\hat{a}$ which is not an evidence favourable to the stationarity alternative at all. Our algorithm searches the ML estimate $\hat{a}$ in the range $-\infty < \hat{a} \leq -\varepsilon_1$.

Remark 2.2. Another, computationally simpler, method of contracting the eigenvalues is the contraction of the coefficients C_i to $C_i^\dagger \equiv \lambda^i C_i$, $i = 1, \ldots, a$ for $\lambda = (1+\varepsilon)^{-1} \max_{1 \leq j \leq q} |\omega_j|$. Since then $C^\dagger(z) = \sum_{i=0}^{a} C_i(\lambda z)^i$, the roots of $\det C^\dagger(z) = 0$ are constituted of $\lambda^{-1}\omega_1, \ldots, \lambda^{-1}\omega_q$ so that all the roots are on or inside the circle $\{z : |z| \leq (1+\varepsilon)^{-1}\}$. As for the MA coefficients, they may be modified by the canonical factorization of the spectral density which has the advantage of keeping the spectral density matrix invariant under the modification. (See Hosoya (1997).) Those approaches are not pursued in this paper.

3. Three-step estimation procedure

This section aims at implementing the procedure given by Takimoto and Hosoya (2004) with a mechanism which guarantees the estimated coefficients in Steps 2 and 3 to satisfy the root conditions automatically. Our modified three-step algorithm is organized as this: (1) Step 1 produces a consistent initial estimate of β and residuals; (2) Step 2 estimates the ARMA coefficients for each pair of ARMA orders (a, b) by substituting the disturbances with the residuals produced in Step 1. If the root conditions of the estimated model are violated, the estimated set of ARMA coefficients is modified so as to be located inside the admissible set by the root-contraction method presented in Section 2; (3) For each pair of lag orders, Step 3 sets the estimated parameters in Step 2 as the initial values for the maximizing iteration. In order to maximize the Whittle likelihood, our method uses an optimality algorithm in which penalty functions are added to the log likelihood in order to keep the ARMA estimate satisfying the root conditions. We identify the ARMA orders in the final stage by means of the BIC, producing the final selection of the pair $(\hat{a}, \hat{b})$.

Remark 3.1. The consistency of lag orders and parameters involved in the model, which is required in deriving the asymptotic distribution of the WLR statistic, is guaranteed by our way of application of information criteria. Namely, we use the AIC in Step 1 for approximating the virtually infinite-order AR model by a truncated version and the BIC in Step 3 for selection of the finite lag orders of ARMA model. For stationary ARMA series, Theorem 3 of Hannan and Rissanen (1982) showed that, under a set of regularity conditions, if the maximal lag order within which selection is made in Step 1 increases monotonically to infinity by a rate faster than $\log T$ but not faster than $(\log T)^b$ for some $b > 1$, their three-step method based on the BIC produces $\sqrt{T}$-consistent estimators of the model parameters. The OLS estimator of β is T-consistent as Stock (1987) showed.

To be more specific, our three-step procedure is conducted as follows:

Step 1 (Estimation of the disturbance series). The purpose of this step is to produce an estimate of the unobserved innovation sequence by means of a consistent initial estimate of β. We set $\beta' = [I_r, \beta_0]$ for $\Pi = \alpha\beta'$ as in Ahn and Reinsel (1990), where β_0 is an $r \times (p - r)$ matrix of unspecified parameters. A consistent initial estimate $\hat{\beta}$ of β can be obtained by the OLS procedure thanks to its superconsistency. If Π has full rank, we set $\hat{\beta} = I_p$ for which the data generation process is reduced to a stationary VARMA model. We set the fitted AR order as n, and denoting by $\hat{\varepsilon}_n(t)$ the orthogonal projection residual of $Z(t)$ onto the linear span of $\hat{\beta}Z(t-1), \triangle Z(t-1), \ldots, \triangle Z(t-n+1), 1$ and $1\{\text{rank}(\Pi) = p\}t$, set $\hat{\Omega} = 1/(T-n)\sum_{t=n+1}^{T} \hat{\varepsilon}_n(t)\hat{\varepsilon}_n'(t)$. We then apply the information criterion

$$AIC(n) = \log\det\hat{\Omega} + \frac{2p((n-1)p + r + 1 + 1\{\text{rank}(\Pi) = p\})}{T}.$$

The order n which minimizes $AIC(n)$ is selected as the order estimate $\hat{n}$, and the corresponding residual sequence $\{\hat{\varepsilon}(t), t = \hat{n} + 1, \ldots, T\}$ is obtained.

Remark 3.2. Another identifying restriction is provided by Johansen (1991) who uses the normalization $\hat{\alpha}' R\hat{\alpha} = I_n$, where R is the sample moment matrix of residuals. For a relationship between those two identifying restrictions, see Watson (1994), p. 2891.

Step 2 (Estimation and its modification). In this step, based on Hosoya's representation theorem (Hosoya (2003, Theorem 2.1)), the AR part of a cointegration model is separated into the unit-root component and the stationary component so that we can focus only on the latter component. The stationarity component is expressed in the Jordan form in which the eigenvalues exceeding unity in modulus are contracted so as to be inside the unit circle.

- Step 2.1. The OLS coefficient estimate is obtained by the regression of $\triangle Z(t)$ on $\hat{\beta}'Z(t-1), \triangle Z(t-1), \ldots, \triangle Z(t-a+1), 1, 1\{\text{rank}(\Pi) = p\}t, \hat{\varepsilon}(t-1), \ldots, \hat{\varepsilon}(t-b)$.

It is not necessarily the case that the coefficient estimate obtained in Step 2.1 belongs to the admissible set. Under H_r : rank$(\Pi) = r$, for $0 \leq r < p$, the roots of $\det A(z) = 0$ must consist of r unit roots and $(p-r)$ roots which lie outside the unit circle. But it is not guaranteed by Step 2.1 estimation, since no restriction is imposed in the estimation procedure. The ARMA coefficient modification procedure proceeds as follows:

- Step 2.2. Express the cointegrated AR coefficients $A(z)$ in terms of $C(z)$ given in (2.1).
- Step 2.3. Express coefficients of $C(z)$ in the Jordan form representation as in (2.3).
- Step 2.4. Modify the eigenvalues so that all the values fall inside the unit circle as in (2.5).
- Step 2.5. Produce new $C(z)$ based on the modified Jordan form as in (2.6).
- Step 2.6. Produce new $A(z)$ from new $C(z)$ as in (2.7).
- Step 2.7. Express MA coefficients in the Jordan form.
- Step 2.8. Modify the eigenvalues of the Jordan form and produce new MA coefficients.

Remark 3.3. We assume that each component of $\{Z(t)\}$ is an $I(1)$ process. If, under the hypothesis of non-cointegrated $I(1)$, we need to modify the estimated coefficients in Step 2.1, it indicates the possibility that $\{Z(t)\}$ may be an $I(2)$ process. In such a situation, we may require estimation and testing procedures which allow for the integration order bigger than one. This paper does not go into that issue.

By the root-contraction method of Steps 2.2 to 2.8, the coefficient estimate of the ARMA model is pulled into the admissible set and we use it as the initial value for the optimization iteration in Step 3.

Step 3 (MWL estimation). Step 3 consists of a quasi-Newton maximization of the Whittle likelihood function augmented by penalty functions. The procedure proposed in this paper differs essentially from Takimoto and Hosoya (2004) in that two penalty functions are added to the objective function so that the maximizer of the argumented objective function automatically satisfies the root conditions. The (possibly modified) coefficient estimate in Step 2 is used as the initial value for the iteration of this step. Denote by $f_{\xi\xi}$ the spectral density of the MA process $\xi(t)=\sum_{l=0}^{b}\Theta(l)\varepsilon(t-l)$; namely

$$f_{\xi\xi}(\lambda)=\frac{1}{2\pi}\left\{\sum_{l=0}^{b}\Theta(l)e^{il\lambda}\right\}\Omega\left\{\sum_{l=0}^{b}\Theta(l)e^{il\lambda}\right\}^{*}.$$

Let $W_1^{(1)}(\lambda)$ and $W_2^{(1)}(\lambda)$ be p and $(r+p\times(a-1)+s)$-vector functions respectively defined on $[-\pi,\pi]$ by

$$\begin{aligned}W_1^{(1)}(\lambda)&=(2\pi(T-c+1))^{-1/2}\sum_{t=c}^{T}e^{it\lambda}\triangle Z(t),\\ W_2^{(1)}(\lambda)&=(2\pi(T-c+1))^{-1/2}\sum_{t=c}^{T}e^{it\lambda}\{(\hat{\beta}'Z(t-1))',\triangle Z'(t-1),\ldots,\\ &\qquad\triangle Z'(t-a+1),1,1\{\text{rank}(\Pi)=p\}t\}',\end{aligned}$$

where $c=\max(\hat{n}+a+1,\hat{n}+b+1)$; $s=1$ for H_r, $0\le r\le p-1$, and $s=2$ for H_p in view of the additional trend term under the full rank hypothesis. Let $Q_1(\psi^{(1)},\Omega^{(1)})$ be the real-valued function defined by

$$\begin{aligned}Q_1(\psi^{(1)},\Omega^{(1)})\;&=\;\log\det\Omega^{(1)}+\frac{1}{2\pi}\int_{-\pi}^{\pi}\{W_1^{(1)}(\lambda)-(W_2^{(1)}(\lambda)'\otimes I_p)\psi^{(1)}\}^{*}\\ &\qquad\cdot f_{\xi\xi}(\lambda)^{-1}\{W_1^{(1)}(\lambda)-(W_2^{(1)}(\lambda)'\otimes I_p)\psi^{(1)}\}d\lambda\end{aligned}$$

where $\psi^{(1)}=(\text{vec}(\alpha)',\text{vec}(\Gamma(1))',\ldots,\text{vec}(\Gamma(a-1))',\mu',(\nu 1\{\text{rank}(\Pi)=p\})')'$ is a $p\times(r+p\times(a-1)+s)$-vector. Setting

$$LR(r)=-(T-c+1)/2\times Q_1(\psi^{(1)},\Omega^{(1)}). \tag{3.1}$$

We call it Type 1 Whittle likelihood (L1), whereas Type 2 Whittle likelihood (L2) uses the full sample data, exploiting the assumption that $\triangle Z(t)$ and $\beta'Z(t-1)$ are stationary under H_r. Set

$$\begin{aligned}W_1^{(2)}(\lambda)&=(2\pi T)^{-1/2}\left\{I_p-\sum_{j=1}^{a-1}\Gamma(j)e^{-ij\lambda}\right\}\sum_{t=1}^{T}e^{it\lambda}\triangle Z(t),\\ W_2^{(2)}(\lambda)&=(2\pi T)^{-1/2}\sum_{t=1}^{T}e^{it\lambda}\{(\hat{\beta}'Z(t-1))',1,1\{\text{rank}(\Pi)=p\}t\}',\\ \psi^{(2)}&=(\text{vec}(\alpha)',\mu',(\nu 1\{\text{rank}(\Pi)=p\})')',\end{aligned}$$

where $W_1^{(2)}(\lambda)$, $W_2^{(2)}(\lambda)$ and $\psi^{(2)}$ are $p,(r+s)$ and $p(r+s)$-vector functions respectively. Type 2 Whittle likelihood is then defined by

$$LR(r)=-T/2\times Q_2(\psi^{(2)},\Omega^{(2)}) \tag{3.2}$$

where Q_2 is defined by replacing $W_1^{(1)},W_2^{(1)},\psi^{(1)},\Omega^{(1)}$ in Q_1 by $W_1^{(2)},W_2^{(2)}$, $\psi^{(2)},\Omega^{(2)}$. Setting $\Theta(\lambda)=\sum_{l=0}^{b}\Theta(l)e^{i\lambda l}$, the minimizer $\Omega^{(i)}$ of Q_i for $i=1,2$ is given by

$$\hat{\Omega}^{(i)} = \int_{-\pi}^{\pi} Re[\{\Theta(\lambda)^{-1}W_1^{(i)}(\lambda) - (W_2^{(i)}(\lambda)' \otimes \Theta(\lambda)^{-1})\psi^{(i)}\} \cdot \{\Theta(\lambda)^{-1}W_1^{(i)}(\lambda) - (W_2^{(i)}(\lambda)' \otimes \Theta(\lambda)^{-1})\psi^{(i)}\}^*]d\lambda,$$

whence we have the relation

$$Q_i(\psi^{(i)}, \hat{\Omega}^{(i)}) = \log\det\hat{\Omega}^{(i)} + p.$$

Setting the coefficient estimate obtained in Step 2 as the initial value, we solve the minimizing problem of $\log\det\hat{\Omega}$ for the parameter vector ψ by means of a quasi-Newton method. For the likelihood to be maximized in the time-domain representation, it is necessary to evaluate residual sequences in each iteration step to estimate the covariance matrix Ω. In contrast, the Whittle approach reduces the computational amount by omitting the exploitation of residuals. The root modification method in Step 2 places the initial estimate in the admissible set. To keep the coefficient estimate remaining inside the admissible set during the optimization iteration, we add two penalty terms to the objective function $\log\det\hat{\Omega}$. For a suitably small positive ε_2 and a suitably fixed positive value d, define a new objective function by

$$\log\det\hat{\Omega} + 1\{\max|\omega_{i_1}| > 1-\varepsilon_2\}de^{-1/u_1} + 1\{\max|\omega_{i_2}| > 1-\varepsilon_2\}de^{-1/u_2}, \tag{3.3}$$

where ω_{i_1} and ω_{i_2} are eigenvalues of the companion matrix for the AR and MA parts respectively, $i_1 = 1, \ldots, (a-1)p + r$ and $i_2 = 1, \ldots, bp$, and $u_j = \max|\omega_{i_j}| - (1-\varepsilon_2)$. Namely, the u_j expresses the distance between the maximal eigenvalue and unity. Therefore, the further $u_j > 0$ is away from zero, the heavier the penalty e^{-1/u_j} becomes. If u_j is negative, all of the eigenvalues are inside the unit circle and these penalty terms do not work. The constant d should be chosen so that $\log\det\Omega$ and the penalties are comparable magnitudes.

After obtaining the MWL estimates for each combination of the ARMA orders (a, b), we select the lag structure of the ARMA model by means of the BIC, which is given by

$$BIC(a,b) = \log\det\hat{\Omega} + p\cdot\log\frac{1}{2\pi} + p + \frac{p(p(a-1) + bp + r + 1 + 1\{\text{rank}(\Pi) = p\})\log T}{T},$$

for $a \le \bar{a}$ and $b \le \bar{b}$, where $\bar{a}$ and $\bar{b}$ are chosen beforehand to be sufficiently large.

4. Testing procedure

This section explains the testing procedure for the rank r hypothesis against the rank p hypothesis based on the Whittle likelihood, where the full-rank alternative is supposed to be given by the stationary model involving the linear time trend.

The LR statistic for cointegrated ARMA models is not reducible, in finite-sample circumstances, to the eigenvalue problem as Johansen (1988, 1995) posed for VAR models, but rather a direct computation of the WLR statistics is required. For testing against the cointegrating rank p, we evaluate the p-value based on the asymptotic distribution given in Corollary 5.1 of Hosoya and Takimoto (2003) which gives the asymptotic distribution of the WLR statistics in case the full-rank alternative explicitly involves a linear time trend.

Remark 4.1. The BIC applied to different cointegration-rank models would possibly choose different ARMA orders. For convenience of inferential purposes, we set ARMA lag orders for the hypotheses of $r = 0, \ldots, p-1$ to equal to the one for the full-rank trend-stationary model. Namely, only in estimating the full-rank hypothesis, the lag structure is identified by the information criterion BIC after evaluating the likelihoods for each combination of the lag orders and for other hypotheses of smaller ranks, we simply apply that set of lag orders.

Suppose that the null hypothesis of cointegration rank r $(0 \le r < p)$ is given by

$$H_r : \Pi = \alpha\beta',$$

whereas the alternative hypothesis of cointegration rank p is given by

$$H_p : \Pi = \alpha\beta' + \alpha^{(p-r)}\beta_{\perp},$$

where α and β are $p \times r$ full rank matrices, and $\alpha^{(p-r)}$ is a full rank $p \times (p-r)$ matrix and $\beta_{\perp}$ is a $p \times (p-r)$ matrix which is a full rank $p \times (p-r)$ matrix orthogonal to β. Define the log-WLR statistic by $LR(r,p) = 2\{LR(p) - LR(r)\}$, where $LR(\cdot)$ is defined by either (3.1) or (3.2).

Given a standard Brownian motion $\{B(u, I_{p-r}), 0 \le u \le 1\}$, define

$$G(u) = \begin{bmatrix} B(u, I_{p-r}) - \bar{B} \\ u - 1/2 \end{bmatrix},$$

where $G(u)$ is the $(p-r+1)$-vector process and $\bar{B} = \int_0^1 B(u, I_{p-r})du$. The log-WLR statistic $LR(r,p)$ for testing the rank r against the rank p is asymptotically distributed as

$$LR_A(r,p) \equiv \mathrm{tr}\Big[\int_0^1 dB(u, I_{p-r})G(u)'\Big\{\int_0^1 G(u)G(u)'du\Big\}^{-1} \cdot \int_0^1 G(u)dB(u, I_{p-r})'\Big]. \tag{4.1}$$

(See for those limit results Hosoya and Takimoto (2003). Note that (4.1) differs from what Johansen (1995) gave. In the latter, $G(u)$ is a $(p-r)$-vector process with $B(\cdot)-\bar{B}$ replaced by $B^{(p-r-1)}(\cdot)-\bar{B}^{(p-r-1)}$ which is a subprocess constituted of $B_k(\cdot, I_{p-r}), 1 \le k \le p-r-1$.)

For the purpose of evaluating p-value based on this asymptotic expressions, we use Monte Carlo simulation of the stochastic integrals in (4.1) which is conducted as follows:

- Step 1. By means of standard-normal random numbers, generate the trace statistic given in (4.1) by setting the number of partitions for numerical integration equal to 400 and the number of iterations equal to 100000.
- Step 2. Based on 100000 observed trace statistics, calculate p-value by counting the number of the simulated trace statistics exceeding the observed WLR.

Remark 4.2. For the limiting result (4.1) to hold, a weaker set of conditions for $\varepsilon(t)$ than the ones in (1.1) suffices; see Hosoya (2005).

Remark 4.3. In respect of discrete approximation of the stochastic integrals in (4.1), our trials by several combinations of the number of partition and iteration confirm that 400-step Gaussian random walks suffice to approximate stochastic integrals.

5. Small sample performance

To investigate the effects of deterministic trend under the full-rank alternative hypothesis, this section examines the rejection probability of the cointegration rank test of Section 4 in small-sample circumstances and the performance of the root-contraction mechanism in Steps 2 and 3 by means of Monte Carlo simulation. Many authors investigate the finite sample performance of various cointegrating rank tests intensively in the 1990s. They report that in some situation the small-sample properties of Johansen's trace statistics are at variance with what the asymptotic theory predicts and the asymptotic distributions tend to have large size distortions. In particular, Hubrich *et al.* (2001) reviews the literature on cointegration tests and presumptions for the asymptotics based on a unifying framework, and provide new simulation studies to evaluate the performance of rank tests for the same data generating processes (DGP). Furthermore, Johansen (2002) and Nielsen (2004) provide some correction procedures to improve the finite sample bias. But, in this paper, we focus on the finite-sample performance of our eigenvalue contraction procedure in a simple model and do not go into the correction issue.

The empirical sizes are known to be influenced by cointegrating vectors, covariances of innovation and the dimension of the model. In this paper, we deal with the case of a bivariate random walk with a drift as the DGP. Consider the following simple bivariate series $\{Z(t)\}$ generated by

$$\triangle Z(t) = \Pi Z(t-1) + \mu + 1\{\mathrm{rank}(\Pi) = 2\}\nu t + \varepsilon(t). \tag{5.1}$$

Suppose that the null hypothesis of cointegration rank zero is given by H_0 : $\mathrm{rank}(\Pi) = 0$, whereas the alternative hypothesis of rank two is given by H_{2t} : $\mathrm{rank}(\Pi) = 2$. Assume that the true generating process is a two-dimensional random walk with $\Pi = 0$, $\mu = (0.5, 0.5)'$ and the $\varepsilon(t)$'s are $i.i.d.N\ (0, I_2)$ in (5.1).

We are concerned with rejection probabilities of rank tests against the full-rank trend-stationary hypothesis and the effectiveness of our root-contraction procedure. For comparing with usual rank tests against the constant-mean stationary alternative hypothesis, we conduct another simulation: The two dimensional model is given by

$$\triangle Z(t) = \Pi Z(t-1) + \mu + \varepsilon(t).$$

Assume that the DGP is a two-dimensional random walk with a drift as in the previous setting. Under this model, the alternative hypothesis is denoted by H_{2c} : rank$(\Pi) = 2$ to discriminate it from H_{2t}. To produce the empirical size, the sample sizes are chosen to be $T = 50$ and 100 and generate 1000 sets of data series. In order to eliminate the effect of initial values in data generation, data from the 101th observation on are used.

We investigate the two WLR tests; namely, (i) H_0 vs H_{2c} and (ii) H_0 vs H_{2t}. Table 1 exhibits the quantiles of rejection probability, where rejection probabilities are observed to be roughly close to the nominal sizes expected from the asymptotic theory in each case, but it seems that empirical sizes against trend-stationary alternatives, H_{2t}, are slightly higher than those against constant-mean ones, H_{2c}. Table 2 shows how many times the root-contraction procedure is employed to let all of the characteristic roots stay outside the unit circle in Step 2 for obtaining initial values and in Step 3 for guaranteeing the root condition

Table 1. Rejection probabilities for rank(Π)=0.

H_0 vs H_{2c}	nominal 10% test	nominal 5% test	nominal 1% test
$T = 50$	0.117	0.062	0.013
$T = 100$	0.100	0.045	0.010
H_0 vs H_{2t}			
$T = 50$	0.118	0.058	0.013
$T = 100$	0.109	0.056	0.010

Note. H_{2c} and H_{2t} mean that the full-rank alternative hypotheses are reduced to a constant-mean and trend stationary models, respectively.

Table 2. Observed number of the root contraction.

Step	sample size	H_{2c}	H_{2t}
Step 2	50	344	14
	100	395	8
Step 3	50	112	14
	100	65	8

Note. The row of Step 2 denotes the observed number of the root-contraction procedure conducted for initial values and the row of Step 3 provide the average observed number of the root-contraction procedure conducted to keep parameter coefficients belonging to the admissible set in 1000 replication.

in the iterative optimization. It provides the observed number of root modification in Step 2 and the average observed number of root modification in Step 3 in 1000 replication. Under the rank zero hypothesis, it does not report the observed number, since no AR coefficients are contained in the model. In estimating the constant-mean stationary model by using 50 and 100 samples, the observed number in Step 2 are 344 and 395 respectively. In Step 3, the penalty term works 112 and 65 times in average to keep estimates to stay in the admissible set. On the other hand, estimates of the trend-stationary model belong to the admissible set in almost all replication. In 50 and 100 samples, 14 and 8 cases need to be contracted for initial values in 1000 replication. The observed numbers in Step 3 are reduced from 112 and 65 to 14 and 8 by containing trend term. Since estimates given by unrestricted estimation may not satisfy the root condition, the LR statistics constructed from these estimates may violate the hypothesis in concern. Characteristic roots derived of coefficient estimates by unrestricted OLS in Step 2 may not satisfy the root condition even for this simple bivariate random walk. Regardless of whether we use the time-domain or the frequency-domain approaches, to obtain the maximized likelihood satisfying the root condition, we need to start from valid initial values. Implementing our root-modification mechanism in Steps 2 and 3 makes the estimates conformable to the hypothesis in concern.

Also, it seems unreasonable to set a constant-mean stationary as the full-rank alternative hypothesis if the DGP has a drift, since we need then to contract characteristic roots more often for estimating coefficients. By including the trend term in the full rank model, the number of root modification is decreased drastically and the frequency increases for obtaining valid coefficient estimate without contraction mechanism.

6. An applied example

We apply our three-step procedure to a trivariate series of the U.S. interest-rates which are the Federal Fund rate (FF), 90-day Treasury Bill rate (GM3) and 1-year Treasury Bill rate (GT1) over the period January, 1960 through December, 1979. (All the three series were obtained from the citibase financial data base). These series were investigated by Stock and Watson (1988) in views of testing for common trends, by Reinsel and Ahn (1992) and Yap and Reinsel (1995) who fitted VAR(2) and VARMA$(1,1)$ models respectively, and also by Takimoto and Hosoya (2004) who fitted VARMA models based on the Whittle likelihood approach. All of those analyses concluded that there is at least one cointegrating relationship among those series in which the alternative hypothesis is posited as the full-rank constant-mean stationary model. But Figures 1 to 3 indicate that those economic series can not be constant-mean stationary processes; rather the mean depends on the time t.

We investigate the same series under the trend-stationary model and compare with the results given by the preceding literature. As in the above cited papers, we use the log-transformed data, which seem to behave more homogeneously over

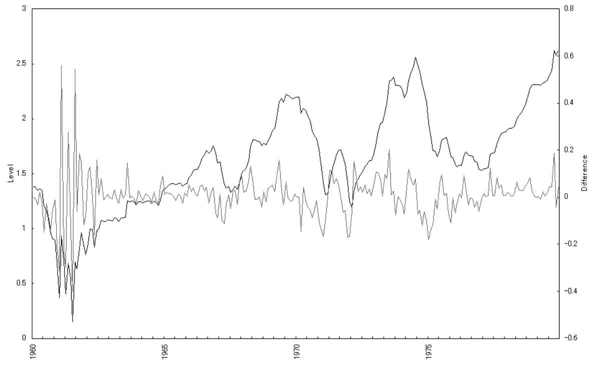

Figure 1. ln FF.

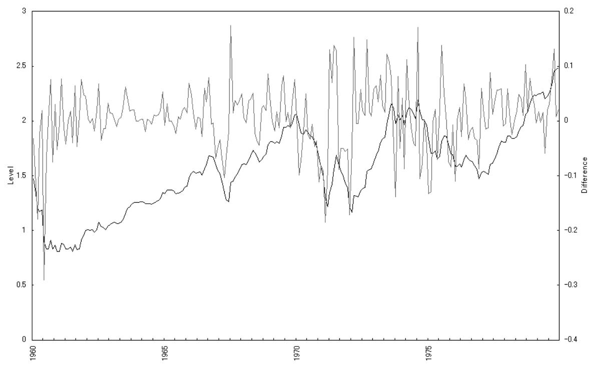

Figure 2. ln GM3.

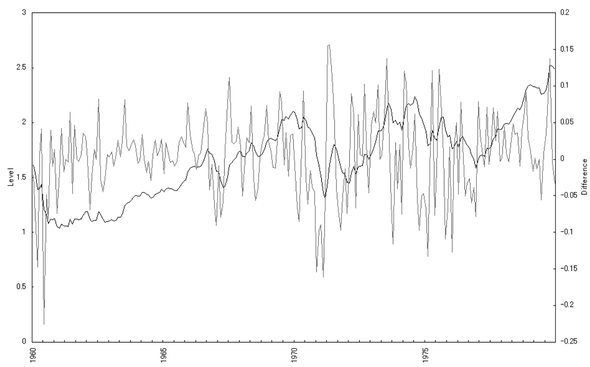

Figure 3. ln GT1.

the observation period than the original series. All series are plotted in levels and differences in Figures 1 through 3. For determining the lag structure of ARMA, we use 220 observations over the period from September, 1961 to December, 1979, because the 20 lags are chosen by the AIC in Step 1 and they need to be kept as initial values. But after choice of the order $(2, 0)$, in order to exploit as much observations as possible, we use the original 240 in the subsequent estimation procedure, since the selected ARMA$(2, 0)$ model involves no MA terms. Because of high volatility in the beginning part of the observation period, we use Type 1 Whittle likelihood in the following analysis.

Before conducting the three-step procedure we must investigate the influence of some constant values, ε_1, ε_2 and d, on the objective function given in Step 3. For numerically maximizing the Whittle likelihood, we set ε_1 equal to 0.05 for Step 2 and ε_2 equal to 0.02 for Step 3. (See (2.5) and (3.3).) The setting $\varepsilon_1 > \varepsilon_2$ seems to work better, since this allows eigenvalues of the third-step estimates to be located nearer to the unit circle. For the penalty terms to work effectively, we set d equal to 10^{30}. The values of u_j for $j = 1, 2$ in the penalty functions indicate the distance between $1 - \varepsilon_2$ and the maximum absolute value of eigenvalues of the Jordan matrix for the AR or MA parts. Table 3 illustrates the cases $u_j = 0.01$, 0.015, 0.02 and 0.03. It shows that if $u_j = 0.01$, that is, the maximum absolute eigenvalue is at most 0.99, influence of the penalty terms 3.7×10^{-14} is negligible, because the leading term of objective function $\log\det\hat{\Omega}$ for each combination of a and b is around -18. But if u_j is 0.015, the penalty term $10^{30}e^{-1/u_j}$ is 1.1×10, which is very large in comparison with $\log\det\hat{\Omega}$. This setting can prevent eigenvalues to be more than 0.995 in Step 3 iteration and it seems to be good enough for the interest-rate series for the purpose of optimization conducted inside the admissible set.

We determine the lag order of ARMA model by evaluating the BIC based on the Whittle likelihoods under the full-rank trend-stationary hypothesis. Maximum lag order $\bar{a}$ and $\bar{b}$ are set as 5 in this illustration. The computation involves ARMA$(5, 5)$ model as the largest possible model and this model has as many as 96 parameters under the alternative hypothesis. All the first derivatives of the objective function evaluated at the estimate does not necessarily vanish for some steps of iterations, but, as Huber (1967) and Pollard (1985) showed, for the asymptotic results of the maximum likelihood estimator to hold, it suffices for the first derivatives divided by the sample size T to go to zero as the sample size increases.

Table 3. Values of u_j and $10^{30}e^{-1/u_j}$.

u_j	$10^{30}e^{-1/u_j}$
0.01	3.7×10^{-14}
0.015	1.1×10
0.02	1.9×10^{8}
0.03	3.3×10^{15}

Before proceeding to Step 3, if necessary, the initial coefficients as for AR and MA parts are modified so as to belong to the admissible set, but, all of initial coefficients of AR and MA parts belong to the admissible set; each eigenvalue is less than unity in estimating full-rank model. In Step 2, eigen-value contraction algorithm was not employed in all of the lag-order combinations.

As regards the convergence criteria in optimization, the relative rates of changes for the objective function and variables and the absolute value of derivatives are set equal to 10^{-4}, 10^{-4} and 1, respectively. We set the absolute value of derivatives relatively large to reduce the computational cost. The observed maximum of those values is 0.195 and the rest ones are less than 10^{-2} and we may judge that the maximum of the Whittle likelihood is attained by our convergence conditions. Although for this data set it is not necessary to contract eigenvalues in Step 2, we observe that the maximum of eigenvalues is larger than unity in the optimization iteration in Step 3, and that penalty terms work well to keep the ARMA coefficient estimate inside the admissible set and thanks to this mechanism, our procedure avoids automatically iteration drift.

Based on the outcomes of Step 3 for each lag structure, the BIC selects the ARMA$(2,0)$ for the trend-stationary model among 30 combinations of the lag structure. Table 4 exhibits the BIC values for AR order from one to three and MA order from zero to three. The BIC is minimized for the ARMA$(2,0)$ model and the following analysis is based on the ARMA$(2,0)$ model.

For testing cointegration rank, in addition to the full-rank likelihood, we must evaluate the Whittle likelihood for ARMA$(2,0)$ model with reduced ranks 0 to 2. We examine how many times the initial coefficients obtained in Step 2 are modified and the penalty terms work to keep coefficients inside the admissible set in Step 3. In all of the models the set of initial coefficient estimates by unrestricted OLS procedure is admissible; any modification of estimates are not observed.

As for the root-contraction procedure in Step 3, except for the rank two

Table 4. The BICs after Step 3 under H_3.

lag order	BIC
$(1,0)$	-21.25
$(1,1)$	-21.32
$(1,2)$	-21.20
$(1,3)$	-21.08
$(2,0)$	-21.45
$(2,1)$	-21.28
$(2,2)$	-21.12
$(2,3)$	-21.00
$(3,0)$	-21.42
$(3,1)$	-21.21
$(3,2)$	-20.99
$(3,3)$	-20.81

model, the eigenvalues evaluated in the optimizing iteration did not exceed unity and the penalty was not employed. But in the rank two model, the observed number of times on which penalty was enforced is 3. In estimating the models with ranks 0 to 3, the number of direction searched are 1, 89, 149 and 1, respectively. The likelihoods of the rank 1 and 2 model, where the initial likelihoods are improved in the process of iteration, gain from -20.517 and -20.552 to -20.535 and -20.596, respectively. We may conclude that our eigen-value contraction algorithm provides pertinent estimates whose characteristic roots satisfy the root condition exactly.

The test results for the ranks 0 to 2 against the full-rank trend stationary alternative hypothesis are listed in Table 5. Our tests strongly reject all the nonstationary hypotheses and supports the trend-stationary hypothesis in contrast to the results reported in the literature. Yap and Reinsel (1995) and Takimoto and Hosoya (2004) dealing with the stationarity hypothesis with constant term, and Reinsel and Ahn (1992) dealing with the one with no constant term, support the possibility of one or two cointegrating relationships for these trivariate series; all of them support that the series involve at least one nonstationary component.

Consequently our finally identified model is given as:

$$\triangle\hat{Z}(t) = \begin{bmatrix} -0.259 & 0.335 & -0.004 \\ 0.049 & -0.114 & 0.015 \\ 0.051 & 0.013 & -0.149 \end{bmatrix} Z(t-1)$$

$$+ \begin{bmatrix} -0.053 & -0.015 & 0.345 \\ 0.064 & 0.021 & 0.275 \\ 0.039 & -0.014 & 0.360 \end{bmatrix} \triangle Z(t-1) + \begin{bmatrix} -0.039 \\ 0.050 \\ 0.105 \end{bmatrix} + \begin{bmatrix} 0.00006 \\ 0.0002 \\ 0.0004 \end{bmatrix} t;$$

$$\hat{\Omega} = \begin{bmatrix} 7.800\times10^{-3} & 2.543\times10^{-3} & 2.166\times10^{-3} \\ 2.543\times10^{-3} & 3.153\times10^{-3} & 2.462\times10^{-3} \\ 2.166\times10^{-3} & 2.462\times10^{-3} & 2.644\times10^{-3} \end{bmatrix}.$$

The graphs of residuals based on this model are given in Figures 4 to 6.

Table 5. The p-value of the rank test for U.S. interest-rates.

null	alt	$L(r,3)$
$r=0$	$j=3$	66.93
		$(0.00003)^{***}$
$r=1$	$j=3$	33.19
		$(0.00443)^{***}$
$r=2$	$j=3$	18.76
		$(0.00390)^{***}$

Note. *** denotes 1% significance level.

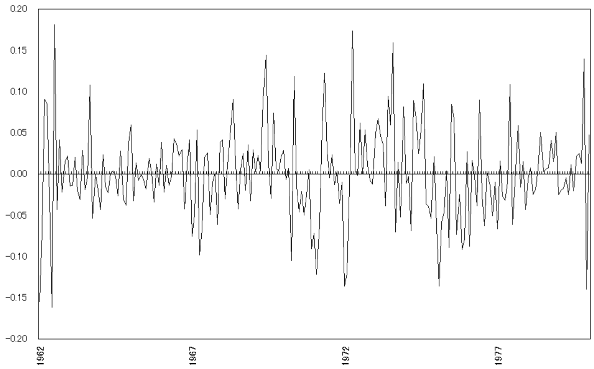

Figure 4. $\ln \mathrm{FF} - \ln \hat{\mathrm{FF}}$.

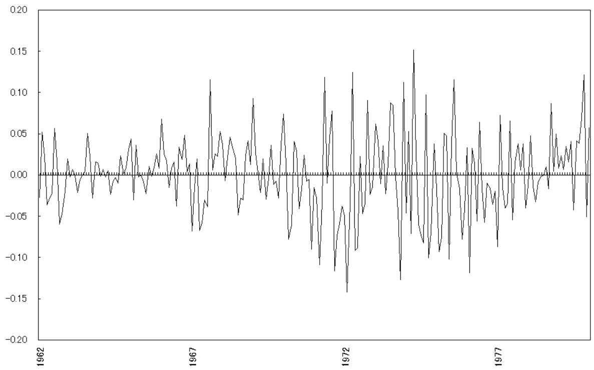

Figure 5. $\ln \mathrm{GM3} - \ln \hat{\mathrm{GM}}3$.

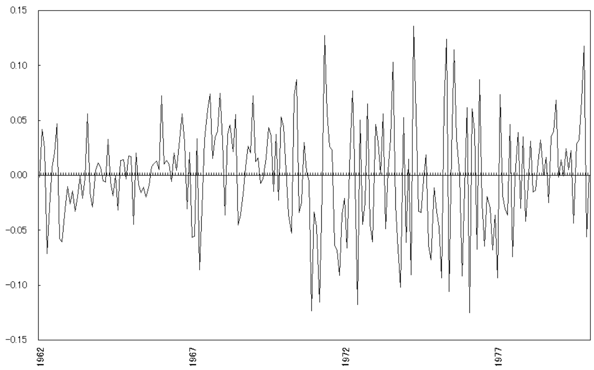

Figure 6. $\ln \mathrm{GT1} - \ln \hat{\mathrm{GT}}1$.

7. Concluding remarks

In this paper, we presented a detailed account of a new numerical procedure for evaluating the MWL estimates. One of the difficulties involved in the numerical evaluation in finite samples is that the maximum of the Whittle likelihood is not necessarily attained inside the admissible parameter set. Based on a representation theorem of cointegrated processes, we proposed an eigen-value contraction method in the case that the root condition is violated and also a numerical iteration procedure for maximization of the Whittle likelihood in each step of which the root condition is implemented by means of penalty functions so that the outcome in each stage of iteration is guaranteed to fall inside the admissible set.

For further development of our approach, we may need to take into account the heteroscedasticity of the error terms. Possible extensions to deal with this situation are to incorporate the Box-Cox transformation into our three-step procedure, to remodel the error process $\{\xi(t)\}$ as a non-linear process to accommodate conditional heteroscedasticity and/or to take structural breaks into account in the model (1.1). As for the literature of the related topics, Ling and McAleer (2003) proposed vector ARMA-GARCH models in the time-domain approach and Zaffaroni (2003) presented univariate nonlinear MA models in the frequency-domain one. (See also Robinson and Zaffaroni (1997, 1998).) In the time-domain method, moreover, by extending Ahn and Reinsel (1990), Li *et al.* (2001) investigated partially nonstationary multivariate AR model with conditional heteroscedasticity. Extension of our approach in those directions remains open.

Another direction of extension is to incorporate fractionally integrated process, whose integrated order can take fractional number. Based on the Whittle likelihood, Hosoya (2004) and Shimotsu and Phillips (2005) dealt with testing the fractional cointegration rank and semiparametric estimation of fractional cointegrating system respectively. By exploiting these approaches, we may extend the presented estimation and testing procedure to the fractional cointegrated time series.

Many authors have pointed out that the finite sample performance of the cointegration rank test has size distortion. For more accurate test in small sample, we may need to implement such correction procedures as suggested Johansen (2002) and Nielsen (2004), if necessary.

Acknowledgements

Earlier versions of the paper were presented at *the autumn meetings of the Japan Statistical Society* (Nagoya, 2003) and *the Japanese Economic Association* (Tokyo, 2005), a seminar at Institute of Statistical Mathematics (Tokyo, 2003) and Econometrics workshop at Kobe University (Kobe, 2005). We are grateful for helpful comments made by participants on those occasions; especially our thanks are directed to Eiji Kurozumi (Hitotsubashi University) for his valuable comments and suggestions. We would also like to express our gratitude to two anonymous referees of this journal for valuable comments. The paper is partially supported by Grant-in-Aid for Scientific Research 18730148 granted to the first author and (C)(2)17530136 granted to the second author. All the computation of this paper was conducted by the supercomputer NEC SX-7 of Tohoku University's Information Synergy Center.

References

Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially nonstationary autoregressive models, *Journal of the American Statistical Association*, **85**, 818–823.

Engle, R. F. and Granger, C. W. J. (1987). Cointegration and error correction: representation, estimation and testing, *Journal of Econometrics*, **55**, 251–276.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, **16**, 121–130.

Hannan, E. J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order, *Biometrika*, **69**, 81–94. Correction, 1983, *Biometrika*, **70**, 303.

Hosoya, Y. (1997). A note on factorization of multivariate ARMA spectra, *The Institute of Statistical Mathematics Cooperative Research Report*, **103**, 60–69.

Hosoya, Y. (2003). An asymptotic theory for inference on general unit-root cointegration, *Annual Report of the Economic Society*, *Tohoku University*, **64**, 37–55.

Hosoya, Y. (2004). An asymptotic test theory of the fractional cointegration rank, Paper presented at *Time-Series Analysis Symposium* held at Waseda University in January.

Hosoya, Y. (2005). Fractional invariance principle, *Journal of Time Series Analysis*, **26**, 463–486.

Hosoya, Y. and Takimoto, T. (2003). Inference on general unit-root cointegration and associated computational methods, *Developments of Statistical Inference*: *Preprint for a Meeting in Honour of Professor Kei Takeuchi on the Occasion of His 70th Birthday* (eds. S. Akahira *et al.*), 45–66.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221–233.

Hubrich, K., Lütkepohl, H. and Saikkonen, P. (2001). A review of systems cointegration tests, *Econometric Reviews*, **20**, 247–318.

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 Auto-regressive Models*, Oxford University Press, Oxford.

Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model, *Econometrica*, **70**, 1929–1961.

Li, W. K., Ling, S. and Wong, H. (2001). Estimation for partially nonstationary multivariate autoregressive models with conditional heteroscedasticity, *Biometrika*, **88**, 1135–1152.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model, *Econometric Theory*, **19**, 280–310.

Lütkepohl, H. and Claessen, H. (1997). Analysis of cointegrated VARMA processes, *Journal of Econometrics*, **80**, 223–239.

Miller, K. S. (1968). *Linear Difference Equations*, W.A. Benjamin, Inc., New York.

Nielsen, B. (2004). On the distribution of likelihood ratio test statistics for cointegration rank, *Econometric Reviews*, **23**, 1–23.

Pollard, D. (1985). New ways to prove central limit theorems, *Econometric Theory*, **1**, 295–314.

Poskitt, D. S. (2003). On the specification of cointegrated autoregressive moving-average forecasting systems, *International Journal of Forecasting*, **19**, 503–519.

Reinsel, G. C. and Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis*, **13**, 353–375.

Robinson, P. M. and Zaffaroni, P. (1997). Modelling nonlinearity and long memory in time series, *Nonlinear Dynamics and Time Series*, Fields Institute Communications.

Robinson, P. M. and Zaffaroni, P. (1998). Nonlinear time series with long memory: A model for stochastic volatility, *Journal of Stochastic Planning and Inference*, **68**, 359–371.

Shimotsu, K. and Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration, *The Annals of Statistics*, **33**, 1890–1933.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica*, **55**, 1035–1056.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends, *Journal of the American Statistical Association*, **83**, 1097–1107.

Takimoto, T. (2001). Computational methods for identification and estimation of VARMA model, *Annual Report of the Economic Society*, *Tohoku University*, **63**, 199–224.

Takimoto, T. and Hosoya, Y. (2004). Three-step procedure for estimating and testing cointegrated ARMAX models, *The Japanese Economic Review*, **55**, 418–450.

Watson, M. W. (1994). Vector autoregressions and cointegration, *Handbook of Econometrics volume 4* (eds. R. F. Engle and D. L. McFadden), 2843–2915, Elsevier, Amsterdam.

Yap, S. F. and Reinsel, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model, *Journal of the American Statistical Association*, **90**, 253–267.

Zaffaroni, P. (2003). Gaussian inference on certain long-range dependent volatility models, *Journal of Econometrics*, **115**, 199–258.