A modified Box-Cox transformation in the multivariate ARMA model

Takahiro Terasaka and Yuzo Hosoya

No.0601

OUC Economic Research Papers

Department of Economics

Otaru University of Commerce

A modified Box-Cox transformation in the multivariate ARMA model *

First submission December, 2004

First revision May, 2005

Second revision June, 2006

Third revision July, 2006

Takahiro Terasaka [1] and Yuzo Hosoya[2]

Abstract

The Box-Cox transformation has been used as a simple method of transforming dependent variable in ordinary-linear regression circumstances for improving the Gaussian-likelihood fit and making the disturbance terms of a model reasonably homoscedastic. The paper introduces a new version of the Box-Cox transformation and investigates how it works in terms of asymptotic performance and application, focusing in particular on inference on stationary multivariate ARMA models. The paper proposes a computational estimation procedure which extends the three-step Hannan and Rissanen method so as to accommodate the transformation and, for the purpose of parameter testing, the paper proposes a Monte-Carlo Wald test. The allied algorithm is applied to a bivariate series of Tokyo stock-price index (Topix) and the call rate.

*Earlier versions of the paper were presented at the autumn meeting of Japan Economic Society (October 2003, Tokyo) and the annual meeting of Japan Statistical Society (September 2004, Hanamaki). The research is partially supported by Grant-in-Aid for Junior Scientific Research (B)15730100 for the first author and (C)(2)15530136 for the second author respectively. The authors wish to thank to Fen Yao of Kagawa Univ. and Taro Takimoto of Kyusyu Univ. for their programming assistance.

[1]Dept. of Economics, Otaru University of Commerce. For correspondence, 3-5-21 Midori, Otaru, 047-8501; e-mail address terasaka@res.otaru-uc.ac.jp.

[2]Dept. of Economics, Meisei University.

1 Introduction

According to Box and Tidwell (1962), the objective of transformation of dependent variable is to achieve the assumption of the identically normally distributed disturbance terms and thereby to stabilize the variance and simplify the function form of the model, whereas the assumption of independence is assumed or treated as if it is satisfied. One of the earlier parametric functional transformation forms known in the literature is the Johnson family of transformations as presented by Johnson (1949). His transformation formula is given by $x^\dagger = \gamma + \delta f\{(x-\xi)/\lambda\}$ where $x^\dagger$ is a standard normal variable and the function f depends on no variable parameters and is a monotonic function of x; the family includes the Pearson type distributions. MacKinnon and Magee (1990) provide applied examples of the transformation $g(x) = \sinh^{-1}(x) = \log\left\{x + \sqrt{x^2+1}\right\}$ which belongs to the Johnson family. In practical use we need to decide upon a particular function form among the Johnson family. Box and Tidwell (1962) and Box and Cox (1964) suggest the following transformation, which is now known as the Box-Cox transformation,

$$f_1(x,\lambda) = \begin{cases} (x^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log x & \lambda = 0 \end{cases} \tag{1.1}$$

where the parameter λ should be usually estimated in a framework of a specific statistical model.

Since the Box-Cox transformation requires not only that the x in (1.1) is positive, but delimits the sample space of the transformed variable in such a way that $f_1(x,\lambda) > -1/\lambda$ if $\lambda > 0$ and $f_1(x,\lambda) < -1/\lambda$ if $\lambda < 0$ so that it is not consistent with the assumption that the transformed variable $f_1(x,\lambda)$ is normally distributed. Therefore we need in its practical applications to assume either that the available observations happen to be in that range or that they take large positive values (due to a large value of the mean level) in comparison with the variance of the transformed variable; see Davidson and MacKinnon (1993). Zarembka (1974) points out “if the probability of such large negative values is quite low, the error term may still be approximately normal”. To resolve the problem of limited range of error term, Bickel and Doksum (1981) propose the transformation $f_2(x,\lambda) = \left\{|x|^\lambda sgn(x) - 1\right\}/\lambda$,where the variable x can take negative value. But as MacKinnon and Magee (1990) point out, even though this transformation solves the difficulty of the inconsistency with the normality assumption of the error term, it has no limit as $\lambda \to 0$ when $x < 0$. We focus in this paper on the variable which takes only positive value. In section 2.1 of this paper we propose a modified Box-Cox transformation, which has the merit that, while it retains the main characteristics of the original Box-Cox transformation, the transformed variable has the range $(-\infty, \infty)$ so that the transformation is consistent with the normality assumption. Section 2.2 investigates the modified transformation in the framework of the stationary multivariate ARMA model, where a computational estimation procedure is discussed in detail, whereas Section 2.3 proposes a Monte-Carlo Wald test for testing the transformation parameter.

Even though innovation Gaussianity, variance stability and model simplicity are desirable properties for time series models to possess, rather few literature deals with data transformation. Exception is Davidson and MacKinnon (1984, 93) who discuss the maximum-likelihood estimation for nonlinear autoregression, but the asymptotic properties of the maximum-likelihood estimation given in Davidson and MacKinnon's theory is not extendable to the modified Box-Cox transformation ARMA model unless the MA part degenerates to the white noise, since their results are essentially based on a Martingale central limit theorem. In section 3, we investigate the limiting properties of the maximum Whittle-likelihood estimators for a set-up where a non-linear transformed series is generated by a stationary Gaussian linear process. The set-up is general enough to include the modified Box-Cox ARMA model as a special case, and we show the asymptotic normality of the estimators via Rozanov's version of central limit theorem for complete regular processes. We show that the asymptotic covariance matrix of the estimators is non-standard [Theorem 3.1] and involves third and forth-order cumulant spectra [Lemma 3.2]. Our asymptotic result indicates that the standard χ^2 asymptotics of the likelihood ratio test is not applicable to testing of the transformation parameter λ in the modified Box-Cox ARMA model set-up. Section 4 provides simulation results exhibiting the performance of the proposed estimation procedure. As for empirical analysis, we investigate in Section 5 a bivariate series composed of the "ratio" series of the Japanese call rate $(r(t)/r(t-1))$

and the Tokyo stock price index $(TOPIX(t)/TOPIX(t-1))$ based on the model (2.3) by application of our numerical methods. In particular, our Monte-Carlo Wald test suggests that the conventional logarithmic data transformation is not supported for both series. Section 6 is for concluding remarks and Appendix is for mathematical proofs.

As for the notations of the paper, A' and A^* denote respectively the transpose and the conjugate transpose of a matrix A, trB indicates the trace of B and $detB$ is the determinant of B. The set of all integers is denoted by $\mathbb{Z}$, whereas $\mathbb{R}^+$ and $\mathbb{R}$ denote the sets of positive and all real numbers respectively.

2 Smooth Modification

2.1 *Modifying the Box-Cox transformation*

For positive numbers x and α, set $\rho(x,\alpha) = (\log x - \log \alpha)/\log x$; then define $x^{[\lambda]}$ for the cases (i) $\lambda = 0$, (ii) $\lambda > 0$, (iii) $\lambda < 0$, respectively by

$$
\begin{aligned}
&\text{(i) } x^{[\lambda]} = \log x,\\
&\text{(ii)} x^{[\lambda]} = \begin{cases} \rho(x,\delta)\left(\log x + \frac{\delta^\lambda - 1}{\lambda} - \log\delta\right) + (1-\rho(x,\delta))(x^\lambda - 1)/\lambda, & \text{if } \ 0 < x \le \delta \\ (x^\lambda - 1)/\lambda & \text{if } \ x > \delta, \end{cases}\\
&\text{(iii) } x^{[\lambda]} = \begin{cases} (x^\lambda - 1)/\lambda & \text{if } \ 0 < x \le M \\ \rho(x,M)\left(\log x + \frac{M^\lambda - 1}{\lambda} - \log M\right) + (1-\rho(x,M))(x^\lambda - 1)/\lambda, & \text{if } \ x > M, \end{cases}
\end{aligned}
\tag{2.1}
$$

where δ and M are positive numbers chosen sufficiently small and large respectively. Note that $x^{[\lambda]}$ thus defined satisfies $\lim_{\lambda\to 0} x^{[\lambda]} = \log x$ for any x, $0 < x < \infty$, and also has the first derivative with respect to x in case $\lambda > 0$ and $0 < x \le \delta$, which is given by

$$
\frac{dx^{[\lambda]}}{dx} = \frac{\log\delta}{(\log x)^2 x}\left\{\log x - \frac{x^\lambda - 1}{\lambda} + \frac{\delta^\lambda - 1}{\lambda} - \log\delta\right\} + \frac{\log x - \log\delta}{\log x}\frac{1}{x} + \frac{\log\delta}{\log x}x^{\lambda-1}
\tag{2.2}
$$

and is seen to be continuous at $x = \delta$, whereas, for $\lambda < 0$ and $x \ge M$, the derivative is given by (2.2) with δ replaced by M. Hence in either case the modified Box-Cox transformation is continuously differentiable over the domain $0 < x < \infty$. The transformation $x^{[\lambda]}$ thus defined has the merit of mapping the set of positive values onto $(-\infty, \infty)$, and is free from the restriction on the range imposed by the original Box-Cox transformation. Not only being formally consistent with the Gaussian error term which takes any value in the real line, it also retains most of the characteristics of the Box-Cox transformation by choosing δ and M sufficiently small and large respectively. Figures 1 and 2 illustrate the modified Box-Cox transformation for certain values of λ; those figures show how the modified version stands to the original transformation.

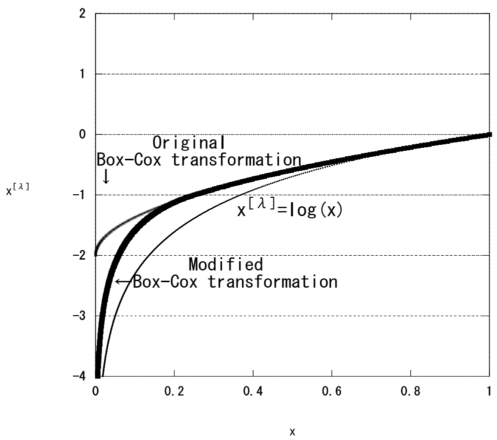

Figure 1: Modified Box-Cox transformation ($\lambda = 0.5, \delta = 0.25$)

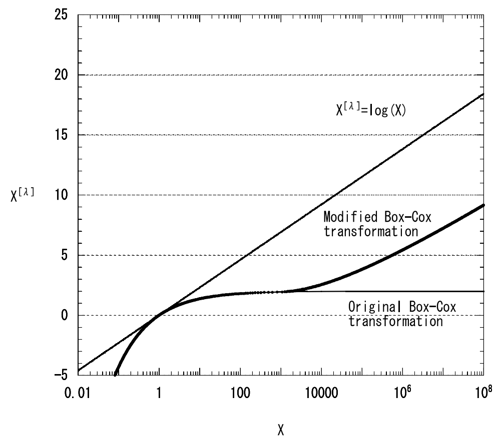

Figure 2: Modified Box-Cox transformation ($\lambda = -0.5, M = 1000$)

In empirical estimation, we predetermine the values δ and M so as that the interval (δ, M) is wide enough to contain the range of observations. Consequently the likelihood based on the modified transformation is equal to the one based on the original transformation. In the test procedure of Section 2.3, we retain the values of δ and M which are used in the estimation and consequently most of the generated data values are included in the interval (δ, M). It is desirable that the test outcome is less sensitive to a particular choice of δ and M. In this paper, we do not go into the problem of adaptive choice of δ and M.

2.2 *Estimation procedure of the stationary ARMA model involving the transformation*

To investigate the stationary ARMA model involving the modified Box-Cox transformation introduced in the previous subsection, let $\{\varepsilon(t), t \in \mathbb{Z}\}$ be a m-vector Gaussian white noise process with mean 0 and covariance matrix Σ, and suppose that the transformed series $\{y^{[\lambda]}(t), t \in \mathbb{Z}\}$ is generated by a trend-stationary ARMA process

$$\sum_{j=0}^{a} A(j) y^{[\lambda]}(t-j) = \mu + \tau t + \sum_{k=0}^{b} B(k)\varepsilon(t-k), \quad t \in \mathbb{Z} \tag{2.3}$$
$$(A(0) = B(0) = I_m),$$

where y is a m-vector and λ is a m-vector of real numbers, $y^{[\lambda]}(t) \equiv (y_1^{[\lambda_1]}(t), \cdots, y_m^{[\lambda_m]}(t))'$, $\varepsilon(t) = (\varepsilon_1(t), \cdots, \varepsilon_t(t))'$, the $A(j)$ and $B(k)$ are $m \times m$ matrices, μ is a m-vector, and τ is a coefficient m-vector for trend term. The zeros of the polynomials $\det\left\{\sum_{j=0}^{a} A(j) z^j\right\}$ and $\det\left\{\sum_{k=0}^{b} B(k) z^k\right\}$ are all assumed to be outside of the unit circle so that $\{y^{[\lambda]}(t)\}$ is an invertible trend stationary Gaussian process and we assume that the polynomials do not share common zeros.

Being conditioned on $y(0) = \cdots = y(-a+1) = 0$, $\varepsilon(0) = \cdots = \varepsilon(-b+1) = 0$, the Gaussian log likelihood function based on (2.3) for given lag-orders a, b is represented by

$$\begin{aligned}\log(L_T) = &-\frac{1}{2} mT \log(2\pi) - \frac{1}{2} T \log\det\Sigma \\ &-\frac{1}{2}\sum_{t=1}^{T}\left\{\left(y^{[\lambda]}(t) + \sum_{j=1}^{a} A(j) y^{[\lambda]}(t-j) - \mu - \tau t - \sum_{k=1}^{b} B(k)\varepsilon(t-k)\right)' \Sigma^{-1}\right. \\ &\times\left.\left(y^{[\lambda]}(t) + \sum_{j=1}^{a} A(j) y^{[\lambda]}(t-j) - \mu - \tau t - \sum_{k=1}^{b} B(k)\varepsilon(t-k)\right)\right\} + \sum_{t=1}^{T}\sum_{l=1}^{m} k_l(y_l(t), \lambda_l)\end{aligned} \tag{2.4}$$

where $k_l(y_l(t), \lambda_l)$ denotes the logarithm of $\left|\partial y_l^{[\lambda]}(t) / \partial y_l(t)\right|$ which is the l-th element of the Jacobian of the transformation from $y(t)$ to $y^{[\lambda]}(t)$. The maximizer $\hat{\Sigma}$ of this log likelihood is provided as

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\left(y^{[\lambda]}(t) + \sum_{j=1}^{a} A(j) y^{[\lambda]}(t-j) - \mu - \tau t - \sum_{k=1}^{b} B(k)\varepsilon(t-k)\right)$$

$$\times\left(y^{[\lambda]}(t)+\sum_{j=1}^{a}A(j)y^{[\lambda]}(t-j)-\mu-\tau t-\sum_{k=1}^{b}B(k)\varepsilon(t-k)\right)' \tag{2.5}$$

[see Anderson (1984, p.62)] and the maximum has the value

$$\max_{\Sigma}\log(L_T)=-\frac{1}{2}mT\{\log(2\pi)+1\}-\frac{T}{2}\log\det\hat{\Sigma}+\sum_{t=1}^{T}\sum_{l=1}^{m}k_l(y_l(t),\lambda_l)$$

whence we have the concentrated log likelihood function

$$l_T\equiv-\frac{T}{2}\log\det\hat{\Sigma}+\sum_{t=1}^{T}\sum_{l=1}^{m}k_l(y_l(t),\lambda_l) \tag{2.6}$$

[see Remark 2.1.]. We use the three-step method proposed by Hannan and Rissanen (1982) and Hannan and Kavalieris (1984) for the purpose of identifying the ARMA orders and estimating the ARMA parameters. Since the data-generating process (DGP) (2.3) involves the parameter λ, we need to modify the algorithm. Our estimation procedure is constituted of the two main steps. In the first step, we determine the ARMA lags $\hat{a},\hat{b}$, estimates of the constant term $\hat{\mu}$ and trend term $\hat{\tau}$. In the second step, we estimate the $A(j)$'s and the $B(k)$'s and λ by means of maximizing (2.6).

Step 1. Set up a system of grid values for the parameter λ. For each grid point, apply the first and second steps of Hannan-Rissanen(1982)'s three steps method. Calculate the BIC in the second step of Hannan-Rissanen method, and determine the ARMA lags $\hat{a},\hat{b}$, the observed $\varepsilon(t)$, and all the model parameters for which the BIC is smallest. More specifically this step consists of the following sub steps (1), (2) and (3) [see Remark 2.2, 2.3].

(1) For each grid point λ, apply the first step of Hannan-Rissanen's method to obtain the residual series:

- An observed series $\{\varepsilon(t)\}$, denoted by$\{\dot{\varepsilon}(t)\}$, is obtained by

$$\dot{\varepsilon}\,(t)=\sum_{j=0}^{n}\dot{A}\,(j)\,y^{[\lambda]}\,(t-j)-\dot{\mu}-\dot{\tau}t\quad t=n+1,\cdots,T$$

where $\dot{A}(0)=I_m$, the $\dot{A}(j)$ are $m\times m$ matrices, $\dot{\mu}$ is a m-vector constant term, $\dot{\tau}$ is a m-vector trend term, and n is AR lag length. We define the coefficient matrix $\dot{C}' = \left(\dot{A}(1), \cdots, \dot{A}(n), \dot{\mu}, \dot{\tau}\right)$, and define, with $N = T - n$, the $N \times m$ matrix $\dot{\mathbf{Y}} = \left(Y^{[\lambda]}(n+1), \cdots, Y^{[\lambda]}(T)\right)'$ and $N \times (mn+2)$ matrix $\dot{\mathbf{X}}$ whose typical row is $\left(-Y^{[\lambda]'}(t-1), \cdots, -Y^{[\lambda]'}(t-n), 1, t\right)$. The $\dot{C}$ is calculated by regressing $\dot{\mathbf{Y}}$ on $\dot{\mathbf{X}}$, namely $\dot{C} = \left(\dot{\mathbf{X}}'\dot{\mathbf{X}}\right)^{-1}\dot{\mathbf{X}}'\dot{\mathbf{Y}}$.

- We use the AIC for choosing the AR lag length.

$$AIC(n) = \log\det\dot{\Sigma}_n - \frac{2}{T}\sum_{t=n+1}^{T}\sum_{l=1}^{m} k_l(y_l(t), \lambda_l) + \frac{2\left(m^2 n + 2m + m(m+1)/2\right)}{T}$$

where $\dot{\Sigma}_{\dot{n}} = \frac{1}{T-\dot{n}}\sum_{\tau=n+1}^{T}\dot{\varepsilon}(\tau)\dot{\varepsilon}(\tau)'$. We denote the selected AR-lag length by $\dot{n}$.

(2) Apply the second step of Hannan-Rissanen's method to estimate the variance-covariance matrix.

- Observed process of $\varepsilon(t)$, denoted $\ddot{\varepsilon}(t)$ is evaluated by

$$\ddot{\varepsilon}(t) = \sum_{j=0}^{a}\ddot{A}(j)\, y^{[\lambda]}(t-j) - \sum_{k=1}^{b}\ddot{B}(k)\,\dot{\varepsilon}(t-k) - \ddot{\mu} - \ddot{\tau}t, \quad t = s+1, \cdots, T$$

where $\ddot{A}(0) = I_m$, the $\ddot{A}(j)$ are $m \times m$ matrices, the $\ddot{B}(k)$ are $m \times m$ matrices, $\ddot{\mu}$ is a m-vector constant term, $\ddot{\tau}$ is a m-vector trend term, and $s = \dot{n} + \max(a, b)$. We define the coefficient matrix $\ddot{C}' = \left(\ddot{A}(1), \cdots, \ddot{A}(a), \ddot{B}(1), \cdots, \ddot{B}(b), \ddot{\mu}, \ddot{\tau}\right)$, and define, with $N = T - s$, the $N \times m$ matrix $\ddot{\mathbf{Y}} = \left(Y^{[\lambda]}(s+1), \cdots, Y^{[\lambda]}(T)\right)'$ and the $N \times \{(a+b)m+2\}$ matrix $\ddot{\mathbf{X}}$ whose typical row is
$\left(-Y^{[\lambda]'}(t-1), \cdots, -Y^{[\lambda]'}(t-a), \dot{\varepsilon}'(t-1), \cdots, \dot{\varepsilon}'(t-b), 1, t\right)$. The $\ddot{C}$ is calculated by regressing $\ddot{\mathbf{Y}}$ on $\ddot{\mathbf{X}}$, namely $\ddot{B} = \left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1}\ddot{\mathbf{X}}'\ddot{\mathbf{Y}}$.

- For determining the ARMA lag lengths, we evaluate BIC defined by

$$BIC(a,b) = \log\det\ddot{\Sigma}_{a,b} - \frac{2}{T}\sum_{t=s+1}^{T}\sum_{l=1}^{m} k_l(y_l(t), \lambda_l)$$

$$+\frac{\log T\left(m^2(a+b) + 2m + m(m+1)/2\right)}{T}$$

to determine a and b. where $\ddot{\Sigma}_{a,b} = \frac{1}{T-s}\sum_{t=s+1}^{T}\ddot{\varepsilon}(n+t)\ddot{\varepsilon}(n+t)'$.

- Re-estimate the innovation sequence by fitting the selected ARMA orders; the re-estimated sequence is denoted also by $\{\ddot{\varepsilon}(t)\}$.

(3) For the grid values of the λ's at intervals of appropriate width, apply the procedures (1) and (2) above:

- For each λ and the lag orders a, b, the candidate value of the constant term $\hat{\mu}$ and the trend term $\hat{\tau}$, candidate initial value of $\ddot{A}(j)$, $\ddot{B}(k)$, and the sequence $\{\ddot{\varepsilon}(t)\}$ for maximizing (2.6) in the next step are determined and the $BIC(a,b)$ is calculated.
- The lag orders $\hat{a}$, $\hat{b}$, and $\hat{\mu}$, $\hat{\tau}$, $\ddot{A}(j)$, $\ddot{B}(k)$, $\ddot{\lambda}$, $\{\ddot{\varepsilon}(t)\}$ are chosen by the smallest $BIC(a,b)$ criterion. We determine the coefficient estimates of the constant term $\hat{\mu}=\ddot{\mu}$, and the trend term $\hat{\tau}=\ddot{\tau}$ [see Remark 2.4].

Step 2 (Hannan-Rissanen Estimation): Apply the third step of Hannan-Rissanen's method to estimate $A(a)$, $B(b)$, and λ by means of maximizing (2.6). Denote $\alpha = vec[\mu,\tau]$, $\beta = vec[A(1),...,A(a),B(1),...,B(b)]$, $\gamma = vec[\beta,\lambda]$, and $\delta = vec[\alpha,\gamma]$. Since the likelihood equations $\partial l_T/\partial\gamma = 0$ are non-linear in γ, we need an iterative method for finding a solution [see Remark 2.4].

(4) For fixed $\lambda=\ddot{\lambda}$ and given $\hat{\mu}$, $\hat{\tau}$, we maximize (2.6) with respect to β and obtain $\hat{\beta}$. Since the likelihood equations $\partial l_T/\partial\beta = 0$ are non-linear in β except when b=0, we use the Gauss-Newton procedure for obtaining the solutions of these equations. At the (j+1)-th iteration, β are updated as follows

$$\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} + \delta_j \left[\sum_{t=1}^{T} \frac{\partial\varepsilon(t)}{\partial\beta}\Sigma^{-1}\frac{\partial\varepsilon(t)}{\partial\beta'}\right]^{-1}_{\hat{\beta}^{(j)},\ddot{\lambda}} \left[\sum_{t=1}^{T} \frac{\partial\varepsilon(t)}{\partial\beta}\Sigma^{-1}\varepsilon(t)\right]_{\hat{\beta}^{(j)},\ddot{\lambda}} \tag{2.7}$$

where $\hat{\beta}^{(j)}$ is an estimate at the j-th iteration, δ^j is an scale factor at the j-th iteration [see Reincel(1997) pp126-127]. We set the scale factor $\delta_j = (1/2)^l$ where l is the smallest natural number when the log-likelihood increase at j-th iteration. The iteration terminates if the increase of the log likelihood is less than 0.1, whereas we regard the iteration not successful if $j>5$ or $l\geq 50$.

(5) For fixed $\hat{\beta}$ and given $\hat{\mu}$, $\hat{\tau}$, we maximize (2.6) with respect to λ and obtain $\hat{\lambda}$. Since the likelihood equations $\partial l_T / \partial \lambda = 0$ are non-linear in λ, we use the Quasi-Newton procedure of Fukushima and Ibaraki (1991) for obtaining the solutions of these equations [see Remark 2.4]. At the k+1-th iteration, λ are updated as follows

$$\hat{\lambda}^{(k+1)} = \hat{\lambda}^{(k)} + \eta_k \left[H_k^{-1}\right]_{\hat{\lambda}^{(k)},\hat{\beta}} \left[\frac{\partial l_T}{\partial \lambda}\right]_{\hat{\lambda}^{(k)},\hat{\beta}}$$

where $\hat{\lambda}_k$ is an estimate at the k-th iteration, η_k is an scale factor at the k-th iteration, and H_k^{-1} is the k-th approximate inverse Hessian matrix $[\partial^2 l_T / \partial \lambda \partial \lambda']^{-1}_{\hat{\lambda}^{(k)}\hat{\beta}}$ which is updated by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula.

(6) If the difference of the log-likelihood attained in (5) and (4) is less than 0.1 in modulus, we stop the optimization, otherwise iterate the step 2 commencing from $\hat{A}(j)$, $\hat{B}(k)$, $\hat{\lambda}$ to $\ddot{A}(j)$, $\ddot{B}(k)$, $\ddot{\lambda}$.

Remark 2.1. The likelihood function (2.4) is not an exact one since it involves unobservable random vectors. See Reincel (1997) for example for the exact likelihood and the allied asymptotics in the case the transformation is not involved. The use of exact likelihood in the present situation makes the computation amount required in the estimation and testing procedure impracticable.

Remark 2.2. In respect to ARMA order selection, there seem to be two basic approaches in the literature; one is based on diagnostic checking (or testing) of the serial and the partial serial covariances basically by the likelihood-ratios and the other is the use of such information criteria as the AIC [Akaike (1973)] or BIC. Hosoya (1989a; 2002, Chapter 9) maintains that there is no notable merit in using the AIC for selecting among finite-parameter models.

Remark 2.3. If the initial values of λ are chosen at intervals of suitable width (for example, the lattice of width of 0.02 or 0.1) in the step 1, $\hat{\lambda}$ obtained in the step 2 does not depart much from the Step 1 estimate $\tilde{\lambda}$. We modified the algorithm in the line search step so as that we reduced the length of the direction vector in case some element of λ exceeds 3 in the simulation analysis and 4 in the empirical analysis.

Remark 2.4. We do not estimate the parameter $\delta = vec[\alpha, \gamma]$ simultaneously because the likelihood equation $\partial l_T / \partial \delta = 0$ are so complicated that numerical optimization is hard to attain. Since the Hessian matrix is not block-diagonal between α and β, α and λ, and β and λ, our separated estimation procedure would not be efficient, but we can control the numerical convergence better by this way. Furthermore, it is a merit of our Monte-Carlo test method that it applies even if an efficient estimation method is not used as long as the estimators are asymptotically normal.

2.3 *Monte-Carlo Wald Test*

As is seen in the next section, the standard asymptotics does not apply to the model (2.3), and also it is not easy to evaluate numerically the asymptotic covariance matrix derived in Theorem 3.1. So we propose a Monte-Carlo Wald test for the purpose of testing the parameters of the model (2.3). Our main concern is to test the validity of a set of g restriction of the element of λ. For the null hypothesis $\lambda = \lambda^0$, our test procedure is given as follows:

Step 1. Estimate the model (2.3) under the restriction $\lambda = \lambda^0$ for a set of observations $\{y(1), \cdots, y(T)\}$.

Step 2. Generate data series based on the model estimated in Step 1:

(1) Generate independent normal random vectors $\{\varepsilon(t)^\dagger\}$ with mean 0 and variance-covariance matrix of the disturbance estimated under the null hypothesis.

(2) For coefficients estimated under the null hypothesis, generate $\{y(t)^\dagger; t = 1, \cdots, T\}$.

Step 3. Estimate λ for the generated series $\{y(t)^\dagger\}$ where the estimate is denoted by $\lambda^\dagger$.

Step 4: Evaluate the sample variance-covariance matrix of $\lambda^\dagger$:

(1) Iterate Steps 2 and 3 N times and obtain $\lambda^\dagger_{(1)}, \cdots, \lambda^\dagger_{(N)}$, and calculate its variance -covariance matrix $\Omega^\dagger = \frac{1}{N}\left\{\sum_{k=1}^{N}(\lambda^\dagger_{(k)} - \bar{\lambda})(\lambda^\dagger_{(k)} - \bar{\lambda})'\right\}$. where $\bar{\lambda} = \sum_{k=1}^{N} \lambda^\dagger_{(k)}/N$

(2) Finally we obtain the Wald statistic as

$$W = \left(\hat{\lambda}_T - \lambda^0\right)' \Omega^{\dagger^{-1}} \left(\hat{\lambda}_T - \lambda^0\right).$$

where $\hat{\lambda}_T$ is the estimate of λ for the nonrestrictive model based on the original data $\{y(t); t = 1, \cdots, T\}$.

(3) We reject the null hypothesis if the calculated Wald statistic is grater than the critical value form a χ^2 distribution with g degrees of freedom, at the chosen level of significance.

Remark 2.5. As for Monte-Carlo Test, see for example Efron and Tibshirani(1993).

3 Asymptotic theory for transformation-linear processes

Davidson and MacKinnon (1984) provide asymptotic properties of the maximum-likelihood estimators in a nonlinear regression model. They specify a set of conditions for the standard asymptotics of estimation to hold whereas Amemiya (1977) discusses only rather in abstract terms. Their information-amount equality, which enables the asymptotic covariance matrix of the estimators to be equated to the inverse of the Fisher information amount, however, relies crucially upon the Martingale property of the first-order derivative of the log likelihood function; see Davidson and MacKinnon (1984, p.499). But that property seems valid only in such special cases as a certain class of nonlinear AR processes with i.i.d. Gaussian error terms, and does not apply to more general circumstances where the error terms are dependent.

With application to the modified Box-Cox model (2.3) in view and taking an approach somewhat different from Davidson and MacKinnon's, this section investigates limiting properties of the maximum Whittle likelihood estimator for a transformed series $F(y(t), \lambda), t = 1, \cdots, T$, in a general linear process set-up. Instead of the time-domain representation, we work with the frequency-domain version of the likelihood function, since the latter approach makes much easier the derivation and enables compact representation of the asymptotic results when a transformed series is generated by a general stationary process. The appendix provides sketchy proofs of the results established in this section; formal proofs are straightforward but require a very lengthy train of arguments. Hosoya and Taniguchi (1982, 1993) and Hosoya (1989b, 1997) provide asymptotic properties of Whittle-likelihood based statistics for linear processes in a very general framework, but the presence of nonlinear transformation necessitates a certain modification to the proofs given in those papers.

Let $F(y, \lambda)$ be a m-vector valued one-to-one function of $y \in (\mathbb{R}^+)^m$ onto $\mathbb{R}^m$ for each λ, and suppose that the functional form of $F(y, \lambda)$ is known except for λ and that $F(y, \lambda)$ and its Jacobian matrix are three times continuously differentiable with respect to λ except for a finite set of y. For the model (2.3), we have $F(y, \lambda) = y^{[\lambda]}$ and this condition is satisfied. Although the modified Box-Cox model (2.3) and its applied model in section 5 contains the time-trend term, we deal only with the case it is absent. The asymptotic theory of this section is extensible to the trend-stationary model, only making the asymptotic covariance matrix expression of Theorem 3.1 a somewhat more complicated one. Denote by θ the generic s-vector parameter subsuming the model parameters except for a m-vector μ and a n-vector λ [for example, $s = m^2(a+b) + \{m(m+1)/2\}$ for the model (2.3)]; we assume that n is not necessarily equal to m for generality. The parameter spaces of θ, μ and λ are denoted by Θ, M and Λ which are compact sets containing open sets to which the true values θ^0, μ^0 and λ^0 belong. Suppose that the m-vector $F(y(t), \lambda)$ is generated by a short-memory linear process

$$F(y(t), \lambda) = \mu + \sum_{j=0}^{\infty} G(j, \theta)\varepsilon(t-j) \tag{3.1}$$

where the $\varepsilon(t),\ t = \pm 1, \pm 2, \cdots$, are Gaussian white noise m-vector process with mean $E(\varepsilon(t)) = 0$ and covariance matrix $Cov(\varepsilon(t)) = \Sigma(\theta)$; the coefficients $G(j, \theta)$ are $m \times m$ matrices such that $G(0, \theta) \equiv I_m$.

To focus on the effect of the introduction of the $F(\cdot, \lambda)$-transformation, we impose rather a stringent set of regularity conditions A(i) through (iii) below paralleling to the one Walker (1964) introduced on the spectral density matrix $g(\omega, \theta)$ of the linear process $F(y(t), \lambda) - \mu = \sum_{j=0}^{\infty} G(j, \theta)\varepsilon(t-j)$ which is represented by

$$g(\omega, \theta) = \frac{1}{2\pi}\{\sum_{j=0}^{\infty} G(j, \theta)e^{i\omega j}\}\Sigma(\theta)\{\sum_{j=0}^{\infty} G(j, \theta)e^{i\omega j}\}^*.$$

While Walker deals only with a scalar-valued process, the extension to vector processes is straightforward. Setting $h(\omega, \theta) = g(\omega, \theta)^{-1}$ in the sequel, we assume the following:

Assumption A:

(i) The first-order derivatives $h^{(j)}(\omega, \theta) = \partial h(\omega, \theta)/\partial \theta_j, j = 1, \cdots, s$, exist and are continuous functions of (ω, θ) on $[-\pi, \pi] \times \Theta$.

(ii) The second and third order derivatives $h^{(i,j)}(\omega, \theta) = \partial^2 h(\omega, \theta)/\partial \theta_i \partial \theta_j$ and $h^{(i,j,k)}(\omega, \theta)$ $= \partial^3 h(\omega, \theta)/\partial \theta_i \partial \theta_j \partial \theta_k, i, j, k = 1, \cdots, s$, exist and are continuous functions of (ω, θ) on $[-\pi, \pi] \times N_\delta(\theta^0)$ where $N_\delta(\theta^0)$ is a neighbourhood of θ^0.

(iii) The coefficients $G(j, \theta^0)$ in the process (3.1) satisfy:

$$\sum_{j=0}^{\infty} j \|G(j, \theta^0)\| < \infty,$$

where $\| \cdot \|$ denotes a Euclidean norm.

(iv) $F(y, \lambda)$ is differentiable with respect to y and the Jacobian matrix $D_y F(y, \lambda) = \{\partial F_j(y, \lambda)/\partial y_k, j, k = 1, \cdots, m\}$ is positive definite for $(y, \lambda) \in (\mathbb{R}^+)^m \times \Lambda$. $D_y F(y, \lambda)$ is third-order differentiable with respect to λ except for a finite set of y-values.

The assumptions A (i) through (iii) are certainly satisfied for identifiable, invertible stationary ARMA models. Denote the finite Fourier transformation of $F(y(t), \lambda), t = 1, \cdots, T$, by

$$\tilde{F}_T(\omega, \lambda) = \frac{1}{\sqrt{2\pi T}} \left(\sum_{t=1}^{T} F(y(t), \lambda) e^{it\omega} \right)$$

and set $1_T(\omega) = (2\pi T)^{-1/2} \sum_{t=1}^{T} e^{it\omega}$. Denote by ξ the vector of the whole parameters involved; namely, $\xi = (\theta', \mu', \lambda')'$ so that $\xi \in \Xi \equiv \Theta \times M \times \Lambda$. Denote by $Q_T(\xi)$ the log Whittle likelihood divided by $(-T/2)$; namely,

$$\begin{aligned} Q_T(\xi) &= Q_T(\theta, \mu, \lambda) \\ &= \log det\Sigma(\theta) + \frac{1}{2\pi} \int_{-\pi}^{\pi} tr[h(\omega, \theta)\{\tilde{F}_T(\omega, \lambda) - 1_T(\omega)\mu\}\{\tilde{F}_T(\omega, \lambda) - 1_T(\omega)\mu\}^*] d\omega \\ &\quad - \frac{2}{T} \sum_{t=1}^{T} J(y(t), \lambda), \end{aligned} \tag{3.2}$$

where $$J(y(t),\lambda) \equiv \log detD_y F(y(t),\lambda). \tag{3.3}$$

The derivatives with respect to λ_i and/or λ_j are denoted by superscripts in parentheses such as $F^{(i)}(y(t),\lambda), F^{(i,j)}(y(t),\lambda)$; so that for instance, $F^{(i,j)}(y(t),\lambda) \equiv \partial^2 F(y(t),\lambda)/\partial\lambda_i\partial\lambda_j$. Set $z(t) = F(y(t),\lambda^0)$ for $y(t)$ generated by (3.1) for $\xi = \xi^0$; similarly, set $z^{(i)}(t) = F^{(i)}(y(t),\lambda^0), z^{(i,j)}(t) = F^{(i,j)}(y(t),\lambda^0), z^{[k]}(t) = J^{(k)}(y(t),\lambda^0)$ and $z^{[k,l]}(t) = J^{(k,l)}(y(t),\lambda^0)$, where $z^{[k]}(t)$ and $z^{[k,l]}(t)$ are both scalars. The co-spectral density matrices are indicated by $c(\omega)$ with superscripts so that, for instance, $c^{\cdot,(i)}(\omega)$ and $c^{(i),(j,k)}(\omega)$ denote the $m \times m$ density matrices between the processes z and $z^{(i)}$ and between $z^{(i)}$ and $z^{(j,k)}$ respectively, whereas $c^{(i),(j),[k]}_{\alpha_1,\alpha_2}(\omega_1,\omega_2)$ denotes the third-order spectral density for $\{z^{(i)}_{\alpha_1}(t), z^{(j)}_{\alpha_2}(t), z^{[k]}(t)\}$. To express fourth-order cumulant spectral densities, we use such notation as $g_{\alpha_1,\alpha_2,\alpha_3,\alpha_4}(\omega_1,\omega_2,\omega_2|\cdot,(i),(j))$ to indicate the fourth-order density for α_1 to α_4-th elements of the joint $(3m)$-vector process $\{z(t)', z^{(i)}(t)', z^{(j)}(\cdot)'\}'$ where $1 \le \alpha_1,\cdots,\alpha_4 \le 3m$ so that the α_j can indicate any component of the $3m$-vector, whereas $g(\omega|\cdot,(i))$ denotes the $2m\times 2m$ spectral density matrix of the joint process $\{z(t)', z^{(i)}(t)'\}'$. We assume the following assumptions B and C to hold under the null DGP.

Assumption B:

(i) For any $i = 1,\cdots,n$, and for any $\lambda_1 \in \Lambda$,

   (1) there exists $\delta_1 > 0$ such that $E\sup_{\lambda\in B_{\delta_1}(\lambda_1)} \parallel F^{(i)}(y(1),\lambda) \parallel^2 < \infty$, and $E\sup_{\lambda\in B_{\delta_1}(\lambda_1)} |J^{(i)}(y(1),\lambda)|^2 < \infty$,

   (2) the process $\{\sup_{\lambda\in B_\delta(\lambda_1)} \parallel F^{(i)}(y(t),\lambda) \parallel\}$ for each i and any λ_1 after mean correction has a spectral density $g_{i,\lambda_1}(\omega)$ such that $\int_{-\pi}^{\pi} |g_{i,\lambda_1}(\omega)|^{1+c} < \infty$ for some $c > 0$ and has a bounded fourth-order cumulant spectral density,

   (3) there exists $\delta_2 > 0$ for which $E \parallel z^{(i)}(t) \parallel^{2+\delta_2} < \infty$, and $E \parallel z^{[i]}(t) \parallel^{2+\delta_2} < \infty$.

(ii) There exists a neighborhood $B_\delta(\lambda^0)$ of the true λ^0 and $c > 0$ such that suprema of $\parallel F^{(i,j)}(y(1),\lambda) \parallel^2, \parallel F^{(i,j,k)}(y(1),\lambda) \parallel^2$ and $|J^{(i,j,k)}(y(1),\lambda)|^2$ over $B_\delta(\lambda^0)$ have finite expectation for any $i,j,k = 1,\cdots,n$.

(iii) Let P_t denotes the product of a pair of series $x(t)$ and $y(t)$ by $P_t(x,y,\tau) = x(t)y(t+$

τ); then for any fixed positive integer T_1, the joint process $\{P_t(z_{i_1}, z_{i_2}, \tau_1), P_t(z_{i_3}, z_{i_4}^{(j_1)}, \tau_2),$ $P_t(1, z_{i_5}, 0), P_t(1, z_{i_6}^{[j_2]}, 0); i \leq i_1, \cdots, i_6 \leq m,\ 1 \leq j_1, j_2 \leq m, -T_1 \leq \tau_1, \tau_2 \leq T_1\}$ has a second-order spectral density which is bounded and continuous.

(iv) For $i, j, k = 1, \cdots, n$, the third-order cumulant spectral densities for the processes $\{z_{\alpha_1}, z_{\alpha_2}^{(i)}\}, \{z_{\alpha_1}, z_{\alpha_2}^{(i)}, z_{\alpha_3}^{(j)}\}$ and $\{z_{\alpha_1}^{(i)}, z_{\alpha_2}, z^{[j,k]}\}(1 \leq \alpha_1, \alpha_2, \alpha_3 \leq m)$ exist, respectively and are bounded. Also the fourth-order cumulant spectral densities for the $(3m)$-vector process $\{z, z^{(i)}, z^{(j)}\}$ exist and are bounded.

Assumption C: (i) For any $\xi \neq \xi^0$, $\lim_{T\to\infty} E(Q_T(\xi) - Q_T(\xi^0)) > 0$.

(ii) $\lim_{T\to\infty} \sqrt{T} E(\partial Q_T(\xi^0)/\partial\lambda) = 0$.

Remark 3.1. Assumption C (i) is an identifiability condition whereas C (ii) is certainly satisfied in case Q_T is given by the exact Gaussian likelihood since then $E(\partial Q_T(\xi^0)/\partial\lambda)$ $= 0$ for all T as long as the likelihood is defined. For the Whittle likelihood function, the assumption is proved under a general condition by Hosoya (1997) in the case where the process in concern does not involve F-transformation and the proof applies to $\partial Q_T(\xi^0)/\partial\theta$ and $\partial Q_T(\xi^0)/\partial\mu$. To prove the assumption explicitly for $\partial Q_T(\xi^0)/\partial\lambda$, certain additional assumptions are required.

In the sequel, Assumptions A, B and C are assumed to hold.

Lemma 3.1. *We have:*

$$p\lim_{T\to\infty} \frac{\partial^2 Q_T(\xi^0)}{\partial\xi\partial\xi'} = \lim_{T\to\infty} E\left(\frac{\partial^2 Q_T(\xi^0)}{\partial\xi\partial\xi'}\right) \equiv \Psi \equiv \begin{bmatrix} \Psi_{\theta\theta} & \cdot & \cdot \\ \Psi_{\theta\mu} & \Psi_{\mu\mu} & \cdot \\ \Psi_{\theta\lambda} & \Psi_{\mu\lambda} & \Psi_{\lambda\lambda} \end{bmatrix},$$

where the respective components are given by

$$
\begin{aligned}
\Psi_{\theta_\alpha\theta_\beta} &= \int_{-\pi}^{\pi}\left[tr\{h^{(\alpha,\beta)}(\omega,\theta^0)g(\omega,\theta^0)\} + \frac{\partial^2}{\partial\theta_\alpha\partial\theta_\beta}\log det\Sigma(\theta^0)\right]d\omega; && (3.4)\\
\Psi_{\theta_\alpha\mu_l} &= 0; && (3.5)\\
\Psi_{\theta_\alpha\lambda_j} &= \frac{1}{2\pi}tr\int_{-\pi}^{\pi}h^{(\alpha)}(\omega,\theta^0)\{c^{(j),\cdot}(\omega)+c^{\cdot,(j)}(\omega)\}d\omega; && (3.6)\\
\Psi_{\mu_l\mu_n} &= \frac{1}{2\pi}\{h_{ln}(0,\theta^0)+h_{nl}(0,\theta^0)\}; && (3.7)\\
\Psi_{\mu_l\lambda_i} &= \frac{1}{2\pi}tr\left[h(0,\theta^0)\{E(z^{(i)}(1))e_l(m)'+e_l(m)E(z^{(i)}(1))'\}\right] && (3.8)\\
\Psi_{\lambda_i\lambda_j} &= \frac{1}{2\pi}tr\int_{-\pi}^{\pi}h(\omega,\theta^0)\left[c^{(i,j),\cdot}(\omega)+c^{(i),(j)}(\omega)+c^{(j),(i)}(\omega)+c^{\cdot,(i,j)}(\omega)\right]d\omega \\
&\quad+\frac{1}{2\pi}tr[h(0,\theta^0)\{E(z^{(i)}(1))E(z^{(j)}(1))'+E(z^{(j)}(1))E(z^{(i)}(1))'\}] \\
&\quad+E\{z^{[i,j]}(1)\}, && (3.9)
\end{aligned}
$$

where $e_l(m)$ denotes the unit column m-vector whose l-th element is unity.

Let $H^1(\omega)$ be the $2m\times 2m$ block diagonal matrix with $h(\omega)$ in the first block and 0 in the second block, and let $H^2(\omega)$ be the $2m\times 2m$ matrix given by

$$H^2(\omega) = \begin{bmatrix} 0 & h(\omega) \\ h(\omega) & 0 \end{bmatrix},$$

where $h(\omega,\theta^0)$ is abbreviated as $h(\omega)$ or h. Let $H^3(\omega)$ and $H^4(\omega)$ be the $3m\times 3m$ matrices respectively given by

$$H^3(\omega) = \begin{bmatrix} 0 & h(\omega) & 0 \\ h(\omega) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad\text{and}\quad H^4(\omega) = \begin{bmatrix} 0 & 0 & h(\omega) \\ 0 & 0 & 0 \\ h(\omega) & 0 & 0 \end{bmatrix}.$$

The next lemma establishes the asymptotic normality of the estimator $\hat{\xi}$.

Lemma 3.2.

$$T^{1/2}\frac{\partial Q_T(\xi^0)}{\partial\xi} \xrightarrow{d} N(0,\Phi) \qquad (3.10)$$

where Φ is the asymptotic covariance matrix whose block components are denoted by

$$\Phi = \begin{bmatrix} \Phi_{\theta\theta} & \cdot & \cdot \\ \Phi_{\theta\mu} & \Phi_{\mu\mu} & \cdot \\ \Phi_{\theta\lambda} & \Phi_{\mu\lambda} & \Phi_{\lambda\lambda} \end{bmatrix},$$

where by means of the notations $g's$ and $c's$ introduced in the paragraph preceding to Assumption B the components are given specifically by

$$\begin{aligned}
\Phi_{\theta_\alpha\theta_\beta} &= \Psi_{\theta_\alpha\theta_\beta}; && (3.11)\\
\Phi_{\theta_\alpha\mu_j} &= \Psi_{\theta_\alpha\mu_j} = 0; && (3.12)\\
\Phi_{\theta_\alpha\lambda_i} &= 4\pi tr\int_{-\pi}^{\pi} H^1(\omega)g(\omega|\cdot,(i))H^2(\omega)g(\omega|\cdot,(i))d\omega \\
&\quad +2\pi\sum_{\alpha_1,\cdot,\alpha_4=1}^{2m}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} H^{(1)}_{\alpha_1\alpha_2}(\omega_1)H^{(2)}_{\alpha_3\alpha_4}(\omega_2)g_{\alpha_1,\cdots,\alpha_4}(-\omega_1,\omega_2,-\omega_2|\cdot,(i))d\omega_1 d\omega_2 \\
&\quad +\sum_{\alpha_1,\alpha_2=1}^{m}\sqrt{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} h^{(\alpha)}_{\alpha_1,\alpha_2}(\omega)c^{\cdot,\cdot,[j]}_{\alpha_1,\alpha_2}(-\omega_1,\omega_2)d\omega_1 d\omega_2; && (3.13)
\end{aligned}$$

$$\begin{aligned}
\Phi_{\mu_l\mu_n} &= \Psi_{\mu_l\mu_n}; && (3.14)\\
\Phi_{\mu_i\lambda_j} &= \sum_{\alpha_1,\alpha_2=1}^{m}\int_{-\pi}^{\pi} h_{\alpha_1,i}(\omega)c^{\cdot,[j]}_{\alpha_2}(\omega)d\omega \\
&\quad +\sum_{\alpha_1,\cdots,\alpha_4=1}^{m}\sqrt{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} h_{\alpha_1,i}(\omega_1)h_{\alpha_2,\alpha_3}(\omega_2)[c^{\cdot,(j),\cdot}_{\alpha_1,\alpha_2,\alpha_3}(-\omega_1,\omega_2) \\
&\quad +c^{\cdot,\cdot,(j)}_{\alpha_1,\alpha_2,\alpha_3}(-\omega_1,\omega_2)]d\omega_1 d\omega_2; && (3.15)\\
\Phi_{\lambda_i\lambda_j} &= 4\pi tr\int_{-\pi}^{\pi} H^3(\omega)g(\omega|\cdot,(i),(j))H^4(\omega)g(\omega|\cdot,(i),(j))d\omega + c^{[i],[j]}(0) \\
&\quad +2\pi\sum_{\alpha_1,\cdots,\alpha_4=1}^{3m}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} H^3_{\alpha_1\alpha_2}(\omega_1)H^4_{\alpha_3\alpha_4}(\omega_2)g_{\alpha_1,\cdots,\alpha_4}(-\omega_1,\omega_2,-\omega_2|\cdot,(i),(j))d\omega_1 d\omega_2 \\
&\quad +\sum_{(k,l)=(i,j),(j,i)}\sum_{\alpha_1,\alpha_2=1}^{m}\sqrt{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} h_{\alpha_1\alpha_2}(\omega)c^{(k),\cdot,[l]}_{\alpha_1,\alpha_2}(\omega_1,\omega_2)d\omega_1 d\omega_2; && (3.16)
\end{aligned}$$

Theorem 3.1. *Suppose that Φ and Ψ are non-singular; then, under Assumptions A, B, C, $p\lim_{T\to\infty}\hat{\xi} = \xi^0$ and $\sqrt{T}(\hat{\xi}-\xi^0)$ is asymptotically distributed as $N(0,\Psi^{-1}\Phi\Psi^{-1})$.*

The representations of Φ and Ψ in Lemmas 3.1 and 3.2 indicate that, although Ψ is expressed only in terms of the second-order spectral density of $F(y(t),\lambda^0)$ and their derivatives, Φ involves the third and fourth-order spectra and so in general $\Phi \neq \Psi$. The result of Theorem 3.1 is thus at variance with the standard asymptotic theory according to which the asymptotic distribution should be $N(0, \Phi^{-1})$. The consequence is that the likelihood-ratio test statistics in general are not asymptotically χ^2-distributed under the null hypothesis $\lambda = \lambda^0$. For example, suppose that we are interested in testing the null hypothesis $H_0 : \xi = \xi^0$ for the process $\{y(t)\}$ generated by (3.1); then the statistic $2 \max_\xi L_T(\xi)/L_T(\xi^0)$ is distributed asymptotically as $z'\Psi z$ where z is a normal random-vector with mean 0 and covariance matrix $\Psi^{-1}\Phi\Psi^{-1}$.

4 Simulation analysis

In this section, we examine the performance of our algorithm proposed in Section 2.2 by applying it to two examples of the DGP (2.3). For assessment of the performance of the algorithm, we focus on the lag-orders selection, the small sample distributions of the estimates of λ, (1,1) element estimates of $A(1)$ and Σ. We also apply a portmanteau Gaussianity test by Doornik and Hansen (1994) to examine whether the residuals from the fitted Box-Cox model are reasonably normally distributed. We conducted those numerical evaluations by means of the super computer SX-7 of Tohoku University Information Synergy Center. The programs were written in FORTRAN. We set the number of replication to be 1500 times, in each of which T is set to be 500. One execution of the estimation procedure took about 10 minutes.

Remark 4.1. Generation of $y(t)$ series requires inversion of $y^{[\lambda]}(t)$ to $y(t)$ for each given value of λ. We used the bisection method for numerical inversion.

Case 1. The ARMA coefficients in this example are the same as in Reinsel (1997, p.318), except that we add the constant term. We set $(\lambda_1, \lambda_2) = (-0.5, -0.5)$ to see how the negative exponents affect the results whereas the variance-covariance matrix is set as below. Specifically, suppose that the DGP is given by the bivariate $ARMA\,(2, 1)$ model $y^{[\lambda]}\,(t) + A(1)y^{[\lambda]}\,(t-1) + A(2)y^{[\lambda]}\,(t-2) = \mu + \varepsilon\,(t) + B(1)\varepsilon\,(t-1)$ where the coefficients are

$$A(1) = \begin{bmatrix} -1.5 & 0.6 \\ -0.3 & -0.2 \end{bmatrix}, A(2) = \begin{bmatrix} 0.5 & -0.3 \\ -0.7 & 0.2 \end{bmatrix}, \mu = \begin{bmatrix} -2.0 \\ -2.0 \end{bmatrix}, B(1) = \begin{bmatrix} -0.4 & -0.3 \\ 0.5 & -0.8 \end{bmatrix},$$

$$\varepsilon(t) \sim \text{I.I.D.}N \begin{bmatrix} 2.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}.$$

The parameters of the modification are chosen as $(\delta, M) = (0.25, 1000)$. The average rate of generated data which exceed M is 7.8 percent. The histograms of the estimates are exhibited in Figures 3 through 6 for the cases where the true ARMA orders are estimated.

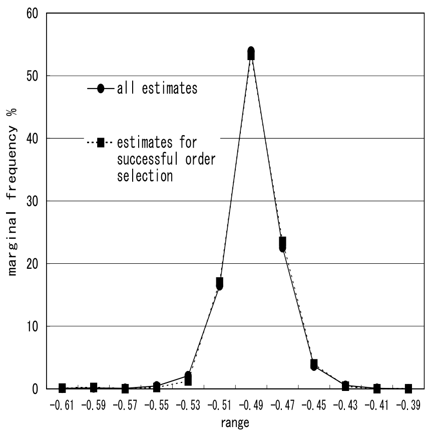

Figure 3: Histogram of λ_1

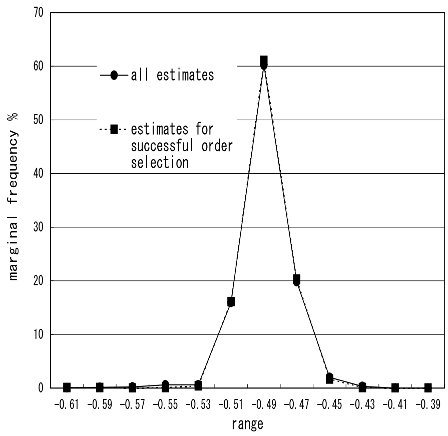

Figure 4: Histogram of λ_2

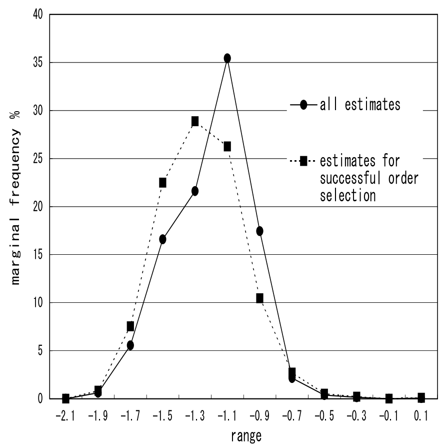

Figure 5: Histogram of coefficient (1,1) of AR(1)

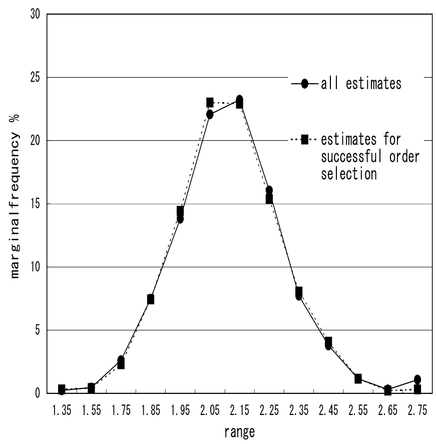

Figure 6: Histogram of variance (1,1) element

Figures 3 and 4 show that the estimates $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are reasonably centered around the true value $(\lambda_1, \lambda_2) = (-0.5, -0.5)$. Figure 5 exhibits the histogram of the (1,1) element of the estimated coefficient matrix $\hat{A}(1)$ where we observe that the estimates tend to have a small bias. The histogram of the (1,1) element of the estimated covariance matrix $\hat{\Sigma}$ in Figure 6 indicates that the estimates also have a small positive bias. The histograms based on all estimates are quite similar to those based only on successful order selection.

Case 2. To look into the performance of logarithmic transformation, we set $(\lambda_1, \lambda_2) = (0, 0)$ and $\mu = (1.0, 1.0)'$ retaining the other aspects as in case 1. Since the corresponding histograms of the estimates are very similar to the ones for case 1, they are not exhibited. Tables 1 through 5 summarize the computation performance.

Table 1: Simulation Performance

Simulation case	Iteration	calculation successful	successful order selection
Case 1	1500	1296	956
Case 2	1500	1208	952

Table 2: Lag-order selection (case 1)

		AR order										
		0	1	2	3	4	5	6	7	8	9	10
MA order	0	0	0	0	18	13	3	2	0	9	6	2
	1	0	0	956	6	0	0	0	0	0	0	0
	2	0	277	3	1	0	0	0	0	0	0	0

Table 3: Lag-order selection (case 2)

		AR order										
		0	1	2	3	4	5	6	7	8	9	10
MA order	0	0	0	0	14	11	1	2	2	0	1	0
	1	0	2	952	1	0	0	0	0	0	0	0
	2	0	215	6	0	0	0	0	0	0	0	0
	3	0	1	0	0	0	0	0	0	0	0	0

Table 4: Means and Mean Square errors of estimates

	Parameter		λ_1	λ_2	coefficient (1,1) of AR(1)	variance (1,1) element
Case 1	estimates for successful order selection	mean	-0.498	-0.497	-1.26	1.98
		MSE	0.000274	0.000348	0.119	0.0990
	all estimates	mean	-0.499	-0.498	-1.20	1.98
		MSE	0.000298	0.000353	0.150	0.0986
	true value		-0.5	-0.5	-1.5	2.0
Case 2	estimates for successful order selection	mean	-0.00127	-0.000890	-1.24	1.97
		MSE	0.249	0.249	0.116	0.0194
	all estimates	mean	-0.00122	0.000907	-1.19	1.97
		MSE	0.249	0.249	0.142	0.0223
	true value		0.0	0.0	-1.5	2.0

Table 5: Doornik and Hansens' normality test

Simulation case		Number of rejection Significance level 5 %	1 %
Case 1	calculation successful	16	3
	successful order selection	11	3
Case 2	calculation successful	10	1
	successful order selection	10	1

Table 1 shows the number of successful simulations, and Table 2 and 3 exhibit the selected order. They indicate that the order selection is mostly successful. Table 4, which summarizes the means and mean-square errors of the estimates, indicates that the estimation procedure we propose works reasonably well. In regard to the residual normality, we applied the Doornik and Hansen test to the residuals obtained in Cases 1 and 2 simulations. From the results shown in Table 5, we may conclude that our transformation is efficient for residual normality.

5 Empirical results

This section deals with the time series of bivariate monthly data of Japanese call rate $r(t)$ and the Tokyo stock price index $TOPIX(t)$. We fit the ARMA model (2.3) to the ratio data $(r(t)/r(t-1))$ and $(TOPIX(t)/TOPIX(t-1)) \times 100$ under two hypotheses. The model a) leaves (λ_1, λ_2) to be estimated whereas the model b)imposes the constraint $(\lambda_1, \lambda_2) = (0.0, 0.0)$. The Monte-Carlo Wald test introduced in Section 2.3 is carried out to test the model b) against a). Notice that the ratio data are related to the growth rate of the level data and the Box-Cox transformed ratio data $(x(t)/x(t-1))$ is the growth rate of the transformed level data $x^\lambda(t)$; namely, if $f_1(y, \lambda)$ denotes the ordinary Box-Cox transformation,

$$f_1(x(t)/x(t-1), \lambda) = \begin{cases} \{x^\lambda(t) - x^\lambda(t-1)\}/\{\lambda x^\lambda(t-1)\} & \text{if} \quad \lambda \neq 0 \\ \log x(t) - \log x(t-1) & \text{if} \quad \lambda = 0. \end{cases}$$

We conducted the estimation of Section 2.2 for the 161 observations over the period stating from August 1985 through December 1998. The results for a) are as follows: The selected ARMA lags are ARMA$(0, 1)$, and the estimates are

$$\hat{\lambda} = \begin{bmatrix} 3.3 \\ 0.1 \end{bmatrix}, \hat{\mu} = \begin{bmatrix} 0.00527 \\ 5.88 \end{bmatrix}, \hat{\tau} = \begin{bmatrix} 0.000180 \\ 0.000344 \end{bmatrix}, \hat{B}(1) = \begin{bmatrix} 0.138 & 0.186 \\ -0.153 & 0.303 \end{bmatrix},$$

$$\hat{\Sigma} = \begin{bmatrix} 4.73 \times 10^{-3} & -2.78 \times 10^{-4} \\ -2.78 \times 10^{-4} & 5.20 \times 10^{-3} \end{bmatrix}$$

and Doornik and Hansens' normality test statistic is 3.3947. As regards the results for b), the selected ARMA lags are ARMA$(0, 1)$. The parameter estimates are

$$\hat{\mu} = \begin{bmatrix} 0.00306 \\ 4.62 \end{bmatrix}, \hat{\tau} = \begin{bmatrix} 0.000284 \\ 0.000214 \end{bmatrix}, \hat{B}(1) = \begin{bmatrix} 0.198 & 0.307 \\ -0.0841 & 0.301 \end{bmatrix},$$

$$\hat{\Sigma} = \begin{bmatrix} 6.78 \times 10^{-3} & -2.07 \times 10^{-4} \\ -2.07 \times 10^{-4} & 2.07 \times 10^{-3} \end{bmatrix}$$

and the normality test statistic is 37.212.

Remark 5.1. Table 6 lists the BIC values of the model (b) for respective lags. The ARMA lags (1,0) might be a reasonable candidate, but the BIC of ARMA lags (0,1) is as small as that of ARMA(1,0). Since the same set of lags is selected for model (a) by the BIC, we consequently selected the model ARMA(0,1).

Table 6: The BIC-values under the null hypothesis

AR order	MA order	BIC
1	0	-1.5895
0	1	-1.5875
1	1	-1.5211
2	0	-1.5067
0	2	-1.4950

Figures 7 and 8 respectively exhibit the ratio series of the Japanese call rate $r(t)$ and Tokyo stock price index $TOPIX(t)$ together with the residuals for the model (a), where notable heteroscedasticity in the residuals is observed.

Table 7 lists the results of Monte-Carlo Wald test for various δ and M. The test rejects the model (b) against the model (a) for every (δ, M), and so we may conclude that model (a) is favoured to model (b). The test shows that it is not suitable to assume that the bivariate series of the PER and the growth rate of the interest rate are assumed to be normally distributed; namely the test indicates these are not log-normally distributed. Characteristically, Doornik and Hansens' normality test also indicates that model (a) is superior to model (b).

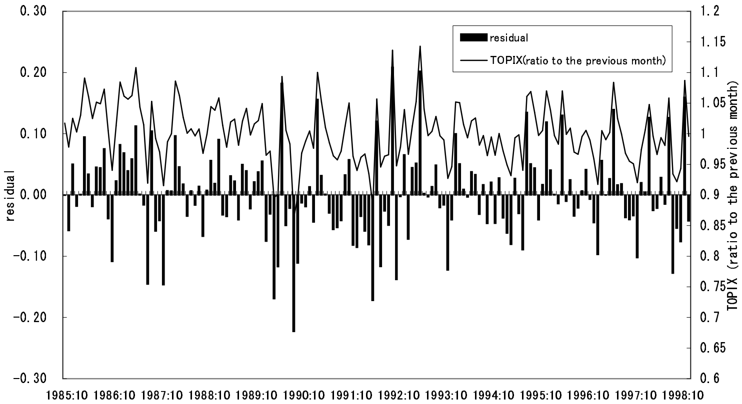

Figure 7: TOPIX (the original ratio series and the residuals)

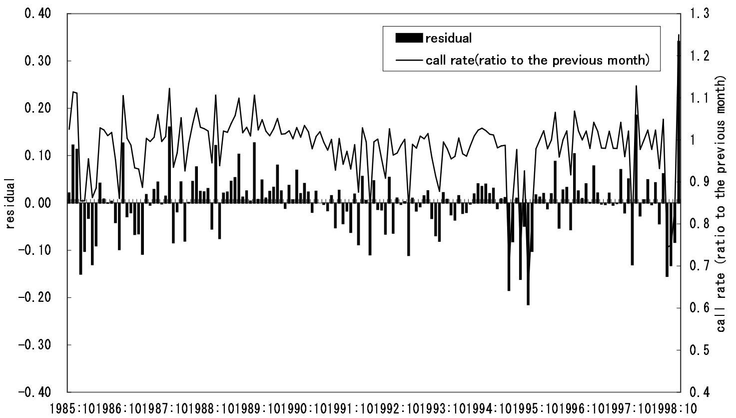

Figure 8: Call Rate (the original ratio series and the residuals)

Table 7: Monte-Carlo Wald test statistic

Modified Box-Cox transformation parameter δ	M	Iteration successful (out of 500)	Wald test statistic
0.85	1000	442	15.980**
0.90	1000	441	12.696**
0.10	110	462	16.743**

** : 1% significance

6 Conclusion

In this paper, we dealt with transformation linear stationary processes. We exhibit the method of finding the Gaussian ARMA model of best fit to data by using a modified Box-Cox transformation. Especially we show that the assumption of the Gaussian ARMA model is not appropriate for the bivariate series of the PER and the growth rate of the interest rate.

Franses and McAleer (1998) investigate the Box-Cox transformation on the augmented Dickey-Fuller regression. They however, leave open the issue of what is the most appropriate non-linear transformation. Recently de Jong (2003) discusses the regression estimation of the logarithmically transformed unit-root processes. The extension to non-stationary process of our approach remains open. Yang (2006) proposes a modified family of power transformation for positive x

$$h(x, \lambda) = \begin{cases} (x^{\lambda} - x^{-\lambda})/2\lambda & \lambda \neq 0 \\ \log x & \lambda = 0 \end{cases}$$

which removes the bound of the sample space in the original Box-Cox transformation.

Furthermore, there are two problems remaining to be investigated. If the transformation linear process (3.1) is autoregressive, the asymptotic result of Theorem 3.1 should be reduced to the standard one as in Davidson and MacKinnon (1984). Since the Whittle likelihood reduces to the likelihood Davidson and MacKinnon dealt with except for asymptotically negligible terms, the equivalence is anticipated, but the rigorous proof $\Psi = \Phi$ is required. Another problem of interest is how to evaluate numerically the terms given in formulas of Lemmas 3.1 and 3.2 so that numerical comparison of the asymptotic covariance matrix with the standard one can be made. Spectral densities involved in those formulas could be estimated by means of Monte-Carlo generated $y(t)$'s, but no practicable numerical procedure is yet available.

A. Appendix

Proof of Lemma 3.1. The relation (3.4) follows from Hosoya and Taniguchi (1982, p.137). Since $2\pi|1_T(\omega)|^2$ is nothing but the Fejér kernel, the relation (3.7) is the consequence of the equalities:

$$\lim_{T\to\infty} E\left[\frac{\partial^2 Q_T}{\partial\mu_l\partial\mu_n}\right] = \lim_{T\to\infty} E\left[\frac{1}{2\pi}tr\int_{-\pi}^{\pi} h(\omega,\theta^0)|1_t(\omega)|^2 \left\{e_l(m)e_n(m)' + e_n(m)e_l(m)'\right\} d\omega\right]$$

$$= \frac{1}{2\pi}\{h_{nl}(0,\theta^0) + h_{ln}(0,\theta^0)\}.$$

Similarly the property of the Fejér kernel implies (3.5) and (3.8). As for (3.9),

$$\lim_{T\to\infty} E\left[\frac{\partial^2 Q_T}{\partial\lambda_i\partial\lambda_j}\right]$$

$$\begin{aligned} = & \lim_{T\to\infty} E[\frac{1}{2\pi}tr\int_{-\pi}^{\pi} h(\omega,\theta^0)\{\tilde{F}^{(i,j)}(\tilde{F}-\mu 1_T)^* + (\tilde{F}^{(i)} - E(\tilde{F}^{(i)}))(\tilde{F}^{(j)} - E(\tilde{F}^{(j)}))^* \\ & +(\tilde{F}^{(j)} - E(\tilde{F}^{(j)}))(\tilde{F}^{(i)} - E(\tilde{F}^{(i)}))^* + (\tilde{F}-\mu 1_T)\tilde{F}^{(i,j)*}\}]d\omega \\ & +\frac{1}{2\pi}tr[h(0,\theta^0)\{E(z^{(i)}(1))E(z^{(j)}(1))' + E(z^{(j)}(1))E(z^{(i)}(1))'\}] + E(z^{[i,j]}(1)) \end{aligned}$$

where the first member on the right hand side is equal to

$$\frac{1}{2\pi}tr\int_{-\pi}^{\pi} h(\omega,\theta^0)[c^{(i,j),\cdot}(\omega) + c^{(i),(j)}(\omega) + c^{(j),(i)}(\omega) + c^{\cdot,(i,j)}(\omega)]d\omega.$$

Lastly (3.6) follows from the relations:

$$\lim_{T\to\infty} E\left[\frac{\partial^2 Q_T}{\partial\theta_\alpha\partial\lambda_j}\right] = \lim_{T\to\infty} E\left[\frac{1}{2\pi}tr\int_{-\pi}^{\pi} h^{(\alpha)}\{\tilde{F}^{(j)}(\tilde{F}-\mu 1_T)^* + (\tilde{F}-\mu 1_T)\tilde{F}^{(j)*}\}d\omega\right]$$

$$= \frac{1}{2\pi}tr\int_{-\pi}^{\pi} h^{(\alpha)}\{c^{(j),\cdot}(\omega) + c^{\cdot,(j)}(\omega)\}d\omega. \quad \square$$

Before proceeding to the proof of Lemma 3.2, it is useful to introduce some concepts related to central limit theorem for strictly stationary processes and to establish Lemma A below. In view that the process $\{y(t)\}$ generated by (3.1) is strictly stationary and so are $\{g(y_t)\}$ for any measurable function g, the proof of the central limit theorem relies essentially upon the regularity concepts defined for strictly stationary processes.

Set $z(t) \equiv F(y(t), \lambda^0) - \mu^0$ when $y(t)$ is generated by (3.1) for $\theta = \theta^0$ and $\mu = \mu^0$; then evidently the process $\{z(t)\}$ is a strictly stationary Gaussian process. To a strictly stationary process $\{w(t)\}$, there is associated a set of shift transformations $U = \{U_t, t \in \mathbb{Z}\}$ such that $U_t w(s) = w(t+s)$ for any $s \in \mathbb{Z}$. The process $\{w(t)\}$ is said to be metrically transitive if for any set A invariant with respect to the transformation set U, it holds either $Pr(A) = 1$ or $Pr(A) = 0$; see Rozanov (1967). If $\{w(t)\}$ is metrically transitive and $E|w(t)| < \infty$, then $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} w(t) = E(w(0))$ with probability 1; namely a strong law of large number holds. Denote by $\mathcal{F}_{-\infty}^{t}$ and $\mathcal{F}_{t+\tau}^{\infty}$ the σ-fields generated by $\{w(s),\ s \leq t\}$ and $\{w(s),\ s \geq t+\tau\}$ respectively. Set $\alpha_w(\tau) = \sup|P(A \cap B) - P(A)P(B)|$, where the supremum is over $A \in \mathcal{F}_{-\infty}^{t}$ and $B \in \mathcal{F}_{t+\tau}^{\infty}$. The process $\{w(t)\}$ is said be completely regular if $\lim_{\tau\to\infty} \alpha_w(\tau) = 0$. If the assumption A (iii) holds, $\{z(t)\}$ is a metrically transitive and completely regular Gaussian process. Furthermore $\alpha_z(\tau) = O(\tau^{-3})$ as is shown in Lemma A below.

Denote $H(t-)$ and $H(t+)$ be the closed linear subspaces spanned by $\{z(s), s \leq t\}$ and $\{z(s), s \geq t\}$ in the Hilbert space of all random variables with second moment defined on the probability space on which $\{z(t), t \in \mathbb{Z}\}$ is defined. Also denote by $\mathcal{F}_z(t-)$ and $\mathcal{F}_z(t+)$ the σ-fields generated by $\{z(s), s \leq t\}$ and $\{z(s), s \geq t\}$ respectively. Let the notation $\xi \in \mathcal{F}_z(t-)$ indicate that ξ is measurable with respect to $\mathcal{F}_z(t-)$. Define the index ρ by

$$\rho[\mathcal{F}_z(t-), \mathcal{F}_z((t+\tau)+)] = \sup_{\xi_1 \in \mathcal{F}_z(t-), \xi_2 \in \mathcal{F}_z((t+\tau)+)} corr(\xi_1, \xi_2)$$

where *corr* denotes the ordinary correlation coefficient. In view of Theorems 10.1 and 10.2 of Rozanov (1964, p.181), since the process $\{z(t)\}$ is Gaussian,

$$\rho[\mathcal{F}_z(t-), \mathcal{F}_z((t+\tau)+)] = \rho[H_z(t-), H_z((t+\tau)+)] \geq \alpha_z(\tau). \tag{A.1}$$

Moreover Rosanov's Lemma 10.2 (1964, p.182) gives the following important result:

$$\sup_{f,g} E\{f(z(t))g(z(t+s))\} = corr(z(t), z(t+s)) \tag{A.2}$$

where f and g real-valued functions such that $E(f(z(t))) = E(g(z(s))) = 0$ and $Var(f(z(t))) = Var(g(z(s))) = 1$; namely, correlation between transformed Gaussian variables does not exceed the correlation between original Gaussian variables.

Lemma A. *If Assumption A(iii) holds, then* $\alpha_F(\tau) = O\left(\tau^{-3}\right)$.

Proof. Let $\{a_j, j = 0, 1, 2, \cdots,\}$ be a sequence of real numbers for which $\sum_{j=0}^{\infty} j|a_j| < \infty$; then it is easy to see that there is M such that $j^2|a_j| < M$ for all j. Set $\nu(t) = \sum_{j=0}^{\infty} a_j \varepsilon_k(t-j)$, for any $1 \leq k \leq m$. Then we have

$$|Cov(\nu(t), \nu(t+\tau)| \leq \sum_{j=\tau}^{\infty} a_j^2,$$

where the right-hand side member is of order

$$O(\sum_{j=\tau}^{\infty} j^{-4}) = O(\tau^{-3}).$$

Consequently, $\rho[H_F(t-), H_F((t+\tau)+)] = O(\tau^{-3})$. The lemma follows then from (A.1). $\square$.

The next theorem was given by Rosanov (1967, p191).

Theorem A: *Suppose a strictly stationary process* $\{\xi(t)\}$ *satisfies the conditions* $\alpha_\xi(\tau) = O(\tau^{-1-\delta})$ *and* $E\|\xi(0)\|^{2+\varepsilon} < \infty$ *for some* $\delta > 0$ *and* $\varepsilon > 4/\delta$ *and suppose the spectral density matrix is bounded and continuous and nondegenerate at zero, then the central limit theorem is applicable to* $\{\xi(t)\}$.

Proof of Lemma 3.2. (1) At first, we deal with the relations (3.11)-(3.16). The proof for the relation (3.11) proceeds quite in parallel to Walker (1964, p.371-6). Since $\tilde{F}(\omega, \lambda^0)$ is normally distributed, we have (3.12). The relation $\Phi_{\mu_l\mu_n} = \Psi_{\mu_l\mu_n}$ follows from:

$$\Phi_{\mu_l\mu_m} = \lim_{T\to\infty} TE\left[tr\int_{-\pi}^{\pi} he_l(m)1_T\{\tilde{F} - 1_T\mu\}^* d\omega \cdot tr\int_{-\pi}^{\pi} he_n(m)1_T\{\tilde{F} - 1_T\mu\}^* d\omega\right]$$
$$= 2\pi\{h_{ln}(0, \theta^0) + h_{nl}(0, \theta^0)\}.$$

We have (3.13) by directly evaluating

$$\Phi_{\theta_\alpha\lambda_i} = \lim_{T\to\infty} TCov\left[tr\int_{-\pi}^{\pi} h^{(\alpha)}(\tilde{F} - 1_T\mu)(\tilde{F} - 1_T\mu)^* d\omega,\right.$$
$$\left. tr\int_{-\pi}^{\pi} h\{(\tilde{F}^{(i)} - E\tilde{F}^{(i)}))(\tilde{F} - 1_T\mu)^* + (\tilde{F} - 1_T\mu)(\tilde{F}^{(i)} - E(\tilde{F}^{(i)})^*)\}d\omega + \frac{1}{\sqrt{2\pi T}}\int_{-\pi}^{\pi} \tilde{J}^{(i)} d\omega\right].$$

where $\tilde{F}, \tilde{F}^{(i)}, \tilde{J}^{(j)}$ are all evaluated at $\lambda = \lambda^0$. By means of the matrices $H^3(\omega)$ and $H^4(\omega)$, we have (3.16) by evaluating

$$\Phi_{\lambda_i,\lambda_j} = \lim_{T\to\infty} TCov\left[tr\int_{-\pi}^{\pi} h\{(\tilde{F}^{(i)} - E(\tilde{F}^{(i)})\}(\tilde{F} - 1_T\mu)^* + (\tilde{F} - 1_T\mu)(\tilde{F}^{(i)} - E(\tilde{F}^{(i)}))^*\}d\omega\right.$$

$$+\frac{1}{\sqrt{2\pi T}}\int_{-\pi}^{\pi}\tilde{J}^{(i)}d\omega,$$

$$\left. tr\int_{-\pi}^{\pi} h\{(\tilde{F}^{(j)} - E(\tilde{F}^{(j)})(\tilde{F} - 1_T\mu)^* + (\tilde{F} - 1_T\mu)(\tilde{F}^{(j)} - E(\tilde{F}^{(j)}))^*\}d\omega + \frac{1}{\sqrt{2\pi T}}\int_{-\pi}^{\pi}\tilde{J}^{(j)}d\omega\right].$$

Finally we have (3.15) by directly evaluating

$$\Phi_{\mu_i,\lambda_j} = \lim_{T\to\infty} TCov\left[tr\int_{-\pi}^{\pi} he_m(i)(\tilde{F} - 1_T\mu)^*,\right.$$

$$\left. tr\int_{-\pi}^{\pi} h\{(\tilde{F}^{(j)} - E(\tilde{F}^{(j)})(\tilde{F} - 1_T\mu)^* + (\tilde{F} - 1_T\mu)(\tilde{F}^{(j)} - E(\tilde{F}^{(j)}))^*\}d\omega + \frac{1}{\sqrt{2\pi T}}\int_{-\pi}^{\pi}\tilde{J}^{(j)}d\omega\right].$$

(2) Secondly, asymptotic normality is proved as follows. Denote by $S_t(a)$ the linear combination of $\{P_t(F_{i_1}, F_{i_2}, \tau_1), P_t(F_{i_3}, F_{i_4}^{(j_1)}, \tau_2), P_t(1, F_{i_5}, 0), P_t(1, J_{i_6}^{(j_2)}, 0); i \le i_1, \cdots, i_6 \le m,\ 1 \le j_1, j_2 \le m, -T_1 \le \tau_1, \tau_2 \le T_1\}$ with coefficient vector a; see the definition of P_t in assumption B(iii). As in Walker (1964) and Hosoya (1989b, 1997), the asymptotic normality of $T^{1/2}\left\{\frac{\partial Q_T(\xi^0)}{\partial\xi} - E\frac{\partial Q_T(\xi_0)}{\partial\xi}\right\}$ holds if for every $a \ne 0$, $\sum_{t=1}^{T}\{S_t(a) - E(S_t(a))\}/\sqrt{T}$ has a limiting normal distribution by means of the Bernstein lemma. Thanks to Assumption B (i)(3) and Gaussianity of F, the moment condition is satisfied, whereas the spectral condition is assumed in B(iii). Since $S_t(a)$ is a function of $F(t, \lambda^0)$, it follows form (A.2) and Lemma A that $\alpha_s(\tau) = O(\tau^{-(1+\delta)})$. Hence Theorem A is able to be applied to the process $\{S_t(a)\}$. Finally, the asymptotic normality of $\sqrt{T}\partial Q_T(\xi^0)/\partial\xi$ follows from Assumption C (ii) and Remark 3.1. □

Proof of Theorem 3.1. (1) By a train of arguments parallel to Walker (1964, pp.367-370), the consistency proof is carried out in view of Assumption C (i) by showing that for all $\xi_1, \xi_2 \in \Xi$ such that $|\xi_2 - \xi_1| < \delta$ there exists $H_{\delta,T}(y, \xi_1)$ such that $|Q_T(\xi_2) - Q_T(\xi_1)| < H_{\delta,T}(y, \xi_1)$ and $\lim_{\delta\to 0} EH_{\delta,T}(y, \xi_1) = 0$ uniformly in T and also $\lim_{T\to\infty} VarH_{\delta,T}(y, \xi_1) = 0$ for fixed δ. Let l be the dimension of the vector ξ and $Q_T^{(i)}$ the partial derivative with respect to ξ_i. If $|\xi_2 - \xi_1| < \delta$, we have $|Q_T(\xi_2) - Q_T(\xi_1)| \le$

$\delta \sum_{i=1}^{l} |Q_T^{(i)}(\xi_\kappa)|$ where $\xi_\kappa = \kappa\xi_2 + (1-\kappa)\xi_1$ $(0 < \kappa < 1)$, but Assumptions A (i) and B (ii) guarantee that $E \sum_{i=1}^{l} \sup_{\xi \in B_\delta(\xi_1)} |Q_T^{(i)}(\xi)|$ is bounded and also it follows from Assumption B (ii) in view of Hosoya (1997, p.131, Lemma 3.3) that $Var \sum_{i=1}^{l} \sup_{\xi \in B_\delta(\xi_1)} |Q_T^{(i)}(\xi)|$ tends to 0 as $T \to \infty$, where $B_\delta(\xi_1)$ denotes the ball of center ξ_1 and radius δ. Hence if we set $H_{\delta,T}(y, \xi_1) = \delta \sum_{i=1}^{l} \sup_{\xi \in B_\delta(\xi_1)} |Q_T^{(i)}(\xi)|$, the consistency of $\hat{\xi}$ follows. (2) The limiting distribution of $\sqrt{T}(\hat{\xi} - \xi^0)$ is derived by means of the standard technique of the Taylor expansion of $\partial Q(\hat{\xi})/\partial \xi = 0$ around $\xi = \xi^0$ in view that $\hat{\xi}$ is consistent and that $\partial^3 Q_T(\xi)/\partial \xi \partial \xi' \partial \xi_k$ is bounded in probability in a neighborhood of ξ^0 thanks to Assumption B (ii)(1) and (2). Thus the theorem is concluded. $\square$

References.

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Proc. 2nd Int. Symp. Information Theory (eds. Petriv, B.N. and Czaki, F.)*, 267-281, Academiai Kiado, Budapest.

[2] Amemiya, T. (1977). The Maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model, *Econometrica*, 45, 955-968.

[3] Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis(second edition)*, Wiley, New York.

[4] Bickel, P.J. and Doksum, K.A. (1981). An analysis of transformations revisited, *Jour. Amer. Statist. Assoc.*, 76, 296-311.

[5] Box, G.E.P and Cox, D.R (1964). An analysis of transformations, *J. Roy. Statist. Soc. B.*26, 211-252.

[6] Box, G.E.P. and Tidwell, Paul W. (1962). Transformation of the independent variables, *Technometrics*, 4, 531-550.

[7] Davidson, R. and MacKinnon, J.G. (1984). Model specification tests based on artificial linear regressions, *International Economic Review*, 25, 485-502.

[8] Davidson, R. and MacKinnon, J.G. (1990). Specification tests based on artificial regressions, *Journal of the American Statistical Association*, 85, 220-227.

[9] Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, New York.

[10] de Jong, R.M. (2003). Logarithmic spurious regressions, *Economics Letters*, 81, 13-21.

[11] Doornik, J.A, and Hansen, H. (1994). An omnibus test for univariate and multivariate normality. mimeo. Nuffield College, Oxford.

[12] Efron, B and Tibshirani, R J.(1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

[13] Franses, P.H. and McAleer, M. (1998). Testing for unit roots and non-linear transformations, *Journal of Time Series Analysis*, 19, 147-164.

[14] Hannan, E.J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order, *Biometrika*, 69, 81-94.

[15] Hannan, E.J and Kavalieris, L. (1984). A method for autoregressive-moving average estimation, *Biometrika*, 71, 273-280.

[16] Hosoya, Y. (1989a). Hierarchical statistical models and a generalized likelihood ratio test. *J. Roy. Statist. Soc. B.* 51, 435-447.

[17] Hosoya, Y. (1989b). The bracketing condition for limit theorems on stationary linear processes. *Ann. Statist.* 17, 401-418.

[18] Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models, *Ann. Statist.* 25, 105-137.

[19] Hosoya, Y. (2002). *Statistical Evidence and Its Interpretations (enlarged edition)* (in Japanese), Makino Publishing Company, Tokyo.

[20] Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, 10, 132-153: Correction, 1993, *Ann. Statist.* 21, 1115-1117.

[21] Ibaraki, T. and Fukushima, M. (1991). *FORTRAN 77 Optimization Programming* (in Japanese), Iwanamishoten, Tokyo.

[22] Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation, *Biometrika*, 36, 149-176.

[23] MacKinnon, J.G. and Magee, L. (1990). Transforming the dependent variable in regression models, *International Economic Review*, 31, 315-339.

[24] Reinsel, G.C. (1997). *Elements of Multivariate Time Series Analysis (second edition)*, Springer, New York.

[25] Rosanov, Yu.A. (1967). *Stationary Random Processes*, Holden-Day, San Francisco.

[26] Walker, A.M. (1964). Asymptotic properties of least-square estimates of parameters of the spectrum of a stationary non-deterministic time series. *Journal of Australian Mathematical Society*, 4, 363-384.

[27] Yang, Z. (2006). A modified family of power transformations. *Economics Letters*, 92, 14-19.

[28] Zarembka, P. (1974). Transformation of variables in econometrics, in P. Zarembka, *Frontiers in Econometrics*, Academic Press, New York.