数理解析研究所講究録
916 巻 1995 年 169-179

Discriminant Analysis for Regression Models with Long-memory Linear Disturbances

阪大基礎工 張 国強 (G. Zhang)

We shall consider the problems of classifying an observation from regression model with stationary long-memory linear disturbances into one of two populations described by the mean functions of the model. We mainly use the log-likelihood ratio(LR) as discriminant statistic which is optimal in the sense of minimizing the misclassification probabilities. Then we propose a new discriminant statistic related to the Best Linear Unbiased Estimator(BLUE) and the Least Squares Estimator(LSE). Finally, we discuss the asymptotics of quatratic discriminant statistics in the non-Gaussian case.

Keywords: Discriminant analysis, Regression model, Long-memory linear disturbances, Misclassification probability, BLUE, LSE.

1 INTRODUCTION

There are a variety of problems related to regression analysis in time series. One is that of discriminanting between two regressions. For simplicity, let us consider the two hypotheses

$$H_j: \quad Y_t = \sum_{i=1}^{k} X_i(t)\beta_i^{(j)} + \varepsilon_t, \qquad j = 1, 2. \tag{1.1}$$

where $X_1(t), \cdots, X_k(t)$ are nonstochastic regressors, $\boldsymbol{\beta}^{(j)} = (\beta_1^{(j)}, \cdots, \beta_k^{(j)})'$ are vectors of regression parameters, and $\{\varepsilon_t\}$ is assumed to be a zero mean Gaussian stationary process with spectral density $f(\lambda)$. Assume that the covariance matrix $\boldsymbol{\Sigma}$ of the disturbances $\{\varepsilon_t\}$ is nonsingular. Then the two p.d.f.s corresponding to H_1 and H_2 are

$$p_j(\boldsymbol{Y}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{(j)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{(j)})\}, \quad j = 1, 2 \tag{1.2}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \boldsymbol{X}_r = \begin{pmatrix} X_r(1) \\ \vdots \\ X_r(n) \end{pmatrix}, \qquad \boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_k).$$

In this paper, we consider the problem of classifying $\boldsymbol{Y}$ from H_1 or H_2 into one of two populations H_1 and H_2. We assign $\boldsymbol{Y}$ to H_1 if $\boldsymbol{Y}$ falls into region R_1; otherwise we assign it to H_2, where R_1 and R_2 are exclusive and exhaustive regions in $\boldsymbol{R}^n$. Let the probability of misclassifying $\boldsymbol{Y}$ from H_i into H_j be $P(j|i) = \int_{R_j} p_i(\boldsymbol{Y})d\boldsymbol{Y}$. We choose R_1 and R_2 to minimize $P(2|1) + P(1|2)$. Then it is known that the regions based on LR

$$R_j = \left\{\boldsymbol{Y} : (-1)^{j-1}\mathrm{LR} > 0\right\}, \quad j = 1, 2 \tag{1.3}$$

give the optimal classification, where

$$\mathrm{LR} = (\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y} - \frac{1}{2}\boldsymbol{X}(\boldsymbol{\beta}^{(2)} + \boldsymbol{\beta}^{(1)})\right). \tag{1.4}$$

In most of time series, the dependence between distant observations is quite weak, and we usually use ARMA and other stationary short-memory processes to discribe their characteristics, and a series of works for the discriminant analysis of such time series have been done. In the problems of discriminanting two Gaussian processes by linear fultering, Shumway and Unger (1974) gave certain spectral approximations of Kullback-Leibler information and J-divergence. Zhang and Taniguchi (1994) used an approximation of LR as a classification statistic for the non-Gaussian vector time series classification problems. Zhang (1994) discussed the higher-order asymptotic theories of discriminant analysis for stationary ARMA processes. Shumway (1982) gave an extensive review of various discriminant problems in time series.

However, time series classification problems are not restricted to the physical sciences, but occur under many and varied circumstances in the other fields. For instance, in many empirical time series, especially those of economics and geophysics, the dependence between distant observations is so strong that ARMA models are unable to express the spectral densities of low frequencies adequately. For such long-memory(or strongly dependent) processes, Adenstedt (1974) found that for a large class of spectral densities, the variance of the BLUE for the mean is seen to depend asymptotically only on the spectral density near the origin. Giraitis and Surgailis (1990) gave a central limit theorem for quadratic forms in strongly dependent linear variables and applied it to prove the asymptotic normality of Whittle's estimator of parameters of strongly dependent linear sequences. Yajima (1988, 1991) considered estimation of a regression model with long-memory stationary errors by LSE and BLUE, and gave their asymptotic properties.

In Section 2, we use LR as a discriminant statistic for classifying two regression models (1.1) with short-memory stationary distubances. We show that LR is a consistent classification statistic under Grenander's conditions on $\boldsymbol{X}_t$.

In Section 3, we use LR as a classification statistic for regression models (1.1) with stationary Gaussian long-memory disturbances and discuss its asymptotic properties of misclassification probabilities. We study a classification statistic which is based on a linear combination of LSE and BLUE.

In Section 4, we propose a discriminant statistic of quadratic form for a simple regression model with stationary "non-Gaussian" long-memory disturbances. Then we elucidate some results which are different from those for short-memory disturbances.

2 SHORT-MEMORY CLASSIFICATION

In this section, we summarize the discriminant theory of regression model (1.1) when $\{\varepsilon_t\}$ is a zero mean short-memory stationary process with positive and continuous spectral density $f(\lambda), \lambda \in [-\pi, \pi]$.

Following Grenander (1954) we assume that the nonstochastic regressors $\{X_r(t)\}$ possess the following properties:

(G.1) $\chi_r(n) = \sum_{t=1}^{n} |X_r(t)|^2 \to \infty, \quad \text{as} \quad n \to \infty, \quad \text{for } r = 1, \cdots, k,$

(G.2) $\lim_{n\to\infty} \frac{\chi_r(n+1)}{\chi_r(n)} = 1, \text{ for } r = 1, \cdots, k,$

(G.3) $\lim_{n\to\infty} \frac{\sum_{t=1}^{n} X_r(t+h)X_s(t)}{\{\chi_r(n)\chi_s(n)\}^{1/2}} = R_{rs}(h), \quad \text{(say)}$

exists for each $r, s = 1, \cdots, k$, and all integers $h \geq 0$. Defining $X_r(t) = 0, t < 0$, (for each $r = 1, \cdots, k$), the above definition of $R_{rs}(h)$ can now be extended to all integer values of h, and we can write

$$\boldsymbol{R}(h) = \{R_{rs}(h)\}, \qquad h = 0, \pm 1, \pm 2, \cdots,$$

i.e. $\boldsymbol{R}(h)$ is a $q \times q$ matrix with entry $R_{rs}(h)$ in the rth row and sth column.

(G.4) The matrix $\boldsymbol{R}(0)$ is nonsingular.

It is not difficult to show that $\boldsymbol{R}'(h) = \boldsymbol{R}(-h)$ (for each h), and that $\boldsymbol{R}(0)$ is positive definite(cf Grenander and Rosenblatt (1957)). It then follows that, there exists a Hermitian matrix function, $\boldsymbol{M}(\lambda) = \{M_{rs}(\lambda)\}$ such that $R_{rs}(h)$ admits a spectral representation of the form,

$$R_{rs}(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dM_{rs}(\lambda)$$

where matrix form $\boldsymbol{M}(\lambda)$ has positive semi-definite matrix increments. Let

$$\boldsymbol{D}_n = \text{diag}\{\chi_1^{1/2}(n), \cdots, \chi_k^{1/2}(n)\}$$

i.e. the $\boldsymbol{D}_n$ is the diagonal matrix with entry $\chi_r^{1/2}(n)$ in the rth row and column.

[Theorem 2.1] Suppose that the spectral density $f(\lambda)$ is a positive and continuous function on $[-\pi, \pi]$. For classifying $\boldsymbol{Y}$ into one of two hypotheses H_1 and H_2, we use the classification rule (1.3) based on LR. Then under Assumptions (**G.1**)-(**G.4**), the misclassification probabilities satisfy

$$\lim_{n\to\infty} P_{\text{LR}}(2|1) = \lim_{n\to\infty} P_{\text{LR}}(1|2) = 0, \quad \text{for} \quad \boldsymbol{\beta}^{(1)} \not\equiv \boldsymbol{\beta}^{(0)}. \tag{2.1}$$

We put the proof of theorems and propositions in Section 5 if they are not straightforward.

Next, to evaluate the goodness of classification statistic LR, we consider the case when $\boldsymbol{\beta}^{(2)}$ is contiguous to $\boldsymbol{\beta}^{(1)}$:

$$\begin{aligned} H_1 &: \boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}; \\ H_2 &: \boldsymbol{\beta}^{(2)} = \boldsymbol{\beta} + \boldsymbol{D}_n^{-1}\boldsymbol{H}, \end{aligned} \tag{2.2}$$

where $\boldsymbol{H} = (H_1, \cdots, H_k)'$ is a constant vector. Then

[Theorem 2.2] Suppose that $f(\lambda)$ is a positive and continuous function on $[-\pi, \pi]$ and that **(G.1)-(G.4)** hold. Then under contiguous assumption (2.2), the misclassification probabilities of LR are evaluated as

$$\lim_{n\to\infty} P_{\mathrm{LR}}(2|1) = \lim_{n\to\infty} P_{\mathrm{LR}}(1|2) = \Phi\left(-\frac{1}{2}\sqrt{\boldsymbol{H}'\boldsymbol{\Omega}\boldsymbol{H}}\right). \tag{2.3}$$

where

$$\boldsymbol{\Omega} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f^{-1}(\lambda) d\boldsymbol{M}(\lambda).$$

Let us consider the case of polynomial regression.

[Example 2.1] Suppose that $X_r(t) = t^{r-1}, r = 1, \cdots, k$ and $\{\varepsilon_t\}$ is a stationary process with zero mean and spectral density $f(\lambda)$ that is a positive and continuous function on $[-\pi, \pi]$. Then $\boldsymbol{M}(\lambda)$ has only a jump at $\lambda = 0$ of $\boldsymbol{M}_0 = \left\{\frac{[(2r-1)(2s-1)]^{\frac{1}{2}}}{(r+s-1)}\right\}_{r,s=1,\cdots,k}$ (e.g. Anderson(1971)). Thus under contiguous assumption (2.2), the misclassification probabilities of LR are evaluated as

$$\lim_{n\to\infty} P_{\mathrm{LR}}(2|1) = \lim_{n\to\infty} P_{\mathrm{LR}}(1|2) = \Phi\left(-\frac{1}{2}\sqrt{\boldsymbol{H}'\boldsymbol{\Omega}_0\boldsymbol{H}}\right). \tag{2.4}$$

where

$$\boldsymbol{\Omega}_0 = \left\{\frac{[(2r-1)(2s-1)]^{\frac{1}{2}}}{2\pi f(0)(r+s-1)}\right\}_{r,s=1,\cdots,k}.$$

3 LONG-MEMORY CLASSIFICATION

In this section, we consider the discriminant problem of regression model (1.1) when $\{\varepsilon_t\}$ is a Gaussian long-memory linear process of the form

$$\varepsilon_t = \sum_{j=0}^{\infty} \alpha_j \omega_{t-j}, \quad \omega_t \sim N(0,1), \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty. \tag{3.1}$$

We introduce an important Lemma that plays key role in the following discussion. Then we show the discriminant properties of LR and a new classification statistic based on LSE and BLUE. Let

$$g_{jl}^n(\lambda) = \frac{\left(\sum_{t=1}^n X_j(t)e^{-it\lambda}\right)\left(\sum_{t=1}^n X_l(t)e^{it\lambda}\right)}{2\pi\chi_r(n)\chi_s(n)}, \quad M_{jl}^n(\lambda) = \int_{-\pi}^{\lambda} g_{jl}^n(\omega) d\omega, \tag{3.2}$$

Besides **(G.1)-(G.4)**, we add the following assumptions:

(G.5) $X_r(t) = t^{r-1}$, for $r = 1, \cdots, p,\ 1 \le p \le k$,

(G.6) $\{\varepsilon_t\}$ has the spectral density $f(\lambda) = \frac{f^*(\lambda)}{|1-e^{i\lambda}|^{2d}}, 0 < d < \frac{1}{2}$, where $f^*(\lambda)$ is a positive and continuous function on $[-\pi, \pi]$.

(G.7) For every $\delta > 0$ there exists c such that $\int_{|\lambda| \le c} f(\lambda) dM_{jj}^n(\lambda) < \delta$, for every n and $j = p+1, \cdots, k$,

(G.8) $\frac{\max_{1 \le t \le n} X_r^2(t)}{\chi_r^{1/2}(n)} = o(n^{-\mu})$, $r = p+1, \cdots, k$, for some $\mu > 1 - 2d$.

From Assumption **(G.5)**, we have

$$\chi_r(n) = \sum_{t=1}^{n} t^{2(r-1)} \sim \frac{n^{2r-1}}{2r-1}, \qquad r = 1, \cdots, p, \tag{3.3}$$

so that Assumptions **(G.1)** and **(G.2)** are certainly satisfied. Define

$$\widetilde{\boldsymbol{D}}_n = \operatorname{diag}\left(\frac{\chi_1^{1/2}(n)}{n^d}, \cdots, \frac{\chi_p^{1/2}(n)}{n^d}, \chi_{p+1}^{1/2}(n), \cdots, \chi_k^{1/2}(n) \right), \tag{3.4}$$

The following Lemma follows from Yajima (1991).

[Lemma 3.1] Under Assumptions **(G.1)-(G.8)**,

$$\lim_{n \to \infty} \widetilde{\boldsymbol{D}}_n^{-1} \boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{X} \widetilde{\boldsymbol{D}}_n^{-1} = \begin{pmatrix} \boldsymbol{W}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{W}_2 \end{pmatrix} \tag{3.5}$$

and

$$\lim_{n \to \infty} \boldsymbol{D}_n^{-2} \widetilde{\boldsymbol{D}}_n \boldsymbol{X}' \boldsymbol{\Sigma} \boldsymbol{X} \widetilde{\boldsymbol{D}}_n \boldsymbol{D}_n^{-2} = \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}_2 \end{pmatrix} \tag{3.6}$$

where $\boldsymbol{W}_1$, $\boldsymbol{B}_1$ are $p \times p$ matrix with (i,j)th entry

$$w_1^{ij} = \frac{\Gamma(i-d)\Gamma(j-d)\{(2i-1)(2j-1)\}^{\frac{1}{2}}}{2\pi f^*(0)\Gamma(i-2d)\Gamma(j-2d)(i+j-1-2d)}$$

$$b_1^{ij} = \frac{2\pi f^*(0)\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \left[\{(2i-1)(2j-1)\}^{\frac{1}{2}} \int_0^1 \int_0^1 x^{i-1} y^{j-1} |x-y|^{2d-1} dx dy \right]$$

and $\boldsymbol{W}_2$, $\boldsymbol{B}_2$ are $(k-p) \times (k-p)$ matrix with (i,j)th entry

$$w_2^{ij} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^{-1}(\lambda) d\boldsymbol{M}_{i+p\ j+p}(\lambda), \qquad b_2^{ij} = 2\pi \int_{-\pi}^{\pi} f(\lambda) d\boldsymbol{M}_{i+p\ j+p}(\lambda)$$

respectively.

Now, to classify $\boldsymbol{Y}$ (with long-memory disturbances) into one of two hypotheses H_1 and H_2, besides using criterion LR, we propose a new classification criterion. Let

$$\widehat{\boldsymbol{\beta}}_\alpha = \alpha \widehat{\boldsymbol{\beta}}_L + (1-\alpha) \widehat{\boldsymbol{\beta}}_B, \qquad (0 \le \alpha \le 1) \tag{3.7}$$

where

$$\widehat{\boldsymbol{\beta}}_L = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{Y}, \quad \widehat{\boldsymbol{\beta}}_B = (\boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{Y} \tag{3.8}$$

are the LSE and BLUE of $\boldsymbol{\beta}$, respectively. We use the discriminant rule: assign Y to H_1 if

$$(\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^{(1)})'(\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^{(1)}) < (\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^{(2)})'(\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^{(2)})$$

to H_2, otherwise. That is, assign Y to H_1 if

$$I_\alpha = (\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})' \left[2\widehat{\boldsymbol{\beta}}_\alpha - (\boldsymbol{\beta}^{(2)} + \boldsymbol{\beta}^{(1)})\right] < 0, \tag{3.9}$$

to H_2, otherwise.

[Theorem 3.1] Under Assumptions **(G.1)-(G.8)**,

$$\left.\begin{array}{ll} \text{(i)} & \lim_{n\to\infty} P_{\mathrm{LR}}(2|1) = \lim_{n\to\infty} P_{\mathrm{LR}}(1|2) = 0 \\ \text{(ii)} & \lim_{n\to\infty} P_{\mathrm{I}_\alpha}(2|1) = \lim_{n\to\infty} P_{\mathrm{I}_\alpha}(1|2) = 0 \end{array}\right\} \text{ for } \boldsymbol{\beta}_1 \not\equiv \boldsymbol{\beta}_2 \tag{3.10}$$

Now, let the regression parameters associated with H_1 and H_2 be

$$\begin{aligned} H_1 &: \boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}; \\ H_2 &: \boldsymbol{\beta}^{(2)} = \boldsymbol{\beta} + \widetilde{\boldsymbol{D}}_n^{-1}\boldsymbol{H}. \end{aligned} \tag{3.11}$$

where $\boldsymbol{H} = (H_1, \cdots, H_k)'$ is a constant vector, then

[Theorem 3.2] Under Assumptions **(G.1)-(G.8)** and under contiguous condition (3.11),

$$\lim_{n\to\infty} P_{\mathrm{LR}}(2|1) = \lim_{n\to\infty} P_{\mathrm{LR}}(1|2) = \Phi\left(-\frac{1}{2}\left\{\boldsymbol{H}'\begin{pmatrix} \boldsymbol{W}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{W}_2 \end{pmatrix}\boldsymbol{H}\right\}^{\frac{1}{2}}\right) \tag{3.12}$$

$$\Phi\left(-\frac{1}{2\sqrt{k_1}}\right) \geq \lim_{n\to\infty} P_{I_\alpha}(2|1) = \lim_{n\to\infty} P_{I_\alpha}(1|2) \geq \Phi\left(-\frac{1}{2\sqrt{k_2}}\right), \tag{3.13}$$

where k_1 and k_2 are chosen so that $k_1\boldsymbol{H}\boldsymbol{H}' \geq \boldsymbol{F}_\alpha \geq k_2\boldsymbol{H}\boldsymbol{H}'$. Here

$$\boldsymbol{F}_\alpha = \alpha^2\boldsymbol{A}\begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}_2 \end{pmatrix}\boldsymbol{A} + (1-\alpha^2)\begin{pmatrix} \boldsymbol{W}_1^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{W}_2^{-1} \end{pmatrix},$$

where

$$\boldsymbol{A} = \lim_{n\to\infty} \widetilde{\boldsymbol{D}}_n(\boldsymbol{X}'\boldsymbol{X})^{-1}\widetilde{\boldsymbol{D}}_n^{-1}\boldsymbol{D}_n^2.$$

[Example 3.1] In Theorem 3.2, if $X_r(t) = 1, r = 1$, Then

$$\begin{aligned} \lim_{n\to\infty} P_{\mathrm{LR}}(2|1) &= \lim_{n\to\infty} P_{\mathrm{LR}}(1|2) = \lim_{n\to\infty} P_{I_0}(2|1) = \lim_{n\to\infty} P_{I_0}(1|2) \\ &= \Phi\left(-\frac{|H|\Gamma(1-d)}{2\sqrt{(1-2d)2\pi f^*(0)\Gamma(1-2d)}}\right), \end{aligned} \tag{3.14}$$

and

$$\lim_{n\to\infty} P_{I_1}(2|1) = \lim_{n\to\infty} P_{I_1}(1|2) = \Phi\left(-\frac{|H|}{2\sqrt{G_f}}\right), \tag{3.15}$$

where

$$G_f = \frac{2\pi f^*(0)\Gamma(1-2d)}{d(1+2d)\Gamma(d)\Gamma(1-d)}.$$

4 NON-GAUSSIAN CLASSIFICATION

In this section, we consider the following stochastic model

$$H_j: \quad Y_t = m^{(j)} + \varepsilon_t, \ \ j = 1, 2, \tag{4.1}$$

where $m^{(j)}, j = 1, 2$ is the constant mean of $\{Y_t\}$ under H_j, and the linear disturbances $\{\varepsilon_t\}$ satisfy

$$\varepsilon_t = \sum_{j=0}^{\infty} \alpha_j \omega_{t-j}, \qquad \sum_{j=0}^{\infty} \alpha_j^2 < \infty, \tag{4.2}$$

where $\{\omega_t\}$ is a sequence of i.i.d. random variables with zero mean and $\mathrm{var}\{\omega_t\} = 1$.

Suppose that $\{\varepsilon_t\}$ has the spectral density

$$f(\lambda) = f^*(\lambda)|1 - e^{i\lambda}|^{-2d}, \qquad 0 < d < 1/2. \tag{4.3}$$

Since we do not assume the Gaussianity of $\{\omega_t\}$, the following condition is imposed.

(G.9) For every $r = 1, 2, \cdots$, the rth order cumulant κ_r^{ω} of $\{\omega_t\}$ exists.

Next we introduce a Toeplitz matrix $\boldsymbol{A} = \{a_{s-l}\}$ with

$$a_\tau = \int_{-\pi}^{\pi} e^{i\tau x} \widehat{a}(x) dx, \tag{4.4}$$

and assume that

(G.10) $\widehat{a}(x) < c|x|^{-2\gamma}, \quad \gamma < \frac{1}{2}$.

(G.11) $\gamma + d < \frac{1}{4}$.

The following lemma is due to Giraitis and Surgailis (1990),

[Lemma 4.1] Let $\boldsymbol{\epsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$ be the n consective stretch of $\{\varepsilon_t\}$ in regression models (4.1) satisfy (4.2). Then under Assumptions **(G.9)-(G.11)**,

$$\frac{\boldsymbol{\epsilon}'\boldsymbol{A}\boldsymbol{\epsilon} - E_j\{\boldsymbol{\epsilon}'\boldsymbol{A}\boldsymbol{\epsilon}\}}{\sqrt{n}} \xrightarrow{\mathcal{L}} N\left(0, \sigma_a^2\right), \ \ j = 1, 2 \tag{4.5}$$

where

$$\sigma_a^2 = 16\pi^3 \int_{-\pi}^{\pi} \left(\widehat{a}(x) f(x)\right)^2 dx + \kappa_4^{\omega} \left(2\pi \int_{-\pi}^{\pi} \widehat{a}(x) f(x) dx\right)^2$$

and κ_4^{ω} is the fourth cumulant of $\{\omega_t\}$.

Now, to classify $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ into one of two hypotheses H_1 and H_2, instead of using LR, we adopt the following rule: assign $\boldsymbol{Y}$ to H_1 if

$$\left[\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} - E_1(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y})\right]^2 < \left[\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} - E_2(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y})\right]^2$$

to H_2, otherwise. Setting $\boldsymbol{m}_j = (m^{(j)}, \cdots, m^{(j)})'$ and $\nabla_j = \boldsymbol{m}_j'\boldsymbol{A}\boldsymbol{m}_j$, without loss of generality, we assume that $\nabla_1 > \nabla_2$, then the classification rule can be written as: assign $\boldsymbol{Y}$ to H_1 if

$$D = 2\boldsymbol{\epsilon}'\boldsymbol{A}\boldsymbol{\epsilon} + 4\boldsymbol{m}_j'\boldsymbol{A}\boldsymbol{\epsilon} + 2\nabla_j - (\nabla_1 + \nabla_2) - 2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) > 0$$

to H_2, otherwise.

[Theorem 4.1] Suppose that the stationary linear disturbances $\{\varepsilon_t\}$ in regression models (4.1) satisfy (4.2). Then under Assumptions **(G.9)-(G.11)**,
(i) If $|a_\tau| = O(\rho^\tau), 0 \le \rho < 1$, (i.e., $\widehat{a}(x) = g(x)$ is a short-memory spectral density), then

$$\frac{D - E_j D}{n^{\frac{1}{2}+d}} \xrightarrow{\mathcal{L}} N\left(0, 16(m^{(j)})^2 G_{fg}\right), \quad j = 1, 2 \tag{4.6}$$

where

$$G_{fg} = \frac{2\pi f^*(0) g^2(0) \Gamma(1-2d)}{d(1+2d)\Gamma(d)\Gamma(1-d)}.$$

(ii) If $\gamma < -\frac{d}{2}$, then

$$\frac{D - E_j D}{n^{\frac{1}{2}}} \xrightarrow{\mathcal{L}} N\left(0, 4\sigma_a^2\right), \quad j = 1, 2 \tag{4.7}$$

[Remark 4.1] Theorem 4.1 holds only for the long-memory ($d \ne 0$) processes. From this theorem, we can see that the distribution of D depends heavily on the selection of γ. When $\gamma = 0$, that is $\widehat{a}(x)$ is a positive and continuous function on $[-\pi, \pi]$, D becomes a robust statistic with respect to non-Gaussianity of $\{\varepsilon_t\}$.

The following theorem and proposition follow from Theorem 4.1.

[Theorem 4.2] Suppose that the stationary linear disturbances $\{\varepsilon_t\}$ in regression models (4.1) satisfy (4.2). Then under Assumptions **(G.9)-(G.11)**,
(i) If $\widehat{a}(x) = g(x)$, where $g(x)$ is a positive and continuous function on $[-\pi, \pi]$, then

$$P_D(2|1) + P_D(1|2) = 1 + \Phi\left(-\frac{n^{-\frac{1}{2}-d}(\nabla_1 - \nabla_2)}{4|m^{(1)}|\sqrt{G_{fg}}}\right) - \Phi\left(\frac{n^{-\frac{1}{2}-d}(\nabla_1 - \nabla_2)}{4|m^{(2)}|\sqrt{G_{fg}}}\right) + o(1), \tag{4.8}$$

(ii) If $\gamma < -\frac{d}{2}$, then

$$P_D(2|1) = P_D(1|2) = \Phi\left(-\frac{n^{-\frac{1}{2}}(\nabla_1 - \nabla_2)}{2\sigma_a}\right) + o(1). \tag{4.9}$$

Next, let us consider the classification effects of D under contiguous condition

$$\begin{cases} H_1: & m^{(1)} = m_0; \\ H_2: & m^{(2)} = m_0 + Hn^{-\frac{1}{2}+d}. \end{cases} \tag{4.10}$$

[Proposition 4.1] Suppose that the stationary linear disturbances $\{\varepsilon_t\}$ in regression models (4.1) satisfy (4.2). Then under Assumptions **(G.9)-(G.11)** and contiguous condition (4.10), if $\widehat{a}(x) = g(x)$, then

$$\lim_{n\to\infty} P_D(2|1) = \lim_{n\to\infty} P_D(1|2) = \lim_{n\to\infty} P_{I_1}(2|1) = \lim_{n\to\infty} P_{I_1}(1|2) = \Phi\left(-\frac{|H|}{2\sqrt{G_f}}\right). \tag{4.11}$$

5 APPENDIX: PROOFS OF THEOREMS

[Proof of Theorem 2.1] Since the mean and variance of LR are

$$E_j\mathrm{LR} = \frac{(-1)^j}{2}(\beta^{(2)} - \beta^{(1)})'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\beta^{(2)} - \beta^{(1)}) = (-1)^j\mu, \quad j = 1,2,$$

$$\mathrm{Var}_j\mathrm{LR} = (\beta^{(2)} - \beta^{(1)})'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\beta^{(2)} - \beta^{(1)}) = v^2,$$

where E_j and Var_j stand for the expectation and variance under H_j. Hence

$$\mathrm{LR} \sim N((-1)^j\mu, v^2), \; j = 1,2,$$

we have

$$P_{\mathrm{LR}}(2|1) = P_{\mathrm{LR}}(1|2) = \Phi\left(-\frac{\mu}{v}\right) = \Phi\left(-\frac{1}{2}\left\{(\beta^{(2)} - \beta^{(1)})'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\beta^{(2)} - \beta^{(1)})\right\}^{\frac{1}{2}}\right).$$

From Grenander and Roseblatt (1957), we have

$$\lim_{n\to\infty} \boldsymbol{D}_n^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{D}_n^{-1} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f^{-1}(\lambda)d\boldsymbol{M}(\lambda) = \boldsymbol{\Omega},$$

that is

$$P_{\mathrm{LR}}(2|1) = P_{\mathrm{LR}}(1|2) \sim \Phi\left(-\frac{1}{2}\left\{(\beta^{(2)} - \beta^{(1)})'\boldsymbol{D}_n\boldsymbol{\Omega}\boldsymbol{D}_n(\beta^{(2)} - \beta^{(1)})\right\}^{\frac{1}{2}}\right).$$

By Assumption **(G.1)**,

$$\lim_{n\to\infty} P_{\mathrm{LR}}(2|1) = \lim_{n\to\infty} P_{\mathrm{LR}}(1|2) = 0, \quad \text{for} \quad \beta^{(2)} \not\equiv \beta^{(1)}.$$

[Proof of Theorem 2.2] From the proof of Theorem 1, we have

$$P_{\mathrm{LR}}(2|1) = P_{\mathrm{LR}}(1|2) \sim \Phi\left(-\frac{1}{2}\left\{(\beta^{(2)} - \beta^{(1)})'\boldsymbol{D}_n\boldsymbol{\Omega}\boldsymbol{D}_n(\beta^{(2)} - \beta^{(1)})\right\}^{\frac{1}{2}}\right),$$

hence, putting contiguous condition (2.2) into the above, we get (2.3) immediately.

[Proof of Theorem 3.1] (i) From Lemma 3.1, we have

$$P_{\mathrm{LR}}(2|1) = P_{\mathrm{LR}}(1|2) \sim \Phi\left(-\frac{1}{2}\left\{(\beta^{(2)} - \beta^{(1)})'\widetilde{\boldsymbol{D}}_n\begin{pmatrix}\boldsymbol{W}_1 & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{W}_2\end{pmatrix}\widetilde{\boldsymbol{D}}_n(\beta^{(2)} - \beta^{(1)})\right\}^{\frac{1}{2}}\right),$$

since

$$\frac{\chi_r^{1/2}(n)}{n^d} \sim \frac{n^{r-\frac{1}{2}-d}}{(2r-1)^{1/2}} \to \infty, \text{ as } n \to \infty,$$

for $r = 1, \cdots, p$, and considering Assumption **(G.1)** for $r = p+1, \cdots, k$, we have the result.

(ii) We first calculate the mean and variance of I_α under hypotheses H_j,

$$E_j I_\alpha = (-1)^j(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})'(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}), \quad j = 1,2$$

and

$$\begin{aligned}\mathrm{Var}_j I_\alpha &= 4(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})'[\alpha^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &\quad +(1-\alpha^2)(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}](\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)}) \\ &\sim 4(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})'\widetilde{\boldsymbol{D}}_n^{-1}\boldsymbol{F}_\alpha\widetilde{\boldsymbol{D}}_n^{-1}(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})\end{aligned}$$

$$P_{I_\alpha}(2|1) = P_{I_\alpha}(1|2) \sim \Phi\left(-\frac{(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})'(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})}{2\left\{(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})'\widetilde{\boldsymbol{D}}_n^{-1}\boldsymbol{F}_\alpha\widetilde{\boldsymbol{D}}_n^{-1}(\boldsymbol{\beta}^{(2)} - \boldsymbol{\beta}^{(1)})\right\}^{\frac{1}{2}}}\right),$$

since all the elements of $\widetilde{\boldsymbol{D}}_n^{-1}$ tend to 0 under **(G.1)** and **(G.5)** as $n \to \infty$, we get the result.

[Proof of Theorem 3.2] Under contiguous condition (2.2), the fact for LR is clear, we only prove (3.13). From the proof of Theorem 3.1, we know that under (2.2),

$$P_{I_\alpha}(2|1) = P_{I_\alpha}(1|2) \sim \Phi\left(-\frac{1}{2}\frac{\boldsymbol{H}'\widetilde{\boldsymbol{D}}_n^{-2}\boldsymbol{H}}{\{\boldsymbol{H}'\widetilde{\boldsymbol{D}}_n^{-2}\boldsymbol{F}_\alpha\widetilde{\boldsymbol{D}}_n^{-2}\boldsymbol{H}\}^{1/2}}\right).$$

Since $k_1\boldsymbol{H}\boldsymbol{H}' \geq \boldsymbol{F}_\alpha \geq k_2\boldsymbol{H}\boldsymbol{H}'$, we have

$$k_1\{\boldsymbol{H}'\widetilde{\boldsymbol{D}}_n^{-2}\boldsymbol{H}\}^2 \geq \boldsymbol{H}'\widetilde{\boldsymbol{D}}_n^{-2}\boldsymbol{F}_\alpha\widetilde{\boldsymbol{D}}_n^{-2}\boldsymbol{H} \geq k_2\{\boldsymbol{H}'\widetilde{\boldsymbol{D}}_n^{-2}\boldsymbol{H}\}^2.$$

Hence

$$\Phi\left(-\frac{1}{2\sqrt{k_1}}\right) \geq \lim_{n\to\infty} P_{I_\alpha}(2|1) = \lim_{n\to\infty} P_{I_\alpha}(1|2) \geq \Phi\left(-\frac{1}{2\sqrt{k_2}}\right).$$

[Proof of Theorem 4.1] (i) Under Assumptions **(G.9)** and **(G.10)**, we have

$$E_j[\boldsymbol{\epsilon}'\boldsymbol{A}\boldsymbol{\epsilon} - \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})] = 0,$$

$$\mathrm{Var}_j[n^{-\frac{1}{2}-d}\boldsymbol{\epsilon}'\boldsymbol{A}\boldsymbol{\epsilon}] = n^{-2d}\sigma_a^2 + o(n^{-2d}), \quad \text{(from Lamma 4.1)}.$$

Thus

$$\begin{aligned}n^{-\frac{1}{2}-d}(D - E_jD) &= n^{-\frac{1}{2}-d}4\boldsymbol{m}_j'\boldsymbol{A}\boldsymbol{\epsilon} + o_p(1) \\ &:= 4m^{(j)}n^{-\frac{1}{2}-d}\boldsymbol{U}'\boldsymbol{A}\boldsymbol{\epsilon} + o_p(1),\end{aligned}$$

where $\boldsymbol{U} = (1,\cdots,1)'$. Define $\widetilde{\varepsilon}_t = \sum_{k=-\infty}^{\infty} a_k\varepsilon_{t+k}$. Then we can show that

$$n^{-\frac{1}{2}-d}E|\boldsymbol{U}'\boldsymbol{A}\boldsymbol{\epsilon} - \boldsymbol{U}'\widetilde{\boldsymbol{\epsilon}}| = n^{-\frac{1}{2}-d}\sum_{t=1}^{n}\sum_{|l|>n}|a_{l-t}|E|\varepsilon_l| \to 0 \quad \text{as} \quad n \to \infty$$

where $\widetilde{\boldsymbol{\epsilon}} = (\widetilde{\varepsilon}_1,\cdots,\widetilde{\varepsilon}_n)'$. Note that $\{\widetilde{\varepsilon}_t\}$ has the spectral density $f(\lambda)g^2(\lambda)$ with $g(\lambda) = \frac{1}{2\pi}\sum_{-\infty}^{\infty} a_\tau e^{-i\tau\lambda}$. Thus applying Yajima(1991) to $n^{-\frac{1}{2}-d}\boldsymbol{U}'\widetilde{\boldsymbol{\epsilon}}$, we obtain

$$n^{-\frac{1}{2}-d}(D - E_jD) \xrightarrow{\mathcal{L}} N\left(0, 16(m^{(j)})^2G_{fg}\right), \quad j = 1,2.$$

(ii) In this case, since

$$n^{-1}E_j|\boldsymbol{U}'\boldsymbol{A}\boldsymbol{\epsilon}|^2 = n^{-1}\boldsymbol{U}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{U} = O(n^{2\gamma+d}) = o(1), \qquad (\text{From} \quad 2\gamma + d < 0).$$

Then

$$\begin{aligned} n^{-\frac{1}{2}}(D - E_jD) &= 2n^{-\frac{1}{2}}[\boldsymbol{\epsilon}'\boldsymbol{A}\boldsymbol{\epsilon} - \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})] + o_p(1) \\ &\xrightarrow{\mathcal{L}} N\left(0, 4\sigma_a^2\right), \quad (\text{from Lemma 4.1}). \end{aligned}$$

REFERENCES

Adenstedt, R.K. (1974) On large-sample estimation for the mean of a stationary random sequence. *Ann. Statist.* **2**, 1095-107.

Anderson, T.W. (1971) *The Statistical Analysis of Time Series.* Wiley: New York.

Giraitis, L. and Surgailis, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Th. Rel. Fields.* **86**, 87-104.

Grenander U. and Rosenblatt M. (1957) *Statistical Analysis of Stationary Time Series.* Wiley: New York.

Liggett, W.S. (1971) On the asymtotic optimality of spectral analysis for testing hypotheses about time series. *Ann. Math. Statist.* **42**, 1348-58.

Shumway, R.H. (1982) Discriminant analysis for time series, in: *Handbook of Statistics*, Vol. 2, ed. P.R. Krishnaiah and L.N. Kanal, Amsterdam: North-holland, pp. 1-46.

Shumway, R.H. and Unger, A.N. (1974) Linear discriminant functions for stationary time series. *J. Am. Statist. Assoc.* **69**, 948-56.

Yajima, Y. (1988) On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16**, 791-807.

Yajima, Y. (1991) Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.* **19**, 158-77.

Zhang, G. and Taniguchi, M. (1994) Discriminant analysis for stationary vector time series. *J. Time Series Anal.* **15**, 117-26.

Zhang, G. (1994) Higher-order asymptotic theories for discriminant analysis of Gaussian ARMA processes. *Statist. Probab. Lett.* **20**, 259-68.