<table>
<thead>
<tr>
<th>Title</th>
<th>Discriminant Analysis for Regression Models with Long-memory Linear Disturbances (Information and Statistical Inference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Zhang, G.</td>
</tr>
<tr>
<td>Citation</td>
<td>数理解析研究所講究録 1995年版 916-169-179</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1995-07</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/59631">http://hdl.handle.net/2433/59631</a></td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Textversion</td>
<td>publisher</td>
</tr>
</tbody>
</table>

Kyoto University
Discriminant Analysis for Regression Models with Long-memory Linear Disturbances

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 quadratic 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

\[ H_j : \ Y_t = \sum_{i=1}^{k} X_i(t) \beta^{(j)}_i + \epsilon_t, \quad j = 1, 2. \]  

(1.1)

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

\[ p_j(Y) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (Y - \mathbf{X}\beta^{(j)})' \Sigma^{-1} (Y - \mathbf{X}\beta^{(j)}) \right\}, \quad j = 1, 2 \]

(1.2)

where

\[
\begin{align*}
\mathbf{Y} &= \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, & \mathbf{X}_r &= \begin{pmatrix} X_r(1) \\ \vdots \\ X_r(n) \end{pmatrix}, & \mathbf{X} &= (X_1, \cdots, X_k).
\end{align*}
\]

In this paper, we consider the problem of classifying \( Y \) from \( H_1 \) or \( H_2 \) into one of two populations \( H_1 \) and \( H_2 \). We assign \( Y \) to \( H_1 \) if \( Y \) falls into region \( R_1 \); otherwise we assign it to \( H_2 \), where \( R_1 \) and \( R_2 \) are exclusive and exhaustive regions in \( \mathbb{R}^n \). Let the probability of
misclassifying $Y$ from $H_i$ into $H_j$ be $P(j|i) = \int_{R_j} p_i(Y) dY$. 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

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

(1.3)

give the optimal classification, where

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

(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 describe 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 filtering, 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 disturbances. We show that LR is a consistent classification statistic under Grenander's conditions on $X_t$.

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:

\begin{align*}
(G.1) \quad \chi_r(n) &= \sum_{t=1}^{n} |X_r(t)|^2 \rightarrow \infty, \quad \text{as} \quad n \rightarrow \infty, \quad \text{for} \quad r = 1, \cdots, k, \\
(G.2) \quad \lim_{n \rightarrow \infty} \frac{\chi_r(n+1)}{\chi_r(n)} &= 1, \quad \text{for} \quad r = 1, \cdots, k, \\
(G.3) \quad \lim_{n \rightarrow \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)}
\end{align*}

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

\[
R(h) = \{R_{rs}(h)\}, \quad h = 0, \pm 1, \pm 2, \cdots,
\]

i.e. \( R(h) \) is a \( q \times q \) matrix with entry \( R_{rs}(h) \) in the \( r \)th row and \( s \)th column.

\( (G.4) \) The matrix \( R(0) \) is nonsingular.

It is not difficult to show that \( R'(h) = R(-h) \) (for each \( h \)), and that \( R(0) \) is positive definite(cf Grenander and Rosenblatt (1957)). It then follows that, there exists a Hermitian matrix function, \( M(\lambda) = \{M_{rs}(\lambda)\} \) 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 \( M(\lambda) \) has positive semi-definite matrix increments. Let

\[
D_n = \text{diag}\{\chi_1^{1/2}(n), \cdots, \chi_k^{1/2}(n)\}
\]

i.e. the \( D_n \) is the diagonal matrix with entry \( \chi_r^{1/2}(n) \) in the \( r \)th row and column.

[Theorem 2.1] Suppose that the spectral density \( f(\lambda) \) is a positive and continuous function on \([\pi, \pi]\). For classifying \( Y \) 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

\[
\lim_{n \rightarrow \infty} P_{LR}(2|1) = \lim_{n \rightarrow \infty} P_{LR}(1|2) = 0, \quad \text{for} \quad \beta^{(1)} \neq \beta^{(0)}.
\]

(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 \( \beta^{(2)} \) is contiguous to \( \beta^{(1)} \):

\[
H_1 : \beta^{(1)} = \beta; \\
H_2 : \beta^{(2)} = \beta + D_n^{-1}H,
\]

(2.2)
where \( H = (H_1, \cdots, H_k)' \) is a constant vector.

**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

\[
\lim_{n \to \infty} P_{LR}(2|1) = \lim_{n \to \infty} P_{LR}(1|2) = \Phi \left( -\frac{1}{2} \sqrt{H'\Omega H} \right).
\]

(2.3)

where

\[
\Omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^{-1}(\lambda)dM(\lambda).
\]

Let us consider the case of polynomial regression.

**Example 2.1** Suppose that \( X_r(t) = t^{r-1}, r = 1, \cdots, k \) and \( \{\epsilon_t\} \) is a stationary process with zero mean and spectral density \( f(\lambda) \) that is a positive and continuous function on \([-\pi, \pi]\). Then \( M(\lambda) \) has only a jump at \( \lambda = 0 \) of \( M_0 = \left\{ \frac{[(2r-1)(2s-1)]^{1/2}}{2\pi f(0)(r+s-1)} \right\}_{r,s=1,\cdots,k} \) (e.g. Anderson(1971)). Thus under contiguous assumption (2.2), the misclassification probabilities of LR are evaluated as

\[
\lim_{n \to \infty} P_{LR}(2|1) = \lim_{n \to \infty} P_{LR}(1|2) = \Phi \left( -\frac{1}{2} \sqrt{H'\Omega_0 H} \right).
\]

(2.4)

where

\[
\Omega_0 = \left\{ \frac{[(2r-1)(2s-1)]^{1/2}}{2\pi f(0)(r+s-1)} \right\}_{r,s=1,\cdots,k}.
\]

### 3 LONG-MEMORY CLASSIFICATION

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

\[
\epsilon_t = \sum_{j=0}^{\infty} \alpha_j \omega_{t-j}, \quad \omega_t \sim N(0,1), \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty.
\]

(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

\[
g_{jl}(\lambda) = \frac{\sum_{t=1}^{n} X_j(t)e^{-it\lambda}}{2\pi X_r(n)X_s(n)}, \quad M_{jl}(\lambda) = \int_{-\pi}^{\lambda} g_{jl}(\omega)d\omega,
\]

(3.2)

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

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

(G.6) \( \{\epsilon_t\} \) 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 $\int_{|\lambda|<c} f(\lambda)dM_{jj}^{n}(\lambda) < \delta$, for every $n$ and $j = p + 1, \cdots, k$,

\[ \frac{\max_{1 \leq r \leq k} X_{r}^{2}(t)}{\chi_{r}^{1/2}(n)} = o(n^{-\mu}), \quad r = p + 1, \cdots, k, \text{ for some } \mu > 1 - 2d. \]

From Assumption (G.5), we have

\[ \chi_r(n) = \sum_{t=1}^{n} t^{2(r-1)} \sim \frac{n^{2r-1}}{2r-1}, \quad r = 1, \cdots, p, \quad (3.3) \]

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

\[ \overline{D}_{n} = \text{diag}\left( \frac{\chi_{1}^{1/2}(n)}{n^{d}}, \cdots, \frac{\chi_{p}^{1/2}(n)}{n^{d}}, \chi_{p+1}^{1/2}(n), \cdots, \chi_{k}^{1/2}(n) \right), \quad (3.4) \]

The following Lemma follows from Yajima (1991).

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

\[ \lim_{n \to \infty} \overline{D}_{n}^{-1} X' \Sigma^{-1} X \overline{D}_{n}^{-1} = \begin{pmatrix} W_1 & O \\ O & W_2 \end{pmatrix} \]  
and

\[ \lim_{n \to \infty} D_{n}^{-2} \overline{D}_{n} X' \Sigma X \overline{D}_{n} D_{n}^{-2} = \begin{pmatrix} B_1 & O \\ O & B_2 \end{pmatrix} \]  

where $W_1, B_1$ are $p \times p$ matrix with $(i,j)$th entry

\[ w_{1}^{ij} = \frac{\Gamma(i-d)\Gamma(j-d)\{(2i-1)(2j-1)\}^{1/2}}{2\pi f^{*}(0)\Gamma(i-2d)\Gamma(j-2d)(i+j-1-2d)} \]

and $W_2, B_2$ are $(k-p) \times (k-p)$ matrix with $(i,j)$th entry

\[ w_{2}^{ij} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^{-1}(\lambda)dM_{i+p,j+p}(\lambda), \quad b_{2}^{ij} = 2\pi \int_{-\pi}^{\pi} f(\lambda)dM_{i+p,j+p}(\lambda) \]

respectively.

Now, to classify $Y$ (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

\[ \hat{\beta}_{\alpha} = \alpha \hat{\beta}_{L} + (1 - \alpha) \hat{\beta}_{B}, \quad (0 \leq \alpha \leq 1) \]

where

\[ \hat{\beta}_{L} = (X'X)^{-1}X'Y, \quad \hat{\beta}_{B} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \]

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

\[ (\hat{\beta}_{\alpha} - \beta^{(1)})'(\hat{\beta}_{\alpha} - \beta^{(1)}) < (\hat{\beta}_{\alpha} - \beta^{(2)})'(\hat{\beta}_{\alpha} - \beta^{(2)}) \]
to \( H_2 \), otherwise. That is, assign \( Y \) to \( H_1 \) if
\[
I_{\alpha} = (\beta^{(2)} - \beta^{(1)})' \left[ 2\hat{\beta}_\alpha - (\beta^{(2)} + \beta^{(1)}) \right] < 0,
\]
to \( H_2 \), otherwise.

**[Theorem 3.1]** Under Assumptions (G.1)-(G.8),
\[
\begin{align*}
(\text{i}) & \quad \lim_{n \to \infty} P_{LR}(2|1) = \lim_{n \to \infty} P_{LR}(1|2) = 0 \quad \text{for } \beta_1 \neq \beta_2 \\
(\text{ii}) & \quad \lim_{n \to \infty} P_{I_\alpha}(2|1) = \lim_{n \to \infty} P_{I_\alpha}(1|2) = 0 \quad \{ \text{for } \beta_1 \neq \beta_2 \}
\end{align*}
\]

Now, let the regression parameters associated with \( H_1 \) and \( H_2 \) be
\[
H_1 : \beta^{(1)} = \beta; \\
H_2 : \beta^{(2)} = \beta + \overline{D}_n^{-1}H.
\]

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

**[Theorem 3.2]** Under Assumptions (G.1)-(G.8) and under contiguous condition (3.11),
\[
\begin{align*}
\lim_{n \to \infty} P_{LR}(2|1) = \lim_{n \to \infty} P_{LR}(1|2) &= \Phi \left( -\frac{1}{2} \left[ H' \left( \begin{array}{cc} W_1 & O \\ O & W_2 \end{array} \right) H \right]^{\frac{1}{2}} \right) \\
\Phi \left( -\frac{1}{2\sqrt{k_1}} \right) &\geq \lim_{n \to \infty} P_{I_\alpha}(2|1) = \lim_{n \to \infty} P_{I_\alpha}(1|2) \geq \Phi \left( -\frac{1}{2\sqrt{k_2}} \right),
\end{align*}
\]

where \( k_1 \) and \( k_2 \) are chosen so that \( k_1H H' \geq F_{\alpha} \geq k_2H H' \). Here
\[
F_{\alpha} = \alpha^2 A \left( \begin{array}{cc} B_1 & O \\ O & B_2 \end{array} \right) A + (1 - \alpha^2) \left( \begin{array}{cc} W_1^{-1} & O \\ O & W_2^{-1} \end{array} \right),
\]

where
\[
A = \lim_{n \to \infty} \overline{D}_n(X'X)^{-1} \overline{D}_n^{-1} D_n^2.
\]

**[Example 3.1]** In Theorem 3.2, if \( X_r (t) = 1, \ r = 1 \), then
\[
\begin{align*}
\lim_{n \to \infty} P_{LR}(2|1) = \lim_{n \to \infty} P_{LR}(1|2) &= \lim_{n \to \infty} P_{I_\alpha}(2|1) = \lim_{n \to \infty} P_{I_\alpha}(1|2) \\
&= \Phi \left( -\frac{|H|\Gamma(1-d)}{2\sqrt{(1-2d)2\pi f^*(0)\Gamma(1-d)}} \right),
\end{align*}
\]

and
\[
\begin{align*}
\lim_{n \to \infty} P_{I_\alpha}(2|1) = \lim_{n \to \infty} P_{I_\alpha}(1|2) &= \Phi \left( -\frac{|H|}{2\sqrt{G_f}} \right),
\end{align*}
\]

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

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

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

$$\epsilon_t = \sum_{j=0}^{\infty} a_j \omega_{t-j}, \quad \sum_{j=0}^{\infty} a_j^2 < \infty, \quad (4.2)$$

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

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

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

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

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

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

$$a_r = \int_{-\pi}^{\pi} e^{i\pi r x} \tilde{a}(x) dx, \quad (4.4)$$

and assume that

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

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

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

[Lemma 4.1] Let $\epsilon = (\epsilon_1, \cdots, \epsilon_n)'$ be the $n$ consective stretch of $\{\epsilon_t\}$ in regression models (4.1) satisfy (4.2). Then under Assumptions (G.9)-(G.11),

$$\frac{\epsilon'A \epsilon - E_j(\epsilon'A \epsilon)}{\sqrt{n}} \overset{\mathcal{L}}{\to} N \left( 0, \sigma_\epsilon^2 \right), \quad j = 1, 2 \quad (4.5)$$

where

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

and $\kappa_4^\omega$ is the fourth cumulant of $\{\omega_t\}$.

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

$$[Y' A Y - E_1(Y' A Y)]^2 < [Y' A Y - E_2(Y' A Y)]^2$$

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

$$D = 2\epsilon'A \epsilon + 4m_j' A \epsilon + 2\nabla_j - (\nabla_1 + \nabla_2) - 2\text{tr}(A \Sigma) > 0$$
to $H_2$, otherwise.

[Theorem 4.1] Suppose that the stationary linear disturbances \{\epsilon_t\} in regression models (4.1) satisfy (4.2). Then under Assumptions (G.9)-(G.11),

(i) If $|a_r| = O(\rho^r), 0 \leq \rho < 1$, (i.e., $\hat{a}(x) = g(x)$ is a short-memory spectral density), then

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

(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

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

(4.7)

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

The following theorem and proposition follow from Theorem 4.1.

[Theorem 4.2] Suppose that the stationary linear disturbances \{\epsilon_t\} in regression models (4.1) satisfy (4.2). Then under Assumptions (G.9)-(G.11),

(i) If $\hat{a}(x) = g(x)$, where $g(x)$ is a positive and continuous function on $[-\pi, \pi]$, then

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

(4.8)

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

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

(4.9)

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

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

(4.10)

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

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

(4.11)
5 APPENDIX: PROOFS OF THEOREMS

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

\[ E_j LR = \frac{(-1)^j}{2} \left( (\beta^{(2)} - \beta^{(1)})' X' \Sigma^{-1} X (\beta^{(2)} - \beta^{(1)}) \right) = (-1)^j \mu, \quad j = 1, 2, \]

\[ \text{Var}_j LR = (\beta^{(2)} - \beta^{(1)})' X' \Sigma^{-1} X (\beta^{(2)} - \beta^{(1)}) = \nu^2, \]

where \( E_j \) and \( \text{Var}_j \) stand for the expectation and variance under \( H_j \). Hence

\[ LR \sim N((-1)^j \mu, \nu^2), \quad j = 1, 2, \]

we have

\[ P_{LR}(2|1) = P_{LR}(1|2) = \Phi \left( -\frac{\mu}{\nu} \right) = \Phi \left( -\frac{1}{2} \left( (\beta^{(2)} - \beta^{(1)})' X' \Sigma^{-1} X (\beta^{(2)} - \beta^{(1)}) \right)^{1/2} \right). \]

From Grenander and Roseblatt (1957), we have

\[ \lim_{n \to \infty} D_n^{-1} X' \Sigma^{-1} X D_n^{-1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^{-1}(\lambda) dM(\lambda) = \Omega, \]

that is

\[ P_{LR}(2|1) = P_{LR}(1|2) \sim \Phi \left( -\frac{1}{2} \left( (\beta^{(2)} - \beta^{(1)})' D_n \Omega D_n (\beta^{(2)} - \beta^{(1)}) \right)^{1/2} \right). \]

By Assumption (G.1),

\[ \lim_{n \to \infty} P_{LR}(2|1) = \lim_{n \to \infty} P_{LR}(1|2) = 0, \quad \text{for} \quad \beta^{(2)} \neq \beta^{(1)}. \]

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

\[ P_{LR}(2|1) = P_{LR}(1|2) \sim \Phi \left( -\frac{1}{2} \left( (\beta^{(2)} - \beta^{(1)})' D_n \Omega D_n (\beta^{(2)} - \beta^{(1)}) \right)^{1/2} \right), \]

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

\[ P_{LR}(2|1) = P_{LR}(1|2) \sim \Phi \left( -\frac{1}{2} \left( (\beta^{(2)} - \beta^{(1)})' W_1 \begin{bmatrix} O & O \\ O & W_2 \end{bmatrix} D_n (\beta^{(2)} - \beta^{(1)}) \right)^{1/2} \right), \]

since

\[ \frac{\chi_{r}^{1/2}(n)}{n^d} \sim \frac{n^{r-\frac{d}{2}}}{(2r-1)^{1/2}} \to \infty, \quad \text{as} \quad n \to \infty, \]

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 $I_{\alpha}$ under hypotheses $H_{j}$,

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

and

$$\text{Var}_{j}I_{\alpha} = 4(\beta^{(2)} - \beta^{(1)})'[\alpha^{2}(X'X)^{-1}X'\Sigma X(X'X)^{-1} \alpha^{2}(X'X)^{-1} + (1 - \alpha^{2})(X'\Sigma^{-1}X)^{-1}] (\beta^{(2)} - \beta^{(1)})$$

$$\sim 4(\beta^{(2)} - \beta^{(1)})' \overline{D}_{n}^{-1} F_{\alpha} \overline{D}_{n}^{-1} (\beta^{(2)} - \beta^{(1)})$$

since all the elements of $\overline{D}_{n}^{-1}$ 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),

$$P_{I_{\alpha}}(2|1) = P_{I_{\alpha}}(1|2) \sim \Phi \left( -\frac{1}{\frac{1}{2 \sqrt{k_{1}}} \overline{D}_{n}^{-2} F_{\alpha} \overline{D}_{n}^{-2} H} \right)$$

Since $k_{1}HH' \geq F_{\alpha} \geq k_{2}HH'$, we have

$$k_{1}\{\overline{D}_{n}^{-2} H\}^{2} \geq \overline{D}_{n}^{-2} F_{\alpha} \overline{D}_{n}^{-2} H \geq k_{2}\{\overline{D}_{n}^{-2} H\}^{2}.$$  

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

$$E_{j}[\epsilon'A \epsilon - \text{tr}(A \Sigma)] = 0,$$

$$\text{Var}_{j}[n^{-\frac{1}{2}-d} \epsilon'A \epsilon] = n^{-2d}\sigma_{a}^{2} + o(n^{-2d}), \quad \text{from Lamma 4.1}.$$  

Thus

$$n^{-\frac{1}{2}-d}(D - E_{j}D) = n^{-\frac{1}{2}-d} 4m_{j}'A \epsilon + o_{p}(1)$$

$$:= 4m^{(j)} n^{-\frac{1}{2}-d} U'A \epsilon + o_{p}(1),$$

where $U = (1, \cdots , 1)'$. Define $\tilde{\epsilon}_{t} = \sum_{k=1}^{\infty} a_{k} \epsilon_{t+k}$. Then we can show that

$$n^{-\frac{1}{2}-d}E[U'A \epsilon - U'\tilde{\epsilon}] = n^{-\frac{1}{2}-d} \sum_{t=1}^{n} \sum_{|l|>|n} |a_{l-t}|E[\epsilon_{l}] \to 0 \text{ as } n \to \infty$$

where $\tilde{\epsilon} = (\tilde{\epsilon}_{1}, \cdots , \tilde{\epsilon}_{n})'$. Note that $\{\tilde{\epsilon}_{t}\}$ 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 $n^{-\frac{1}{2}-d}U'\tilde{\epsilon}$, we obtain

$$n^{-\frac{1}{2}-d}(D - E_{j}D) \xrightarrow{C} N \left( 0, 16(m^{(j)})^{2} G_{\epsilon} \right), \quad j = 1, 2.$$
In this case, since
\[ n^{-1}E_j|U'A\epsilon|^2 = n^{-1}U'A\Sigma AU = O(n^{2\gamma+d}) = o(1), \quad \text{(From} \quad 2\gamma + d < 0). \]
Then
\[ n^{-\frac{1}{2}}(D - E_jD) = 2n^{-\frac{1}{2}}[\epsilon'\epsilon - \text{tr}(A\Sigma)] + o_p(1) \]
\[ \xrightarrow{\mathcal{L}} N(0, 4\sigma_a^2), \quad \text{(from Lemma 4.1)}. \]

REFERENCES


