SOME STATISTICAL APPLICATIONS FOR LOCALLY STATIONARY PROCESSES

K. SAKIYAMA

Received November 26, 2001

ABSTRACT. Time series analysis under stationary assumption is well established. However it is not sufficient for stationary time series models to describe the real world. A class of locally stationary processes was introduced by Dahlhaus. By using nonstationary models with time varying spectra, we attempt to analyze some data from mining explosions, natural earthquakes and financial time series. Although many researchers used the ordinary autoregressive or autoregressive moving average models, we fit a time varying autoregressive (TVAR) model of order p to the data whose coefficients are polynomials with respect to time. Moreover we select a favorable model by use of AIC, and compare the results with the others. Finally, some numerical problems of discriminant and cluster analysis will be discussed.

1 Introduction. Although stationary time series analysis is well established, stationary time series models are not plausible to describe the real world. Priestly introduced a class of nonstationary processes based on the concept of "*evolutionary spectra*". The evolutional spectral density functions are time dependent and generalize the usual definition of spectra for stationary processes. Recently Dahlhaus introduced a class of locally stationary processes. Then he elucidated some fundamental results of the statistical inference.

In many applications for the seismic records, modelling based on autoregressive moving average (ARMA) processes has proven effective (Polhemus and Cakmak (1981); Chang et al. (1982); Gersch and Kitagawa (1985); Cakmak et al. (1985); Dargahi-Noubary (1992)). In the work by Chang et al. (1982), several ARMA models of different orders were identified, parameters were estimated and statistical measures were evaluated to test the goodness of fit between the models and actual data. Then it was found that ARMA(2,1) or ARMA(4,1) models provided good fits to time segments of the earthquake acceleration time histories. Polhemus and Cakmak (1981) also used ARMA(2,1) and ARMA(4,1) models for the earthquake data.

Time series modelling based on nonstationary processes is possible. Adak (1998) discussed applications of the locally stationary processes to speech signals and earthquake data. Sakiyama and Taniguchi (2001) investigated the problems of classifying a multivariate non-Gaussian locally stationary process $\{X_{t,T}\}$ into one of two categories described by two hypotheses: $\pi_1 : f(u,\lambda)$, $\pi_2 : g(u,\lambda)$, where $f(u,\lambda)$ and $g(u,\lambda)$ are time varying spectral density matrices. They used a Gaussian approximation of the likelihood ratio (GLR) for this problem, and showed that GLR is consistent. Also the misclassification probabilities were evaluated under contiguous hypotheses. In this paper we analyze some data from mining explosions, natural earthquakes and financial time series. The theoretical results of our

2000 *Mathematics Subject Classification.* Primary 62M10, 62H30; Secondary 62H12, 62P99.
Key words and phrases. Locally stationary processes; classification criterion; time varying spectral density; natural earthquake; mining explosion; daily return.

discriminant analysis are applied to them. In comparison with the literature in this field we report some interesting features. This paper is organized as follows. In Section 2, we review the discriminant analysis for locally stationary processes. In Section 3, we attempt to fit time varying autoregressive models to some real data. Then, using AIC we compare our results with the literature in this field. Section 4 discusses some data analysis of classifying earthquakes and mining explosions. Moreover, we discuss the problem of clustering five daily returns data.

2 Discriminant Analysis for Locally stationary Processes.

Dahlhaus(1997) developed asymptotic theory for univariate locally stationary processes. We start with the definition of a multivariate locally stationary process. Since we discuss the discriminant analysis for multivariate locally stationary processes, we extend some of his results to the case when the process concerned is multivariate.

Definition 1 *A sequence of multivariate stochastic processes* $X_{t,T} = (X_{t,T}^{(1)}, \dots, X_{t,T}^{(d)})'$ $(t = 1, \dots, T)$ *is called locally stationary with transfer function matrix* $A_{t,T}(\lambda) = \{A_{t,T}(\lambda)_{a,b} : a, b = 1, \dots, d\}$ *and mean* $\mathbf{0}$ *if there exist a representation*

$$X_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,T}(\lambda) d\xi(\lambda) \tag{1}$$

where the following holds.

(i) $\xi(\lambda) = (\xi_1(\lambda), \dots, \xi_d(\lambda))'$ *is a complex valued vector process on* $[-\pi, \pi]$ *with* $\overline{\xi_a(\lambda)} = \xi_a(-\lambda)$, $E\xi_{a_j}(\lambda) = 0$ *and*

$$cum\{d\xi_{a_1}(\lambda_1), \dots, d\xi_{a_k}(\lambda_k)\} = \eta\left(\sum_{j=1}^{k} \lambda_j\right) g_{a_1,\dots,a_k}(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1, \dots, d\lambda_k$$

where $cum\{\cdots\}$ *denotes the cumulant of* k*-th order,* $g_a(\lambda) = 0$, $g_{a,b}(\lambda) = \delta(a, b)$, $|g_{a_1,\dots,g_{a_k}}(\lambda_1, \dots, \lambda_{k-1})| \le C(k)$ *(*$C(k)$ *is constant) for all* $a_1, \dots, a_k \in \{1, \dots, d\}$ *and* $\eta(\lambda) = \sum_{l=-\infty}^{\infty} \delta(\lambda + 2\pi l)$ *is the period* 2π *extension of the Dirac delta function.*

(ii) There exists a constant K *and a* 2π*-periodic matrix valued function* $A(u, \lambda) = \{A(u, \lambda)_{a,b} : a, b = 1, \dots, d\} : [0, 1] \times \boldsymbol{R} \to \boldsymbol{C}^{d \times d}$ *with* $\overline{A(u, \lambda)} = A(u, -\lambda)$ *and*

$$\sup_{t,\lambda} \left| A_{t,T}(\lambda)_{a,b} - A\left(\frac{t}{T}, \lambda\right)_{a,b} \right| \le KT^{-1}$$

for all $a, b \in \{1, \dots, d\}$ *and* $T \in \boldsymbol{N}$*, where* $A(u, \lambda)$ *is assumed to be continuous in* u*.*

We call $f(u, \lambda) \equiv A(u, \lambda)\overline{A(u, \lambda)}'$ the time varying spectral density matrix of $\{X_{t,T}\}$. Letting

$$d_N^{(a)}(u, \lambda) = \sum_{s=0}^{N-1} X_{[uT]-N/2+s+1,T}^{(a)} \exp(-i\lambda s) , \tag{2}$$

we introduce the periodogram matrix $I_N(u, \lambda) = \{I_N(u, \lambda)_{a,b} : a, b = 1, \dots, d\}$ over a segment of length N with midpoint $[uT]$, where

$$I_N(u, \lambda)_{a,b} = \frac{1}{2\pi N} d_N^{(a)}(u, \lambda) d_N^{(b)}(u, -\lambda) . \tag{3}$$

The shift from segments to segment is denoted by N. $I_N(u_j,\lambda)$ is calculated over segments with midpoints $u_jT = t_j = N(j-1/2), (j=1,\dots,M)$ where $T = NM$.
In what follows we briefly review the results by Sakiyama and Taniguchi (2001) (Propositions 1-3). Consider the problems of classifying a multivariate locally stationary process $\{X_{t,T}\}$ into one of two categories described by two hypotheses:

$$\pi_1 : f(u,\lambda) \quad , \quad \pi_2 : g(u,\lambda)$$

where $f(u,\lambda)$ and $g(u,\lambda)$ are $d\times d$ time varying spectral density matrices. For this discriminant problem, we use

$$D(f:g) = \frac{1}{4\pi M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\Bigg[\log\left\{\frac{|g(u_j,\lambda)|}{|f(u_j,\lambda)|}\right\} + \mathrm{tr}\left[I_N(u_j,\lambda)\left\{g^{-1}(u_j,\lambda) - f^{-1}(u_j,\lambda)\right\}\right]\Bigg]d\lambda \tag{4}$$

as a classification statistic. That is, if $D(f:g) > 0$ we choose category π_1. Otherwise we choose category π_2. this criterion is an approximation of Gaussian log-likelihood ratio between π_1 and π_2 (Dahlhaus (1997)). We set down the following assumption.

Assumption 1 *There exists $C > 0$ such that the minimum eigenvalues of $f(u,\lambda)$ and $g(u,\lambda)$ are grater than C for all u and λ.*

The following proposition describes the asymptotics of $D(f:g)$ under π_1 and π_2.

Proposition 1 *Suppose that Assumption 1 holds. Then, as $T\to\infty$, under π_1*

$$\sqrt{T}\left[D(f:g) - E\left\{D(f:g)|\pi_1\right\}\right] \xrightarrow{\mathcal{D}} N\left\{0,\sigma^2(f,g)\right\} \tag{5}$$

and under π_2

$$\sqrt{T}\left[D(f:g) - E\left\{D(f:g)|\pi_2\right\}\right] \xrightarrow{\mathcal{D}} N\left\{0,\sigma^2(g,f)\right\} \tag{6}$$

where

$$\begin{aligned}\sigma^2(f,g) = {}& \frac{1}{4\pi}\int_0^1\Bigg[\int_{-\pi}^{\pi} tr\left\{f(u,\lambda)g^{-1}(u,\lambda) - I_d\right\}^2 d\lambda \\ & +\frac{1}{8\pi}\sum_{b_1,b_2,b_3,b_4=1}^{d}\Bigg[\left[\int_{-\pi}^{\pi} A(u,\lambda)^*\left\{g^{-1}(u,\lambda) - f^{-1}(u,\lambda)\right\}A(u,\lambda)d\lambda\right]_{b_2,b_1} \\ & \times\left[\int_{-\pi}^{\pi} A(u,\mu)^*\left\{g^{-1}(u,\mu) - f^{-1}(u,\mu)\right\}A(u,\mu)d\mu\right]_{b_3,b_4} \\ & \times g_{b_1,b_2,b_3,b_4}(\lambda,-\lambda,-\mu)\Bigg]\Bigg]du,\end{aligned}$$

and $[M]_{a,b}$ is the (a,b) element of matrix M.

If we use $D(f:g)$ as a classification criterion, the misclassification probabilities are

$$P(2|1) = P\{D(f:g)\le 0|\pi_1\} \quad , \quad P(1|2) = P\{D(f:g) > 0|\pi_2\}.$$

The following proposition shows that the classification statistic is asymptotically consistent.

Proposition 2 *Under Assumption 1,*

$$\lim_{T\to\infty} P(2|1) = 0 \ , \ \lim_{T\to\infty} P(1|2) = 0.$$

To evaluate the goodness of $D(f:g)$ we assume that $g(u,\lambda)$ is contiguous to $f(u,\lambda)$. Now we set the spectral densities as

$$\pi_1 : f(u,\lambda) = f_\theta(u,\lambda) \ , \ \pi_2 : g(u,\lambda) = f_{\theta+h/\sqrt{T}}(u,\lambda) \tag{7}$$

where $\theta \in \Theta \subset \boldsymbol{R}^q$ and $h = (h_1, \dots, h_q)'$.

Assumption 2 *(i) We observe the realization $X_{1,T}, \dots, X_{T,T}$ of a d-dimensional locally stationary process with mean* $\mathbf{0}$ *and transfer function matrix $A_{t,T}(\lambda)$. The time varying spectral density matrix is $f_\theta(u,\lambda) = A_\theta(u,\lambda)\overline{A_\theta(u,\lambda)}'$, $\theta \in \Theta \subset \boldsymbol{R}^q$, where Θ is compact.*

(ii) All the eigenvalues of $f_\theta(u,\lambda)$ are bounded from below by some constant $C > 0$ uniformly in θ, u and λ.

(iii) The components of $f_\theta(u,\lambda)$, $\nabla f_\theta(u,\lambda)$ and $\nabla^2 f_\theta(u,\lambda)$ are continuous on $\Theta \times [0,1] \times [-\pi,\pi]$ (∇ denotes the gradient with respect to θ).

(iv) N and T fulfill the relations $T^{1/4} \ll N \ll T^{1/2}/\ln T$.

Proposition 3 *Under (7), we suppose Assumption 2. If we use $D(f:g)$ as a classification criterion, then*

$$\lim_{T\to\infty} P(2|1) = \lim_{T\to\infty} P(1|2) = \Phi\left[\frac{-\frac{1}{2}F(\theta)}{\{F(\theta)+D(\theta)\}^{\frac{1}{2}}}\right] \tag{8}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and

$$F(\theta) = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} tr\left\{\sum_{i=1}^q h_i \nabla_i f_\theta(u,\lambda) f_\theta^{-1}(u,\lambda)\right\}^2 du d\lambda$$

$$\begin{aligned} D(\theta) = \frac{1}{8\pi}\int_0^1 \sum_{b_1,b_2,b_3,b_4=1}^d & \\ \Bigg[\sum_{i=1}^q h_i \Bigg[\int_{-\pi}^{\pi} & A_\theta(u,\lambda)^* \left\{ f_\theta^{-1}(u,\lambda)\nabla_i f_\theta(u,\lambda) f_\theta^{-1}(u,\lambda)\right\} A_\theta(u,\lambda) d\lambda\Bigg]_{b_2,b_1} \\ \times \sum_{j=1}^q h_j \Bigg[\int_{-\pi}^{\pi} & A_\theta(u,\mu)^* \left\{ f_\theta^{-1}(u,\mu)\nabla_j f_\theta(u,\mu) f_\theta^{-1}(u,\mu)\right\} A_\theta(u,\mu) d\mu\Bigg]_{b_3,b_4} \\ & \times g_{b_1,b_2,b_3,b_4}(\lambda,-\lambda,-\mu)\Bigg]\Bigg] du. \end{aligned} \tag{9}$$

In Section 4 we discuss the discriminant problem in the following parametric form:

$$\pi_1 : f_\theta(u,\lambda) \ , \ \pi_2 : g_\theta(u,\lambda),$$

where $f_\theta(u,\lambda)$ and $g_\theta(u,\lambda)$ are time varying spectral density functions. Consider the problem of classifying a locally stationary process $\{X_{t,T}\}$ into one of two categories described by two hypotheses:

$$\pi_1 : f_\theta(u,\lambda) \quad , \quad \pi_2 : g_\theta(u,\lambda)$$

where $f_\theta(u,\lambda)$ and $g_\theta(u,\lambda)$ are parametric time varying spectral densities. We can construct the estimated spectral density $h_{\widehat{\theta}}(u,\lambda)$ from $\{X_{t,T}\}$. Let

$$D(h_\theta : g_\theta) = \frac{1}{4\pi M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left[\log\left\{\frac{g_\theta(u_j,\lambda)}{f_\theta(u_j,\lambda)}\right\} + h_\theta(u_j,\lambda)g_\theta(u_j,\lambda)^{-1} - 1\right]d\lambda.$$

Then the inequality

$$D(h_{\widehat{\theta}} : g_\theta) > D(h_{\widehat{\theta}} : f_\theta) \tag{10}$$

implies that $h_{\widehat{\theta}}(u,\lambda)$ is nearer to $f_\theta(u,\lambda)$ than $g_\theta(u,\lambda)$ in the sense of spectral divergence measure $D(\ :\ )$. Write

$$\begin{aligned} D(h_{\widehat{\theta}}) &\equiv D(h_{\widehat{\theta}} : g_{\widehat{\theta}}) - D(h_{\widehat{\theta}} : f_{\widehat{\theta}}) \\ &= \frac{1}{4\pi M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left[\log\left\{\frac{g_{\widehat{\theta}}(u_j,\lambda)}{f_{\widehat{\theta}}(u_j,\lambda)}\right\}\right. \\ &\qquad\qquad \left. + h_{\widehat{\theta}}(u_j,\lambda)\left\{g_{\widehat{\theta}}^{-1}(u_j,\lambda) - f_{\widehat{\theta}}^{-1}(u_j,\lambda)\right\}\right]d\lambda. \end{aligned} \tag{11}$$

We propose a rule to classify $\{X_{t,T}\}$ into π_1 or π_2 according as $D(h_{\widehat{\theta}}) > 0$ or $D(h_{\widehat{\theta}}) \le 0$, respectively.
It should be recognized that $D(f_\theta, g_\theta)$ is a functional of the product $f_\theta(u_j,\lambda)$ $\times g_\theta(u_j,\lambda)^{-1}$ and can be extended to a more general divergence measure;

$$\overline{D}(f_\theta : g_\theta) = \frac{1}{4\pi}\sum_{j=1}^{M}\int_{-\pi}^{\pi} F_j\{f_\theta(u_j,\lambda)g_\theta(u_j,\lambda)^{-1}\}d\lambda, \tag{12}$$

where $F_j\{\ \cdot\ \}$ is a suitable function, which is specified below. To ensure that general mesures of the form $\overline{D}(f_\theta : g_\theta)$ in (12) have the divergence property, we shall require $\overline{D}(f_\theta : g_\theta) \ge 0$ with equality if and only if $f_\theta = g_\theta$ *a.e.* Therefore the function $F_j(x)$ must have a unique minimum at $x = 1$. Generally $\overline{D}(f_\theta : g_\theta)$ is not symmetric with respect to f_θ and g_θ but can easily be made so by defining

$$\overline{F}_j(x) = F_j(x) + F_j(x^{-1}),$$

and we obtain the symmetric spectral divergence measure

$$\begin{aligned} \overline{D}(f_\theta : g_\theta) &= \frac{1}{4\pi M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left[F_j\{f_\theta(u_j,\lambda)g_\theta(u_j,\lambda)^{-1}\}\right. \\ &\qquad\qquad \left. + F_j\{g_\theta(u_j,\lambda)f_\theta(u_j,\lambda)^{-1}\}\right]d\lambda \end{aligned} \tag{13}$$

(see Taniguchi (1991) and Kakizawa (1996)).
The measures of disparity between spectral densities can be used for clustering locally stationary processes. For example, let $f_{\widehat{\theta}}$ and $g_{\widehat{\theta}}$ be estimated spectral densities for two different processes. We define the symmetric spectral distance between $f_{\widehat{\theta}}$ and $g_{\widehat{\theta}}$ by $\overline{D}(f_{\widehat{\theta}} : g_{\widehat{\theta}})$.
For example, by using $F_j(x) = x + x^{-1}$, we obtain the symmetric spectral distance

$$
\begin{aligned}
(14)\ \overline{\overline{D}}(f_{\widehat{\theta}} : g_{\widehat{\theta}}) &\equiv \frac{1}{2}\overline{D}(f_{\widehat{\theta}} : g_{\widehat{\theta}}) \\
(15)\qquad &= \frac{1}{4\pi M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left[\{f_\theta(u_j,\lambda)g_\theta(u_j,\lambda)^{-1}\} + \{g_\theta(u_j,\lambda)f_\theta(u_j,\lambda)^{-1}\}\right]d\lambda.
\end{aligned}
$$

In Section 4 we will use this procedure.

3 Fitting Locally Stationary Models to Data. We consider the following time varying AR model

$$
\sum_{j=0}^{p} a_j^\theta\left(\frac{t}{T}\right)X_{t-j,T} = \sigma\left(\frac{t}{T}\right)e_t \tag{16}
$$

where $a_0(u) \equiv 1$ and the $\{e_t\}$ is a sequence of *i.i.d.* random variables with mean zero and variance 1. Suppose that $a_\theta(u) = (a_1^\theta(u), \ldots, a_p^\theta(u))'$ depend on a finite dimensional parameter. We further suppose that $a_j^\theta(u)$'s are parameterized as

$$
a_j^\theta(u) = \sum_{k=1}^{K}\theta_{jk}u^{k-1}. \tag{17}
$$

Let $\theta = (\theta_{11}, \ldots, \theta_{1K}, \cdots, \theta_{p1}, \ldots, \theta_{pK})'$ and $f_k(u) = u^{k-1}$. Write $F(u) = \{f_i(u)f_j(u)\}_{i,j=1,\ldots,K}$ and $f(u) = (f_1(u), \ldots, f_K(u))'$. We denotes by $A \otimes B$ the right direct product of the matrix A and B. Let

$$
L_T(\theta) = \frac{1}{4\pi}\frac{1}{M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left[\log\{f(u,\lambda)\} + I_N(u,\lambda)f^{-1}(u,\lambda)\right]d\lambda.
$$

It is easily seen that the quasi-maximum likelihood estimator $\widehat{\theta}$ of θ defined by $L_T(\widehat{\theta}) = \max_{\theta\in\Theta} L_T(\theta)$ is given by

$$
\begin{aligned}
&\widehat{\theta}\{\sigma(u_1)^2, \ldots, \sigma(u_M)^2\} \\
&= -\left(\frac{1}{M}\sum_{j=1}^{M}\sigma(u_j)^{-2}\Sigma_N(u_j)\otimes F(u_j)\right)^{-1}\left(\frac{1}{M}\sum_{j=1}^{M}\sigma(u_j)^{-2}C_N(u_j)\otimes f(u_j)\right)
\end{aligned} \tag{18}
$$

where

$$
\begin{aligned}
c_N(u,j) &= \int_{-\pi}^{\pi} I_N(u,\lambda)\exp(i\lambda j)d\lambda \\
(19)\qquad &= \frac{1}{N}\sum_{s,t=0,s-t=j}^{N-1} X_{[Tu]-N/2+s+1,T}X_{[Tu]-N/2+t+1,T},
\end{aligned}
$$

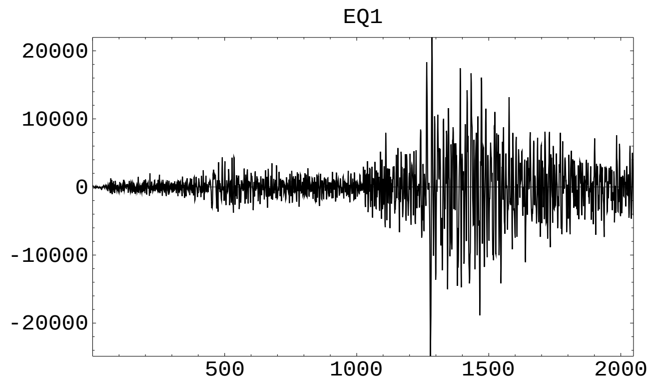

Figure 1: Natural Earthquake (EQ1)

$C_N(u) = (c_N(u,1), \ldots, c_N(u,p))'$ and $\Sigma_N(u) = \{c_N(u, i-j)\}_{i,j=1,\ldots,p}$. In the definition od $\widehat{\theta}$ we have to know the innovation variances $\sigma(u_j)^2$. Because on each time segment with midpoint u_jT, we may suppose the process $X_{t,T}$ is stationary, it is possible to estimate $\sigma(u_j)^2$ by

$$\widehat{\sigma}(u_j)^2 = 2\pi \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \alpha I_N(u_j,\lambda) d\lambda\right], \tag{20}$$

where $\alpha = \exp\gamma$ ($\gamma \simeq 0.57721$, Euler's constant) (see Taniguchi (1980)). Then we can estimate θ by $\widehat{\theta} \equiv \widehat{\theta}\{\widehat{\sigma}(u_1)^2, \ldots, \widehat{\sigma}(u_M)^2\}$. On each time segment, we can calculate the residual variance $\widehat{\sigma}^2(u_j, p, K)$ of the time varying AR model by means of the relation (16) with coefficients $a_j^{\widehat{\theta}}\left(\frac{t}{T}\right)$.

Let us discuss seismic data with length 2048. Figures 1-4 show records of seismic data from natural earthquakes (EQ1, EQ3, EQ4, EQ5). Figures 5-7 show records of seismic data from mining explosions. (EX3, EX4, EX8). Suppose $N = 2^4 = 16$ and $M = 2^7 = 128$.

From these Figures we may suppose that these seismic data have a structural change of innovation variance. Hence we divide the time interval $[1, 2^{11}]$ into $[1, 2^{10}]$ and $[2^{10}+1, 2^{11}]$. Then using (20) we estimate the variances. For these seismic data we fit the time varying AR model (16) which minimizes the AIC criterion

$$\mathrm{AIC}(p,k) = \frac{1}{M}\sum_{j=1}^{M} \log \widehat{\sigma}^2(u_j, p, K) + 2(p + p(K-1))/T, \tag{21}$$

where

$$\widehat{\sigma}^2(u_j, p, K) \approx \frac{1}{N}\sum_t \left(X_{t,T} + \sum_{k=1}^{p} a_j^{\widehat{\theta}}(u_j) X_{t-k,T}\right)^2, \quad (t \in [2^4(j-1)+1, 2^4 j]).$$

Tables 1-4 show the values of $\mathrm{AIC}(p,K)$ for EQ1 and EQ3-5.

Dargahi-Noubary (1995) claimed that many seismic source function models can be regarded as having been generated by the one-parameter third-order autoregressive model

$$(1 - \exp(-k_0)B)^3 s_j(t) = w_j(t), \tag{22}$$

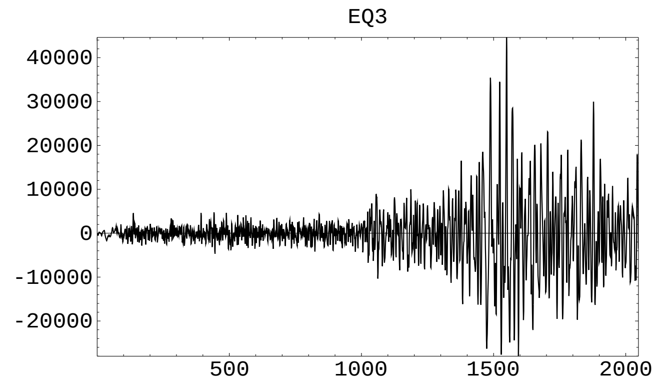

Figure 2: Natural Earthquake (EQ3)

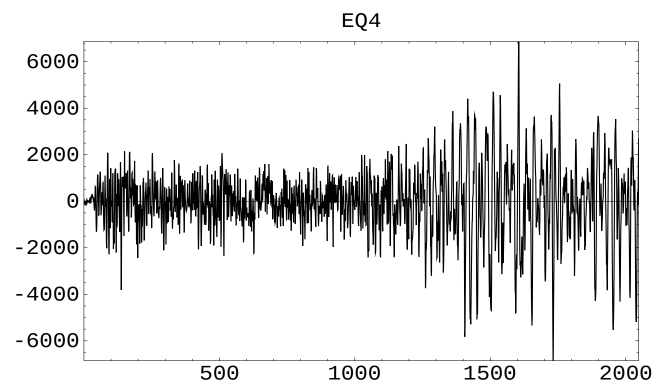

Figure 3: Natural Earthquake (EQ4)

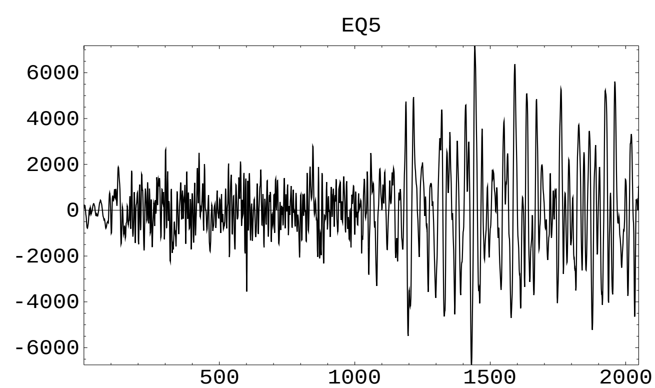

Figure 4: Natural Earthquake (EQ5)

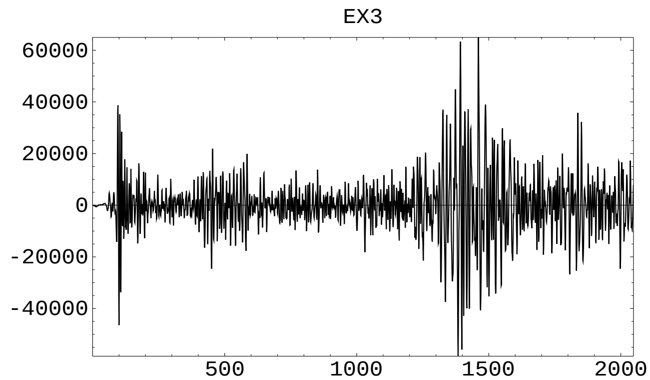

Figure 5: Mining Explosion (EX3)

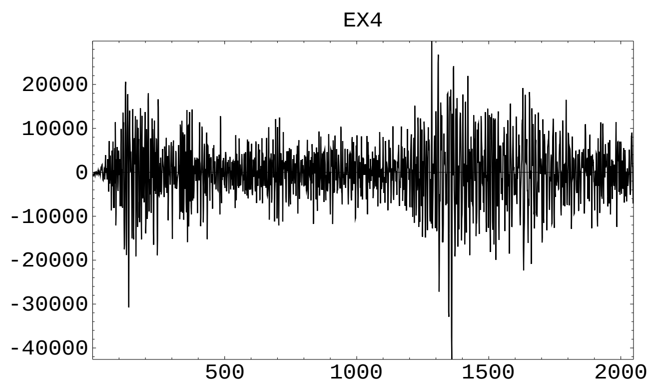

Figure 6: Mining Explosion (EX4)

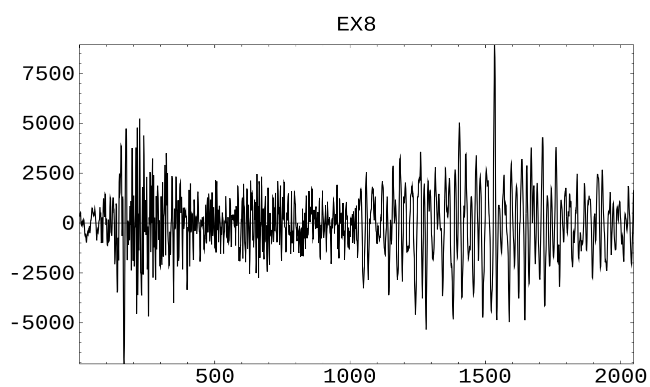

Figure 7: Mining Explosion (EX8)

Table 1: AIC values for EQ1

Poly.\TVAR	3	4	5	6	7
3	241.988	169.937	193.745	178.96	*155.307
4	288.046	260.35	159.966	206.146	183.848
5	318.18	303.784	273,105	163.471	211.676
6	345.561	345.513	308.187	279.117	173.62

Table 2: AIC values for EQ3

Poly.\TVAR	3	4	5	6	7
3	273.284	182.639	225.03	207.508	187.236
4	317.616	297.716	182.864	234.907	208.65
5	330.8	330.736	307.846	*170.304	236.491
6	390.072	379.595	337.997	314.333	189.522

Table 3: AIC values for EQ4

Poly.\TVAR	3	4	5	6	7
3	226.662	131.898	197.224	173.893	161.212
4	273.413	261.414	*129.283	216.172	187.55
5	288.146	295.972	280.505	171.2553	222.217
6	346.551	342.194	302.916	289.672	179.694

Table 4: AIC values for EQ5

Poly.\TVAR	3	4	5	6	7
3	229.313	147.462	188.155	176.621	167.685
4	268.331	257.616	*142.118	200.008	189.754
5	279.879	293.038	276.006	147.251	209.835
6	340.536	333.888	309.381	282.786	147.03

Table 5: AIC values for EX3

Poly.\TVAR	3	4	5	6	7
3	287.158	206.429	232.634	214.451	*191.373
4	340.751	318.489	230.325	248.651	217.249
5	353.192	352.109	321.352	215.89	245.88
6	405.621	394.062	350.934	322.261	226.381

Table 6: AIC values for EX4

Poly.\TVAR	3	4	5	6	7
3	274.403	209.979	219.539	201.264	*174.026
4	319.257	289.905	196.622	232.871	202.562
5	346.861	328.034	296.605	198.32	232.76
6	308.26	373.516	331.105	305.312	200.14

where B is the usual backshift operator, $\{w_j(t)\}$ is a sequence of white noise, and k_0 is an unknown parameter. By combining (22) and

$$y_j(t) = s_j(t) + \sum_{k=1}^{n} \theta_k s_j(t - kd),$$

Shumway et.al. (1998) discussed detecting of the delay d which corresponds to the spacing of the charges in the ripple-fired event.
Let us return to our results. From Tables 3 and 4, it is seen that the TVAR(5) model is preferred. However, TVAR(7) and TVAR(6) are selected in Tables 1 and 2, respectively. For mining explosions (EX3, EX4 and EX8), Tables 5 and 6 show that TVAR(7) is preferred. On the other hand, in Table 7, TVAR(5) is preferred. These results show an interesting feature of locally stationary modelling.

4 Numerical Analysis of Discriminant and Clustering Problem.

4.1 Example of Classification Problem in Seismic Data Here discriminant procedures in Section 2 are applied to the problem of discriminating seismic records from earthquakes and mining explosions. Suppose that we know EQ1 is recorded from the group of earthquakes while EX4 is from the group of explosions. Assuming that we do not know EX3 is from mining explosions, we are classify it into one of the two groups (the group of

Table 7: AIC values for EX8

Poly.\TVAR	3	4	5	6	7
3	248.423	162.696	211.047	195.06	182.264
4	285.955	276.518	*160.633	218.764	199.003
5	298.775	311.795	291.594	162.479	221.832
6	353.558	350.328	321.3	296.253	175.508

earthquakes and the group of explosions). Rules for classification are based on the distance between the estimated time varying spectral densities of the data EQ1, EX4, and EX3. We estimate these spectral densities by the method of Section 3. We fit the TVAR(7) model for EQ1, EX4 and EX3. Let $\widehat{f}_{EQ1}$, $\widehat{f}_{EX4}$, and $\widehat{f}_{EX3}$ be the estimated time varying spectral densities of EQ1, EX4, and EX3, respectively. Then the value of the criterion

$$D(\widehat{f}_{EX3}) = D(\widehat{f}_{EX3} : \widehat{f}_{EQ1}) - D(\widehat{f}_{EX3} : \widehat{f}_{EX4})$$

is positive, which means that EX3 is classified into the group of mining explosions. Next, let $\widehat{f}_{EQ5}$ be the spectral density of EQ5. We use TVAR(5) for EQ5. Suppose that we do not know EQ5 is from natural earthquakes. Similarly, we classify it into one of two groups (EQ1 and EX4). Then we observe

$$D(\widehat{f}_{EQ5}) = D(\widehat{f}_{EQ5} : \widehat{f}_{EQ1}) - D(\widehat{f}_{EQ5} : \widehat{f}_{EX4})$$

is negative. From the result, EQ5 is classified into the group of natural earthquakes. Therefore the classification rule based on $D(\,\cdot\,)$ works well.

4.2 Example of Cluster Analysis for Some Daily Returns We discuss a clustering problem for New York stock exchange data. The data are daily returns of AMOCO, FORD, HP, IBM, and MERCK companies. The individual time series are the last 1024 data points from the data representing the daily returns for the five companies from February 2, 1984, to December 31, 1991. Figures 8-12 show the graphs of the data.

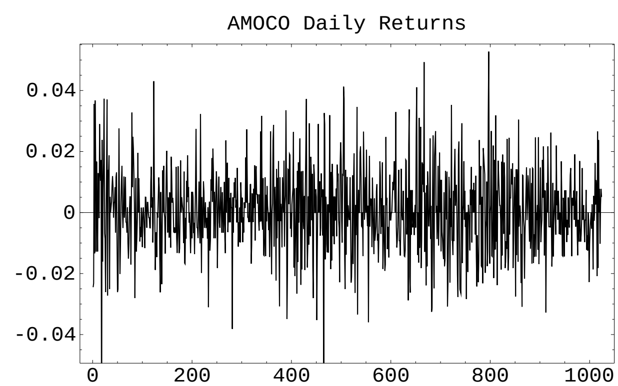

Figure 8: AMOCO Daily Returns

In the figures we can see the change of variance (volatility) with time. For such data GARCH (generalized autoregressive conditionally heteroscedastic) models are often used. Engle (1982) introduced the ARCH (autoregressive conditionally heteroscedastic) (p) model defined as

$$X_t = e_t\sqrt{U_t}, \tag{23}$$

where $\{e_t\}$ is a sequence of $i.i.d.(0,1)$ random variables, e_t is independent of $X_s,\ s<t$, and U_t evolves according to

$$U_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2. \tag{24}$$

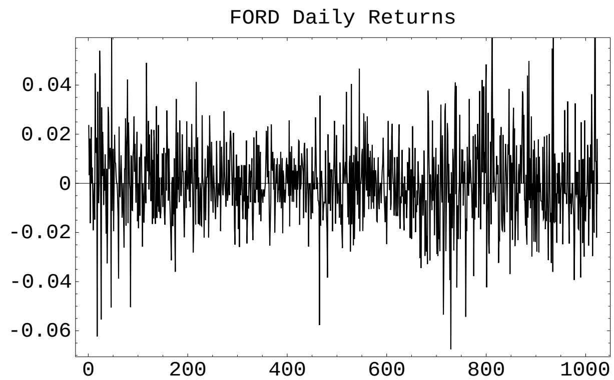

Figure 9: FORD Daily Returns

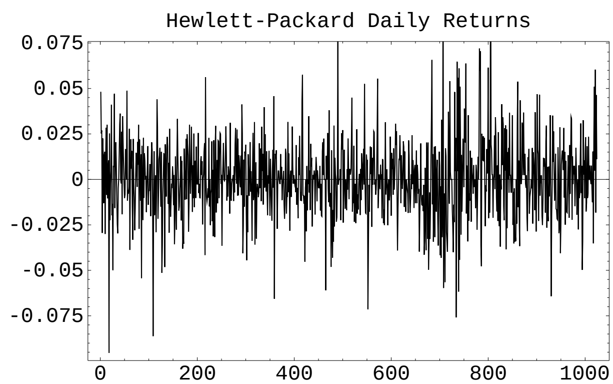

Figure 10: Hewlett-Packard Daily Returns

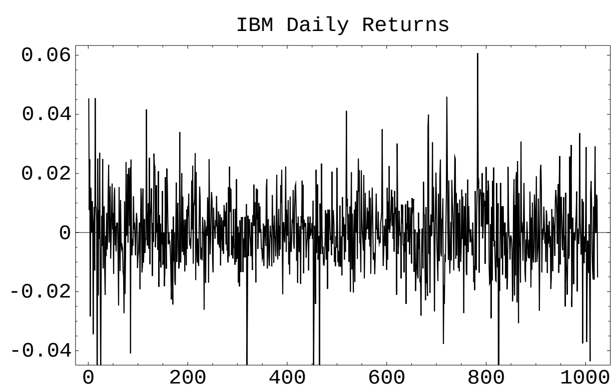

Figure 11: IBM Daily Returns

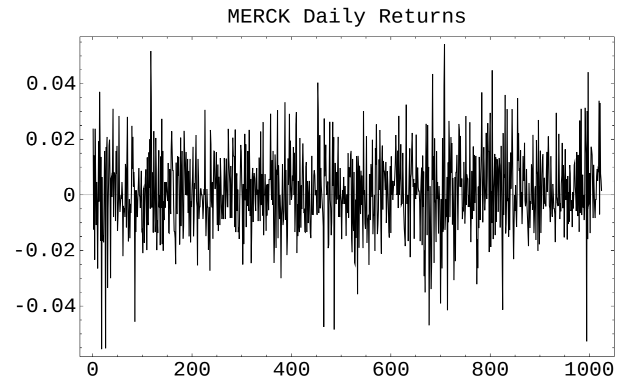

Figure 12: MERCK Daily Returns

The model in (23) can be rewritten in the form

$$Y_t = U_t + \eta_t,\ U_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} \tag{25}$$

where $Y_t = X_t^2$ and $\eta_t = U_t(e_t^2 - 1)$. Henceforth, denote by $\mathcal{F}_t$ the σ-field generated by $\{Y_t, Y_{t-1}, \dots, \}$. We note that the disturbance term, η_t, in (25) is a martingale difference since $E[\eta_t|\mathcal{F}_{t-1}] = U_t E[(e_t^2 - 1)|\mathcal{F}_{t-1}] = 0$.
Bollerslev (1986) introduced the GARCH model:

$$X_t = e_t\sqrt{U_t}, \tag{26}$$

$$U_t = \alpha_0 + \sum_{j=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j U_{t-j}, \tag{27}$$

where $\alpha_0 > 0, \alpha_j \geq 0,\ j = 1, \dots, p$, and $\beta_j \geq 0,\ j = 1, \dots, q$. Similarly as in (25), it is well known that any GARCH(p,q) process can be represented as an ARMA[max$\{p, q\}, q$] in the squared processes. In the case

$$Y_t = \alpha_0 + \sum_{j=1}^{p} (\alpha_j + \beta_j) Y_{t-j} + \eta_t - \sum_{j=1}^{q} \beta_j \eta_{t-j},$$

where $Y_t = X_t^2$, and $\{\eta_t\}$ is a martingale difference sequence (see Gouriéroux (1997)).
In Figure 13, we plot the sample autocorrelation of Hewlett- Packard. From this figure we may suppose that the time series is uncorrelated. Figure 14 shows the sample autocorrelation of the square transformed data. Here the square transformed process of $\{X_t\}$ implies the process $\{X_t^2 - \sum_{t=1}^{2^{10}} X_t^2/2^{10}\}$. From Figure 13, we may suppose that these five time series are correlated. By using AIC we attempt to fit a stationary AR(p) model to the square transformed data. The results are shown in Table 8.

Next we try to use TVAR models for the square transformed data. Let $T = 2^{10} = 1024$, $M = 2^6 = 64$, and $N = 2^4 = 16$. Assuming that innovation variances of the data are constant over time, we choose the orders of models by minimizing the AIC criterion.

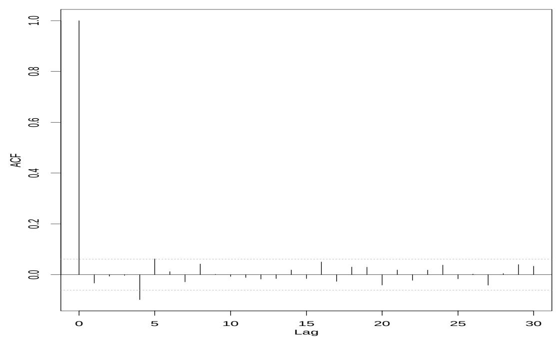

Figure 13: sample ACF (Hewlett-Packard)

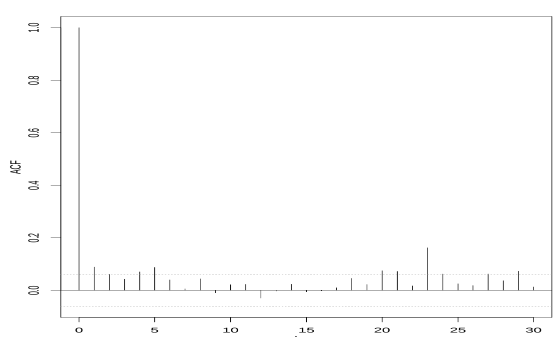

Figure 14: sample ACF of squared data (Hewlett-Packard)

Table 8: Suitable stationary AR model

	AMOCO	FORD	Hewlett-Packard	IBM	MERCK
order p	2	1	23	0	8

Table 9: AIC values for AMOCO

Poly.\TVAR	1	2	3	4	5	6
2	-288.450	-288.335	-266.884	-242.727	-222.136	-197.963
3	-310.044	-282.16	-234.657	-242.176	-203.554	-182.695
4	*-310.086	-262.427	-203.016	-188.624	-174.03	-152.357
5	-310.044	-204.142	-181.211	-135.385	-127.096	-112.651
6	-310.044	-201.677	-134.992	-107.834	-119.275	-79.2127

Tables 9-13 show these values of selected orders of TVAR model. From the results TVAR(1) model is preferred. Tables 9, 12, and 13 show that polynomial of order 4 is good. On the other hand, from Tables 10 and 11 it is seen that polynomials of order 6 and 2 are preferred, respectively. For a locally stationary, it seems that TVAR(1) model with parametric polynomial function of time shows good fitting, which makes a sharp contrast with the usual AR fitting (see Table 8).

Now we discuss hierarchical clustering techniques for locally stationary processes as follows. First, find the two elements which are closest in the sense of the distance (14). Then these two items become a cluster. Next the distance between nonclustered items and a current cluster is calculated as the average of the distances to elements in the cluster (note that many different ideas of defining the "distance between two clusters" are possible). Again, we combine the objects that are closest together, and then compute the new distance between two different clusters. Repeat the above step until all the items are merged into one cluster. This is a hierarchical clustering method.
We estimate the group (cluster :$\{f_{\hat{\theta}}^{(1)}, \dots , f_{\hat{\theta}}^{(k)}\}$) spectral density by

$$\frac{1}{k}\sum_{l=1}^{k} f_{\hat{\theta}}^{(l)}.$$

Namely, it is obtained by averaging the estimators.
Table 14 shows the results of hierarchical clustering based on (14) for the daily return data.

Table 10: AIC values for FORD

Poly.\TVAR	1	2	3	4	5	6
2	-278.103	-275.368	-258.455	-229.051	-211.99	-196.746
3	-298.134	-246.87	-231.298	-216.533	-186.549	-164.000
4	-298.351	-217.577	-193.108	-163.54	-146.726	-130.351
5	-299.038	-155.89	-149.703	-114.743	-118.136	-96.5533
6	*-299.115	-156.354	-100.024	-80.1962	-92.1226	-57.349

Table 11: AIC values for Hewlett-Packard

Poly.\TVAR	1	2	3	4	5	6
2	*-282.028	-258.570	-243.746	-217.755	-195.525	-180.637
3	-281.689	-245.256	-209.535	-194.079	-178.416	-155.544
4	-281.725	-210.843	-169.988	-168.801	-141.719	-131.395
5	-281.379	-178.851	-149.765	-123.515	-107.899	-110.007
6	-281.281	-147.924	-92.8831	-79.5077	-61.4551	-52.8566

Table 12: AIC values for IBM

Poly.\TVAR	1	2	3	4	5	6
2	-291.820	-290.825	-268.559	-247.353	-224.004	-201.401
3	-313.595	-280.943	-242.065	-207.356	-195.296	-174.433
4	*-313.716	-253.278	-200.466	-174.157	-150.874	-144.108
5	-313.448	-198.912	-175.485	-136.009	-125.312	-109.32
6	-313.337	-199.194	-124.449	-98.16	-90.831	-78.8435

Table 13: AIC values for MERCK

Poly.\TVAR	1	2	3	4	5	6
2	-289.726	-283.336	-259.329	-237.138	-216.174	-196.429
3	-310.536	-253.777	-211.043	-192.543	-175.745	-163.341
4	*-310.79	-199.811	-183.383	-173.809	-128.886	-109.336
5	-310.73	-154.172	-96.1708	-101.388	-69.8109	-60.1667
6	-313.759	-134.611	-32.3654	-29.7983	-17.4483	-23.0049

Table 14: Results of hierarchical clustering based on $\overline{\overline{D}}(\,\cdot : \cdot\,)$

No.	Minimum Distance	Hierarchical Clustering
4		(AMOCO,IBM,FORD,H-P,MERCK)
*3	9.15908	(AMOCO,IBM,FORD,H-P), (MERK)
2	8.01698	(AMOCO,IBM,FORD), (H-P), (MERCK)
1	5.7082	(AMOCO,IBM), (FORD), (H-P), (MERCK)
0	3.42466	(AMOCO), (FORD), (H-P), (IBM), (MERCK)

For the clustering problem for seismic data by use of stationary modelling, Kakizawa et al.(1998) considered the following distance measures computed from the symmetric Chernoff information divergence,

$$JB_\alpha(f:g) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\log\left\{\frac{\alpha f + (1-\alpha)g}{g}\right\} + \log\left\{\frac{\alpha g + (1-\alpha)f}{f}\right\}\right]d\lambda,$$

where $f \equiv f(\lambda)$, $g \equiv g(\lambda)$ are spectral densities for two different stationary processes. Similarly, for the locally stationary processes, we can introduce the distance measure

$$\begin{aligned} DJ_\alpha(f_{\widehat{\theta}} : g_{\widehat{\theta}}) \quad = \quad & \frac{1}{4\pi M}\sum_{j=1}^{M}\int_{-\pi}^{\pi}\left[\log\left\{\frac{\alpha f_{\widehat{\theta}}(u_j,\lambda) + (1-\alpha)g_{\widehat{\theta}}(u_j,\lambda)}{g_{\widehat{\theta}}(u_j,\lambda)}\right\}\right. \\ & \left. + \log\left\{\frac{\alpha g_{\widehat{\theta}}(u_j,\lambda) + (1-\alpha)f_{\widehat{\theta}}(u_j,\lambda)}{f_{\widehat{\theta}}(u_j,\lambda)}\right\}\right]d\lambda. \end{aligned} \tag{28}$$

For $\alpha = 0.1$, 0.3, and 0.5, Tables 15-17 show the results of hierarchical clustering based on (28) for the daily returns data. Then the following tables show that the results for $\overline{\overline{D}}(\,\cdot : \cdot\,)$ are different from those for $DJ_\alpha(\,\cdot : \cdot\,)$. Namely, in No.3 of Table 14 we get the cluster of AMOCO, IBM, FORD, and H-P. In No.3 of Tables 15-17, however, AMOCO, IBM, FORD, and MERCK are clustered.

Table 15: Results of hierarchical clustering based on $DJ_\alpha(\,\cdot : \cdot\,)$ $(\alpha = 0.1)$

No.	Minimum Distance	Hierarchical Clustering
4		(AMOCO,IBM,FORD,MERCK,H-P)
*3	0.148727	(AMOCO,IBM,FORD,MERCK), (H-P)
2	0.122676	(AMOCO,IBM,FORD), (H-P), (MERCK)
1	0.0232	(AMOCO,IBM), (FORD), (H-P), (MERCK)
0	0.0136	(AMOCO), (FORD), (H-P), (IBM), (MERCK)

Table 16: Results of hierarchical clustering based on $DJ_\alpha(\cdot : \cdot)$ ($\alpha = 0.3$)

No.	Minimum Distance	Hierarchical Clustering
4		(AMOCO,IBM,FORD,MERCK,H-P)
*3	0.295102	(AMOCO,IBM,FORD,MERCK), (H-P)
2	0.182384	(AMOCO,IBM,FORD), (H-P), (MERCK)
1	0.052211	(AMOCO,IBM), (FORD), (H-P), (MERCK)
0	0.0141	(AMOCO), (FORD), (H-P), (IBM), (MERCK)

Table 17: Results of hierarchical clustering based on $DJ_\alpha(\cdot : \cdot)$ ($\alpha = 0.5$)

No.	Minimum Distance	Hierarchical Clustering
4		(AMOCO,IBM,FORD,MERCK,H-P)
*3	0.335568	(AMOCO,IBM,FORD,MERCK), (H-P)
2	0.202399	(AMOCO,IBM,FORD), (H-P), (MERCK)
1	0.061422	(AMOCO,IBM), (FORD), (H-P), (MERCK)
0	0.020970	(AMOCO), (FORD), (H-P), (IBM), (MERCK)

In financial engineering, there is an important problem which classifies several companies into some class of credit from their financial data. Usually people in this field use the usual multivariate method for *i.i.d.* data. Thus the above analysis will give a new approach in this field.

Acknowledgements

The author would like to express his sincere thanks to Professors Shingo Shirahata and Masanobu Taniguchi for their constant encouragement and comments. Thanks are also due to Professor R.H.Shumway, University of California Davis for offering the author his earthquakes and mining explosions data.

References

[1] S.Adak, Time-dependent spectral analysis of nonstationary time series, *J. Amer. Statist. Assoc.* **93** (1998), 1488-1501.

[2] A.S.Cakmak, R.I.Sherif, and G.Ellis, Modelling earthquakes ground motions in California using parametric time series methods, *Soil Dynamics and Earthquake Engineering.* **4** (1985), 124-131.

[3] M.K.Chang, J.W.Kwiatokowski, R.F. Nau, R.M. Oliver, and K.S.Pister, ARMA models for earthquake ground motions, (1981). *Tech.Report ORC97*

[4] R.Dahlhaus, Fitting time series models to nonstationary processes, *Ann. Statist.* **25** (1997), 1-37.

[5] R.Dahlhaus, A likelihood approximation for locally stationary process, *Ann. Statist.* **28** (2000), 1762-1794.

[6] G.R.Dargahi-Noubary, "A Uniformly Modulated Nonstationary Model for Seismic Records", *In* Nonstationary Stochastic Processes and Their Applications, A.G.Miamee,(Ed.) World Scientific Publishing Co. (1992)

[7] G.R.Dargahi-Noubary, Stochastic modeling and identification of seismic records based on established deterministic formulation,*J. Time Ser. Anal.* **16** (1995), 201-220.

[8] W.Gersch, and G.Kitagawa, A time varying AR coefficient model for modeling and simulating earthquake ground motion, *Earthquake Engrg. and Struct. Dynamics.* **13** (1985), 124-131

[9] M.Y.Hussain, and T.S.Rao, The estimation of autoregressive, moving average and mixed autoregressive moving average systems with time-dependent parameters of non-stationary time series, *Int.J.Control* **23** (1976), 647-656.

[10] Y.Kakizawa, Discriminant analysis for non-Gaussian vector stationary processes. *J. Nonpara. Statist.* **7** (1996), 187-203.

[11] Y.Kakizawa, R.H.Shumway, and M.Taniguchi, Discrimination and clustering for multivariate time series, *J. Amer. Statist. Assoc.* **93** (1998), 328-340.

[12] J.R.Magnus, and H.Neudecker, "Matrix Differential Calculus with Applications in Statistics and Econometrics," Chichester: Wiley, (1988).

[13] N.W.Polhemus, and A.S.Cakmak, Simulation of earthquake ground motions using autoregressive moving average (ARMA) models, *Earthquake Eng. and Struct. Dynamics*, **9** (1981), 343-354.

[14] K.Sakiyama, and M.Taniguchi, Discriminant analysis for locally stationary process, Preprint Series S-58, Osaka University, Department of Mathematics Science. (2001)

[15] R.H.Shumway, and A.N.Unger, Linear discriminant functions for stationary time series, *J. Am. Stat. Assoc.* **69** (1974), 948-956.

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

[17] R.H.Shumway, and D.S.Stoffer, "Time Series Analysis and Its Applications," Springer-Verlag, New York, 2000.

[18] R.H.Shumway, D.R.Baumgardt, and Z.A.Der, A cepstral F statistic for detecting delay-fired seismic signals, *Technometrics* **40** (1998), 100-110.

[19] M.Taniguchi, On estimation of the integrals of certain functions of spectral density, *J. Appl. Prob.* **17** (1980), 73-83.

[20] M.Taniguchi, and Y.Kakizawa, "Asymptotic Theory of Statistical Inference for Time Series," Springer-Verlag, New York, 2000.

[21] J.Yuan, and T.S.Rao, Classification of textures using second-order spectra, *J. Time Ser. Anal.* **13** (1992), 547-562.

[22] G.Q.Zhang, and M.Taniguchi, Discriminant analysis for stationary vector time series, *J. Time Ser. Anal.* **15**, (1994), 117-126.

[23] G.Q.Zhang, and M.Taniguchi, Nonparametric approach for discriminant analysis in time series. *J. Nonpara. Statist.* **5** (1995), 91-10

Department of Mathematical Science Faculty of Engineering Science, Osaka University,
Machikaneyama 1-3 Toyonaka, Osaka 560-8531, Japan
Email: sakiyama@sigmath.es.osaka-u.ac.jp