Recursive segmentation procedure based on the Akaike information criterion test

Aki-Hiro SATO
Department of Applied Mathematics and Physics
Graduate School of Informatics
Kyoto University
aki@i.kyoto-u.ac.jp
JAPAN
Outline

• Background and Motivation
• Segmentation procedure based on Akaike Information Criterion
• Numerical Analysis
• Empirical analysis and discussion
• Summary


Motivation

• Capitalism has no purposes
• This is just a mechanism to develop a system by using profit which was gained from the system.
• Firms in modern capitalism make a business cycle.
• The boom-bust cycle is observed in financial markets.
Business cycle

development
saturation
destruction
construction

time

size of system

construction
destruction
saturation
development
construction
destruction
saturation
development
construction
Log returns of foreign exchange rates

- EUR/JPY (Jan, 2000 to Dec 2009)
- The nonstationarity of financial time series is one of the important properties
How do we deal with nonstationary time series

• How do we treat the nonstationary time series observed in financial markets?
• Both simple and reliable method regarding statistical significance level of test statistics should be desired.
• The time series segmentation can be used for this purpose.
A method to segment non-stationary time series into stationary segments

- Goldfeld and Quandt (1973)

\[ h(y) = \frac{\lambda}{(2\pi)^{\frac{1}{2}} \sigma_1} \exp \left\{ -\frac{(y-a_1-b_1x)^2}{2\sigma_1^2} \right\} + \frac{1-\lambda}{(2\pi)^{\frac{1}{2}} \sigma_2} \exp \left\{ -\frac{(y-a_2-b_2x)^2}{2\sigma_2^2} \right\} \]

- A recursive entropic scheme (2012)

\[ \Delta(t) = \log L_2(t) - \log L_1 \]

\[ = \sum_{s=1}^{t} \log \frac{1}{\sqrt{2\pi}\sigma_L^2} \exp \left\{ -\frac{(r(s)-\mu_L)^2}{2\sigma_L^2} \right\} + \sum_{s=t+1}^{T} \log \frac{1}{\sqrt{2\pi}\sigma_R^2} \exp \left\{ -\frac{(r(s)-\mu_R)^2}{2\sigma_R^2} \right\} - \sum_{t=1}^{T} \log \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left\{ -\frac{(r(s)-\mu)^2}{2\sigma^2} \right\} \]


One-dimensional case

Nonstationary time series are assumed to consist of several locally stationary time series with different statistics.

The time series consisting of 4 segments. Each segment is sampled from a zero-mean normal distribution with different variance. The variance is set as 2, 3, 1, and 4 from the left segment. The length of each segment is set as 500.
Segmentation Procedure

- Let \( g(x; \mu, \sigma^2) \) be a Gaussian density function parameterized by \( \mu \) and \( \sigma \):
  \[
  g(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]
  \]

- Assume two types of likelihood functions for observations \( x_s \) \((s=1, \ldots, T)\):

  **Null hypothesis:**
  \[
  L_1 = \prod_{s=1}^{T} g(x_s; \mu, \sigma^2)
  \]

  **Alternative hypothesis:**
  \[
  L_2(t) = \prod_{s=1}^{t} g(x_s; \mu_L, \sigma^L_s) \prod_{s=t+1}^{T} g(x_s; \mu_R, \sigma^R_s)
  \]

- Define the log-likelihood ratio between \( L_1 \) and \( L_2(t) \):
  \[
  \Delta(t) = \log L_1 - \log L_2(t)
  \]
Approximation of Likelihood-ratio

\[ \Delta(t) = \sum_{s=1}^{t} \log g(x_s; \mu_L, \sigma^2_L) + \sum_{s=t+1}^{n} \log g(x_s; \mu_R, \sigma^2_R) - \sum_{s=1}^{n} \log g(x_s; \mu, \sigma^2) \]

\[ \approx nH[g(x_s; \mu, \sigma^2)] - tH[g(x_s; \mu_R, \sigma^2_R)] - (n-t)H[g(x_s; \mu_L, \sigma^2_L)] \]

where \( H[g(x; \mu, \sigma^2)] = -\int_{-\infty}^{\infty} g(x; \mu, \sigma^2) \log g(x; \mu, \sigma^2) \, dx \)

\[ \Delta(t) = n \log \sigma - t \log \sigma_L - (n-t) \log \sigma_R \geq 0 \]
Segmentation procedure

Replacing $\sigma$, $\sigma_L$, and $\sigma_R$ as the maximum likelihood estimator

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{s=1}^{n} \left( x_s - \frac{1}{n} \sum_{s=1}^{n} x_s \right)^2,$$

$$\hat{\sigma}^2_L = \frac{1}{t} \sum_{s=1}^{t} \left( x_s - \frac{1}{t} \sum_{s=1}^{t} x_s \right)^2, \quad \hat{\sigma}^2_R = \frac{1}{n-t} \sum_{s=t+1}^{n} \left( x_s - \frac{1}{n-t} \sum_{s=t+1}^{n} x_s \right)^2,$$

one gets

$$\Delta(t) = n \log \hat{\sigma} - t \log \hat{\sigma}_L - (n-t) \log \hat{\sigma}_R$$

$$t^* = \arg \max_t \Delta(t)$$
Recursive segmentation procedure

The segmentation procedure is applied to each segment recursively. The termination condition is given by $\Delta_c$. If $\max \Delta(t)$ is less than $\Delta_c$, then we do not apply the segmentation procedure any more.
Wilks’s theorem

- The threshold value $\Delta_c$ is related to the significance level of acceptance. According to Wilks’s theorem, the probability density of $2\Delta(t)$ is approximately sampled from a chi-squared density with a degree of freedom equal to the difference $k$ between the number of free parameters.

$$p(2\Delta) = \frac{1}{2^k k} \left(2\Delta\right)^{k-1} e^{-\Delta}$$
Information criterion test for M-dimensional multiple time series

• The Akaike information criterion of a model with $K$ model parameters for $T$ observations,

$$AIC = -2L(\hat{\theta}) + 2(K + 1)$$

where $L(\hat{\theta})$ is the likelihood value of the model with the maximum likelihood estimator $\hat{\theta}$. 
Significance level

\[ \Delta_{AIC}(t) = AIC_2(t) - AIC_1 = 2\Delta(t) + 2(k_2 - k_1) \]

\[
\Pr[\Delta_{AIC} \leq x] = \frac{1}{2^{(k_2-k_1)/2}} \frac{1}{\Gamma\left(\frac{k_2-k_1}{2}\right)} \int_0^x \left(y - 2(k_2 - k_1)\right)^{\frac{k_2-k_1}{2}-1} e^{-\frac{y+2(k_2-k_1)}{2}} dy
\]

\[
= \gamma\left(x - 2(k_2 - k_1); \frac{k_2-k_1}{2}\right)
\]

\(\gamma(x; a): \text{Regularized incomplete gamma function}\)
Artificial time series

Let assume the time series consisting of 3 segments. The first 100 points are sampled from a normal distribution with mean 0.1 and standard deviation 1.0. The second 150 points are sampled from a normal distribution with mean -0.2 and standard deviation 1.5, and the third 200 points are sampled from a normal distribution with zero-mean and standard deviation 3.0.
Numerical Study

From the time series, how we can determine segment boundary?

The segmentation procedure is applied to this artificial time series with $\Delta_c=10$.

The segmentation procedure can separate the time series into 3 elements.

<table>
<thead>
<tr>
<th>segment</th>
<th>start</th>
<th>end</th>
<th>length</th>
<th>mean</th>
<th>std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>85</td>
<td>85</td>
<td>0.256684</td>
<td>1.024991</td>
</tr>
<tr>
<td>2</td>
<td>86</td>
<td>299</td>
<td>214</td>
<td>-0.267535</td>
<td>1.404834</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>550</td>
<td>251</td>
<td>-0.131082</td>
<td>3.018483</td>
</tr>
</tbody>
</table>
Empirical Analysis of Japanese securities

• The data on daily prices of companies listed on the first section of the Tokyo Stock Exchange for the period from 4\textsuperscript{th} January, 2000 to 11\textsuperscript{th} December 2010 (2,675 business days).

• 1,413 companies which last more than 10 years are selected; Major Japanese companies such as Toyota, Hitachi, Panasonic, and so on, are included.
Data and parameters

• Log-returns of stock $i$ are defined as

$$x_{i,t} = \log E_{i,t} - \log O_{i,t} \quad (i = 1, \ldots, M; t = 1, \ldots, N),$$

where $E_{i,t}$ and $O_{i,t}$ are ending and opening prices at day $t$.

• The number of stocks: $M=1,413$

• The number of observations: $n=2,675$

• $\Delta_c$ is fixed as 10.
The daily number of segment boundaries

(I) 2003 to 2007
(II) The end of 2007
(III) September 2008
(IV) March 2090
(V) April 2011
(a) June 2000
(b) April 2004
(c) February 2006
(d) 2007 to 2009
(e) March 11, 2011
Quintile statistics

• Classify segments into 5 categories from the smallest variance:
  – Category 1 is 1%-20%
  – Category 2 is 21%-40%
  – Category 3 is 41%-60%
  – Category 4 is 61%-80%
  – Category 5 is 81%-100%
The number of stocks in each category

Results from time series for the period from 2000 to 2012

(I) 2003 to 2007
(II) The end of 2007
(III) September 2008
(IV) March 2990
(V) April 2011

Results from time series for the period from 2000 to 2012
Results

• From the end of 2002, the number of the fourth and fifth quintiles decreased (Recovery phase)
• From the beginning of 2004, the number of the first quintile increased (Recovery phase)
• From 2004 to 2007, the number of the first and second quintiles took high values (boom phase)
• From the end of 2007, the number of the first quintile sharply dropped (Crash phase)
• At the end of 2008, the number of the fifth quintile sharply increased (Crisis phase)
• From the end of 2009, the number of the first quintile increased and the number of the fifth quintile decreased (Recovery phase)
Stability index

Stability index = #1 + #2 - #4 - #5
Stability index = #1+#2-#4-#5
Multivariate case

Segmentation procedure for multi-dimensional time series

• Let \( r_i(s) (i=1, \ldots M; s=1, \ldots, T) \) be the \( M \)-dimensional multiple log-return time series defined as \( r_i(s) = R_i(s+\Delta t) - R_i(s) \), where \( R_i(s) (s=1, \ldots, T+1) \) is the exchange rate of \( i \)-th currency pair at time \( s \).

• Let us assume that the multivariate time series consists of \( n \) segments sampled from \( n \) different multivariate normal distributions \((k=1, \ldots, n)\).

\[
p(r; \mu^{(k)}, C^{(k)}) = \frac{1}{(2\pi)^{M/2}|C^{(k)}|^{1/2}} \exp \left[ -\frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} (C^{(k)})^{-1}_{ij} (r_i - \mu_i^{(k)})(r_j - \mu_j^{(k)}) \right]
\]
Likelihood-ratio

- To determine the $n$ stationary segments from the given multiple time series $r(s) = (r_1(s), r_2(s), \ldots, r_M(s))$, we employ the recursive segmentation procedure.

\[ L_1 = \prod_{s=1}^{T} p(r(s); \mu, C), \]

The null model:

\[ L_2(t) = \prod_{s=1}^{t} p(r(s); \mu^{(L)}, C^{(L)}) \prod_{s=t+1}^{T} p(r(s); \mu^{(R)}, C^{(R)}) \]

The alternative model:
Likelihood-ratio

\[ \Delta(t) = \log L_1 - \log L_2(t) \]

\[ = \sum_{s=1}^{T} \log p(r(s); \mu, C) - \sum_{s=1}^{t} \log p(r(s); \mu^{(L)}, C^{(L)}) - \sum_{s=t+1}^{T} \log p(r(s); \mu^{(R)}, C^{(R)}) \]

\[ \approx \frac{t}{2} \log|C^{(L)}| + \frac{T-t}{2} \log|C^{(R)}| - \frac{T}{2} \log|C| \]
Likelihood-ratio

\[ \Delta(t) \approx \frac{t}{2} \log \left| \mathbf{C}^{(L)} \right| + \frac{T-t}{2} \log \left| \mathbf{C}^{(R)} \right| - \frac{T}{2} \log \left| \mathbf{C} \right| \]

\[ \approx \frac{t}{2} \log \left| \hat{\mathbf{C}}^{(L)} \right| + \frac{T-t}{2} \log \left| \hat{\mathbf{C}}^{(R)} \right| - \frac{T}{2} \log \left| \hat{\mathbf{C}} \right| \]

\[ \hat{\mu}_i^{(L)} = \frac{1}{t} \sum_{s=1}^{t} r_i(s), \quad \hat{\mathcal{C}}_{ij}^{(L)} = \frac{1}{t} \sum_{s=1}^{t} \left( r_i(s) - \mu_i^{(L)} \right) \left( r_j(s) - \mu_j^{(L)} \right) \]

\[ \hat{\mu}_i^{(R)} = \frac{1}{T-t} \sum_{s=t+1}^{T} r_i(s), \quad \hat{\mathcal{C}}_{ij}^{(R)} = \frac{1}{T-t} \sum_{s=t+1}^{T} \left( r_i(s) - \mu_i^{(L)} \right) \left( r_j(s) - \mu_j^{(R)} \right) \]

\[ \hat{\mu}_i = \frac{1}{T} \sum_{s=1}^{T} r_i(s), \quad \hat{\mathcal{C}}_{ij}^{(L)} = \frac{1}{T} \sum_{s=1}^{T} \left( r_i(s) - \mu_i \right) \left( r_j(s) - \mu_j \right) \]
General case

- By using M-dimensional probability density \( p(r; \theta) \) parameterized by \( \theta, \theta_L, \theta_R \), we can write \( \Delta(t) \) as

\[
\Delta(t) = \sum_{s=1}^{T} \log p(r; \theta) - \sum_{s=1}^{t} \log p(r; \theta_L) - \sum_{s=t+1}^{T} \log p(r; \theta_R)
\]

\[
\approx tH[p; \theta_L] + (T-t)H[p; \theta_R] - TH[p; \theta]
\]

where \( H[p; \theta] \) is Shannon entropy defined as

\[
H[p; \theta] = - \int_{-\infty}^{\infty} dr_1 \cdots \int_{-\infty}^{\infty} dr_M p(r; \theta) \log p(r; \theta)
\]
Likelihood-ratio

Empirically the log-likelihood-ratio is computed from maximum likelihood estimators.

\[ \Delta(t) \approx tH[p; \theta_L] + (T-t)H[p; \theta_R] - TH[p; \theta] \]

\[ \approx tH[p; \hat{\theta}_L] + (T-t)H[p; \hat{\theta}_R] - TH[p; \hat{\theta}] \]

\[ \hat{\theta}_L = \arg \max_{\theta} \sum_{s=1}^{t} \log p(r(s); \theta_L) \]

\[ \hat{\theta}_R = \arg \max_{\theta} \sum_{s=t+1}^{T} \log p(r(s); \theta_R) \]

\[ \hat{\theta} = \arg \max_{\theta} \sum_{s=1}^{T} \log p(r(s); \theta) \]
AIC for the \( i.i.d. \) \( M \)-dimensional Gaussian distribution with \( T \) observations

\[
AIC_1 = T \log |\hat{C}| + TM \log(2\pi) + TM^2 + (M^2 + 3M + 2)
\]

\[
AIC_2(t) = t \log |\hat{C}_L| + (T - t) \log |\hat{C}_R| + TM \log(2\pi) + TM^2 + 2(M^2 + 3M + 2)
\]

\[
\Delta AIC(t) = AIC_2(t) - AIC_1
\]

\[
= t \log |\hat{C}_L| + (T - t) \log |\hat{C}_R| - T \log |\hat{C}| + (M^2 + 3M + 2)
\]

\[
= 2\Delta(t) + (M^2 + 3M + 2)
\]
Recursive segmentation procedure

• The spectrum of $\Delta(t)$ has a minimum at some time that we denote by $t^*$,

$$t^* = \arg \min_t \Delta_{AIC}(t)$$

• As the termination condition, a statistical significance level computed from the bootstrap distribution.
Let $\gamma (0 < \gamma < 1)$ be a ratio to determine the Jackknife sequences $x_L(s) (s = \tau_k, \ldots, \tau_k + \lfloor t^* \gamma \rfloor)$, where $\tau_k$ is randomly selected with the same probability for $0 < \tau_k < t - \lfloor t \gamma \rfloor + 1$; $x_R(s) (s = \tau_k', \ldots, \tau_k' + \lfloor (T-t^*) \gamma \rfloor)$, where $\tau_k$ is randomly selected with the same probability for $t^* < \tau_k' < T - \lfloor (T-t^*) \gamma \rfloor + 1$.

From Jackknife sequences the Jackknife test statistic through Jackknife variance-covariance matrices.
The Jackknife test statistic

- The Jackknife test statistics is computed from

\[
\Delta AIC(t^*) = t^* \log |\tilde{C}_L| + (T - t^*) \log |\tilde{C}_R| - T \log |\tilde{C}| + (M^2 + 3M + 2)
\]

where

\[
\tilde{\mu}_i^{(L)} = \frac{1}{t^*} \sum_{s = \tau_k}^{\tau_k + t^* \gamma} r_i(s)
\]

\[
\tilde{C}_{ij}^{(L)} = \frac{1}{t^*} \sum_{s = \tau_k}^{\tau_k + t^* \gamma} (r_i(s) - \mu_i^{(L)})(r_j(s) - \mu_j^{(L)})
\]

\[
\tilde{\mu}_i^{(R)} = \frac{1}{(T - t^*) \gamma} \sum_{s = \tau_k}^{\tau_k + (T - t^*) \gamma} r_i(s)
\]

\[
\tilde{C}_{ij}^{(R)} = \frac{1}{(T - t^*) \gamma} \sum_{s = \tau_k}^{\tau_k + (T - t^*) \gamma} (r_i(s) - \mu_i^{(L)})(r_j(s) - \mu_j^{(R)})
\]

\[
\tilde{\mu}_i = \frac{1}{T \gamma} \left( \sum_{s = \tau_k}^{\tau_k + t^* \gamma} r_i(s) + \sum_{s = \tau_k}^{\tau_k + (T - t^*) \gamma} r_i(s) \right)
\]

\[
\tilde{C}_{ij}^{(L)} = \frac{1}{T \gamma} \left( \sum_{s = \tau_k}^{\tau_k + t^* \gamma} (r_i(s) - \mu_i^{(L)})(r_j(s) - \mu_j^{(L)}) + \sum_{s = \tau_k}^{\tau_k + (T - t^*) \gamma} (r_i(s) - \mu_i^{(L)})(r_j(s) - \mu_j^{(R)}) \right)
\]
The estimation error of the test statistics

The statistical significance level is estimated from the historical probability density

$$\Pr[\Delta_{AIC}(t^*) < 0] \approx \frac{K[\hat{\Delta}_{AIC}(t^*) < 0]}{K}$$

where $K[\hat{\Delta}_{AIC}(t^*) < 0]$ is the number of events where $\Delta_{AIC}(t^*) < 0$ is satisfied.

$K$ : the number of bootstrap trials.
Recursive segmentation procedure

\[ t^* = \arg \min_t \Delta(t) \]
Sample error

• The sample (empirical) variance-covariance matrix is noise-dressed because of finiteness of time series. Such sample errors can be evaluated from eigenvalue distributions.

• The eigenvalue distribution of an empirical variance-covariance matrix also depends on the ratio between the length of the data set $T$ and the dimension of multivariate time series $M$. 
Eigenvalue density representation

\[ \Delta_{AIC}(t) \approx \frac{T}{2} \log |C| - \frac{t}{2} \log |C^{(L)}| - \frac{T-t}{2} \log |C^{(R)}| + (M^2 + 3M + 2) \]

\[ = \frac{T}{2} \sum_{i=1}^{M} \log \lambda_i - \frac{t}{2} \sum_{i=1}^{M} \log \lambda_i^{(L)} - \frac{T-t}{2} \sum_{i=1}^{M} \log \lambda_i^{(R)} + (M^2 + 3M + 2) \]

\[ = \frac{MT}{2} \left( \int_{0}^{\infty} \rho(\lambda) \log \lambda d\lambda - \frac{t}{T} \int_{0}^{\infty} \rho_L(\lambda) \log \lambda d\lambda - \frac{T-t}{T} \int_{0}^{\infty} \rho_R(\lambda) \log \lambda d\lambda \right) + (M^2 + 3M + 2) \]

where \( \lambda, \lambda_i^{(L)}, \) and \( \lambda_i^{(R)} \) represent \( i \)-th eigenvalues of the corresponding variance-covariance matrices: \( C, C^{(L)} \) and \( C^{(R)} \), respectively. \( \rho(\lambda), \rho_L(\lambda), \) and \( \rho_R(\lambda) \), respectively, represent spectra of these matrices.
Sample error

• In the case of $M$-dimensional multivariate Gaussian uncorrelated identically distributed random variables ($M/T < 1$), the density of the eigenvalues of the sample variance-covariance matrix is approximated by the Marčenko-Pastur density:

$$
\rho(\lambda) = \begin{cases} 
\frac{T}{M} \frac{\sqrt{(\lambda_- - \lambda)(\lambda_+ - \lambda)}}{2\pi\sigma^2 \lambda} & (\lambda_- \leq \lambda \leq \lambda_+) \\
0 & \text{(otherwise)}
\end{cases}
$$

where $\lambda_{\pm} = \sigma^2 \left(1 \pm \sqrt{M/T}\right)^2$ and $\sigma^2$ is a scale factor related to the variance of individual degree of freedom.
Sample error

- At $M = T$, the eigenvalue density can be approximated as
  \[ \rho(\lambda) = \frac{1}{2\pi\sigma^2} \sqrt{\frac{\lambda_+ - \lambda}{\lambda}} \]

where $\lambda_+ = 4\sigma^2$. In this case, the integrand in $\Delta(t)$ becomes singular at $\lambda = 0$ and in effect, $\Delta(t)$ is ill-defined. From statistical uncertainty when $M$ approaches $T$ and, thus, this makes it practically impossible to estimate $t^*$ properly.
The situation is even worse for $M/T > 1$, since then the density has a peak at $\lambda = 0$.

The integrand of $\Delta(t)$ is well defined only if $T > M$

Typically, to distinguish two eigenvalues of the covariance matrix one needs $M > aT$, where the coefficient $a$ is inversely proportional to the square of the difference of the eigenvalues in equations.

Throughout this analysis we set $a = 3$.

The length of each segment should be greater than $3M$. 
Numerical Analysis

• Test data

\[
x(t) = \begin{cases} 
    A_1 \sigma_1 \xi(t) & (1 \leq t \leq 100) \\
    A_2 \sigma_2 \xi(t) & (101 \leq t \leq 200) \\
    A_3 \sigma_3 \xi(t) & (201 \leq t \leq 300) \\
    A_4 \sigma_4 \xi(t) & (301 \leq t \leq 400)
  \end{cases}
\]

\(\xi_i(t)\) is drawn from \(i.i.d\). standard normal distributions:

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \xi_i(t) \xi_j(t) = 0, \quad \lim_{T \to \infty} \frac{1}{T} \sum_{i=1}^{T} \xi_i(t)^2 = 1
\]

• The variance-covariance matrix at segment \(k\)

\[
C^{(k)} = \frac{1}{100} \sum_{t=1+100(k-1)}^{100k} A_k \sigma_k \xi(t) \xi(t)^t \sigma_k A_k^t \approx \sigma_k^2 A_k E A_k^t = \sigma_k^2 A_k A_k^t
\]
Numerical simulation

\[ M = 10 \]
\[ T = 400 \]
\[ \gamma = 0.3 \]
\[ \alpha_{th} = 0.01 \]
\[ K = 1,000 \]
Empirical analysis

• 30 currency pairs consisting of AUD, BRL, CAD, CHF, EUR, GBP, JPY, MXN, NZD, SGD, USD, and ZAR


• Durations: January 3, 2001 to December 30, 2011.

• There are 2,760 data points in this multiple time series.
## Segments

<table>
<thead>
<tr>
<th>k</th>
<th>Start date</th>
<th>End date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2001-01-03</td>
<td>2001-10-15</td>
</tr>
<tr>
<td>2</td>
<td>2001-10-16</td>
<td>2002-12-31</td>
</tr>
<tr>
<td>3</td>
<td>2003-01-02</td>
<td>2004-07-07</td>
</tr>
<tr>
<td>4</td>
<td>2004-07-08</td>
<td>2007-01-02</td>
</tr>
<tr>
<td>5</td>
<td>2007-01-03</td>
<td>2007-07-26</td>
</tr>
<tr>
<td>6</td>
<td>2007-07-27</td>
<td>2008-09-11</td>
</tr>
<tr>
<td>7</td>
<td>2008-09-12</td>
<td>2009-05-04</td>
</tr>
<tr>
<td>8</td>
<td>2009-05-05</td>
<td>2010-03-02</td>
</tr>
<tr>
<td>9</td>
<td>2010-03-15</td>
<td>2010-12-31</td>
</tr>
<tr>
<td>10</td>
<td>2011-01-04</td>
<td>2011-05-10</td>
</tr>
<tr>
<td>11</td>
<td>2011-05-11</td>
<td>2011-12-30</td>
</tr>
</tbody>
</table>

- $M=30$
- $K=1000$
- $\alpha_{th}=0.1$
- $\gamma=0.3$
- 11 segments

- Paribas shock
- Lehman shock
- Euro shock
- US debt ceiling crisis
<table>
<thead>
<tr>
<th>k</th>
<th>Start date</th>
<th>End date</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2001-01-03</td>
<td>2001-10-15</td>
<td>0.0002121010</td>
<td>0.0000612156</td>
</tr>
<tr>
<td>2</td>
<td>2001-10-16</td>
<td>2002-12-31</td>
<td>0.0001779307</td>
<td>0.0000416606</td>
</tr>
<tr>
<td>3</td>
<td>2003-01-02</td>
<td>2004-07-07</td>
<td>-0.0002222775</td>
<td>0.0000383826</td>
</tr>
<tr>
<td>4</td>
<td>2004-07-08</td>
<td>2007-01-02</td>
<td>-0.0000401190</td>
<td>0.0000208055</td>
</tr>
<tr>
<td>5</td>
<td>2007-01-03</td>
<td>2007-07-26</td>
<td>-0.0004537986</td>
<td>0.0000141872</td>
</tr>
<tr>
<td>6</td>
<td>2007-07-27</td>
<td>2008-09-11</td>
<td>0.0003910212</td>
<td>0.0000526633</td>
</tr>
<tr>
<td>7</td>
<td>2008-09-12</td>
<td>2009-05-04</td>
<td>0.0002183230</td>
<td>0.0002930646</td>
</tr>
<tr>
<td>8</td>
<td>2009-05-05</td>
<td>2010-03-12</td>
<td>-0.0008621163</td>
<td>0.0000433654</td>
</tr>
<tr>
<td>9</td>
<td>2010-03-15</td>
<td>2010-12-31</td>
<td>-0.0006908856</td>
<td>0.0000584935</td>
</tr>
<tr>
<td>10</td>
<td>2011-01-04</td>
<td>2011-05-10</td>
<td>0.0001761222</td>
<td>0.0000457062</td>
</tr>
<tr>
<td>11</td>
<td>2011-05-11</td>
<td>2011-12-30</td>
<td>-0.0003064625</td>
<td>0.0000477169</td>
</tr>
</tbody>
</table>
Conclusion

• An information criterion (AIC) test was proposed for a mixture of multivariate Gaussian distribution.

• The proposed method was confirmed to detect the segmented boundary with 6% relative error.

• The proposed method was applied for a log-return time series consisting of 30 currency pairs and 11 segments were obtained.

• Some segments were confirmed to correspond to critical events such as the Paribas shock in 2007, the Lehman shock in 2008, the European Sovereign debt crisis in 2010, and the US debt-ceiling crisis in 2011.
Thank you for your kind attention

Aki-Hiro Sato (sato.akihiro.5m@kyoto-u.ac.jp)

Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University