한국통계학회논문집 제3권 1호
1996년 4월 pp. 207-217

Change Analysis with the Sample Fourier Coefficients

Jaehee Kim[1)]

Abstract

The problem of detecting change with independent data is considered. The asymptotic distribution of the sample change process with the sample Fourier coefficients is shown as a Brownian Bridge process. We suggest to use dynamic statistics such as a sample Brownian Bridge and graphs as statistical animation. Graphs including change PP plots are given by way of illustration with the simulated data.

1. Introduction

The change problem is of great practical importance. It tests the hypothesis that they are identically distributed. Parzen (1979) attempted an approach to statistical data analysis in parametric and nonparametric ways and used his idea for change analysis (1992). Kander and Zacks (1966) suggested test procedures for possible changes in the location parameters when the change-point is unknown. Zacks (1983) conducted a survey of tests for the change problem with fixed sample and sequential procedures. Eubank and Hart (1992) showed that the goodness-of-fit test can be used in testing for possible changes.

In this paper, we are concerned with the change problem. Detection of change using dynamic statistics which is Wiener or Brownian Bridge process will be introduced. The sample Fourier series as a data transformation is used in the change analysis. Also we approach change analysis graphically.

2. The Basic Change Problem and the Analysis

Let $Y_1, \dots, Y_n$ be a sequence of independent continuous random variables with mean μ_i's and common finite variance. For the basic change problem, the hypotheses of interest are

$$H_0:\ \mu_1 = \mu_2 = \cdots = \mu_n = \mu \tag{2.1}$$

1) Assistant Professor, Department of Statistics, Duksung Women's University, Seoul, 132-714, Korea

which is the homogeneity or no change against

$$H_1: \mu_1 = \mu_2 = \cdots = \mu_m = \mu, \ \mu_{m+1} = \cdots = \mu_n = \mu + \delta, \quad m, \delta \text{ unknown.} \tag{2.2}$$

Let μ and σ denote the true mean and denote the standard deviation under H_0.

The sample Brownian Bridge is defined as a piecewise constant process on $0 < t < 1$:

$$B(t) = \frac{1}{n} \sum_{j=1}^{[nt]} (Y_j - \hat{\mu})/\hat{\sigma} \quad \text{with jumps at } t_j = \frac{j}{n}$$

where $\hat{\mu} = \hat{E}(Y) = \frac{1}{n} \sum_{j=1}^{n} Y_j = \overline{Y}$, $\widehat{\sigma^2} = \widehat{Var}(Y) = \hat{E}((Y - \hat{\mu})^2)$ and $[nt]$ is the largest integer that does not exceed nt.

Let us define normalized variable $\tilde{Y} = (Y - \hat{\mu})/\hat{\sigma}$. Following Parzen (1992) dynamic statistics on $0 < t < 1$ called sample change density (or sample Brownian derivative) is defined as

$$\tilde{c}(t) = \tilde{Y}_j, \ (j-1)/n \le t < j/n \quad \text{for } j = 1, \ldots, n. \tag{2.3}$$

Sample change process (or sample Brownian Bridge) is defined

$$\tilde{C}(t) = \int_0^t \tilde{c}(s) ds. \tag{2.4}$$

Sample change test process is defined as

$$\widetilde{CT}(t) = \tilde{C}(t)/\sqrt{t(1-t)}, \quad 0 < t < 1. \tag{2.5}$$

The distribution of $\sqrt{n}\tilde{C}(t)$ (or $\sqrt{n}B(t)$) is asymptotically a Brownian Bridge process with mean zero and covariance kernel $\min(s, t) - st$.

Graphs of the dynamic statistics $\tilde{c}$, $\tilde{C}$, and $\widetilde{CT}$ provide graphical diagnostics to test whether there is change in the process. But if there is cyclic change or more than one change-point, the cusum type sample change process with the data does not work well. In the following section this problem will be dealt with the sample Fourier coefficients. We propose to use the sample Fourier coefficients in the change analysis for any knid of change, that is, the change is abrupt, cyclic or smooth.

3. The Proposed Procedures

Suppose one observes data $Y_1, \ldots, Y_n$ obeying the model

$$Y_i = \mu_j + \varepsilon_i, \quad t_{j-1} \le i < t_j, \quad j = 1, \ldots, m+1 \tag{3.1}$$

where $1 = t_0 < t_1 < t_2 < \cdots < t_m < t_{m+1} = n$ are unknown change-points, ε_i's are white noises with mean zero and finite variance σ^2. This model is a general expression of the change-point model. The above model represents that the regression function is a stepwise function and $\mu_j=\mu_j(x_i)$ which depends on the design points. Therefore if the regression function is continuous with smooth change, it can be expressed as

$$Y=f(x)+\varepsilon, \tag{3.2}$$

where $f(x)$ is nonconstant.

Using sample Fourier coefficients, the null hypothesis of no change against the alternative hypothesis of at least one level change in the process can be tested. When the alternative hypothesis is true, a sample Fourier series estimator applied to data will be sensitive to level changes. Tests with the sample Fourier series estimator might be more sensitive in detecting cyclic change and change with multiple change-points.

Define the sample Fourier coefficients with the cosine bases as $\widehat{\phi}_0=\bar{y}$ and

$$\hat{\phi}_j=\frac{1}{n}\sum_{i=1}^{n} Y_i\sqrt{2}\cos(\pi j x_i), \quad x_i=\frac{i-0.5}{n}, \quad j=1,\ldots,n-1. \tag{3.3}$$

Mathematical properties of Fourier coefficients are discussed in Tolstov (1976). We apply change analysis (Parzen (1992)) with the sample Fourier coefficients as follows. Define normalized sample Fourier coefficient as $\tilde{a}=\sqrt{n}\hat{\phi}/\hat{\sigma}$. Sample change density, sample change process and sample change test process with the sample Fourier coefficients can be defined analogously to the ones in Section 2 as

$$\widetilde{ca}(t)=\tilde{a}_j,\quad (j-1)/n\le t<j/n \quad \text{for } j=1,\ldots,n,$$

$$\widetilde{CA}(t)=\int_0^t \widetilde{ca}(s)ds, \tag{3.4}$$

$$\widetilde{CTA}(t)=\widetilde{CA}(t)/\sqrt{t(1-t)},\quad 0<t<1.$$

Theorem 3.1 The asymptotic distribution of $\sqrt{n}\,\hat{\phi}_j$ is $N(0,\sigma^2)$ under H_0. And the asymptotic distribution of $n\ \hat{\phi}_j^{\,2}/\sigma^2$ is Chi-square with the degree of freedom 1.

Remark. $\sqrt{n}\,\hat{\phi}_j$'s are asymptotically mutually independent.

Theorem 3.2 The asymptotic distribution of $\sqrt{n}\widetilde{CA}(t)$ is a Brownian Bridge process with mean zero and covariance kernel $\min(s,t)-st$.

Let us define equally spaced design points as $x_i=(i-0.5)/n$. For large n the Fourier series estimator estimates a piecewise constant function. The underlying function can be estimated with sample Fourier coefficients with $\lambda<n$,

$$\hat{f}_\lambda(x)=\hat{\phi}_0+\sum_{j=1}^{\lambda}\hat{\phi}_j\sqrt{2}\cos(\pi jx). \tag{3.5}$$

We can consider the integrated squared error,

$$ISE(\lambda) = \int_0^1 (\hat{f}_\lambda(x) - f(x))^2 = (\hat{\phi}_0 - \phi_0)^2 + \sum_{j=1}^{k} (\hat{\phi}_j - \phi_j)^2 + \sum_{j=1}^{\infty} \phi_j^2 - \sum_{j=1}^{k} \phi_j^2.$$

We can choose λ which minimizes the integrated squared error. Minimizing $ISE(k)$ is equivalent to minimizing

$$r(\lambda) = \sum_{j=1}^{\lambda} \hat{\phi}_j^{\,2} - 2 \sum_{j=1}^{\lambda} \hat{\phi}_j \phi_j.$$

The expected value of $r(\lambda)$ can be written as

$$\begin{aligned} E(r(k)) &= E\left(\sum_{j=1}^{\lambda} \hat{\phi}_j^{\,2} \right) - 2 \sum_{j=1}^{\lambda} E\, \hat{\phi}_j \phi_j \\ &\approx \sum_{j=1}^{\lambda} E(\hat{\phi}_j^{\,2}) - 2 \sum_{j=1}^{\lambda} \phi_j^2 \quad \text{as } n \to \infty \\ &\approx \sum_{j=1}^{\lambda} E(\hat{\phi}_j^{\,2}) - 2 \sum_{j=1}^{\lambda} \left(E(\hat{\phi}_j^{\,2}) - Var(\hat{\phi}_j) \right) \\ &= - \sum_{j=1}^{\lambda} \left(E(\hat{\phi}_j^{\,2}) - 2 Var(\hat{\phi}_j) \right), \end{aligned}$$

using the fact that $E(\hat{\phi}_j) \to \phi_j$ as $n \to \infty$ for piecewise continuous functions f. Define the risk as a function of λ

$$R(\lambda) = E \int_0^1 (\hat{f}_\lambda(x) - f(x))^2 dx, \quad \lambda = 0, 1, 2, \dots, n-1. \tag{3.6}$$

An approximately unbiased estimator of $C - R(\lambda)$ is

$$\hat{M}(\lambda) = \sum_{j=1}^{\lambda} \hat{\phi}_j^{\,2} - \frac{2 \widehat{\sigma^2} \lambda}{n},$$

where C is a constant not depending on λ and $\widehat{\sigma^2}$ is any consistent estimator of σ^2. Let $\hat{\lambda}$ be the maximizer of $\hat{M}(\lambda)$. We can consider the test that rejects the null hypothesis if $\hat{\lambda} \geq \widehat{\lambda_\alpha}$ with level α. The level is controlled by the distribution of $\hat{\lambda}$. $\hat{\lambda}_\alpha$ is defined to be the maximizer of

$$m(\lambda, C_\alpha) = \sum_{j=1}^{\lambda} \widehat{\phi_j^2} - \frac{C_\alpha \widehat{\sigma^2} \lambda}{n}, \tag{3.7}$$

where, $C_{0.05} = 4.18$, and $C_{0.10} = 3.22$ (from Eubank and Hart (1992)).

If the null hypothesis is true, then it is expected that $\hat{\phi}_j = 0,\ j = 1, \dots, n-1$. The graphs of $m(\lambda, C_\alpha)$, $\lambda = 0, 1, \dots, n-1$ can be used for checking if there is change in the process. The maximum of $m(\lambda, C_\alpha)$ should occur at $\lambda \geq 1$ if the underlying function is nonconstant.

We consider the sample quantile function of the sample Fourier coefficients $Q^n(u)$, $0<u<1$. $Q^n(u)$ is the inverse of the sample distribution function $F^{(n)}(y)$, $-\infty<y<\infty$. To compute the sample quantile function u_j, v_j for $j=1,\dots,n$, should be determined. The cumulative relative frequency is denoted by

$$u_j=F^{(n)}(v_j)=\text{fraction of sample Fourier coefficients} \le v_j. \tag{3.8}$$

Note $u_n=1$, and let $u_0=0$. If all values in the sample Fourier coefficients are distinct, the distinct values are the order statistics $\hat{\phi}_{(1)}<\cdots<\hat{\phi}_{(n)}$. The sample quantile function $Q^{(n)}$ can be determined as

$$Q^{(n)}(u)=v_j, \quad u_{j-1}<u\le u_j, \tag{3.9}$$

or equivalently

$$Q^{(n)}(u_j)=v_j, \quad j=1,\dots,n.$$

The sample distribution function $F^{(n)}$ of data is discrete. To estimate a continuous distribution function, define initial value v_0^m and mid-values by

$$v_j^m=(v_j+v_{j+1})/2, \quad j=1,\dots,n-1, \quad v_0^m=v_1, \quad v_n^m=v_n, \tag{3.10}$$

$$u_j=(j-0.5)/n, \quad j=1,\dots,n-1,$$

$$Q^{(n)}(u_j)=v_j^m, \tag{3.11}$$

$$w_j=F(Q^{(n)}(u_j))=F(v_j^m),$$

where $F(\cdot)$ is the cumulative distribution function of $N(0,1)$ distribution. Since the asymptotic distribution of the normalized sample Fourier coefficients is standard normal. We suggest to go over the change PP plot graphs of (u_j, $\sqrt{n}(w_j-u_j)$) which are expected to behave as a Brownian Bridge under H_0 (see Ross (1983)).

If the normalized data are used for the change PP plot, the distribution from which the data are obtained should be known. In practice, it is hard to know the true distribution. However when the sample Fourier coefficients are used for the change PP plot, we can use $N(0,1)$ distribution function if the sample size is large enough. We also recommend to check the normal quantile plots since the distribution of sample Fourier coefficients is asymptotically identically normal under H_0.

Remark. w_j's show the value for the distribution and u_j's represent the empirical distribution. Under H_0, $\sqrt{n}(w_j-u_j)$ is a random walk. If the alternative hypothesis is true, there might be a trend in the change PP plot.

4. Simulation Study

A simulation is conducted to investigate the behavior of the sample Fourier coefficients in the change analysis. The critical values for the sample change test can be obtained from basic calculations. But the main interest in this article is to detect change from the graphical point of view. Models considered are as follows

(i) No change model

$$f(x)=0, \quad 0\le x\le 1,$$

(ii) One change-point model $(\beta=1.0)$

$$f(x)=\begin{cases}0, & 0\le x<0.5,\\ \beta, & 0.5\le x\le 1,\end{cases}$$

(iii) Two change-points model $(\beta=1.0)$

$$f(x)=\begin{cases}0, & 0\le x<0.3,\\ \beta, & 0.3\le x<0.7,\\ 0, & 0.7\le x\le 1,\end{cases}$$

(iv) Cyclic change model

$$f(x)=\cos(\pi jx), \quad 0\le x\le 1, \quad j=3.$$

Data are generated from

$$y_i=f(x_i)+\varepsilon_i, \quad x_i=\frac{i-0.5}{n}, \quad n=50,$$

where ε_i's are iid from $N(0,1)$ with a fixed sample size 50. We used a nonparametric variance estimator

$$\sigma^2=\frac{1}{2(n-1)}\sum_{i=1}^{n-1}(y_{i+1}-y_i)^2,$$

which is less sensitive whether the null hypothesis is true or not.

Figures 1, 2, 3 and 4 show the comparison of change analysis with the sample Fourier coefficients for given models. If the null hypothesis is true, plots of the change process with the data and the change process with the sample Fourier coefficients show no trends. However when there is change in the process of means, we see some trends in plots of change process in Figures 2,3, and 4. For plots of m, $\alpha=0.10$ was used. Maximums of m occur at 0,1,2, and 3 respectively in Figures 1,2,3, and 4. The change PP plot in Figure 1 is most likely to behave as a Brownian Bridge. Change PP plots in Figures 2,3, and 4 are not shown as Brownian Bridges. The normal quantile plots show outliers in Figures 2,3 and 4 with the change model.

5. Discussions

We used the sample Fourier coefficients to detect change of the process. There are some advantages when we use the sample Fourier coefficients in change analysis: (i) For any kind of change the sample Fourier coefficients can be used. (ii) We can estimate the underlying change function, if exist, after the order of the Fourier series is obtained. (iii) Since each sample Fourier coefficient is asymptotically normally distributed, we can use the standard normal distribution function in change PP plots. Also we may use various orthogonal bases to get the sample Fourier coefficients.

Graphs of change process, change test, and change PP plot are presented to identify changes graphically. In practice graphical methods are easy to understand and appeal to non-statisticians. Further development of theoretical and graphical methods are anticipated in the change analysis.

References

[1] Eubank, R. and Hart, J. (1992). Testing Goodness-of-fit in Regression via Order Selection Criteria, *Annals of Statistics*, 20, 1412-1425.

[2] Kander, Z. and Zacks, S. (1966). Test Procedures for Possible Changes in Parameters of Statistical Distributions Occurring at Unknown Time Points, *Journal of American Statistical Association*, 1196-1210.

[3] Parzen, E. (1979). Nonparametric Statistical Data Modeling, *Journal of American Statistical Association*, 74, 105-122.

[4] Parzen, E. (1992). Comparison Change Analysis, *Nonparametric Statistics and Related Topics*.

[5] Ross, S. M. (1983). *Stochastic Processes*, John Wiley & Sons, New York.

[6] Tolstov, G. P. (1976). *Fourier Series*, Dover, New York.

[7] Zacks, S. (1983). Survey of Classical and Bayesian Approaches to the Change-Point Problem: Fixed Sample and Sequential Procedures of Testing and Estimation, *Recent Advances in Statistics*, 245-269.

Figure 1. Change Analysis with No Change Model

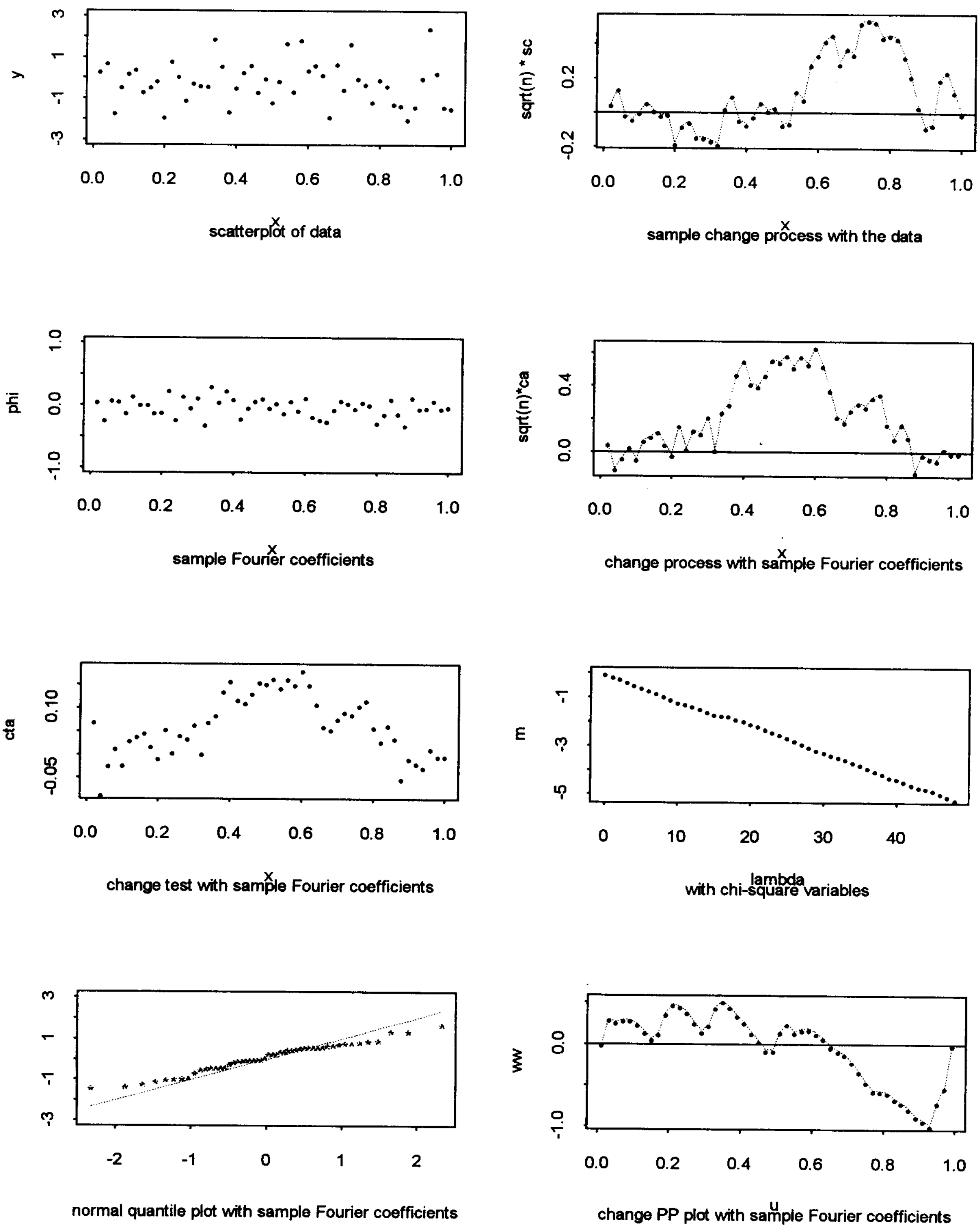

Figure 2. Change Analysis with One Change-point Model

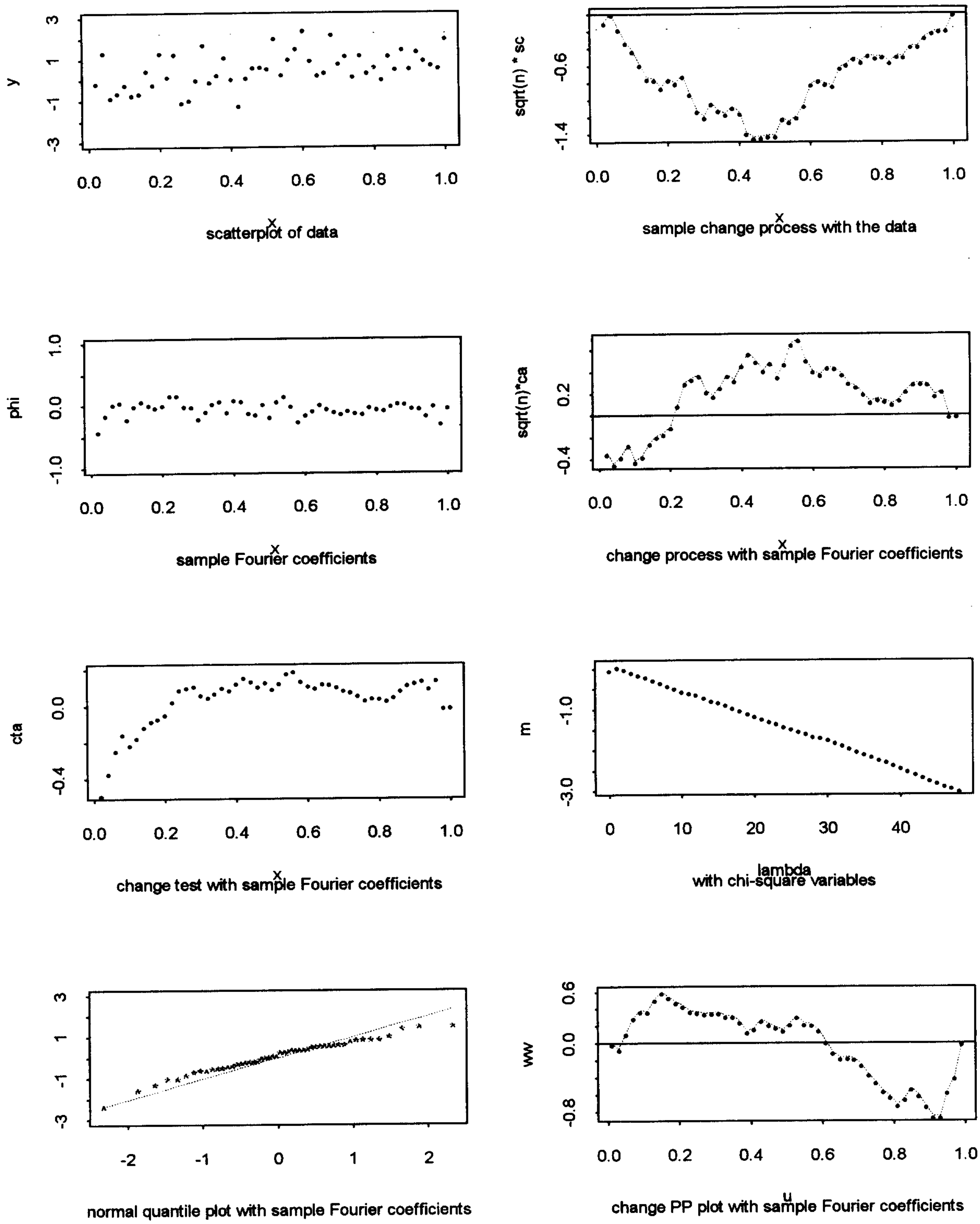

Figure 3. Change Analysis with Two Change-points Model

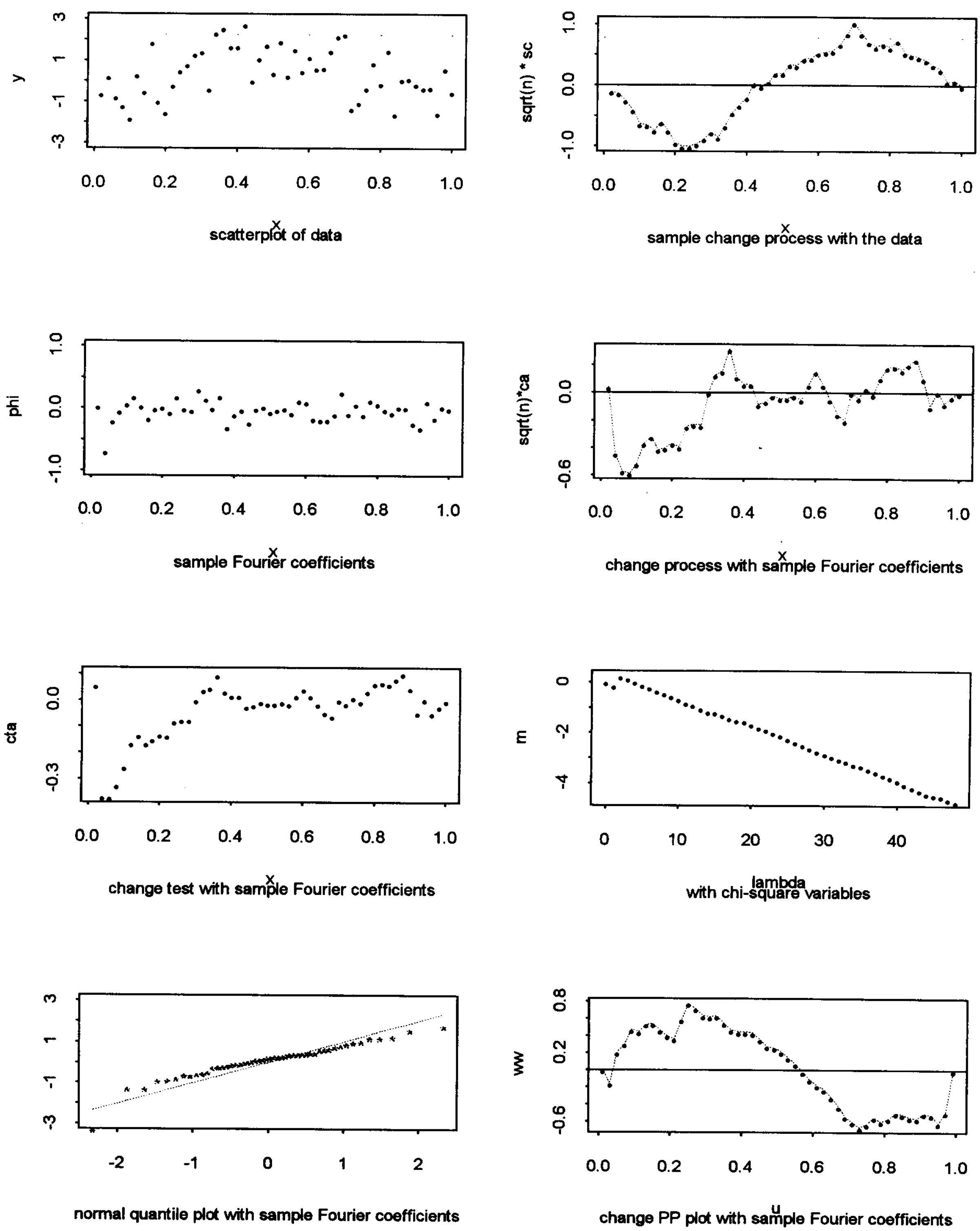

Figure 4. Change Analysis with Cyclic Change Model