SUMSRM: A New Statistic for the Structural Break Detection in Time Series

Kwok Pan PANG and Kai Ming Ting

Gippsland School of Computing and Information Technology
Monash University, Victoria 3842, Australia
Email : {ben.pang, kaiming.ting}@infotech.monash.edu.au,

Abstract. Structural break is one of the important concerns in non-stationary time series prediction. The cumulative sum of square (CUSUMS) statistic proposed by Brown et al (1975) has been developed as a general method for detecting a structural break. To better understand CUSUMS, this paper analyses the relationship among the bias of the break location estimation, pre-break data size and the decay rate of square residual. Our analysis reveals that small pre-break data size or low decay rate will greatly increase the bias of the break location estimation when there is a change of the mean. Based on the analysis, the paper proposes a new statistic SUMSRM to improve the performance of structural break detection and to reduce the bias of break location estimation. Our empirical evidence confirms that our intended design of the new statistic performs better than the CUSUMS statistic when there is a change of mean in the time series.

1. Introduction

In forecasting time series, ignoring structural breaks which often occur in the time series significantly reduces the accuracy of the forecast (Pesaran and Timmermann, 2003). Since the classical Chow test (1960) was developed, the past decade has seen considerable empirical and theoretical research on structural break detection in time series. Cumulative Sum of Recursive Residual (CUSUM) and Cumulative Sums of Square (CUSUMS) statistics (Brown et al. 1975) have been developed as general methods for single structural break detection. Kramer and Schotman (1992) proposed a modified statistic from CUSUM; the structural change is detected based on the range of the CUSUM rather than the maximum point of the absolute value of CUSUM. Chu et al (1995) proposed a test for the structural change based on the moving sums (MOSUMS) of the recursive residual. Inclan and Tiao (1994) used the CUSUMS for multiple structural breaks detection. Pesaran and Timmermann (2002) proposed Reverse CUSUM for detecting the most recent break. They showed that the accuracy of the forecast can be improved only if the data after the most recent break is selected as the training set, instead of using all available data for training in the time series which contains structural breaks. Pang and Ting (2003) further extended the idea of Reverse CUSUM and proposed a data selection method for time series prediction: all segments that have the same structure as the most recent segment will be grouped together to form an accumulated segment to be used as the new training set. Further, Pang and Ting (2004) provided an analysis about the centered version of CUSUMS. Their analysis reveals that the structural break detection performance can be improved if we can increase the pre-break data size or decrease the post-break data size, resulting a modified Centered CUSUMS that overcome the existing weakness.

This paper first analyses Centered CUSUMS by examining the relationship among the pre-break data size, decay rate of square residual and the bias of the structural break estimation. The analysis shows that small pre-break data size or low decay rate will increase the bias of the break location estimation when there is the structural change of mean or trend. The decay rate of square residual is defined as the change rate of the Centered CUSUMS, decay rate = (square residual at time b - square residual at time a)/(b-a), where $b>a$. Then, it proposes a new statistic SUMSRM by using square deviation about the median and the sliding window prediction residual. Our analysis shows that the square deviation about the median has higher sensitivity to the structural change compared with the square deviation about the mean used in Centered CUSUMS. The analysis also finds that the sliding window prediction residual can provide a higher decay rate than recursive residual. The empirical evidence shows that the proposed statistic can effectively improve the break detection, and eliminate the bias of the break location estimation, especially when there is a mean change. In the paper, we evaluate the performance of the proposed statistic when there are structural changes with single or multiple breaks.

Section 3 briefly describes the background of the Centered CUSUMS. We present the bias analysis for the Centered CUSUMS in section 4, and the new statistic SUMSRM is proposed in section 5. The experiments and results are reported in section 6.

2. Structural change

The parameters of the predictive model are assumed to be consistent and constant over time. If these conditions cannot be met, it is said that the structural change has occurred in the time series.

Let us take a linear regression as an example. Suppose a time series can be explained by:

$y_t = x_t\beta_t + \sigma_t\varepsilon_t$, where $\varepsilon_t \sim N(0,1)$

The time series is regarded to have "no structural change" if the parameters β_t and σ_t are constant and consistent over time.

3. Background of CUSUMS

Let y_1, y_2,…y_n be the time series under consideration. We first convert the series into input and output pairs to be used for ordinary linear regression.

The basic linear regression model we used is having the output y_t with k input variables:

$$y_t = \lambda_0 + \lambda_1 y_{t-1} + \lambda_2 y_{t-2} + ... + \lambda_k y_{t-k} .$$

We use the following notation to denote the observation matrices $Y_{m,n}$ and $X_{m,n}$ which consist of n observations in the time series.

$$\boldsymbol{Y}_{\mathrm{m,n}} = \begin{bmatrix} y_m \\ y_{m+1} \\ \cdots \\ \cdots \\ y_n \end{bmatrix}, \quad \beta_n = \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ .. \\ .. \\ \lambda_k \end{bmatrix} \quad \text{and}$$

$$\boldsymbol{X}_{\mathrm{m,n}} = \begin{bmatrix} x_m \\ x_{m+1} \\ . \\ . \\ x_n \end{bmatrix} = \begin{bmatrix} 1 & y_{m-1} & y_{m-2} & . & . & y_{m-k-2} & y_{m-k} \\ 1 & y_{m-2} & y_{m-3} & . & . & . & y_{m-k-1} \\ .. & \cdots & \cdots & . & . & . & \cdots \\ .. & \cdots & \cdots & . & . & . & \cdots \\ 1 & y_{n-1} & y_{n-2} & . & . & . & y_{n-k} \end{bmatrix}$$

where $m < n$, and the $x_m, x_{m+1,...,}x_n$ are the row vectors.

Using the observations as the training data, the least square coefficients β_n can be estimated by

$$\hat{\beta}_n = (X_{1,n}^{/} X_{1,n})^{-1} X_{1,n}^{/} Y_{1,n} .$$

The CUSUM and CUSUMS statistics (Brown et al 1975) are defined as follows. The CUSUM statistic is based on the standardized recursive residual w_r:

$$w_r = (y_r - x_r\hat{\beta}_{r-1}) / d_r , \quad r = 2k+1,\ldots,n\text{-}1,n \qquad (1)$$

where

$$\hat{\beta}_r = (X_{1,r}^{/} X_{1,r})^{-1} X_{1,r}^{/} Y_{1,r} \qquad (2)$$

$$d_r = 1 + x_r (X_{1,r}^{/} X_{1,r})^{-1} x_r^{/}$$

The CUSUMS is defined in terms of w_r:

$$s_r = \frac{\sum_{i=1}^{r} w_i^2}{\sum_{i=1}^{n} w_i^2} \quad ,r = 2k+1,\ldots,n\text{-}1,n \qquad (3)$$

The Centered CUSUMS is defined as:

$$s_{r,n}^{*} = \frac{\sum_{i=1}^{r} w_i^2}{\sum_{i=1}^{n} w_i^2} - \frac{r}{n} \quad , \ r = 2k+1,\ldots,n\text{-}1,n \qquad (4)$$

Note that $s_{r,n}^{*}$ has zero mean.

The test statistic for structural break detection is:

$$T = \sqrt{\frac{n}{2}} \max_r \left| s_{r,n}^{*} \right|$$

The estimated break location, if T is above a critical value (for a specific confidence level), is defined as:

$$\hat{r} = \arg\max_r \left| s_{r,n}^{*} \right|$$

Under variance homogeneity, Inclan and Tiao (1994) show that $s_{r,n}^{*}$ behaves like a Brownian Bridge asymptotically.

Pang and Ting (2004) show that small post-break data size or large pre-break data size can improve the structural break detection performance, and they propose the modified Centered CUSUMS statistic as follows:

$$MT_{n_1} = \max_{0 \le r \le n_1} \sqrt{\frac{n_i}{2}} \left| \frac{\sum_{i=1}^{r} w_i^{\,2}}{\sum_{i=1}^{n_1} w_i^{\,2}} - \frac{r}{n_1} \right| \quad , \text{ where } r \le n_1 \le n$$

The estimated break location will be obtained by:

$$\hat{r} = \arg\max_{0 \le r \le n_1} \sqrt{\frac{n_i}{2}} \left| \frac{\sum_{i=1}^{r} w_i^{\,2}}{\sum_{i=1}^{n_1} w_i^{\,2}} - \frac{r}{n_1} \right| \quad , \text{ where } r \le n_1 \le n$$

In the modified Centered CUSUMS statistic, we will use an adjustable data size n_1 instead of data size n. n_1 is always within the range $r \le n_1 \le n$, and it is not fixed. MT_{n_1} will be optimized by selecting an appropriate n_1. As S_{r,n_1}^{*} behaves like the Brownian Bridge asymptotically, we can use the same critical value for the specified n_1 that are tabulated in Inclan and Tiao (1994).

4. Bias analysis for the Centered CUSUMS

In this section, we concentrate on analyzing the bias of the break location estimation when changing the parameter β or σ. We will split the analysis into two parts: (a) change of β_t and (b) change of σ_t, after we show the effect of a step change of the Centered CUSUMS below.

The Centered CUSUMS in (4) can be rewritten as

$$C(r) = \frac{1}{nV}\left(\sum_{i=1}^{r} w_i^2 - rV\right)$$

where $V = (\sum_{i=1}^{n} w_i^2)/n$ can be interpreted as the average of the square residual.

The step change of the Centered CUSUMS can be written as:

$$C(r) - C(r-1) = \frac{1}{nV}\left(w_r^2 - V\right) \qquad (5)$$

It shows that the change of the Centered CUSUMS can be explained using the difference between the value of the square residual and the average square residual value. If $w_r^2 > V$, the Centered CUSUMS will increase at time r. However, if $w_r^2 < V$, the Centered CUSUMS will decrease at time *r*.

(a) The bias of the break location estimation when there is a change of β_t

We suppose the time series is composed of two segments. We assume the variance σ_t^2 is constant over the time, and only the mean or the trend β_t changes.

$$\beta_t = \begin{cases} \beta_A & t = 1,2,\ldots,r \\ \beta_B & t = r+1, r+2, \ldots, n \end{cases}$$

where $\beta_A \neq \beta_B$

Based on the above structural change, we generate the square residual plot (the square of the recursive residual vs. time) in figure 1. It produces the peak at t = *r*+1, and a downward slope after the peak. Then we take the average of the square of the recursive residual (i.e. the horizontal dotted line in figure 1).

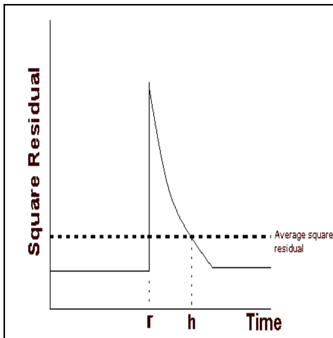

Figure 1: square residual vs. time (when the mean or the trend change)

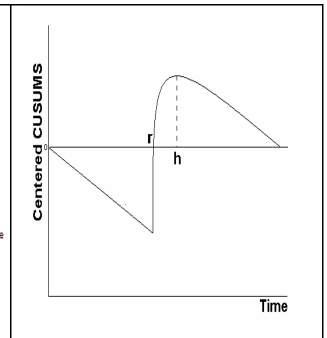

Figure 2: the Centered CUSUMS (when the mean or trend changes)

Based on the equations (4) and (5), we convert the square residual to the Centered CUSUMS, and produce the plot of the Centered CUSUMS in figure 2. The value of the Centered CUSUMS starts to decrease until it reaches the minimum point at t = *r*. Then the value of Centered CUSUMS starts increasing until t = *h*, and the value decreases after t = *h*.

After taking the absolute value to the Centered CUSUMS, the estimated break location can be determined by :

$$\text{Break point} = \begin{cases} r \text{ when } |C(r)| \geq |C(h)| \\ h \text{ when } |C(r)| < |C(h)| \end{cases}$$

Let us consider the first condition $|C(r)| \geq |C(h)|$

$$|C(r)| \geq |C(h)| \Rightarrow \left|\frac{1}{nV}\left(\sum_{i=1}^{r} w_i^2 - rV\right)\right| \geq \left|\frac{1}{nV}\left(\sum_{i=1}^{h} w_i^2 - hV\right)\right|$$

$$\Rightarrow \left|\sum_{i=1}^{r} w_i^2 - rV\right| \geq \left|\sum_{i=1}^{h} w_i^2 - hV\right|$$

As $C(h) > 0$ and $C(r)<0$, the above formula can be rewritten as :

$$|C(r)| \geq |C(h)| \Rightarrow \left(rV - \sum_{i=1}^{r} w_i^2\right) \geq \left(\sum_{i=1}^{h} w_i^2 - hV\right)$$

$$\Rightarrow rV - \sum_{i=1}^{r} w_i^2 \geq \sum_{i=1}^{r} w_i^2 + \sum_{i=r+1}^{h} w_i^2 - hV$$

$$\Rightarrow 2rV - 2\sum_{i=1}^{r} w_i^2 \geq \sum_{i=r+1}^{h} w_i^2 - (h-r)V$$

$$\Rightarrow 2rV - 2\sum_{i=1}^{r} w_i^2 \geq nV\left(C(h) - C(r)\right)$$

Using the similar approach for the second condition, we can write it as:

$$|C(r)| < |C(h)| \Rightarrow 2rV - 2\sum_{i=1}^{r} w_i^2 < nV\left(C(h) - C(r)\right)$$

Based on the above analysis, a bias estimate of the break location, instead of r, will be introduced when $\left(2rV - 2\sum_{i=1}^{r} w_i^2\right) < nV\big(C(h) - C(r)\big)$

The bias of the break location estimation can be reduced by the following two factors:

(i) Pre-break data size
(ii) Decay rate of square residual after the break.

If the pre-break data size is large enough, then $2rV - 2\sum_{i=1}^{r} w_i^2 \geq nV\big(C(h) - C(r)\big)$, and the estimated break location is expected to be *r*. Unfortunately, it is not possible to increase the pre-break data size in many situations. Increasing the decay rate is the other way to eliminate the bias of the break location estimation by forming a steeper slope after the break, which makes the value *h* closer to the value *r*.

In order to increase the decay rate, we propose to use sliding window prediction residual instead of the recursive residual. More details will be discussed in section 5.2.

b) Break location estimation when there is a change of σ_t

We suppose the time series are composed of two segments. We assume the mean or trend β_t is constant over the time and only the variance σ_t^2 changes.

$$\sigma_t = \begin{cases} \sigma_A & t = 1,2,\ldots,r \\ \sigma_B & t = r+1, r+2, \ldots, n \end{cases}$$

Where: $\sigma_A \neq \sigma_B$

Based on the above structural change, let us consider the situation : $\sigma_A < \sigma_B$. Its square residual plot (the square of the recursive residual vs. time) is indicated in figure 3. The square residual plot consists of two horizontal lines. The lower horizontal line represents the square residual of the first segment, and the upper horizontal line represents the square residual of the second segment. The average of the square of the recursive residual is shown as the horizontal dotted line in figure 3.

Based on the equations (4) and (5), we convert the square residual to the Centered CUSUMS, and produce a plot of the Centered CUSUMS in figure 4. The value of the Centered CUSUMS starts to decrease until it reaches the minimum point at t = *r*. Then the value starts increasing until t = *n*. The estimated break location = *r* will be determined after taking the maximum absolute value of the Centered CUSUMS.

It is interesting to note that the square residual after the break point (t = *r*) keeps constant and all stay above the average of the square residual. The estimated break location is thus expected to be at t = *r*. We have provided the evidence to show that changing the pre-break data size or the decay rate won't make any alternation on the bias of the break location estimation.

Same approach can be applied into the situation $\sigma_A > \sigma_B$. The bias of break location estimation won't be affected by the change of the variance.

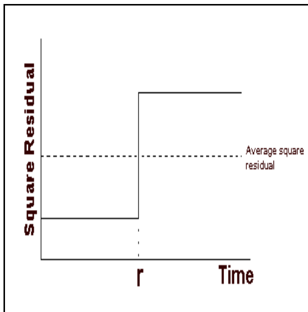

Figure 3: square residual vs. time (when the variance change)

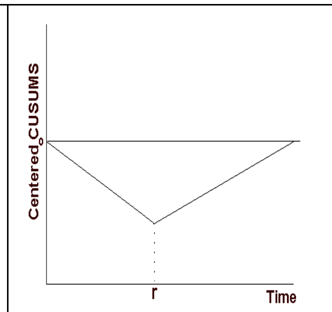

Figure 4: the Centered CUSUMS (when the variance changes)

5. A new statistic

Based on the above analysis, the paper proposes a new statistic using the square deviation about the median and sliding window prediction residual. The details of discussion are listed in the followings.

5.1 Square deviation about the median

Let us consider the mean of square deviation about the parameter θ

$$E[(x-\theta)^2] = E[(x - E(x) + E(x) - \theta)^2]$$
$$= E\{[x - E(x)]^2 + 2[x - E(x)][E(x) - \theta] + [E(x) - \theta]^2\}$$
$$= E[(x - E(x))^2] + (E(x) - \theta)^2$$
$$= Var(x) + (Bias)^2$$

Note that the Bias here refers to the bias the residual estimation, which is different from the bias of the break location estimation discussed in the previous sections.

If we let $\theta = mean$, square deviation about the mean will be exactly the same as the *var(x)*. If we let $\theta = Median$, the Median will be approximately equal to the *E*(*x*) when *x* is symmetrically distributed and when the data size is large enough. However, when the structure changes, the median starts to deviate from *E(x).* It is found that the structural change intensifies the bias, which leads to larger value of square deviation about the median. There is an implication that the square deviation about the median is more sensitive to the structural change (i.e. change of the mean or trend) comparing with the square deviation about the mean. For instance, we let x be the residual. The mean of residual is always assumed to be zero in the ordinary linear regression. The square deviation for the residual will be the same as the sum of square residual no matter the structural change occurs or not. However, if we use the square deviation about the median, we expect the difference between the value of square deviation about the median and square deviation about the mean will be small if no structural change occurs, and the difference will be enlarged if the structure changes

5.2 Sliding window prediction residual

We propose to use the sliding window prediction residual instead of the recursive residual because the sliding window prediction can increase the decay rate. When using the sliding window approach, only the windows that cover the break point will be affected by the structural change. Suppose the data size of the time series is n, the size of slide window is m and the break is located at c. If we use the recursive residual, all prediction residual after point c will be affected by the structural change. However, if we use the sliding window prediction residual instead, only the prediction residuals $\{w_i, c \le i \le c+m\}$ will be affected by the structural change. That means less data or residual will be affected by the structural change when using sliding window prediction residual. It also implies that higher decay rate (i.e. the slope after the peak become steeper in the square residual plot) is obtained.

The sliding window prediction residual can be described as figure 5. For each window, the size of the each window (p) is fixed. The prediction residuals are obtained by using sliding window. All data in the window will be used as training set to train the model, and then we use the trained model to make a one step ahead prediction. After obtaining the prediction residual, we slide the window one step ahead to make the next prediction residual. This approach is different from the Centered CUSUMS which adopts the recursive residual.

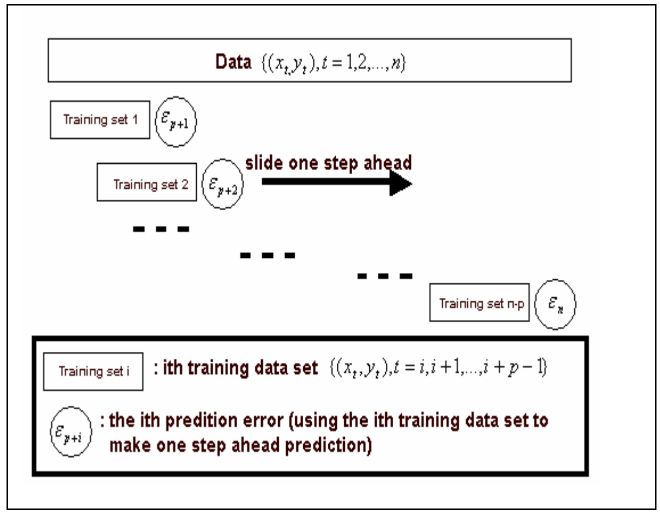

Figure 5 : Sliding window prediction residual

5.3. SUMSRM statistic

As mentioned in section 5.1, when the structure changes, the square deviation about the median increases as the bias of the expected residual estimation increases. Also, using the slide window prediction residual, instead of the recursive prediction residual, will increase the decay rate of square residual after the peak caused by the structural change. We propose a new statistic by combining these two ideas and the main idea from the modified Centered CUSUMS, searching for an appropriate post-break data size for structural break detection. Empirical evidence from Pang and Ting (2004) has shown that large post-break data size weakens the structural break detection performance, so it is important to select appropriate post break data size. We call the new proposed statistic "Sum of Square Sliding Residual about the Median" (SUMSRM).

Suppose the time series $\{(x_t, y_t), t = 1,2,\ldots,n\}$, where x_t is the row vector, $x_t = (y_{t-1}, y_{t-2},\ldots, y_{t-k})$ and p is the size of each sliding window.

The SUMSRM test statistic for the break detection is:

$$S_{n_1} = \max_{0 \le r \le n_1 - p} \left| D_{r,n_1} \right|, \quad r \le n_1 - p \le n - p$$

where

$$D_{r,n_1} = \sqrt{\frac{n_1 - p}{2}} \left(\frac{\sum_{i=1}^{r} \gamma_i^2}{\sum_{i=1}^{n_1-p} \gamma_i^2} - \frac{r}{n_1 - p} \right), \quad r \le n_1 - p \le n - p$$

$$\gamma_i = \frac{(\varepsilon_i - Med_i)}{d_i},$$

$$d_i = 1 + x_i (X'_{i-p,i-1} X_{i-p,i-1})^{-1} x'_i, \quad i = p+1, p+2, \ldots, n$$

ε_i is the prediction residual of the i[th] slide window:

$\varepsilon_i = y_i - x_i \hat{\beta}_i$, $i = p+1, p+2, \ldots, n$

Med_i is the median of the i[th] window training set residual:

$Med_i = Median(Y_{i-p,i-1} - X_{i-p,i-1}\hat{\beta}_i)$, $i = p+1, p+2, \ldots, n$

If S_{n_1} is above the critical value for the adjusted data size n_1, the estimated break location is defined as

$$\hat{k} = p + \underset{0 \le r \le n_1 - p}{\arg\max} \left| \sqrt{\frac{n_1 - p}{2}} \left(\frac{\sum_{i=1}^{r} \gamma_i^2}{\sum_{i=1}^{n_1 - p} \gamma_i^2} - \frac{r}{n_1 - p} \right) \right|, \; r \le n_1 - p \le n - p$$

The parameter β of each sliding windows can be obtained by:

$\hat{\beta}_i = (X'_{i-p,i-1} X_{i-p,i-1})^{-1} X'_{i-p,i-1} Y_{i-p,i-1}$, $i = p+1, p+2, \ldots, n$

We assume that the residual is normally distributed. Based on this assumption, D_{r,n_1} behaves like the Brownian Bridge Asymptote, and its proof is shown in the appendix A. According to the above statistic, the critical value at different data size and window size are estimated through the simulation. The critical value tables are shown in the appendix B. For the simulation, we use a Gaussian random walk with varied data size and varied window size to approximate the wiener process and replicate 10000 times with different seeds.

6. Experiment

We divide the experiments into two sections. The first section of our experiments is designed to verify the analysis conducted in section 4, showing the relationship between the pre-break data size and the bias of the break location estimation. The second section aims to evaluate the performance of the SUMSRM test statistic when single break or multiple breaks exist. In the experiments, we use the significant level α at 0.01 for the statistical test. The performance will be measured in terms of (a) accuracy, (b) mean of square deviation, (c) estimated bias and (d) mode of the estimated break location. We define the first three measures below.

(a) Accuracy is the percentage of correct classification. Every predicted break location falls into one of the following categories:

(i) **Correct Classification**: If the estimated break location is within the boundary (i.e. actual break location ± 10), we classify it to be correct classification.

(ii) **Incorrect Classification**: If the estimated break location is outside the boundary (i.e. actual break location ± 10), we classify it to be incorrect classification.

(b) Mean of Square Deviation (MSD) is defined as:

$$MSD = \frac{\sum_{i=1}^{v} (\hat{b}_i - Actb_i)^2}{v}$$

where

$\hat{b}$ is the estimated break location, $Actb$ is the actual break location which is the nearest break point to the estimated break location, and v is the number of simulation used in the experiment.

(c) The bias of the estimated break location (EB) is defined as:

$$\text{Estimated Bias (EB)} = \frac{\sum_{i=1}^{v} (\hat{b}_i - Actb_i)}{v}$$

As some simulated time series data with the structural change cannot be detected in some cases (i.e. the estimated break location will be defined as zero when the series with the structural change cannot be detected), these undetected observations may make EB generate the misleading measurement. Therefore, we mainly use the mode of the estimated location for measuring the bias of estimated break location, and EB is used for the reference.

For each experiment, we simulate the series 3000 times with different seeds, and we report the average performance over 3000 runs. In the experiments, we specify the number of input variables to be 3, and the size of each sliding window to be 40.

6.1 Validating the analysis result for the Centered CUSUMS

This experiment aims to validate the relationship between the pre-break data size and the bias of the break point estimation when the mean or variance changes. The procedure is described in the following.

Procedure

1. Two single break time series are designed to evaluate the break point estimation, as shown in the

table 1. The first time series is designed for the mean change, and the second one is for the variance change. Both series have the pre-break data size 300, and post break data size 300. We specify σ to be 0.2 for both segments in the first series, and in the second series, σ is specified to be 0.5 and 2.0 respectively for the first and second segment of the second series. We record their break detection performance and break location estimation.

2. We use the same time series as in (1), but the pre-break data size is reduced to 100 (the post-break data size remains the same). We record their break detection performance and break point estimation, and then compare the result with (1).

Table 1. Description of the two single-break series (when mean or variance changes): each series is composed of two segments. The whole sequence of the series is $\{y_t, t = 1,2,...,600\}$

Category of the structural change	**Segment 1** $\{y_t, t = 1,2,...,300\}$	**Segment2** $\{y_t, t = 301,302,...,600\}$
Mean change	$y_t = 20 + 0.6y_{t-1} + 0.3y_{t-2} + 0.1y_{t-3} + \sigma_1\varepsilon_t$, $\sigma_1 = 0.2$	$y_t = 30 + 0.6y_{t-1} + 0.3y_{t-2} + 0.1y_{t-3} + \sigma_1\varepsilon_t$, $\sigma_1 = 0.2$
Variance change	$y_t = 0.6y_{t-1} + 0.3y_{t-2} + 0.1y_{t-3} + \sigma_2\varepsilon_t$, $\sigma_2 = 0.5$	$y_t = 0.6y_{t-1} + 0.3y_{t-2} + 0.1y_{t-3} + \sigma_3\varepsilon_t$, $\sigma_3 = 2.0$

Experiment Results

The results are summarized as follows:

The effect of Pre-break data size to the estimation of the break location (when the mean changes).
As shown in figures 6a and 6b, the result shows the bias of the break location estimation is affected by the pre-break data size. When the pre-break data is 300, the mode of the estimated break location is 300 that is exactly the same as the actual break location. However, when we reduce the pre-break data size to 100, we find the mode of estimated break location is 122 and the estimated bias of break location is 21.7 where the actual break location is 100. The bias of estimate break location is increased to 22 after reducing the pre-break data size.

The effect of the pre-break data size to the estimation of the break location (when the variance changes)
As shown in figures 7a and 7b, the result shows that pre-break data size has a small impact on the bias of the break location estimation. When the pre-break data size is 300, the mode of the estimated break location is 300 that is exactly same as the actual break location. When we reduce the data size to 100, the mode of the estimated break location is 100 that is also the same as the actual break location.

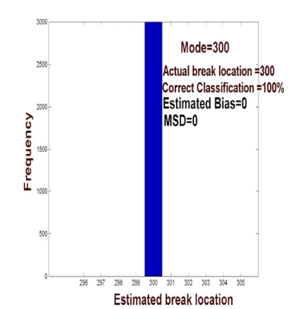

Figure 6a : The distribution of the estimated break location (change of mean with the pre-break data size =300)

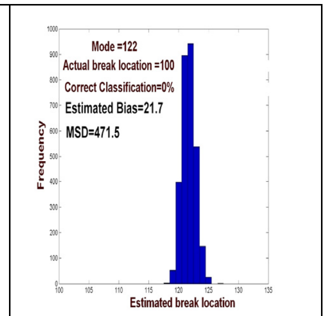

Figure 6b : The distribution of the estimated break location (change of the mean with the pre-break data size =100)

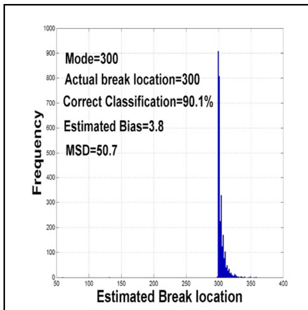

Figure 7a: the distribution of the estimated break location (change of variance with the pre-break data size =300)

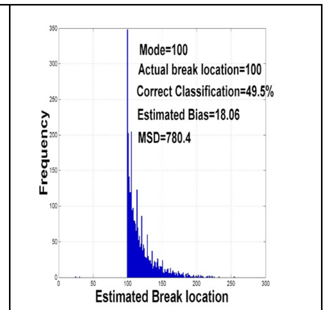

Figure 7b: the distribution of the estimated break location (change of variance with the pre-break data size =100)

6.2 Evaluating the effectiveness of the proposed statistic

The following experiment is designed to evaluate the performance of the proposed statistic and compare it with the traditional Centered CUSUMS and the modified Centered CUSUMS statistics when single break or multiple breaks exist.

6.2.1 Time series with a single break

The experiment is designed to evaluate the performance of the proposed SUMSRM statistic for the time series in which a single break exists. In the experiment, we will use the same time series as in the experiment in section 6.1 with the pre-break data size 100 and post break data size 300. We record the break detection performance and break point estimation using the traditional centered CUSUMS, modified centered CUSUMS and SUMSRM statistic.

Experiment Results

The results are summarized as follows:

- As shown in figure 8a, when the mean changes, the SUMRM significantly reduces the bias of the break location estimation. The result clearly shows that the SUMSRM statistic outperforms the traditional Centered CUSUMS and the modified Centered CUSUMS, and is significantly better in all four performance measures. The difference between the mode of estimated break location and actual break location is reduced from 22 to 4 after switching to SUMSRM from the traditional Centered CUSUMS or modified Centered CUSUMS.

- As shown in figure 8b, when variance changes, the SUMSRM statistic performs better than the traditional Centered CUSUMS, but is slightly worse than the modified Centered CUSUMS, in terms of the percentage of correct classification and MSD. Also, the modes of the estimated break location derived by those three selected statistics are almost identical, and they are the same or almost the same as the actual break location.

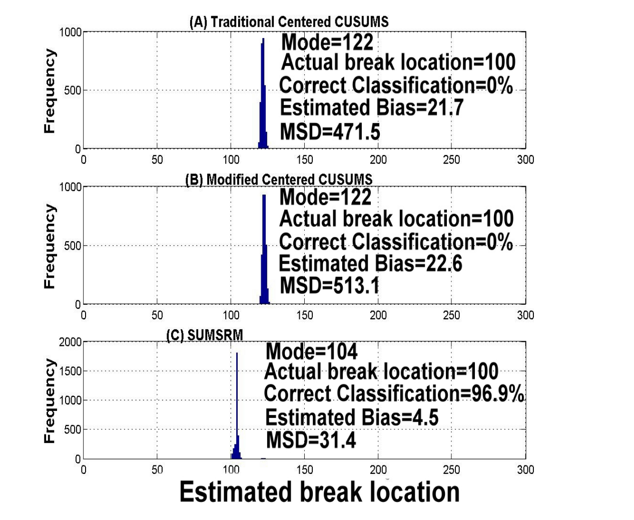

Fig 8a : Comparison of the break detection performance using traditional Centered CUSUMS, modified Centered CUSUMS and SUMSRM statistics (the change of mean with pre-break data size =100)

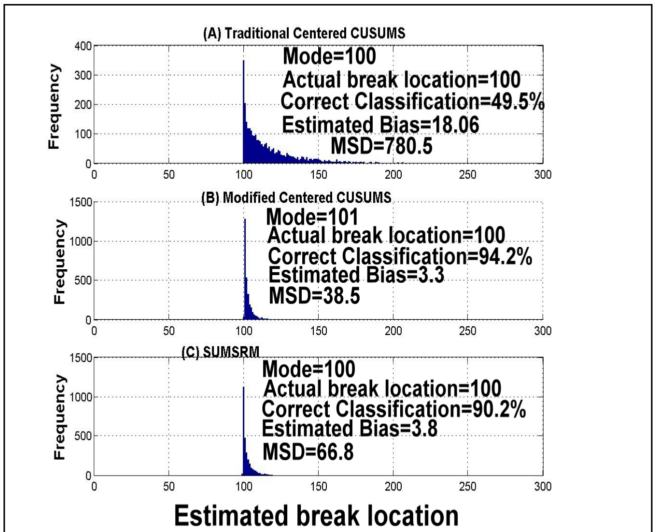

Fig 8b : Comparison of the break detection performance using traditional Centered CUSUMS, modified Centered CUSUMS and SUMSRM statistics (the change of variance with pre-break data size =100)

6.2.2 Time series with multiple breaks

This part of the experiments is to evaluate the performance of the SUMSRM statistic for the time series with multiple breaks when the mean or the variance changes. It also examines whether the SUMSRM can still detect any one of the structural changes and whether the estimated break location still falls into the acceptable range of break location. The time series is composed of 7 segments {seg_1, seg_2,…, seg_7}. The characteristics of the series with the change of mean and variance are described in tables 2 and 3 respectively.

Table 2: Descriptions of 7 segment series with the change of mean: the sequence of the series is {seg_1, seg_2, seg_3, seg_4, seg_5, seg_6, seg_7}. The length of each segment is specified to be 100, and the initial value of the series are specified as: $y_1 = y_2 = y_3 = 1$. The distribution of ε_t is $\varepsilon_t \sim N(0,1)$.

seg_1, seg_3, seg_5, seg_7	seg_2, seg_4, seg_6
$y_t = 20 + 0.6 y_{t-1} +$ $0.3 y_{t-2} + 0.1 y_{t-3} + \sigma_1 \varepsilon_t$, $\sigma_1 = 0.2$	$y_t = 30 + 0.6 y_{t-1} +$ $0.3 y_{t-2} + 0.1 y_{t-3} + \sigma_1 \varepsilon_t$, $\sigma_1 = 0.2$

Table 3: Description of 7 segment series with the variance change: the sequence of series is {seg_1, seg_2, seg_3, seg_4, seg_5, seg_6, seg_7}. The length of each segment is specified to be 100, and the initial value of the series are specified as: $y_1 = y_2 = y_3 = 1$. The distribution of ε_t is $\varepsilon_t \sim N(0,1)$

seg1, seg3, seg5, seg7	seg2, seg4, seg6
$y_t = 0.6y_{t-1} + 0.3y_{t-2} + 0.1y_{t-3} + \sigma_2\varepsilon_{1t}$, $\sigma_2 = 0.5$	$y_t = 0.6y_{t-1} + 0.3y_{t-2} + 0.1y_{t-3} + \sigma_3\varepsilon_t$, $\sigma_3 = 2.0$

Experiment Result

The results are summarized as follows:

The performance of the three statistics in the multiple-break time series when the mean changes.

As shown in table 4a, the SUMSRM statistic outperforms the traditional Centered CUSUMS and the modified Centered CUSUMS. The SUMSRM obtains 100% correct classification rate and has the smallest bias of the estimated break location in terms of the estimated bias and the mode among three statistics. A large bias of the estimated break location has been observed in the experiment using the traditional Centered CUSUMS.

The performance of the three statistics in the multiple-break time series when the variance changes.

As shown in table 4b, all three statistics perform comparably in this experiment. While Modified Centered CUSUMS performs the best in terms of the correct classification; Centered CUSUMS is the best in terms of estimated bias; SUMSRM is the best in terms of MSD. Nevertheless, the differences are small.

Table 4a: Comparison of the break detection performance among the Centered CUSUMS, the modified Centered CUSUMS and the SUMSRM statistic when the mean changes in the time series with the multiple breaks. The actual break location are : 100, 200, 300, 400, 500 and 600.

	Traditional Centered CUSUMS	Modified Centered CUSUMS	SUMSRM
mode	213	101	100
Correct Classification	13.5%	100%	100%
Estimated Bias	11.8	1	0
MSD	145.5	1	0

Table 4b: Comparison of the break detection performance among the Centered CUSUMS, the modified Centered CUSUMS and the SUMSRM statistic when the variance changes in the time series with the multiple breaks points. The actual break location are: 100, 200, 300, 400, 500 and 600.

	Traditional Centered CUSUMS	Modified Centered CUSUMS	SUMSRM
mode	600	100	100
Correct Classification	92.7%	95.2%	93.7%
Estimated Bias	-0.4	2	0.8
MSD	36	35	33

8. Conclusions

This paper makes the following contributions in the structural break detection and break location estimation.

- It provides a better understanding on how the pre-break data size and the decay rate of square residual affect the bias of the break location estimation in CUSUMS. Large pre-break data size and high decay rate of the square residual can effectively reduce the bias of the structural break location estimation.

- We identify a key weakness of the CUSUMS statstic: high bias of the break location estimation. We proposed a new statistic SUMSRM which has a low bias.

- Our experimental results show that the proposed SUMSRM statistic can effectively minimize the bias of the break location estimation and provide better structural break detection performance when there is a change of mean in the time series with single or multiple breaks. This result confirms our claim about the SUMSRM statistic - using the square deviation about the median and sliding window prediction residual can improve the break detection performance and eliminate the bias of the break location estimation.

- SUMSRM significantly outperforms the Centered CUSUMS and modified Centered CUSUMS, when there is a structural change of mean with a single break or multiple breaks in all four performance measures.

- Empirical experiments have demonstrated that when there is a structural change of variance in the time series with single or multiple breaks, the SUMSRM and modified Centered CUSUMS outperform the Centered CUSUMS. Both SUMSRM and modified Centered CUSUMS can improve the structural break detection by selecting appropriate post break data size.

- When there is a variance change in the time series with single or multiple breaks, SUMSRM performs a bit less satisfactory than the modified Centered CUSUMS. Sliding window approach cannot help SUMSRM to improve the structural break detection in this situation. As mentioned in sections 4b and 5.2, sliding window approach can help increase the decay rate. However, empirical evidence has shown that the decay rate has no relationship with the bias of the break location estimation when a variance change occurred in the time series. Besides, sliding window approach adopted by SUMSRM generates fewer pre-break prediction residuals and thus resulted in less satisfactory performance. This result can be supported and explained by our previous research analysis (Pang and Ting 2004) which clearly demonstrates that fewer pre-break data size (number of pre-break residual) will weaken the structural break detection performance, especially when the data set with small pre-break data size is used.

Appendix A: Proof of the proposed statistics SUMSRM having the Brownian asymptotes.

The notion below will be used in the following proof.

$\xrightarrow{p}$: Probability Convergence

$\xrightarrow{D}$: Distribution Convergence

W : Brownian Motion

W^o : Brownian Bridge

Let $\xi_t = (w_t - M_t)^2 - E[(w_t - M_t)^2]$, $t = p+1, p+2, \ldots, n$

Let $i = t - p$, then

$\xi_i = (w_i - M_i)^2 - E[(w_i - M_i)^2]$, $i = 1,2,3,\ldots,n-p$

$$E(\xi) = E[(w_i - M_i)^2 - E[(w_i - M_i)^2]$$
$$= E[(w_i - M_i)^2] - E[(w_i - M_i)^2]$$
$$= 0$$

$$Var(\xi_t) = Var\{(w_i - M_i)^2 - E[(w_i - M_i)^2]\}$$
$$= Var[(w_i - M_i)^2] - 0$$
$$= Var[(e_i - M_i)^2]$$

when data size for median is large enough, $\xi \sim \chi^2$ distribution.

Let $E[(e_i - M_i)^2] = \sigma_\psi{}^2$

$$Var(\xi_i) = Var[(e_i - M_i)^2] = \sigma = 2\sigma_\psi{}^2 = 2E[(e_i - M_i)^2]$$

where $E[(e_{i_t} - M_i)^2]$ is the mean of square deviation from the median; e_i is the residual, $e \sim N(0, \sigma^2)$

Let $X_{n-p}(t) =$

$$\frac{1}{\sigma\sqrt{(n-p)}} \sum_{i=0}^{[(n-p)t]} \xi_i + ((n-p)t - [(n-p)t]) \frac{1}{\sigma\sqrt{(n-p)}} \xi_{[(n-p)t]+1}$$

where t, $0 < t \le 1$

Based on the Donsker's Theorem (Billingsley 1968 Theorem 10.1) , $X_n \xrightarrow{d} W$,

and based on the theorem 5.1 of Billingsley (1968), so $\{X_n(t) - tX_n(1)\} \xrightarrow{D} W^o$

then

$$X_{n-p}(t) - tX_{n-p}(1) = \frac{1}{\sigma\sqrt{(n-p)}} \sum_{i=0}^{[(n-p)t]} \xi_i + ((n-p)t - [(n-p)t]) \frac{1}{\sigma\sqrt{n-p}} \xi_{[(n-p)t]+1} - \frac{t}{\sigma\sqrt{n-p}} \sum_{i=0}^{n-p} \xi_i$$

$$= \frac{1}{\sigma\sqrt{n-p}} \left(\sum_{i=0}^{[(n-p)t]} \xi_i - t \sum_{i=0}^{n-p} \xi_i \right) + ((n-p)t - [(n-p)t]) \frac{1}{\sigma\sqrt{n-p}} \xi_{[(n-p)t]+1}$$

$$= \frac{1}{\sigma\sqrt{n-p}} \left(\sum_{i=0}^{[(n-p)t]} \xi_i - \frac{(n-p)t}{n-p} \sum_{i=0}^{n-p} \xi_i \right) + ((n-p)t - [(n-p)t]) \frac{1}{\sigma\sqrt{n-p}} \xi_{[(n-p)t]+1}$$

Let the integer $r = (n-p)t$, then

$$X_{n-p}(t) - tX_{n-p}(1) = \frac{1}{\sigma\sqrt{n-p}} \left(\sum_{i=0}^{r} \xi_i - \frac{r}{n-p} \sum_{i=0}^{n-p} \xi_i \right) + ((n-p)t - [(n-p)t]) \frac{1}{\sigma\sqrt{n-p}} \xi_{n-p+1}$$

When $n \to \infty$,

$$\lim_{n\to\infty}((n-p)t-[(n-p)t])\frac{1}{\sigma\sqrt{n-p}}\,\xi_{n-p+1}\overset{p}{\to}0$$

and

$$X_{n-p}(t)-tX_{n-p}(1)\to W^0$$

According to the theorem (Billingsley theorem 4.1, p.25), it shows that:

If $X_n \overset{D}{\to} X$ and distance $\rho(X_n, Y_n)\overset{p}{\to}0$ then $Y_n \overset{D}{\to} X$

Therefore,

$$X_{n-p}(t)-tX_{n-p}(1)=\frac{1}{\sigma\sqrt{n-p}}\left(\sum_{i=0}^{r}\xi_i-\frac{r}{n-p}\sum_{i=0}^{n-p}\xi_i\right)$$

substitute $\xi_i=(e_i-Median_i)^2-E[(e_i-Median_i)^2]$ and $\sigma^2=2\sigma_\psi{}^2$ into above equation.

$$X_{n-p}(t)-tX_{n-p}(1)=$$

$$\frac{1}{\sigma_\psi\sqrt{2(n-p)}}\left(\sum_{i=0}^{r}\{(e_i-Median_i)^2-E[(e_i-Median_i)^2]\}-\right.$$

$$\left.\frac{r}{n-p}\sum_{i=0}^{n-p}\{(e_i-Median_i)^2-E[(e_i-Median_i)^2]\}\right)$$

$$=\frac{1}{\sigma_\psi\sqrt{2(n-p)}}\left(\sum_{i=1}^{r}(e_i-Median_i)^2-\frac{r}{n-p}\sum_{i=1}^{n-p}(e_i-Median_i)^2\right)$$

$$=\frac{\sum_{i=1}^{n-p}(e_i-Median_i)^2}{\sigma_\psi\sqrt{2(n-p)}}\left(\frac{\sum_{i=1}^{r}(e_i-Median_i)^2}{\sum_{i=1}^{n-p}(e_i-Median_i)^2}-\frac{r}{n-p}\frac{\sum_{i=1}^{n-p}(e_i-Median_i)^2}{\sum_{i=1}^{n-p}(e_i-Median_i)^2}\right)$$

$$=\frac{\sum_{i=1}^{n-p}(e_i-Median_i)^2}{\sigma_\psi\sqrt{2(n-p)}}\left(\frac{\sum_{i=1}^{r}(e_i-Median_i)^2}{\sum_{i=1}^{n-p}(e_i-Median_i)^2}-\frac{r}{n-p}\right)$$

When $n\to\infty$, $\frac{1}{n-p}\sum_{i=1}^{n-p}(e_i-Median_i)^2\to\sigma_\psi^2$

Therefore,

$$\sqrt{\frac{n-p}{2}}\left(\frac{\sum_{i=1}^{r}(e_i-Median_i)^2}{\sum_{i=1}^{n-p}(e_i-Median_i)^2}-\frac{r}{n-p}\right)\overset{D}{\to}W^o$$

And the distribution of $\sup_t\left|W_t^o\right|$ was given in equation (11.39) of Billingsley (1968)

$$P\{\sup_t\left|W_t^o\right|\le b\}=1+2\sum_{k=1}^{\infty}(-1)^r e^{-2r^2b^2}\quad,\quad b>0$$

Appendix B. Critical Value

singnificant level :0.1

data size / win_width	200	300	400	500	600	700	800	900	1000	2000	4000
20	1.21	1.23	1.24	1.25	1.25	1.26	1.26	1.26	1.26	1.28	1.27
30	1.18	1.21	1.22	1.23	1.23	1.24	1.24	1.24	1.25	1.26	1.26
40	1.17	1.19	1.21	1.21	1.21	1.23	1.22	1.23	1.23	1.24	1.25
50	1.15	1.18	1.19	1.19	1.21	1.21	1.22	1.22	1.22	1.24	1.24
60	1.15	1.18	1.19	1.19	1.2	1.21	1.21	1.22	1.22	1.23	1.24
70	1.15	1.17	1.17	1.19	1.2	1.2	1.21	1.22	1.21	1.23	1.24
80	1.15	1.17	1.17	1.19	1.2	1.19	1.21	1.21	1.21	1.23	1.23
90	1.14	1.17	1.17	1.18	1.19	1.19	1.21	1.21	1.21	1.23	1.23
100	1.14	1.16	1.17	1.18	1.19	1.19	1.2	1.21	1.21	1.22	1.22
200	———	1.15	1.17	1.18	1.18	1.18	1.18	1.2	1.2	1.21	1.21

singnificant level :0.05

data size / win_width	200	300	400	500	600	700	800	900	1000	2000	4000
20	1.35	1.37	1.38	1.39	1.38	1.4	1.4	1.41	1.41	1.43	1.42
30	1.32	1.35	1.36	1.36	1.37	1.38	1.37	1.39	1.38	1.4	1.39
40	1.31	1.33	1.35	1.35	1.35	1.37	1.37	1.37	1.37	1.38	1.39
50	1.28	1.31	1.32	1.33	1.34	1.34	1.35	1.36	1.36	1.37	1.38
60	1.28	1.31	1.31	1.33	1.33	1.34	1.34	1.36	1.36	1.37	1.37
70	1.29	1.31	1.31	1.32	1.33	1.33	1.34	1.36	1.35	1.37	1.37
80	1.27	1.31	1.3	1.32	1.33	1.33	1.34	1.36	1.35	1.36	1.37
90	1.27	1.31	1.3	1.32	1.32	1.33	1.34	1.35	1.34	1.36	1.36
100	1.27	1.29	1.3	1.31	1.32	1.32	1.34	1.35	1.34	1.36	1.36
200	--------	1.29	1.3	1.3	1.31	1.31	1.33	1.34	1.33	1.35	1.35

singnificant level :0.01

data size / win_width	200	300	400	500	600	700	800	900	1000	2000	4000
20	1.64	1.65	1.67	1.67	1.67	1.7	1.68	1.71	1.68	1.71	1.7
30	1.6	1.63	1.65	1.64	1.62	1.66	1.65	1.65	1.65	1.68	1.7
40	1.58	1.61	1.63	1.63	1.61	1.64	1.64	1.65	1.64	1.66	1.67
50	1.57	1.6	1.6	1.6	1.61	1.61	1.62	1.63	1.63	1.66	1.66
60	1.55	1.59	1.6	1.6	1.61	1.61	1.62	1.61	1.64	1.66	1.65
70	1.55	1.58	1.6	1.59	1.6	1.61	1.61	1.6	1.64	1.64	1.64
80	1.55	1.58	1.58	1.59	1.6	1.61	1.61	1.6	1.62	1.64	1.64
90	1.54	1.57	1.58	1.58	1.59	1.6	1.61	1.6	1.63	1.64	1.64
100	1.52	1.57	1.58	1.59	1.6	1.61	1.61	1.6	1.62	1.64	1.64
200	———	1.54	1.57	1.57	1.59	1.57	1.61	1.6	1.59	1.62	1.62

References

[1] Billingsley, P. (1968), Convergency of Probability Measures, New York : John Wiley.

[2] Brown, R.L., Durbin, J., and Evans, J.M. (1975) "Techniques for Testing the constancy of Regression Relationship over Time", Journal of Royal Statistical Society, Series B, 37, 149-192.

[3] Chow, G. (1960), Tests of equality between sets of coefficients in two linear regressions. Econometrica, Vol. 28, No. 3, 591-605

[4] Inclan, C. and Tiao, G. C.. (1994) “Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance”, Journal of the American Statistical Association, September 1994. Vol. 89, no. 427, 913-923.

[5] Krämer, Walter and Schotman, Peter (1992) “Range vs Maximum in OLS-based version of CUSUM test”, Economics letter 40, pp 379-381

[6] Pang, K. P. and Ting, K. M. (2003) “Improving Time series prediction by data selection” International Conference on Computational Intelligence for Modeling, Control & Automation, February 2003, 803-813

[7] Pang, K. P. and Ting, K. M. (2004) “Improving the Centered CUSUMS statistic for Structural Break Detection in Time Series”, Forthcoming of 17th Australian Joint conference on Artificial Intelligence.

[8] Pesaran, H. and Timmermann, A. (2002) “Market timing and Return Prediction under Model instability”, Journal of Empirical Finance, December 2002. Vol. 9, 495-510

[9] Chu, Chia Shang J., Hornik, Kurt and Kuan, Chung Ming (1995) “MOSUM Tests for Parameter constancy ”, Biometrika 82, 3 pp. 603-617

[10] Pesaran, H. and Timmermann, A., (2003) "How Costly is it to Ignore Breaks when Forecasting the Direction of a Time Series?," Cambridge Working Papers in Economics 0306, Department of Applied Economics, University of Cambridge.

