

Wavelet based Segmentation in Detecting Multiple Mean Changes in Time Series

Abstract

Aims/ objectives: Multiple mean break detection problem in time series is considered. A segmentation based on detecting turning points is applied to the original time series and its scaling coefficients series resulting from the maximal overlapped discrete wavelet transform (MODWT). Using a segmentation level along with a minimal distance parameter between two successive turning points we select a small number of segments within each series. A change point statistical test is then run separately within each series and over each segment. The simulation experiment shows that the multiple mean break detection procedure offers very good practical performance

Keywords: Discrete Wavelet Transform, Multiple Mean break, Segmentation, Turning Points, Time series

2010 Mathematics Subject Classification: 53C25; 83C05; 57N16

1 Introduction

Change point detection is a fundamental problem in the analysis of nonstationary time series. There is an extensive literature in identifying single break in the mean and other parameters such as the variance. Detection of several change-points are of interest in a wide range of applications such as stock prices and climatic data. In this paper we consider the problem of detecting simultaneously several mean breaks of a discrete time series by segmenting the series using the maximal overlapped discrete wavelet transform (MODWT) Antoniadis, Oppenheim (1995), Coifman and Donoho (1995), Donald and Andrew (2000) & Daubechies (1992), and applying a segmentation procedure to both the original series and to the smoothed resulting scaling coefficients from the transform. Our approach is nonparametric and is quite different from the other related work such as the Bayesian approach in Paul (2006) which suggest a recursive algorithm to identify change-points using segmentation

and assuming independence between segments, and Kim and Cheon (2013) which follows the same bayesian approach and make use of the advanced Monte Algorithm (SAMC). We can mention as well the wild binary segmentation (WBS) proposed by Piotr (2014) where the basic model is a deterministic piecewise-constant signal with multiple breaks plus a white noise. The method is efficient but make use of many technical assumptions. A good review and discussion of multiple break detection methods can be found in Killick et al. (2012) which an online repository of publications and software dedicated to break detections in data.

We assume that the break locations k_i^* and their numbers is unknown, but they are assumed to be relatively away from each other so that the sample size of any retained segment allows us to run a change-point statistical test designed for a single break detection. We emphasize here only the situation where a change in trends is occurring over different time periods. Several segmentation methods are available in the litterature, and in this work we explore the segmentation proposed in Jiangling et al. (2011) which does not rely on any probabilistic assumption. The approach for obtaining such automatic segmentation is based on locating turning points which are defined first by identifying local maximum and minimum. In general these turning points indicate a change in the direction of local trends of the series. The basic idea of this procedure is to split the series into different periods such that each extracted segment is associated to a single trend. In order to gain the full benefit of such segmentation, we first apply it the original series and then to their scaling coefficients up to some level J of the MODWT transform. The advantage of the MODWT transform is twofold, first we do not loose the existing different local trends in the original series and secondly the data is smoothed to some level so that we nearly avoid the burden of how to choose the segmentation level. In our simulation we show that a good the segmentation level is often either 2 or 3. Its important to note here that this procedure does not require to run the wavelet transform to a higher level, the segmentation works better for lower scale of the MODWT transform as can be seen in figure 2.

The proposed procedure begin first by segmenting the series X_t and its scaling coefficients by identifying separately the existing turning points and counting their numbers within each series. Then based on these points, as described in the algorithm in section 3 a new segmentation is derived where each turning point is within or in the middle of these new segments. The corresponding time interval of these new segments obtained from each series are then used to segment both the series and its scaling coefficients. Each turning point within each series is then subject to a statistical test and checked whether its a true change in the mean.

The outline of this paper is as follows. Section 2 provides a brief review of the MODWT tranform in the literature. The scaling coefficients for a series with multiple mean break are depicted in this section. Section 3 introduces how turning points are selected along with a segmentation algorithm. In section 4 we give a summary of one of the mean statistical test used in a single change point detection problem. Section 5 presents some Monte Carlo experiments. Section 6 provides a some conclusions.

Let $\mathbf{X} = (X_1, \dots, X_N)$ be an observed time serie for which we want to test for several breaks in the mean. Assume that $\mathbf{X}$ is regarded under H_0 as a realization of discrete time stochastic process with zero mean, and constant variance.

The null hypothesis of our test problem would be

$$H_0 : E(X_t) = \mu \quad t = 1, \dots, N \quad (1.1)$$

which we wish to test against the alternative hypothesis

$$H_1 : E(X_t) = \mu_i \quad \text{for} \quad k_{i-1} \leq t < k_i, \quad i = 1, \dots, K < N \quad (1.2)$$

where $k_0 = 1$ and the times $t = k_i$ of mean change are unknown.

2 The MODWT transform

The maximal-overlap discrete wavelet transform (MODWT) also called the undecimated or shift invariant discret wavelet transform has been discussed in the wavelet literature see Coifman and Donoho (1995); Daubechies (1992); Donald and Andrew (2000). For the class of discrete compactly supported Daubechies wavelets, we denote $\{\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$, $l = 0, \dots, L_j - 1\}$ the level j of a wavelet and scaling filters of length $L_j = (2^j - 1)(L - 1) + 1$. Theses filters are discussed and given in more details in Donald and Andrew (2000). The stochastic processes resulting from applying these filters to $\{X_t\}$ are respectively given by the level j wavelet and scaling coefficients

$$V_{j,t} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} X_{t-l} \quad j = 1, 2 \dots J \quad (2.1)$$

The MODWT wavelet and scaling coefficient based on a finite sample size are given by

$$\text{and } \tilde{V}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \dots, N-1 \quad (2.2)$$

obtained as a result of circularly filtering $X_0, X_1, \dots, X_{N-1}$ with the filters $\{\tilde{h}_{j,l}\}$ and $\{\tilde{g}_{j,l}\}$. Where $X_{t \bmod N} = X_t$ if $t \geq 0$ and $X_{t \bmod N} = X_{N-|t|}$ if $t < 0$. Note that $\tilde{W}_{j,t} = W_{j,t}$ for $t \geq L_j - 1$ and $j = 1, \dots, J$ where J is the maximum level up to which we run the MODWT.

Its important to note that in general the wavelet coefficients remove trends and demean the process X_t whereas the scaling coefficients $V_{j,t}$ preserve the overall shape of X_t and keep track of all important data points, and particularly the turning points that detect direction change in the local trends. These changes should be regarded as a change in the mean of the original series X_t as shown figure 1.

3 Segmentation procedure

Turning points are defined as local minimum or maximum points in a time series, they are identified as the start and end of a local trend. They are used to segment the series into different periods such that each extracted segment is expected to cover a single trend over that period. In this work we explore the segmentation procedure described by the Algorithm 2 in Jiangling et al. (2011). The proposed method generates segments at different levels of details up to a specified segmentation level, and does not depend on any threshold. A set of turning points is selected at a specified level of the segmentation. This segmentation level aim to reduce the number of tuning points by combining small trends into large ones. The series X_t and the scaling coefficients $V_{j,t}$ $j = 1, \dots, J$ are then segmented separately according to the previous method, and a small number of turning points are identified within each series. These are regarded as possible change points to be tested. To this end we use these set of identified turning points to split separately the series X_t and $V_{j,t}$ $j = 1, \dots, J$ into new set of segments such that for each series each point of the selected set of turning point belongs to at least one of these segments. Obviously for a given series this procedure allows for a situation where one turning point might belong to more than one segment, which might occurs particularly if some of these identified points are too close. To avoid this problem, we introduce a parameter λ_d that specifies the minimal distance we should have between two successive turning points. This help to further reduce the number of turning points. A practical choice of λ_d would depend on the segment size N_s , in our case we set it so that $\lambda_d \geq 6$ which correspond to the smallest integer value of $N_s/15$. When a turning point is within two different segments that are not disjoint, then it will be subject twice to the same statistical test but with different data. This should strengthen the chance of good detection. As illustrated in the example of simulated models with three breaks, the number of turning

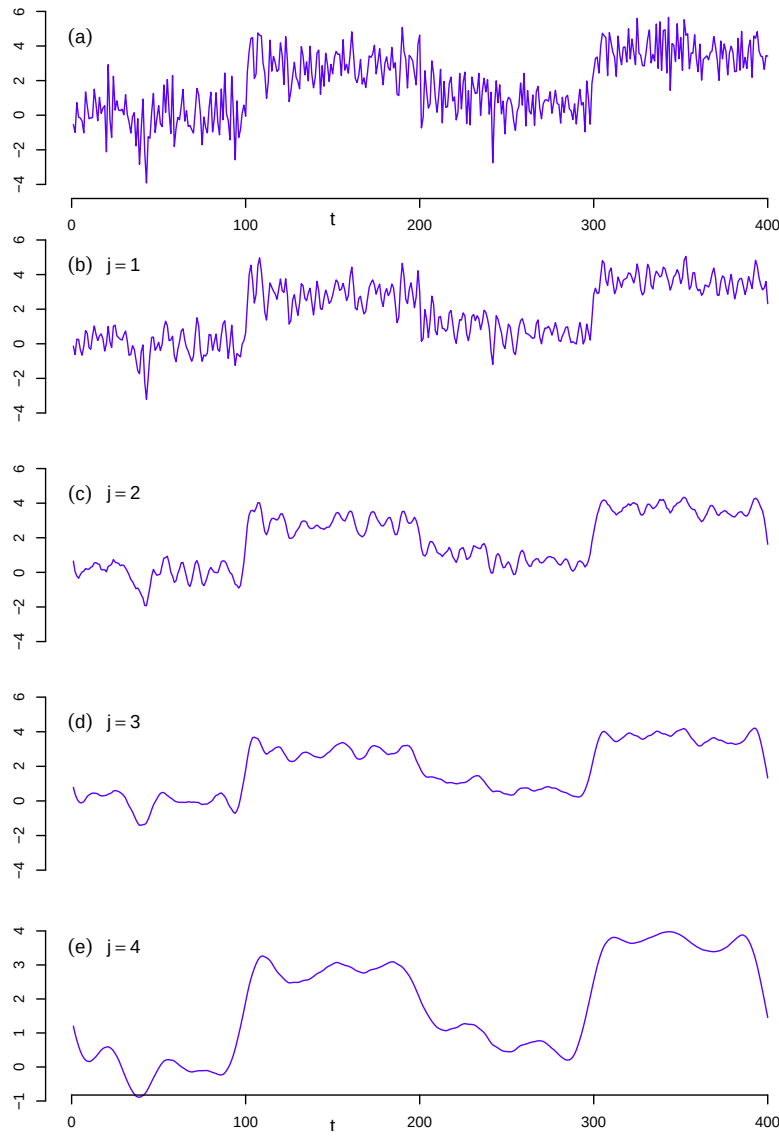


Figure 1: (a) Gaussian G_t serie with multiple breaks at $k_1 = 100$, $k_2 = 200$ and $k_3 = 300$. (b)-(e) are the scaling coefficients $\tilde{V}_{j,t}$ respective for level $j = 1, 2, 3, 4$

points was in general less than 20 for a sample size of $N = 400$, and the correct detection of the true breaks were very high and models dependent as shown in the table 1. The next algorithm describe the main steps of the new segmentation to split the original series X_t and the coefficients $V_{j,t}$.

3.1 Algorithm

Let $\{X_t, t = 1, \dots, N\}$ be an observed time series and $\{p_k, k = 1, \dots, K\}$ be the set of turning points locations ordered in times retained from its segmentation with $K > 2$ according to the procedure in Algorithm 2 of Jiangling et al. (2011). An arbitrary choice of the minimum distance paramter $\lambda_d = 5$ was used, and if two successive turning points are too close we discard the smallest in magnitude. Denote by N_s the minimum segment size to be used so that the power of the statistical test is not affected. Then given the random location of these turning points, one way to split up X_t into segments Y_k is as follow

Algorithm

```
(1) for  $k = 1$ 
    if  $(p_1 \geq N_s/2)$ 
         $Y_1 = X[1 : (p_1 + N_s/2)]$ 
    else  $Y_1 = X[1 : (p_1 + N_s)]$ 
(2) set  $k = 2$ 
while  $(k \leq K)$ 
    if  $(p_k \geq p_{k-1} + N_s/2)$  and  $(p_k < N - N_s/2)$ 
         $Y_k = X[(p_{k-1} + 1) : (p_k + N_s/2)]$ 
    else  $Y_k = NULL$ 
     $k = k + 1$ .
```

Note that the same segmentation procedure is applied also to the scaling coefficients $V_{j,t}$ for $j = 1, \dots, J$ where X_t is replaced with $V_{j,t}$. For a fixed series, based on the locations of the set of selected turning points a new segmentation is derived such that each turning point is within a segment of size bigger than N_s .

4 Change point statistic

The constructed segments $Y_k, k = 1, \dots, K$ are assumed to contain all mean break points and are then subject to a statistical test. Due to the random fluctuations in the data that affect both segmentations, the number of retained segments Y_k according to the previous algorithm should be higher than the true number of breaking points. This only increases the probability of covering all the mean break points by these segments. Note that each turning point within a segments Y_k is subject to a statistical test to check whether its in fact a true change point. There are many statistical tests available in the literature that can be applied for this purpose. We choose to apply a nonparametric test proposed by Lobato (2001), and Xiaofeng and Xianyang (2010). The basic idea of this test suggests to bypass the estimation problem of the long run variance as discussed in Vogelsang (1998, 1999) by applying a self normalizer to the test statistic. Two types of self normalisation were introduced in this context, the first is based on the proposed test statistic in Lobato (2001) and the second type is proposed in Xiaofeng and Xianyang (2010). We make use of the second type, and in our setting, we define the partial sums computed from the scaling coefficients $V_{j,t}$ for $j = 1, \dots, J$.

$$S_j(t_1, t_2) = \sum_{t=t_1}^{t_2} V_{j,t}, \text{ if } L_j - 1 \leq t_1 < t_2 \text{ and } 0 \text{ otherwise}$$

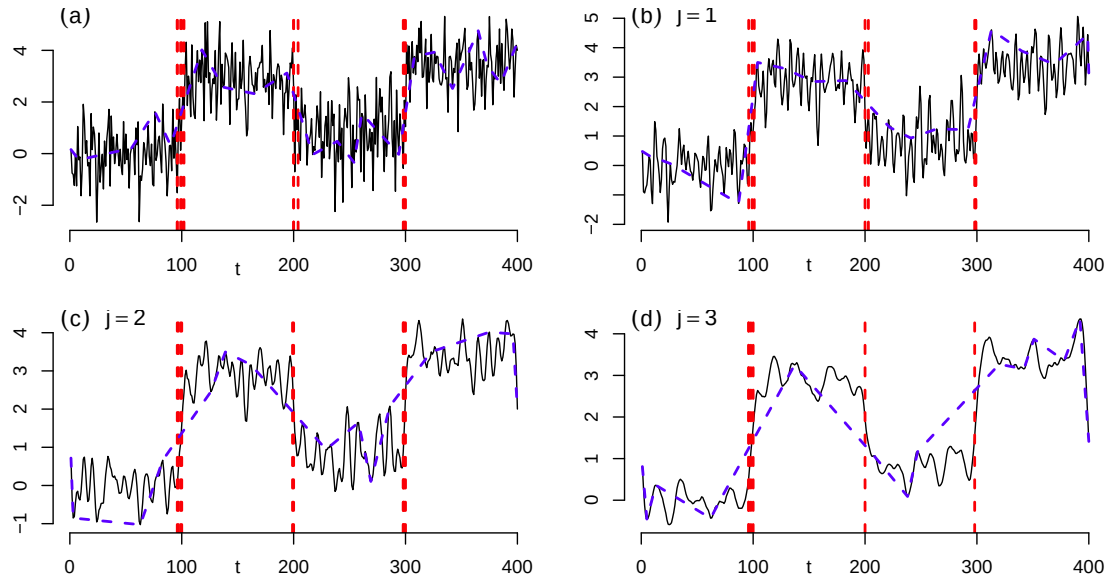


Figure 2: (a) Gaussian series G_t with multiple mean breaks. (b)-(d) are the scaling coefficients $\hat{V}_{j,t}$ respective for level $j = 1, 2, 3$. The segmentation of each series is plotted in dashed lines. The vertical lines are the detected breaks after running the statistical test over the set of segments within each series

Then the normalization process is defined for $k = L_j - 1, L_j, \dots, N - 1$ by

$$R_j(k) = \frac{1}{N_j^2} \left[\sum_{t=L_j-1}^k (S_j(1, t) - \frac{t}{k} S_j(1, k))^2 + \sum_{t=k+1}^{N-1} (S_j(t, N) - \frac{N-t+1}{N-k} S_j(k+1, N))^2 \right] \quad (4.1)$$

The statistic is then given by

$$Q_{V_j} = \max_{L_j-1 \leq k \leq N-1} \frac{T_j^2(k)}{R_j(k)} \quad (4.2)$$

where $T_j(k) = \frac{1}{\sqrt{N_j}} \sum_{t=L_j-1}^k (\hat{V}_{j,t} - \bar{V}_j)$ $k = L_j - 1, \dots, N$

A similar statistic Q_X based on X_t is computed where $V_{j,t}$ is replaced with X_t .

The asymptotic distribution of Q_{V_j} or Q_X is not standard, its critical values are computed by means of simulations, and are tabulated in Xiaofeng and Xianyang (2010). Based on the scaling coefficients V_{j,t_i} , the test statistic Q_{V_j} is computed separately over each segment Y_k and compared to its critical values. For instance, for $\alpha = 5\%$, then the simulated 95% quantile value is 40.1.

Any turning point that is selected by this statistical test as a breaking point is retained. If the same point is selected more than one time because its belong to two different non disjoint segments, and if the estimated locations in time are different, then we can adopt a rule to choose the point that correspond to the highest value of statistic Q_{V_j} .

5 Simulations

To illustrate the proposed procedure, we consider three different simulated models each with sample size $N = 400$ given in the example below. We set $\mu_1 = 0$, $\mu_2 = 3$, $\mu_3 = 1.5$ and $\mu_4 = 3.5$ so that we have three breaks in the mean at times $k_1 = 100$, $k_2 = 200$ and $k_3 = 300$. The process $\{\epsilon_t\}$ is a Gaussian uncorrelated sequence with mean zero and variance 1. The Least Asymmetric wavelet filter $LA(8)$ of length $L = 8$ was used to run the transform up to level $J = 3$. The choice of the wavelet filter was arbitrary with moderate length. In order to get the set of turning points we apply the segmentation procedure in Jiangling et al. (2011) at segmentation level 3, then by using the proposed algorithm in section 3 with $N_s = 100$ and $\lambda_d = 5$ we derive a new set of segments Y_k for $k = 1, \dots, K$ separately for the original series X_t and $V_{j,t}$, $j = 1, 2, 3$. We then run the statistical test given in (7) over each segment Y_k . Based on the test result we decide whether there exists a change point in a given segment. Any break detection is then retained and plotted as shown in figure 2 for the Gaussian model.

This process of generating data sets, running the above procedure separately for each series and then testing at significance level $1 - \alpha = 95\%$ over each segment Y_k was repeated a total of $n = 100$ times for each sample size. The proportion of true break detection for these replications is summarized in 1 for three models. For illustration, the results for a single replication for model (a) are shown in figure 2.

5.1 Example

(a) Gaussian model

$$X_t = \mu_1 I_{(0 \leq t < 100)} + \mu_2 I_{(100 \leq t < 200)} + \mu_3 I_{(200 \leq t < 300)} + \mu_4 I_{(300 \leq t \leq 400)} + \epsilon_t \quad (5.1)$$

(b) Non-linear model

$$X_t = \mu_1 I_{(0 \leq t < 100)} + \mu_2 I_{(100 \leq t < 200)} + \mu_3 I_{(200 \leq t < 300)} + \mu_4 I_{(300 \leq t \leq 400)} + \epsilon_t^2 \quad (5.2)$$

(c) $AR(1)$ model

$$X_t = 0.7X_{t-1} + \mu_1 I_{(0 \leq t < 100)} + \mu_2 I_{(100 \leq t < 200)} + \mu_3 I_{(200 \leq t < 300)} + \mu_4 I_{(300 \leq t \leq 400)} + \epsilon_t \quad (5.3)$$

6 CONCLUSIONS

The problem of detecting multiple mean changes in time series seems to be a difficult issue to solve particularly if we adopt the nonparametric approach without any assumption on the time series model. The wavelet transform comes as a handy very useful mathematical tool that can provide a smoother representation of the series X_t without losing details in the shape of the existing local trends. In fact as can be seen from figure 2, the scaling coefficients $V_{j,t}$ offers to run the same statistical test using different data set for the same test problem. However the approach used by the segmentation starts to lose the tracking of the local up trends and down trends when we move to higher scales $J \geq 4$. Therefore the choice of the MODWT transform level J should be kept lower.

In order to search and locate the mean breaks, a segmentation procedure was used to subdivide in an appropriate way the sample time interval of each series so that each break belongs to at least one segment. Because we do not have a prior information about the number of breaks, we could end up with a large number of turning points if we rely on segmentation alone from the Algorithm in Jiangling et al. (2011). The minimum distance parameter λ_d was introduced to overcome this problem and therefore reduces the number of turning points which helps in getting a small number of segments.

Table 1: Proportion of true mean break detection from $n = 100$ replications each with sample size $N = 400$. The true time location of breaks are k_1 , k_2 and k_3

Model	$k_1 = 100$	$k_2 = 200$	$k_3 = 300$
G_t	100%	100%	100%
$V_{1,t}$	100%	98%	100%
$V_{2,t}$	100%	98%	100%
$V_{3,t}$	98%	89%	84%
G_t^2	100%	96%	99%
$V_{1,t}$	100%	98%	97%
$V_{2,t}$	100%	96%	100%
$V_{3,t}$	98%	88%	85%
$AR(1)$	80%	53%	66%
$V_{1,t}$	79%	45%	67%
$V_{2,t}$	75%	45%	56%
$V_{3,t}$	68%	33%	37%

As illustrated in ??, the proportion of true mean breaks detection is very high and around 100% for the gaussian model (a) and its scaling coefficients up to level $j = 2$. This proportion get slightly lower for level $j = 3$. We get a very similar result for the square of the Gaussian model (b). The lower proportions in the $AR(1)$ model particularly for the second break $k_2 = 200$ are likely due to the differentiation operation in $AR(1)$ model which affects the different means. It should be noted as well that the power of the statistical test used here may be affected due to the varying small sample size of a segment Y_k . In general the performances of this procedure are good enough, and can be improved by allowing a larger sample size for Y_k , which is arbitrarily chosen in our procedure. It should be noted here as well that the choice of the wavelet filter was also arbitrary and any other wavelet filter could well be applied.

References

- Antoniadis, A. Oppenheim, G. (1995). Translation-invariant Denoising. In Wavelet and Statistics: Lecture Notes in Statistics 103, Springer-Verla, 125.
- Coifman R. and Donoho D. L. (1995). Translation-invariant Denoising. In: Wavelet and Statistics: Lecture Notes in Statistics. Antoniadis, A. Oppenheim, G., New York: Springer-Verla Vol. 103, 125–50.
- Donald B. Percival and Andrew T. Walden (2000) Wavelet Methodes for Time Series Analysis. Cambridge Series in Statistical and Probabilistic Mathematics.
- Daubechies I. (2000), Ten Lectures on wavelets, Philadelphia: SIAM.

- Paul Fearnhead (2006), Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16, 203–213.
- Jiangling Y., Yain-Whar Si and Zhiguo G.(2011) Financial Time Series Segmentation Based On Turning Points Proceedings of 2011 International Conference on System Science and Engineering, Macau, China - June 2011.
- Kim J. and Cheon S. (2013), Bayesian Multiple Change-Point Estimation and Segmentation *Communications for Statistical Applications and Methods* 20:6, 439454.
- Killick, R., Nam, C., Aston, J. and Eckley, I. (2012). changepoint.info: The changepoint repository. Available at <http://changepoint.info>.
- Lobato, I. N. (2001), Testing That a Dependent Process Is Uncorrelated, *J. of the American Stat. Association* 96, 1066–1076.
- Piotr Fryzlewicz (2014), Wild Binarysegmentation for multiple Change-Point Detection, *The Annals of Statistics* 42,(6), 2243–2281.
- Vogelsang T. J. (1998) , Testing for a Shift in Mean Without Having to Estimate Serial-Correlation Parameters, *Journal of Business and Economic Statistics*, 16, 73–80.
- Vogelsang T. J. , Sources of Nonmonotonic Power When Testing for a Shift in Mean of a Dynamic Time Series, *Journal of Econometrics*, 88, 283–299.
- Xiaofeng Shao and Xianyang Zhang (2010), Testing for Change Points in Time Series, *J. American Stat. Association*, 105:491, 1228–1240.