Stony Brook University

OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

Multiple Change-points Estimation in GARCH Models

A Dissertation presented

by

Sichen Zhou

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

August 2015

Stony Brook University

The Graduate School

Sichen Zhou

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation

Haipeng Xing - Dissertation Advisor
Associate Professor, Department of Applied Mathematics and Statistics

Song Wu - Chairperson of Defense
Assistant Professor, Department of Applied Mathematics and Statistics

Xinyun Chen
Assistant Professor, Department of Applied Mathematics and Statistics

Yixin Fang
Assistant Professor, Division of Biostatistics, School of Medicine
New York University

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Multiple Change-points Estimation in GARCH Models

by

Sichen Zhou

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2015

The generalized conditional heteroscedastic (GARCH) models are often used to estimate volatility in financial markets as they mimic the patterns in real world with volatility clustering as well as high excess kurtosis. However, in applications to asset return series, they usually possess undesired persistence in volatility, which can be explained by structure changes in parameters associated with significant economic events such as financial crises. From this motivation, we provide an estimation procedure for multiple parameter changes in GARCH models. By introducing the specified forward and backward filtration and combining them with Bayes' theorem, our estimation procedure has attractive statistical and computational properties and yields explicit recursive formulas to provide semi-parametric estimates for the piecewise constant parameters. Based on the estimates given above with the quasi-likelihood of our model and the modified Bayesian information criterion (MBIC), we also develop a segmentation procedure to give inference on the number and locations of the change-points that partition the unknown parameter sequence into segments of equal values. Furthermore, we propose an expectation-maximization (EM) algorithm to estimate the change-points probability p in our model. Simulation studies are used to compare our performance to the existing procedure and the "oracle"

estimates, which assume that the change-points are already known. The mean Euclidean error (EE), the Kullback–Leibler divergence (KL), the goodness of fit and the accuracy rate of the numbers of change-points detected are given. Finally, illustrative applications to the S&P 500 index and the IBM stock returns are shown to give an insight how our estimation results coincide with the real financial crises.

Contents

1 Introduction **1**

1.1 Literature on Change-points Problems 2

1.2 Previous Applications to Financial Time Series 3

1.3 Introduction of Existing Procedures 4

1.3.1 Markov Regime-switching Models 4

1.3.2 Cumulative-sums (CUSUM) Type Statistics 5

1.3.3 Existing Segmentation Procedures 5

1.4 Motivation 6

1.5 Outline 8

2 Modeling Change-points in GARCH and Semi-parametric Estimation Procedure **10**

2.1 Modeling Multiple Change-points in GARCH (1,1) 10

2.2 Local Likelihood Mixture 11

2.3 Forward Filtration 12

2.4 Backward Filtration 13

2.5 Semi-parametric Estimation 14

2.6 BCMIX Approximation 16

2.7 Use of Blocks 17

2.8 Estimation of Change-points Probability p 18

3 Unconditional Variance Based Segmentation Procedure **21**

3.1 Intuition 21

3.2 Unconditional Variance 22

3.3 Δ_t for Detecting Multiple Change-points 22

3.4 Modified Bayesian Information Criterion 25

3.5 Top Down Approach 29

3.6 Segmentation Procedure 29

3.7 Comparing to Binary Segmentation Algorithm 31

4 Simulation Studies **35**

4.1 Illustrative Example 35

4.1.1 Semi-parametric Estimates 36

4.1.2 Segmentation Results 37

4.1.3 Choice of Minimum Possible Distance m 39

4.1.4 Choice of Change-points Probability p 39

4.2 Simulation Studies with Simultaneous Changes 41

4.2.1 Semi-parametric Estimates 43

4.2.2 Segmentation Results 52

4.2.3 Choice of Change-points Probability p 61

4.3 Simulation Studies with Individual Parameter Shifts 62

4.3.1 Semi-parametric Estimates 63

4.3.2 Segmentation Results 68

5 Real Data Analysis **71**

5.1 Daily Log Return Series 71

5.1.1 S&P 500 Index 71

5.1.2 IBM Stock Return 77

5.2 Weekly Log Return Series 82

5.2.1 S&P 500 Index 82

5.2.2 IBM Stock Return 86

6 Conclusion and Future Work **91**

Bibliography **93**

List of Figures

1.1 S&P 500: daily log return series from 1999 to 2014 6

1.2 S&P 500: estimated volatility with the GARCH(1,1) model 7

3.1 Illustration: detecting possible change-points by Δ_t with $m = 100$ 25

4.1 Illustration: semi-parametric estimates with no blocks and a block of 4 . 36

4.2 Illustration: segmentation results with no blocks and a block of 4 37

4.3 Illustration: segmentation results with $m = 100$ and 200 38

4.4 Illustration: semi-parametric estimates given different p with a block of 4 40

4.5 Simulation I: typical estimates in Scenario 1 43

4.6 Simulation I: typical estimates in Scenario 2 44

4.7 Simulation I: typical estimates in Scenario 3 45

4.8 Simulation I: typical estimates in Scenario 4 46

4.9 Simulation I: typical estimates in Scenario 5 47

4.10 Simulation I: typical estimates in Scenario 6 48

4.11 Simulation I: typical segmentation results in Scenario 1 53

4.12 Simulation I: typical segmentation results in Scenario 2 54

4.13 Simulation I: typical segmentation results in Scenario 3 54

4.14 Simulation I: typical segmentation results in Scenario 4 55

4.15 Simulation I: typical segmentation results in Scenario 5 55

4.16 Simulation I: typical segmentation results in Scenario 6 56

4.17 Simulation II: typical estimates in Scenario 1 64

4.18 Simulation II: typical estimates in Scenario 2 64
4.19 Simulation II: typical estimates in Scenario 3 65
4.20 Simulation II: typical estimates in Scenario 4 65
4.21 Simulation II: typical estimates in Scenario 5 66
4.22 Simulation II: typical estimates in Scenario 6 66

5.1 S&P 500: semi-parametric estimates of daily log return series with p_0 . . 72
5.2 S&P 500: Δ_t of daily log return series given $\widehat{p}$ with one update and p_0 . 73
5.3 S&P 500: segmentation results of daily log return series 74
5.4 S&P 500: post-segmentation estimated volatility of daily log return series 77
5.5 IBM: daily log return series from 1999 to 2014 78
5.6 IBM: semi-parametric estimates of daily log return series with p_0 78
5.7 IBM: Δ_t of daily log return series given $\widehat{p}$ with one update and p_0 79
5.8 IBM: segmentation results of daily log return series 80
5.9 IBM: post-segmentation estimated volatility of daily log return series . . 82
5.10 S&P 500: weekly log return series from 1999 to 2014 83
5.11 S&P 500: semi-parametric estimates of weekly log return series with p_0 . 84
5.12 S&P 500: segmentation results of weekly log return series 84
5.13 S&P 500: post-segmentation estimated volatility of weekly log return series 86
5.14 IBM: weekly log return series from 1999 to 2014 87
5.15 IBM: semi-parametric estimates of weekly log return series with p_0 88
5.16 IBM: segmentation results of weekly log return series 89
5.17 IBM: post-segmentation estimated volatility of weekly log return series . 90

List of Tables

3.1 MBIC: range of penalty given the number of change-points k and the length of the series n with $m = 100$ 27

3.2 MBIC: range of penalty given the number of change-points k and the length of the series n with $m = 200$ 28

3.3 LM: asymptotic critical values from Andrew(1993) 32

4.1 Illustration: convergence of change-points probability p and corresponding numbers of change-points k 39

4.2 Simulation I: comparing semi-parametric estimates with "oracle" estimates 51

4.3 Simulation I: comparing segmentation results between UVB and BSA . . 57

4.4 Simulation I: comparing ΔMBIC between UVB and BSA 59

4.5 Simulation I: comparing $\widehat{p}$ with one update between different scenarios . 61

4.6 Simulation II: comparing semi-parametric estimates with "oracle" estimates 67

4.7 Simulation II: comparing segmentation results between UVB and BSA . 69

5.1 S&P 500: comparing post-segmentation maximized log-likelihood and BIC of daily log return series 74

5.2 S&P 500: comparing pre- and post-segmentation piecewise estimates . . 75

5.3 S&P 500: detected structure changes of daily log return series 76

5.4 IBM: comparing post-segmentation maximized log-likelihood and BIC of daily log return series 81

5.5 IBM: comparing pre- and post-segmentation piecewise estimates 81

5.6 IBM: detected structure changes of daily log return series 81
5.7 S&P 500: comparing post-segmentation maximized log-likelihood and BIC of weekly log return series 85
5.8 S&P: detected structure changes of weekly log return series 85
5.9 IBM: comparing post-segmentation maximized log-likelihood and BIC of weekly log return series 89
5.10 IBM: detected structure changes of weekly log return series 89

Acknowledgements

I would like to thank my advisor, Prof. Haipeng Xing, for suggesting this exciting topic, for interesting and enlightening discussions and for his guidance and continuous support.

Also, I would like to thank Prof. Song Wu, Prof. Xinyun Chen and Prof. Yixin Fang for their valuable experience and advice. It is my honor to have them on my dissertation committee.

Lastly, I would like to thank all my fellow group members and friends, especially Ke Wang, Shanshan Li, Tiantian Li and Qi Fan, for all their support and help.

Chapter 1

Introduction

The generalized conditional heteroscedastic (GARCH) models were proposed by Bollerslev(1986)[1] to characterize observed time series by modeling the error terms,

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{s} \alpha_i y_{t-i}^2 + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2 \tag{1.1}$$

where ε_t is symmetric i.i.d. with mean equal to zero and variance equal to one and have finite kurtosis. GARCH models are often used to fit financial time series since they are consistent with the volatility clustering as well as the high excess kurtosis. However, most asset return series cannot be adequately described by GARCH models, where undesired persistence in volatility shows up. Many researchers attribute this to structure changes in parameters associated with significant economic events such as financial crises. In this dissertation research, we introduce an estimation procedure for multiple parameter changes in GARCH-type volatility models, where change-points take place with constant parameters in between, corresponding to the sudden shocks caused by extraordinary economic events.

In this chapter, first we review some classic works on change-points problems, where difficulty increases dramatically from single change-point problems to multiple ones. Then previous applications to financial time series are listed in the second section. The Markov

regime-switching models, the cumulative-sums (CUSUM) type statistics, and the existing segmentation procedures are introduced to solve problems in ARCH or GARCH settings. After discussing the limitation of existing methods, we introduce the motivation of our study in the fourth section. Finally, the outline of this dissertation is given.

1.1 Literature on Change-points Problems

The initial change-point problem was proposed by Page(1955)[2], where the parameters can have at most one change and partial solution was given only when initial parameters are known. Quandt(1960)[3] and Hinkley(1970)[4] first provided frequentist solutions to this problem by using likelihood ratio statistic for testing the hypotheses related to the position of the change-point. Both methods can be applied to detect structure changes in linear regression. Andrews(1993)[5] further provided Wald test, Lagrange multiplier (LM) test, and likelihood-ratio(LR)-like test based on generalized method of moments (GMM) estimators, where asymptotic critical values are provided from a Bessel process of order q corresponding to the number of shifting parameters. Meanwhile, Bayesian approach dates back to Shiryaev(1963)[6], in which concepts of "mean delay time" and "false alarm" were introduced. Later on, Carlin, Gelfand and Smith(1992)[7] proposed a hierarchical Bayesian change-point model and utilized an associated Gibbs sampler to obtain the desired marginal posterior densities.

Problems of multiple change-points are difficult due to computational complexity. One frequentist approach, proposed by Bai and Perron(1998)[8], is to compute the least squares estimates of piecewise constant regression parameters by dynamic programming. Later, Qu and Perron(2007)[9] generalized the procedure to multivariate regression with arbitrary restrictions using restricted quasi-likelihood maximization. As for Bayesian approach, the successive developments are in works of Liu and Lawrence(1999)[10], Wang and Zivot(2000)[11], both of which specify priors for parameter distributions and use the Gibbs sampler with a Metropolis-Hastings procedure. Another approach, based on the

binary segmentation procedure introduced by Sen and Srivastava(1975)[12], is the circular binary segmentation (CBS) method proposed by Olshen, et al.(2004)[13], which makes computation more conveniently.

1.2 Previous Applications to Financial Time Series

Researchers have shown great interest in modelling financial time series with time-varying parameters. Inclan and Tiao(1994)[14] considered a simple variance-change model and then introduced the iterated cumulative sums of squares (ICSS) algorithm which can detect multiple changes. Later, Chen and Gupta(1997)[15] provided similar results using unbiased Schwarz information criterion (SIC) with an application to stock prices. Chib, Nardari and Shephard(2002)[16], making use of the reversible jump Markov chain Monte Carlo (MCMC) introduced by Green(1995)[17], which can simulate posterior distributions even with unknown number of parameters, developed a procedure to detect structural changes in stochastic volatility (SV) models. Meanwhile, Bai and Perron(2003)[18] in their paper provided an illustrative example to determine the break dates of US ex-post real interest rate. Recently, Xing, Sun and Chen(2012)[19] modeled credit rating with piecewise homogeneous Markov chains, which is proved to provide better forecasts than the corresponding time-homogeneous model.

Specific to GARCH-type volatility models, change-points problems are even harder to solve because volatility cannot be directly observed from the series and there exists non-linear dependence of the conditional variance on the past observations. Hamilton and Susmel(1994)[20], Cai(1994)[21] proposed the Markov regime-switching models to analyze the volatility in ARCH models. In these models, parameters are allowed to change within several different regimes, where the transitions are governed by a hidden Markov chain. Gray(1996)[22], Dueker(1997)[23] further extended the regime-switching models to GARCH settings. Kokoszka and Leipus(2000)[24] considered ARCH models and derived a cumulative-sums (CUSUM) type statistic to detect a single change point.

It was generalized to GARCH settings by Berkes, et al.(2004)[25], using a statistic based on the approximate likelihood scores. On reviewing the previous developments, Andreou and Ghysels(2002)[26] evaluated the performance of several tests for structural breaks in the conditional variance dynamics of asset returns. Recently, Galeano and Tsay(2010)[27] studied the case in which multiple changes occur in individual parameters of the GARCH (1,1) model and provided an binary segmentation algorithm.

1.3 Introduction of Existing Procedures

1.3.1 Markov Regime-switching Models

Let s_t denote an unobservable state variable with value taken from 1 to N. A Markov regime-switching model in the GARCH (1,1) setting can be presented as follows,

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega_{s_t} + \alpha_{s_t} y_{t-1}^2 + \beta_{s_t} \widehat{\sigma}_{t-1}^2 \tag{1.2}$$

where

$$\widehat{\sigma}_{t-1}^2 = \sum_{n=1}^{N} P(s_{t-1} = n | \mathcal{Y}_{1,t-1})(\omega_n + \alpha_n y_{t-2}^2 + \beta_n \widehat{\sigma}_{t-2}^2)$$

Another way to establish the model is to assume N separate GARCH processes and $\sigma_{s_t,t}$ is chosen from the corresponding process given s_t for any time t,

$$y_t = \sigma_{s_t,t} \varepsilon_t, \quad \sigma_{n,t}^2 = \omega_{n,t} + \alpha_{n,t} y_{t-1}^2 + \beta_{n,t} \sigma_{n,t-1}^2 \tag{1.3}$$

Both models promise that $\sigma_{s_t,t}$ only depends on $\{s_t, \mathcal{Y}_{1,t}\}$ rather than $\{s_1, \ldots, s_{t-1}\}$, and therefore the embedded process $\{s_t\}$ satisfies the Markov property. A modified Kalman filter was proposed by Kim(1994)[28], which is followed by recent numerical methods to provide solutions.

1.3.2 Cumulative-sums (CUSUM) Type Statistics

The concept of CUSUM was first mentioned in Page(1955)[2] to detect changes in statistical quality control. For a sequence $\{x_1, \ldots, x_n\}$, a parameter change is detected when the cumulative sum S exceeds a certain threshold, which is defined as follows,

$$S_0 = 0, \quad S_{t+1} = \max\left(0, S_t + x_t - \widehat{x}_t\right)$$

where $\widehat{x}_t$ is an estimate of x_t provided that no parameter changes would take place.

Later on, more complex CUSUM type statistics are introduced to solve different types of change-points problems. One example is the Lagrange multiplier (LM) test developed by Andrews(1993)[5]. It uses the cumulative sum of score functions $\sum_{i=1}^{t} S_i(x)$ as the essence of the test statistic and determine there exists a structure change when $\max_{1 \leqslant t \leqslant n} LM(\sum_{i=1}^{t} S_i(x))$ exceeds the critical value. If so, choose t as the change-point with the largest $LM(\sum_{i=1}^{t} S_i(x))$.

1.3.3 Existing Segmentation Procedures

Numerous segmentation procedures are proposed to generalize single change-point detection methods to solve multiple change-points problems. One example is the binary segmentation algorithm proposed by Galeano and Tsay(2010)[27]. They usually consist of three steps,

1. Determine whether there exists a single change-point and if so, choose that with the largest test statistic from the single detection method as the change-point.

2. Repeating the single detection method in each segment divided by the detected structure changes until there are no extra change-points to be detected.

3. Refine the segmentation results in each segment by certain pairs of the detected structure changes, and repeat until the number and locations of all possible change-points remain unchanged.

1.4 Motivation

One of our motivation to develop a multiple change-points estimation procedure for GARCH models is the increasing importance to determine structure changes in financial time series. In real world, the economic environment shifted dramatically during 2008 Global Financial Crisis, where the fundamentals of global markets are impractical to remain unchanged. Meanwhile, recent research studies have shown time series models with homogeneous parameters would often lead to doubtful results, which is illustrated in the following example.

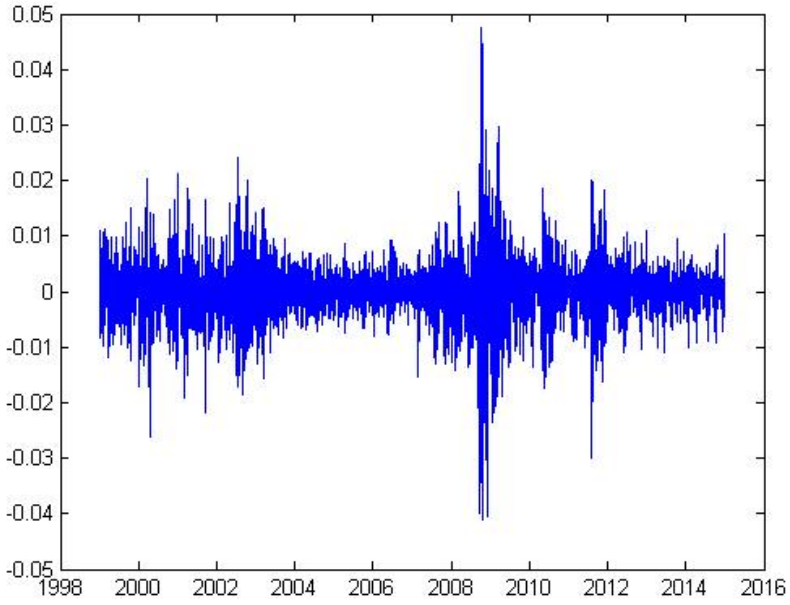

Figure 1.1: S&P 500: daily log return series from 1999 to 2014

Figure 1.1 shows the daily log return series of the S&P 500 index from 1999 to 2014. The data come from *Yahoo! Finance*. We apply the time-homogeneous GARCH(1,1) model to fit the series, which results in $\mu = 2.07 \times 10^{-4}$, $\omega = 2.69 \times 10^{-7}$, $\alpha = 0.905$, $\beta = 0.086$. Figure 1.2 shows the estimated volatility with the time-homogeneous GARCH(1,1) model. We can tell the volatility is much higher during the crisis. Notice that $\alpha + \beta = 0.991$ is close to 1, suggesting high volatility persistence. This both provides an

evidence to question the stationarity of the volatility and leads to poor estimation of the unconditional variance. One possible explanation is that the parameters are shifted during financial crises. In comparison, we simulate ten sequences of equal length of 300 from the time-homogeneous GARCH(1,1) model with different parameters and combine them together. The exact setting will be given in Section 4.1. The estimated parameters of the new series are $\alpha = 0.694,\ \beta = 0.305$, which presents the same pattern as the daily log return series, where $\alpha + \beta = 0.999$.

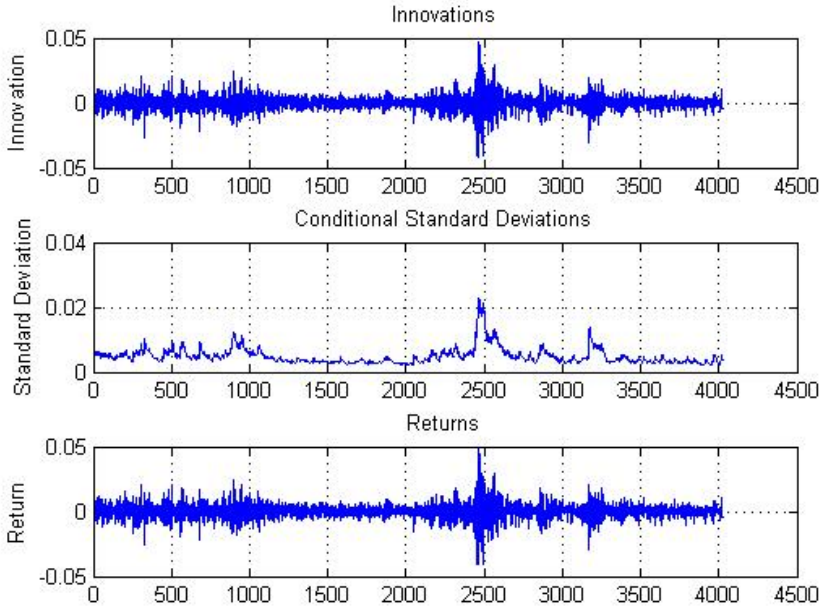

Figure 1.2: S&P 500: estimated volatility with the GARCH(1,1) model

The other reason for our interest is the limitations of existing methods in detecting change-points in GARCH settings. Markov chain Monte Carlo (MCMC) and the use of Gibbs samplers are computationally expensive, which are not good choices to solve multiple change-points problems. Markov regime-switching models have limited regimes, which cannot well reflect the impact of different extraordinary events. Moreover, though we can define the markets as "bull" and "bear", parameters can vary for different periods even in the same state. Cumulative-sums (CUSUM) type statistics are generally fast to compute and the critical values can be determined either theoretically under the null

hypnosis of absence of change-points or empirically using simulation studies to compare the power of the tests. However, the effectiveness of such approaches are greatly affected by the locations of the true structural changes and the estimation errors. Lastly, results of segmentation procedures are often sub-optimal, leading to underestimate the number of structural changes.

Thus, we would like to develop a procedure effectively in computation, of which the time cost should grow linearly corresponding to the length of the time series. Rather than implementing segmentations directly, we would first provide an overall estimation of the piecewise parameters, and then segment the sequence given our estimate results so as to provide inference on the number and locations of the structure changes. Though it is well understood that structure changes at the beginning and the end of the series are impossible to detect, we hope to develop a consistent procedure for the rest of the sequence however the change-points are distributed.

1.5 Outline

In this dissertation, for simplicity, we focus on the GARCH (1,1) model, which is the most commonly used GARCH model in financial time series analysis. Further extensions to general GARCH-type volatility models, e.g. exponential GARCH, would be made possible by modifying the corresponding likelihood functions.

In Chapter 2, we set up a model of multiple change-points in the GARCH (1,1) setting. After presenting the forward and the backward filtration, we provide explicit recursive formulas for the semi-parametric estimates of the piecewise constant parameters. Then, the bounded complexity mixture (BCMIX) approximation and the use of blocks are introduced to achieve higher computational speed. Later, an expectation-maximization (EM) algorithm is developed to provide estimation of the change-points probability p.

Chapter 3 presents a segmentation procedure based on a function Δ_t in regard to the impact of our estimates over the unconditional variance, where we use a modified

Bayesian information criterion (MBIC) inspired from Zhang and Siegmund(2007)[29] to determine the number and positions of structure changes. We further apply a top down approach to improve the results. After we summarize our segmentation procedure, the binary segmentation algorithm by Galeano and Tsay(2010)[27] is introduced, where we implement their procedure to the case in which parameters change simultaneously so as to compare with our own procedure.

Chapter 4 first presents an illustrative example of a series which consists of ten time-homogeneous GARCH(1,1) sequences, where we apply our estimation and segmentation procedure and also study the choices of the minimum possible distance m and the change-points probablity p. Then simulation studies of our proposed estimates are given, where the estimation errors are evaluated by the mean Euclidean error (EE), the Kullback-Leibler (KL) divergence and the goodness of fit comparing to the "oracle" estimates, which assume that the structure changes are already known. We also list the accuracy rate of the numbers of change-points detected from our own procedure, in comparison with the results from the binary segmentation algorithm (BSA) by Galeano and Tsay(2010)[27] and our modification for simultaneous changes. Later we compare with their procedure under the circumstances of individual parameter shifting.

In Chapter 5, we apply our procedure to the daily and weekly log return series of the S&P 500 index and the IBM stock returns, where estimated change-points are shown to be coincide with the major economic events. Further discussion and some concluding remarks are given in Chapter 6.

Chapter 2

Modeling Change-points in GARCH and Semi-parametric Estimation Procedure

2.1 Modeling Multiple Change-points in GARCH (1,1)

We consider a GARCH (1,1) model,

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega_t + \alpha_t y_{t-1}^2 + \beta_t \sigma_{t-1}^2 \tag{2.1}$$

where ε_t is independent and identically distributed with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$ and have finite kurtosis. Here we assume ε_t follows standard normal distribution, which can be generalized to elliptically symmetric distributions to fit heavy tails. Instead of being time-homogeneous, the parameter vector $\theta_t = (\omega_t, \alpha_t, \beta_t)'$ is supposed to have occasional changes such that for $t > 1$, the indicator variables,

$$I_t = \mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}}$$

are i.i.d. Bernoulli random variables with the change-points probability p, which describes how often structure changes would take place. The piecewise constant parameter θ_t is assumed to take a value under the regularity condition of (2.1), i.e. $\alpha_t + \beta_t < 1$ and $(\alpha_t + \beta_t)^2 + 2\alpha_t^2 < 1$. Such constraints promise that the series is piecewise stationary, where θ_t can be estimated piecewisely by quasi-maximum likelihood estimation (QMLE) introduced by Bollerslev and Wooldridge(1992)[30].

2.2 Local Likelihood Mixture

Our target is to develop an estimation procedure for the parameter vector θ_t to help detecting multiple change-points in the previous time-varying GARCH(1,1) model, i.e. find time t such that $I_t = 1$. Since θ_t remains constant between adjacent structure changes, we consider the closest structure changes that happen before and after time t. Define $K_t = \max\{s : I_s = 1,\ s \leqslant t\}$ and $\widetilde{K}_t = \min\{s : I_s = 1,\ s > t\}$, i.e. K_t and $\widetilde{K}_t$ are the most recent change time before (including the same as) and after time t. We further denote by C_{ij} the event that there is no change from $i+1$ to j but changes do occur at time i and $j+1$, i.e. $C_{ij} = \{I_i = I_{j+1} = 1,\ I_{i+1} = \ldots = I_j = 0\}$. Both i and j take values from 1 to n, where $i = 1$ and $j = n$ suggest there exists no change-point before and after time t respectively.

As introduced by Lai and Xing(2011)[31], in light of the law of total probability, the overall likelihood of θ_t can be written as the weighted average of the local likelihood given any combinations of the most recent changes. By denoting the partial sequence $(y_i, \ldots, y_j)$ with Y_{ij} for all $i \leqslant j$, we consider the local likelihood mixture (LLM) for θ_t given $Y_{1,n}$,

$$L(\theta_t; Y_{1,n}) = \sum_{1 \leqslant i \leqslant t \leqslant j \leqslant n} w_{ijt} L(\theta_t; Y_{ij}) \tag{2.2}$$

where $w_{ijt} = P(C_{ij}|Y_{1,n})$ is the probability that the closest change-points before and after time t are i and $j+1$. We will develop a recursive formula for computing w_{ijt} in the following sections.

$L(\theta_t; Y_{ij})$ is the likelihood function of θ_t given Y_{ij}, which can be further written as the quasi-likelihood function of the time-homogeneous GARCH(1,1) model under the assumption of no changes,

$$L(\theta_t; Y_{ij}) = \exp\{-\frac{1}{2}\sum_{t=i}^{j}(\log(2\pi) + \log\sigma_t^2 + \frac{y_t^2}{\sigma_t^2})\} \tag{2.3}$$

2.3 Forward Filtration

To derive a recursive formula for w_{ijt}, we need to compute the conditional probabilities given any recent changes before and after time t respectively and then combine them together. Therefore, first we introduce the following forward filtration to calculate the conditional probabilities based on the last change-point before time t. To start with, we denote $f_{ij} = L(\theta_t; Y_{ij}, C_{ij})$ the likelihood for Y_{ij} given C_{ij} of a constant θ_t from time i to j and $p_{it} = P(K_t = i|Y_{1,t})$ the probability that the most recent change time before and including time t is i given $Y_{1,t}$.

To compute the probability p_{it}, we decompose the likelihood $L(\theta_t; Y_{1,t})$ into conditional probabilities given each possible change time before and including time t,

$$L(\theta_t; Y_{1,t}) = f(\theta_t|Y_{1,t}) = \sum_{i=1}^{t} p_{it} f(\theta_t|Y_{i,t}, K_t = i) \tag{2.4}$$

and then we break down the conditional probability $f(\theta_t, y_t|Y_{1,t-1})$ based on whether structure change takes place at time t or not, i.e. $I_t = 0$ or 1,

$$f(\theta_t, y_t|Y_{1,t-1}) = pf(\theta_t, y_t|Y_{1,t-1}, I_t = 1) + (1-p)f(\theta_t, y_t|Y_{1,t-1}, I_t = 0) \tag{2.5}$$

Note that the first term on the right-hand side of (2.5) can be expressed as,

$$pf(\theta_t, y_t|Y_{1,t-1}, I_t = 1) = pf_{t,t}f(\theta_t|Y_{1,t}, I_t = 1) = pf_{t,t}f(\theta_t|Y_{t,t}, K_t = t)$$

and we can further expand the second term by conditioning on each possible change time before time t,

$$
\begin{aligned}
&(1-p)f(\theta_t, y_t|Y_{1,t-1}, I_t = 0) \\
&= (1-p)\sum_{i=1}^{t-1} P(K_{t-1} = i|Y_{1,t-1}, I_t = 0) f(\theta_t, y_t|Y_{1,t-1}, I_t = 0, K_{t-1} = i) \\
&= (1-p)\sum_{i=1}^{t-1} p_{i,t-1} f(y_t|Y_{1,t-1}, I_t = 0, K_{t-1} = i) f(\theta_t|Y_{1,t}, I_t = 0, K_{t-1} = i) \\
&= (1-p)\sum_{i=1}^{t-1} p_{i,t-1} \frac{f(Y_{i,t}, K_t = i)}{f(Y_{i,t-1}, K_{t-1} = i)} f(\theta_t|Y_{i,t}, K_t = i) \\
&= (1-p)\sum_{i=1}^{t-1} p_{i,t-1} \frac{f_{i,t}}{f_{i,t-1}} f(\theta_t|Y_{i,t}, K_t = i)
\end{aligned}
$$

Since $f(\theta_t|Y_{1,t}) \propto f(\theta_t, y_t|Y_{1,t-1})$, we compare the coefficients of $f(\theta_t|Y_{i,t}, K_t = i)$ between (2.4) and (2.5) for $1 \leqslant i \leqslant t$. Noting that $\sum_{i=1}^{t} p_{it} = 1$, the recursive formula for probability p_{it} is given by $p_{it} = p^*_{it} / \sum_{k=1}^{t} p^*_{kt}$, where

$$
p^*_{it} = \begin{cases} p f_{t,t} & if\ i = t, \\ (1-p) p_{i,t-1} f_{i,t} / f_{i,t-1} & if\ i < t, \end{cases} \tag{2.6}
$$

2.4 Backward Filtration

As for the backward filtration, we denote $q_{t,j} = P(\widetilde{K}_t = j+1|Y_{t+1,n})$ the probability that the most recent change time after t is $j+1$ given $Y_{t+1,n}$. By decomposing the likelihood $L(\theta_t; Y_{t+1,n})$ into conditional probabilities given each possible change time after time t,

$$
L(\theta_t; Y_{t,n}) = f(\theta_t|Y_{t,n}) = \sum_{j=t}^{n} q_{t,j} f(\theta_t|Y_{t,j}, \widetilde{K}_t = j+1) \tag{2.7}
$$

We break down the conditional probability $f(\theta_t, y_t|Y_{t+1,n})$ as follows,

$$
f(\theta_t, y_t|Y_{t+1,n}) = p f(\theta_t, y_t|Y_{t+1,n}, I_{t+1} = 1) + (1-p) f(\theta_t, y_t|Y_{t+1,n}, I_{t+1} = 0) \tag{2.8}
$$

where the first term on the right-hand side can be expressed as,

$$pf(\theta_t, y_t|Y_{t+1,n}, I_{t+1}=1) = pf_{t,t}f(\theta_t|Y_{t,n}, I_{t+1}=1) = pf_{t,t}f(\theta_t|Y_{t,t}, \widetilde{K}_t = t+1)$$

and the second term can be expanded as,

$$\begin{aligned}
&(1-p)f(\theta_t, y_t|Y_{t+1,n}, I_{t+1}=0) \\
&= (1-p)\sum_{j=t+1}^{n} P(\widetilde{K}_{t+1}=j+1|Y_{t+1,n}, I_{t+1}=0)f(\theta_t, y_t|Y_{t+1,n}, I_{t+1}=0, \widetilde{K}_{t+1}=j+1) \\
&= (1-p)\sum_{j=t+1}^{n} q_{t+1,j}f(y_t|Y_{t+1,n}, I_{t+1}=0, \widetilde{K}_{t+1}=j+1)f(\theta_t|Y_{t,n}, I_{t+1}=0, \widetilde{K}_{t+1}=j+1) \\
&= (1-p)\sum_{j=t+1}^{n} q_{t+1,j}\frac{f(Y_{t,j}, \widetilde{K}_t=j+1)}{f(Y_{t+1,j}, \widetilde{K}_{t+1}=j+1))}f(\theta_t|Y_{t,j}, \widetilde{K}_t=j+1) \\
&= (1-p)\sum_{j=t+1}^{n} q_{t+1,j}\frac{f_{t,j}}{f_{t+1,j}}f(\theta_t|Y_{t,j}, \widetilde{K}_t=j+1)
\end{aligned}$$

Similarly to the previous section, we compare the coefficients of $f(\theta_t|Y_{t,n}, \widetilde{K}_t = j+1)$ in (2.7) and (2.8) for $t \leqslant j \leqslant n$, as $f(\theta_t|Y_{t,n}) \propto f(\theta_t, y_t|Y_{t+1,n})$. Since $\sum_{j=t}^{n} q_{t,j} = 1$, probability $q_{t,j}$ can be computed recursively by $q_{t,j} = q^*_{t,j}/\sum_{k=t}^{n-1} q^*_{t,k}$, where

$$q^*_{t,j} = \begin{cases} pf_{t,t} & if\ j=t, \\ (1-p)q_{t+1,j}f_{t,j}/f_{t+1,j} & if\ j>t, \end{cases} \tag{2.9}$$

2.5 Semi-parametric Estimation

By the Bayes' theorem,

$$f(\theta_t|Y_{1,n}) \propto f(\theta_t|Y_{1,t})f(\theta_t|Y_{t+1,n})/f(\theta_t) \tag{2.10}$$

where $f(\theta_t)$ is the uninformative prior of θ_t. There is no need to specify $f(\theta_t)$ as it will be cancelled out when we derive the recursive formula for w_{ijt}.

An alternative way to break down $L(\theta_t; Y_{t+1,n})$ is to condition the likelihood on whether structure change takes place at time $t+1$ and if not, further decompose the likelihood based on the next possible change-point after time $t+1$. Thus the likelihood $L(\theta_t; Y_{t+1,n})$ can be written as,

$$\begin{aligned} L(\theta_t; Y_{t+1,n}) = f(\theta_t|Y_{t+1,n}) &= pf(\theta_t|Y_{t+1,n}, I_{t+1}=1) + (1-p)f(\theta_t|Y_{t+1,n}, I_{t+1}=0) \\ &= pf(\theta_t) + (1-p)\sum_{j=t+1}^{n} q_{t+1,j} f(\theta_t|Y_{t+1,j}, \widetilde{K}_{t+1} = j+1) \end{aligned} \tag{2.11}$$

Replacing $f(\theta_t|Y_{1,t})$ and $f(\theta_t|Y_{t,n})$ in (2.10) with (2.4) and (2.11),

$$\begin{aligned} L(\theta_t; Y_{1,n}) &\propto \sum_{1\leqslant i\leqslant t=j\leqslant n} pp_{it} f(\theta_t|Y_{i,t}, K_t = i) \\ &\quad + \sum_{1\leqslant i\leqslant t<j\leqslant n} (1-p)p_{it}q_{t,j+1} f(\theta_t|Y_{i,t}, K_t = i) f(\theta_t|Y_{t+1,j}, \widetilde{K}_t = j)/f(\theta_t) \\ &= \ldots + \sum_{1\leqslant i\leqslant t<j\leqslant n} (1-p)p_{it}q_{t,j+1} \frac{f_{ij} f(\theta_t, Y_{i,t}, K_t = i) f(\theta_t, Y_{t+1,j}, \widetilde{K}_t = j)}{f_{i,t} f_{t+1,j} f(\theta_t, Y_{ij}, C_{ij}) f(\theta_t)} L(\theta_t; Y_{ij}) \\ &= \sum_{1\leqslant i\leqslant t=j\leqslant n} pp_{it} L(\theta_t; Y_{ij}) + \sum_{1\leqslant i\leqslant t<j\leqslant n} (1-p)p_{it}q_{t,j+1} \frac{f_{ij}}{f_{i,t} f_{t+1,j}} L(\theta_t; Y_{ij}) \end{aligned}$$

By comparing the coefficients of the likelihood $L(\theta_t; Y_{ij})$ in the equation above with those in (2.2) and making use of $\sum_{i=1}^{t} p_{it} = 1$ and $\sum_{1\leqslant i\leqslant t\leqslant j\leqslant n} w_{ijt} = 1$, the conditional probabilities w_{ijt} can be determined by the following equations,

$$w_{ijt} \propto \ w^*_{ijt} = \begin{cases} pp_{it} & if \ i \leqslant t = j, \\ (1-p)p_{it}q_{t+1,j} f_{ij}/f_{it} f_{t+1,j} & if \ i \leqslant t < j, \end{cases} \tag{2.12}$$

where $w_{ijt} = w^*_{ijt}/P_t$ and $P_t = p + \sum_{1\leqslant i\leqslant t<j\leqslant n} w^*_{ijt}$.

As for the local likelihood mixture (2.2), we can maximize $L(\theta_t; Y_{1,n})$ by maximizing $L(\theta_t; Y_{ij})$ for $1 \leqslant i \leqslant t$ and $t \leqslant j \leqslant n$. Since $L(\theta_t; Y_{ij})$ is the likelihood with constant θ_t, classic likelihood maximization procedure can be used. We denote $\widetilde{\theta}_{ij}$ the maximum likelihood estimate of θ_t from $L(\theta_t; Y_{ij})$.

In light of Lai and Xing(2011)[31], we propose the semi-parametric estimates $\widehat{\theta}_t$, which provide estimation to the parameter vector θ_t by the following mixture,

$$\widehat{\theta}_t = \sum_{1 \leqslant i \leqslant t \leqslant j \leqslant n} w_{ijt} \widetilde{\theta}_{ij} \tag{2.13}$$

Since the evaluation of p_{it},$q_{t+1,j}$ and w_{ijt} involves the likelihood f_{ij} for Y_{ij} given C_{ij} of which θ_t is an unknown constant, we replace f_{ij} with the maximized local likelihood $L(Y_{ij}; \widetilde{\theta}_{ij})$. Denote the probabilities p_{it}, $q_{t,j}$ and w_{ijt} after such replacement by $\widetilde{p}_{it}$,$\widetilde{q}_{t,j}$ and $\widetilde{w}_{ijt}$, respectively. We can approximate (2.13) by,

$$\widehat{\theta}_t = \sum_{1 \leqslant i \leqslant t \leqslant j \leqslant n} \widetilde{w}_{ijt} \widetilde{\theta}_{ij} \tag{2.14}$$

2.6 BCMIX Approximation

To reduce the computational complexity, we use the bounded complexity mixture (BCMIX) approximation introduced by Lai and Xing(2011)[31], which includes $M(p)$ components and keeps the most recent $m(p)$ weights p_{kn} with $n - m(p) < k \leqslant n$ and $m(p) < M(p)$, to obtain the conditional probability p_{it} in (2.6). Values of $M(p)$ and $m(p)$ are presets in regard to the change-points probability p. In this dissertation, we consider to use $M(p) = 15$ and $m(p) = 10$. Let $K_{t-1}(p)$ be the set of indices i for which $p_{i,t-1}$ is kept at stage $t-1$, we have $K_{t-1}(p) \supset \{t-1, \ldots, t-m(p)\}$. At stage t, define p^*_{it} as in (2.6) for $i \in \{t\} \bigcup K_{t-1}(p)$ and we only consider the index i_t not belonging to $\{t, \ldots, t-m(p)+1\}$ such that,

$$p^*_{i_t,t} = \min \{p^*_{it} : i \in K_{t-1}(p) \; and \; i \leqslant t - m(p)\}$$

where we choose i_t to be the smallest if there exist two or more minimizers. Define $K_t(p) = \{t\} \bigcup (K_{t-1}(p) - \{i_t\})$ and let $p_{it} = p^*_{it} / \sum_{k \in K_t(p)}^{t} p^*_{kt}$ for $i \in K_t(p)$, which yields a BCMIX approximation to p_{it}.

Similarly, we can obtain a BCMIX approximation to $q_{t,j}$, i.e. $q_{t,j} = q^*_{t,j} / \sum_{k\in\widetilde{K}_t(p)}^{t} q^*_{t,k}$, with a corresponding set $\widetilde{K}_t(p)$. Let $\widetilde{K}_{t+1}(p)$ be the set of indices j for which $q_{t+1,j}$ is kept at stage $t+1$, we have $\widetilde{K}_{t+1}(p) \supset \{t+1, \ldots, t+m(p)\}$. At stage t, define $q^*_{t,j}$ as in (2.9), for $j \in \{t\} \bigcup \widetilde{K}_{t+1}(p)$ and we only consider the index j_t not belonging to $\{t, \ldots, t+m(p)-1\}$ such that,

$$q^*_{t,j_t} = \min \{q^*_{t,j} : j \in \widetilde{K}_{t+1}(p) \; and \; j \geqslant t+m(p)\}$$

where j_t is chosen to be the largest when there exist two or more minimizers. Define $\widetilde{K}_t(p) = \{t\} \bigcup (\widetilde{K}_{t+1}(p) - \{j_t\})$, which provides a BCMIX approximation to $q_{t,j}$.

Thus, the BCMIX approximation to (2.14) can be obtained by replacing $1 \leqslant i \leqslant t \leqslant j \leqslant n$ with $i \in K_t(p)$, $j \in \widetilde{K}_{t+1}(p)$, i.e.

$$\widehat{\theta}_t = \sum_{i\in K_t(p), j\in\widetilde{K}_{t+1}(p)} \widetilde{w}_{ijt}\widetilde{\theta}_{ij} \tag{2.15}$$

By implementing the BCMIX approximation, we make sure there are at most $M(p)$ components for each iteration of p_{it} and $q_{t,j}$ given time t. Therefore, our semi-parametric estimation procedure achieves linear complexity.

2.7 Use of Blocks

To further improve the computational efficiency, we consider to use blocks in our estimation procedure. Using a block of b represents that we only consider i, j and t that are multipliers of b when computing p_{it}, $q_{t,j}$ and w_{ijt}, therefore greatly reduces the computational time. Specially, using a block of 1 indicates no use of blocks. Suppose $K_{b,t}(p)$ and $\widetilde{K}_{b,t+1}(p)$ are the sets in (2.15) corresponding to the block of b, the parameter vector θ_t can be estimated by,

$$\widehat{\theta}_t = \begin{cases} \widehat{\theta}_{b\lceil t/b \rceil} & if \ b \nmid t, \\ \sum\limits_{\substack{b|t, i \in K_{b,t}(p) \\ j \in \widetilde{K}_{b,t+1}(p)}} \widetilde{w}_{ijt} \widetilde{\theta}_{ij} & if \ b \mid t, \end{cases} \tag{2.16}$$

By using the block of b, we reduce the computational time to $1/b$ at the cost of the ability to locate the exact positions of the structure changes as we treat every consecutive b points to be identical. The illustrative example in Section 4.1 compares the result with a block of 4 and that with no blocks, where it shows that, with a small b, such simplification will not deteriorate the estimation results too much. However, in consideration of the further segmentation based on our semi-parametric estimates, the result with a block of 4 slightly differs from that with no blocks. Thus, we use a block of 4 in our simulation studies to reduce the computational time and use no blocks in the real data analysis for accuracy in segmentations.

2.8 Estimation of Change-points Probability p

The semi-parametric estimate given in (2.14) involves the hyperparameter, change-points probability p for the change-time process $\{I_t\}$, which is usually not provided by the data set. Therefore, we can either choose to set an arbitrary p in our estimation procedure or estimate the change-points probability p through the following empirical Bayes approach. By the definition of p^*_{it} in (2.6), the conditional probability of y_t given $Y_{1,t-1}$ can be written as,

$$f(y_t|Y_{1,t-1}) = \sum_{i=1}^{t} p^*_{it}$$

where p^*_{it} are functions of the change-points probability p. Thus the log-likelihood function given the overall series $Y_{1,n}$ of θ_t can also be expressed as a function of p,

$$l(p; Y_{1,n}) = \sum_{t=1}^{n} \log f(y_t|Y_{1,t-1}) = \sum_{t=1}^{n} \log \{\sum_{i=1}^{t} p^*_{it}\} \tag{2.17}$$

Since p^*_{it} have to be computed recursively for $1 \leqslant i \leqslant t$, direct maximization of (2.17) would be computationally expensive. Instead, we can use the expectation–maximization (EM) algorithm which provides a much simpler structure of the log-likelihood $l_c(p)$ with the complete data $\{y_t, I_t, \theta_t, 1 \leqslant t \leqslant n\}$,

$$\begin{aligned} l_c(p) &= \sum_{t=1}^{n} \{\log f(\theta_t|y_t) + \mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}} \log p(I_t = 0) + \mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}} \log p(I_t = 1)\} \\ &= -\frac{1}{2} \sum_{t=1}^{n} (\log(2\pi) + \log \sigma_t^2 + \frac{y_t^2}{\sigma_t^2}) + \sum_{t=1}^{n} \{\mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}} \log p + \mathbf{1}_{\{\theta_t \neq \theta_{t-1}\}} \log (1-p)\} \end{aligned}$$

The E-step of the EM algorithm involves computing $E(\log \sigma_t^2|Y_{ij})$ and $E(\frac{y_t^2}{\sigma_t^2}|Y_{ij})$ where it stands that $C_{ij} = \{I_i = I_{j+1} = 1,\ I_{i+1} = \ldots = I_j = 0\}$ for time t. Neither of them can be easily formulated. However, since both terms are conditioned on the last update $\widehat{p}_{old}$ in the EM algorithm, the M-step of the EM algorithm only involves the maximization of the second term in (2.18), which results in the following update formula,

$$\begin{aligned} \widehat{p}_{new} &= \frac{1}{n} \sum_{t=1}^{n} P(I_t = 1|Y_{1,n}) = \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{n} P(C_{it}|Y_{1,n}) = \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{n} w_{itt} \\ &= \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{n} w^*_{itt}/P_t = \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{n} \widehat{p}_{old} p_{it}/P_t = \frac{1}{n} \sum_{t=1}^{n} \widehat{p}_{old}/P_t \end{aligned} \tag{2.18}$$

where P_t is a function of $\widehat{p}_{old}$ given by (2.12). However, it is very time consuming for $\widehat{p}$ to converge as we need to repeat our procedure for every update. The illustrative example in Section 4.1 shows that $\widehat{p}$ does converge but converges slowly for our EM algorithm, which also has an undesired upward bias. Moreover, as long as our preset p_0 does not deviate from the true change-points probability p too much, similar results can be obtained with or without the estimation of p. Simulation studies in Section 4.2 present that the estimation results given $\widehat{p}$ with one update from (2.18) are even worse than those given the initial value $p_0 = 0.001$ when the length of the series n is small , since a larger change-points probability p we use in our semi-parametric estimation would lead to more fluctuations in the estimates.

Therefore, the bottom line is that we use a preset $p_0 = 0.001$ and $\widehat{p}$ with only one update to acquire some knowledge of the true change-points probablity p, and then compare the estimation and segmentation results to choose which one to apply, as shown in both the simulation studies and the real data analysis. Further discussion will be made in Chapter 6.

Chapter 3

Unconditional Variance Based Segmentation Procedure

3.1 Intuition

In order to provide inference on the number and locations of the structure changes, we would like to develop a segmentation procedure given the estimated parameter vector $\widehat{\theta}_t = (\widehat{\omega}_t, \widehat{\alpha}_t, \widehat{\beta}_t)'$. One simple point to start with is to consider the Euclidean distance between $\widehat{\theta}_t$ and $\widehat{\theta}_{t+1}$ for $1 \leqslant t < n$, i.e.

$$\|\widehat{\theta}_t - \widehat{\theta}_{t+1}\|_2 = [(\widehat{\omega}_t - \widehat{\omega}_{t+1})^2 + (\widehat{\alpha}_t - \widehat{\alpha}_{t+1})^2 + (\widehat{\beta}_t - \widehat{\beta}_{t+1})^2]^{1/2} \tag{3.1}$$

However, it does not work well as $\widehat{\omega}_t$ varies greatly in scale with different data inputs, e.g., $\widehat{\omega}_t$ of the S&P 500 weekly log return series is about 5 times that of the S&P 500 daily log return series in the same period. Therefore, applying the Euclidean distance to structure change detection would only lead to using partial information of either $\widehat{\omega}_t$ or $\widehat{\alpha}_t$ and $\widehat{\beta}_t$. Thus we need to find another function to detect multiple structure changes, which is affected by $\widehat{\omega}_t$, $\widehat{\alpha}_t$ and $\widehat{\beta}_t$ at the same time.

3.2 Unconditional Variance

For a time-homogeneous GARCH model, the unconditional variance $E(\sigma_t^2)$ is a constant value to which the square of volatility would revert in the long term. Therefore, we can use the unconditional variance as a signal to divide the series into segments with constant parameters. $E(\sigma_t^2)$ can be written as a function of θ_t,

$$E(\sigma_t^2) = \omega_t/(1-\alpha_t-\beta_t) \tag{3.2}$$

Since we pay more attention to the relative changes of the unconditional variances rather than the absolute changes, we would take $log(E(\sigma_{t+1}^2)/E(\sigma_t^2))$ into consideration, which can be approximated by the first order Taylor expansion as follows,

$$\begin{aligned}
&log(E(\sigma_{t+1}^2)/E(\sigma_t^2)) \\
&= log(E(\sigma_{t+1}^2)) - log(E(\sigma_t^2)) \\
&= \frac{1}{E(\sigma_t^2)}\{\frac{\partial E(\sigma_t^2)}{\partial \omega_t}(\omega_{t+1}-\omega_t) + \frac{\partial E(\sigma_t^2)}{\partial \alpha_t}(\alpha_{t+1}-\alpha_t) + \frac{\partial E(\sigma_t^2)}{\partial \beta_t}(\beta_{t+1}-\beta_t)\} \\
&= \frac{1-\alpha_t-\beta_t}{\omega_t}\{\frac{1}{1-\alpha_t-\beta_t}(\omega_{t+1}-\omega_t) + \frac{\omega_t}{(1-\alpha_t-\beta_t)^2}[(\alpha_{t+1}-\alpha_t)+(\beta_{t+1}-\beta_t)]\} \\
&= (\omega_{t+1}-\omega_t)/\omega_t - [(\alpha_{t+1}-\alpha_t)+(\beta_{t+1}-\beta_t)]/(1-\alpha_t-\beta_t)
\end{aligned} \tag{3.3}$$

If we further replace the parameters in (3.3) with our estimated parameter vector $\widehat{\theta}_t$, we can estimate $log(E(\sigma_{t+1}^2)/E(\sigma_t^2))$ by,

$$log(E(\sigma_{t+1}^2)/E(\sigma_t^2)) = (\widehat{\omega}_{t+1}-\widehat{\omega}_t)/\widehat{\omega}_t + [(\widehat{\alpha}_{t+1}-\widehat{\alpha}_t)+(\widehat{\beta}_{t+1}-\widehat{\beta}_t)]/(1-\widehat{\alpha}_t-\widehat{\beta}_t) \tag{3.4}$$

3.3 Δ_t for Detecting Multiple Change-points

Before we propose a function Δ_t for detecting multiple change-points, two notations are introduced for our segmentation procedure, which are the minimum possible distance

between change-points m and the maximum possible number of change-points M. The minimum possible distance m is a preset value which reflects how separately we believe the structure changes would take place or the least interval we need to separate two adjacent change-points given the specified model. It is easily understood that we cannot separate two adjacent change-points if they are too close to each other. In the extreme case, if two change-points are exactly next to each other, we do not even have the subsequence in between to provide inference. The numbers of points before the first structure change and after the last structure change are also governed by m. Our simulation studies and real data analysis suggest that we may consider to leave an even large space before the first structure change and after the last structure change. Further discussion will be made in Chapter 6. Hwang and Pereira(2006)[32] advises that at least 500 observations are required for GARCH(1,1) models in consideration of the size of biases and convergence errors. In this dissertation, we use $m = 100$ or 200 as we focus more on the structure change detection than the parameter estimation. Later, we will show that m can also be chosen by the length of the series n given our modified Bayesian information criterion (MBIC). Meanwhile, the maximum possible number M provides a limit for the number of change-points, which is given by,

$$M = \lfloor n/m \rfloor - 1 \tag{3.5}$$

As we focus on the individual impact of $\widehat{\omega}_t$, $\widehat{\alpha}_t$ and $\widehat{\beta}_t$ over $log(E(\sigma^2_{t+1})/E(\sigma^2_t))$ without neglecting each other, we propose the following function Δ_t for detecting multiple change-points in our segmentation procedure,

$$\Delta_t = |\widehat{\omega}_{t+1} - \widehat{\omega}_t|/\widehat{\omega}_t + (|\widehat{\alpha}_{t+1} - \widehat{\alpha}_t| + |\widehat{\beta}_{t+1} - \widehat{\beta}_t|)/(1 - \widehat{\alpha}_t - \widehat{\beta}_t) \tag{3.6}$$

where we take absolute values of each term in the right-hand side of (3.4) so as to make sure each individual impact to be positive.

Lai and Xing(2011)[31] suggests the use of "bandwidth" d by replacing the estimated parameters by their local average within a width of d in order to make Δ_t smooth. For instance, we replace $\widehat{\omega}_t$ in the denominator in (3.6) by $\sum_{i=t-d+1}^{t+d}\widehat{\omega}_i/2d$ and $\widehat{\omega}_{t+1}$ and $\widehat{\omega}_t$ in the numerator by $\sum_{i=t+1}^{t+d}\widehat{\omega}_i/d$ and $\sum_{i=t-d+1}^{t}\widehat{\omega}_i/d$ respectively. Such replacement can mitigate the impact of extreme values in $\widehat{\omega}_t$, $\widehat{\alpha}_t$ and $\widehat{\beta}_t$ over Δ_t. The d we choose should be smaller than $m/2$, which promises that there exists at most one change-point within the width. In this dissertation, we use $d = 40$.

Thus, for $m < t \leqslant n - m$, Δ_t after smoothing can be written as,

$$\Delta_t = \frac{2|(\sum_{i=t+1}^{t+d} - \sum_{i=t-d+1}^{t})\widehat{\omega}_i|}{\sum_{i=t-d+1}^{t+d}\widehat{\omega}_i} + \frac{2|(\sum_{i=t+1}^{t+d} - \sum_{i=t-d+1}^{t})\widehat{\alpha}_i| + 2|(\sum_{i=t+1}^{t+d} - \sum_{i=t-d+1}^{t})\widehat{\beta}_i|}{2d - \sum_{i=t-d+1}^{t+d}\widehat{\alpha}_i - \sum_{i=t-d+1}^{t+d}\widehat{\beta}_i} \tag{3.7}$$

We can locate the first possible change-point by choosing t with the largest Δ_t, which is denoted by l_1. As for the second possible change-point, noting that it should be at least m points away from l_1, we can find it by maximizing Δ_t excluding the local region of l_1 with a width of m, i.e. $l_2 = \arg\max_{m<t\leqslant n-m}\{\Delta_t : t \notin \{l_1 - m < t < l_1 + m\}\}$. Similarly, the ith possible change-point can be determined by,

$$l_i = \arg\max_{m<t\leqslant n-m}\{\Delta_t : t \notin \bigcup_{j=1}^{i-1}\{l_j - m < t < l_j + m\}\} \tag{3.8}$$

We keep the process above running until the number of possible change-points reaches K, such that $\{m < t \leqslant n - m\}/\bigcup_{j=1}^{K}\{l_j - m < t < l_j + m\} = \emptyset$, which indicates we cannot find more possible change-points after that. Such K is smaller or equal to the maximum maximum possible number M. Thus, we have obtained K possible change-points $\{l_1, \ldots, l_K\}$, which can be used to establish K time-varying models each with a different number of structure changes from 1 to K, i.e. we choose the first k possible change-points $\{l_1, \ldots, l_k\}$ as the change-points for the kth model for $1 \leqslant k \leqslant K$. By applying model selection to these K models and the time-homogeneous model, which is corresponding to choose the number of change-points, we can determine the number and

locations of the structure changes. In the following section, we will develop a modified Bayesian information criterion (MBIC) to choose from the $(K+1)$ models, which would be further improved by a top down approach.

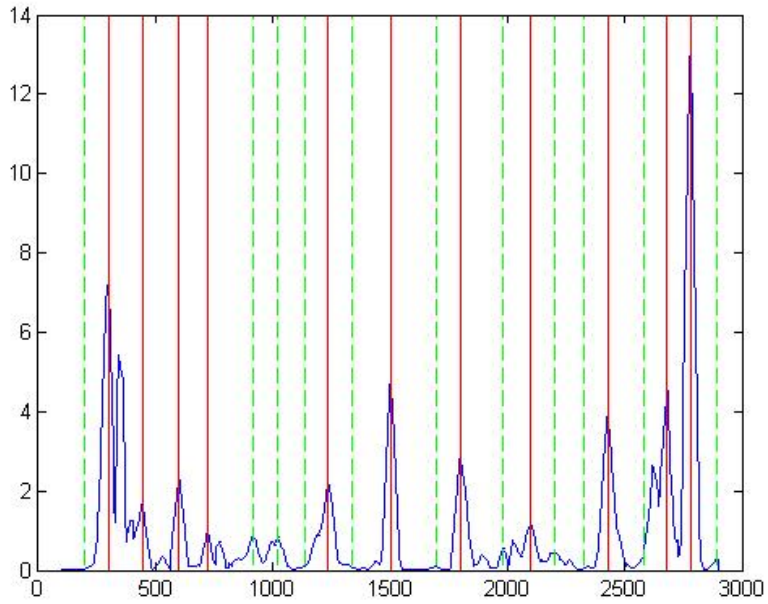

Figure 3.1: Illustration: detecting possible change-points by Δ_t with $m = 100$

Figure 3.1 shows the 22 possible change-points detected by Δ_t with $m = 100$ in the illustration example in Section 4.1. The solid lines represent the first 11 possible change-points while the dash lines represent the rest 11 possible change-points. We can see that possible change-points are detected in succession according to their scale in Δ_t with the minimum possible distance m in between.

3.4 Modified Bayesian Information Criterion

A Bayesian information criterion (BIC) for model selection among a finite set of models is given by,

$$BIC = -2l(\widehat{\theta}_t) + r\log(n) \tag{3.9}$$

where $l(\widehat{\theta}_t)$ is the maximized value of the log-likelihood function of the model, r is the number of free parameters to be estimated and n is the number of data points in the observed data. In our case, $r = 4k$, where k is the number of structure changes. The intuition behind this is that for each change-point added, we introduce a parameter for the location of the new change-point as well as a parameter vector θ_t for the additional segment which contains three individual parameters ω_t, α_t and β_t shifting simultaneously.

Lavielle(2005)[33] proposed to replace the penalty term in (3.9) by $\delta r \log(n)$ for change-points problems, where δ is a shrinkage parameter chosen by the user. In our case, we choose $\delta = \frac{1}{2}$, since the change-points are mostly determined by the local estimates and the piecewise constant parameter vector θ_t can be estimated by the subsequences between adjacent structure changes. Thus, the Bayesian information criterion (BIC) for our model selection can be written as,

$$BIC = -2l(\widehat{\theta}_t) + 2k \log(n) \tag{3.10}$$

Zhang and Siegmund(2007)[29] proposed a modified Bayesian information criterion (MBIC) for Gaussian change-point models, of which the penalty term is in the form of $3k \log(n) + \sum_{i=1}^{k+1} \log(l_i - l_{i-1})$, where $l_0 = 0$, $l_{k+1} = n$ and $\{l_1, \ldots, l_k\}$ are the total k change-points. Since $\log(\cdot)$ is a concave function, the penalty term is maximized when the change-points are evenly distributed and minimized when the distances between adjacent change-points reach the minimum possible distance m. Such property well responses to the fact that the maximum log-likelihood is much improved when a change-point is detected in the middle of a subsequence than close to the edges.

In light of this, we propose our modified Bayesian information criterion (MBIC),

$$MBIC = -2l(\widehat{\theta}_t) + 2k \log(n) + \frac{2k}{k+2} \log(m)(\sum_{i=1}^{k+1} \log(l_i - l_{i-1}) - (k+1) \log(n/(k+1))) \tag{3.11}$$

In our MBIC, we consider the following adjustment to the penalty term in (3.10),

$$\sum_{i=1}^{k+1} \log(l_i - l_{i-1}) - (k+1)\log(n/(k+1)) \tag{3.12}$$

which reaches its maximum 0 when the change-points are evenly distributed. Thus the new penalty term is separated into $2k\log(n)$, which is solely based on the number of change-points, and $\frac{2k}{k+2}\log(m)(\sum_{i=1}^{k+1}\log(l_i - l_{i-1}) - (k+1)\log(n/(k+1)))$, which is mainly based on where these structure changes take place. The reason why we put a coefficient $\frac{2k}{k+2}\log(m)$ before (3.12) is that structure changes tend to spread evenly with a large minimum possible distance m. Also, extreme values of (3.12) are more frequently reached with a small number of change-points k.

So as to make our MBIC work, we need to make sure that the extra penalty for an additional change-point is always positive. In the extreme case, (3.12) reaches its minimum $k\log(m) + \log(n - km) - (k+1)\log(n/(k+1))$ when the change-points are as close together as possible with the minimum possible distance m in between.

	n=1000		n=3000		n=5000		n=7000		n=9000	
k	min	max	min	max	min	max	min	max	min	max
1	10.68	13.82	9.72	16.01	9.22	17.03	8.88	17.71	8.62	18.21
2	20.57	27.63	15.56	32.03	13.03	34.07	11.33	35.42	10.05	36.42
3	31.95	41.45	21.71	48.04	16.55	51.10	13.09	53.12	10.49	54.63
4	44.98	55.26	29.05	64.05	20.95	68.14	15.53	70.83	11.45	72.84
5	59.50	69.08	37.71	80.06	26.52	85.17	19.03	88.54	13.38	91.05
6	75.22	82.89	47.66	96.08	33.28	102.21	23.63	106.24	16.38	109.26
7	91.79	96.71	58.80	112.09	41.16	119.24	29.33	123.95	20.42	127.47
8	108.64	110.52	71.04	128.10	50.10	136.28	36.04	141.66	25.46	145.68
9	124.34	124.34	84.27	144.11	60.01	153.31	43.71	159.37	31.43	163.89

Table 3.1: MBIC: range of penalty given the number of change-points k and the length of the series n with $m = 100$

Table 3.2: MBIC: range of penalty given the number of change-points k and the length of the series n with $m = 200$

	n=2000		n=4000		n=6000		n=8000		n=10000	
k	min	max	min	max	min	max	min	max	min	max
1	11.59	15.20	10.72	16.59	10.16	17.40	9.75	17.97	9.43	18.42
2	22.28	30.40	18.34	33.18	15.85	34.80	14.05	35.95	12.63	36.84
3	34.67	45.61	26.85	49.76	21.91	52.20	18.32	53.92	15.51	55.26
4	48.98	60.81	36.97	66.35	29.32	69.60	23.76	71.90	19.40	73.68
5	64.99	76.01	48.76	82.94	38.27	87.00	30.63	89.87	24.63	92.10
6	82.39	91.21	62.10	99.53	48.69	104.39	38.91	107.85	31.22	110.52
7	100.75	106.41	76.84	116.12	60.49	121.79	48.52	125.82	39.11	128.94
8	119.45	121.61	92.85	132.70	73.54	139.19	59.37	143.80	48.22	147.37
9	136.82	136.82	109.99	149.29	87.74	156.59	71.35	161.77	58.45	165.79

Table 3.1 and 3.2 give the ranges of the penalties with $m = 100$ and 200 respectively. In both tables, the minimum and the maximum of the penalties increase with the number of change-points k with different lengths of series n, which indicates that our MBIC works in these scenarios. This also provides us an idea to choose m for a given series with length n, since a table like these can always be created for a specific m to check whether it is suitable for the corresponding n and different levels of k. Moreover, we can even modify the coefficients in the penalty term of our MBIC and make the penalties best satisfy our interest. Further discussion will be made in Chapter 6.

Given our MBIC, model selection is made among the $(K + 1)$ models in the previous section. If the time-homogeneous model is selected, we claim that there exist no structure changes. Otherwise, suppose the kth model is selected with the minimum MBIC, which contains k possible change-points $\{l_1, \ldots, l_k\}$, we can perform the following top down approach to determine the number and locations of the structure changes.

3.5 Top Down Approach

The k possible change-points $\{l_1, \ldots, l_k\}$ might involve some misleading points which does not improve the overall log-likelihood as much as it is supposed to by dividing the series into segments. This is due to the estimation errors of our estimate $\widehat{\theta}_t$ as well as the function Δ_t and our previous rules for determine change-points may not perfectly match the characteristic of the true structure changes. Therefore, we consider to use the conventional top down approach in model selection to eliminate these misleading points, which helps to improve our segmentation result. Given the k possible change-points $\{l_1, \ldots, l_k\}$, we compute the MBIC for the rest possible change-points without l_i for $1 \leqslant i \leqslant k$ and choose l_i with the smallest MBIC. If the result improves the MBIC for $\{l_1, \ldots, l_k\}$ by at least $log(n)$, i.e. $\Delta MBIC > log(n)$, we remove l_i and the rest $k-1$ points make up the new set of all possible change-points. The reason why we set up a threshold $log(n)$ for $\Delta MBIC$ is that we still have some faith for our initial order of possible change-points and the penalty term of the MBIC is roughly proportion to $log(n)$ for each change-point added. Also, Kass and Raftery (1995)[34] suggests the strength of the evidence against the model with the higher BIC value is strong when $\Delta BIC = 6$ and very strong when $\Delta BIC = 10$, where our threshold usually falls in between as $\log(1000) = 6.9$ and $\log(10000) = 9.2$. Thus, as we continue the top down approach until it stops, we can obtain $k^\star$ points from $\{l_1, \ldots, l_k\}$, which are the structure changes detected by our procedure.

3.6 Segmentation Procedure

In this section, we summarize our unconditional variance based (UVB) segmentation procedure in the previous sections, which include a function Δ_t for detecting possible change-points, the modified Bayesian information criterion (MBIC) for model selection and a top down approach for improvement. Given the semi-parametric estimate $\widehat{\theta}_t$, we can determine the number and locations of structure changes in the following steps,

Unconditional variance based segmentation procedure

1. Compute Δ_t in (3.7) and choose t with the largest Δ_t as the first possible change-point. By maximizing Δ_t at least m points away from the selected change-points, the rest possible change-points can be determined sequentially by (3.8). Establish K time-varying models each with the first k possible change-points $\{l_1, \ldots, l_k\}$ for $1 \leqslant k \leqslant K$ and combine them with the time-homogeneous model.

2. Compute the modified Bayesian information criterion (MBIC) in (3.11) for each of the $(K+1)$ models given by step 1 and choose the model with the smallest MBIC. There are three possibilities,

   (a) If the time-homogeneous model is selected, then there are no structure changes and the segmentation process is terminated.

   (b) If the model selected has only one change-point, then there exists a unique structure change as presented in the model and the segmentation process is terminated.

   (c) If the model selected has multiple change-points, suppose it has k structure changes, then $\{l_1, \ldots, l_k\}$ are our possible change-points, which can be further improved by step 3.

3. Compute the MBIC for the rest possible change-points in $\{l_1, \ldots, l_k\}$ without l_i for $1 \leqslant i \leqslant k$ and choose l_i with the smallest MBIC, where $\{l_1, \ldots, l_{i-1}, l_{i+1}, \ldots, l_k\}$ are the structure changes in the new model. Compute $\Delta MBIC$ as the difference of our MBIC between the old model, which is initially given by step 2 and the new model. if $\Delta MBIC > log(n)$, update the old model with the new model and then repeat this step. Otherwise, terminate the process and label the change-points in the old model as the structure changes detected. Suppose there exist $k^\star$ change-points, then $\{l_1^\star, \ldots, l_{k^\star}^\star\}$ are our structure changes.

3.7 Comparing to Binary Segmentation Algorithm

The parameter estimation of the time-homogeneous GARCH(1,1) model can be achieved by maximizing the quasi-loglikelihood function,

$$l(\theta) = \sum_{t=1}^{n} l_t(\theta) = -\frac{1}{2}\sum_{t=1}^{n}(\log(2\pi) + \log\sigma_t^2 + \frac{y_t^2}{\sigma_t^2}) \tag{3.13}$$

Let $S_t(\widehat{\theta})$ and $H_t(\widehat{\theta})$ be the score function and the Hessian matrix of $l_t(\theta)$ evaluated at $\widehat{\theta}$, which is the quasi-maximum likelihood estimate of the true parameter vector θ. Though $H_t(\widehat{\theta})$ has a more complicated form, $E(H_t(\widehat{\theta}))$ can be easily written as,

$$\begin{aligned}
E(H_t(\widehat{\theta})) &= E(S_t(\widehat{\theta})S_t(\widehat{\theta})') \\
&= E((\frac{y_t^2-\sigma_t^2}{2\sigma_t^4})^2 \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]' \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]) \\
&= \frac{1}{4\sigma_t^4}E(\frac{y_t^2}{\sigma_t^2}-1)^2 \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]' \cdot [1, y_{t-1}^2, \sigma_{t-1}^2] \\
&= \frac{1}{2\sigma_t^4} \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]' \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]
\end{aligned}$$

where $\frac{y_t^2}{\sigma_t^2}$ follows the chi-square distribution with degree of freedom 1.

Galeano and Tsay(2010)[27] proposed an binary segmentation algorithm in the case where structure changes take place in individual parameters of a GARCH model. It makes use of the LM test statistics introduced by Andrews(1993)[5],

$$LM(\nu) = \frac{n}{\nu(1-\nu)}\bar{l}_\nu(\widehat{\theta})'B_n^{-1}A_n(A_n'B_n^{-1}A_n)^{-1}A_n'B_n^{-1}\bar{l}_\nu(\widehat{\theta}) \tag{3.14}$$

where ν is the relative position of the structure change to be tested, which is govern by $\nu_{\min}$ and $\nu_{\max} = 1 - \nu_{\min} \in (0,1)$ to avoid the situation where the LM test for detecting single change point is made too close to the edges of the series. $\bar{l}_\nu(\widehat{\theta})$, A_n and B_n are denoted by the following equations,

Table 3.3: LM: asymptotic critical values from Andrew(1993)

	q=1			q=2			q=3		
$\nu_{\min}$	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	2.71	3.84	6.63	4.61	5.99	9.21	6.25	7.81	11.34
.45	4.38	5.91	9.00	6.60	8.11	11.77	8.50	10.15	14.23
.40	5.10	6.57	9.82	7.45	9.02	12.91	9.46	11.17	14.88
.35	5.59	7.05	10.53	8.06	9.67	13.53	10.16	12.05	15.71
.30	6.05	7.51	10.91	8.57	10.19	14.16	10.76	12.58	16.24
.25	6.46	7.93	11.48	9.10	10.75	14.47	11.29	13.16	16.60
.20	6.80	8.45	11.69	9.59	11.26	15.09	11.80	13.69	17.28
.15	7.17	8.85	12.35	10.01	11.79	15.51	12.27	14.15	17.68
.10	7.63	9.31	12.69	10.50	12.27	16.04	12.81	14.62	18.28
.05	8.19	9.84	13.01	11.20	12.93	16.44	13.47	15.15	19.06

$$\bar{l}_\nu(\widehat{\theta}) = \frac{1}{n}\sum_{t=1}^{[\nu n]} S_t(\widehat{\theta}) = \frac{1}{n}\sum_{t=1}^{[\nu n]} \frac{y_t^2 - \sigma_t^2}{2\sigma_t^4} \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]'$$

$$A_n = -\frac{1}{n}\sum_{t=1}^{n} E[H_t(\widehat{\theta})] = -\frac{1}{n}\sum_{t=1}^{n} \frac{1}{2\sigma_t^4} \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]' \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]$$

$$B_n = \frac{1}{n}\sum_{t=1}^{n} S_t(\widehat{\theta}) S_t(\widehat{\theta})' = -\frac{1}{n}\sum_{t=1}^{n} (\frac{y_t^2 - \sigma_t^2}{2\sigma_t^4})^2 \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]' \cdot [1, y_{t-1}^2, \sigma_{t-1}^2]$$

In their paper, as they focus on the shifts in individual parameters where only one parameter ω, α or β can change at a time, they build the following LM test statistics for $i = \omega$, α and β,

$$LM_i(\nu) = \frac{n}{\nu(1-\nu)} \frac{(\bar{l}_\nu(\widehat{\theta})' B_n^{-1} A_{i,n})^2}{A_{i,n}' B_n^{-1} A_{i,n}} \tag{3.15}$$

where $\nu_{\min} \leqslant \nu \leqslant \nu_{\max}$ and $A_{i,n}$ is the first, second and third colomn of the matrix A_n in regard to ω, α and β.

Andrew(1993)[5] provided asymptotic critical values for the LM tests with the upper and lower bounds $\nu_{\min}$ and $\nu_{\max}$ for ν, which is given by the square of a Bessel process of order q corresponding to the number of shifting parameters.

Table 3.3 shows a part of the asymptotic critical values for the LM tests demonstrated in Andrew(1993)[5]. It requires a certain minimum value $\nu_{\min}$, otherwise we fail to obtain the asymptotic critical values as the limiting distributions of the Bessel processes would not converge. When in comparison with our segmentation procedures, we use $q = 3$ for simultaneous changes in all parameters and $q = 1$ for individual parameter shifts, which is based on the definition of order q.

Thus, we modify the binary segmentation algorithm from Galeano and Tsay(2010)[27] by simplifying the procedure with the test statistics $LM(\nu)$ for simultaneous changes so as to compare with our unconditional variance based segmentation procedure. The exact steps of the modified procedure are listed as follows,

Binary segmentation algorithm modified from Galeano and Tsay(2010)

1. Compute $LM(\nu)$ in (3.14) for $\nu_{\min} \leqslant \nu \leqslant \nu_{\max}$ and choose ν with the largest $LM(\nu)$. If the test statistics $LM(\nu)$ is insignificant, then there are no structure changes and the segmentation process is terminated. Otherwise, if the test statistics $LM(\nu)$ is significant, take ν as the first possible change-point.

2. Segment the series by ν into two subsequences $\{y_1, \ldots, y_{\lfloor \nu n \rfloor}\}$ and $\{y_{\lfloor \nu n \rfloor + 1}, \ldots, y_n\}$. We take step 1 in each of the two subsequences, and repeat both steps until no extra possible change-points can be detected.

3. Suppose we have detected k possible change-points, $\{\nu_1, \ldots, \nu_k\}$ and set $\nu_0 = 0$ and $\nu_{k+1} = 0$, we can check each possible change-point ν_i by computing the test statistics $LM(\nu)$ in the interval $\nu_{i-1} < \nu \leqslant \nu_{i+1}$ for $1 \leqslant i \leqslant k$. If the test statistics $LM(\nu)$ is significant for some ν, update the ith possible change-point ν_i with ν. Otherwise, eliminate the ith possible change-point ν_i from the list. Repeat this step until the number and locations of the possible change-points remain changed, which results in the detected structure changes.

Comparing to our procedure, the binary segmentation algorithm (BSA) is particular easy to compute as the only maximum likelihood estimate $\widehat{\theta}$ required for each step is that of the overall subsequence. Therefore, the computational complexity only grows with the number of change-points, whereas ours is proportion to the length of the series n, which has been explained in Section 2.6. However, such procedure is likely to underestimate the number of structure changes as each single change-point is less significant and thus harder to detect among multiple change-points. Also, it is very tricky to choose the critical values for the LM test statistics, since $\nu_{\min}$ keeps increasing whenever the length of the subsequence shrinks in regard to a certain minimum possible distance m. Galeano and Tsay(2010)[27] suggests to use a fixed $\nu_{\min} = 0.1$. In our practice, we select the minimum $\nu_{\min}$ that satisfies both the minimum possible distance m and $\nu_{\min} \geqslant 0.1$ in each step. A corresponding critical value from Table 3.3 is used with the significance level 10%, as we want to have more power in detecting structure changes. In the following simulation studies and the real data analysis, we will compare the results of our unconditional variance based (UVB) segmentation procedure with those of the binary segmentation algorithm (BSA).

Chapter 4

Simulation Studies

4.1 Illustrative Example

We present an illustrative example of a series which consists of ten time-homogeneous GARCH(1,1) sequences with equal length of 300. The exact settings of the sequences are listed as follows,

1. $\theta_t = \{0.4, 0.2, 0.7\}$ for $1 \leqslant t \leqslant 300$;
2. $\theta_t = \{0.2, 0.6, 0.2\}$ for $301 \leqslant t \leqslant 600$;
3. $\theta_t = \{0.8, 0.3, 0.6\}$ for $601 \leqslant t \leqslant 900$;
4. $\theta_t = \{0.6, 0.5, 0.3\}$ for $901 \leqslant t \leqslant 1200$;
5. $\theta_t = \{0.4, 0.2, 0.7\}$ for $1201 \leqslant t \leqslant 1500$;
6. $\theta_t = \{0.2, 0.6, 0.2\}$ for $1501 \leqslant t \leqslant 1800$;
7. $\theta_t = \{0.8, 0.3, 0.6\}$ for $1801 \leqslant t \leqslant 2100$;
8. $\theta_t = \{0.6, 0.5, 0.3\}$ for $2101 \leqslant t \leqslant 2400$;
9. $\theta_t = \{0.4, 0.2, 0.7\}$ for $2401 \leqslant t \leqslant 2700$;
10. $\theta_t = \{0.2, 0.6, 0.2\}$ for $2701 \leqslant t \leqslant 3000$.

4.1.1 Semi-parametric Estimates

We fit the series with the time-homogeneous GARCH(1,1) model, which results in $\mu = 2.63 \times 10^{-2}$, $\omega = 8.29 \times 10^{-2}$, $\alpha = 0.694$, $\beta = 0.305$, suggesting high volatility persistence. Therefore, we perform our semi-parametric estimation over the series given a preset $p_0 = 0.001$ as the change-points probability p. In order to study the impact of the use of blocks, we also provide the estimation with a block of 4. We compare our estimation results with the "oracle" estimates, which assume the structure changes are already known.

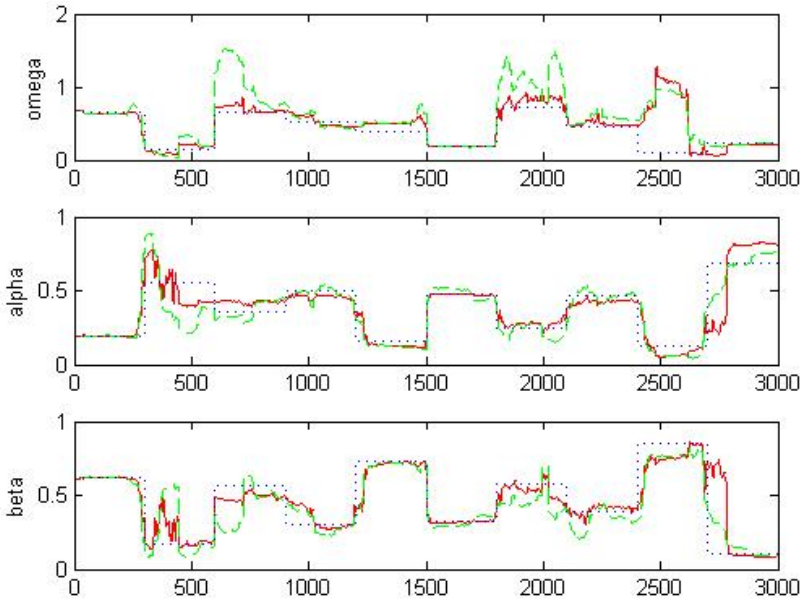

Figure 4.1: Illustration: semi-parametric estimates with no blocks and a block of 4

In Figure 4.1, the solid line represents the semi-parametric estimates with no blocks and the dash line represents those with a block of 4, where the dot line represents the "oracle" estimates. We can see both our estimates are close to the "oracle" estimates with a slight delay in shifting. However, the estimates with a block of 4 are less stable compared to the estimates with no blocks, disturbing greatly before and after the true structure changes. This is lead by simplifying the indexes i, j and t in (2.15) with only multipliers of 4, which reduces the numbers of the terms to be summed up to compute

the estimated $\widehat{\theta}$. Therefore, we only use the block of 4 in the following simulation studies to reduce the computational time but adopt no blocks for the real data analysis so as to provide better estimation.

4.1.2 Segmentation Results

Given the estimated parameters above with no blocks, we can compute Δ_t with $m = 100$ as shown in Figure 3.1 to obtain possible change-points, where we have located 22 structure changes in succession. Then, we apply the modified Bayesian information criterion to select from the corresponding models and use the top down approach to eliminate the misleading points. Our segmentation procedure results in the following structure changes, where we compare to those from the estimated parameters with a block of 4.

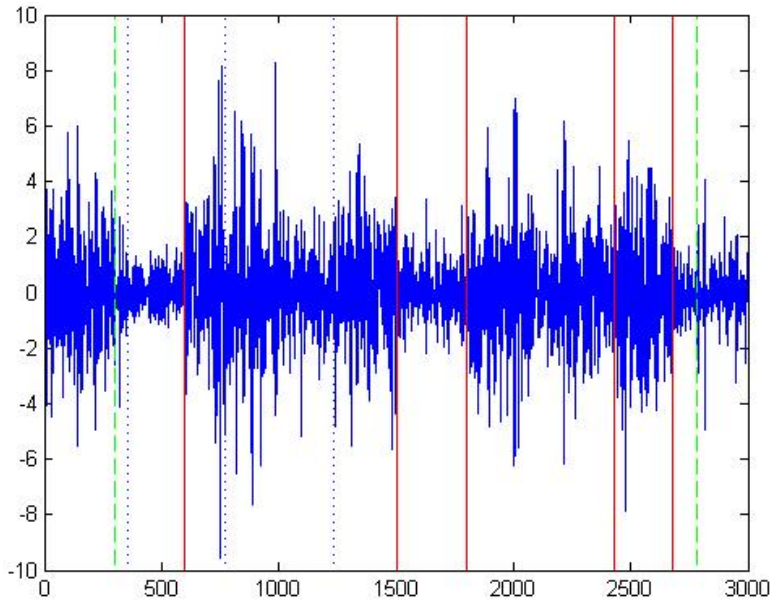

Figure 4.2: Illustration: segmentation results with no blocks and a block of 4

In Figure 4.2, the solid lines represent the structure changes detected in both settings, where the dash lines represent the rest change-points detected with no blocks and the dot lines represent those with a block of 4 only. Here we assume change-points are identical

within the range of 4. Both results fail to detect the change-points located at $t = 900$ and 2100, which is partially lead by using a small $p_0 = 0.001$ whereas the true change-points probability $p = 0.003$. Later, we will show that the number of detected structure changes is positively related to the change-points probability we use. Also, such scenario makes our MBIC not so useful as the true structure changes are evenly located.

In comparison, we implement the binary segmentation algorithm (BSA) modified from Galeano and Tsay(2010)[27] for simultaneous changes. The result is impressive as it detects 8 change-points located at $t = \{289, 595, 924, 1238, 1500, 1795, 2443, 2668\}$, which suggests our implementation is successful. However, for each change-point detected, the result from BSA deviates more from the corresponding true change-point than that from our procedure. Moreover, in the following simulation studies, we will show BSA results in much worse segmentations when the length of the series n is small and the true structure changes are more separate, where none of them are located in the middle of the sequence.

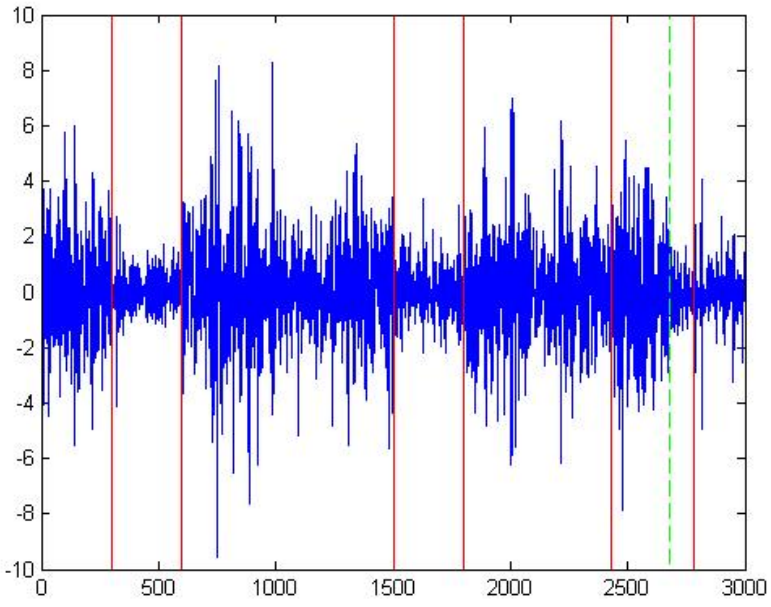

Figure 4.3: Illustration: segmentation results with $m = 100$ and 200

4.1.3 Choice of Minimum Possible Distance m

In this section, we compare the segmentation results given the different minimum possible distance m, i.e. $m = 100$ and 200. In Figure 4.3, the solid lines represent the structure changes detected in both settings, while the dash line represents the additional change-point detected with $m = 100$. In most cases, a large m would only lead to omission of the structure changes within the width m of others, as our MBIC does not change much in response to m. However, we should choose m carefully in regard to our interest and the estimation errors, especially when it turns out some of the detected structure changes are barely m points away from each other as shown in the case of $m = 100$. This is either a signal that we overestimate the number of change-points as the penalty term in our MBIC is close to its minimum in such circumstances, or suggesting that we can locate the structure changes more accurately with a smaller m.

4.1.4 Choice of Change-points Probability p

An expectation-maximization (EM) algorithm is introduced to estimate the change-points probability p in Section 2.8. In this scenario, we will perform the estimation of p with a block of 4. We are also interested in how the choice of p would affect the number of the detected structure changes.

Table 4.1: Illustration: convergence of change-points probability p and corresponding numbers of change-points k

#	1	2	3	4	5	6	7	8	9	10
$p(\times 10^{-3})$	1.0	2.9	3.6	4.1	4.4	4.6	4.7	4.8	4.9	4.9
$k(m = 100)$	8	10	9	10	10	14	14	14	14	14
$k(m = 200)$	10	8	8	8	9	9	9	9	8	8

Table 4.1 shows ten iterations of the estimated change-points probability $\widehat{p}$ and the corresponding numbers of change-points k with $m = 100$ and 200. The estimated $\widehat{p}$ does converge but converges slowly as the difference between that of the ninth iteration

and the tenth iteration is smaller than 10^{-4}. However, the estimated $\widehat{p}$ has a upward bias as the true parameter $p = 0.003$. The reason behind is that the persistence within each time-homogeneous GARCH(1,1) sequence can also be considered as if there exist additional structure changes.

As for $m = 100$, the number of change-points k generally increase with the change-points probability p. However, this is not true for $m = 200$, as k is close to the maximum number of possible change-points M. Under such circumstance, the number of change-points k is very sensitive to the choice of m, which may eliminate the key structure changes in the segmentation process. It is very suspicious to see a smaller k for $m = 100$ than for $m = 200$ as shown in Table 4.1 when $p = 0.001$. Therefore, we should make sure k is at most $\lfloor M/2 \rfloor$, which allows the detected structure changes to have a chance to be evenly distributed.

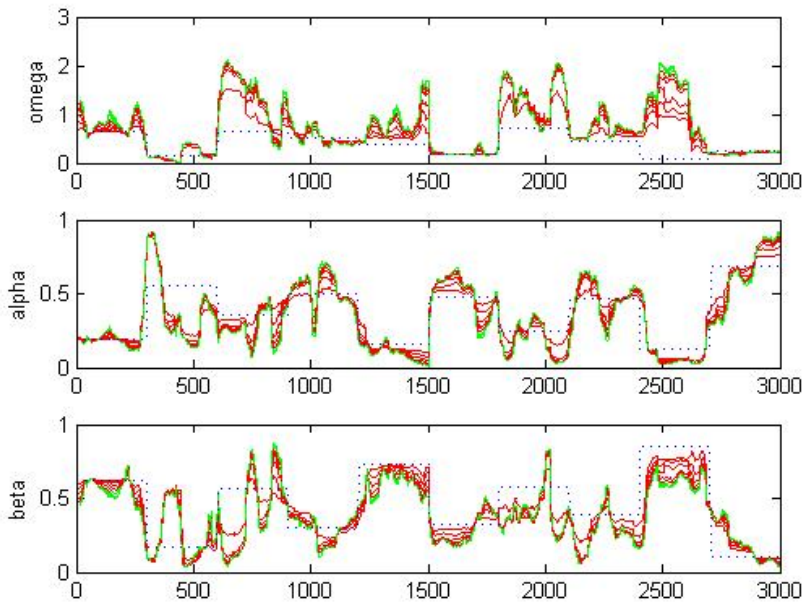

Figure 4.4: Illustration: semi-parametric estimates given different p with a block of 4

In Figure 4.4, the solid lines represent the semi-parametric estimates from the first five iterations of the estimated change-points probability $\widehat{p}$ and the dash lines represent those from the last five iterations, where the dot line represents the "oracle" estimates.

From the figure above, with the estimated $\widehat{p}$ increasing above the true change-points probability p, our semi-parametric estimates are more volatile, and thus deviate more from the piecewise constant "oracle" estimates. Therefore, choice of the change-points probability p does affect the segmentation results and structure changes are best detected when p is known. However, p is usually not given by the data set and it is time consuming to estimate p with the EM algorithm as we need to repeat our estimation procedure until p slowly converges. Moreover, the estimated $\widehat{p}$ from our EM algorithm has a upward bias, especially when then length of the series n is small. In light of this, we only consider the initial value $p_0 = 0.001$ and $\widehat{p}$ with one update in our simulation studies as well as the real data analysis. The reason we choose a small initial value $p_0 = 0.001$ is due to the competitive advantage of underestimating p over overestimating p in regard to the number of structure changes detected, as an underestimated p would only smooth the semi-parametric estimates so as to omit some insignificant change-points during the process, whereas an overestimated p would make the estimates so volatile as to involve too many misleading points in the segmentation procedure, as shown in Table 4.1. We will compare the estimation and segmentation results of the initial value $p_0 = 0.001$ and those of $\widehat{p}$ with one update in the following simulation studies.

4.2 Simulation Studies with Simultaneous Changes

We study six scenarios each including 500 series with equal length $n = 1000$, where the parameters ω_t, α_t and β_t change simultaneously. In our semi-parametric estimation, we consider the initial value $p_0 = 0.001$ and also perform the estimation for $\widehat{p}$ with one update, where a block of 4 is used in both procedures. Later in our segmentation procedure, we choose $m = 100$. The first three scenarios have fixed change-points, while structure changes take place randomly as in the time-varying GARCH(1,1) model for the last three scenarios. In the setup, we require that the "oracle" estimates $\widehat{\theta}_{oracle}$ do not deviate from the true parameter vector θ_t too much, such that $\|\widehat{\theta}_{oracle} - \widehat{\theta}_t\|_\infty \leqslant 0.2$, i.e.

$|\widehat{\omega}_{oracle} - \widehat{\omega}_t| \leqslant 0.2$, $|\widehat{\alpha}_{oracle} - \widehat{\alpha}_t| \leqslant 0.2$ and $|\widehat{\beta}_{oracle} - \widehat{\beta}_t| \leqslant 0.2$ for $1 \leqslant t \leqslant n$. Otherwise, our estimation and segmentation results would be pointless, since the "oracle" estimates are provided by the true structure changes and should be close to the true parameter vector θ_t. Moreover, without such restrictions, the sequences generated from the time-homogeneous GARCH(1,1) model might not even be stationary when the length of the sequences are small. The exact settings of the six scenarios are listed as follows,

Scenario 1. The series are generated from three time-homogeneous GARCH(1,1) models piecewisely and there exist two structure changes at $t = 301$ and 701, where $\theta_t = \{0.4, 0.2, 0.7\}$ for $1 \leqslant t \leqslant 300$, $\theta_t = \{0.2, 0.6, 0.2\}$ for $301 \leqslant t \leqslant 700$, and $\theta_t = \{0.8, 0.3, 0.6\}$ for $701 \leqslant t \leqslant 1000$.

Scenario 2. The series are generated from four time-homogeneous GARCH(1,1) models piecewisely and there exist three structure changes at $t = 201$, 501 and 701, where $\theta_t = \{0.4, 0.2, 0.7\}$ for $1 \leqslant t \leqslant 200$, $\theta_t = \{0.2, 0.6, 0.2\}$ for $201 \leqslant t \leqslant 500$, $\theta_t = \{0.8, 0.3, 0.6\}$ for $501 \leqslant t \leqslant 700$, and $\theta_t = \{0.6, 0.5, 0.3\}$ for $701 \leqslant t \leqslant 1000$.

Scenario 3. The series are generated from one time-homogeneous GARCH(1,1) model, where $\theta_t = \{0.4, 0.2, 0.7\}$ for $1 \leqslant t \leqslant 1000$.

Scenario 4. The series are generated from the time-varying GARCH(1,1) model specified by (2.1) and (2.2), where the true change-points probability $p = 0.001$. A minimum possible distance $m = 100$ is required for the structure changes to stay away from each other. Also, a minimum jump size of 0.2 is required for each parameter ω_t, α_t and β_t to shift at the change-points.

Scenario 5. Same as Scenario 4, except that $p = 0.002$.

Scenario 6. Same as Scenario 4, except that $p = 0.005$.

4.2.1 Semi-parametric Estimates

In this section, we compare the semi-parametric estimates using the initial value $p_0 = 0.001$ and $\widehat{p}$ with one update in our estimation procedure to the "oracle" estimates, which assume the structure changes are already known. We also apply the bounded complexity mixture (BCMIX) approximation and use a block of 4 in the simulation studies to reduce the computational time. The typical estimates in the six scenarios are shown in the following figures.

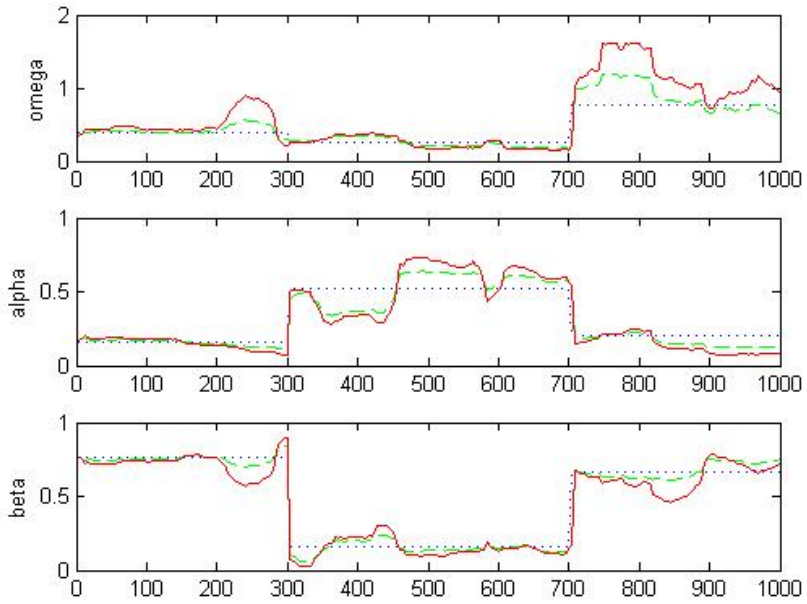

Figure 4.5: Simulation I: typical estimates in Scenario 1

Figure 4.5 shows the typical estimates in Scenario 1 with two fixed change-points at $t = 301$ and 701. The solid line represents the semi-parametric estimates using $\widehat{p}$ with one update and the dash line represents those with $p_0 = 0.001$, where the dot line represents the "oracle" estimates.

We can see both our estimates are close to the "oracle" estimates, where the estimates using $\widehat{p}$ with one update deviate more from the the "oracle" estimates. One possible explanation is that the estimated $\widehat{p}$ from our expectation-maximization (EM) algorithm tends to overestimate the true change-points probability p when the length of the series

n is small. With the initial value $p_0 = 0.001$ to start with, which is equal to the true change-points probability p, our $\widehat{p}$ with one update is already larger than p, resulting in more fluctuations in the estimates. As long as such fluctuations are smaller in scale than the true structure changes and located in the middle of the series, they will not affect our segmentation results. However, if such fluctuations are large and close to the edges of the series, they would result in misleading change-points to be detected, which will be discussed later when we analyze the segmentation results.

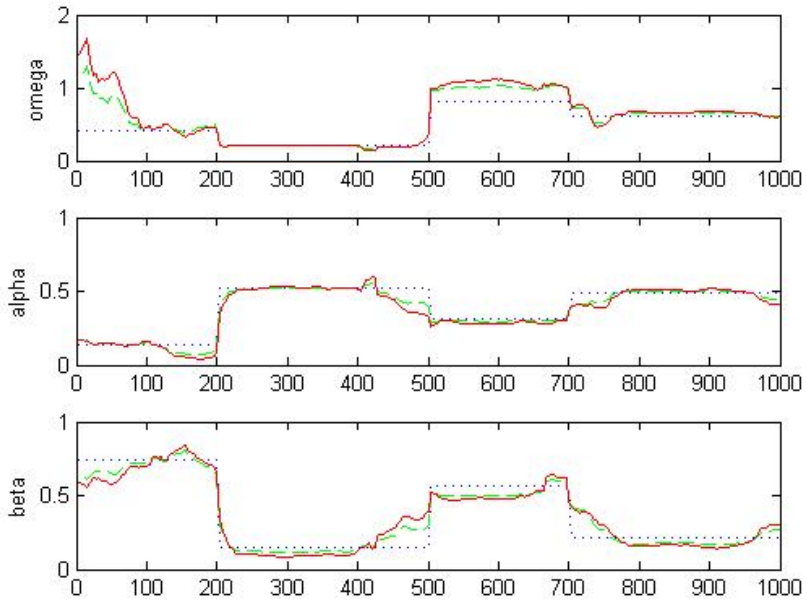

Figure 4.6: Simulation I: typical estimates in Scenario 2

Figure 4.6 shows typical estimates in Scenario 2 with three fixed change-points at $t = 201$, 501 and 701. The solid line represents the semi-parametric estimates using $\widehat{p}$ with one update and the dash line represents those with $p_0 = 0.001$, where the dot line represents the "oracle" estimates.

We can see both our estimates are close to the "oracle" estimates, where the differences are maximized at the start of the sequence. This is often true as the estimation errors are large at the beginning and the ending of the series as when we consider the local likelihood, the sample size is quite small when it is limited to both edges. Therefore we

have introduced the minimum possible distance m to also govern the interval in which the first and the last change-points are allowed when detecting structure changes. Meanwhile, our estimated parameters $\widehat{\omega}_t$, $\widehat{\alpha}_t$ and $\widehat{\beta}_t$ do not always shift at the same pace at the true structure changes, where some changes abruptly and others changes more smoothly. This is why we need to consider the impact of $\widehat{\omega}_t$, $\widehat{\alpha}_t$ and $\widehat{\beta}_t$ simultaneously, otherwise it would be difficult for us to locate the true structure changes if we only consider the estimates that shift smoothly.

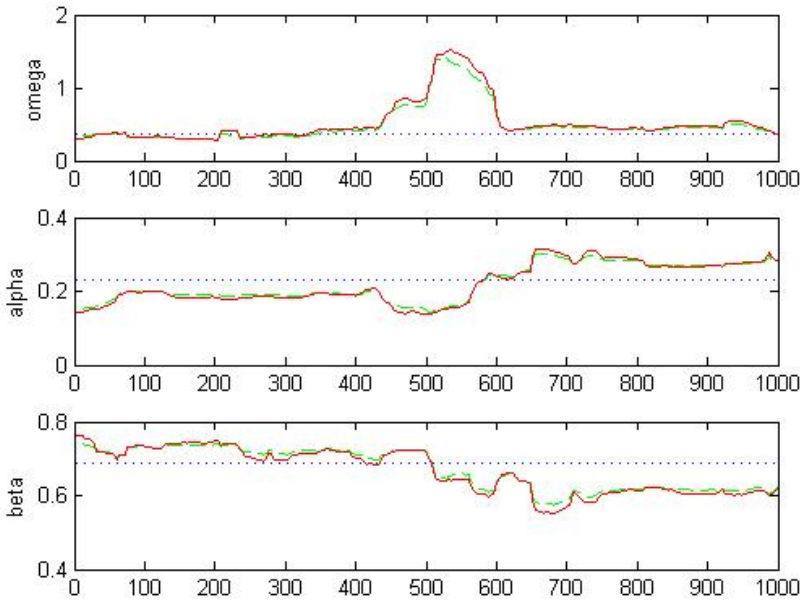

Figure 4.7: Simulation I: typical estimates in Scenario 3

Figure 4.7 shows typical estimates in Scenario 3 with no change-points at all, i.e. the series are generated from a time-homogeneous GARCH(1,1) model. The solid line represents the semi-parametric estimates using $\widehat{p}$ with one update and the dash line represents those with $p_0 = 0.001$, where the dot line represents the "oracle" estimates.

We can see both our estimates are close to the "oracle" estimates, except that there exist a large fluctuation in the middle of the sequence, which is due to the estimation errors. However, as we later apply the modified Bayesian information criterion (MBIC) to choose from the possible change-points, it is unlikely to offset the penalty to be selected

as a change-point. Even it does, it is more likely to be an outlier from the generator rather than indicating the failure of our procedure. In fact, from Table 4.3, it is very rare for our procedure to reject the true time-homogeneous model. Therefore, though we consider the MBIC in replace of the conventional BIC to detect more structure changes which are either at the edges or close to each other, the compensation paid for detecting additional structure changes is minimum.

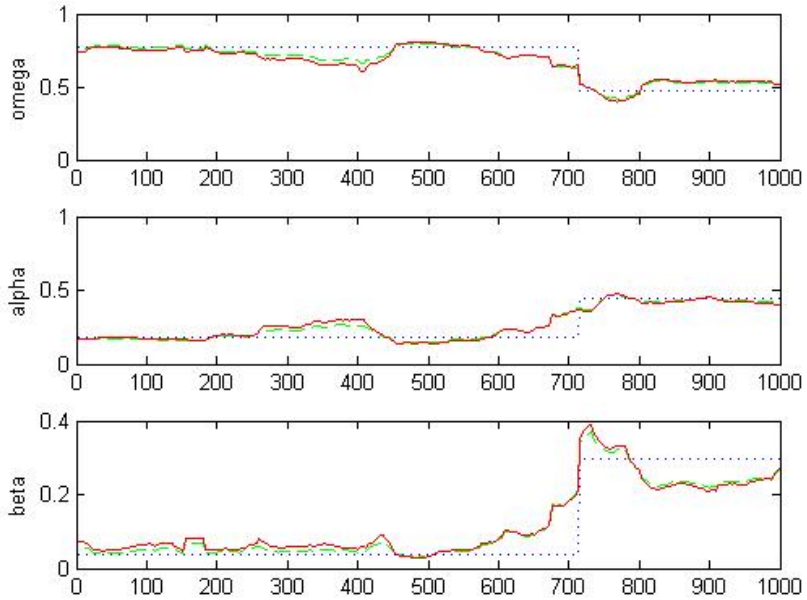

Figure 4.8: Simulation I: typical estimates in Scenario 4

Figure 4.8 shows typical estimates in Scenario 4 with random change-points which correspond to a change-points probability $p = 0.001$. The solid line represents the semi-parametric estimates using $\widehat{p}$ with one update and the dash line represents those with $p_0 = 0.001$, where the dot line represents the "oracle" estimates.

The figure presents in specific the case of single change-point, where our estimates well fit the "oracle" estimates, which indicates our procedure has enough power to detect single change-point. Though it is easily understood that our procedure must be inferior to the Wald test, the Lagrange multiplier (LM) test, and the likelihood-ratio(LR)-like test provided by Andrews(1993)[5] in detecting single change-point, as ours rely on both

the estimation and the segmentation procedures to work well. Moreover, our focus is to discover the structure changes jointly rather than making individual tests. Therefore, in the segmentation procedure, we first apply the MBIC to discover as many structure changes as possible and then use a top down approach to eliminate the misleading change-points. It is strongly recommended that not use our procedure for single change-point detection only.

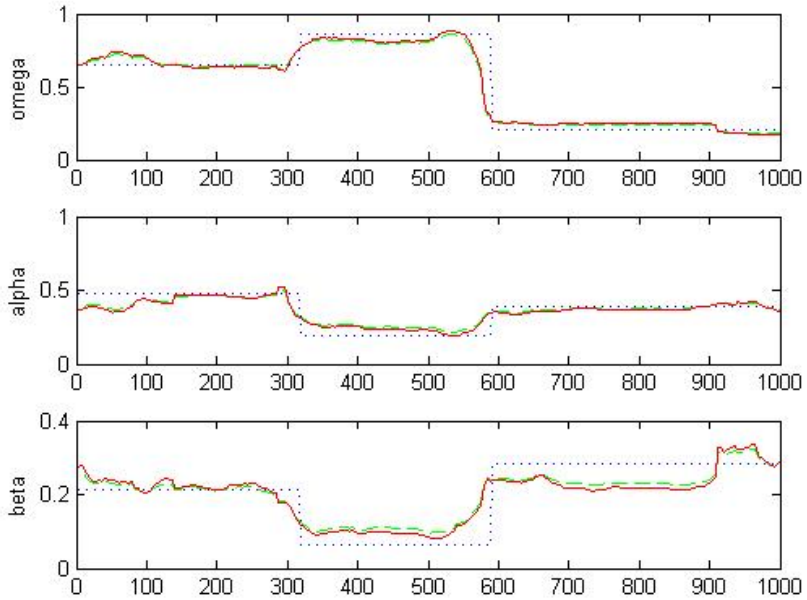

Figure 4.9: Simulation I: typical estimates in Scenario 5

Figure 4.9 shows typical estimates in Scenario 5 with random change-points which correspond to a change-points probability $p = 0.002$. The solid line represents the semi-parametric estimates using $\widehat{p}$ with one update and the dash line represents those with $p_0 = 0.001$, where the dot line represents the "oracle" estimates.

The figure presents in specific the case of two change-points, where our estimates well fit the "oracle" estimates, indicating good performance of our estimation procedure. As shown in Table 4.3, the binary segmentation algorithm (BSA) is not good enough for detecting paired structure changes, especially when they are distributed symmetrically. This is because BSA requires that either change-point to be significant to continue the

detection process but usually they are both significant or insignificant at the same time. It is never a problem for us as we can always discover the possible change-points sequentially by using a function Δ_t, which is based on the semi-parametric estimates. The only concern would be applying a proper modified Bayesian information criterion (MBIC) for choosing from the possible change-points. Further discussion will be made in Chapter 6.

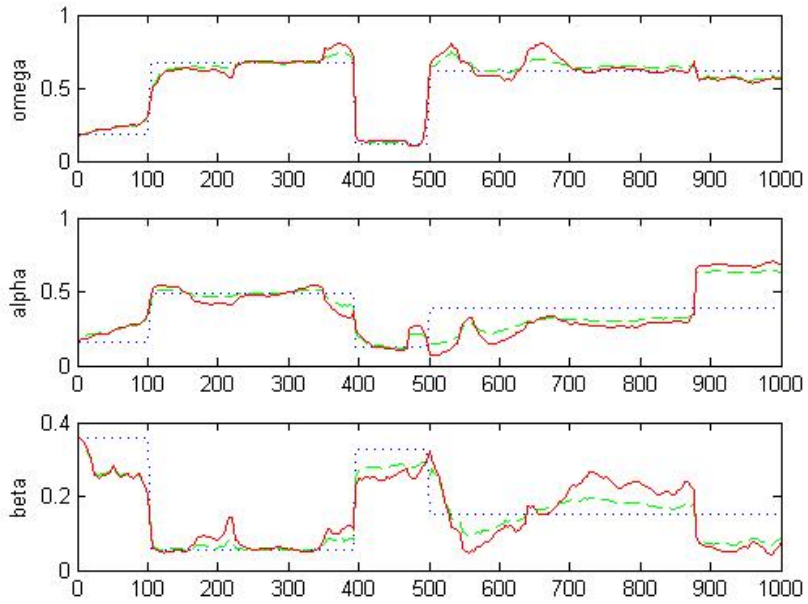

Figure 4.10: Simulation I: typical estimates in Scenario 6

Figure 4.10 shows typical estimates in Scenario 6 with random change-points which correspond to a change-points probability $p = 0.005$. The solid line represents the semi-parametric estimates using $\widehat{p}$ with one update and the dash line represents those with $p_0 = 0.001$, where the dot line represents the "oracle" estimates.

The figure presents in specific the case of three change-points, where our estimates well fit the "oracle" estimates, indicating good performance of our estimation procedure. The case is quite extreme, with one change-point located near the start of the series and the other two structure changes are barely the minimum possible distance m away. Nevertheless, our estimates are still impressive, indicating that our estimation procedure can cover most cases, no matter the structure changes are evenly distributed or extremely

separate. However, we can see there also exist a large fluctuation near the end of the sequence. Therefore, it is very difficult for us to tell whether there exist change-points close to both edges of the series. We should be very careful for this, as the penalty in our MBIC is small in such cases. Further discussion will be made in Chapter 6.

In order to compare our semi-parametric estimates with the "oracle" estimates, we evaluate the estimation errors by the mean Euclidean error (EE), the Kullback-Leibler (KL) divergence and the goodness of fit (GOF). Each of them will be introduced in the following before we reach our conclusions.

The mean Euclidean error (EE) is simply the average Euclidean distance between our semi-parametric estimates $\widehat{\theta}_t$ and the true parameter vector θ_t for $1 \leqslant t \leqslant n$, i.e.

$$EE = \frac{1}{n}\sum_{t=1}^{n} \|\widehat{\theta}_t - \theta_t\|_2 = \frac{1}{n}\sum_{t=1}^{n}[(\widehat{\omega}_t - \omega_t)^2 + (\widehat{\alpha}_t - \alpha_t)^2 + (\widehat{\beta}_t - \beta_t)^2]^{1/2} \tag{4.1}$$

In our simulation study, the mean Euclidean error properly describes how close the estimates $\widehat{\theta}_t$ are to the true parameter vector θ_t as ω_t, α_t and β_t are of the same scale, where all of them are within the range between 0.1 and 1.0.

The Kullback-Leibler (KL) divergence, or the relative entropy, measures the difference of the probability distributions given the estimates $\widehat{\theta}_t$ from the probability distributions given the true parameter vector θ_t, which describes the information lost when $\widehat{\theta}_t$ is used to approximate θ_t. It is written as,

$$KL = E_{\theta_t}(\log \frac{f(\widehat{\theta}_t)}{f(\theta_t)}) \tag{4.2}$$

Since $\log(\cdot)$ is a concave function, the Kullback-Leibler divergence is always non-negative and reaches its minimum zero only when the two probability distributions are the same almost everywhere,

$$KL = E_{\theta_t}(\log \frac{f(\widehat{\theta}_t)}{f(\theta_t)}) \geqslant \log E_{\theta_t}(\frac{f(\widehat{\theta}_t)}{f(\theta_t)}) = 0 \tag{4.3}$$

Given the log-likelihood function of the time-homogeneous GARCH(1,1) model in (3.13), we can compute the Kullback-Leibler divergence as follows,

$$\begin{aligned} KL &= E_{\theta_t}(\log f(\widehat{\theta}_t)) - E(\log f(\theta_t)) \\ &= E[-\frac{1}{2}\sum_{t=1}^{n}(\log(2\pi) + \log\widehat{\sigma}_t^2 + \frac{y_t^2}{\widehat{\sigma}_t^2})] - E[-\frac{1}{2}\sum_{t=1}^{n}(\log(2\pi) + \log\sigma_t^2 + \frac{y_t^2}{\sigma_t^2})] \\ &= E[-\frac{1}{2}\sum_{t=1}^{n}(\log\frac{\sigma_t^2}{\widehat{\sigma}_t^2} + \frac{y_t^2}{\sigma_t^2} - \frac{y_t^2}{\widehat{\sigma}_t^2})] \end{aligned} \tag{4.4}$$

where the expectation in (4.4) can be replaced by the mean value of all simulations in each scenario and σ_t^2 and $\widehat{\sigma}_t^2$ are given by the true parameter vector θ_t and the corresponding estimates $\widehat{\theta}_t$ according to (2.1),

$$\begin{cases} \sigma_1^2 = y_1^2,\ \sigma_t^2 = \omega_t + \alpha_t y_{t-1}^2 + \beta_t\sigma_{t-1}^2 \quad for\ 1 < t \leqslant n, \\ \widehat{\sigma}_1^2 = y_1^2,\ \widehat{\sigma}_t^2 = \widehat{\omega}_t + \widehat{\alpha}_t y_{t-1}^2 + \widehat{\beta}_t\widehat{\sigma}_{t-1}^2 \quad for\ 1 < t \leqslant n, \end{cases} \tag{4.5}$$

Lastly, the goodness of fit (GOF) describes how well the estimates $\widehat{\theta}_t$ fit the simulated observations. We consider the following test statistic,

$$GOF = \frac{1}{n}\sum_{t=1}^{n}\frac{y_t^2}{\widehat{\sigma}_t^2} \tag{4.6}$$

Given the true parameter vector θ_t, it stands that $\frac{y_t^2}{\sigma_t^2}$ follows the chi-square distribution with degree of freedom 1, where $y_t = \sigma_t\varepsilon_t$ and ε_t follows the standard normal distribution. Therefore, since ε_t is independent and identically distributed, $\sum_{t=1}^{n}\frac{y_t^2}{\sigma_t^2}$ follows the chi-square distribution with degree of freedom n with mean equal to n and variance equal to $2n$. Thus, should our estimates be good enough, the goodness of fit and its standard deviation of the mean will be close to 1 and $\sqrt{1/500}\cdot\sqrt{2/1000} = 0.002$ respectively.

Table 4.2 compares our estimates given $\widehat{p}$ with one update and $p_0 = 0.001$ with the "oracle" estimates in the six scenarios in regard to the mean Euclidean error (EE), the Kullback-Leibler (KL) divergence and the goodness of fit (GOF).

Table 4.2: Simulation I: comparing semi-parametric estimates with "oracle" estimates

		EE	KL	GOF
Scenario 1	$\widehat{p}$	0.4801 (0.0075)	0.0240 (4.21E-04)	1.0025 (0.0019)
	p_0	0.3445 (0.0061)	0.0176 (3.29E-04)	1.0051 (0.0018)
	oracle	0.0998 (0.0006)	0.0028 (1.29E-04)	1.0194 (0.0015)
Scenario 2	$\widehat{p}$	0.4225 (0.0060)	0.0247 (3.52E-04)	0.9981 (0.0014)
	p_0	0.3062 (0.0050)	0.0177 (2.76E-04)	1.0018 (0.0014)
	oracle	0.1014 (0.0005)	0.0027 (1.10E-04)	1.0185 (0.0014)
Scenario 3	$\widehat{p}$	0.5557 (0.0107)	0.0226 (3.72E-04)	0.9943 (0.0007)
	p_0	0.4098 (0.0085)	0.0172 (2.53E-04)	0.9948 (0.0006)
	oracle	0.0645 (0.0013)	0.0010 (4.26E-05)	1.0031 (0.0003)
Scenario 4	$\widehat{p}$	0.2275 (0.0049)	0.0142 (3.41E-04)	0.9854 (0.0006)
	p_0	0.1961 (0.0039)	0.0112 (2.36E-04)	0.9867 (0.0006)
	oracle	0.0746 (0.0012)	0.0013 (5.90E-05)	1.0020 (0.0003)
Scenario 5	$\widehat{p}$	0.2614 (0.0051)	0.0156 (3.63E-04)	0.9873 (0.0009)
	p_0	0.2196 (0.0041)	0.0121 (2.63E-04)	0.9888 (0.0008)
	oracle	0.0778 (0.0010)	0.0016 (7.56E-05)	1.0048 (0.0007)
Scenario 6	$\widehat{p}$	0.4533 (0.0167)	0.0186 (5.20E-04)	0.9908 (0.0019)
	p_0	0.2615 (0.0047)	0.0137 (4.23E-04)	0.9931 (0.0017)
	oracle	0.0822 (0.0008)	0.0022 (9.68E-05)	1.0116 (0.0012)

For each item listed above, the table presents the mean values and the corresponding standard deviations of the means, which are shown in the parentheses. In all six scenarios, the performance of our estimates given $\widehat{p}$ with one update is worse than that from $p_0 = 0.001$. The reason behind is that the estimated $\widehat{p}$ is larger than p_0 in each scenario as shown is Table 4.5. Since p_0 is close to the true change-points probability p, the larger $\widehat{p}$ would incur more fluctuations in the estimates, resulting in further deviation from the "oracle" estimates. Though the results of our estimates are significantly large than those of the oracle estimates in regard to the mean Euclidean error (EE) and the Kullback-

Leibler (KL) divergence, which is lead by the estimation errors and huge fluctuations of our estimates at the beginning and the ending of the series, the goodness of fit (GOF) of both our estimates and the "oracle" estimates is quite similar, where the differences between them and the theoretical mean value 1 are at the same scale. This is because our semi-parametric estimates are provided by the local likelihood mixture, where $\frac{y_t^2}{\sigma_t^2}$ is estimated by quasi-maximum likelihood estimation (QMLE) in each local likelihood for $1 \leqslant t \leqslant n$. However, the standard deviations of the means of the goodness of fit (GOF) are all larger than the theoretical value 0.002, only reaching minimum value 0.003 with the "oracle" estimates in Scenario 3 and 4, which indicates high estimation errors of all estimates. The differences in the mean Euclidean error (EE) and the Kullback-Leibler (KL) divergence between our estimates and the "oracle" estimates reach maximum in Scenario 3, due to $\widehat{p}$ and p_0 are both larger than the corresponding true change-points probability $p = 0$. Such differences reach minimum in Scenario 4, when p_0 is equal to the true change-points probability $p = 0.001$, indicating good performance of our estimates in the case of single change-point. In the table, we do not have the post-segmentation estimates provided by the binary segmentation procedure (BSA) in comparison, since such estimates are extremely volatile and sensitive to the detected structure changes when the length of the series n is small. Therefore, we have only got a slight idea of how good our estimates are, which will be further clarified by the segmentation results in the following section.

4.2.2 Segmentation Results

In this section, we compare the segmentation results of our unconditional variance based (UVB) segmentation procedure with those of the binary segmentation algorithm (BSA). We denote our procedure given $\widehat{p}$ with one update and the initial value $p_0 = 0.001$ by UVB($\widehat{p}$) and UVB(p_0). Also, we denote the binary segmentation algorithm for simultaneous changes modified from Galeano and Tsay(2010)[27] by BSA(S), and their own procedure for shifts in individual parameters by BSA(I). The reason why we include

BSA(I) is that, in their paper, they suggests simultaneous changes in parameters of a time-varying GARCH model can be explained by individual parameter shifts where only one parameter ω, α or β can change at a time. Moreover, we would like to compare the segmentation results between BSA(I) and our modification BSA(S) so as to select one as reference in the real data analysis. In our study, we choose a significance level 10% for BSA(S) as discussed in Section 3.7 so as to detect more structure changes. Meanwhile, we choose a significance level 5% for BSA(I) since it leads to a similar Type II error as for BSA(S) to reject the true time-homogeneous model, which will be presented in Table 4.3. The reason behind is that BSA(I) considers the maximum of the three test statistics $LM_i(\nu)$, where $i = \omega$, α and β, leading to higher significance level in the setting of simultaneous changes. The typical segmentation results in the six scenarios are shown in the following figures.

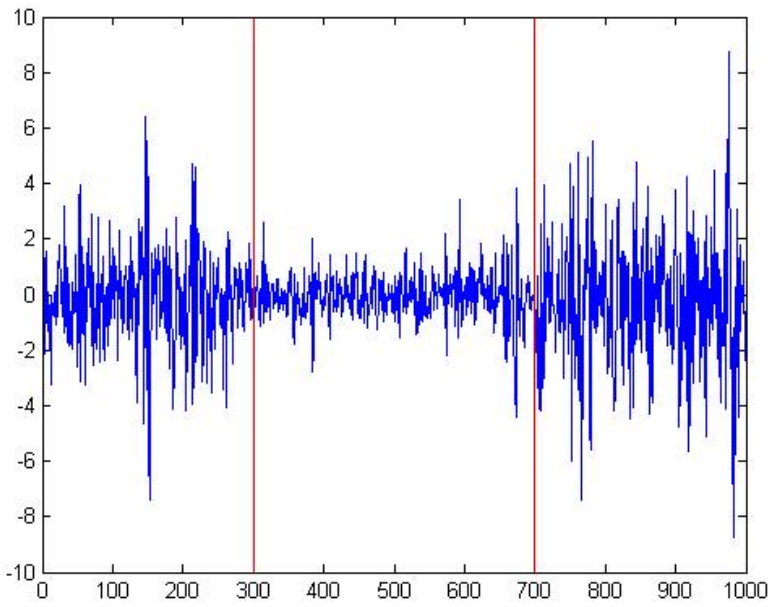

Figure 4.11: Simulation I: typical segmentation results in Scenario 1

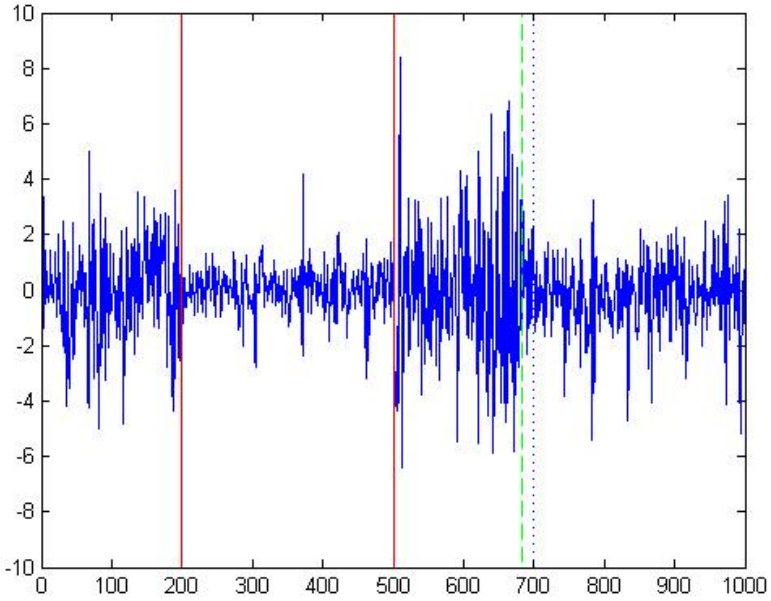

Figure 4.12: Simulation I: typical segmentation results in Scenario 2

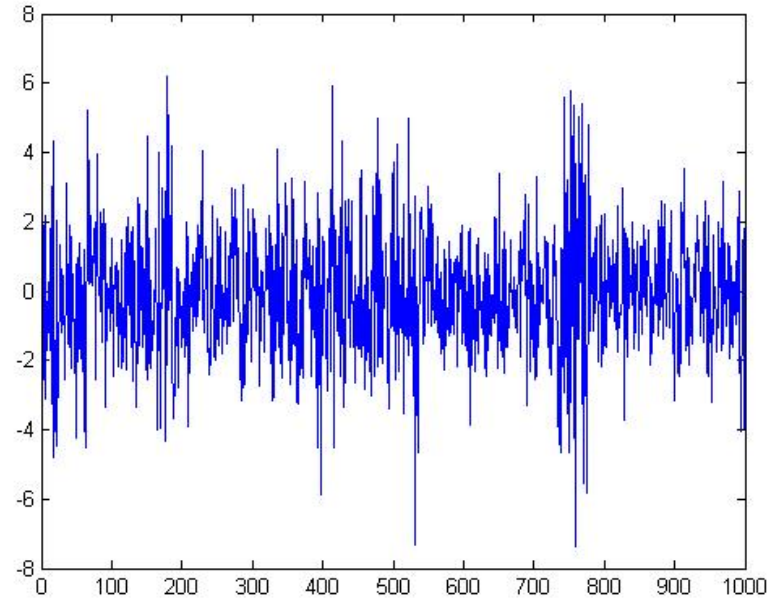

Figure 4.13: Simulation I: typical segmentation results in Scenario 3

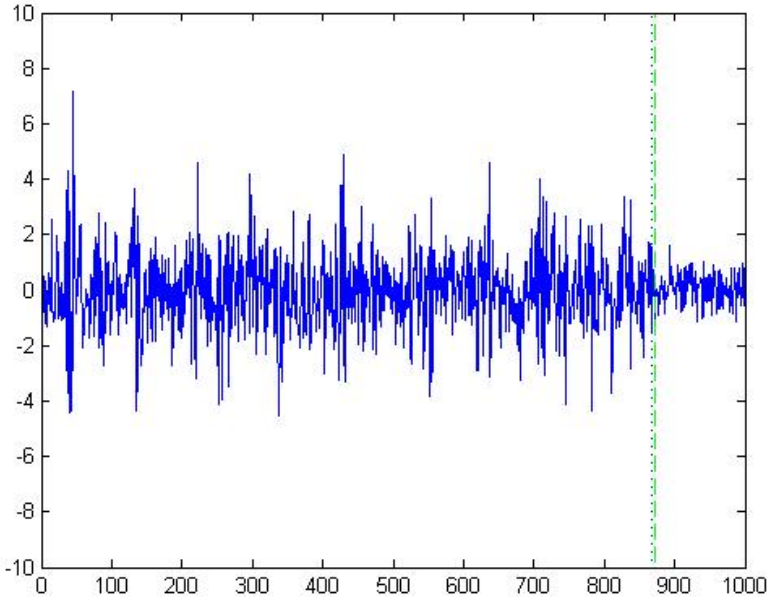

Figure 4.14: Simulation I: typical segmentation results in Scenario 4

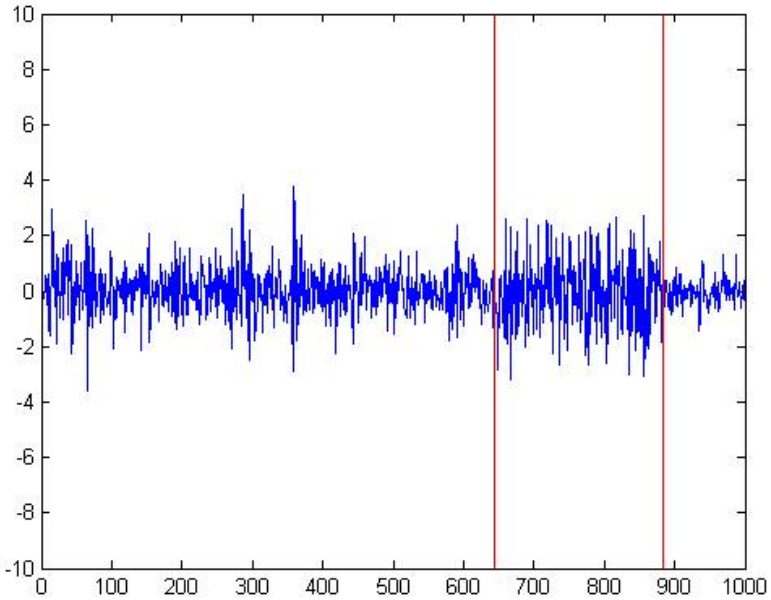

Figure 4.15: Simulation I: typical segmentation results in Scenario 5

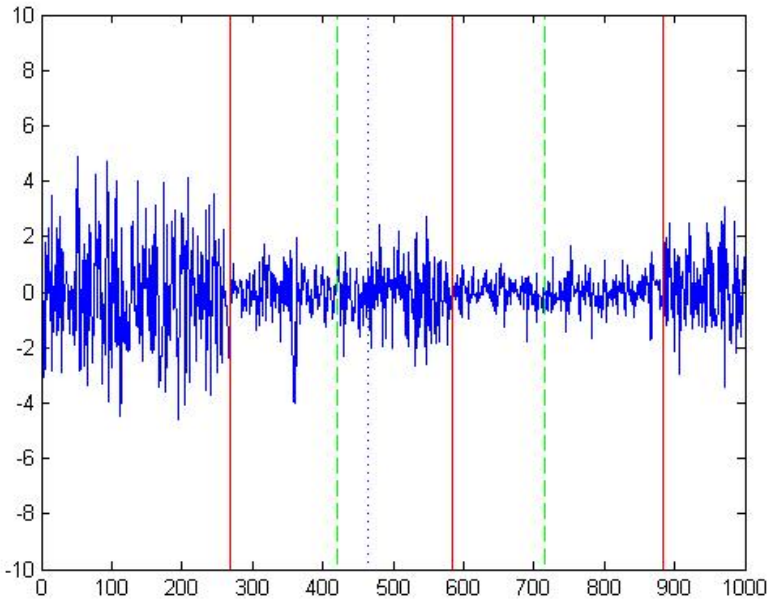

Figure 4.16: Simulation I: typical segmentation results in Scenario 6

Figure 4.11 to 4.16 demonstrate the typical segmentation results in all six scenarios of our unconditional variance based segmentation procedure (UVB) given $\widehat{p}$ with one update. The solid lines represent the detected structure changes which perfectly match the true structure changes, where the dash lines represent the rest change-points detected with our procedure and the dot lines represent the true structure changes not being detected. Here we assume change-points are identical within the range of 4. From the first five figures, our segmentations have the same number of change-points as in the actual settings, where the detected structure changes either perfectly match the true structure changes or take place extremely close to them, indicating great performance of our segmentation procedure. In Figure 4.16, we obtain an additional misleading point from our segmentation procedure, which is partially lead by a small penalty in MBIC when the misleading point is close to other detected change-points. However, it is inevitable to have certain possibility to select misleading points when the true structure changes are evenly distributed, otherwise it would be very difficult for us to detect all the true structure changes when they take place separately, which is often true for the existing methods.

Table 4.3: Simulation I: comparing segmentation results between UVB and BSA

		$\Delta = \widehat{k} - k$		$\lvert\Delta\rvert = 0$	$\lvert\Delta\rvert = 1$	$\lvert\Delta\rvert = 2$
Scenario 1 ($k = 2$)	UVB($\widehat{p}$)	0.392	(0.030)	0.686	0.244	0.054
	UVB(p_0)	0.210	(0.022)	0.812	0.162	0.022
	BSA(S)	-1.336	(0.045)	0.252	0.060	0.688
	BSA(I)	-1.264	(0.048)	0.260	0.068	0.672
Scenario 2 ($k = 3$)	UVB($\widehat{p}$)	-0.004	(0.032)	0.600	0.370	0.028
	UVB(p_0)	-0.036	(0.026)	0.722	0.262	0.014
	BSA(S)	-1.248	(0.059)	0.320	0.242	0.176
	BSA(I)	-0.638	(0.056)	0.558	0.206	0.068
Scenario 3 ($k = 0$)	UVB($\widehat{p}$)	0.044	(0.011)	0.964	0.028	0.008
	UVB(p_0)	0.056	(0.012)	0.952	0.040	0.008
	BSA(S)	0.116	(0.015)	0.890	0.104	0.006
	BSA(I)	0.122	(0.016)	0.890	0.098	0.012
Scenario 4 ($p = 0.001$)	UVB($\widehat{p}$)	-0.100	(0.028)	0.786	0.158	0.054
	UVB(p_0)	-0.094	(0.027)	0.802	0.152	0.042
	BSA(S)	-0.110	(0.029)	0.754	0.194	0.048
	BSA(I)	-0.126	(0.029)	0.764	0.178	0.054
Scenario 5 ($p = 0.002$)	UVB($\widehat{p}$)	-0.286	(0.040)	0.604	0.276	0.096
	UVB(p_0)	-0.268	(0.036)	0.662	0.236	0.090
	BSA(S)	-0.454	(0.045)	0.570	0.262	0.122
	BSA(I)	-0.506	(0.045)	0.546	0.290	0.118
Scenario 6 ($p = 0.005$)	UVB($\widehat{p}$)	-0.648	(0.053)	0.436	0.310	0.178
	UVB(p_0)	-0.700	(0.050)	0.474	0.300	0.148
	BSA(S)	-1.498	(0.062)	0.252	0.270	0.234
	BSA(I)	-1.400	(0.059)	0.262	0.306	0.220

Table 4.3 compares the segmentation results between our unconditional variance based segmentation procedure (UVB) and the binary segmentation algorithm (BSA) given the difference between the number of the detected structure changes and that of the true structure changes. In all six scenarios, our UVB is superior to BSA in terms of accuracy rate of the numbers of change-points detected. The difference of accuracy rates between UVB and BSA reaches maximum in Scenario 1, due to ineffectiveness of BSA to detect paired structure changes, especially when they are distributed symmetrically. The reason behind is that BSA may fail to detect either of the paired structure changes even though both of them are significant in the local subsequences. Such difference reaches minimum in Scenario 4, since BSA should perform at least as good as our UVB in detecting single change-point, where it degenerates to the Lagrange multiplier (LM) test. If we consider the average number of change-points detected, BSA, like other existing methods, always underestimate the number except the case there exist no structure changes, whereas UVB overestimate the number when the true structure changes are evenly distributed as in the scenarios with fixed change-points, underestimate the number when the true structure changes are more separate as in the scenarios with random change-points. Such trade-off has been made possible though the penalty term of our MBIC. As for the comparison between UVB($\widehat{p}$) and UVB(p_0), UVB(p_0) is generally superior to UVB($\widehat{p}$) since our semi-parametric estimates are more smooth with p_0. However, the difference between UVB($\widehat{p}$) and UVB(p_0) are more subtle in the scenarios with random change-points. Meanwhile, BSA(I) is generally superior to BSA(S) though they have a similar type II error as shown in Scenario 3. This is lead by the use of multiple LM tests in BSA(I), which is more versatile than applying one single LM test. The difference of the accuracy rates between BSA(I) and BSA(S) reaches maximum in Scenario 2, where the single threshold in BSA(S) is less likely to exceed, therefore terminating the segmentation process too early. Such difference reaches minimum in the scenarios with random change-points, indicating similar performances of BSA(I) and BSA(S) generally. Thus, we would use BSA(I) with a significant level 5% as the reference in the real data analysis.

Table 4.4: Simulation I: comparing ΔMBIC between UVB and BSA

		$\Delta = \Delta$MBIC	$\Delta < -\log(n)$	$\Delta > \log(n)$
Scenario 1 ($k = 2$)	UVB($\widehat{p}$)	5.172 (0.653)	0.332	0.100
	UVB(p_0)	1.636 (0.572)	0.148	0.132
	BSA(S)	46.101 (1.259)	0.920	0.008
	BSA(I)	43.214 (1.364)	0.852	0.022
Scenario 2 ($k = 3$)	UVB($\widehat{p}$)	3.731 (0.452)	0.340	0.156
	UVB(p_0)	-0.577 (0.333)	0.130	0.250
	BSA(S)	26.479 (1.069)	0.824	0.052
	BSA(I)	17.911 (1.093)	0.638	0.092
Scenario 3 ($k = 0$)	UVB($\widehat{p}$)	-0.069 (0.022)	0.000	0.004
	UVB(p_0)	-0.113 (0.029)	0.000	0.006
	BSA(S)	0.279 (0.084)	0.028	0.006
	BSA(I)	0.170 (0.069)	0.026	0.008
Scenario 4 ($p = 0.001$)	UVB($\widehat{p}$)	-0.517 (0.266)	0.076	0.146
	UVB(p_0)	-1.249 (0.191)	0.036	0.158
	BSA(S)	1.503 (0.532)	0.116	0.152
	BSA(I)	1.064 (0.452)	0.112	0.150
Scenario 5 ($p = 0.002$)	UVB($\widehat{p}$)	-1.312 (0.377)	0.116	0.254
	UVB(p_0)	-2.341 (0.344)	0.050	0.282
	BSA(S)	4.275 (0.795)	0.248	0.256
	BSA(I)	4.131 (0.818)	0.224	0.252
Scenario 6 ($p = 0.005$)	UVB($\widehat{p}$)	-4.030 (0.473)	0.144	0.420
	UVB(p_0)	-5.804 (0.409)	0.072	0.484
	BSA(S)	8.540 (1.229)	0.406	0.368
	BSA(I)	6.381 (1.150)	0.350	0.404

Table 4.4 compares the segmentation results between our unconditional variance based segmentation procedure (UVB) and the binary segmentation (BSA) given the difference between the modified Bayesian information criterion (MBIC) of the detected structure changes and that of the true structure changes. It demonstrates how close the detected structure changes are to the true structure changes, and provides us some thoughts about whether our MBIC is appropriate. The results are consistent with those in the previous table, where ΔMBIC between the detected structure changes and the true structure changes of our UVB is significantly smaller than BSA, indicating greater performance of our procedure. ΔMBIC reaches maximum for BSA in Scenario 1, where BSA usually successfully detects both the change-points or fail to detect either of them, the latter leading to a large ΔMBIC comparing to the true structure changes. Meanwhile, ΔMBIC reaches minimum for BSA in Scenario 4, where no structure changes take place. As for our segmentation procedure UVB, ΔMBIC reaches maximum in Scenario 1 whereas it reaches minimum in Scenario 6, since the penalty term in our MBIC is maximized when the structure changes are evenly distributed and minimized when the structure changes are extremely separate. The average ΔMBIC of our UVB is in the range of $-\log(n)$ and $\log(n)$, where $log(1000) = 6.9$, indicating harmony between our procedure and the modified Bayesian information criterion (MBIC), as no strong evidence exists to tell apart our detected structure changes and the true structure changes in the eye of MBIC. However, in Scenario 6, ΔMBIC has a substantial chance to be greater than $log(n)$, which is close to the accuracy rate of our UVB in the previous table, indicating we overestimate the number of structure changes due to the use of our MBIC in certain cases. However, this is a trade-off need to be made as the conventional BIC would certainly underestimate the number of structure changes as in the existing methods. Therefore, the modified Bayesian information criterion (MBIC) is appropriate for our segmentation procedure, though we wish the true structure changes would always have the smallest MBIC, which can be approached by modifying the coefficients in the penalty term of our MBIC. Further discussion will be made in Chapter 6.

4.2.3 Choice of Change-points Probability p

In this section, we study the estimated $\widehat{p}$ with one update from our expectation-maximization (EM) algorithm to have some thoughts about how they are related to the true change-points probability p.

Table 4.5: Simulation I: comparing $\widehat{p}$ with one update between different scenarios

Scenario	1	2	3	4	5	6
$\widehat{p}$	0.0018 (1.19E-05)	0.0020 (1.10E-05)	0.0015 (1.48E-05)	0.0014 (1.69E-05)	0.0015 (1.64E-05)	0.0018 (1.47E-05)
p	0.002	0.003	0	0.001^*	0.002^*	0.005^*

Table 4.5 provides the estimated $\widehat{p}$ with one update in our simulation studies. Note that the actual numbers of structure changes are smaller than those according to the true change-points probability p we set in the scenarios with random change-points, due to the required minimum distance $m = 100$. We can see the estimated $\widehat{p}$ in all six scenarios are greater than the initial value $p_0 = 0.001$, which causes more fluctuations in our semi-parametric estimates, therefore leading to inferior estimation and segmentation results. The estimated $\widehat{p}$ increases with true change-points probability p both in the scenarios with fixed change-points and those with random change-points, as shown in Scenario 1 to 3 and 4 to 6 respectively. However, the estimated $\widehat{p}$ reaches its minimum not in Scenario 3, where no structure changes take place, but in Scenario 4. The reason behind is that, in our settings with fixed change-points, each time-homogeneous GARCH(1,1) subsequence has somewhat high persistence, where $\alpha + \beta = 0.8$ or 0.9, which can be further explained by additional structure changes within the subsequence. Therefore, our estimated $\widehat{p}$ from the EM algorithm tends to overestimate the true change-points p. This causes a problem to choose the change-points probability p in our segmentation procedure. One solution is to consider all the estimated $\widehat{p}$ in the EM algorithm and the initial value p_0, and choose the best estimation and segmentation results among them. In our real data analysis, we only consider the estimated $\widehat{p}$ with one update and the initial value p_0 for simplicity.

4.3 Simulation Studies with Individual Parameter Shifts

We also study the case where only one parameter ω, α or β can change at a time so as to compare with the binary segmentation algorithm (BSA) introduced by Galeano and Tsay(2010)[27] for individual parameters shifts. In this study, we have six scenarios each including 500 series with equal length $n = 1000$. We only consider the initial value $p_0 = 0.001$ in our semi-parametric estimation, where a block of 4 is used. This is because, in the previous study, our estimates with the initial value p_0 outperform those with $\widehat{p}$ with one update when the length of the series n is small and p_0 is close to the true change-points probability p. In all six scenarios, the structure changes take place at the fixed locations. The first three scenarios have single change-point for each of ω, α and β , while structure changes take place twice for individual parameters in the last three scenarios. In the setup, we have the same constraints for the "oracle" estimates $\widehat{\theta}_{oracle}$ as in the previous study, such that $\|\widehat{\theta}_{oracle} - \widehat{\theta}_t\|_\infty \leqslant 0.2$. The exact settings of the six scenarios are listed as follows,

Scenario 1. The series are generated from two time-homogeneous GARCH(1,1) models piecewisely and there exist one structure change for ω at $t =$ 501 and 701, where $\theta_t = \{0.8, 0.3, 0.5\}$ for $1 \leqslant t \leqslant 500$ and $\theta_t = \{1.0, 0.3, 0.5\}$ for $501 \leqslant t \leqslant 1000$.

Scenario 2. The series are generated from two time-homogeneous GARCH(1,1) models piecewisely and there exist one structure change for α at $t =$ 501 and 701, where $\theta_t = \{0.8, 0.3, 0.3\}$ for $1 \leqslant t \leqslant 500$ and $\theta_t = \{0.8, 0.5, 0.3\}$ for $501 \leqslant t \leqslant 1000$.

Scenario 3. The series are generated from two time-homogeneous GARCH(1,1) models piecewisely and there exist one structure change for β at $t =$ 501 and 701, where $\theta_t = \{0.8, 0.1, 0.3\}$ for $1 \leqslant t \leqslant 500$ and $\theta_t = \{0.8, 0.1, 0.5\}$ for $501 \leqslant t \leqslant 1000$.

Scenario 4. The series are generated from three GARCH(1,1) models piecewisely and there exist two structure changes for ω at $t = 301$ and 701, where $\theta_t = \{0.8, 0.3, 0.3\}$ for $1 \leqslant t \leqslant 300$, $\theta_t = \{1.0, 0.3, 0.3\}$ for $301 \leqslant t \leqslant 700$, and $\theta_t = \{1.2, 0.3, 0.3\}$ for $701 \leqslant t \leqslant 1000$.

Scenario 5. The series are generated from three GARCH(1,1) models piecewisely and there exist two structure changes for α at $t = 301$ and 701, where $\theta_t = \{1.0, 0.1, 0.3\}$ for $1 \leqslant t \leqslant 300$, $\theta_t = \{1.0, 0.3, 0.3\}$ for $301 \leqslant t \leqslant 700$, and $\theta_t = \{1.0, 0.5, 0.3\}$ for $701 \leqslant t \leqslant 1000$.

Scenario 6. The series are generated from three GARCH(1,1) models piecewisely and there exist two structure changes for β at $t = 301$ and 701, where $\theta_t = \{1.0, 0.3, 0.1\}$ for $1 \leqslant t \leqslant 300$, $\theta_t = \{1.0, 0.3, 0.3\}$ for $301 \leqslant t \leqslant 700$, and $\theta_t = \{1.0, 0.3, 0.5\}$ for $701 \leqslant t \leqslant 1000$.

4.3.1 Semi-parametric Estimates

In this section, we compare the semi-parametric estimates with the initial value $p_0 = 0.001$ in our estimation procedure to the "oracle" estimates. We also apply the bounded complexity mixture (BCMIX) approximation and use a block of 4 in the simulation studies to reduce the computational time. The typical estimates in the six scenarios are shown in Figure 4.17 to 4.22. The solid lines represent the semi-parametric estimates with $p_0 = 0.001$, where the dash lines represent the "oracle" estimates and the dot lines represent the true parameter vector θ_t. We can see our estimates are very close to the "oracle" estimates, where they are most apart from each other either near the true structure changes or at the edges of the series. However, it is very difficult to locate the structure changes as the scale of individual parameter shifting is considerably small in our settings. With the length of the series $n = 1000$, the jump size of each shift 0.2 is often offset by the estimation errors, where it is still possible for our "oracle" estimates to ignore such shifts given our constraints.

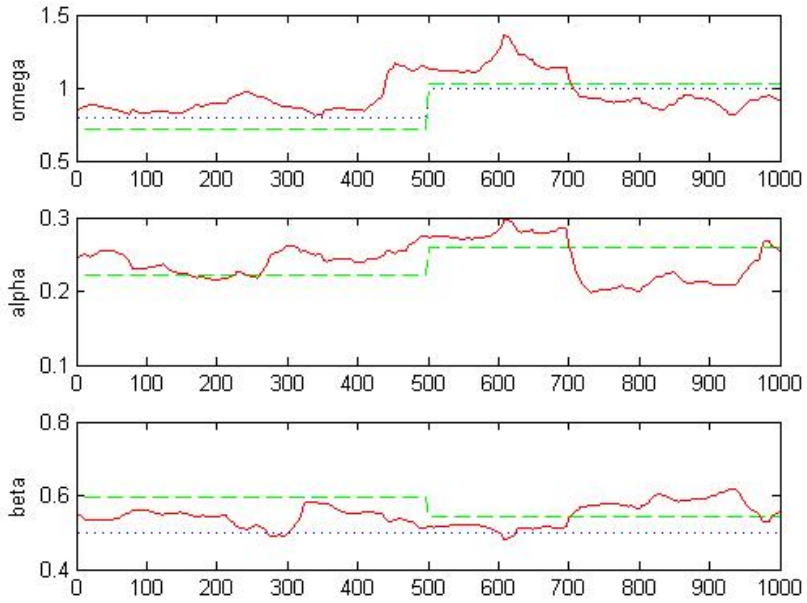

Figure 4.17: Simulation II: typical estimates in Scenario 1

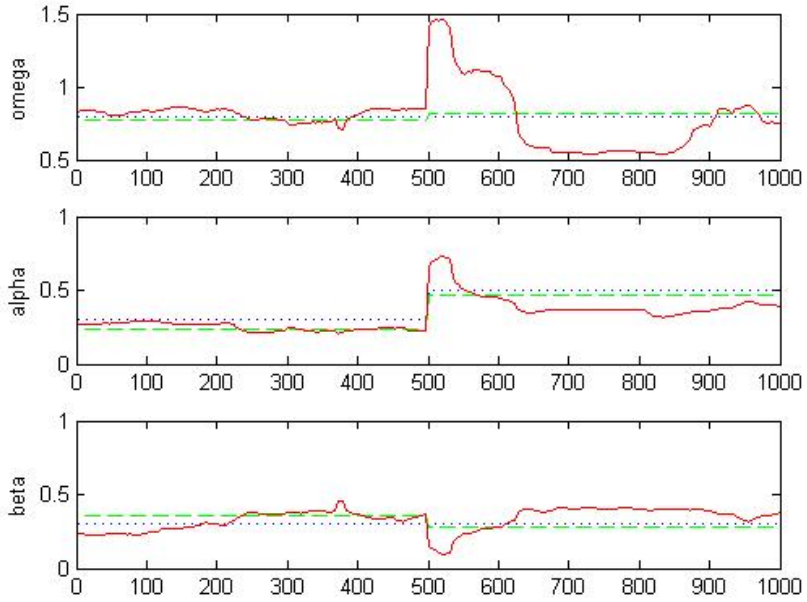

Figure 4.18: Simulation II: typical estimates in Scenario 2

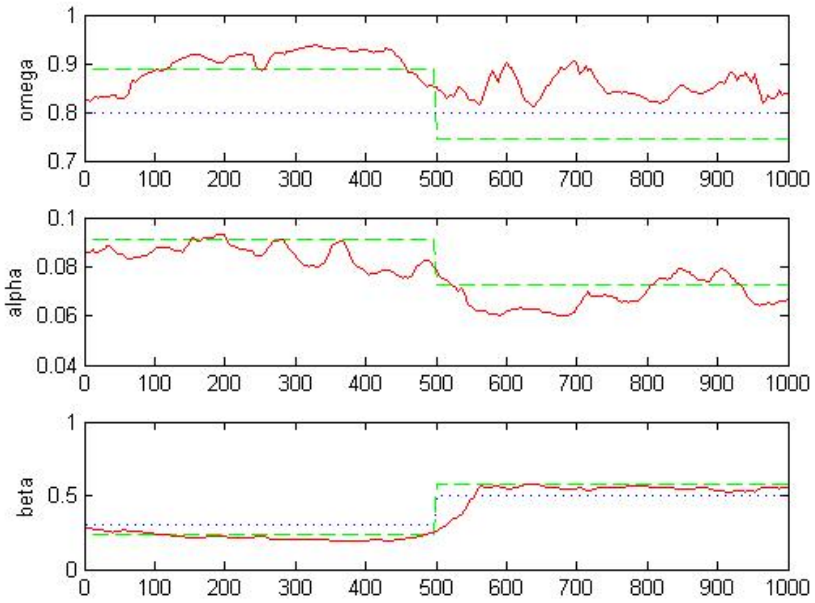

Figure 4.19: Simulation II: typical estimates in Scenario 3

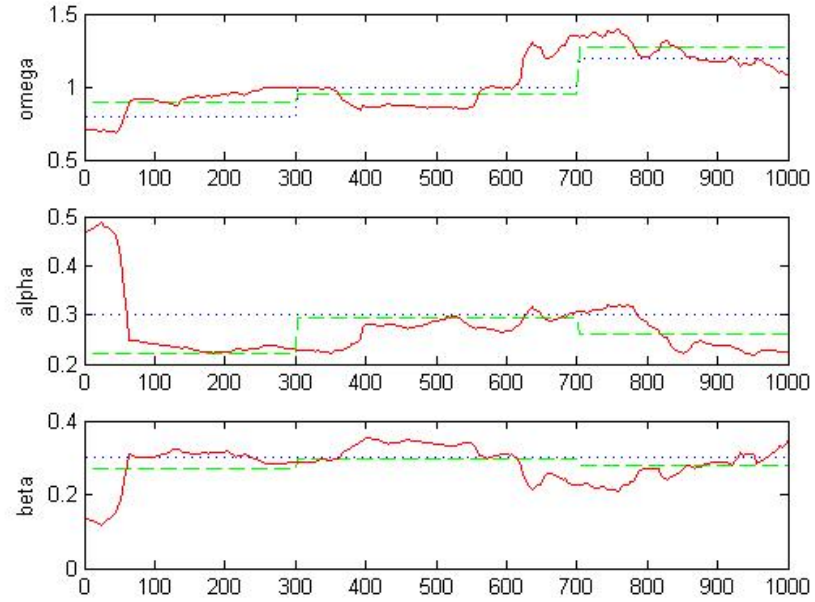

Figure 4.20: Simulation II: typical estimates in Scenario 4

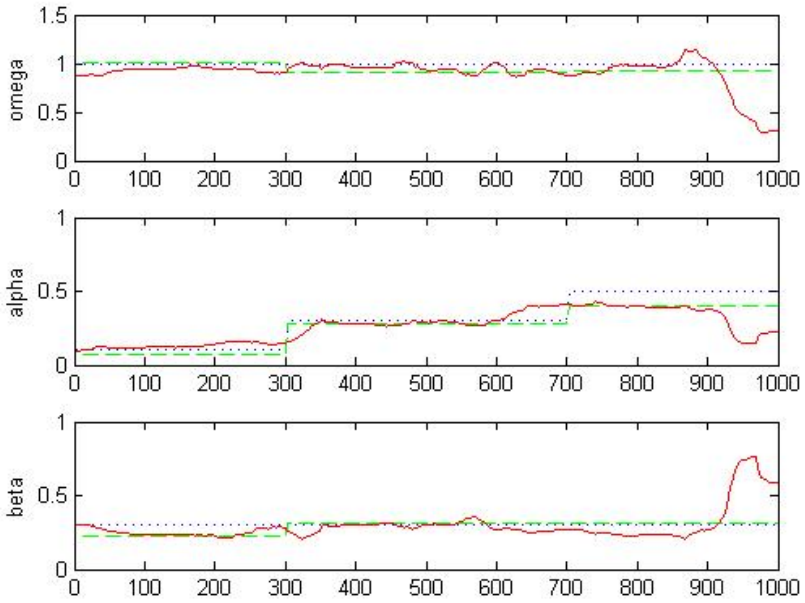

Figure 4.21: Simulation II: typical estimates in Scenario 5

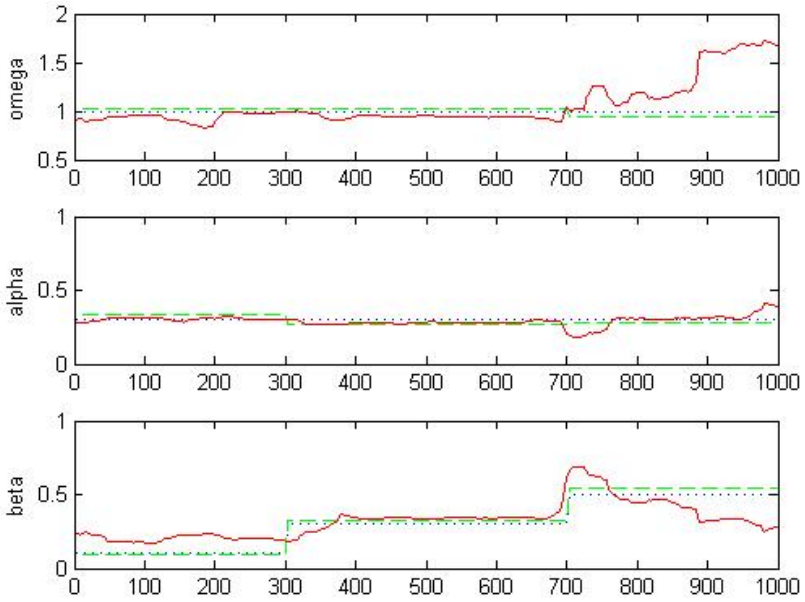

Figure 4.22: Simulation II: typical estimates in Scenario 6

Table 4.6: Simulation II: comparing semi-parametric estimates with "oracle" estimates

		EE	KL	GOF
Scenario 1	p_0	0.3114 (0.0061)	0.0129 (2.25E-04)	0.9906 (0.0004)
	oracle	0.0833 (0.0009)	0.0015 (4.32E-05)	1.0019 (0.0002)
Scenario 2	p_0	0.2083 (0.0029)	0.0118 (2.48E-04)	0.9870 (0.0005)
	oracle	0.0905 (0.0008)	0.0015 (3.99E-05)	1.0014 (0.0001)
Scenario 3	p_0	0.2010 (0.0040)	0.0087 (1.61E-04)	0.9911 (0.0004)
	oracle	0.0825 (0.0009)	0.0015 (4.55E-05)	1.0011 (0.0001)
Scenario 4	p_0	0.1851 (0.0023)	0.0096 (1.71E-04)	0.9879 (0.0004)
	oracle	0.0886 (0.0007)	0.0017 (4.33E-05)	1.0015 (0.0001)
Scenario 5	p_0	0.2010 (0.0027)	0.0098 (1.63E-04)	0.9868 (0.0004)
	oracle	0.0912 (0.0007)	0.0019 (4.52E-05)	1.0020 (0.0002)
Scenario 6	p_0	0.2405 (0.0037)	0.0117 (1.76E-04)	0.9889 (0.0004)
	oracle	0.0880 (0.0007)	0.0017 (4.16E-05)	1.0029 (0.0002)

Table 4.6 compares our estimates given $p_0 = 0.001$ with the "oracle" estimates in the six scenarios in regard to the mean Euclidean error (EE), the Kullback-Leibler (KL) divergence and the goodness of fit (GOF). For each item listed above, the table presents the mean values and the corresponding standard deviations of the means, which are shown in the parentheses. The differences of the mean Euclidean error (EE) and the the Kullback-Leibler (KL) divergence between our estimates and the "oracle" estimates are close to those of the scenarios with random change-points in the previous study, as shown in Table 4.2. Therefore, the estimation errors of our procedure is mostly determined by the persistence of the GARCH(1,1) subsequences in the setup, which is leveraged by our choice of change-points probability p. As for our semi-parametric estimates, both the mean values and the standard deviations of the means of the goodness of fit (GOF) are close to the theoretical values 1 and 0.002, indicating great performance in fitting the volatility. Therefore, though the structure changes are difficult to be detected in these scenarios as the scale of the shifts is small compared to the estimation errors, our estimates

are very close to the "oracle" estimates, providing certain inference on the number and locations of the change-points. Thus we can still use the function Δ_t given our estimates to detect the structure changes.

4.3.2 Segmentation Results

In this section, we apply our segmentation procedure to the case where only one parameter ω, α or β can change at a time. Remember that the first part of the penalty term in the modified Bayesian information criterion (MBIC), $2k\log(n)$, is solely based on the number of free parameters to be estimated, as shown in (3.11). Therefore, we should reduce such penalty into half in our application to individual parameter shifting, since for each change-point added, we introduce a parameter for the location of the new change-point and only one parameter that corresponds to the type of the structure change, i.e. ω, α or β, rather than totally four parameters as in the case of simultaneous changes. Our modified Bayesian information criterion (MBIC) for shifts in the individual parameters of a GARCH(1,1) model can be written as,

$$MBIC^* = -2l(\widehat{\theta}_t) + k\log(n) + \frac{2k}{k+2}\log(m)(\sum_{i=1}^{k+1}\log(l_i - l_{i-1}) - (k+1)\log(n/(k+1))) \tag{4.7}$$

Table 4.7 compares the segmentation results between our unconditional variance based segmentation procedure (UVB) and the binary segmentation algorithm (BSA) introduced by Galeano and Tsay(2010)[27]. In our BSA, we replace MBIC by MBIC* and also use a minimum possible distance $m = 100$. Meanwhile, we consider both the significant levels of 5% and 10% for BSA, where the critical values are listed in Table 3.3 with a corresponding $\nu_{\min}$. Here $\nu_{\min}$ is made to satisfy the minimum possible distance m. The table shows the numbers of change-points detected in each of the six scenarios.

We can see our UVB is superior to BSA in terms of accuracy rate of the numbers of change-points detected in all six scenarios except Scenario 3. This is because BSA works best when detecting single change-point, where it degenerates to the Lagrange multiplier

Table 4.7: Simulation II: comparing segmentation results between UVB and BSA

		$\widehat{k}=0$	$\widehat{k}=1$	$\widehat{k}=2$	$\widehat{k}=3$	$\widehat{k}$	
Scenario 1 $(\omega, k=1)$	UVB(p_0)	0.316	0.312	0.200	0.110	1.302	(0.055)
	BSA(5%)	0.800	0.184	0.016	0.000	0.216	(0.020)
	BSA(10%)	0.682	0.266	0.044	0.006	0.380	(0.028)
Scenario 2 $(\alpha, k=1)$	UVB(p_0)	0.262	0.330	0.222	0.122	1.410	(0.054)
	BSA(5%)	0.768	0.210	0.020	0.002	0.256	(0.022)
	BSA(10%)	0.598	0.318	0.076	0.008	0.494	(0.030)
Scenario 3 $(\beta, k=1)$	UVB(p_0)	0.050	0.564	0.276	0.092	1.468	(0.037)
	BSA(5%)	0.216	0.732	0.052	0.000	0.836	(0.022)
	BSA(10%)	0.126	0.726	0.130	0.016	1.042	(0.026)
Scenario 4 $(\omega, k=2)$	UVB(p_0)	0.210	0.466	0.192	0.090	1.296	(0.047)
	BSA(5%)	0.546	0.422	0.030	0.000	0.490	(0.026)
	BSA(10%)	0.342	0.582	0.066	0.004	0.750	(0.029)
Scenario 5 $(\alpha, k=2)$	UVB(p_0)	0.034	0.576	0.242	0.106	1.550	(0.040)
	BSA(5%)	0.212	0.718	0.066	0.004	0.862	(0.023)
	BSA(10%)	0.090	0.758	0.128	0.024	1.086	(0.025)
Scenario 6 $(\beta, k=2)$	UVB(p_0)	0.000	0.320	0.376	0.180	2.134	(0.047)
	BSA(5%)	0.004	0.746	0.220	0.030	1.276	(0.023)
	BSA(10%)	0.000	0.518	0.354	0.114	1.626	(0.033)

(LM) test. The numbers of detected structure changes are mostly accurate in the scenarios when only β changes, indicating the shifts in the autocorrelation are easier to detect. As in the previous study, our UVB tends to overestimate the number of change-points when the true structure changes are evenly distributed, due to the penalty term in our MBIC*. Meanwhile, BSA always underestimates the number of change-points, especially when the scale of the size of each jump is small compared to the estimation errors compared to the estimation errors as in these scenarios. There are two reasons: first, the structure changes that are significant locally may fail to be detected in the overall series; second, the

detected structure changes may deviate from the location of the true structure changes or even include some misleading change-points, which can result in the elimination of the true structure changes being detected, as BSA involves the further refinement in the local subsequences. The latter is really a curse for the change-points estimation problems in the GARCH settings, as the estimation errors are huge. During the refinement process of BSA, a significant change-point can be easily eliminated in the subinterval of two less significant change-points, which will be further eliminated as of their lower significance. Thus BSA may often end up detecting only one single change-point, as shown in the following real data analysis.

Chapter 5

Real Data Analysis

5.1 Daily Log Return Series

5.1.1 S&P 500 Index

For the first example, we analyze the daily log return series of the S&P500 index from January 4, 1999 to December 31, 2014 as shown in Figure 1.1, which consists of $n = 4024$ data points. The data come from *Yahoo! Finance*. The sample mean, variance, skewness, and excess kurtosis of the return series are 5.58×10^{-5}, 3.07×10^{-5}, -0.1754, and 7.73, respectively. Thus, the data presents negative skewness and high excess kurtosis, the latter of which can be well fit by GARCH models. The estimated coefficients by fitting the return series with the GARCH (1,1) model are which results in $\mu = 2.07 \times 10^{-4},\ \omega = 2.69 \times 10^{-7},\ \alpha = 0.905,\ \beta = 0.086$, with the corresponding errors 6.11×10^{-5}, 3.72×10^{-8}, 0.0071 and 0.0066, respectively. Note that $\hat{\alpha} + \hat{\beta} = 0.991$, the return series shows high persistence in volatility.

Since we are more interested in the patterns of the volatility, the error terms of the return series are studied by simply removing the drift term $\mu = 5.58 \times 10^{-5}$. We provide our semi-parameter estimates with $\widehat{p}$ with one update and the initial value $p_0 = 0.001$, the latter of which are shown in Figure 5.1. No blocks are used in our estimation procedure, where $\widehat{p} = 0.0034$ is given after the first iteration.

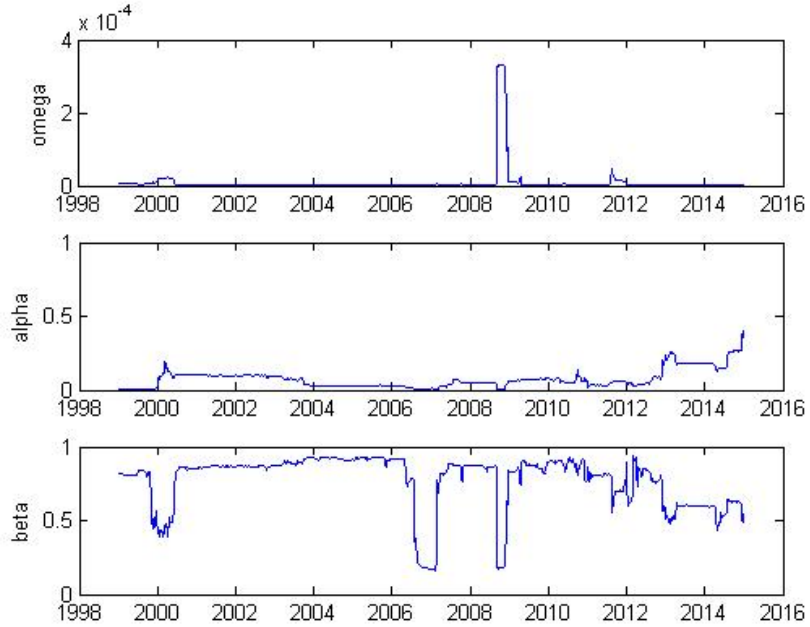

Figure 5.1: S&P 500: semi-parametric estimates of daily log return series with p_0

Figure 5.1 shows our semi-parameter estimates of the S&P 500 index daily log return series from 1999 to 2014 with $p_0 = 0.001$, where the labels under the horizontal axis represent the start of the corresponding year. We can see there exist a huge shift for ω and β prior to 2009, while the parameters are quite stable between 2001 to 2006. Also, α increases dramatically after 2013.

In Figure 5.2, we compare the function Δ_t given $\widehat{p}$ with one update and that given $p_0 = 0.001$, which are computed from our semi-parameter estimates as in (3.7). The solid line represents Δ_t given $\widehat{p}$ with one update, while the dash line represents that given $p_0 = 0.001$. From the figure, we can see, though these two functions are similar in shape, the local maxima of them are not necessarily the same. For instance, the function Δ_t given $\widehat{p}$ with one update reaches its maximum between 2008 and 2009 while Δ_t given p_0 reaches its maximum between 2011 and 2012. Therefore, Δ_t given different change-points probability p would probably result in different possible change-points in order, which may further leads to different structure changes being detected. The reason behind is that there exist no clear boundaries to separate the structure changes and

others in the real data analysis, where the order of significance for each data point may slightly differ given our estimates with different change-points probability p. Therefore, we need to choose from $\widehat{p}$ with one update and p_0 to determine which sets of change-points as our detected structure changes. We provide a solution to this by considering the overall maximized log-likelihood and the Bayesian information criterion (BIC) as shown in (3.10) in regard to the post-segmentation piecewise GARCH(1,1) sequences given our unconditional variance based (UVB) segmentation procedure. We also compare the results from the binary segmentation algorithm (BSA) introduced by Galeano and Tsay(2010)[27], with a significant level 5%, and the time-homogeneous GARCH(1,1) model. Here, we choose the minimum distance $m = 200$ in our segmentation procedure.

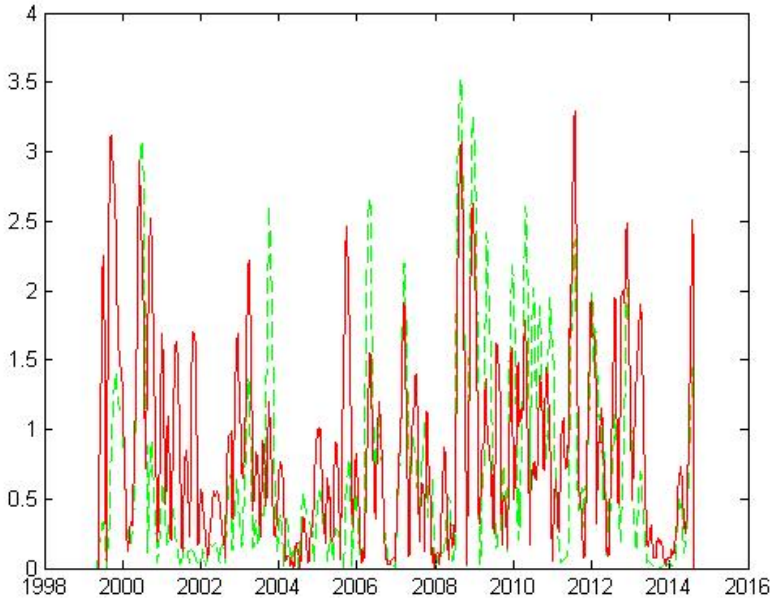

Figure 5.2: S&P 500: Δ_t of daily log return series given $\widehat{p}$ with one update and p_0

Figure 5.3 shows our segmentation results of the S&P 500 daily log return series with $\widehat{p}$ with one update and $p_0 = 0.001$. The solid line represents the structure change detected in both settings, where the dash lines represent the rest change-points detected with $\widehat{p}$ with one update. Here we assume change-points given $\widehat{p}$ and $p_0 = 0.001$ are identical within the range of 4.

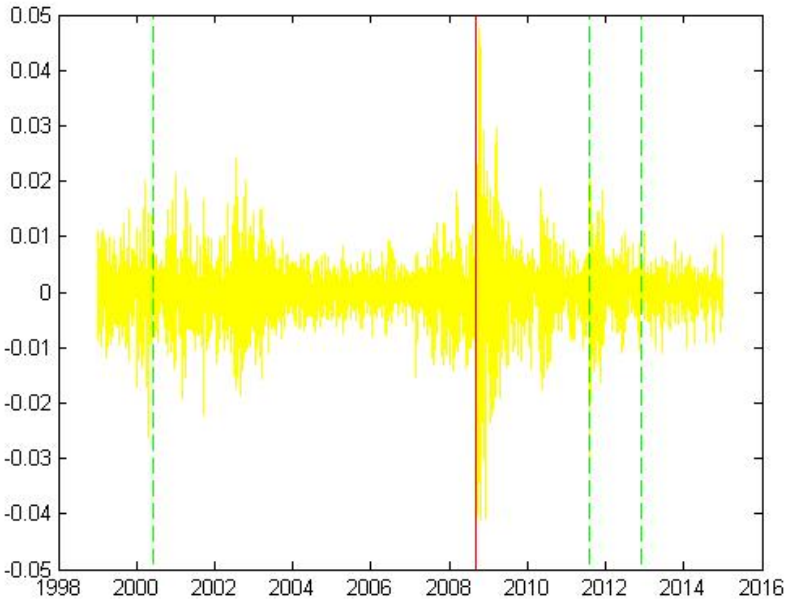

Figure 5.3: S&P 500: segmentation results of daily log return series

From the figure, we can see there exist four detected structure changes from our procedure with $\widehat{p}$ with one update, where $t = 359$, 2431, 3162 and 3495, while only one change-point is detected with $p_0 = 0.001$, where $t = 2431$. Therefore, choice of the change-points probability p would greatly affect the number of detected structure changes. However, in both cases, it appears that the detected change-points are reasonable by comparing to the time plot of the return series.

	UVB($\widehat{p}$)	UVB(p_0)	BSA	N/A
LLM	16112	16083	16085	16074
BIC	-32279	-32180	-32184	-32148

Table 5.1: S&P 500: comparing post-segmentation maximized log-likelihood and BIC of daily log return series

Table 5.1 compares the post-segmentation maximized log-likelihood and BIC from different segmentation procedures. Here, "LLM" stands for the maximized log-likelihood and "N/A" stands for no segmentation procedure is applicable, i.e. the results come from the time-homogeneous GARCH(1,1) model.

In the table, the binary segmentation algorithm (BSA) results in only one change-point at $t = 403$, which provides an even smaller BIC than our procedure with $p_0 = 0.001$. This is because the Lagrange multiplier (LM) test is superior for detectind single change-point in regard of the maximized log-likelihood as it is based on an overall test statistic. However, the detected change-points from our procedure, which are based on the local relative changes of the unconditional variances, seem to be more reasonable. One possible explanation is that the unconditional variances are more stable than the corresponding estimated parameters. Therefore, our procedure has superiority over the binary segmentation algorithm (BSA) to detect multiple structure changes when the estimation errors are large, which is often true in the GARCH settings. In this case, we choose our procedure given $\widehat{p}$ with one update, as it provides the smallest BIC among all the segmentation results.

Table 5.2: S&P 500: comparing pre- and post-segmentation piecewise estimates

Period	ω		α		β	
$[1, 358]$	2.00E-07	(6.63E-07)	0.9748	(0.0310)	0.0207	(0.0122)
$[359, 2430]$	2.00E-07	(4.14E-08)	0.9258	(0.0096)	0.0653	(0.0086)
$[2431, 3161]$	3.63E-07	(1.36E-07)	0.8972	(0.0169)	0.0939	(0.0170)
$[3162, 3494]$	5.30E-07	(2.77E-07)	0.8757	(0.0329)	0.1020	(0.0253)
$[3495, 4024]$	1.99E-06	(7.10E-07)	0.5775	(0.1135)	0.2189	(0.0590)
$[1, 4024]$	2.82E-07	(3.91E-08)	0.9034	(0.0072)	0.0861	(0.0066)

Table 5.2 compares the pre- and post-segmentation piecewise estimates of the S&P 500 daily log return series. We can see considerably higher autocorrelation of the volatility in the last period of our segmentation, therefore the recent movements of the volatility are more predictable. However, in other periods, there still exists high persistence, due to the additional possible structure changes within each period as well as the huge estimation errors given a smaller sample size in each period. Thus, as previously mentioned, we are more interested in the segmentation results.

Table 5.3: S&P 500: detected structure changes of daily log return series

Change-point	Date	Event	Stage
359	June 5, 2000	internet bubble	mid
2431	September 2, 2008	global financial crisis of 2008	mid
3162	July 27, 2011	United States debt-ceiling crisis of 2011	start
3495	November 21, 2012	United States debt-ceiling crisis of 2011	end

Table 5.3 presents the detected structure changes of the S&P 500 daily log return series from our procedure with corresponding dates in real world. We also list the related extraordinary economic events, which are supported by the following facts,

May 16, 2000 Federal funds rate raised to the highest level 6.5% since 1991.

September 7, 2008 The U.S. government took over Fannie Mae and Freddie Mac.

August 5, 2011 S&P downgraded U.S. credit rating.

January 1, 2013 Congress approved a budget deal to avoiding the "fiscal cliff."

The facts listed are from California Department of Finance at the webpage *http* : *//www.dof.ca.gov/HTML/FS_DATA/LatestEconData/Chronology/chronology.htm*. Though these ex-post economic interpretations are very speculative, our estimation and segmentation procedures successfully identify the parameter changes that coincide with the external events affecting the US financial market.

Figure 5.4 shows the post-segmentation estimated volatility of the S&P 500 daily log return series. We can see our detected structure changes well response to the estimated volatility. The estimated volatility reaches its maximum prior to 2009, where it is less volatile at a low value between 2003 and 2007, which presents the similar patterns as our semi-parametric estimates.

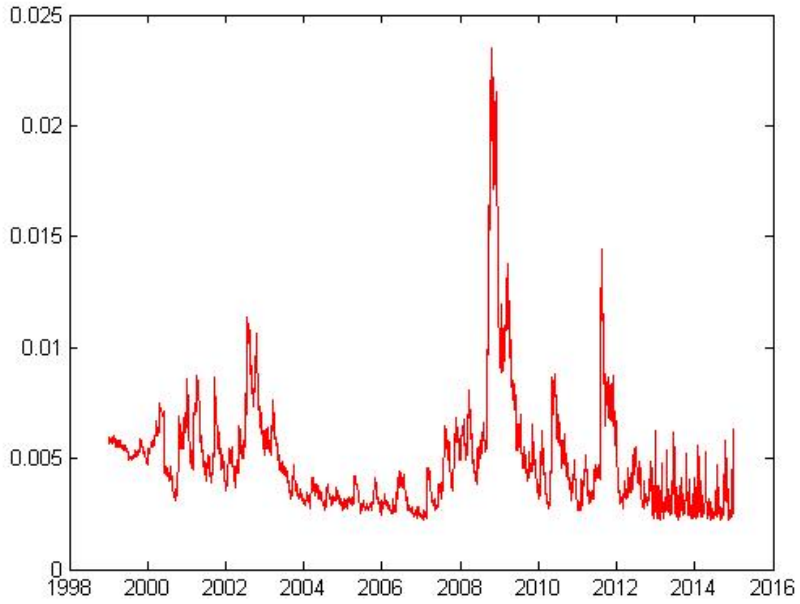

Figure 5.4: S&P 500: post-segmentation estimated volatility of daily log return series

5.1.2 IBM Stock Return

In the second example, we illustrate the performance of our proposed procedure by analyzing the daily log return series of the IBM stock from January 4, 1999 to December 31, 2014 as shown in Figure 5.5, which consists of $n = 4024$ data points. The data come from *Yahoo! Finance*. The sample mean, variance, skewness, and excess kurtosis of the return series are 8.29×10^{-5}, 5.98×10^{-5}, -0.1453, and 8.56, respectively. Thus, the data presents negative skewness and high excess kurtosis, the latter of which can be well fit by GARCH models. The estimated coefficients by fitting the return series with the GARCH (1,1) model are which results in $\mu = 2.41 \times 10^{-4}$, $\omega = 8.19 \times 10^{-7}$, $\alpha = 0.891$, $\beta = 0.099$, with the corresponding errors 7.95×10^{-5}, 6.23×10^{-8}, 0.0051 and 0.0054, respectively. Note that $\hat{\alpha} + \hat{\beta} = 0.990$, the return series shows high persistence in volatility.

The error terms of the return series are studied by simply removing the drift term $\mu = 8.29 \times 10^{-5}$. We provide our semi-parameter estimates with $\widehat{p}$ with one update and the initial value $p_0 = 0.001$, the latter of which are shown in Figure 5.6. No blocks are used in our estimation procedure, where $\widehat{p} = 0.0042$ is given after the first iteration.

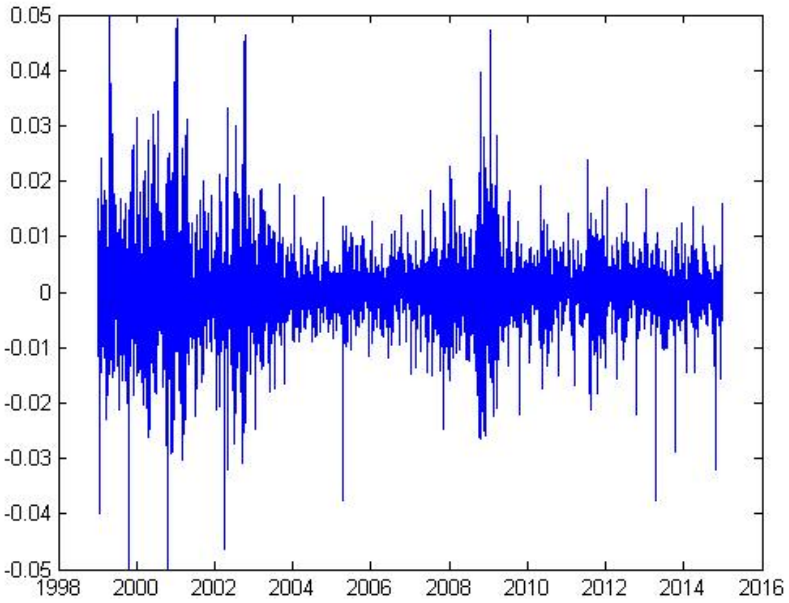

Figure 5.5: IBM: daily log return series from 1999 to 2014

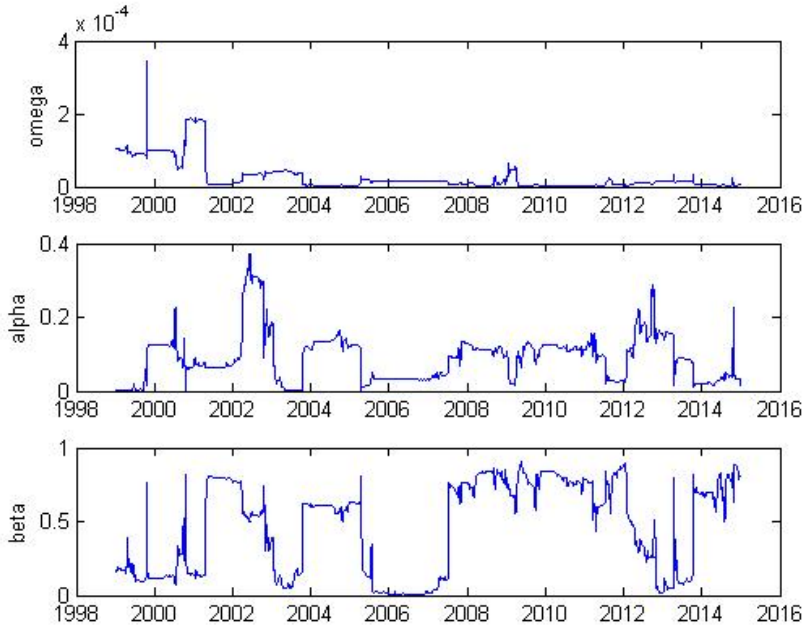

Figure 5.6: IBM: semi-parametric estimates of daily log return series with p_0

Figure 5.6 shows our semi-parameter estimates of the IBM stock daily log return series from 1999 to 2014 with $p_0 = 0.001$, where the labels under the horizontal axis represent the start of the corresponding year.

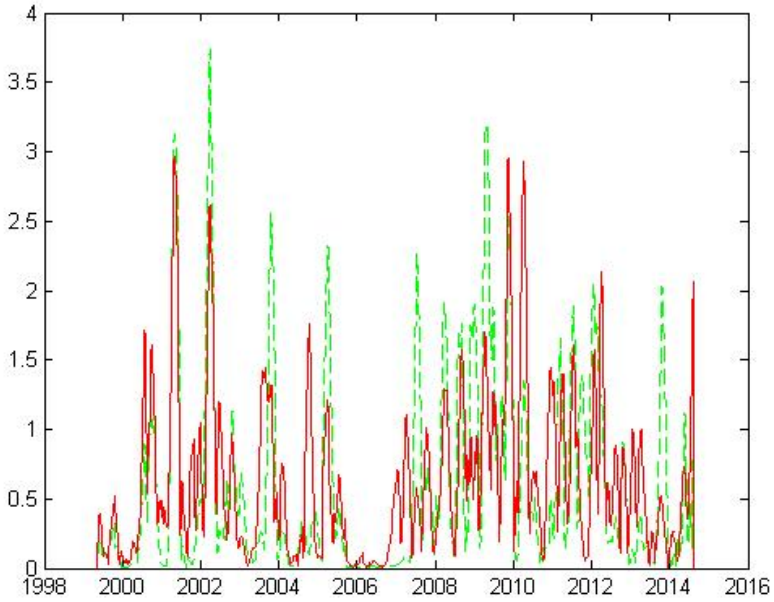

Figure 5.7: IBM: Δ_t of daily log return series given $\widehat{p}$ with one update and p_0

In Figure 5.7, we compare the function Δ_t given $\widehat{p}$ with one update and that given $p_0 = 0.001$, which are computed from our semi-parameter estimates as in (3.7). The solid line represents Δ_t given $\widehat{p}$ with one update, while the dash line represents that given $p_0 = 0.001$. From the figure, we can see, though these two functions are similar in shape, the local maxima of them are not necessarily the same. Here, we choose the minimum distance $m = 200$ in our segmentation procedure.

Figure 5.8 shows our segmentation results of the IBM stock daily log return series with $\widehat{p}$ with one update and $p_0 = 0.001$. The solid lines represent the structure changes detected in both settings, where the dash lines represent the rest change-points detected with $\widehat{p}$ with one update and the dot lines represent those with $p_0 = 0.001$ only. Here we assume change-points given $\widehat{p}$ and $p_0 = 0.001$ are identical within the range of 4. We can see there exist four detected structure changes from our procedure with $\widehat{p}$ with one update, where

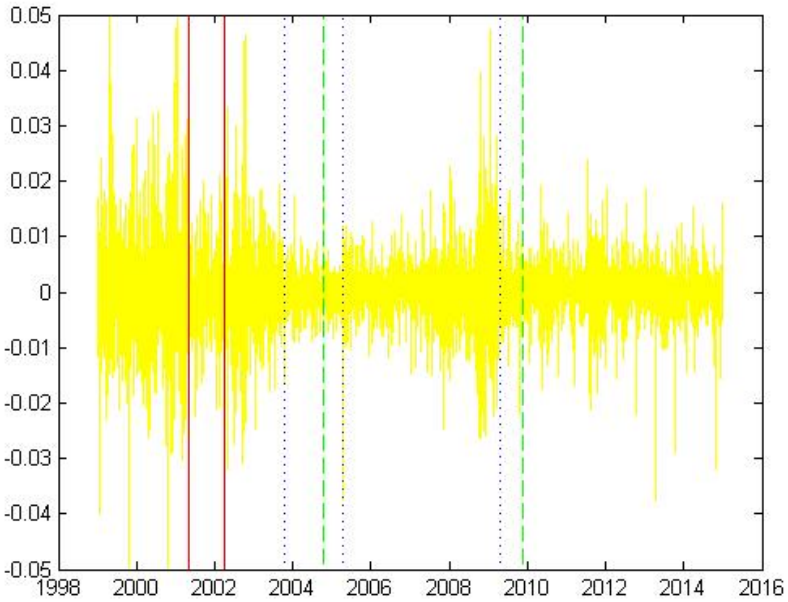

Figure 5.8: IBM: segmentation results of daily log return series

$t = 587$, 814, 1454 and 2733, while five change-points are detected with $p_0 = 0.001$, where $t = 588$, 815, 1205, 1574 and 2591. Therefore, choice of the change-points probability p would greatly affect the number of detected structure changes. However, in both cases, it appears that the detected change-points are reasonable by comparing to the time plot of the return series.

Table 5.4 compares the post-segmentation maximized log-likelihood and BIC from different segmentation procedures. Here, "LLM" stands for the maximized log-likelihood and "N/A" stands for no segmentation procedure is applicable, i.e. the results come from the time-homogeneous GARCH(1,1) model. In this case, we choose our procedure given $p_0 = 0.001$, as it provides the smallest BIC among all the segmentation results.

In Table 5.5, we compares the pre- and post-segmentation piecewise estimates of the IBM stock daily log return series. Table 5.6 presents the detected structure changes of the IBM stock daily log return series from our procedure with corresponding dates in real world. Though these ex-post economic interpretations are very speculative, we successfully identify the parameter changes that coincide with the external events.

Table 5.4: IBM: comparing post-segmentation maximized log-likelihood and BIC of daily log return series

	UVB($\widehat{p}$)	UVB(p_0)	BSA	N/A
LLM	14646	14684	14618	14576
BIC	-29347	-29437	-29250	-29152

Table 5.5: IBM: comparing pre- and post-segmentation piecewise estimates

Period	ω		α		β	
$[1, 587]$	9.56E-06	(5.92E-06)	0.9215	(0.0442)	0.0194	(0.0091)
$[588, 814]$	7.41E-06	(6.01E-06)	0.7972	(0.1442)	0.0678	(0.0571)
$[815, 1204]$	3.77E-06	(1.41E-06)	0.8574	(0.0339)	0.0995	(0.0250)
$[1205, 1573]$	3.67E-06	(1.91E-06)	0.6188	(0.1729)	0.1295	(0.0574)
$[1574, 2590]$	2.90E-07	(8.94E-08)	0.9440	(0.0075)	0.0487	(0.0075)
$[2591, 4024]$	4.38E-06	(4.30E-06)	0.7409	(0.1806)	0.1216	(0.0823)
$[1, 4024]$	8.20E-07	(6.09E-08)	0.8921	(0.0050)	0.0976	(0.0053)

Table 5.6: IBM: detected structure changes of daily log return series

Change-point	Date	Event	Stage
588	May 2, 2001	internet bubble	end
815	April 3, 2002	stock market downturn	start
1205	October 17, 2003	stock market downturn	end
1574	April 7, 2005	bull market	start
2591	April 22, 2009	global financial crisis	end

Figure 5.9 shows the post-segmentation estimated volatility of the IBM stock daily log return series. We can see our detected structure changes well response to the estimated volatility as it reaches local maxima close to 2003 and 2009.

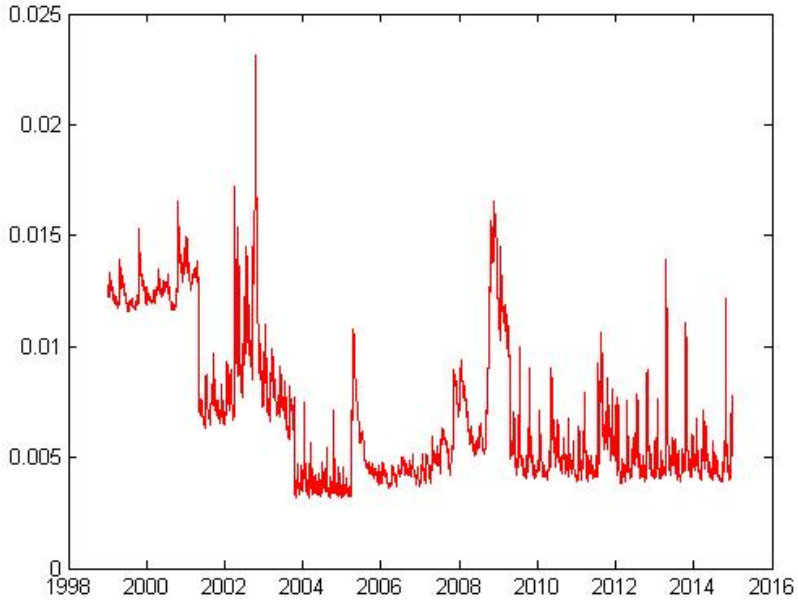

Figure 5.9: IBM: post-segmentation estimated volatility of daily log return series

5.2 Weekly Log Return Series

5.2.1 S&P 500 Index

We are also interested in whether our procedure can be applied to the financial time series where more fluctuations exist in the volatility. Therefore, for the third example, we analyze the weekly log return series of the S&P500 index from January 4, 1971 to December 29, 2014 as shown in Figure 5.10, which consists of $n = 2294$ data points. The data come from *Yahoo! Finance*. The sample mean, variance, skewness, and excess kurtosis of the return series are 5.88×10^{-4}, 9.73×10^{-5}, -0.5628, and 5.50, respectively. Thus, the data presents negative skewness and high excess kurtosis, the latter of which can be well fit by GARCH models. The estimated coefficients by fitting the return series with the GARCH (1,1) model are which results in $\mu = 8.97 \times 10^{-4}$, $\omega = 4.04 \times 10^{-6}$, $\alpha =$

0.820, $\beta = 0.086$, with the corresponding errors 1.62×10^{-4}, 7.70×10^{-7}, 0.141 and 0.0113, respectively. Note that $\hat{\alpha} + \hat{\beta} = 0.961$, the return series also shows high persistence in volatility as compared to the daily log return series.

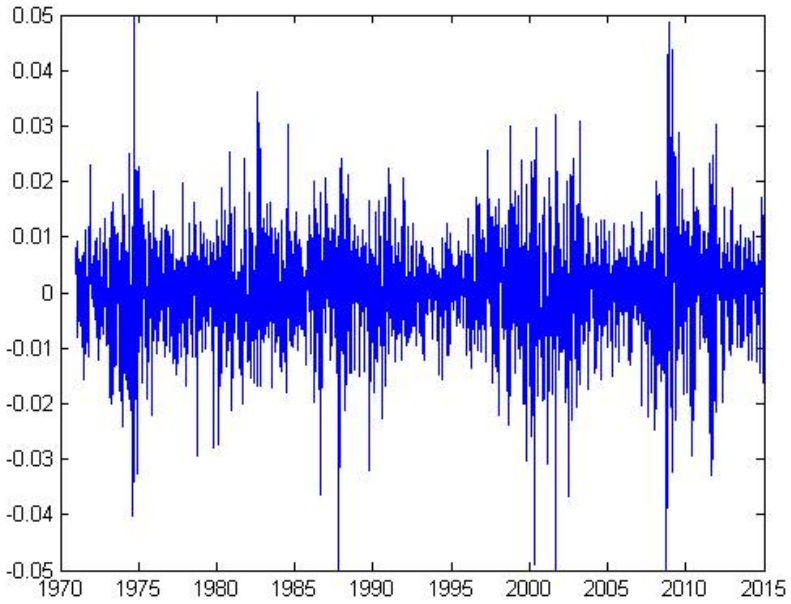

Figure 5.10: S&P 500: weekly log return series from 1999 to 2014

The error terms of the return series are studied by simply removing the drift term $\mu = 5.88 \times 10^{-4}$. We provide our semi-parameter estimates with $\widehat{p}$ with one update and the initial value $p_0 = 0.001$, the latter of which are shown in Figure 5.11. No blocks are used in our estimation procedure, where $\widehat{p} = 0.0045$ is given after the first iteration.

Figure 5.11 shows our semi-parameter estimates of the S&P 500 weekly log return series from 1971 to 2014 with $p_0 = 0.001$, where the labels under the horizontal axis represent the start of the corresponding year and Figure 5.12 shows our segmentation results of the S&P 500 weekly log return series with $\widehat{p}$ with one update and $p_0 = 0.001$. The solid line represent the structure change detected in both settings, where the dot lines represent those with $p_0 = 0.001$ only. Here we choose the minimum distance $m = 100$ in our segmentation procedure and assume change-points given $\widehat{p}$ and $p_0 = 0.001$ are identical within the range of 4.

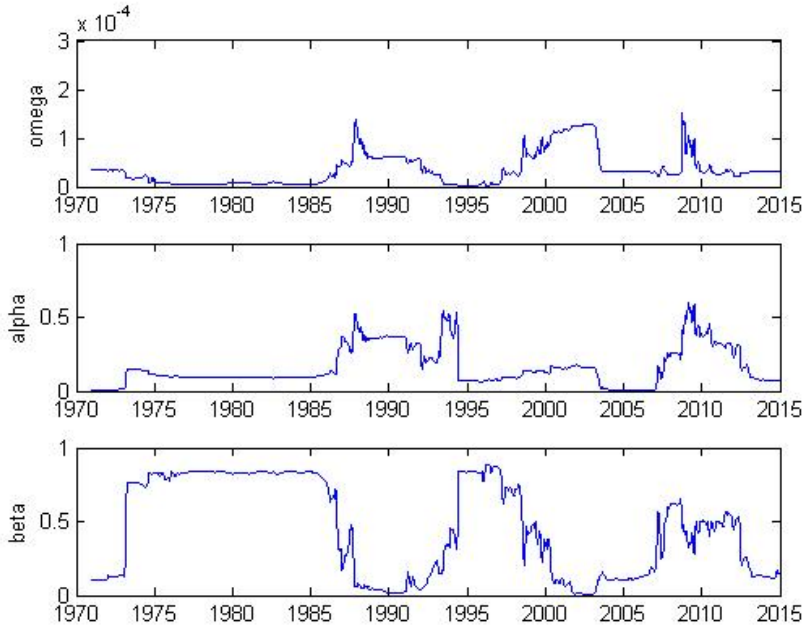

Figure 5.11: S&P 500: semi-parametric estimates of weekly log return series with p_0

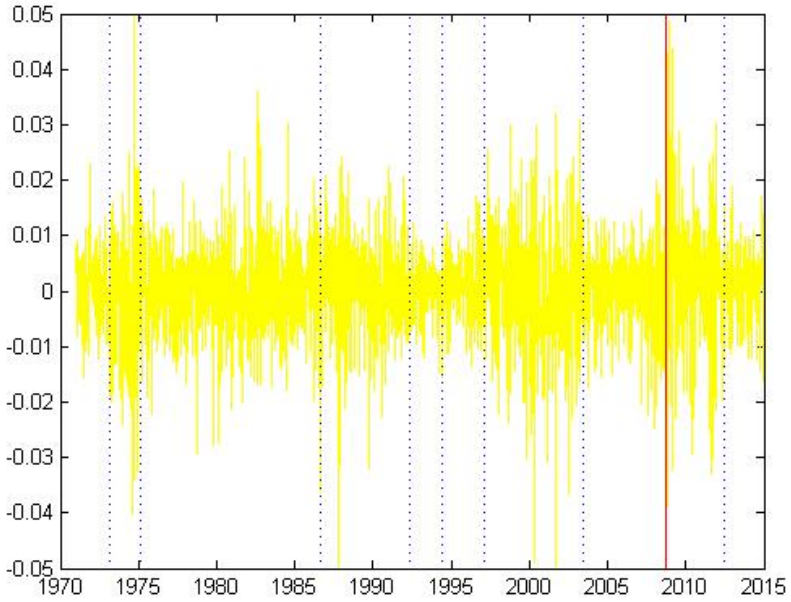

Figure 5.12: S&P 500: segmentation results of weekly log return series

We can see there exist only one detected structure change from our procedure with $\widehat{p}$ with one update, where $t = 1965$, while nine change-points are detected with $p_0 = 0.001$, where $t = 114$, 215, 817, 1112, 1223, 1361, 1690, 1967 and 2160. Therefore, choice of the change-points probability p would greatly affect the number of detected structure changes. However, in both cases, it appears that the detected change-points are reasonable by comparing to the time plot of the return series.

Table 5.7 compares the post-segmentation maximized log-likelihood and BIC from different segmentation procedures. Here, "LLM" stands for the maximized log-likelihood and "N/A" stands for no segmentation procedure is applicable, i.e. the results come from the time-homogeneous GARCH(1,1) model. In this case, we choose our procedure given $p_0 = 0.001$, as it provides the smallest BIC among all the segmentation results.

Table 5.7: S&P 500: comparing post-segmentation maximized log-likelihood and BIC of weekly log return series

	UVB($\widehat{p}$)	UVB(p_0)	BSA	N/A
LLM	7595	7655	7609	7584
BIC	-15203	-15435	-15260	-15168

Change-point	Date	Event	Stage
114	March 5, 1973	oil crisis	start
215	February 10, 1975	oil crisis	end
817	August 25, 1986	black Monday	start
1112	April 20, 1992	black Wednesday	start
1223	June 6, 1994	economic crisis in Mexico	start
1361	January 27, 1997	Asian Financial Crisis	start
1690	May 27, 2003	stock market downturn	end
1967	September 15, 2008	global financial crisis	mid
2160	May 29, 2012	global financial crisis	end

Table 5.8: S&P: detected structure changes of weekly log return series

Table 5.8 presents the detected structure changes of the S&P 500 weekly log return series from our procedure with corresponding dates in real world. Though these ex-post economic interpretations are very speculative, we successfully identify the parameter changes that coincide with the external events.

Figure 5.13 shows the post-segmentation estimated volatility of the S&P 500 weekly log return series. We can see our detected structure changes well response to the estimated volatility as it reaches local maxima close to 1975, 1987 and 2009.

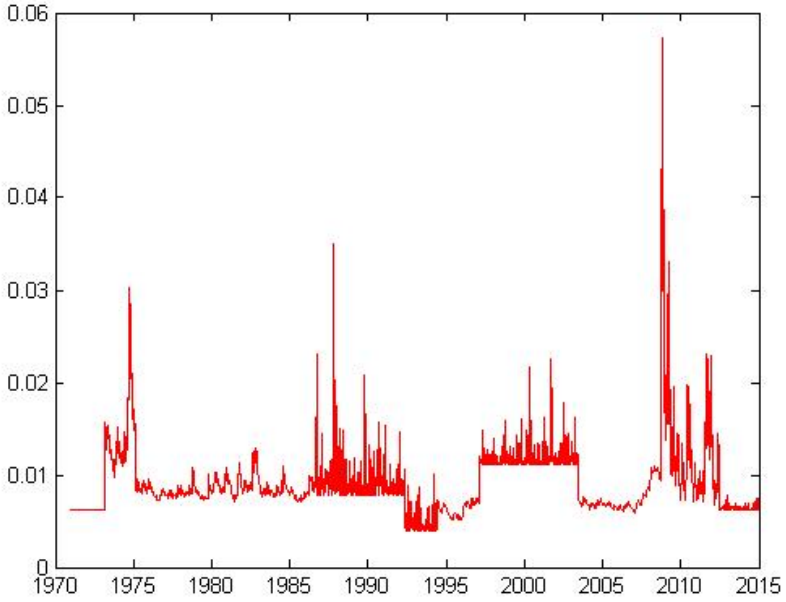

Figure 5.13: S&P 500: post-segmentation estimated volatility of weekly log return series

5.2.2 IBM Stock Return

In the last example, we illustrate the performance of our proposed procedure by analyzing the weekly log return series of the IBM stock from January 4, 1971 to December 29, 2014 as shown in Figure 5.14, which consists of $n = 2294$ data points. The data come from *Yahoo! Finance*. The sample mean, variance, skewness, and excess kurtosis of the return series are 6.50×10^{-4}, 2.27×10^{-4}, -0.1621, and 3.22, respectively. Thus, the data presents negative skewness and high excess kurtosis, the latter of which can be well fit by

GARCH models. The estimated coefficients by fitting the return series with the GARCH (1,1) model are which results in $\mu = 7.31 \times 10^{-4},\ \omega = 4.45 \times 10^{-6},\ \alpha = 0.924,\ \beta = 0.057,$ with the corresponding errors 2.81×10^{-4}, 7.58×10^{-7}, 0.0085 and 0.0063, respectively. Note that $\hat{\alpha} + \hat{\beta} = 0.981$, the return series shows high persistence in volatility.

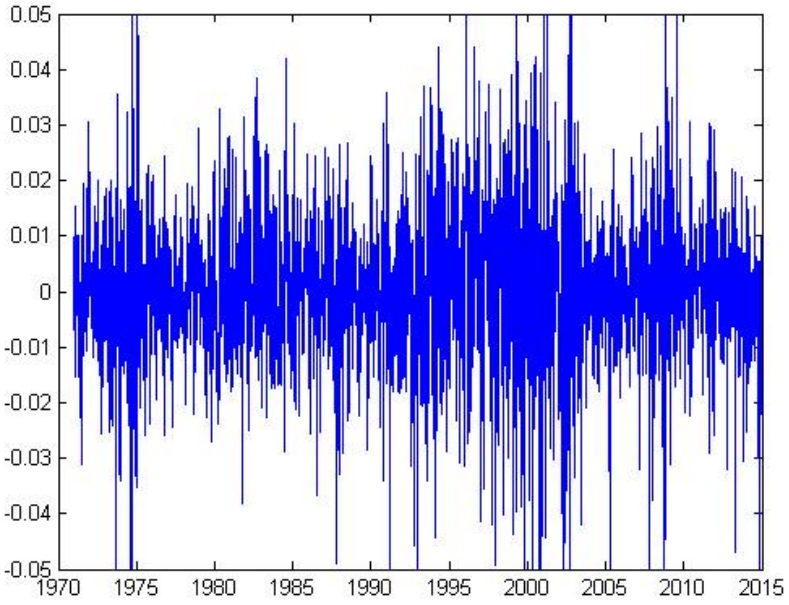

Figure 5.14: IBM: weekly log return series from 1999 to 2014

Since we are more interested in the patterns of the volatility, the error terms of the return series are studied by simply removing the drift term $\mu = 5.88 \times 10^{-4}$. We provide our semi-parameter estimates with $\widehat{p}$ with one update and the initial value $p_0 = 0.001$, the latter of which are shown in Figure 5.15. No blocks are used in our estimation procedure, where $\widehat{p} = 0.0043$ is given after the first iteration.

Figure 5.15 shows our semi-parameter estimates of the IBM stock weekly log return series from 1971 to 2014 with $p_0 = 0.001$, where the labels under the horizontal axis represent the start of the corresponding year and Figure 5.16 shows our segmentation results of the IBM stock weekly log return series with $\widehat{p}$ with one update and $p_0 = 0.001$. The solid lines represent the structure changes detected in both settings, where the dash lines represent the rest change-points detected with $\widehat{p}$ with one update and the dot lines

represent those with $p_0 = 0.001$ only. Here we choose the minimum distance $m = 100$ in our segmentation procedure and assume change-points given $\widehat{p}$ and $p_0 = 0.001$ are identical within the range of 4.

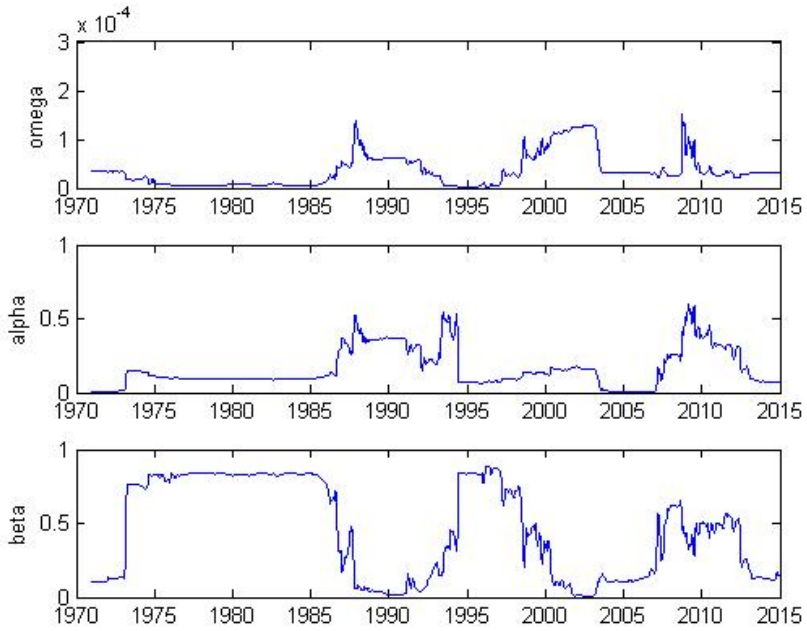

Figure 5.15: IBM: semi-parametric estimates of weekly log return series with p_0

We can see there exist eight detected structure changes from our procedure with $\widehat{p}$ with one update, where $t = 215$, 919, 1069, 1582, 1692, 1908, 2009 and 2110, while six change-points are detected with $p_0 = 0.001$, where $t = 214$, 1148, 1302, 1421, 2009, and 2160. Therefore, choice of the change-points probability p would greatly affect the number of detected structure changes. However, in both cases, it appears that the detected change-points are reasonable by comparing to the time plot of the return series.

Table 5.9 compares the post-segmentation maximized log-likelihood and BIC from different segmentation procedures. Here, "LLM" stands for the maximized log-likelihood and "N/A" stands for no segmentation procedure is applicable, i.e. the results come from the time-homogeneous GARCH(1,1) model. The binary segmentation algorithm (BSA) results in only one change-point at $t = 230$. In this case, we choose our procedure given $\widehat{p}$ with one update, as it provides the smallest BIC among all the segmentation results.

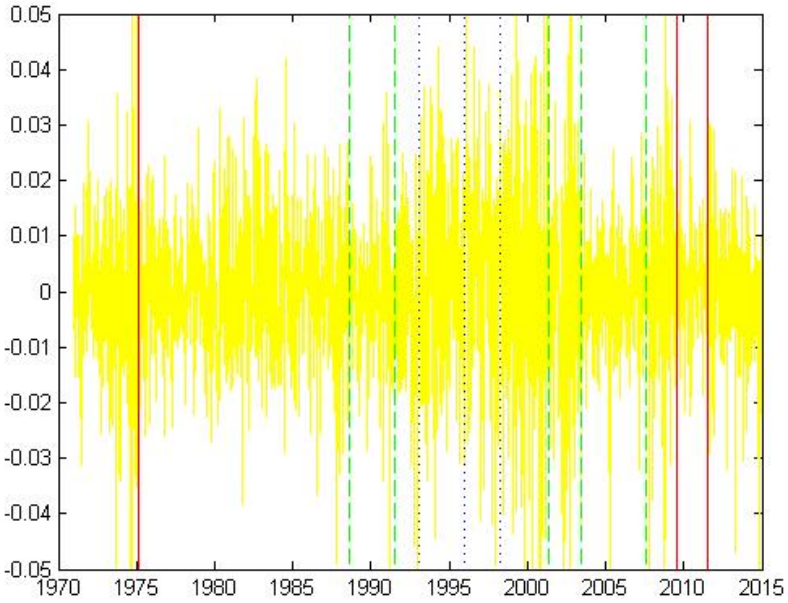

Figure 5.16: IBM: segmentation results of weekly log return series

Table 5.9: IBM: comparing post-segmentation maximized log-likelihood and BIC of weekly log return series

	UVB($\widehat{p}$)	UVB(p_0)	BSA	N/A
LLM	6594	6572	6534	6527
BIC	-13298	-13226	-13082	-13053

Change-point	Date	Event	Stage
215	February 10, 1975	oil crisis	end
919	August 8, 1988	black Monday	end
1069	June 24, 1991	early 1990s recession	mid
1582	April 23, 2001	internet bubble	end
1692	June 9, 2003	stock market downturn	end
1908	July 30, 2007	global financial crisis	start
2009	July 6, 2009	global financial crisis	end
2110	June 13, 2011	United States debt-ceiling crisis	start

Table 5.10: IBM: detected structure changes of weekly log return series

Table 5.10 presents the detected structure changes of the IBM stock weekly log return series from our procedure with corresponding dates in real world. Though these ex-post economic interpretations are very speculative, we successfully identify the parameter changes that coincide with the external events.

Figure 5.17 shows the post-segmentation estimated volatility of the IBM stock weekly log return series. We can see our detected structure changes well response to the estimated volatility as it reaches local maxima close to 1975, 2003 and 2009.

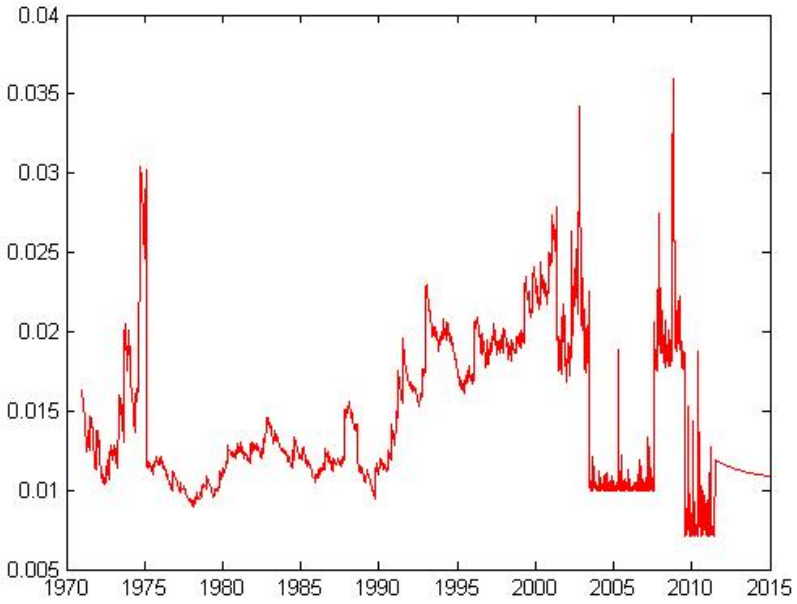

Figure 5.17: IBM: post-segmentation estimated volatility of weekly log return series

Chapter 6

Conclusion and Future Work

In the analysis of asset return series, we modeled change-points in the GARCH(1,1) setting and provided an estimation procedure for multiple parameter changes in GARCH models. By introducing the forward and backward filtration and combining them with Bayes' theorem, our estimation procedure has attractive statistical and computational properties and yields explicit recursive formulas to provide semi-parametric estimates for the piecewise constant parameters. Moreover, we proposed an expectation-maximization (EM) algorithm to estimate the change-points probability p in our model. Later, we also developed a segmentation procedure based on our semi-parametric estimates, where we first consider the individual impact of the estimated parameters to the unconditional variance to choose possible change-points, then applied the modified Bayesian information criterion (MBIC) and the conventional top down approach for model selection. Simulation studies were used to compare our performance to the binary segmentation algorithm (BSA) introduced by Galeano and Tsay(2010)[27] and the "oracle" estimates. The mean Euclidean error (EE), the Kullback–Leibler divergence (KL), the goodness of fit and the accuracy rate of the numbers of change-points detected are given. We implement our estimation and segmentation procedures to the daily and weekly log returns of the S&P 500 index and the IBM stock to give an insight how our estimation results coincide with the real financial crises.

In the end, we would like to discuss the following improvements, which can be made in the future work.

From our simulation studies and real data analysis, we can see the choice of the change-points probability p has a small impact on the semi-parametric estimates, but well affect the number of detected structure changes. Though we have proposed an expectation-maximization (EM) algorithm for the change-points probability p, the estimated $\widehat{p}$ has an upward bias due to the estimation errors and the persistence in the piecewise GARCH(1,1) subsequences. Also, in real practice, the definition of the true change-points probability p is quite vague as we cannot simply set the boundaries for determining structure changes. Therefore, we consider to use a trial and error procedure to choose from different p given a Bayesian information criterion (BIC). In future work, we would like to study a better way to choose the change-points probability p.

In our study, we have shown the minimum possible distance m can be determined by our interest and the length of the series n. However, if some of the detected structure changes are exactly m points away from each other, it either shows we overestimate the number of structure changes, or we can improve the accuracy of the locations by choosing a smaller m. In future work, we would like to study a better way to choose the minimum possible distance. Also, we may consider to leave more space without structure changes at the beginning and the ending of the series as simulation studies have shown the estimates are more volatile at the edges of the series.

We have proposed a modified Bayesian information criterion (MBIC) to uniformly determine the number and locations of structure changes. However, given certain range of the numbers of structure changes we expect and the length of the series n, we can modify the coefficients in the penalty term of our MBIC. Moreover, though we would like to have increasing power to detect structure changes at the edges of the subsequences, we are reluctant to see multiple structure changes detected close to each other. Therefore, we may even modify the penalty term itself in future work.

Bibliography

[1] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307-327.

[2] Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. Biometrika 42, 523-527.

[3] Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. J. Amer. Statist. Assoc. 55, 324-330.

[4] Hinkley, D. (1970). Inference about the change point in a sequence of random variables. Biometrika 57, 1-17.

[5] Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. Econometrica 61, 821-856.

[6] Shiryaev, A. N. (1970). On optimum methods in quickest detection problems. Theory Probab. Appl. 8, 22-46.

[7] Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. Appl. Statist. 41, 389-405.

[8] Bai, J. and Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica 66, 47-78.

[9] Qu, Z. and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. Econometrica 75, 459-502.

[10] Liu, J. S. and Lawrence, C. E. (1999). Bayesian inference on biopolymer models. Bioinformatics 15, 38-52.

[11] Wang, J. and Zivot, E. (2000). A Bayesian time series model of multiple structural changes in level, trend, and variance. J. Bus. Econ. Statist. 18, 374-386.

[12] Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean. Ann. Statist. 3, 98-108.

[13] Olshen, A. B., Venkatraman, E. S., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5, 557-572.

[14] Inclan, C. and Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of change of variance. J. Amer. Statist. Assoc. 89, 913-923.

[15] Chen, J. and Gupta, A. K. (1997). Testing and locating variance changepoints with application to stock prices. J. Amer. Statist. Assoc. 92, 739-747.

[16] Chib, S., Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. J. Econometrics 108, 281-316.

[17] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711-732.

[18] Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. J. Appl. Econometrics 18, 1-22.

[19] Xing, H., Sun, N. and Chen, Y. (2012). Credit rating dynamics in the presence of unknown structural breaks. J. Bank. Financ. 36, 78-89.

[20] Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. J. Econometrics 64, 307-333.

[21] Cai, J. (1994). A Markov model of switching regime ARCH. J. Bus. Econ. Statist. 12, 309-316.

[22] Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. J. Financ. Econ. 42, 27-62.

[23] Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stockmarket volatility. J. Bus. Econ. Statist. 15, 26-34.

[24] Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH models. Bernoulli 6, 513-539.

[25] Berkes, I., Gombay, E., Horvath, L. and Kokoszka, P. (2004). Sequential change-point detection in GARCH (p, q) models. Econometric Theory 20, 1140-1167.

[26] Andreou, E. and Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics. J. Appl. Econometrics 17, 579-600.

[27] Galeano, P. and Tsay, R. S. (2010). Shifts in individual parameters of a GARCH model. J. Financ. Econometrics 8, 122-153.

[28] Kim, C.J. (1994), Dynamic linear models with Markov-switching, J. Econometrics 60, 1-22.

[29] Zhang, N.R. and Siegmund, D.O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. Biometrics 63, 22-32.

[30] Bollerslev, T. and Wooldridge, J.M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Rev. 11, 143-172.

[31] Lai, T.L. and Xing, H. (2011). A simple Bayesian approach to multiple change-points. Statistica Sinica 21, 539-569.

[32] Hwang, S. and Pereira, P.L.V. (2006). Small sample properties of GARCH estimates and persistence. Europ. J. Finance 12, 473-494.

[33] Lavielle, M. (2005). Using penalized contrasts for the change-point problem. Signal Process. 85, (2005), 1501–1510.

[34] Kass, R.E. and Raftery, A.E. (1995). Bayes factors. J. Am. Statist. Assoc. 90, 773–795.