Review of Economics & Finance
Submitted on 24/09/2016
Article ID: 1923-7529-2017-02-44-17
Haipeng Xing, Hongsong Yuan, and Sichen Zhou

A Mixtured Localized Likelihood Method for GARCH Models with Multiple Change-points[1]

Prof. ***Haipeng Xing***
Department of Applied Mathematics and Statistics,
SUNY at Stony Brook, Stony Brook, 11790, U.S.A.

Prof. ***Hongsong Yuan*** (Corresponding author)
School of Information Management and Engineering,
Shanghai University of Finance and Economics, Shanghai, 200433, CHINA
Tel: +86 21 65901462; E-mail: yuan.hongsong@shufe.edu.cn

Dr. ***Sichen Zhou***
Worldquant LLC, Shanghai, 200040, CHINA

Abstract: This paper discusses GARCH models with multiple change-points and proposes a mixture localized likelihood method to estimate the piecewise constant GARCH parameters. The proposed method is statistically and computationally attractive as it synthesizes two degenerated and basic inference procedures. A bounded complexity mixture approximation, whose computational complexity is linear only, is also proposed for the estimates of time-varying GARCH parameters. These procedures are further applied to solve challenging problems such as inference on the number and locations of change-points that partition the unknown parameter sequence into segments of constant values. An illustrative analysis of the S&P500 index is provided.

Keywords: Localized likelihood; GARCH; Multiple change-points; Segmentation

JEL Classifications: C14, C22, C51

1. Introduction

Volatility modeling has been one of the most successful areas of research in econometrics and economic forecasting in the past decades. Since the seminal works of Engle (1982) and Bollerslev (1986), autoregressive conditionally heteroscedastic (ARCH) and generalized autoregressive conditionally heteroscedastic (GARCH) models have been widely used to study empirical characteristics of asset returns and exchange rates. Stylized fact of volatility persistence, which means the sum of estimated autoregressive parameters of a GARCH model is close to 1, is commonly observed. It has been documented that if the model parameters contain structural breaks or change-points, then the fitted models that overlook the parameter changes tend to exhibit long memory; see Diebold (1986), Perron (1989), Lamoureux and Lastrapes (1990), Mikosch and Starica (2004), Hillebrand (2005), and reference therein. Many authors, for example, Galeano and Tsay

[1] The first author's research is supported by the National Science Foundation under grant DMS-1206321 and DMS-1612501 at SUNY at Stony Brook. The second author's research is supported by the China Scholarship Council (File No. 201505990277).

(2010) and Lai and Xing (2013), further argued that the structural breaks or change-points in model parameters are usually associated with extraordinary economic and political events such as financial crisis, economic recessions, and changes of central banks' monetary policies. To model and make inference on such nonlinearity in GARCH models, Kokoszka and Leipus (2000) and Berkes *et al*. (2004a) introduced a cumulative-sum type statistic to estimate a single change-point in ARCH and GARCH models, respectively. Fearnhead and Liu (2007) used forward filtering recursions for the posterior distribution of the Markov chain to analyze on-line change-point problems. Galeano and Tsay (2010) proposed an iterative procedure to test change-point in individual parameters of a GARCH model and segment the time series accordingly.

In this paper, we consider a class of GARCH models with unknown multiple change-points and develop a mixtured localized likelihood method to estimate the time-varying GARCH parameters. To circumvent the complication by the intertwining of jumps in GARCH parameters and GARCH dynamics, we decompose the likelihood into a mixture of localized likelihood and develop a recursive scheme for the mixture probabilities. We then use a combination of forward and backward filtering recursions to compute the mixture probabilities; see Yao (1984), Lai *et al*. (2005), and Lai and Xing (2011, 2013). Extending such an approach and combining it with maximum likelihood procedures for GARCH models, we obtain an efficient estimation procedure. Since the probability of change-points is assumed at each time point in our model, the likelihood of our model at each time point can be written as a mixture of localized likelihood given the most recent change-points.

Besides the issue of estimating time-varying parameters, one may also be interested in segmenting the data. Common procedures for the segmentation problem rely on test statistics, such as the Lagrange multiplier (LM), the cumulative-sum (CUSUM), and likelihood based statistics, to segment the time series. In particular, the LM test statistics have been discussed by Andrews (1993) for general nonlinear models, and Chu (1995), Lundbergh and Teräsvirta (2002), and Galeano and Tsay (2010) have applied the LM statistic to GARCH models. The CUSUM statistics have been used to detect change-points in ARCH models by Kokoszka and Leipus (2000) and Andreou and Ghysels (2002, 2004). Fryzlewicz and Subba Rao (2014) propose a segmentation approach to detect multiple change-points in ARCH models. The likelihood-based statistics include the likelihood ratio statistics in Andrews (1993) and Lundbergh and Ter ä svirta (2002) and approximated quasi-likelihood score in Berkes *et al*. (2004b). These approaches involve maximizing the log-likelihood over the locations of the change-points and the piecewise constant parameters when a fixed number of change-points is assumed. Such an optimization problem can be solved by dynamic programming, and then combined with minimization of a model selection criterion to choose the number of change-points. To avoid the computational complexity, we in this paper further develop a segmentation procedure that combines a distance measure based on unconditional variances to locate the change-points, and a Bayesian information criterion to select the number of change-points. Asymptotic properties of this segmentation procedure is discussed.

The rest of the paper is organized as follows. In Section 2, we introduce our model assumption and develop a mixtured localized likelihood estimation procedure for time-varying GARCH parameters. We also propose a bounded complexity mixture algorithm to simplify the estimation procedure. In Section 3, we develop a segmentation procedure based on the unconditional variance of the time series. To show the performance of our inference and segmentation procedure, we present a simulation study to compare various features of our algorithm with an oracle algorithm in Section 4. We then apply our model and inference procedures to the S&P 500 weekly log return series in Section 5 and discuss the economic implications of estimated change-points. Finally, Section 6 provides some concluding remarks.

2. GARCH Models with Multiple Change-points

Consider a GARCH(1,1) model with time-varying coefficients $\boldsymbol{\theta}_t = (\omega_t, \alpha_t, \beta_t)^{\mathrm{T}}$

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

where ϵ_t are independent and identically distributed random variables with mean 0 and variance 1. For convenience, we assume that ϵ_t follows either the standard normal or a standardized Student t-distribution. Suppose the parameters $\boldsymbol{\theta}_t$ undergo occasional changes such that, for $t > 1$, the indicator variables $I_t \coloneqq 1_{\{\boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t-1}\}}$ are independent Bernoulli random variables with $P(I_t = 1) = p$. That is, with probability p, the GARCH parameters jump to a new set of values at each time point. Whenever there is such a jump in GARCH parameters at time t, we assume that the new value $\boldsymbol{\theta}_t$ satisfy the regularity conditions of GARCH model with constant coefficients given in Bollerslev and Wooldridge (1992), so that the quasi-maximum likelihood estimation of the new GARCH model exists.

We note that, the dynamics of time-varying GARCH parameters are actually not specified, except that they jump over time with probability p. Such specification avoids the concern whether one can correctly model the dynamics of GARCH parameters, and gives us more freedom. One potential concern on this setup is how the issue of volatility forecast should be addressed. We will first present an inference procedure in this section and discuss the forecasting issue in Section 3.

2.1 A mixtured localized likelihood for $\boldsymbol{\theta}_t$

Denote $Y_{i,j} = \{y_i, \ldots, y_j\}$ and let $Y_n = Y_{1,n}$. To estimate the GARCH parameter $\boldsymbol{\theta}_t$ given Y_n, we note that the parameter changes in $\boldsymbol{\theta}_t$ can occur before and after time t. To better explain this idea, we let $K_t = \max\{s\colon I_s = 1, s \leq t\}$ and $\widetilde{K}_t = \min\{s\colon I_s = 1, s > t\}$, i.e. K_t and $\widetilde{K}_t$ are the most recent change time before or after t, and $C_{i,j} = \{I_i = I_{j+1} = 1,\; I_{i+1} = \cdots = I_j = 0\}$ for $i \leq t \leq j$, i.e. $C_{i,j}$ is the event that there is no change from i to j and the most recent change time before and after t are i and j, respectively. Then the likelihood for $\boldsymbol{\theta}_t$ can be decomposed as a mixture of localized likelihood

$$L(\boldsymbol{\theta}_t;\, Y_n) = \sum_{1 \leq i \leq t \leq j \leq n} \xi_{ijt}\, L\big(\boldsymbol{\theta}_t; C_{i,j}, Y_{i,j}\big) \tag{2}$$

in which $L\big(\boldsymbol{\theta}_t; C_{i,j}, Y_{i,j}\big)$ is the likelihood of observations $Y_{i,j}$ for constant coefficient $\boldsymbol{\theta}_t$ given the event $C_{i,j}$, and $\xi_{ijt} = P(C_{ij} | Y_n)$ is the mixture probability. Making use of $\sum_{1 \leq i \leq t \leq j \leq n} \xi_{ijt} = 1$, one can show that the mixture probabilities ξ_{ijt} can be calculated by the recursions

$$\xi_{\mathrm{ijt}} \propto \xi_{ijt}^* = \begin{cases} p p_{i,t}, & \text{if } i \leq t = j \\ (1-p) p_{i,t} q_{t+1,j} f_{i,j} / (f_{i,t} f_{t+1,j}), & \text{if } i \leq t < j \end{cases} \tag{3}$$

where $\xi_{ijt} = \xi_{ijt}^* / P_t$, $P_t = p + \sum_{1 \leq i \leq t < j \leq n} \xi_{ijt}^*$, $f_{i,j}$ is the value of the likelihood function of $Y_{i,j}$ given $C_{i,j}$ and a constant $\boldsymbol{\theta}$, $p_{i,t} = P(K_t = i | Y_{1,t})$, and $q_{t+1,j} = P(\widetilde{K}_{t+1} = j | Y_{t+1,n})$. Note that the probabilities $p_{i,t}$ and $q_{t+1,j}$ can be further computed via recursions. One can also show that the recursion for $p_{i,t}$ is given by $p_{i,t} = p_{i,t}^* / \sum_{k=1}^{t} p_{k,t}^*$, where

$$p_{i,t}^* = \begin{cases} p f_{t,t}, & \text{if } i = t, \\ (1-p) p_{i,t-1} f_{i,t} / f_{i,t-1}, & \text{if } i < t. \end{cases} \tag{4}$$

Similarly, the probability $q_{t+1,j}$ can be computed recursively by

$q_{t+1,j} = q_{t+1,j}^* / \sum_{k=t+1}^{n} q_{t+1,k}^*$, where

$$q_{t,j}^* = \begin{cases} pf_{t,t}, & \text{if } j = t, \\ (1-p)q_{t+1,j}f_{t,j}f_{t+1,j}, & \text{if } j > t. \end{cases} \quad (5)$$

Given the mixtured localized likelihood (2), we construct an estimate of $\boldsymbol{\theta}_t$ in the following way. We denote $\widetilde{\boldsymbol{\theta}}_{i,j}$ the maximum likelihood estimate of $\boldsymbol{\theta}_t$ for the localized likelihood $L(\boldsymbol{\theta}_t; C_{i,j}, Y_{i,j})$, then in the spirit of (2), an estimator of $\boldsymbol{\theta}_t$ can be constructed as

$$\widetilde{\boldsymbol{\theta}}_t = \sum_{1\le i\le t\le j\le n} \xi_{ijt}\widetilde{\boldsymbol{\theta}}_{i,j}\ ,\ \ t = 1, \dots, n-1 \quad (6)$$

For the case $t = n$, since

$$l(\boldsymbol{\theta}_n; Y_n) = \sum_{1\le i\le n} p_{i,n} L(\boldsymbol{\theta}_n; K_n = i, Y_{i,n}),$$

we construct the estimator in the same spirit and obtain

$$\widetilde{\boldsymbol{\theta}}_n = \sum_{1\le i\le n} p_{i,n}\widetilde{\boldsymbol{\theta}}_{i,n} \quad (7)$$

Furthermore, since the evaluation of $p_{i,t}, q_{t+1,j}$ and ξ_{ijt} involves the likelihood $f_{i,j}$ of $Y_{i,j}$ in which $\boldsymbol{\theta}_t$ is an unknown constant, we replace $f_{i,j}$ by the maximized likelihood $L(\widetilde{\boldsymbol{\theta}}_{i,j}; C_{i,j}, Y_{i,j})$. Denote the probabilities $p_{i,t}, q_{t+1,j}$ and ξ_{jit} after such replacement by $\tilde{p}_{i,t}, \tilde{q}_{t+1,j}$ and $\tilde{\xi}_{jit}$, respectively. We then approximate (6) and obtain an estimator of $\boldsymbol{\theta}_t$

$$\widehat{\boldsymbol{\theta}}_t = \begin{cases} \sum_{1\le i\le t\le j\le n} \tilde{\xi}_{ijt}\widetilde{\boldsymbol{\theta}}_{i,j}\ , & \text{for } t = 1, \dots, n-1, \\ \sum_{1\le i\le n} \tilde{p}_{i,n}\widetilde{\boldsymbol{\theta}}_{i,n}\ , & \text{for } t = n. \end{cases} \quad (8)$$

Note that the estimates $\widetilde{\boldsymbol{\theta}}_t$ and $\widehat{\boldsymbol{\theta}}_t$ depend on p, but for notational simplicity we ignore the dependency for now.

2.2 Estimation of change-point probability p

The above estimation procedure assumes the change-point probability p is known. In practice, p is unknown and we can estimate it from the data. From the definition of $p_{i,t}^*$, we can express the log-likelihood function as a function of p:

$$l(p) = \sum_{t=1}^n \log f(y_t|Y_{1,t-1}) = \sum_{t=1}^n \log(\sum_{i=1}^t p_{i,t}^*) \quad (9)$$

in which $p_{i,t}^*$ is a function of p, and hence p can be estimated by maximizing (9). Another method is to use the expectation-maximization (EM) algorithm, which provides a much simpler structure of the log-likelihood $l_c(p)$ with the complete data $\{y_t, I_t, \boldsymbol{\theta}_t, 1 \le t \le n\}$,

$$l_c(p) = \sum_{t=1}^n \{\log f(\boldsymbol{\theta}_t|y_t) + 1_{\{\boldsymbol{\theta}_t\ne\boldsymbol{\theta}_{t-1}\}} \log p + 1_{\{\boldsymbol{\theta}_t\ne\boldsymbol{\theta}_{t-1}\}} \log(1-p)\} \quad (10)$$

The E-step of the EM algorithm involves the computing expectation of (10), and the M-step maximizes the expectation of (10), which yields an estimator of p,

$$\hat{p}_{new} = \frac{1}{n}\sum_{t=1}^n P(I_t = 1|Y_n, \hat{p}_{old}) = \frac{1}{n}\sum_{t=1}^n \sum_{i=1}^t P(C_{i,t}|Y_n, \hat{p}_{old}) = \frac{1}{n}\sum_{t=1}^n \hat{p}_{old}/P_t \quad (11)$$

2.3 Bounded complexity mixture approximation for $\widehat{\theta}_t$

The proposed estimator (8) has quadratic and linear complexity for $t < n$ and $t = n$, respectively. This results in rapid increasing computational complexity and memory requirements for estimating $\boldsymbol{\theta}_t$ as n increases. A natural idea to reduce the complexity and to facilitate the use of parallel algorithms for the recursions is to keep only a fixed number $M(p)$ of weights at every t. Following Lai and Xing (2011), we keep in (4) the most recent $m(p)$ weights $p_{j,n}$ and the largest $M(p) - m(p)$ of the remaining weights, where $1 \le m(p) < M(p)$. Specifically, let $\mathcal{K}_{t-1}(p)$ be the set of indices i for which $p_{i,t-1}$ is kept at stage $t-1$; thus, $\mathcal{K}_{t-1}(p) \supset \{t-1, \ldots, t-m(p)\}$. At stage t, define $p^*_{i,t}$ as in (4) for $i \in \{t\} \cup \mathcal{K}_{t-1}(p)$ and let i_t be the index not belonging to $\{t, \ldots, t-m(p)+1\}$ such that

$$p^*_{i_t,t} = \min\{p^*_{j,t}: j \in \mathcal{K}_{t-1}(p) \text{ and } j \le t-m(p)\},$$

choosing i_t to be the minimizer farthest from t if the above set has two or more minimizers. Define $\mathcal{K}_t(p) = \{t\} \cup (\mathcal{K}_{t-1}(p) - \{i_t\})$ and let

$$p_{i,t} = \left(p^*_{i,t} / \sum_{j \in \mathcal{K}_t(p)} p^*_{j,t} \right), \qquad i \in \mathcal{K}_t(p).$$

Similarly, to obtain a BCMIX approximation to $q_{t+1,j}$, let $\widetilde{\mathcal{K}}_{t+1}(p)$ denote the set of indices j for which $q_{t+1,j}$ in (3) is kept at stage $t+1$; thus $\widetilde{\mathcal{K}}_{t+1}(p) \supset \{t+1, \ldots, t+m\}$. At stage t, define $q^*_{t,j}$ as in (5) for $j \in \{t\} \cup \widetilde{\mathcal{K}}_{t+1}(p)$ and let j_t be the index not belonging to $\{t, \ldots, t+m(p)-1\}$ such that

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

choosing j_t to be the minimizer farthest from t if the above set has two or more minimizers. Define $\widetilde{\mathcal{K}}_t(p) = \{t\} \cup \left(\widetilde{\mathcal{K}}_{t+1}(p) - \{j_t\}\right)$ and let $q_{t,j} = \left(q^*_{t,j} / \sum_{j \in \widetilde{\mathcal{K}}_t(p)} q^*_{t,j}\right), j \in \widetilde{\mathcal{K}}_t(p)$, which yields a BCMIX approximation to $q_{t,j}$ in (5).

The BCMIX approximation to (8) can be obtained by replacing $1 \le i \le t \le j \le n$ with $i \in \mathcal{K}_t(p)$ and $j \in \widetilde{\mathcal{K}}_{t+1}(p)$, i.e.,

$$\widehat{\boldsymbol{\theta}}_t(p) = \sum_{i \in \mathcal{K}_t(p), j \in \widetilde{\mathcal{K}}_{t+1}(p)} \tilde{\xi}_{ijt} \widetilde{\boldsymbol{\theta}}_{i,j} \qquad (12)$$

in which $\tilde{\xi}_{ijt} = \xi^*_{ijt}/\tilde{P}_t$, $\tilde{P}_t = p + \sum_{i \in \mathcal{K}_t(p), j \in \widetilde{\mathcal{K}}_{t+1}(p)} \xi^*_{ijt}$ and ξ^*_{ijt} given by (3) for $i \in \mathcal{K}_t(p)$ and $j \in \widetilde{\mathcal{K}}_{t+1}(p)$. Here we have restored the dependency on p in the notation. Note that the computational complexity of the BCMIX approximation (12) is linearly bounded. We can show the following Proposition 1, which assumes conditions (C1) and (C2) that are similar to those of Yao (1988) for piecewise constant normal means:

(C1) The true change-points occur at $t_1^{(n)} < \cdots < t_k^{(n)}$ such that $\liminf_{n \to \infty} n^{-1}\left(t_i^{(n)} - t_{i-1}^{(n)}\right) > 0$ for $1 \le i \le k+1$, with $t_0^{(n)} = 0$ and $t_{k+1}^{(n)} = n$.

(C2) Let $U_{t_i^{(n)}} = \omega_{t_i^{(n)}}/(1 - \alpha_{t_i^{(n)}} - \beta_{t_i^{(n)}})$ be the unconditional variance at $t_i^{(n)}$ for $1 \le i \le k$. There exists $\delta > 0$, which does not depend on n, such that $\min_{1 \le i \le k}\left|U_{t_i^{(n)}}/U_{t_{i-1}^{(n)}} - 1\right| \ge \delta$ for all large n.

Proposition 1. *Assume (C1), (C2) and that* $m(p) \sim |log\, p|^{1+\epsilon}$ *and* $1 \le M(p) - m(p) = O(1)$ *as* $p \to 0$, *for some* $\epsilon > 0$. *Then*

$$\max_{1 \le t \le n:\, \min_{1 \le i \le k}\left|t - t_i^{(n)}\right| \ge m(p)} ||\widehat{\boldsymbol{\theta}}_t(p) - \widehat{\boldsymbol{\theta}}_t|| \xrightarrow{P} 0 \text{ as } n \to \infty$$

uniformly in $a_1/n \le p \le a_2/n$.

As noted in Section 2.2, the parameter p can be estimated by EM algorithm. We can use the BCMIX approximation to replace $\sum_{i=1}^{t} p^*_{i,t}$ by $\sum_{i \in \mathcal{K}_t(p)} p^*_{i,t}$ in the log-likelihood function (9) and thereby estimate p by iterating along the EM algorithm. Putting this estimated $\hat{p}$ in $\widehat{\boldsymbol{\theta}}_t(p)$ in (12) yields the empirical Bayes estimator $\widehat{\boldsymbol{\theta}}_t(\hat{p})$, which by Proposition 1 also satisfies the consistency property

$$\max_{1\le t\le n:\ \min_{1\le i\le k}\left|t-t_i^{(n)}\right|\ge m(p)} \left|\left|\widehat{\boldsymbol{\theta}}_t(\hat{p})-\widehat{\boldsymbol{\theta}}_t\right|\right| \xrightarrow{P} 0 \ as\ n\to\infty. \qquad (13)$$

3. Segmentation and Volatility Forecast

One important problem in multiple change-point analysis is to estimate the location of change-points via maximization of log-likelihood. As explained in Section 1, this optimization problem can be solved by dynamic programming, which involves two layers of optimization. The inner layer of the optimization maximizes the log-likelihood of k change-points,

$$l_n(k)=\sup_{1\le t_1<\cdots<t_k<n,\boldsymbol{\theta}^{(1)},\dots,\boldsymbol{\theta}^{(k+1)}}\sum_{j=1}^{k+1}\sum_{t=t_{j-1}}^{t_j-1}\log f_{\boldsymbol{\theta}^{(j)}}(y_t) \qquad (14)$$

in which $t_0=1$ and $t_{k+1}-1=n$. The outer layer maximizes $l_n(k)$ together with a penalty function of k. The relative simplicity of our estimator for $\boldsymbol{\theta}_t$ in our model opens up new possibilities in resolving this challenging problem and determining the segmentation. In this section we use an appropriately chosen parameter p in our model to tackle this problem. Note that one usually assumes that k is small relative to n and that adjacent change-points are sufficiently far apart so that the segments are identifiable except for relatively small neighborhoods of the change-points. Motivated by this assumption, we restrict p to an interval $[c_1/n, c_2/n]$ in our approach so that the arrival of change-points is approximately Poisson.

Denote $\widehat{\boldsymbol{\theta}}_t$ as our BCMIX estimator of $\boldsymbol{\theta}_t$. A natural idea is to compare the unconditional variances $E(\sigma_t^2)$ at adjacent time points. Since $E(\sigma_t^2)=\omega_t/(1-\alpha_t-\beta_t)$, a first order Taylor approximation of $\log[E(\sigma_{t+1}^2)/E(\sigma_t^2)]$ can be expressed as

$$\log[E(\sigma_{t+1}^2)/E(\sigma_t^2)]\approx(\widehat{\omega}_{t+1}-\widehat{\omega}_t)/\widehat{\omega}_t+[(\hat{\alpha}_{t+1}-\hat{\alpha}_t)+(\hat{\beta}_{t+1}-\hat{\beta}_t)]/(1-\hat{\alpha}_t-\hat{\beta}_t) \qquad (15)$$

This motivates us to consider the following distance measure of $E(\sigma_t^2)$ at times t and $t+1$,

$$\Delta_t := |\widehat{\omega}_{t+1}-\widehat{\omega}_t|/\widehat{\omega}_t+\left(|\hat{\alpha}_{t+1}-\hat{\alpha}_t|+\left|\hat{\beta}_{t+1}-\hat{\beta}_t\right|\right)/(1-\hat{\alpha}_t-\hat{\beta}_t) \qquad (16)$$

Assume that the minimal distance of two adjacent change-points is d_0, and there are at most K change-points. We estimate the change-times sequentially by making use of $\{\Delta_t: d_0<t\le n-d_0\}$. Let $\hat{t}_1$ be the maximizer of Δ_t over $d_0<t\le n-d_0$. After $\hat{t}_1,\dots,\hat{t}_{j-1}$ have been defined, we define $\hat{t}_j$ as the maximizer of Δ_t over t that lies outside the d_0-neighborhood of $\hat{t}_i$ for $1\le i\le j-1$, i.e.,

$$\Delta_{\hat{t}_j}=\max\{\Delta_t: d_0<t\le n-d_0, \min_{1\le i\le j-1}|t-\hat{t}_i|\ge d_0\} \qquad (17)$$

The procedure terminates with k change-points whenever there is no points outside the d_0-neighborhoods of $\{\hat{t}_i: 1\le i\le k\}$, or the maximum number of change-points K is reached. Note that the estimates $\hat{t}_j$ of the locations of the change-points in (17) do not depend on k. Under the model of k change-points, we can take $\hat{t}_1,\dots,\hat{t}_k$ and order them as $\hat{t}_{(1),k}<\cdots<\hat{t}_{(k),k}$ to provide estimates of the k change-points. To simplify notations, we use $l_1,\dots,l_k$ to represent $\hat{t}_{(1),k},\dots,\hat{t}_{(k),k}$ in the sequel. We know that f_{l_{j-1},l_j-1} is the maximum likelihood over the estimated segment $[l_{j-1}, l_j-1]$, $1\le j\le k+1$, where $l_0=1$ and $l_{k+1}-1=n$. This yields the following approximation to (14):

$$\Lambda_{\mathrm{n}}(k)=\sum_{j=1}^{k+1}\log f_{l_{j-1},l_j-1} \qquad (18)$$

With an upper bound K on the number k of change-points, we propose to estimate k by

$$\hat{k}_n = \text{argmax}_{1\le k\le K}\{\Lambda_n(k) - (k+1)C_n\} \tag{19}$$

where C_n is the common penalty for each segment that satisfies

$$C_n \to \infty \text{ and } \frac{C_n}{n} \to 0 \text{ as } n \to \infty \tag{20}$$

We have the following proposition 2:

Proposition 2. *Under (20) and the assumptions of Proposition 1, together with* $d_0 \sim Cm(p)$ *for some* $C > 0$, $\hat{k}_n \xrightarrow{P} k$.

The penalty function C_n can be chosen in different ways. For instance, a Bayesian information criterion (BIC) criterion is obtained if $(k+1)C_n = 2k\log n$. In the spirit of a modified BIC that was proposed by Siegmund and Zhang (2006) for Guassian mean shift models where $(k+1)C_n = 3k\log n + \sum_{i=1}^{k+1}\log(l_i - l_{i-1})$, we propose our version of the modified BIC as follows:

$$(k+1)C_n = k\log(n) + \frac{k}{k+2}\log(d_0)\left(\sum_{i=1}^{k+1}\log(l_i - l_{i-1}) - (k+1)\log\left(\frac{n}{k+1}\right)\right) \tag{21}$$

In other words, we have replaced $\sum_{i=1}^{k+1}\log(l_i - l_{i-1})$ in Siegmund and Zhang (2006) with

$$\sum_{i=1}^{k+1}\log(l_i - l_{i-1}) - (k+1)\log\left(\frac{n}{k+1}\right)$$

which reaches its maximum 0 when the change-points are evenly distributed. The reason for the multiplier $\frac{k}{k+2}\log(d_0)$ is that structure changes tend to spread evenly with a large minimum distance d_0. Moreover, Lavielle (2005) proposed to replace C_n in (19) by $\delta_0 C_n$ for change-point problems, where δ_0 is a parameter chosen by the user. In our study, we follow his suggestion and use $\delta_0 = 1/2$.

We shall point out that, since the dynamics of the time-varying GARCH parameters are specified non-parametrically, the segmentation procedure here also serves for the purpose of forecasting. Actually, after the data are segmented, one could use the segmented data to estimate a GARCH model with constant coefficients. For forecasting purposes a constant-parameter GARCH model needs to be estimated on the last segmentation only. Then the forecasting issue can be addressed by classical discussions for GARCH models with constant coefficients.

4. Simulation Studies

To evaluate the performance of our proposed inference procedure, we conduct extensive simulation studies in this section. To measure the difference between the estimated and true parameters, we consider three types of distance measures here. We define the mean squared error (MSE) for estimate $\{\boldsymbol{\theta}_t : 1 \le t \le n\}$ as

$$\text{MSE} \coloneqq n^{-1}\sum_{i=1}^{n}\left|\left|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t\right|\right|_2 \tag{22}$$

We also consider the Kullback-Leibler (KL) divergence or the relative entropy, which is estimated by

$$\text{KL} \coloneqq \frac{1}{2n}\sum_{t=1}^{n}\left(\frac{y_t^2}{\sigma_t^2} - \frac{y_t^2}{\hat{\sigma}_t^2} + \log\frac{\sigma_t^2}{\hat{\sigma}_t^2}\right) \tag{23}$$

and a goodness-of-fit measure

$$\mathrm{GOF}(\boldsymbol{\theta}_t) \coloneqq \frac{1}{n}\sum_{t=1}^{n}\frac{y_t^2}{\sigma_t^2} \qquad (24)$$

Note that, given the true parameter $\boldsymbol{\theta}_t$, $n\mathrm{GOF}(\boldsymbol{\theta}_t)$ follows a chi-square distribution with n degrees of freedom. Thus, should our estimate $\widehat{\boldsymbol{\theta}}_t$ be close to $\boldsymbol{\theta}_t$, $\mathrm{GOF}(\widehat{\boldsymbol{\theta}}_t)$ and its standard deviation are close to 1 and $\sqrt{2/n}$ respectively.

4.1 Performance of the BCMIX estimator

We consider two types of parameter changes here. The first is that all parameters change at the same time, and the second is that change-point only occurs in one parameter, while other parameters remain constant.

Example 1. Assuming all parameters change simultaneously, we consider the following six scenarios. Among them, the first scenario has no change-points, and the second and the third scenarios have two and three change-points, respectively. The last three scenarios are similar to our model except $\boldsymbol{\theta}_t$ are generated randomly.

(S1) $\boldsymbol{\theta}_t = (0.4, 0.2, 0.7)^{\mathrm{T}}$ for $1 \le t \le 1000$.

(S2) $\boldsymbol{\theta}_t = (0.4, 0.2, 0.7)^{\mathrm{T}}$ for $1 \le t \le 300$, $\boldsymbol{\theta}_t = (0.2, 0.6, 0.2)^{\mathrm{T}}$ for $301 \le t \le 700$, and $\boldsymbol{\theta}_t = (0.8, 0.3, 0.6)^{\mathrm{T}}$ for $701 \le t \le 1000$.

(S3) $\boldsymbol{\theta}_t = (0.4, 0.2, 0.7)^{\mathrm{T}}$ for $1 \le t \le 200$, $\boldsymbol{\theta}_t = (0.2, 0.6, 0.2)^{\mathrm{T}}$ for $201 \le t \le 500$, $\boldsymbol{\theta}_t = (0.8, 0.3, 0.6)^{\mathrm{T}}$ for $501 \le t \le 700$, and $\boldsymbol{\theta}_t = (0.6, 0.5, 0.3)^{\mathrm{T}}$ for $701 \le t \le 1000$.

(S4) Change-points occur randomly with $p = 0.001$. When a change-point occurs at time t, $\boldsymbol{\theta}_t$ follows a truncated uniform distribution on $(0, 1)^3$ such that σ_t^2 is covariance stationary. For identification purposes, we assume a minimum jump size of 0.1 for each parameter ω_t, α_t and β_t at the change-points.

(S5) Same setting as in S4 except $p = 0.002$.

(S6) Same setting as in S4 except $p = 0.005$.

Example 2. We now consider the case of change-points in individual parameters in the following six scenarios, which assume one and two change-points in ω_t, α_t and β_t, respectively.

(S7) $\boldsymbol{\theta}_t = (0.8, 0.3, 0.5)^{\mathrm{T}}$ for $1 \le t \le 500$ and $\boldsymbol{\theta}_t = (1.0, 0.3, 0.5)^{\mathrm{T}}$ for $501 \le t \le 1000$.

(S8) $\boldsymbol{\theta}_t = (0.8, 0.3, 0.3)^{\mathrm{T}}$ for $1 \le t \le 500$ and $\boldsymbol{\theta}_t = (0.8, 0.5, 0.3)^{\mathrm{T}}$ for $501 \le t \le 1000$.

(S9) $\boldsymbol{\theta}_t = (0.8, 0.1, 0.3)^{\mathrm{T}}$ for $1 \le t \le 500$ and $\boldsymbol{\theta}_t = (0.8, 0.1, 0.5)^{\mathrm{T}}$ for $501 \le t \le 1000$.

(S10) $\boldsymbol{\theta}_t = (0.8, 0.3, 0.3)^{\mathrm{T}}$ for $1 \le t \le 300$, $\boldsymbol{\theta}_t = (1.0, 0.3, 0.3)^{\mathrm{T}}$ for $301 \le t \le 700$, and $\boldsymbol{\theta}_t = (1.2, 0.3, 0.3)^{\mathrm{T}}$ for $701 \le t \le 1000$.

(S11) $\boldsymbol{\theta}_t = (1.0, 0.1, 0.3)^{\mathrm{T}}$ for $1 \le t \le 300$, $\boldsymbol{\theta}_t = (1.0, 0.3, 0.3)^{\mathrm{T}}$ for $301 \le t \le 700$, and $\boldsymbol{\theta}_t = (1.0, 0.5, 0.3)^{\mathrm{T}}$ for $701 \le t \le 1000$.

(S12) $\boldsymbol{\theta}_t = (1.0, 0.3, 0.1)^{\mathrm{T}}$ for $1 \le t \le 300$, $\boldsymbol{\theta}_t = (1.0, 0.3, 0.3)^{\mathrm{T}}$ for $301 \le t \le 700$, and $\boldsymbol{\theta}_t = (1.0, 0.3, 0.5)^{\mathrm{T}}$ for $701 \le t \le 1000$.

Table 1 and Table 2 below summarize the MSE, KL and GOF of the BCMIX, oracle (labeled "Oracle"), binary segmentation (labeled "BS"), and standard GARCH (labeled "Standard") estimators with their standard errors in parentheses for S1-S12. While GOF does not differ too much among different estimators, MSE and KL can reveal some information. When the true number of change-points is less than or equal to 1 (namely, scenarios S1, S7, S8, S9), the standard GARCH model tend to do relatively well. When the true number of change-points becomes larger than 1, standard GARCH performs poorly, and BCMIX outperforms both standard GARCH and binary segmentation in most scenarios.

Table 1. Performance of the BCMIX(15, 10) and competing estimators for scenarios S1-S6

Scenario	Method	MSE	KL	GOF	Aver(Δ_t)
S1 0.000	BCMIX	0.556 (1.07e-2)	0.023 (3.72e-4)	0.994 (7.00e-4)	0.024 (6.8e-3)
	Oracle	0.065 (1.30e-3)	0.001 (4.26e-5)	1.003 (3.00e-4)	0
	BS	0.416 (9.48e-2)	0.095 (2.36e-2)	1.010 (1.60e-2)	NA
	Standard	0.353 (2.11e-2)	0.066 (1.22e-2)	1.028 (1.54e-2)	NA
S2 0.016	BCMIX	0.480 (7.50e-3)	0.024 (4.21e-4)	1.003 (1.90e-3)	0.031 (6.2e-3)
	Oracle	0.100 (6.00e-4)	0.003 (1.29e-4)	1.019 (1.50e-3)	0.014 (1.7e-3)
	BS	0.427(8.28e-2)	0.144(2.47e-2)	1.002(4.80e-3)	NA
	Standard	0.533 (3.90e-2)	0.105(2.08e-2)	1.009(1.12e-2)	NA
S3 0.021	BCMIX	0.423 (6.00e-3)	0.025 (3.52e-4)	0.998 (1.40e-3)	0.033 (8.2e-3)
	Oracle	0.101 (5.00e-4)	0.003 (1.10e-4)	1.019 (1.40e-3)	0.019 (2.0e-3)
	BS	0.478 (6.43e-2)	0.131(2.40e-2)	1.000 (2.00e-3)	NA
	Standard	0.463 (3.43e-2)	0.130 (2.37e-2)	1.000 (1.70e-3)	NA
S4 0.004	BCMIX	0.228 (4.90e-3)	0.014 (3.41e-4)	0.985 (6.00e-4)	0.011 (3.7e-3)
	Oracle	0.075 (1.20e-3)	0.001 (5.90e-5)	1.002 (3.00e-4)	0.004 (4.5e-3)
	BS	0.278 (2.38e-1)	3.11e-4(1.26e-2)	1.002(1.42e-2)	NA
	Standard	0.354 (2.73e-1)	0.014(2.32e-2)	1.006(2.47e-2)	NA
S5 0.007	BCMIX	0.261 (5.10e-3)	0.016 (3.63e-4)	0.987 (9.00e-4)	0.015 (5.7e-3)
	Oracle	0.078 (1.00e-3)	0.002 (7.56e-5)	1.005 (7.00e-4)	0.007 (4.6e-3)
	BS	0.318(2.20e-1)	0.020(1.46e-2)	1.004(1.75e-2)	NA
	Standard	0.456 (2.67e-1)	0.025(2.50e-2)	1.011(3.44e-2)	NA
S6 0.015	BCMIX	0.453 (1.67e-2)	0.019 (5.20e-4)	0.991 (1.90e-3)	0.019 (5.5e-3)
	Oracle	0.082 (8.00e-4)	0.002 (9.68e-5)	1.012 (1.20e-3)	0.015 (7.4e-3)
	BS	0.523 (2.15e-1)	0.020 (2.33e-2)	1.009 (2.95e-2)	NA
	Standard	0.572 (2.15e-1)	0.039 (2.61e-2)	1.019 (4.48e-2)	NA

In each of these twelve scenarios, we run 500 simulations and in each simulation, we first estimate the change-point probability p via the EM algorithm and then compute the BCMIX estimator $\widehat{\boldsymbol{\theta}}_t$. Since the BCMIX estimator involves two tuning parameters $M(p)$ and $m(p)$, we try different combinations of $(M(p), m(p))$ = (20, 15), (15, 10), and (10, 7). As we didn't find significant difference among the combinations, we only report the result of $(M(p), m(p))$ = (15, 10) in the sequel. To demonstrate the performance, the BCMIX estimator is compared with

1. an oracle estimator which assumes the change-times are known;
2. the binary segmentation procedure in Galeano and Tsay (2010) that is based on the Lagrange multiplier (LM) test, where the significance level is 5% for the LM test;
3. a standard GARCH model that assumes no change-points.

Table 2. Performance of the BCMIX(15, 10) and competing estimators for scenarios S7-S12

Scenario	Method	MSE	KL	GOF	Aver(Δ_t)
S7 0.0004	BCMIX	0.311 (6.10e-3)	0.013 (2.25e-4)	0.991 (4.00e-4)	0.007 (1.7e-3)
	Oracle	0.083 (9.00e-4)	0.002 (4.32e-5)	1.002 (2.00e-4)	0.001 (1.2e-4)
	BS	0.267(5.77e-2)	0.025(6.00e-3)	1.000(6.83e-4)	NA
	Standard	0.216 (3.39e-2)	0.020(5.10e-3)	1.000(4.38e-4)	NA
S8 0.0008	BCMIX	0.208 (2.90e-3)	0.012 (2.48e-4)	0.987 (5.00e-4)	0.011 (4.3e-3)
	Oracle	0.091 (8.00e-4)	0.002 (3.99e-5)	1.001 (1.00e-4)	0.001 (1.1e-4)
	BS	0.266 (8.05e-2)	0.008 (4.10e-3)	1.000(8.75e-4)	NA
	Standard	0.166 (3.61e-2)	1.36e-4 (2.30e-3)	1.000 (3.58e-4)	NA
S9 0.0008	BCMIX	0.201 (4.00e-3)	0.009 (1.61e-4)	0.991 (4.00e-4)	0.012 (2.5e-3)
	Oracle	0.083 (9.00e-4)	0.002 (4.55e-5)	1.001 (1.00e-4)	0.001 (1.3e-4)
	BS	0.352 (8.06e-2)	0.057(1.49e-2)	1.001(1.20e-3)	NA
	Standard	0.311 (7.60e-2)	0.040(1.30e-2)	1.000(5.99e-4)	NA
S10 0.0003	BCMIX	0.185 (2.30e-3)	0.010 (1.71e-4)	0.988 (4.00e-4)	0.015 (3.3e-3)
	Oracle	0.089 (7.00e-4)	0.002 (4.33e-5)	1.002 (1.00e-4)	8e-4 (2.0e-4)
	BS	0.298(8.37e-2)	0.036(1.02e-2)	1.000(1.80e-3)	NA
	Standard	0.292 (4.60e-2)	0.034(1.01e-2)	1.000(1.20e-3)	NA
S11 0.0005	BCMIX	0.201(2.70e-3)	0.010 (1.63e-4)	0.987 (4.00e-4)	0.012 (2.9e-3)
	Oracle	0.091 (7.00e-4)	0.002 (4.52e-5)	1.002 (2.00e-4)	6e-4 (1.1e-4)
	BS	0.271(4.99e-2)	0.012(5.70e-3)	1.000(6.98e-4)	NA
	Standard	0.287 (4.34e-2)	0.009(5.70e-3)	1.000(5.04e-4)	NA
S12 0.0003	BCMIX	0.241 (3.70e-3)	0.012 (1.76e-4)	0.989 (4.00e-4)	0.007 (1.5e-3)
	Oracle	0.088 (7.00e-4)	0.002 (4.16e-5)	1.003 (2.00e-4)	5e-4 (7.2e-5)
	BS	0.396(2.19e-1)	0.026(5.30e-3)	1.000(2.50e-3)	NA
	Standard	0.579 (2.31e-1)	0.018 (5.80e-3)	0.999(2.60e-3)	NA

Since we will use the distance measure (16) for segmentation, we also show whether (16) could capture parameter changes rather than time-varying variances. Define $\text{Aver}(\Delta_t) = \sum_{t=1}^{n-1} \Delta_t \,/(n-1)$. We show the value of this measure for the true parameters in S1-S12 in the first column of Table 1 and Table 2. We find that, although Δ_t is dependent on the values of GARCH parameters, it can capture parameter changes. We also compute $\text{Aver}(\Delta_t)$ for the BCMIX and oracle estimators, and show them in the last column of Table 1 and Table 2, as we will use them for segmentation purpose in the next section.

4.2 Performance of the segmentation procedure

We next evaluate the performance of our segmentation procedure. For scenarios S1-S12, we use our and binary segmentation (BS) procedures to segment the data and estimate the locations of change-points. For scenarios S1-S6 in which change-points are simultaneous for all parameters, we compute the estimation errors for the number of change-points. Let k be the true number of change-points in each scenario, and $\hat{k}$ are the estimated number of changes from segmentation procedures. We denote $\varphi := \hat{k} - k$ as the estimation error and $\%(|\varphi| = i)$ the frequency that the estimation error has absolute value i. Table 3 shows some summary statistics for φ in our and binary segmentation procedures. We can see that our procedure performs better than the binary segmentation procedure. Furthermore, it seems that the binary segmentation procedure tends to underestimate the number of change-points more than ours when multiple change-point are present.

Table 3. The estimation errors φ for scenarios S1-S6

Scenario	Method	$\boldsymbol{\varphi}$	$\%(\lvert\boldsymbol{\varphi}\rvert = \mathbf{0})$	$\%(\lvert\boldsymbol{\varphi}\rvert = \mathbf{1})$	$\%(\lvert\boldsymbol{\varphi}\rvert = \mathbf{2})$
S1	BCMIX	0.044 (0.011)	0.964	0.028	0.008
	BS	0.116 (0.015)	0.890	0.104	0.006
S2	BCMIX	0.392 (0.030)	0.686	0.244	0.054
	BS	-1.336 (0.045)	0.252	0.060	0.688
S3	BCMIX	-0.004 (0.032)	0.600	0.370	0.028
	BS	-1.248 (0.059)	0.320	0.242	0.176
S4	BCMIX	-0.100 (0.028)	0.786	0.158	0.054
	BS	-0.110 (0.029)	0.754	0.194	0.048
S5	BCMIX	-0.286 (0.040)	0.604	0.276	0.096
	BS	-0.454 (0.045)	0.570	0.262	0.122
S6	BCMIX	-0.648 (0.053)	0.436	0.310	0.178
	BS	-1.498 (0.062)	0.252	0.270	0.234

For scenarios S7-S12 that involves one and two change-points in individual parameters, we show the estimated number of change-points and their frequencies in Table 4. We notice that our procedure still performs better than the binary segmentation procedure except in Scenario 9.

To test the out-of-sample performance of a procedure, we use a measure based on volatility forecasts. Let σ_{T+t}^2 be the t-step ahead variance forecast for the GARCH model with the true coefficients of the last segmentation, and let $\hat{\sigma}_{T+t}^2$ be the forecasted variance for some estimation procedure, where T is the length of the time series. We consider multi-step ahead prediction, and define the average absolute deviance as

$$D := \frac{1}{h}\sum_{t=1}^{h}\left|\hat{\sigma}_{T+t}^2 - \sigma_{T+t}^2\right| \tag{25}$$

where h is pre-specified number, which we set as 10. Notice that high persistence might lead to extraordinary large values of variance forecasts. To alleviate the influence of such extreme cases, we report the median rather than the mean, of D over all 500 simulation runs. The results are summarized in Table 5, where one remarkable fact is that standard GARCH tend to forecast ultra high volatilities, especially for S2 and S3 where both the true number of change-points and the magnitude of the parameter change are relatively large. In these cases, ignoring change-points generally produce high persistent GARCH models, where the sum of the estimates $\hat{\alpha}$ and $\hat{\beta}$ is too close to 1. As a result, in multi-step ahead prediction, as the number of forecasting steps increases, the predicted volatility converges to $\hat{\omega}/(1-\hat{\alpha}-\hat{\beta})$, a potentially very large number. The prediction performances of BCMIX and binary segmentation are comparable in most scenarios.

Table 4. Estimated change-points for scenarios S7-S12

Scenario	Method	$\hat{k}=0$	$\hat{k}=1$	$\hat{k}=2$	$\hat{k}=3$	$\hat{k}$
S7 $k=1$	BCMIX	0.316	0.312	0.2	0.11	1.302 (0.055)
	BS	0.8	0.184	0.016	0	0.216 (0.020)
S8 $k=1$	BCMIX	0.262	0.33	0.222	0.122	1.41 (0.054)
	BS	0.768	0.21	0.02	0.002	0.256 (0.022)
S9 $k=1$	BCMIX	0.05	0.564	0.276	0.092	1.468 (0.037)
	BS	0.216	0.732	0.052	0	0.836 (0.022)
S10 $k=2$	BCMIX	0.21	0.466	0.192	0.09	1.296 (0.047)
	BS	0.546	0.422	0.03	0	0.49 (0.026)
S11 $k=2$	BCMIX	0.034	0.576	0.242	0.106	1.55 (0.040)
	BS	0.212	0.718	0.066	0.004	0.862 (0.023)
S12 $k=2$	BCMIX	0	0.32	0.376	0.18	2.134 (0.047)
	BS	0.004	0.746	0.22	0.03	1.276 (0.023)

Table 5. Median of D for all scenarios S1-S12

Method	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**
BCMIX	0.328	1.181	0.317	0.189	0.189	0.283
Oracle	0.347	1.078	0.337	0.197	0.151	0.246
BS	0.415	1.21	0.316	0.775	0.099	0.284
Standard	0.347	2.11e+4	2.80e+4	0.381	0.903	4.196
Method	**S7**	**S8**	**S9**	**S10**	**S11**	**S12**
BCMIX	0.685	0.47	0.099	0.288	0.586	0.392
Oracle	0.441	0.505	0.078	0.179	0.675	0.399
BS	0.794	0.324	0.077	0.21	0.563	0.37
Standard	0.672	0.97	0.34	0.542	1.618	1.832

5. Application in Analyzing Weekly Returns of S&P 500

We apply the proposed methodology to analyze the weekly log return r_t of the SP500 index, from the trade week of January 4, 1971 to the trade week of December 29, 2014. The data consists of $n = 2295$ closing prices P_t on the last day of week from which the returns $r_t = \log(P_t/P_{t-1})$ are computed. Figure 1 top panel plots the weekly returns r_t. The mean, variance, skewness, and kurtosis of the return series are $1.360 \times 10^{-3}, 5.156 \times 10^{-4}, -0.564$, and 5.504, respectively. We fit the GARCH(1,1) model with multiple change model

$$r_t = \mu + \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega_t + \alpha_t r_{t-1}^2 + \beta_t \sigma_{t-1}^2 \tag{26}$$

to these data, assuming unknown multiple change-points for $(\omega_t, \alpha_t, \beta_t)$. For comparison, we also fit the standard GARCH(1,1) model

$$r_t = \zeta + \sigma_t \epsilon_t, \quad \sigma_t^2 = \xi + \psi r_{t-1}^2 + \phi \sigma_{t-1}^2 \tag{27}$$

with constant parameters, to these data. In both (26) and (27), ϵ_t are assumed to be i.i.d. standard normal. The maximum likelihood estimates of the parameters of (27) with their standard errors in parenthesis, based on the entire time series r_t, are $\hat{\zeta} = 2.07\text{e}-3\ (3.78\text{e}-4)$, $\hat{\xi} = 2.168\text{e}-5\ (5.81e-6)$, $\hat{\psi} = 0.142\ (0.0197)$, $\hat{\phi} = 0.819\ (0.0265)$. Note that $\hat{\psi} + \hat{\phi} = 0.961$ is very close to 1, indicating high volatility persistence.

We apply our inference procedure to fit the model (26). We first estimate the probability of change-point via the EM method in Section 2.2 and obtain that $\hat{p} = 0.0036$. Then we estimate the model parameters of (27) via the mixtured localized likelihood method in Section 2.3. Figure 2 plots the estimates $(\hat{\omega}_t, \hat{\alpha}_t, \hat{\beta}_t)$ over time, which shows that these parameters undergo several abrupt changes during the whole sample period. This can be regarded as the evidence of the existence of multiple change-points. We further compute the volatilities based on the estimated parameters for models (26) and (27), respectively, shown in Figure 3. We find that GARCH models that incorporates the possibility of multiple change-points better capture high volatility in the market, especially when the market suffers instability.

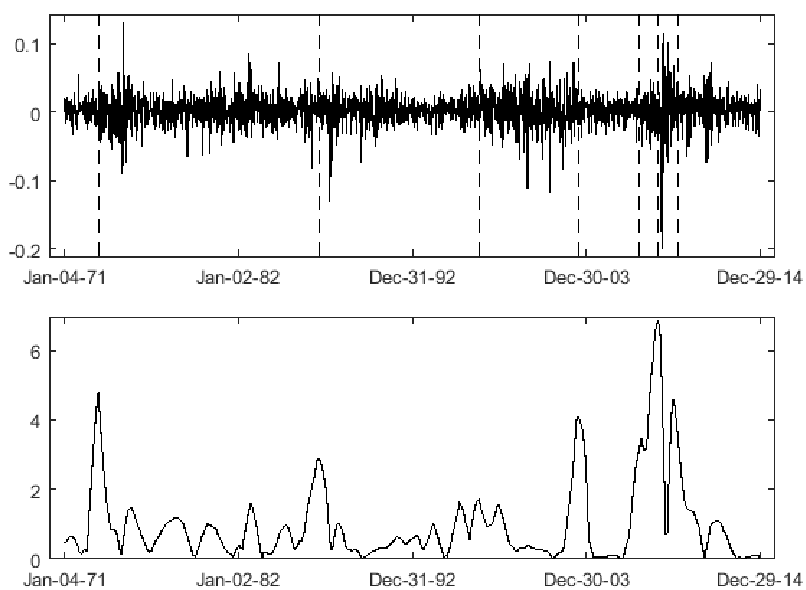

Figure 1. Weekly log returns of S&P 500 index (top) and the distance measure Δ_t (bottom)
Dashed lines on the top panel represent segmentation result.

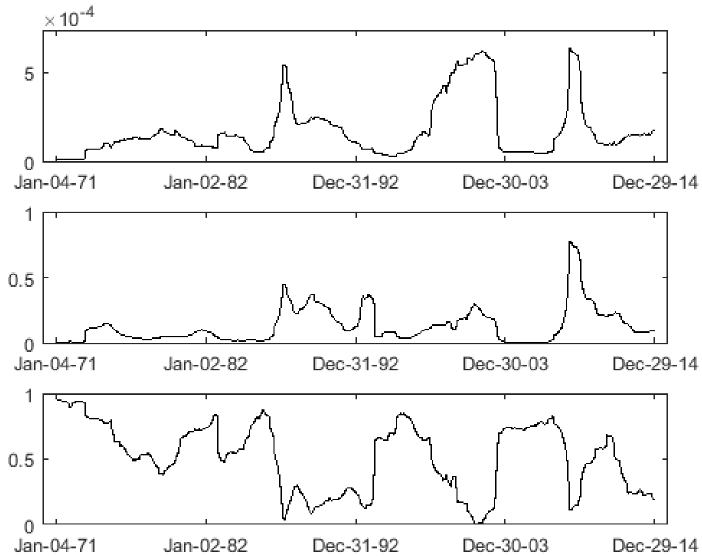

Figure 2. Estimated time-varying parameters $\widehat{\omega}_t$ (top), $\hat{\alpha}_t$ (middle), and $\hat{\beta}_t$ (bottom)

With the estimated time-varying parameters, we compute the distance measure Δ_t via (16) for segmentation purpose, as shown in the bottom panel of Figure 1. Note that the spikes in $\{\Delta_t\}$ indicate potential change-points or structural changes of the series, we then apply the segmentation procedure in Section 3 and obtain the estimated number of change-points $\hat{k} = 7$ and their locations at $t = 115, 840, 1370, 1695, 1895, 1960$ and 2025, which are overlaid with dashed vertical lines on the top panel of Figure 1. These locations correspond to the following dates: March 12, 1973, February 02, 1987, March 31, 1997, June 30, 2003, April 30, 2007, July 28, 2008 and October 26, 2009. Table 6 shows the estimated GARCH parameters in each segment, with the estimated parameters for (27) in the bottom line. Compared to GARCH models without change-points, most segmented data demonstrate smaller values for $\hat{\alpha} + \hat{\beta}$, indicating lower volatility persistence during each segmented period. The only exception is the period from February 2, 1987 to March 31, 1997, during which $\hat{\alpha} + \hat{\beta}$ is 0.973. As NBER documented that the US economy moved out of recession from the end of 1991, it suggests that our segmentation procedure may have missed certain structural change during that period.

Though speculative, it is intriguing to explain these change-points with historical economic events. The first change-point is on March 12, 1973, which is in the beginning period of the 1973 oil crisis and the concurrent 1973-74 stock market crash. The change-point on February 02, 1987 could be considered as the effect of the early stage of the 1987 market crash. The third change-point on March 31, 1997 could be associated with the Asian financial crisis, which had a server impact on the global market due to financial contagion. The change-point on June 30, 2003 is towards the end of the 2002-03 stock market turndown caused by the internet bubble bursting. The change-points on April 30, 2007, July 28, 2008 and October 26, 2009 all lie in the 2008 global financial crisis, and they overlap with the beginning, the peak, and the end of the crisis, respectively.

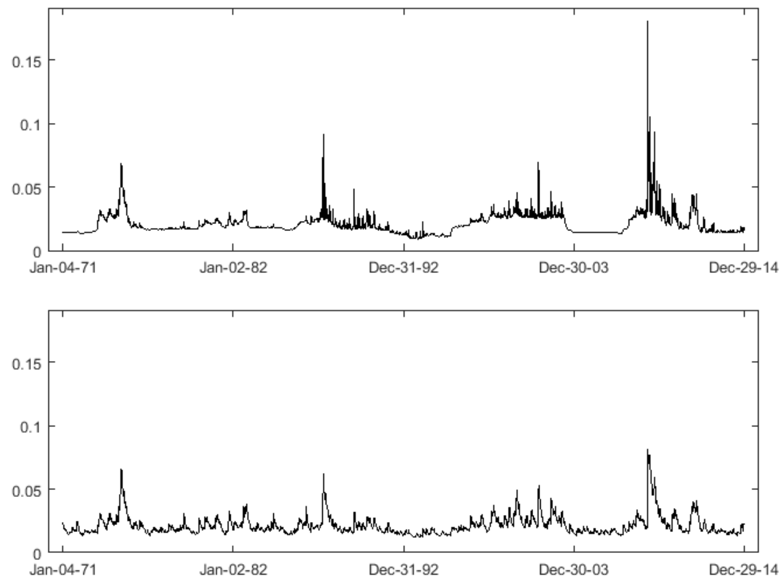

Figure 3. Estimated volatilities in models (26) (Top) and (27) (Bottom)

To investigate predictive power, we do a one-step ahead volatility prediction based on the model estimated using the first n_1 samples of the data. This predicted volatility is compared with benchmark volatilities at time $n_1 + 1$, which are volatilities fitted from the full data (namely the volatilities in Figure 3). We carry out this procedure for both BCMIX method and the standard GARCH for n_1 ranging from 2245 (corresponding to January 13, 2014) to 2294 (corresponding to December 22, 2014), and overlay the predicted and fitted volatilities in Figure 4. From Figure 4 it is obvious that the standard GARCH produces forecasts that are much further away from the benchmark volatilities than those generated from BCMIX. This observation, again can be explained by high persistence in the standard GARCH model when ignoring any change-points.

Table 6. Estimated GARCH parameters with or without segmentation

Period	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$
01/04/1971–03/12/1973	2.764e-5	0.103	0.807	0.91
03/12/1973–02/02/1987	2.877e-5	0.091	0.849	0.94
02/02/1987–03/31/1997	1.025e-5	0.121	0.852	0.973
03/31/1997–06/30/2003	2.536e-4	0.128	0.539	0.667
06/30/2003–04/30/2007	4.164e-5	0.025	0.782	0.807
04/30/2007–07/28/2008	6.191e-5	0.015	0.879	0.894
07/28/2008–10/26/2009	6.417e-4	0.816	0.049	0.865
10/26/2009–12/29/2014	5.561e-5	0.202	0.662	0.864
01/04/1971–12/29/2014	**2.168e-5**	**0.142**	**0.819**	**0.961**

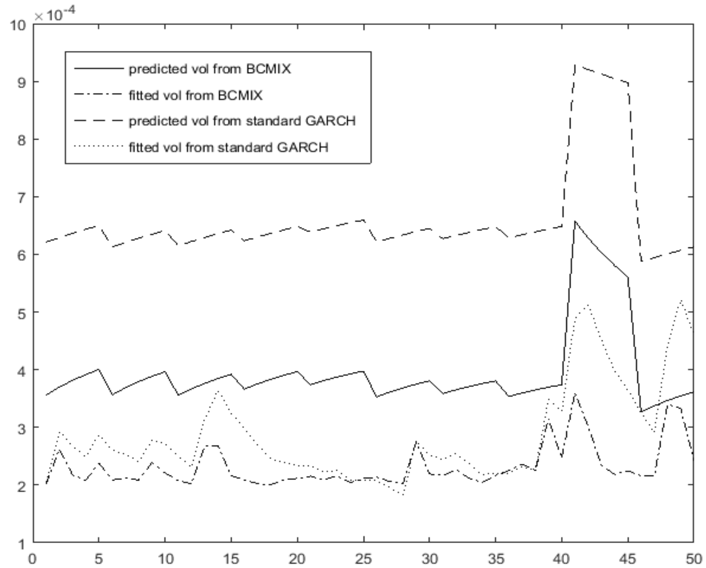

Figure 4. Predicted and fitted volatilities from the BCMIX method and the standard GARCH

6. Conclusion

Change-point GARCH models have been widely discussed to explain the high persistence and long memory phenomena in asset return volatilities. In contrast to many discussions on testing procedures for GARCH models with a single change-point or multiple change-points in individual parameters, the research on direct estimation procedures for GARCH models with multiple change-points has been hampered by the intertwining of nonlinearity of GARCH dynamics and multiple change-points. To overcome this difficulty, we decompose the likelihood to a mixture of localized likelihood and develop a recursive algorithm to compute the mixture probabilities, which is further used to construct an estimate for time-varying GARCH parameters. The developed estimation and associated testing and segmentation procedures are statistically and computationally simple and attractive, as it combines the standard estimation procedure for GARCH models and recent advance in multiple change-points inference. Furthermore, the developed procedure yields consistent and efficient estimates of GARCH parameters and change-points, which are demonstrated via simulation and empirical studies in the paper.

Acknowledgments: The authors would like to thank two anonymous referees for their helpful suggestions. The usual disclaimer applies.

References

[1] Andreou, E., and Ghysels, E. (2002). "Detecting multiple breaks in financial market volatility dynamics", *Journal of Applied Econometrics*, 17 (5): 579-600.

[2] Andreou, E., and Ghysels, E. (2004). "The impact of sampling frequency and volatility estimators on change point tests", *Journal of Financial Econometrics*, 2 (2): 290-318.

[3] Andrews, D. W. K. (1993). "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 61 (4): 821-856.

[4] Berkes, I., Gombay, E., Horvath, L., and Kokoszka, P. (2004a). "Sequential change-point detection in GARCH(*p, q*) models", *Econometric Theory*, 20 (6): 1140-1167.

[5] Berkes, I., Horvath, L., and Kokoszka, P. (2004b). "Testing for parameter constancy in GARCH(p, q) models", *Statistics & Probability Letters*, 70 (4): 263-273.

[6] Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31 (3): 307-327.

[7] Bollerslev, T., and Wooldridge, J.M. (1992). "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews*, 11 (2): 143-172.

[8] Chu, C. S. (1995). "Detecting parameter shift in GARCH models", *Econometric Reviews*, 14 (2): 241-266.

[9] Diebold, F. X. (1986). "Comment on modeling the persistence of conditional variance", *Econometric Reviews*, 5 (1): 51-56.

[10] Engle, R. F. (1982). "Autoregressive conditional heteroskedasticity with estimates of variance of United Kingdom inflation", *Econometrica*, 50 (4): 987-1008.

[11] Fearnhead, P., and Liu, Z. (2007). "On-line inference for multiple change points problems", *Journal of the Royal Statistical Society, Series B*, 69 (4): 589-605.

[12] Fryzlewicz, P., and Subba Rao, S. (2014). "Multiple-change-point detection for autoregressive conditional heteroscedastic processes", *Journal of the Royal Statistical Society, Series B*, 76 (5): 903-924.

[13] Galeano, P., and Tsay, R. (2010). "Shifts in individual parameters of a GARCH model", *Journal of Financial Econometrics*, 8 (1): 122-153.

[14] Hamilton, J. D., and Susmel, R. (1994). "Autoregressive conditional heteroskedasticity and changes in regime", *Journal of Econometrics*, 64 (1-2): 307-333.

[15] Hillebrand, E. (2005). "Neglecting parameter changes in GARCH models", *Journal of Econometrics*, 129 (1-2): 121-138.

[16] Kokoszka, P., and Leipus, R. (2000). "Change-point estimation in ARCH models", *Bernoulli*, 6 (3): 513-539.

[17] Lai, T. L., Liu, H., and Xing, H. (2005). "Autoregressive models with piecewise constant volatility and regression parameters", *Statistica Sinica*, 15 (2): 279-301.

[18] Lai, T.L., and Xing, H. (2011). "A simple Bayesian approach to multiple change-points", *Statistica Sinica*, 21 (2): 539-569.

[19] Lai, T. L., and Xing, H. (2013). "Stochastic change-point ARX-GARCH models and their applications to econometric time series", *Statistica Sinica*, 23 (4): 1573-1594.

[20] Lamoureux, C. G., and Lastrapes, W. D. (1990). "Persistence in variance, structural change and the GARCH model", *Journal of Business and Economic Statistics*, 8 (2): 225-234.

[21] Lavielle, M. (2005). "Using penalized contrasts for the change-point problem", *Signal Processing*, 85 (8): 1501-1510.

[22] Lundbergh, S., and Teräsvirta, T. (2002). "Evaluating GARCH models", *Journal of Econometrics*, 110 (2): 417-435.

[23] Mikosch, T., and Starica, C. (2004). "Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects", *The Review of Economics and Statistics*, 86 (1): 378-390.

[24] Perron, P. (1989). "The great crash, the oil price shock, and the unit root hypothesis", *Econometrica*, 57 (6): 1361-1401.

[25] Siegmund, D., and Zhang, N. (2006). "A modified Bayes information criterion with applications to comparative genomic hybridization data", *Biometrics*, 63 (1): 22-32.

[26] Yao, Y. C. (1984). "Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches", *The Annals of Statistics*, 12 (4): 1434-1447.

[27] Yao Y. C. (1988). "Estimating the number of change-points via Schwarz' criterion", *Statistics & Probability Letters*, 6 (3): 181-189.