Discussion Paper Series A No.471

Asymptotically Efficient Estimation of the Change Point for Semiparametric GARCH Models

Takayuki Shiohama

January 2006

The Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan

Asymptotically Efficient Estimation of the Change Point for Semiparametric GARCH Models*

Takayuki Shiohama

Institute of Economic Research, Hitotsubashi University,
Naka 2-1, Kunitachi, Tokyo 186-0003, Japan

Abstract

The instability of volatility parameters in GARCH models is an important issue for analyzing financial time series. In this paper, we investigate the asymptotic theory for change-point estimators in semiparametric GARCH models. When the parameters of GARCH models have changed within an observed realization, two types of estimators, maximum likelihood estimator (MLE) and Bayesian estimator (BE), are proposed. Then, we derive the asymptotic distributions of these estimators. MLE and BE have different limit laws, and the BE is asymptotically efficient. Monte Carlo studies are conducted on the finite sample behaviors. Further, applications to the Nikkei 225 index are discussed.

Key words: asymptotic efficiency; Bayesian estimator; change point; GARCH process; maximum likelihood estimator.

1 Introduction

Explicit models of heteroskedasticity have received considerable attention in statistics and econometrics literatures during the last two decades. Several models have been proposed to analyze special features of financial data such as log returns of exchange rates and stock prices. The GARCH model, a generalization of the ARCH model introduced by Bolleslev (1986), along with some of its derivative models such as GARCH-M, EGARCH, and GJR are undoubtedly the most successful models used for this analysis. Engle (1995), Gouriéroux (1997), and Campbell, Lo, and MacKinlay (1997) can be referred to for a general overview of the definitions and properties with regard to the ARCH and GARCH models.

Parameter instability in GARCH models may be due to various factors such as policy changes and shocks occurring in the domestic or foreign financial markets. Hence, we need to consider whether the parameters of an observed financial time series are unstable over a period of time. It is well known that the failure to take into account parameter changes that already exsit may lead to incorrect policy implications and predictions. Therefore, it is important to estimate unknown change points in GARCH

*This research is partially supported by the Ministry of Education, Science, Sports and Culture, Grant in Aid for Young Scientist (B), 16730111.

models in order to avoid spurious inferences. As noted in Diebold (1986) and later shown in Lamourex and Lastrapes (1990), a high degree of persistence in GARCH models (IGARCH) can be attributed to a misspecification of the volatility equation. The importance and necessity of testing the stability of volatility parameters are presented in Pagan and Schwert (1990), Bollerslev, Chow, and Kroner (1991), and Lamoureux and Lastrapes (1990).

There is a small but growing interest in testing for and estimating changes in the parameters of ARCH and GARCH models. Chu (1995) and Lundbergh and Teräsvirta (2002) considered Lagrange multiplier tests to detect a parameter shift in GARCH models. Mikosch and Stărică (2002) proposed periodogram-type statistics for testing the goodness-of-fit. Berkes, Horváth, and Kokoszka (2004) studied a test based on approximated likelihood scores to study the parameter constancy in GARCH(p, q) models. Cai (1994) and Hamiltion and Susmel (1994) applied the regime-switching parameters in an ARCH specification in order to account for the possible presence of structural breaks. The application to foreign exchange rates using regime-shift GARCH models can be found in Nakatsuma (2000).

There is little literature available on optimal estimation of change points. For independent and identically distributed observations, Ritov (1990) developed an asymptotically efficient estimation method by using nonparametric setups. For diffusion processes, Kutoyants (1994, 2004) showed that BE is asymptotically optimal. For dependent observations, Shiohama, Taniguchi, and Puri (2003) and Shiohama (2003) studied asymptotically efficient estimations for time series regression models.

In this paper, we consider semiparametric GARCH models with a structural break point. The idea of using semiparametric density in ARCH models was studied by Drost and Klassen (1997). One of the advantages of using this model is that we can treat classical GARCH models as well as GARCH-M models of Engle, Lilien, and Robins (1987) by appropriately choosing the parameters. GARCH-M models have been widely used to analyze the relationship between market returns and their volatilities. To include the risk premium term in the mean equation in GARCH models, we need to investigate the risk managements and option pricing, for example, see Duan (1995) and Heston and Nandi (2000).

This paper is organized as follows. Section 2 defines the semiparametric GARCH models with a structural break point. Further, an asymptotic representation for the log-likelihood ratio between contiguous hypothesis is derived. Section 3 describes the asymptotic estimation theory of change-point estimators. Section 4 explains the cases with regard to a local change, where a problem regard to the shrinking magnitude of a shift is considered. Monte Carlo simulations showing the performances of our theoretical results are given in Section 5. The applications of these theoretical results to the financial markets are presented in Section 6. Finally, Section 7 includes the proofs of the theorems and lemmas given in Sections 2, 3, and 4.

2 Asymptotics of the Likelihood Ratio Process

In this section, we introduce the semiparametric GARCH models with a structural break. Further, the asymptotic representation for the likelihood ratio process are studied. Throughout this paper, we denote the suffix $i = 1, 2$. Let $\mu_i \in \mathbb{R}, \sigma_i > 0, \alpha_i > 0, \omega_i > 0$, and $\beta_i > 0$ be the parameters and let $\{\varepsilon_t : t \in \mathbb{Z}\}$ denote an i.i.d. sequence of innovation errors with location zero, scale one, and density g. We substitute ξ_{it} with $\mu_i + \sigma_i \varepsilon_t$; it should be noted that ξ_{it} is a random variable with location μ_i, scale σ_i, and density $\sigma_i^{-1} g(\{\cdot - \mu_i\}/\sigma_i)$. Further, we introduce the following convention: random variables such as ε and ξ_i denote a typical element in the corresponding sequences $\{\varepsilon_t : t \in \mathbb{Z}\}$ and $\{\xi_{it} : t \in \mathbb{Z}\}$.

The observations $\{y_t, 1, \ldots, n\}$ follow a GARCH(1,1) process with an unknown change point if

$$y_t = \begin{cases} h_{1t}^{1/2} \xi_{1t} = \mu_1 h_{1t}^{1/2} + \sigma_1 h_{1t}^{1/2} \varepsilon_t, & t = 1, \ldots, [\tau n] \\ h_{2t}^{1/2} \xi_{2t} = \mu_2 h_{2t}^{1/2} + \sigma_2 h_{2t}^{1/2} \varepsilon_t, & t = [\tau n] + 1, \ldots, n, \end{cases} \tag{2.1}$$

where the unobservable heteroskedasticity factors h_{1t} and h_{2t} depend on the past values as follows:

$$h_{1t} = \omega_1 + \beta_1 h_{1,t-1} + \alpha_1 y_{t-1}^2, \quad t = 1, \ldots, [\tau n], \tag{2.2}$$

$$h_{2t} = \omega_2 + \beta_2 h_{2,t-1} + \alpha_2 y_{t-1}^2, \quad t = [\tau n] + 1, \ldots, n. \tag{2.3}$$

It should be noted that the Euclidean parameters $\theta_i = (\omega_i, \alpha_i, \beta_i, \mu_i, \sigma_i)'$ and τ are identifiable. In this paper, we assume that equations (2.2) and (2.3) admit a stationary solution $\{h_t : t \in \mathbb{Z}\}$. A necessary and sufficient condition is given by Theorem 2 of Nelson (1990); it is expressed as follows:

Assumption 1 $E \ln\{\beta_i + \alpha_i \xi_i^2\} < 0$ for $i = 1, 2$.

Observe that the model with the autoregression parameters ω_i, α_i, and β_i corresponds to the location-scale model for i.i.d. random variables since the information provided by the observations $h_{01}, y_1, \ldots, y_n$ is equal to that of the random variables $\xi_1, \ldots, \xi_n$. Consequently, the location-scale model is a parametric submodel of our time-series model, and this submodel can be assumed to be regular, for example, see Hájek and Šidák (1967).

Assumption 2 The distribution of ε possesses an absolutely continuous Lebesgue density g with derivative g' and finite Fisher information for the location expressed as

$$I_l(g) = \int \{g'/g\}^2 g(\varepsilon) d\varepsilon \tag{2.4}$$

and that for scale expressed as

$$I_s(g) = \int \{1 + \varepsilon g'/g(\varepsilon)\}^2 d\varepsilon. \tag{2.5}$$

Moreover, the random variable ε has location zero and scale one.

We select the following local parameterizatin:

$$\theta_1^{(n)} = \theta_1 + \frac{\lambda_1}{\sqrt{n}}, \quad \theta_2^{(n)} = \theta_2 + \frac{\lambda_2}{\sqrt{n}}, \quad \text{and} \quad \tau^{(n)} = \tau + \frac{\rho}{n}, \tag{2.6}$$

where $\lambda_1 \in \mathbb{R}^5$, $\lambda_2 \in \mathbb{R}^5$ and $\rho \in \mathbb{R}$ are constants. Hereafter, we assume that $\rho > 0$ without loss of a generality. An analogous discussion for $\rho < 0$ can be derived in a similar mannar. To obtain the asymptotics of our estimators (MLE and BE), we consider the log-likelihood ratio $\Lambda_n(\lambda_1, \lambda_2, \rho)$ of $h_{01}, y_1, \ldots, y_n$ for $\theta^{(n)} = (\theta_1^{(n)'}, \theta_2^{(n)'}, \tau^{(n)})$ with respect to $\theta = (\theta_1', \theta_2', \tau)$ for a fixed g. Note that the residuals and conditional variances up to time t can be calculated recursively from (θ_1, θ_2) and the observations $h_{10}, y_1, \ldots, y_n$, where $h_{10}(\theta_1) = h_{10}$ and $h_{2,[\tau n]}(\theta_2) = h_{2,[\tau n]}$ for $t = 1, 2, \ldots,$ are as follows:

$$\xi_t(\theta) = \begin{cases} y_t / h_{1t}^{1/2}(\theta_1), & t = 1, \ldots, [\tau n], \\ y_t / h_{2t}^{1/2}(\theta_2), & t = [\tau n] + 1, \ldots, n, \end{cases} \tag{2.7}$$

$$\varepsilon_t(\theta) = \begin{cases} \{\xi_t(\theta_1) - \mu_1\}/\sigma_1, & t = 1, \ldots, [\tau n], \\ \{\xi_t(\theta_2) - \mu_2\}/\sigma_2, & t = [\tau n] + 1, \ldots, n, \end{cases} \tag{2.8}$$

$$\begin{cases} h_{1t+1}(\theta_1) = \omega_1 + \beta_1 h_{1t}(\theta_1) + \alpha_1 y_t^2 \\ h_{2t+1}(\theta_2) = \omega_2 + \beta_2 h_{2t}(\theta_2) + \alpha_2 y_t^2. \end{cases} \tag{2.9}$$

Conditionary on h_{10} and $h_{2,[\tau n]}$, the density of $\boldsymbol{y}_n = (y_1, \ldots, y_n)'$ under θ is given by

$$\begin{aligned}
& L_n(\theta_1, \theta_2, \tau) \\
= & \prod_{t=1}^{[\tau n]} \sigma_1^{-1} h_{1t}^{-1/2} g(\sigma_1^{-1}\{h_{1t}^{-1/2} y_t - \mu_1\}) \prod_{t=[\tau n]+1}^{n} \sigma_2^{-1} h_{2t}^{-1/2} g(\sigma_2^{-1}\{h_{2t}^{-1/2} y_t - \mu_2\}) \\
= & \prod_{t=1}^{[\tau n]} \sigma_1^{-1} h_{1t}^{-1/2} g(\{\xi_{1t} - \mu_1\}/\sigma_1) \prod_{t=[\tau n]+1}^{n} \sigma_2^{-1} h_{2t}^{-1/2} g(\{\xi_{2t} - \mu_2\}/\sigma_2) \\
= & \prod_{t=1}^{[\tau n]} \sigma_1^{-1} h_{1t}^{-1/2} g(\varepsilon_{nt}) \prod_{t=[\tau n]+1}^{n} \sigma_2^{-1} h_{2t}^{-1/2} g(\varepsilon_{nt}),
\end{aligned}$$

where $h_{it} = h_{it}(\theta_i)$, $\xi_{it} = \xi_{it}(\theta_i)$, and $\varepsilon_{nt} = \varepsilon_t(\theta)$.

We introduce the following notation with $h_{it}^{(n)} = h_{it}(\theta_i^{(n)})$:

$$l\{\mu_i, \sigma_i\}(x) = \log g(\{x - \mu_i\}/\sigma_i) - \log \sigma_i \tag{2.10}$$

$$\begin{pmatrix} M_{it}^{(n)} \\ S_{it}^{(n)} \end{pmatrix} = n^{1/2}\sigma_i^{-1} h_{it}^{-1/2} \begin{pmatrix} \mu_i^{(n)} {h_{it}^{(n)}}^{1/2} - \mu_i h_{it}^{1/2} \\ \sigma_i^{(n)} {h_{it}^{(n)}}^{1/2} - \sigma_i h_{it}^{1/2} \end{pmatrix} \tag{2.11}$$

and $\varepsilon_t^{(n)} = \varepsilon_t(\theta^{(n)})$. Here, $\log \Lambda_n^s$ denotes the log-likelihood ratio for h_{10}; further, the log-likelihood ratio $\log \Lambda(\lambda_1, \lambda_2, \tau)$ may be written as

$$\log \Lambda(\lambda_1, \lambda_2, \tau) \quad = \quad \log \left\{ \sum_{t=1}^{[\tau n]} A_{1t} + \sum_{t=[\tau n]+1}^{n} A_{2t} + \sum_{t=[\tau n]+1}^{[\tau n+\rho]} (A_{3t} - A_{4t}) \right\} + \Lambda_n^s, \tag{2.12}$$

where

$$\begin{aligned} A_{1t} &= \{l\{(\mu_1, \sigma_1) + \sigma_1 n^{-1/2}(M_{1t}^{(n)}, S_{1t}^{(n)})\}(\xi_{1t}) - l\{(\mu_1, \sigma_1)\}(\xi_{1t})\} \\ &= \{l\{(0,1) + n^{-1/2}(M_{1t}^{(n)}, S_{1t}^{(n)})\}(\varepsilon_{nt}) - l\{(0,1)\}(\varepsilon_{nt})\} \qquad (2.13) \\ A_{2t} &= \{l\{(\mu_2, \sigma_2) + \sigma_2 n^{-1/2}(M_{2t}^{(n)}, S_{2t}^{(n)})\}(\xi_{2t}) - l\{(\mu_2, \sigma_2)\}(\xi_{2t})\}, \\ &= \{l\{(0,1) + n^{-1/2}(M_{2t}^{(n)}, S_{2t}^{(n)})\}(\varepsilon_t) - l\{(0,1)\}(\varepsilon_{nt})\}, \qquad (2.14) \\ A_{3t} &= \log\left[{\sigma_1^{(n)}}^{-1} {h_{1t}^{(n)}}^{-1/2} g(\varepsilon_{nt}^{(n)})\right], \quad \text{and} \quad A_{4t} = \log\left[{\sigma_2^{(n)}}^{-1} {h_{2t}^{(n)}}^{-1/2} g(\varepsilon_{nt}^{(n)})\right]. \qquad (2.15) \end{aligned}$$

To eliminate the initial condition in the log-likelihood ratio statistic, we will use the following regularity condition:

Assumption 3 The density $\bar{g}_\theta$ of the initial value h_{10} under θ satisfies

$$\Lambda_n^s = \log\{\bar{g}_{\theta^{(n)}}/\bar{g}_\theta(h_{01})\} \underset{p}{\longrightarrow} 0 \quad \text{as} \quad n \to \infty.$$

To develop an appropriate expansion Λ_n, it will be convenient to introduce $\dot{l}_{int}$ for the five-dimensional conditional score at time t. More precisely, the three-dimensional vector derivative of the conditional variance is given by

$$H_{it}(\theta_i) = \frac{\partial}{\partial(\omega_i, \alpha_i, \beta_i)} h_{it}(\theta_i) = \beta H_{i,t-1}(\theta_i) + \begin{pmatrix} 1 \\ y_{t-1}^2 \\ h_{it-1}(\theta_i) \end{pmatrix}, \tag{2.16}$$

where $H_{10}(\theta_1) = H_{2[\tau n]}(\theta_2) = (0,0,0)'$. The (5×2)-derivative matrix $W_{it}(\theta_i)$ is defined as

$$W_{it}(\theta_i) = \sigma_i^{-1} \begin{pmatrix} \frac{1}{2} h_{it}^{-1}(\theta_i) H_{it}(\theta_i)(\mu_i, \sigma_i) \\ I_2 \end{pmatrix}. \tag{2.17}$$

The location-scale score is denoted as ($l' = g'/g$)

$$\psi_t(\theta_i) = -\begin{pmatrix} l'(\varepsilon_t(\theta_i)) \\ 1 + \varepsilon_t(\theta_i) l'(\varepsilon_t(\theta_i)) \end{pmatrix}, \tag{2.18}$$

and $\dot{l}_{it}(\theta_i) = W_{it}(\theta_i)\psi_t(\theta_i)$. Then, the conditional score at time t may be denoted by $\dot{l}_{int} = \dot{l}_t(\theta_i)$. It should be noted that $\dot{l}$ is just a heuristic score. We also observe that

$$\begin{aligned}
& \sum_{t=[\tau n]+1}^{[\tau n+\rho]} A_{3t} - A_{4t} \\
= & \sum_{t=[\tau n]+1}^{[\tau n+\rho]} \left[\log\left[\sigma_1^{(n)^{-1}} h_{1t}^{(n)^{-1/2}} g(\varepsilon_{nt}^{(n)})\right] - \log\left[\sigma_2^{(n)^{-1}} h_{2t}^{(n)^{-1/2}} g(\varepsilon_{nt}^{(n)})\right]\right] \\
= & \sum_{t=[\tau n]+1}^{[\tau n+\rho]} \left[\log\left[\sigma_1^{(n)^{-1}} h_{1t}^{(n)^{-1/2}} g(\varepsilon_{nt}^{(n)})/\sigma_1^{-1} h_{1t}^{-1/2} g(\varepsilon_{nt})\right]\right. \\
& \left. + \log\left[\sigma_1^{-1} h_{1t}^{-1/2} g(\varepsilon_{nt})/\sigma_2^{-1} h_{2t}^{-1/2} g(\varepsilon_{nt})\right] - \log\left[\sigma_1^{(n)^{-1}} h_{1t}^{(n)^{-1/2}} g(\varepsilon_{nt}^{(n)})/\sigma_2^{-1} h_{2t}^{-1/2} g(\varepsilon_{nt})\right]\right] \\
= & \sum_{t=[\tau n]+1}^{[\tau n+\rho]} \left\{n^{-1/2}\lambda_1' \dot{l}_{1nt} - n^{-1/2}\lambda_2' \dot{l}_{2nt} + (\theta_2 - \theta_1)' \dot{l}_{1nt}\right\} \\
= & \sum_{t=[\tau n]+1}^{[\tau n+\rho]} (\theta_2 - \theta_1)' \dot{l}_{1nt} + O_p(n^{-1/2}).
\end{aligned} \tag{2.19}$$

An expansion of (2.12) shows that the log-likelihood ratio Λ_n can be alternatively written as

$$\begin{aligned}
& \log \Lambda_n(\lambda_1, \lambda_2, \rho) \\
= & \lambda_1' n^{-1/2} \sum_{t=1}^{[\tau n]} \dot{l}_{1nt} + \lambda_2' n^{-1/2} \sum_{t=[\tau n]+1}^{n} \dot{l}_{2nt} + \sum_{t=[\tau n]+1}^{[\tau n+\rho]} (\theta_2 - \theta_1)' \dot{l}_{1nt} \\
& - \frac{1}{2n}\left[\sum_{t=1}^{[\tau n]} \{\lambda_1' \dot{l}_{1nt}\}^2 + \sum_{t=[\tau n]+1}^{n} \{\lambda_2' \dot{l}_{2nt}\}^2 + \sum_{t=[\tau n]+1}^{[\tau n+\rho]} \{(\theta_2 - \theta_1)' \dot{l}_{1nt}\}^2\right] + R_n.
\end{aligned}$$

The asymptotic representation of this likelihood ratio is stated in the following theorem. The proof of this theorem is presented in Section 7.

Theorem 2.1 *Suppose that Assumptions (A.1)-(A.3) are satisfied. Then, the log-likelihood ratio* $\log \Lambda_n(\lambda_1, \lambda_2, \rho)$*, as defined by (2.12) and (2.20), has an asymptotic representation*

$$\begin{aligned}
& \log \Lambda(\lambda_1, \lambda_2, \rho) \\
= & \sqrt{\tau}\lambda_1' \Delta_{1n} + \sqrt{1-\tau}\lambda_2' \Delta_{2n} + (\theta_2 - \theta_1)' \Delta_{3n} \\
& - \frac{1}{2}\left[\tau \lambda_1' I(\theta_1)\lambda_1 + (1-\tau)\lambda_2' I(\theta_2)\lambda_2 + (\theta_2 - \theta_1)' V(\theta_2 - \theta_1)\right] + o_p(1),
\end{aligned} \tag{2.20}$$

where

$$\Delta_{1n} = \frac{1}{\sqrt{\tau n}} \sum_{t=1}^{[\tau n]} \dot{l}_{1nt} \quad \text{and} \quad \Delta_{2n} = \frac{1}{\sqrt{(1-\tau)n}} \sum_{t=[\tau n]}^{n} \dot{l}_{2nt}$$

are Gaussian random variables with mean 0 and variance $I(\theta_1)$ *and* $I(\theta_2)$*, respectively. Here,* $I(\theta_i)$ *is the probability limit of the averaged score products* $\dot{l}_{it}\dot{l}'_{it}$. $\Delta_{3n} = \sum_{t=[\tau n]+1}^{[\tau n+\rho]} \dot{l}_{1nt}$ *is a random variable with mean 0 and variance* $V \equiv E(\Delta_{3n}\Delta'_{3n})$.

The properties of the likelihood ratio can be obtained by the following lemma:

Lemma 2.1 *Suppose that Assumptions (A.1)-(A.3) hold. Then, for any compact set* $\mathcal{C} \in \boldsymbol{\Theta}$ *to be an open subset of* $\mathbb{R}^5 \times \mathbb{R}^5 \times [0,1]$*, the function* $\Lambda_n(\lambda_1, \lambda_2, \rho)$ *possesses the following properties:*

$$\sup_{\theta} E\Lambda_n^{1/2}(\lambda_1, \lambda_2, \rho) \le \exp\{-g(\lambda_1, \lambda_2, \rho)\}, \tag{2.21}$$

where

$$g(\lambda_1, \lambda_2, \rho) = \lambda_1' \boldsymbol{K}_1 \lambda_1 + \lambda_2' \boldsymbol{K}_2 \lambda_2 + \rho^2 C$$

with some positive definite matrix $\boldsymbol{K}_i$ *and* $C > 0$*. Further, there exists a number* $m > 0$ *such that*

$$\begin{aligned} &\sup_{(\theta_1,\theta_2,\tau)\in\mathcal{C},\lambda_1,\lambda_2<H,\rho<H} \left[\sum_{j=1}^{2} \left\| \lambda_j^{(2)} - \lambda_j^{(1)} \right\|^{2m} + \left\| \rho^{(2)} - \rho^{(1)} \right\|^{2m} \right]^{-1} \\ \times \quad & E\left[\Lambda_n^{1/m}(\lambda_1^{(2)}, \lambda_2^{(2)}, \rho^{(2)}) - \Lambda_n^{1/m}(\lambda_1^{(1)}, \lambda_2^{(1)}, \rho^{(1)}) \right]^m \le B(1+H)^m. \end{aligned}$$

3 Properties of Estimators

We are interested in the behavior of the maximum likelihood estimator (MLE) and the Bayesian estimator (BE) for the parameters of semiparametric GARCH models in the presence of a structural break point. MLE $\hat{\theta}_n^{(ML)} = (\hat{\theta}_{1n}^{(ML)'}, \hat{\theta}_{2n}^{(ML)'}, \hat{\tau}_n^{(ML)})$ is considered as the solution of the following equation:

$$(\hat{\theta}_{1n}^{(ML)'}, \theta_{2n}^{(ML)'}, \hat{\tau}_n^{(ML)}) = \arg\sup_{(\theta_1,\theta_2,\tau)\in\Theta} L_n(\theta_1, \theta_2, \tau). \tag{3.1}$$

To introduce a Bayesian estimator, we need a function $w(y), y \in \mathbb{R}^d$ that is

1. nonnegative, continuous at point 0, and $w(0) = 0$; however, it is not identically zero;
2. symmetric: $w(y) = w(-y)$;
3. the set $\{y : w(y) < c\}$ is convex for all $c > 0$.

The BE (for a quadratic loss function) $\tilde{\theta}_n^{(B)} = (\tilde{\theta}_{1n}^{(B)'}, \tilde{\theta}_{2n}^{(B)'}, \tilde{\tau}_n^{(B)})$ is defined as

$$(\tilde{\theta}_{1n}^{(B)'}, \tilde{\theta}_{2n}^{(B)'}, \tilde{\tau}_n^{(B)}) = \int_\Theta (\theta_1', \theta_2', \tau) \frac{q(\theta_1, \theta_2, \tau) L_n(\theta_1, \theta_2, \tau)}{\int_\Theta q(\theta_1, \theta_2 \tau) L_n(\theta_1, \theta_2, \tau) d(\theta_1, \theta_2, \tau)} d(\theta_1, \theta_2, \tau). \tag{3.2}$$

Let us introduce random fields

$$\Lambda^{(1)}(\lambda_1) = \exp\left\{ \sqrt{\tau}\lambda_1' \Delta_{1n} - \frac{1}{2}\tau\lambda_1' I(\theta_1)\lambda_1 \right\},$$
$$\Lambda^{(2)}(\lambda_2) = \exp\left\{ \sqrt{1-\tau}\lambda_2' \Delta_{2n} - \frac{1}{2}(1-\tau)\lambda_2' I(\theta_2)\lambda_2 \right\},$$

and

$$\Lambda^{(3)}(\rho) = \exp\left\{ (\theta_2 - \theta_1)' \Delta_{3n}(\rho) - \frac{1}{2}(\theta_2 - \theta_1)' V(\rho)(\theta_2 - \theta_1) \right\},$$

where $\boldsymbol{\Delta}_{in}, i = 1, 2, 3$ and V are defined previously in Theorem 2.1. Then, the asymptotic representation of the log-likelihood ratio process is expressed as

$$\Lambda(\lambda_1, \lambda_2, \rho) = \Lambda^{(1)}(\lambda_1) + \Lambda^{(2)}(\lambda_2) + \Lambda^{(3)}(\rho).$$

Let $\boldsymbol{\xi}_i \in \mathbb{R}^5, i = 1, 2,$ be a Gaussian random vector

$$\mathcal{L}\{\boldsymbol{\xi}_i\} = N(\mathbf{0}, I(\theta_i)^{-1})$$

and $\zeta \in \mathbb{R}$ be

$$\zeta = \arg\sup_{\rho\in\mathbb{R}} \exp\{\Lambda^{(3)}(\rho)\}.$$

Therefore, the random vector $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \zeta)$ is defined as

$$(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \zeta) = \arg\sup_{\lambda_1, \lambda_2, \rho \in \mathbb{R}^{11}} \exp\{\Lambda(\lambda_1, \lambda_2, \rho)\}. \tag{3.3}$$

On recalling Theorem 2.1 and Lemma 2.1, it is observed that Theorems 1.10.1 and 1.10.2 of Ibragimov and Has'minskii (1981) can be applied; therefore proof is omitted. The MLE has the following properties:

Theorem 3.1 *Let the parameter set* $\boldsymbol{\Theta}$ *be an open subset of* $\mathbb{R}^5 \times \mathbb{R}^5 \times [0, 1]$. *Then, the MLE is uniformly consistent with* $\boldsymbol{\theta} = (\theta_1, \theta_2, \tau) \in \boldsymbol{\Theta}$ *such that*

$$P - \lim_{n\to\infty} \hat{\boldsymbol{\theta}}_n^{(ML)} = \boldsymbol{\theta}$$

and converges in distribution

$$\mathcal{L}_\theta(\mathcal{A}_n(\hat{\boldsymbol{\theta}}_n^{(ML)} - \boldsymbol{\theta})) \xrightarrow[d]{} \mathcal{L}(\xi_1, \xi_2, \hat{u}),$$

where $\mathcal{A}_n = \mathrm{diag}(\sqrt{\tau n}, \ldots, \sqrt{\tau n}, \sqrt{(1-\tau)n}, \ldots, \sqrt{(1-\tau)n}, n)$*. For any continuous loss function* $w \in W_p$*, we have*

$$\lim_{n\to\infty} E_\theta w(\mathcal{A}_n(\hat{\boldsymbol{\theta}}_n^{(ML)} - \boldsymbol{\theta})) = Ew(\xi_1, \xi_2, \hat{u}).$$

Next, we state the asymptotic properties of the BE $(\tilde{\theta}_{1n}^{(B)}, \tilde{\theta}_{2n}^{(B)}, \tilde{\tau}_n^{(B)})$.

Theorem 3.2 *Let the parameter set* $\boldsymbol{\Theta}$ *be an open subset of* $\mathbb{R}^5 \times \mathbb{R}^5 \times [0,1]$*. Then, the BE is uniformly consistent with* $\boldsymbol{\theta} = (\theta_1, \theta_2, \tau) \in \boldsymbol{\Theta}$ *such that*

$$P - \lim_{n\to\infty} \hat{\boldsymbol{\theta}}^{(B)} = \boldsymbol{\theta}$$

and converges in distribution

$$\mathcal{L}_\theta(A_n(\hat{\boldsymbol{\theta}}_n^{(B)} - \boldsymbol{\theta})) \underset{d}{\longrightarrow} \mathcal{L}(\xi_1, \xi_2, \tilde{u}),$$

where $\mathcal{A}_n = \mathrm{diag}(\sqrt{\tau n}, \ldots, \sqrt{\tau n}, \sqrt{(1-\tau)n}, \ldots, \sqrt{(1-\tau)n}, n)$*. For any continuous loss function* $w \in W_p$*, we have*

$$\lim_{n\to\infty} E_\theta w(\mathcal{A}_n(\tilde{\boldsymbol{\theta}}_n^{(B)} - \boldsymbol{\theta})) = Ew(\xi_1, \xi_2, \tilde{u}).$$

According to Theorem 1.9.1 of Ibragimov and Has'minskii (1981), for any estimator $(\theta_{1n}, \theta_{2n}, \tau_n)$, the inequality

$$\lim_{n\to\infty} \sup_{\theta_1,\theta_2,\tau\in\Theta} E_\theta[w\{\mathcal{A}_n(\theta_{1n}, \theta_{2n}, \tau_n)\}] \geq E[w\{(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \tilde{\zeta})\}]$$

holds. Hence, the BE is asymptotically efficient with respect to the quadratic loss function: however, the MLE is not as asymptotically efficient.

The theorems obtained in this section can be easily extended to semiparametric GARCH(p,q) models and to the problem of multiple structural breaks. Moreover, these results can be extended to the case of ARMA-GARCH models by applying the results of the local asymptotic quadratic (LAQ) form of Ling and McAleer (2003).

4 Local Change Problem

In this section, we consider the asymptotic distributions of the change-point estimators based on a shrinking magnitude of the shift. Since, the limiting distribution of the semiparametric GARCH parameters are identical to those obtained in previous chapters, we will investigate the asymptotic properties for change-point estimators. Hence, we consider the following local parameterization:

$$\tau^{(n)} = \tau + \rho \left\|\delta_n\right\|^{-2}, \tag{4.1}$$

where $\delta_n' = n^{1/2}(\theta_2 - \theta_1)'$, $(\theta_2 - \theta_1)' = O(n^{-\alpha})$ with $0 < \alpha < 1/2$ and $\|\delta_n\| \to \infty$ as $n \to \infty$. The notation $\|x\|$ denotes the Euclidean norm, that is, $\|x\| = (\sum_{j=1}^{p} x_j^2)^{1/2}$ for $x \in \mathbb{R}^p$.

Let

$$\psi(r, \theta_i) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \dot{l}_{int};$$

then based on the general properties of a score function, we have for $c' \in \mathbb{R}^5$,

$$\sqrt{n} c' \psi(1, \theta_i) \underset{d}{\longrightarrow} N(0, c' I(\theta_i) c).$$

To extract the asymptotics, we use the fact that the following functional central limit theorem holds:

$$c' \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \psi(r, \theta_i) = V_L^{1/2} W(r), \tag{4.2}$$

where $W(\cdot)$ represents a Wiener process or standard Brownian motion and V_L is the long-run variance of $c'\psi(1, \theta_i)c$, that is, $V_L \equiv R(0) + 2\sum_{j=1}^{\infty} R(j)$. Here, $R(j)$ is the lag j autocovariance function of $c'\psi(1, \theta_i)c$. The proof is obtained by directly applying Donsker's theorem (see Billingsley (1999) and Phillips and Durlauf (1986)).

Theorem 4.1 *Under the same assumptions of Theorem 2.1 with the condition (4.1), the log-likelihood ratio process* $\log \Lambda_n^L(\rho)$ *has the following asymptotic representation:*

$$\log \Lambda_n^L(\rho) = V_L^{1/2} W(\rho) - \frac{1}{2}|\rho| V_L + o_p(1), \tag{4.3}$$

where $W(\rho); s \in \mathbb{R}$ *is a two-sided standard Wiener process, i.e.,*

$$W(\rho) = \begin{cases} W_1(-\rho) & \rho < 0 \\ W_2(\rho) & \rho \geq 0. \end{cases}$$

Here, $\{W_1(s); s \in [0, \infty)\}$ *and* $\{W_2(s); s \in [0, \infty)\}$ *are independent standard Wiener processes.*

Further, we observe the following lemmas:

Lemma 4.1 *Under the same assumptions of Theorem 2.1 with the condition (4.1), we have*

$$\sup_{\tau \in T} E\Lambda_n^L(\rho) \leq \exp\{-g(\rho)\}, \tag{4.4}$$

where $g(\rho) = \rho C$ *with some constant* $C > 0$. *Further, we have*

$$\sup_{|u|+|v|<R} |u - v|^{-1} E|\Lambda_n^L(u) - \Lambda_n^L(v)|^2 \leq C(1 + R^2). \tag{4.5}$$

The limit process for the likelihood ratio $\Lambda_n^L(\rho)$ is characterized by the process $Z(\rho) = \exp\{W(\rho) - \frac{1}{2}|\rho|\}$. The corresponding random variables $\hat{u}$ and $\tilde{u}$ are defined by the equations

$$\hat{u} = \arg\sup_{u\in\mathbb{R}} Z(u) \quad \text{and} \quad \tilde{u} = \frac{\int uZ(u)du}{\int Z(u)du}. \tag{4.6}$$

Using the arguments from the previous sections, Theorem 4.1 together with Lemma 4.1 indicate a convergence in the distribution of the MLE and BE. The limiting distribution of the MLE and BE for the change point is given by

$$V_L^{-1/2} \|\delta_n\|^2 (\tau_n^{(ML)} - \tau) \xrightarrow[d]{} \mathcal{L}(\hat{u}) \quad \text{and} \quad V_L^{-1/2} \|\delta_n\|^2 (\tau_n^{(B)} - \tau) \xrightarrow[d]{} \mathcal{L}(\tilde{u}). \tag{4.7}$$

The asymptotic efficiency of the BE can be obtained in a similar manner.

Using the limiting distributions obtained above, we can easily construct confidence intervals for the unknown change point τ, while it is difficult to obtain these from Theorems 3.1 and 3.2. By (4.7), we obtain the $100(1-\alpha)$-percent asymptotic confidence interval for the MLE and BE as

$$\left(\hat{\tau}_n^{(ML)} - c^{(ML)}(\alpha/2)\hat{V}_L^{1/2}\left\|\hat{\delta}_n\right\|^{-2}, \hat{\tau}_n^{(ML)} + c^{(ML)}(1-\alpha/2)\hat{V}_L^{1/2}\left\|\hat{\delta}_n\right\|^{-2}\right) \tag{4.8}$$

and

$$\left(\hat{\tau}_n^{(B)} - c^{(B)}(\alpha/2)\hat{V}_L^{1/2}\left\|\hat{\delta}_n\right\|^{-2}, \hat{\tau}_n^{(B)} + c^{(B)}(1-\alpha/2)\hat{V}_L^{1/2}\left\|\hat{\delta}_n\right\|^{-2}\right), \tag{4.9}$$

respectively, where $\hat{V}_L$ and $\hat{\delta}_n$ are the consistent estimators of V_L and δ_n and $c^{(ML)}(\alpha)$ and $c^{(B)}(\alpha)$ are the 100α percent quantiles of the random variables $\hat{u}$ and $\tilde{u}$, respectively. The distribution of $\hat{u}$ is well known; refer to Csörgő and Horváth (1997) and Stryhn (1996) for a detailed description. The distribution of $\tilde{u}$ is investigated by the Monte Carlo simulations. Table 1 provides the asymptotic quantile for $\tilde{u}$ together with $\hat{u}_{MC}$ and $\hat{u}$, where $\hat{u}_{MC}$ denotes the results from the simulations. By comparing the critical values of $\hat{u}$ and $\hat{u}_{MC}$ given in Table 1, we can confirm that the Monte Carlo simulation gives a fairly good approximation of the distribution of $\hat{u}_{MC}$ and $\tilde{u}$. When we construct a 95% confidence interval of τ, the length of the confidence interval using $\hat{\tau}_n^{(ML)}$ is observed to be approximately 28% greater than that of using $\tilde{\tau}_n^{(B)}$.

5 Simulation

In this section, we investigate the finite sample performances of two change-point estimators. The GARCH(1,1) model with an unknown change point is given by

$$y_t = \begin{cases} h_{1t}^{1/2}\varepsilon_t, & t = 1, \dots, [\tau n], \\ h_{2t}^{1/2}\varepsilon_t, & t = [\tau n]+1, \dots, n, \end{cases}$$

where

$$\begin{cases} h_{1t} = \omega_1 + \beta_1 h_{1,t-1} + \alpha_1 y_{t-1}^2, \\ h_{2t} = \omega_2 + \beta_2 h_{2,t-1} + \alpha_2 y_{t-1}^2. \end{cases}$$

We select the parameters $(\omega_1, \alpha_1, \beta_1) = (0.1, 0.1, 0.8)$ and $\beta_2 = (0.4, 0.6)$; the parameters ω_2 and α_2 do not change. Further, we select the sample size as $n = (500, 1000)$ and change point as $\tau = (0.25, 0.5, 0.75)$. The innovation density is performed with $N(0, 1)$ distributions. The prior distribution of the BE is selected as a uniform distribution. Table 2 shows the simulation results with 1,000 replications. We computed the means, standard deviations (S.D.) and the root mean square errors (RMSE) for $\hat{\tau}_n^{(ML)}$ and $\tilde{\tau}_n^{(B)}$. From this table, we observe the following. The RMSE and S.D. of the BE have a better performance than those of the MLE in all the experiments. This is an agreement with the theoretical results given in Section 3. When the sample size n increases, RMSE and S.D. decreases. This verifies the consistency of both the estimators. For a smaller value of $|\beta_1 - \beta_2|$, the values of RMSE and S.D. increase. When the change point is located in the quarter and third quarter of the observations, the bias of the BE and MLE increase.

6 Empirical Results

In this section, we illustrate an application of our theoretical results using the Nikkei 225 index returns. The entire sample consists of the Nikkei 225 returns from January 5, 1997, to March 31, 2005, for a total of 2027 observations. We present the empirical estimates for the GARCH(1,1) model with a normal distribution in the presence of multiple structural breaks[1]. The primary purpose of this section is to investigate the effects of the use of an asymptotically efficient estimator, the BE. To see this, we estimate unknown break points by the MLE, and compare the confidnece intervals of them by using the MLE and BE.

In order to find out the number of break points, we compare the Akaike's information criterion (AIC) and the Schwarz's information criterion (SIC). Such information criterion are often used for model selection: in this case, this refers to the selection of the number of break points. As mentioned in Bai and Perron (2003), the AIC usually overestimates the number of breaks, whereas the SIC sometimes underestimates them. Table 3 shows that the SIC does not selects any structural break point, but the AIC selects a model with three break points; hence we use the AIC selection. The maximum likelihood estimates of the unknown change points with GARCH parameters as well as the case with no structural

[1] We have investigated GARCH-M models with multiple structural breaks, however the satisfactory results on the risk premium term can not obtained.

break point are shown in Table 4. The estimated change points are March 4, 1999, April 14, 2000, and December 17, 2003.

Figure 1 shows the volatility forecasts with unconditional variance. The maximum likelihood break point estimates with 95% confidence intervals based on (4.8) and (4.9) are also given. We made the followings observations. First, it is clear that the unconditional variance is unstable over the given periods. As indicated in Hamiltion and Susmel (1994), a shock on a particular day would produce non-negligible effects on the variance within a few days than that on a year later. This would be the main reason why the GARCH volatility forecasts are sometimes too small and too high across the periods. However, these features can be modified by considering the multiple break point model. Second, both the estimated GARCH models with and without break points exhibit high volatility persistence, which is denoted as $\alpha_i + \beta_i$. This persistence is not attributed to the possible presence of structural breaks in the volatility parameters. We cannot confirm the findings mentioned in Lamoureux and Lastraoes (1990). The change in variance is primarily due to a shift in the parameter ω_i. Third, the length of the confidence interval is affected by the variance of the score functions, while the magnitude of the shift does not affect its length because the nature of GARCH parameters, the magnitude of shift cannot be greater than 1. The estimation of the parameters in the second regime is not statistically significant at the 10% level except for β; widens the confidence interval for the change-point estimates. Fourth, as Tsay (2001) showed, the parameter α_i plays an important role that it determins the kurtosis of the series. Hence, the change in parameter α causes the change in the kurtosis of the given regime. From Table 1, α_2 is close to zero which indicate that the kurtosis is close to the normal destribution, while on the other regimes, α_i for $i = 1, 3, 4$ takes the values beween (0.05,0.10), which indicate the heavy tail phenomima. Finally, the estimated break points –March 4, 1999, April 14, 2000, and December 17, 2003– can be interpreted as an upward trend shift in the level of the Nikkei 225 index, a crash in the New York Dow Jones index, and the turning points that led to the stable level period, respectively. Hence the estimated structural break points relate to the trend shift in the level of market prices.

7 Proofs

Proof of Theorem 2.1 From Theorem 2.1 of Drost and Klaassen (1997) , it can be shown that as $n \to \infty$, $R_n \to 0$,

$$\lambda_1' n^{-1/2} \sum_{t=1}^{[\tau n]} \dot{l}_{1nt} - \frac{1}{2n} \sum_{t=1}^{\tau n} \{\lambda_1' \dot{l}_{1nt}\}^2 \underset{d}{\longrightarrow} N(-\frac{1}{2}\tau\lambda_1' I(\theta_1)\lambda_1, \tau\lambda_1' I(\theta_1)\lambda_1)$$

and

$$\lambda_2' n^{-1/2} \sum_{t=[\tau n]+1}^{n} \dot{l}_{2nt} - \frac{1}{2n} \sum_{t=[\tau n]+1}^{n} \{\lambda_2' \dot{l}_{2nt}\}^2 \xrightarrow[d]{} N(-\frac{1}{2}(1-\tau)\lambda_2' I(\theta_2)\lambda_2, (1-\tau)\lambda_2' I(\theta_2)\lambda_2)$$

as $n \to \infty$. We can easily observe that the random variable Δ_{3n} has a mean $\mathbf{0}$ and variance V. □

The following lemma is used to obtain Lemma 2.1. The proof can be obtained from the classical Kolmogorov exponential inequality (see, for example, Stout (1974), p. 263); hence, we omit the proof.

Lemma 7.1 *Let $X_1, \ldots, X_n$ be a sequence of random variables with mean 0 and finite variance. Let $s_n^2 = \left(\sum_{k=1}^n EX_k\right)^2$ and assume that $|X_k| \le Cs_n^2$, almost surely for each $1 \le k \le n$ and $n \ge 1$. Then, for each $a > 0$ and $n \ge 1$, the assumptions $\epsilon C \le a$ and $0 < \alpha \le a^3/(e^a - 1 - a - a^2/2)$ imply that*

$$E\exp(S_n/s_n^2) \le \exp(-c_1),$$

where $c_1 = \exp\{-(\epsilon/2)(1+\epsilon C/\alpha)\}$ and $S_n = \sum_{k=1}^n X_k$ as usual.

Proof of Lemma 2.1 First, we prove (2.24). From Theorem 2.1, Lemma 7.1 with $2c_1 - 1 > 0$, and the equality $E\exp\zeta^p = \exp(p^2\sigma^2/2)$, which is valid for a Gaussian random variable ζ with parameters $(0, \sigma^2)$, it follows that for $p > 1$

$$\begin{aligned}
& E\Lambda_n(u)^{1/2} = E\left[\exp\left\{\frac{1}{2}(\lambda_1' A_1 + \lambda_2' A_2 + A_3(\rho)) - \frac{1}{4}(\lambda_1' B_1\lambda_1 + \lambda_2' B_2\lambda_2 + B_3(\rho))\right\}\right] \\
\le\ & \exp[-\frac{1}{4}(\lambda_1' B_1\lambda_1 + \lambda_2' B_2\lambda_2 + B_3(\rho))]\left[E\exp\left(\lambda_1' A_1 + \lambda_2' A_2\right)^{p/2}\right]^{1/p}\left[E\exp(A_3(\rho))^{p/2}\right]^{1/p} \\
\le\ & \exp[-\frac{1}{4}(\lambda_1' B_1\lambda_1 + \lambda_2' B_2\lambda_2 + B_3(\rho))]\left[\exp\left(\frac{p}{4}\right)(\lambda_1' B_1\lambda_1 + \lambda_2' B_2\lambda_2) + \frac{c_1}{2}B_3(\rho)\right] \\
=\ & \exp\left\{\left(\frac{p-1}{4}\right)(\lambda_1' B_1\lambda_1 + \lambda_2' B_2\lambda_2) + \frac{2c_1 - 1}{4}B_3(\rho)\right\}
\end{aligned}$$

where

$$A_1 = \sqrt{\tau}\Delta_{1n}, \quad A_2 = \sqrt{1-\tau}\Delta_{2n}, \quad A_3(\rho) = (\theta_2 - \theta_1)'\Delta_{3n},$$
$$B_1 = \tau I(\theta_1), \quad B_2 = (1-\tau)I(\theta_2) \text{ and } B_3(\rho) = (\theta_2 - \theta_1)'V(\theta_2 - \theta_1).$$

These expressions along with $B_3 = O(\rho^2)$ implies (2.24). As for (2.25), we observe that

$$\begin{aligned}
& E_\theta\left|\Lambda_n(\lambda_1^{(1)}, \lambda_2^{(1)}, \rho^{(1)})^{1/2m} - \Lambda_n(\lambda_1^{(2)}, \lambda_2^{(2)}, \rho^{(2)})^{1/2m}\right|^{2m} \\
=\ & E_{\theta_1}\left|Y_n^{1/2m} - 1\right|^{2m}, \qquad (7.1)
\end{aligned}$$

where $Y_n = (\Lambda_n(u^{(2)})/\Lambda_n(u^{(1)}))^{1/m}$, and $u^{(i)} = (\lambda_1^{(i)}, \lambda_2^{(i)}, \rho^{(i)})$ for $i = 1, 2$. The exponent index in Y_n could be

$$\begin{aligned}
\ln Y_n &= (\lambda_1^{(1)} - \lambda_2^{(2)})' A_1 + (\lambda_2^{(1)} - \lambda_2^{(2)})' A_2 + A_3(\rho^{(1)}) - A_3(\rho^{(2)}) \\
&\quad -\frac{1}{2}\left[\lambda_1^{(1)'} B_1 \lambda_1^{(1)} + \lambda_2^{(1)'} B_2 \lambda_2^{(1)} + B_3(\rho^{(1)}) - \lambda_1^{(2)'} B_1 \lambda_1^{(2)} - \lambda_2^{(2)'} B_2 \lambda_2^{(2)} - B_3(\rho^{(2)})\right] \\
&= (\lambda_1^{(1)} - \lambda_2^{(2)})' A_1 + (\lambda_2^{(1)} - \lambda_2^{(2)})' A_2 + A_3(\rho^{(1)}) - A_3(\rho^{(2)}) \\
&\quad -\frac{1}{2}\left[(\lambda_1^{(1)} - \lambda_1^{(2)}) \frac{\partial D_1(\lambda_1^{(*)})}{\partial \lambda_1^{(2)}} + (\lambda_2^{(1)} - \lambda_2^{(2)}) \frac{\partial D_2(\lambda_2^{(*)})}{\partial \lambda_2^{(2)}} + C_{A_3}(\rho^{(1)} - \rho^{(2)})^2)\right] \\
&= F_1 + F_2 + F_3 - \frac{1}{2}[G_1 + G_2 + G_3] \quad \text{(say)} \\
&\equiv \Psi(u^{(1)}, u^{(2)}),
\end{aligned}$$

where $D_i(x) = x' I(\theta_i) x$ for $i = 1, 2$ and $C_{A_3} =$. It can be seen that

$$\begin{aligned}
&E\Psi(u^{(1)}, u^{(2)})^2 = G_1^2 + G_2^2 + G_3^2 + 2(G_1G_2 + G_1G_3 + G_2G_3) \\
&+E\left[F_1^2 + F_2^2 + F_3^2 + 2(F_1F_2 + F_1F_3 + F_2F_3) + \sum_{i=1}^{3} F_i(G_1 + G_2 + G_3)\right].
\end{aligned}$$

We have for $i = 1, 2$ $EF_i^2 = G_i^2 = O\left(\left\|\lambda_i^{(1)} - \lambda_i^{(2)}\right\|^2\right)$ and $EF_3^2 = G_3^2 = O\left((\rho^{(1)} - \rho^{(2)})^2\right)$. Hence from the independence of F_i and F_j for $i \neq j$, we obtain

$$E\Psi(u^{(1)}, u^{(2)})^2 = O\left(\left\|\lambda_1^{(1)} - \lambda_1^{(2)}\right\|^2 + \left\|\lambda_2^{(1)} - \lambda_2^{(2)}\right\|^2 + |\rho^{(1)} - \rho^{(2)}|^2\right). \tag{7.2}$$

Furthremore, we observe that from (2.24) that for $\lambda_j^{(i)}, \rho^{(i)} < H$ with $i, j = i, 2$

$$E\left(\exp\left\{\Psi(u^{(1)}, u^{(2)})\right\}\right) \leq H \tag{7.3}$$

Hence, by using the Schwarz inequality,

$$\begin{aligned}
&E_{\theta_1}\left|Y_n^{1/2m} - 1\right|^{2m} \\
\leq\; & E_{\theta_1}\left|\exp\{\Psi(u^{(1)}, u^{(2)})\} - 1\right|^{2m} \\
\leq\; & E\left|\Psi(u^{(1)}, u^{(2)})\left\{\exp\left\{\Psi(u^{(1)}, u^{(2)})\right\} + 1\right\}\right|^{2m} \\
\leq\; & \left[E\Phi(u^{(1)}, u^{(2)})^2\right]^m \left[2(E\exp\left\{\Psi(u^{(1)}, u^{(2)})\right\} + 1)\right]^m \\
\leq\; & B(1 + H)^m \left\{\left\|\lambda_1^{(2)} - \lambda_1^{(1)}\right\|^{2m} + \left\|\lambda_2^{(2)} - \lambda_2^{(1)}\right\|^{2m} + (\rho^{(2)} - \rho^{(1)})^{2m}\right\},
\end{aligned}$$

where we use the facts (7.2) and (7.3). Hence, the lemma is proved . $\square$

Proof of Theorem 4.1 From (2.20) and the functional limit theorem (4.2), we can see

$$(\theta_2-\theta_1)'\sum_{t=\tau n+1}^{n\tau+\rho n^{2\alpha}}\psi(r,\theta_1)=n^{-\alpha}\sum_{t=1}^{\rho n^{2\alpha}}\psi(r,\theta_1)\underset{d}{\longrightarrow}V_L^{1/2}W_2(\rho),$$

and the variance of $W_2(\rho)$ is expressed as

$$(\theta_2-\theta_1)'\left(\sum_{t=\tau n+1}^{\tau n+\rho n^{2\alpha}}\psi(r,\theta_1)\psi(r,\theta_1)'\right)(\theta_2-\theta_1)'\underset{p}{\longrightarrow}\rho V_L;$$

which concludes the proof. □

Proof of Lemma 4.1 This proof similarly follows from Theorem 1.4 of Kutoyants (1994).

References

[1] Bai, J. & P. Perron (2003) Computation and analysis of multiple structural changes. *Journal of Applied Econometrics* 18, 1-22.

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

[3] Billingsley, P. (1999) *Convergence of Probability Measures.* 2nd eds. New York: Wiley.

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

[5] Bolleslev, T., R. Chou, & K. Kroner (1991) ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52, 5-59.

[6] Campbell, J. Y., A. W. Lo, & A. C. MacKinlay (1997) *The Econometrics of Financial Markets.* Princeton University Press.

[7] Cai, J. (1994) A Markov model of unconditional variance in ARCH. *Journal of Bussiness and Economic Statistics* 12, 309-316.

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

[9] Csörgő, M. & L. Horváth (1997) *Limit Theorems in Change-Point Analysis.* New York: Wiley.

[10] Engle, R. F. (1995) *ARCH Selected Readings.* Oxford University Press

[11] Engle, R. F., D. M. Lilien, & R. P. Robins (1987) Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55, 391-407.

[12] Diebold, F. (1986) Modeling the persistence of conditional variances: A comment. *Econometric Reviews* 5, 51-56.

[13] Drost, F. C. & C. A. J. Klasssen (1997) Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81, 193 -221.

[14] Duan, J. -C. (1995) The GARCH option pricing model. *Mathematical Finance* 5, 13-32.

[15] Gouriéroux, C. (1997) *ARCH Model and Financial Application.* New York: Springer.

[16] Hájek, J. & Z. Šidák (1967) *Theory of Rank Tests.* Prague: Academia.

[17] Hamilton, J. D. & R. Susmel (1994) Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307-33.

[18] Heston, S. L. & S. Nandi (2000) A closed-form GARCH option valuation model. *The Review of Financial Studies* 13, 585-625.

[19] Ibragimov, I. A. & R. Z. Has'minski (1981) *Statistical Estimation.* New York: Springer.

[20] Kutoyants, Yu. A. (1994) *Identification of Dynamical System with Small Noise.* Kluwer Academic Publishers, Dordrecht.

[21] Kutoyants, Yu. A. (2004) *Statistical Inference for Ergodic diffusions.* New York: Springer.

[22] Lamoureux, C. & W. Lastrapes (1990) Persistence in variance, structural change and the GARCH models. *Journal of Business and Economic Statistics* 8, 225-234.

[23] Ling, S. & M. McAleer (2003) On adaptive estimation in nonstationary ARMA models with GARCH errors. *Annals of Statistics* 31, 642-674.

[24] Lundbergh, S. & T. Teräsvirta (2002) Evaluating GARCH models. *Journal of Econometrics* 110, 417-435.

[25] Mikosch, T. & C. Stărică (2002) Long-range dependence effects and ARCH modeling. In *Theory and Applications of Long-range Dependence* eds P. Doukhan, G. Oppenheim and M. S. Taqqu) 439-459. Boston: Birkhauser.

[26] Nakatsuma, T. (2000) Structural changes in volatility of foreign exchange rates after the Asian financial crisis. *Asia-Pacific Financial Markets* 7, 69-82.

[27] Nelson, D. B. (1990) Stationary and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318-334.

[28] Pagan, A. & M. Schwert (1990) Alternative models for conditional stock volatility. *Journal of Econometrics* 3, 87-105.

[29] Phillips, P. C. B & S. N. Durlauf (1986) Multiple time series with integrated variables. *Review of Economic Studies* 53, 473-496.

[30] Ritov, Y. (1990) Asymptotic efficient estimation of the change point with unknown distributions. *Annals of Statistics* 18, 1829-1839.

[31] Shiohama, T. (2003) Asymptotic estimation theory for time series regression models with multiple change points. *Scientiae Mathematicae Japoincae* 57, 299-311.

[32] Shiohama, T., M. Taniguchi, & M. L. Puri (2003) Asymptotic estimation theory of change-point problems for time series regression models and its applications. *Probability, Statistics and their applications: Papers in Honors of Rabi Bhattacharya,* Eds: K. Athreya, M. Majumdar, M. L. Puri, E. Waymire, IMS Lecture Notes Vol. 41 257-284.

[33] Stout, W. F. (1974) *Almost Sure Convergence*. New York: Academic Press.

[34] Stryhn, H. (1996) The location of the maximum of asymptotic two-sided Brownian motion with triangular drift. *Statistics & Probability Letters* 29, 279-284.

[35] Tsay, R. S. (2001) *Analysis of Financial Time Series.* New York: Wiley.

Table 1. Critical values of $\hat{u}$ and $\tilde{u}$.

	Quantile			
	0.01	0.025	0.05	0.10
$\hat{u}$	-15.87	-11.03	-7.69	-4.70
$\hat{u}^{(MC)}$	-15.70	-10.93	-7.66	-4.65
$\tilde{u}$	-11.63	-8.59	-6.29	-4.19

Table 2. Simulation results with $N(0,1)$ distribution.

		MLE			BE		
n	β_2	Mean	S.D.	RMSE	Mean	S.D.	RMSE
				$\tau = 0.25$			
500	0.4	0.2794	0.1235	0.1269	0.2949	0.0985	0.1082
	0.6	0.3363	0.2219	0.2380	0.3457	0.1650	0.1901
1000	0.4	0.2536	0.0385	0.0387	0.2612	0.0357	0.0374
	0.6	0.2690	0.1120	0.1135	0.2830	0.0802	0.0867
				$\tau = 0.5$			
500	0.4	0.4963	0.1139	0.1139	0.4991	0.0864	0.0864
	0.6	0.5075	0.1949	0.1949	0.5079	0.1426	0.1426
1000	0.4	0.4965	0.0667	0.0667	0.5009	0.0411	0.0411
	0.6	0.5044	0.0890	0.0890	0.5056	0.0629	0.0631
				$\tau = 0.75$			
500	0.4	0.7098	0.1550	0.1600	0.7012	0.1261	0.1351
	0.6	0.6689	0.2226	0.2368	0.6555	0.1719	0.1961
1000	0.4	0.7438	0.0618	0.0620	0.7421	0.0492	0.0498
	0.6	0.7367	0.1049	0.1057	0.7268	0.0854	0.0884

Table 3. Results of the AIC and SIC.

	AIC	SIC
$m = 0$	7288	7310
$m = 1$	7274	7336
$m = 2$	7272	7373
$m = 3$	7270	7410
$m = 4$	7272	7454

Notes: (1) m refers to the number of the break points.

(2) AIC is calculated as $-2\log$ likelihood $+ 2(3m + p(m + 1)$, where p denotes the number of parameters in each regime.

(3) SIC is calculated as $-2\log$ likelihood $+ \ln n(3m + p(m + 1))$, where n denotes the sample size.

Table 4. Maximum likelihood estimates of GARCH(1,1) models with and without structural breaks.

	Nikkei 225 Index				
	97/1/7- 99/3/4	99/3/5- 00/4/14	00/4/17- 03/12/17	03/12/18- 05/3/31	97/1/7 05/3/31
τ	0.2698	0.3981	0.8451		-
ω	0.1083**	0.2488	0.1246**	0.0101	0.0557***
	(0.0465)	(0.4265)	(0.0492)	(0.0132)	(0.0176)
α	0.0956***	0.0173	0.0527***	0.0622**	0.0804***
	(0.0241)	(0.0452)	(0.0157)	(0.0252)	(0.0118)
β	0.8687***	0.8000**	0.8991***	0.9285***	0.8971***
	(0.0297)	(0.3500)	(0.0257)	(0.0288)	(0.0144)
$\alpha + \beta$	0.9641	0.8171	0.9517	0.9904	0.9772
$\frac{\omega}{1-(\alpha+\beta)}$	3.0165	1.3604	2.5785	1.0552	2.4479

Standard errors are presented in parentheses. The asterisks indicate significance at 10%(*), 5%(**) and 1%(***) levels.

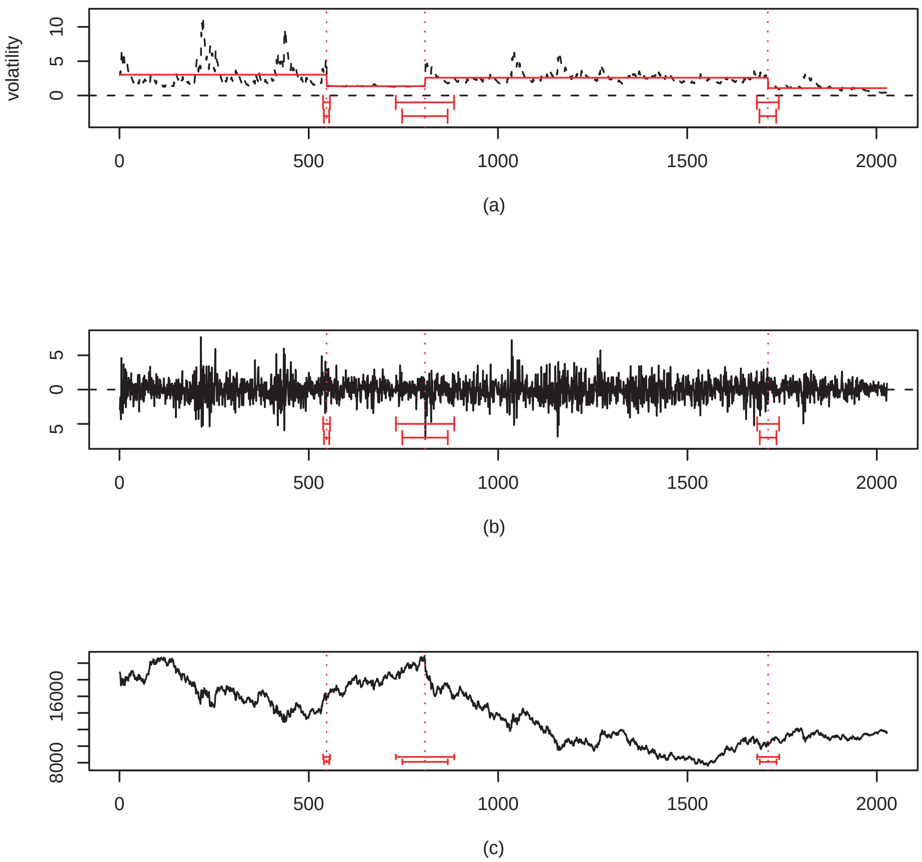

Figure 1. Estimated break points with 95% confidence intervals. The narrow confidence interval corresponds to Bayesian estimators with (a) volatility forecasts, (b) daily log returns of the Nikkei 225 index, and (c) the Nikkei 225 index.