Available online at www.sciencedirect.com
SCIENCE DIRECT
JOURNAL OF Econometrics
ELSEVIER
Journal of Econometrics 126 (2005) 79–114
www.elsevier.com/locate/econbase

Nonparametric estimation of structural change points in volatility models for time series

Gongmeng Chen[a,*], Yoon K. Choi[b], Yong Zhou[a,c]

[a] *School of accounting and Finance, The Hong Kong Polytechnic University, Kowloon, Hong Kong*
[b] *Department of Finance, College of Business Administration, University of Central Florida, Orlando, FL 32816, USA*
[c] *Financial Risk Management Research Group, Chinese Academy of Science, Beijing 100080, China*

Accepted 9 February 2004

Abstract

We propose a hybrid estimation procedure that combines the least squares and nonparametric methods to estimate change points of volatility in time series models. Its main advantage is that it does not require any specific form of marginal or transitional densities of the process. We also establish the asymptotic properties of the estimators when the regression and conditional volatility functions are not known. The proposed tests for change points of volatility are shown to be consistent and more powerful than the nonparametric ones in the literature. Finally, we provide simulations and empirical results using the Hong Kong stock market index (HSI) series.

JEL classification: C1; C22; C5

Keywords: Change points in volatility; Least squares; Nonparametric estimation; Asymptotic properties

1. Introduction

Although financial markets have experienced significant episodes of instability such as the Great Depression, the financial policy regime shifts, and the start of the European Monetary system, econometric models have typically assumed structural stability. In particular, the study of the conditional variance of financial and economic data has drawn much attention due to its importance in hedging strategies and risk management. However, Lamoreux and Lastrapes (1990) have given evidence of structural instability

* Corresponding author. Tel.: +1-00852-2766-7070; fax: +1-00852-2330-9845.
E-mail addresses: acgmchen@inet.polyu.edu.hk (G. Chen), ychoi@bus.ucf.edu (Y.K. Choi), yzhou@bios.unc.edu (Y. Zhou).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2004.02.008

(e.g., jumps) in conditional variance. Many attempts have followed since then to test and estimate jumps and their sizes with conditional heteroskedasticity (Jorion, 1988; Vlaar and Palm, 1993; Drost et al., 1998). All of these models are variants of the popular ARCH (Engle, 1982) and GARCH models (Bollerslev, 1986).

Meanwhile, Pagan and Hong (1991), Pagan and Ullah (1988), and Pagan and Schwert (1990) modelled conditional variance by *nonparametric* procedures.[1] They point out that most parametric models, including ARCH or GARCH models, do not adequately capture the functional relationship between volatility and any underlying economic factors. Nelson (1987) shows, that the most serious limitation of ARCH models of asset pricing is their assumption that only the size, and not the sign, of excess returns determines future conditional variance. Further, Masry and Tjøstheim (1995) estimated and identified the functional structures of nonlinear econometric systems. They also established strong consistency with sharp rates of convergence and asymptotic normality by employing nonparametric kernel estimates.

This paper develops a theory of estimating change points in the conditional variance (volatility) of a *nonparametric* model in which the regression and conditional variance functions are unknown. We expect a significant improvement in describing time series data if we can identify points in time for volatility changes. The correct identification definitely increases the efficiency of parameter estimates with more stable structural regimes. Consider the following nonparametric model:

$$Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t, \tag{1.1}$$

where $\{(X_t, Y_t), t = 1, 2, \ldots\}$ is a sequence of random variables, and $\{\varepsilon_t\}$ is a sequence of stationary errors, with $\mathrm{E}(\varepsilon_t \mid X_t) = 0$ and $Var(\varepsilon_t \mid X_t) = 1$. $\mu(x)$ and $\sigma(X)$ are the regression function (conditional mean) and volatility function (conditional variance), respectively.

For nonparametric regression models, the change point problem has drawn much attention in recent years. Since the inference based on nonparametric models is robust against the misspecification of the underlying regression model, the nonparametric models can effectively avoid the problem of misspecification found in parametric approaches.[2] Very recently, Perron (2001) extends Delgado and Hidalgo (2000)'s nonparametric procedure to detect discontinuities in conditional variance function and proposed an estimator of jumps in the volatility of financial returns. The nonparametric technique is based on one-sided kernel smoothers, first introduced by Müller (1992). The idea is that the left-hand and right-hand side estimates converge to the left and right limit, respectively, at the change points. The difference between these estimates is used to construct the statistic for the detection of a change in volatility. However,

[1] Pagan and Schwert (1990) suggest a simple recursive variance test and show that the data cannot be thought of as homogeneous before and after the Great Depression.

[2] There is a well-developed theory under maximum-likelihood estimation (MLE) for independent and identically distributed observations up to a parametric shift (see Hinkley, 1970; and Bhattacharya, 1987). Bai (1994, 2000) and Antoch et al. (1996) study the estimation of a mean shift in linear processes by the least-squares (LS) method. Bai (1997) consider the LS estimation of a change point in multiple linear regression models. The parametric tests and estimations should perform very well under the correct specifications of the model, but inferences based on misspecified models are not well studied.

the major drawback of the one-sided kernel procedure is that the power of the test is weak, and the rate of convergence is slow.

We propose a hybrid test and estimation procedure for change points in volatility based on the least-squares method in nonparametric time series models. The location of the change point, or change points, is not specified a priori, as has been done in previous studies on change points. We establish the asymptotic properties of the estimators of change points of volatility when the regression and the conditional variance functions are not known. Furthermore, we show that the estimator of the change point is consistent and converges with a rate of $O(T^{-1})$. There are three key features that distinguish our tests and estimation from other literature. First, unlike the MLE, as argued in Bai (1994) and Bai and Perron (1998) the LS method is more flexible in specifying the underlying error distribution function and correlation structure in the data. Our inference procedure for change points possesses the merits of parametric procedures, although the underlying model is nonparametric. Our tests and estimators have the same asymptotic properties as those in parametric models.

Second, the proposed estimators for change points reach the optimal convergence rate of $O(T^{-1})$ in probability; by contrast, Perron's (2001) convergence rate depends on the bandwidth of the nonparametric estimator and is much slower than $O(T^{-1})$. Third, the proposed test is consistent and more powerful than Perron's (2001). Under the alternative hypothesis, the test diverges to infinity at a faster rate than Perron's (2001).

The article is organized as follows. In Section 2, we introduce the LS method. We construct the LS estimator and other transformed estimators and obtain some important asymptotic properties when the regression and conditional variance functions are known. In Section 3, we provide the details of the estimation of the regression and conditional variance functions when they are unknown. We apply local polynomial (linear) smoothers to construct the estimators of the unknown nonparametric regression and conditional variance functions. In Section 4, we propose a new method to select the bandwidths for the estimation of regression and conditional variance functions. In Section 5, we report the simulation and empirical results. The paper concludes with Section 6. Sketches of the proofs and auxiliary results are collected in the appendix.

2. Derivation of estimators

Our derivation of the LS estimator for changes in volatility is based on Bai (1994), who studied the estimation of a shift in mean functions with an unknown shift point in a linear process by the simple LS method. In contrast to the MLE method, the LS procedure does not require specific forms of the marginal or transitional density functions (i.e., the regression and conditional variance functions) or the underlying error distribution function. We start with model (1.1), with a single change point in volatility. The nonparametric model can then be defined as

$$Y_i = \mu(X_i) + \sigma_i(X_i)\varepsilon_i, \quad i = 1, 2, \ldots, T,$$

where

$$\sigma(X_t) = \begin{cases} \tau_1 \sigma_0(X_t) & \text{if} \quad t \leqslant k_0, \\ \tau_2 \sigma_0(X_t) & \text{if} \quad t > k_0, \end{cases}$$

where τ_1, τ_2 and k_0 are unknown parameters. For simplicity, let $k_0 = [T\theta_0]$ for some $0 < \theta_0 < 1$, where $[\,\cdot\,]$ denotes the largest integer less than or equal to its argument. Under H_0 (that is, there is no change in the volatility), it follows from (1.1) that

$$E(Y_i - \mu(X_i))^2 = \sigma^2(X_i),$$

where we assume that $E(\varepsilon_i|X_i) = 0$ and $E(\varepsilon_i^2|X_i) = 1$ and $\{\varepsilon_i\}$ is a sequence of random variables. In fact, the nonparametric volatility from model (1.1) can be re-written as follows:

$$(Y_i - \mu(X_i))^2 = \sigma_0^2(X_i) + \sigma_0^2(X_i)(\varepsilon_i^2 - 1). \tag{2.1}$$

If $\mu(\cdot)$ is known, model (2.1) also is a nonparametric regression similar to (1.1).

In order to simplify the derivation of the ordinary least estimator (OLS) for change points in the regression equation, we first assume that $\sigma_0(x)$ is known. From (2.1), the LS estimator $\hat{k}$ of the change point k_0 in the above model can be defined as

$$\hat{k} = \arg\min_k \left[\min_{\tau_1,\tau_2} \left\{ \sum_{t=1}^{k} (Z_t^2 - \tau_1^2)^2 + \sum_{t=k+1}^{T} (Z_t^2 - \tau_2^2)^2 \right\} \right] \tag{2.2}$$

in which $Z_t = (Y_t - \mu(X_t))/\sigma_0(X_t)$.

The generalized LS estimator $\hat{k}$ of the change point k_0 can also be defined as (2.2) if $\mu(\cdot)$ and $\sigma_0(\cdot)$ were known. Thus, the jump point is estimated by minimizing the sum of squares of residuals among all possible sample slits. Write

$$S_T = \sum_{t=1}^{T} Z_t^2, \quad S_k = \sum_{t=1}^{k} Z_t^2 \quad \text{and} \quad S_{T-k} = \sum_{t=k+1}^{T} Z_t^2.$$

The variance of the first k observations is estimated by $S_k \sigma_0^2(x)/k$, and the variance of the last $T-k$ observations is estimated by $S_{T-k}\sigma_0^2(x)/(T-k)$, as the jump point k and $\sigma_0(x)$ are known. For some k, the LS estimators of $\tau_1^2 (t < k)$ and τ_2^2 $(t > k)$ should be $\bar{Z}_{1,k}$ and $\bar{Z}_{k+1,T}$, respectively, where $\bar{Z}_{1,k} = (1/k)S_k$ and $\bar{Z}_{k+1,T} = (1/(T-k))S_{T-k}$. Then, formula (2.2) can be written as follows:

$$\hat{k} = \arg\min_k \left(\sum_{t=1}^{k} (Z_t^2 - \bar{Z}_{1,k})^2 + \sum_{t=k+1}^{T} (Z_t^2 - \bar{Z}_{k+1,T})^2 \right) = \arg\min U_k^2,$$

where

$$U_k^2 = \sum_{t=1}^{k} (Z_t^2 - \bar{Z}_{1,k})^2 + \sum_{t=k+1}^{T} (Z_t^2 - \bar{Z}_{k+1,T})^2.$$

Simple algebra yields

$$U_k^2 = \sum_{t=1}^{T} (Z_t^2 - \bar{Z})^2 - TV_k^2,$$

where $\bar{Z} = (1/T) \sum_{t=1}^{T} Z_t^2$. Hence, we can obtain an estimator of the jump point in volatility from the definition of the LS estimator as follows:

$$\hat{k} = \arg\max_k |V_k|, \tag{2.3}$$

where

$$V_k = \left(\frac{k(T-k)}{T^2}\right)^{1/2} \left(\frac{1}{T-k} S_{T-k} - \frac{1}{k} S_k\right). \tag{2.4}$$

Thus V_k can be used to detect the change in the volatility. In fact, if H_0 holds, then $S_{T-k}\sigma_0(x)/(T-k)$ and $S_k\sigma_0(x)/k$ are unbiased estimators for the common volatility. Also, the difference, $(1/(T-k))S_{T-k} - (1/k)S_k$ (and thus V_k), is close to 0 under H_0 and will be different from 0 if the volatility changes.

Further, by simple calculations, we obtain

$$\begin{aligned} V_k &= \left(\frac{(T-k)k}{T^2}\right)^{1/2} \left(\frac{1}{T-k} S_{T-k} - \frac{1}{k} S_k\right) \\ &= \left(\frac{1}{k(T-k)}\right)^{1/2} S_T D_k, \end{aligned} \tag{2.5}$$

where

$$D_k = \frac{k}{T} - \frac{S_k}{S_T}. \tag{2.6}$$

Similarly, the maximizer of D_k may be viewed as an estimator for the change point k_0, which can be written as

$$\hat{k}^* = \arg\max_k |D_k| = \arg\max_k (k(T-k))^{1/2} |V_k|. \tag{2.7}$$

D_k is an important statistic for detecting the change point in volatility. Under some conditions, D_k can be viewed as an approximate likelihood ratio statistic for testing the null hypothesis. Following Inclan and Tiao (1994), we can show that for a fixed k, D_k can be written as a function of the usual F-statistics for testing the equality of variances between two independent samples.

Before we examine the asymptotic properties of the statistics, D_k and V_k, we consider a general version of these statistics, as it has some relative advantages to be explained below. In general, we can define a statistic for detecting the change in variance as

$$V_k^{v} = \left(\frac{k}{T}\left(1 - \frac{k}{T}\right)\right)^{1/2-v} V_k, \quad \text{for} \quad 0 \leqslant v \leqslant 1/2. \tag{2.8}$$

Note that $V_k = V_k^{1/2}$ and $D_k = V_k^0$. We only consider the unifying statistic V_k^v because it is flexible.

Now, we can obtain an estimator for the change point k_0 from (2.8), defined by

$$\hat{k}(v) = \arg\max_k |V_k^v|. \tag{2.9}$$

In testing hypothesis H_0: $\lambda_T = 0$ versus H_1: $\lambda_T \neq 0$, where $\lambda_T = \tau_2^2 - \tau_1^2$, the advantage of using factor $(k/T(1-k/T))^{1/2-v}$ with $0 \leqslant v \leqslant 1/2$ is that V_k^v is sensitive with respect to the contiguous alternative $\lambda_T \sim T^{-1/2}$ (see Corollary 2.2 below, or Csörgő and Horváth, 1997; Antoch and Hušková, 1995). We can use both statistics, V_k^v and $\hat{k}(v)$, to deal with tests and estimates of change points.

Given $\hat{k}(v)$, we may obtain estimators of τ_1^2 and τ_2^2, as follows:

$$\hat{\tau}_1^2(v) = \frac{1}{\hat{k}(v)} S_{\hat{k}(v)}, \qquad \hat{\tau}_2^2(v) = \frac{1}{T - \hat{k}(v)} S_{T-\hat{k}(v)}.$$

We can easily show that $\hat{\tau}_1^2(v)$ and $\hat{\tau}_2^2(v)$ are the consistent estimators of τ_1^2 and τ_2^2, respectively. From propositions below, we observe that the asymptotic distribution of $\hat{k}(v)$ $(0 \leqslant v \leqslant 1/2)$ is skewed. In our simulation, we will also examine the performance of the factor $(k/T(1-k/T))^{1/2-v}$ for different values of v, which affects the asymptotic behavior of $\hat{k}(v)$. In the following subsection, we first derive the asymptotic distribution of V_k^v for $0 \leqslant v \leqslant 1/2$ so that we can determine the critical values of the asymptotic distribution. When the value of $\max_{1 \leqslant k \leqslant T}|V_k^v|$ is large, we will reject H_0.

2.1. Asymptotic distribution for V_k^v

Under assumptions (AS.1–AS.8) listed in the appendix, we will show the asymptotic distribution of V_k^v along with consistency, convergence and asymptotic distribution of estimators of change points in volatility. Now, we obtain some basic results for V_k^v when the regression and conditional variance functions are known. Let

$$V_T^v(t) = \begin{cases} 0 & \text{if} \quad 0 \leqslant t < 1/(T+1), \\ V_{[(T+1)t]}^v & \text{if} \quad 1/(T+1) \leqslant t < T/(T+1), \\ 0 & \text{if} \quad T/(T+1) \leqslant t \leqslant 1, \end{cases}$$

where the operator $[\,\cdot\,]$ denotes the largest integer less than or equal to its argument. Similarly, we define $V_T(t)$ and $D_T(t)$ for $0 \leqslant t \leqslant 1$. Therefore, $D_T(t)$, $V_T(t)$ and $V_T^v(t)$ for $0 \leqslant t \leqslant 1$ are three functions which are right continuous with left limits. Let $\{B(t), 0 \leqslant t \leqslant 1\}$ denote a standard Brownian motion on $D[0,1]$. We then obtain the following result:

Proposition 2.1. *Assume* (AS.1) *in the appendix. If* $\tau_1 = \tau_2 = 1$, *i.e., under* H_0 (*that is, there is no change in volatility*), *then*

$$\lim_{T\to\infty} \frac{\sqrt{T}\sigma}{\sigma_w} |V_T^v(t)| \;\; \Rightarrow \;\; (t(1-t))^{-v}|B(t)| \;\, in, \tag{2.10}$$

where $\Rightarrow$ *denotes weak convergence in probability space* $D[\delta, 1-\delta]$ *for some* $\delta > 0$, $\sigma = \lim_{T\to\infty} S_T/T$ *and*

$$\sigma_w^2 = \mathrm{E}(Z_1^2 - \mathrm{E}Z_1^2)^2 + 2\sum_{i=2}^{T} \mathrm{E}((Z_1^2 - \mathrm{E}Z_1^2)(Z_i^2 - \mathrm{E}Z_i^2)). \tag{2.11}$$

The result of Proposition 2.1 gives the asymptotic distribution of the test statistic. This is an application of a multivariate functional central limit theorem of the mixing sequence in Wooldridge and White (1988). Since this distribution is the same as that based on parametric models, we can easily obtain the asymptotic critical values for the proposed tests from the tabulated critical values (see Csörgő and Horváth, 1997; Inclan and Tiao, 1994). We need to estimate the unknown σ_w if we wish to use this result for testing H_0 against H_1 (i.e., there exists a change point in volatility). When $\{Z_i, i = 1, \ldots, T\}$ is independent, it is easy to derive the estimator of σ_w. When dependence is present, we can use the procedure proposed by Peligrad and Shao (1995) for mixing data to construct an estimator for σ_w. An estimator $\hat{\sigma}$ for σ is S_T/T. It follows from Proposition 2.1 that for any $\delta > 0$,

$$\lim_{T\to\infty} \frac{\sqrt{T}\sigma}{\sigma_w} \sup_{\delta<t<1-\delta} |V_T^v(t)| \xrightarrow{\mathrm{d}} \sup_{\delta<t<1-\delta} (t(1-t))^{-v}|B(t)|,$$

$$\lim_{T\to\infty} \frac{\sqrt{T}}{\sigma_w}|D_T(t)| \quad \Rightarrow \quad |B(t)| \quad \text{in} \quad D[0,1],$$

$$\lim_{T\to\infty} \frac{\sqrt{T}}{\sigma_w} \sup_{0<t<1} |D_T(t)| \xrightarrow{\mathrm{d}} \sup_{0<t<1} |B(t)|,$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution. Note that $D[\delta, 1-\delta]$ may be changed into $D[0,1]$ when $v=0$. These results show that there is a smaller loss of the boundary effect for the estimator of change points based on $D_T(t)$ than on $V_T^v(t)$ for different values of v. This point will be addressed again in our simulations. It is easy to show that $\hat{k}(1/2) = \arg\max_k |V_k^v|$ $(v=1/2)$ is the maximum likelihood estimator for k_0 if the errors are independently and normally distributed. But we need a slightly stronger assumption for the jump size λ_T in order to derive the asymptotic properties of estimator $\hat{k}(1/2)$.

According to (1.3.26) of Csörgő and Horváth (1997), the distribution of $\sup_{h<t<1-l}$ $(t(1-t))^{-1/2}|B(t)|$ has an approximation formula. Hence, we can get the asymptotic critical values. The approximated distribution of $sup_{h<t<1-l}(t(1-t))^{-1/2}|B(t)|$ is provided as follows:

$$P\left\{ \sup_{h<t<1-l} \left(\frac{B(t)}{t(1-t)} \right)^{1/2} \geqslant x \right\}$$

$$= \frac{x\exp\{-x^2/2\}}{(2\pi)^{1/2}} \left\{ \log \frac{(1-h)(1-l)}{hl} \right.$$

$$\left. - \frac{1}{x^2} \log \frac{(1-h)(1-l)}{hl} + \frac{4}{x^2} + \mathrm{O}\left(\frac{1}{x^4}\right) \right\},$$

as $x \to \infty$. In addition, to obtain the critical values, one often performs Monte Carlo simulations to find a good approximation for the quantile $z_n = z_n(1-\alpha)$, where

$$z_n = \sup\left\{x : \mathrm{P}\left(\frac{\sqrt{T}}{\sigma_w} \sup_{\delta < t < 1-\delta} |V_T(t)| < x\right) \leqslant 1-\alpha\right\}$$

and σ_w needs to be estimated in applications. A critical issue here is how to choose h and l (e.g., δ in Proposition 2.1). Csörgő and Horváth (1997) found that $h_T = l_T = (\log T)^{3/2}/T$ was a good choice for all of their cases. Therefore, we also employ the same choice for δ.

2.2. Asymptotic properties of estimators of change points in volatility

It is easy to prove the consistency of $\hat{k}(v)$ when the regression function and the conditional variance function $\sigma_0(x)$ are known. Using the same argument as in Antoch and Hušková (1995) and Bai (1994), we can verify the following propositions. We need to impose a few assumptions on the magnitude of a shift $\lambda_T = \tau_2^2 - \tau_1^2$, in order to derive the rate of convergence.

Condition J_1: The magnitude of a jump in the conditional variance function, $\lambda_T = \tau_2^2 - \tau_1^2$, is constant.

Condition J_2: The magnitude of a jump converges to zero as the sample size grows unbounded; i.e., $\lambda_T \to 0$, and $T\lambda_T^2/\log\log T \to \infty$ as $T \to \infty$.

Proposition 2.2. *Under* (AS.1) *and Condition* J_1 *or* J_2,

$$\hat{k}(v) - k_0 = \mathrm{O_P}(1/\lambda_T^2)$$

for $0 \leqslant v \leqslant 1/2$. $k_0 = [\theta_0 T]$ *with some* $0 < \theta_0 < 1$.

For $0 \leqslant v < 1/2$, Condition J_2 can be weakened so that $\lambda_T \to 0$ and $T\lambda_T^2 \to \infty$ as $T \to \infty$. The same assumption regarding λ_T applies to the subsequent propositions and corollaries. In fact, the estimator of the change point, $\hat{\theta}_v (=\hat{k}(v)/T$ for $0 \leqslant v \leqslant 1/2)$, converges to the true change point θ_0 at the rate of $(T\lambda_T^2)^{-1}$. That is,

$$|\hat{\theta}_v - \theta_0| = \mathrm{O_P}((T\lambda_T^2)^{-1}).$$

The rate of convergence not only describes how fast the estimator converges to the true value, but it is also necessary in deriving the asymptotic distribution of the estimator. Similar results are obtained for identically and independently distributed (i.i.d.) models with linear structures having a shift in mean, as in Yao (1987) and Bhattacharya (1987). Bai (1994) also obtained similar results for a mean shift in linear processes, while Bai (2000) had similar results for Gaussian autoregression models. For simplicity, let $\theta{=}\theta_0$.

Corollary 2.1. *Under the assumptions of Proposition* 2.2,

$$\sqrt{T}\begin{pmatrix} \hat{\tau}_1^2(v) - \tau_1^2 \\ \hat{\tau}_2^2(v) - \tau_2^2 \end{pmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Sigma), \tag{2.12}$$

where

$$\Sigma = \begin{pmatrix} \theta^{-1}\sigma_{1w}^2 & Q \\ Q & (1-\theta)^{-1}\sigma_{2w}^{*2} \end{pmatrix}$$

in which

$$\sigma_{1w}^{*2} = \lim_{T\to\infty}\left[\mathrm{E}(Z_1^2 - \mathrm{E}Z_1^2)^2 + 2\sum_{i=2}^{k_0}\mathrm{E}((Z_{k_0}^2 - \mathrm{E}Z_{k_0}^2)(Z_i^2 - \mathrm{E}Z_i^2))\right],$$

$$\sigma_{2w}^{*2} = \lim_{T\to\infty}\left[\mathrm{E}(Z_{k_0+1}^2 - \mathrm{E}Z_{k_0+1}^2)^2 + 2\sum_{i=k_0+2}^{T}\mathrm{E}((Z_{k_0+1}^2 - \mathrm{E}Z_{k_0+1}^2)(Z_i^2 - \mathrm{E}Z_i^2))\right],$$

$$Q = \lim_{T\to\infty}\frac{1}{k_0(T-k_0)}\sum_{i=1}^{k_0}\sum_{j=k_0+1}^{T}(Z_i^2 - \mathrm{E}Z_i^2)(Z_j^2 - \mathrm{E}Z_j^2).$$

Corollary 2.1 shows that the asymptotic distribution of change factors in volatility follows the normal distribution.

Corollary 2.2. *Under the assumptions of Proposition* 2.2,

$$\sqrt{T}\sup_{0<t<1}|V_k^v(t)| - T^{1/2}|\lambda_T|C \xrightarrow{\mathrm{P}} 0$$

under the alternative hypothesis H_1, *where* C *is some positive constant*.

The result of Corollary 2.2 implies that $\sqrt{T}\hat{\sigma}\sup_{0<t<1}|V_k^v(t)|/\hat{\sigma}_w \to \infty$ as $T\lambda_T^2 \to \infty$, where $\hat{\sigma}$ and $\hat{\sigma}_w$ are the consistent estimators of σ and σ_w, respectively. Furthermore, it implies that the proposed tests are consistent with convergence rate of almost $T^{-1/2}$. Our proposed test is more powerful than the one of Perron (2001). Specifically, our proposed test has a power of order $\mathrm{O}(T^{1/2}|\lambda_T|)$ for $0 \leqslant v \leqslant 1/2$ when the test has nontrivial power. We can test the jump with nontrivial significant levels when $\lambda_T = \bar{\mathrm{O}}(T^{-1/2})$ for $0 \leqslant v < 1/2$ and $\lambda_T = \bar{\mathrm{O}}(T^{-1/2}(\log\log T)^{1/2})$ for $v = 1/2$, where $a_T = \bar{\mathrm{O}}(b_T)$ denotes $a_T/b_T \to \infty$ as a_T and $b_T \to 0$. The test by nonparametric procedure has only a local power of order $\bar{\mathrm{O}}((Tb^{p+1})^{1/2}|\lambda_T|)$, where b is a bandwidth in the nonparametric estimator, and p is the dimension of X (see Theorem 1 of Perron 2001). This implies that the test proposed by Perron can only identify the jump with $\lambda_T = \bar{\mathrm{O}}((Tb^{p+1})^{-1/2})$ at nontrivial significant levels. However, our proposed test has a local power of at least order $\bar{\mathrm{O}}(T^{1/2}(\log\log n)^{-1/2}|\lambda_T|)$, which is much better than $\bar{\mathrm{O}}((Tb^{p+1})^{1/2}|\lambda_T|)$ even if $p = 1$.

Next, we shall derive the asymptotic distribution of $\hat{k}(v)$. Let

$$g_v(t) = \begin{cases} (1-v)(1-\theta) + v\theta & \text{if } t \leqslant 0, \\ (1-v)\theta + v(1-\theta) & \text{if } t > 0 \end{cases}$$

and

$$B_s(t) = \begin{cases} \tau_1^2 B_1^*(-t), & t \leqslant 0, \\ \tau_2^2 B_2^*(t), & t > 0, \end{cases}$$

where $B_i^*(t)(i=1,2)$ are two independent standard Brownian motions defined on $[0,\infty)$ with $B_i^*(0)=0,\ i=1,2$. There is an almost surely unique random variable η_v, such that

$$\eta_v = \arg\max_t (B_s(t) - g_v(t)|t|).$$

We define a stochastic process $W^*(m)$ on the set of integers in order to describe our next result. The stochastic process is defined as follows: $W^*(0)=0$, $W^*(m)=W_1(m)$ for $m<0$, and $W^*(m)=W_2(m)$ for $m>0$, in which

$$W_1(m) = 2c_\theta\{g_v(m)m\lambda_T^2 + 2\tau_1^2 \sum_{i=m+1}^{0} (\varepsilon_i^2 - \mathrm{E}\varepsilon_i^2)\lambda_T\}, \quad m=-1,-2,\ldots$$

and

$$W_2(m) = -2c_\theta\{g_v(m)m\lambda_T^2 + 2\tau_2^2 \sum_{i=1}^{m} (\varepsilon_i^2 - \mathrm{E}\varepsilon_i^2)\lambda_T\}, \quad m=1,2,\ldots,$$

where $c_\theta = (\theta(1-\theta))^{1-2v}$, $0 \leqslant v \leqslant 1/2$, provided that $\{\varepsilon_i\}$ is a sequence of strictly stationary random variables. The next result yields the asymptotic distribution of $\hat{k}(v)$ for $0 \leqslant v \leqslant 1/2$.

Proposition 2.3. (a) *Assume* (AS.1) *and condition* J_1. *If* $\{\varepsilon_i\}$ *is a sequence of strictly stationary random variables, then for* $0 \leqslant v \leqslant 1/2$,

$$\frac{\lambda_T^2(\hat{k}(v)-k_0)}{\sigma_w^2} \xrightarrow{\mathrm{d}} \arg\max_m W^*(m).$$

(b) *Under* (AS.1) *and condition* J_2, *for* $0 \leqslant v \leqslant 1/2$,

$$\frac{\lambda_T^2(\hat{k}(v)-k_0)}{\sigma_w^2} \xrightarrow{\mathrm{d}} \eta_v,$$

where

$$\sigma_w^2 = \mathrm{E}(Z_1^2 - \mathrm{E}Z_1^2)^2 + 2\sum_{i=2}^{k_0} \mathrm{E}((Z_1^2 - \mathrm{E}Z_1^2)(Z_i^2 - \mathrm{E}Z_i^2)).$$

This result establishes the asymptotic distributions of the estimators of change point. We can derive the closed form of the distributions of random variable η_v when $\lambda_T \to 0$. Even if λ_T is a constant, we can use the simulation method to approximate the distribution of W^*. The distribution and density of η_0 (i.e. $v=1/2$) have been derived by Yao (1987), Csörgő and Horváth (1997) and Bai (1997). Employing the method of Bai (1997), it is not difficult to derive the distribution and density functions of η_v $(0 \leqslant v \leqslant 1/2)$ along with the confidence intervals of the change point. But all unknown quantities such as λ_T and σ_w^2 need to be consistently estimated (see Peligrad and Shao, 1995). Assume that $\hat{\sigma}_w^2$ is the consistent estimator of σ_w^2, and $\hat{\lambda}_T(=\hat{\tau}_2^2 - \hat{\tau}_1^2)$ is the consistent estimator of λ_T. Then a $100(1-\alpha)\%$ confidence interval is given by

$$[\hat{k}(v) - [c_\alpha/\hat{\Gamma}] - 1, \hat{k}(v) + [c_\alpha/\hat{\Gamma}] + 1],$$

where c_α is the $(1-\alpha/2)$th quantile of the random variable η_v and $\hat{\Gamma} = \hat{\lambda}_T^2/\hat{\sigma}_w^2$.

Although we obtain the explicit distribution function of the asymptotic random variable η_v, it includes many unknown parameters that need to be estimated in order to get the critical values c_α. An alternative procedure to get the critical value c_α is to perform a bootstrapping approximation. Antoch and Hušková (1995) constructed a bootstrapping estimator $\hat{k}^*(v)$ for k_0 which is shown to uniformly converges to η_v. Hence, the quantiles of $\hat{k}^*(v)$ can be used to approximate the quantiles of η_v, from which we can obtain the approximated critical value c_α.

We have not imposed any assumptions on the sequence $\{X_t\}$, implying that $\{X_t\}$ may be a sequence of a multi-dimension vector. The consequences of part (a) of Propositions 2.3 can be extended to nonstationary data. An example is when $\{\varepsilon_t\}$ is a sequence of random variables with the same distribution as ε_{k_0} for $t \leqslant k_0$ and as ε_{k_0+1} for $t > k_0$, respectively.

Finally, an equivalent estimator of change points can be constructed as if there is no heteroscedasticity in the model.[3] The estimator of a change point is defined by

$$\bar{k} = \arg\min_k \left\{ \sum_{t=1}^{k} \left(W_t^2 - k^{-1} \sum_{t=1}^{k} W_t^2 \right)^2 + \sum_{t=k+1}^{T} \left(W_t^2 - (T-k)^{-1} \sum_{t=k+1}^{T} W_t^2 \right)^2 \right\}, \tag{2.13}$$

where $W_t = Y_t - \mu(X_t)$ when $\mu(\cdot)$ is known. Similarly, we can derive a test statistic from $\bar{k}$. That is,

$$\bar{V}_k^v = \left(\frac{k(T-k)}{T^2} \right)^{1-v} \left(\frac{1}{T-k} R_{T-k} - \frac{1}{k} R_k \right),$$

where R_{T-k} and R_k are the same as S_{T-k} and S_k with Z_t replaced by W_t. From the proofs of Propositions 2.1–2.3, we easily show that Propositions 2.1–2.3 hold with the test statistic $\bar{V}_k^v$ and estimator $\bar{k}$, but in Corollary 2.1, τ_1^2 and τ_2^2 should be replaced by $\tau_1^2\sigma_*^2$ and $\tau_2^2\sigma_*^2$, respectively, where $\sigma_*^2 = \mathrm{E}\sigma_0^2(X_i)$. The advantage of this estimator is that it is not necessary to predetermine an estimate of the unknown conditional variance. This has a significant implication for obtaining consistent estimators of change points. We return to this estimator in the next section. One disadvantage is the complicated calculation for σ_w^2 (see (2.11)). The following proposition summarizes the above discussion.

[3] We are grateful to an anonymous referee who has suggested this estimator in order to overcome the difficulty of estimating the conditional variance when there is a break in volatility.

Proposition 2.4. *Suppose that the corresponding assumptions of Propositions* 2.1–2.3 *and Corollaries* 2.1–2.3 *are satisfied. Then the results of Propositions* 2.1–2.3 *and Corollaries* 2.1–2.2 *hold for* $\bar{V}_k^v$ *and,* $\bar{k}$, *respectively, but* τ_1^2 *and* τ_2^2 *in Corollary* 2.1 *should be replaced by* $\tau_1^2\sigma_*^2$ *and* $\tau_2^2\sigma_*^2$.

3. Estimation for change points with unknown volatility

3.1. Estimating change points based on estimated residuals

It is natural to think of nonparametric estimators of change points based on the full sample, such as $\hat{k}(v)$. However, the estimator might not be consistent when we use estimated residuals such as $\hat{S}_k = \sum_{t=1}^{k} \hat{Z}_t^2$ and $\hat{S}_{T-k} = \sum_{t=k+1}^{T} \hat{Z}_t^2$ with $\hat{Z}_t = (Y_i - \hat{\mu}_T(X_t))/\hat{\sigma}_T(X_t)$, where $\hat{\mu}_T(X_t)$ and $\hat{\sigma}_T(X_t)$ are nonparametric estimators of the conditional mean and variance constructed using the full sample. Given a change point, k_0, $sup_k(1/k)\sum_{t=1}^{k}(\hat{Z}_t^2 - Z^2) \neq o_P(1)$ in general, as $\sup_k(1/k)\sum_{t=1}^{k}(\hat{\sigma}_T^2(X_k) - \sigma_T^2(X_k)) \neq o_P(1)$. Therefore, there is no guarantee that we will obtain the same result as in Propositions 2.2–2.4 with $\hat{Z}_t^2$.[4]

Now let $W_t = Y_t - \mu(X_t) = \sigma_0(X_t)\varepsilon_t$ and $\hat{W}_t = Y_t - \hat{\mu}(X_t)$. Note that $E(W_t^2) = E(\sigma_0(X_t)^2 E(\varepsilon_t^2|X_t)) = \tau_i E(\sigma_0(X_t)^2)$ with $\tau_i = \tau_1(\tau_2)$ for $t \leqslant k_0 (t \geqslant k_0)$. Therefore, there is a change in the volatility of W_t when there is a change in the volatility of Z_t. Moreover, the change-point locations in the volatility should be the same for W_t and Z_t. Therefore, we can estimate the change point using $\hat{W}_t$ instead of $\hat{Z}_t$.

We can rewrite (2.13) by replacing W_t with $\hat{W}_t$ as follows:

$$\tilde{k} = \arg\min_k \left\{ \sum_{t=1}^{k} \left(\hat{W}_t^2 - k^{-1} \sum_{t=1}^{k} \hat{W}_t^2 \right)^2 + \sum_{t=k+1}^{T} \left(\hat{W}_t^2 - (T-k)^{-1} \sum_{t=k+1}^{T} \hat{W}_t^2 \right)^2 \right\}, \tag{3.1}$$

where $\hat{W}_t = Y_t - \hat{\mu}(X_t)$, in which $\hat{\mu}(\cdot)$ is the nonparametric estimator in (3.3) in the following subsection. From this formula, we can derive a test statistic ($\tilde{V}_k^v$) as follows:

$$\tilde{V}_k^v = \left(\frac{k(T-k)}{T^2} \right)^{1-v} \left(\frac{1}{T-k} \tilde{R}_{T-k} - \frac{1}{k} \tilde{R}_k \right), \tag{3.2}$$

where $\tilde{R}_{T-k}$ and $\tilde{R}_k$ are the same as R_{T-k} and R_k with W_t replaced by $\hat{W}_t$, respectively. Similarly, we propose the estimator of the change point as $\tilde{k}(v)$, defined by $\tilde{k}(v) = \arg\max_k |\tilde{V}_k^v|$. Hence, we can define $\tilde{V}_T^v(t)$ as $\hat{V}_T^v(t)$, with $\hat{V}_k^v$ replaced by $\tilde{V}_T^v$.

The following proposition is our main result. It shows that even when the unknown functions are replaced by the corresponding nonparametric estimators, the consistency and asymptotic distribution of the proposed estimators remain unchanged.

[4] We thank an anonymous referee for the insightful discussion on the desirability of this estimator.

Proposition 3.1. *Assume that assumptions* (AS.2)–(AS.8) *in the appendix and conditions J_i corresponding to those of Propositions* 2.1–2.4 *and Corollaries* 2.1–2.2 *are satisfied. The consequences of Propositions* 2.1–2.4 *and Corollaries* 2.1–2.2 *then hold for statistics $\tilde{V}_k^v$ and $\tilde{k}(v)$, $0 \leqslant v \leqslant 1/2$, respectively.*

From Proposition 3.1, the estimator of the change point $\tilde{k}(v)$ is consistent. Hence, we can first estimate the change point by $\tilde{k}(v)$, and then the conditional mean and variance, which we will examine in the next section.

3.2. Nonparametric estimation for conditional mean and volatility

Now let us estimate $\mu(x)$ and $\sigma_0(x)$ by a nonparametric technique. For the sake of simplicity, we assume that $\tau_1 = 1$ and $\mu(x)$ and $\sigma_0(x)$ are smooth, and further that X_i has a density function $f(x), x \in [a_1, a_2]$. For model (1.1), along with a change point k_0 in volatility, it is easy to show that $\{Y_i, i \leqslant k_0\}$ and $\{Y_i, i > k_0\}$ are strictly stationary.

Hence, a suitable nonparametric estimator of the regression function can be obtained by

$$\hat{\mu}_n(x) = \frac{\sum_{i=1}^n \mathscr{K}_{n,h}(X_i - x) Y_i}{\sum_{i=1}^n \mathscr{K}_{n,h}(X_i - x)}, \tag{3.3}$$

where $\mathscr{K}_{n,h}(\cdot)$ may take two different values. When $\mathscr{K}_{n,h}(\cdot)$ satisfies

$$\mathscr{K}_{n,h}(x) = \frac{1}{h} K\left(\frac{X_i - x}{h}\right), \tag{3.4}$$

where $K(\cdot)$ is a kernel function and $h{=}h_n$ is a sequence of bandwidths, (3.3) is referred to as the kernel estimator. When $\mathscr{K}_{n,h}(\cdot)$ satisfies

$$\begin{aligned} \mathscr{K}_{n,h}(X_i - x) = K_h(X_i - x) \sum_{j=1}^n K_h(X_j - x)(X_j - x)^2 \\ - K_h(X_i - x)(X_i - x) \sum_{j=1}^n K_h(X_j - x)(X_j - x) \end{aligned} \tag{3.5}$$

with $K_h(\cdot) = K(\cdot/h)$, in which $K(\cdot)$ and h are the same as those of (3.4), then (3.3) is the local linear estimator (Fan and Gijbels, 1996). Furthermore, because $\mu(x)$ is smooth, we can take $n = T$. Under some assumptions, we can easily show that $\hat{\mu}_n(x)$ is a consistent estimator of $\mu(x)$.

There are many studies such as those of Härdle and Tsybakov (1997), Fan and Yao (1998), Pagan and Hong (1991) and Pagan and Ullah (1988) that propose the estimates of volatility $\sigma_0^2(x)$ (conditional variance) in the nonparametric model. However, with a change point in volatility, we propose the following estimator (see Fan and Yao, 1998), assuming $\mathrm{E}(\varepsilon^2|X) = 1$:

$$\hat{\sigma}_s^2(x) = \frac{\sum_{j=1}^n \mathscr{W}_{n,b}(X_i - x)(Y_i - \hat{\mu}(X_i))^2}{\sum_{j=1}^n \mathscr{W}_{n,b}(X_i - x)}, \tag{3.6}$$

where $1 \leqslant n \leqslant k_0, n \to \infty$, as $T \to \infty$ under the alternative hypothesis, and $n = T$ under the null hypothesis; $\hat{\mu}(x)$ is a kernel estimator or local linear estimator of $\mu(x)$; $\mathscr{W}_{n,b}$ is the same as $\mathscr{K}_{n,h}$ with kernel $K(\cdot)$ replaced by $W(\cdot)$. $W(\cdot)$ is a kernel function that may or may not take the same form as $K(\cdot)$ of $\hat{\mu}(x)$; and b is a bandwidth that differs from the one for $\mu(x)$.

However, this estimator, which corresponds to the local constant smoother, is not robust. It is well known that the smoothers of the second-order polynomial Y_i^2 or $(Y_i - \hat{\mu}(X_i))^2$ are sensitive to outliers. In particular, when there are some outliers in the observations or when the distribution function of observation data is heavy-tailed, the estimator will result in a very large bias. Therefore, we may consider the following absolute deviation estimator (see Xia et al., 1998):

$$\hat{\sigma}_{\mathrm{d}}(x) = \frac{\sum_{j=1}^{n} \mathscr{W}_{n,b}(X_i - x)|Y_i - \hat{\mu}(X_i)|}{\sum_{j=1}^{n} \mathscr{W}_{n,b}(X_i - x)}, \tag{3.7}$$

where $W_{n,b}(\cdot)$ and b are the same as in (3.6), and assume $\mathrm{E}(|\varepsilon_1||X) = 1$. Similarly, it contains the local constant smoother (kernel estimator) or local linear estimator.

In summary, when $\mu(x)$ and $\sigma_0(x)$ are unknown, we can reconstruct the statistics $\tilde{k}(v)$ and $\tilde{V}_k^v$ to start the following algorithm. The advantage is that it is not necessary to obtain the estimator of $\sigma_0(x)$ in the first step.

Algorithm X. 1. *Calculate the estimator* $\mu(x)$ *by* (3.3), *using the entire sample*, $X_1, \ldots, X_T$, *and give the values of the estimators*, $\hat{\sigma}$ *and* $\hat{\sigma}_w$. *Calculate the statistics*, $M_T(t) = \sqrt{T}\hat{\sigma}\tilde{V}_T^v(t)/\hat{\sigma}_w$, *for* $\delta_T < t < 1 - \delta_T$, *where* $\delta_T = (\log T)^{3/2}T$.

2. *If* $\sup_{\delta_T < t < 1-\delta_T} \sqrt{T}\hat{\sigma}/\hat{\sigma}_w |M_T(t)| \leqslant c_\alpha$, *where* c_α *is the critical value of the asymptotic distribution* $\sup_{\delta < t < 1-\delta}(t(1-t)^{-v}|B_T(t)|$, *then the test is not significant at the given significant level* (*e.g.*, $\alpha = 0.05$). *Let* $\tilde{k}(v) = T$. *Otherwise*, *we define an estimator* $\tilde{k}(v) = \arg\max_{\delta_T < t < 1-\delta_T} |\tilde{V}_T^v(t)|$. *Then go to Step* 3.

3. *Calculate* $\sigma_0(x)$ *by* (3.7), *and then give the values of estimators*, $\hat{\sigma}$ *and* $\hat{\sigma_w}$, *based on subsample*, $X_1, \ldots, X_{\tilde{k}(v)}$.

4. *Stop the program.*

Hence, we can obtain the estimators $\tilde{k}(v)$, $\hat{\mu}(x)$ and $\hat{\sigma}(x)$.

4. Detection of multiple changes

It is natural to extend our analysis to the detection of multiple changes in volatility. We first construct the detection statistic $\tilde{V}_k^v$ and use the iterated cumulative sums of squares (ICSS) algorithm proposed by Inclan and Tiao (1994) to detect multiple change points. The key to applying the ICSS algorithm in practice is to give the appropriate critical values according to the sizes of the samples. In fact, the asymptotic distributions of the function $\tilde{V}_k^v$ play a key role in utilizing the ICSS algorithm. In the latter part of this section, we give a modified ICSS algorithm based on the statistic $\tilde{V}_k^v$. In the simulation section, we shall use the modified ICSS algorithm to check the change points in volatility of the HSI index prices.

We extend the procedure and results above to general multiple break points in volatility. Let

$$\sigma(X_i) = \begin{cases} \tau_1\sigma_0(X_i) & \text{if } i \leqslant k_1, \\ \tau_2\sigma_0(X_i) & \text{if } k_1 + 1 \leqslant i \leqslant k_2, \\ & \dots \\ \tau_{m+1}\sigma_0(X_i) & \text{if } k_{m+1} \leqslant i \leqslant T, \end{cases} \tag{4.1}$$

where $\tau_i \neq \tau_j, i, j = 1, 2, \dots, m$, $k_i = [T\theta_i]$, $\theta_i \in (0, 1)$ and $\theta_i < \theta_{i+1}$ for $i = 1, 2, \dots, m$ with $\tau_1 = 1$.

We propose a sequence procedure coupled with hypothesis testing and show that this sequential procedure yields a consistent estimate for the true number of change points. We assume that the true number of change points, m, is unknown. The procedure works as follows: When the first change point is identified, say $\tilde{k}_0$, the whole sample is divided into subsamples, with the first subsample consisting of the first $\tilde{k}_0$ observations and the second sample consisting of the rest of the observations. We then apply Proposition 3.1 to perform hypothesis testing to detect the change point for each subsample. If the new change point(s) has been detected, we further divide the corresponding subsample into new subsamples. In the new subsample, we perform the test of change points by Proposition 3.1. These steps are repeated until the null hypothesis test is not rejected for all subsamples. Hence, the number of break points is equal to the number of samples minus 1, $\hat{m}$. The procedure is simple and intuitive in practice. Further, we can prove that $\hat{m}$ converges to m in probability, as shown in the proposition below.

Proposition 4.1. *Suppose that the assumptions of Proposition* 3.1 *are met and that the size of the test* α_T *slowly converges to zero. Then, under model* (4.1) *we have*

$$\mathrm{P}(\hat{m} = m) \to 1.$$

Since the proof of this proposition is similar to that of Proposition 11 in Bai (1997), we omit it.

To end this section, we discuss the ICSS algorithm proposed by Inclan and Tiao (1994). This algorithm is an iterative scheme based on the successive application of $\hat{V}^v_k$ to pieces of the residual errors, dividing consecutively after a possible change point is determined. This is similar to the sequential procedure.

The whole modified ICSS algorithm may be stated here, but to save space we only deal with the *modified part* of the ICSS algorithm proposed by Inclan and Tiao (1994). D_k in the ICSS algorithm should be replaced by $\tilde{V}^v_k$, and critical values c^v_α are given by Table 1 in Inclan and Tiao (1994, p. 914) for $v = 0$, and in Table 1.3.1 in Csörgő and Horváth (1997, p. 25) for $v = 1/2$. The significant confidence level α is usually 0.05. The statistic $M(t_1, T)$ in step 1 of the ICSS algorithm is replaced by

$$M(t_1, T) = \max_{t_1 \leqslant k \leqslant T} \frac{\sigma\sqrt{T - t_1 + 1}}{\hat{\sigma}_w} |\tilde{V}^v_k(a[t_1 : T])|,$$

where the notation $a[t_1 : t_2]$ denotes a sample beginning from t_1 to t_2 ($t_1 < t_2$), and $\tilde{V}^v_k(a[t_1 : t_2])$ denotes the range over which the cumulative sums are obtained. $\hat{\sigma}$ and

Table 1
Estimation of jump points for model (5.7) with different sample sizes T having estimators $\tilde{k}^*$, $\tilde{k}(1/2)$ and $\tilde{k}(1/4)$ when $\mu(x) = 0$

T	τ	θ_0			
		0.3	0.5	0.7	0.85
Panel A: Estimates and standard error for $\tilde{k}(0)$					
100	0.1	0.40(0.112)	0.54(0.064)	0.71(0.052)	0.80(0.116)
	0.2	0.38(0.096)	0.53(0.046)	0.71(0.025)	0.84(0.038)
	0.5	0.37(0.083)	0.53(0.041)	0.71(0.019)	0.85(0.014)
200	0.1	0.36(0.079)	0.52(0.039)	0.71(0.027)	0.83(0.061)
	0.2	0.35(0.059)	0.52(0.029)	0.71(0.014)	0.84(0.017)
	0.5	0.34(0.051)	0.51(0.023)	0.70(0.014)	0.85(0.005)
400	0.1	0.33(0.043)	0.51(0.018)	0.70(0.013)	0.84(0.026)
	0.2	0.32(0.031)	0.50(0.014)	0.70(0.077)	0.85(0.008)
	0.5	0.32(0.029)	0.50(0.011)	0.70(0.049)	0.85(0.003)
Panel B: Estimates and standard error for $\tilde{k}(1/2)$					
100	0.1	0.41(0.166)	0.58(0.113)	0.75(0.081)	0.85(0.106)
	0.2	0.39(0.143)	0.56(0.091)	0.74(0.058)	0.87(0.023)
	0.5	0.37(0.123)	0.55(0.077)	0.74(0.054)	0.88(0.020)
200	0.1	0.36(0.114)	0.54(0.074)	0.73(0.052)	0.87(0.031)
	0.2	0.32(0.065)	0.53(0.051)	0.73(0.038)	0.87(0.024)
	0.5	0.32(0.053)	0.52(0.043)	0.72(0.029)	0.86(0.023)
400	0.1	0.31(0.038)	0.52(0.042)	0.72(0.033)	0.87(0.027)
	0.2	0.31(0.034)	0.51(0.015)	0.70(0.016)	0.86(0.020)
	0.5	0.31(0.017)	0.50(0.014)	0.70(0.014)	0.85(0.017)
Panel C: Estimates and standard error for $\tilde{k}(1/4)$					
100	0.1	0.40(0.129)	0.55(0.079)	0.72(0.059)	0.83(0.101)
	0.2	0.37(0.107)	0.54(0.061)	0.72(0.036)	0.86(0.031)
	0.5	0.36(0.092)	0.54(0.053)	0.72(0.032)	0.86(0.016)
200	0.1	0.36(0.088)	0.53(0.051)	0.72(0.028)	0.86(0.032)
	0.2	0.34(0.055)	0.52(0.035)	0.71(0.019)	0.86(0.016)
	0.5	0.33(0.045)	0.52(0.028)	0.71(0.025)	0.85(0.015)
400	0.1	0.31(0.037)	0.52(0.032)	0.71(0.021)	0.86(0.017)
	0.2	0.31(0.016)	0.51(0.013)	0.70(0.017)	0.86(0.018)
	0.5	0.31(0.013)	0.50(0.011)	0.70(0.015)	0.85(0.019)

In Table 1, $\theta_0 = k_0/T$ and $\tilde{\theta}(v) = \tilde{k}(v)/T$.

$\hat{\sigma}_w$ are the consistent estimators of σ and σ_w, respectively. The other steps are the same as those of the ICSS algorithm.

5. Simulations and real data examples

5.1. Selection of bandwidths in nonparametric estimation

An important issue concerning the application of theoretic findings in this paper is the selection of bandwidths in the estimation of the regression function and the heteroscedastic conditional variance function. The plug-in method based on MISE or MSE involves the additional unknown quantities, the second derivative $\mu''(x)$ and $\sigma_0(x)$. However, the estimation of these unknown quantities involves the selection of other bandwidths, h_1 and b, which are used to estimate the second-order derivatives $\mu''(x)$ and $\sigma_0^2(x)$, respectively. As a result, its application is very difficult, and the calculations are complex. Hence, in order to avoid complicated computations, we use the cross-validation method to select two bandwidths for the estimation of $\mu(x)$ and $\sigma_0^2(x)$. The selected bandwidth turns out to be satisfactory in our simulation and empirical results. The bandwidth selected for the estimator of $\mu(x)$ by the cross-validation method is

$$\hat{h} = \arg\min_h \sum_{i=1}^{n} [Y_i - \hat{\mu}_h(X_i)]^2 w(X_i), \tag{5.1}$$

where $\hat{\mu}_h(X_i)$ is calculated by (3.3) using the data $\{(X_t, Y_t), t \neq i\}$, and $w(x)$ is a given weight function, $0 < n \leqslant T$. After we obtain the bandwidth for the conditional mean (regression) function, we search for the bandwidth for conditional variance. To simplify our simulations, we use the method proposed by Chiou and Müller (1999) to select the bandwidth for the conditional variance function. A "nonparametric" Pearson chi-square statistic, $R^2(y, u, \hat{\sigma})$, is defined as

$$R^2(y, \mu, \hat{\sigma}) = \sum_{i=1}^{n} \frac{(Y_i - \mu(X_i))^2}{\hat{\sigma}_s^2(X_i)}, \tag{5.2}$$

where $y=(Y_1, \ldots, Y_n)$. Chiou and Müller (1999) point out that the expected value of the Pearson chi-square statistic, ER^2, is approximately equal to the degrees of freedom n. Since R^2 depends on the estimated (conditional) variance, this equality can be utilized for the selection of the bandwidth. Assume that $\hat{\mu}_h(\cdot)$ is the estimated value of $\mu(\cdot)$. Let

$$G_p(b, \hat{\mu}_h, \hat{\sigma}_b) = |R^2(y, \hat{\mu}_h, \hat{\sigma}_b) - n|, \tag{5.3}$$

where $\hat{\sigma}_b^2(\cdot)$ is the estimator $\hat{\sigma}_s^2(\cdot)$ with bandwidth b. Hence, the selected bandwidth $\hat{b}_{\text{opt}}$ is

$$\hat{b}_{\text{opt}} = \arg\min_b G_p(b, \hat{\mu}_h, \hat{\sigma}_b),$$

where h is the optimal bandwidth in (5.1).

Alternatively, consider a modified Pearson chi-statistic $D(y, \hat{\mu}, \hat{\sigma})$,

$$D(y, \hat{\mu}_h, \hat{\sigma}_b) = \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_h(X_i))^2}{\hat{\sigma}_b^2(X_i)}. \tag{5.4}$$

This quantity can be used to simultaneously select the bandwidths h and b for the estimators of the conditional mean and the conditional variance functions. Define

$$G_D(h, b, \hat{\mu}, \hat{\sigma}^2) = |D(y, \hat{\mu}, \hat{\sigma}) - n|. \tag{5.5}$$

The data-based bandwidths for h and b are then chosen as

$$(\hat{h}, \hat{b}) = \arg\min_{h,b} G_D(h, b, \hat{\mu}, \hat{\sigma}^2). \tag{5.6}$$

5.2. Simulations and empirical results

In this section, we perform simulations to verify some theoretical properties of the change point estimators in volatility. Also, we consider the effect of estimation when the unknown regression and conditional variance functions are replaced by the corresponding estimators under both i.i.d. errors and dependent errors (i.e., AR(1) process). In our simulations, the standardized Epanechnikov kernel, $K(x) = 3(1 - x^2/5)I(x^2 \leqslant 5)/(4\sqrt{5})$, is used to estimate regression and volatility functions, and the Gaussian kernel, $K(x) = 1/\sqrt{2\pi}\mathrm{e}^{-x^2/2}$ for $x \in (-\infty, \infty)$, is used to select the bandwidths for estimating the conditional mean and conditional variance functions.

Example 5.1. Consider the following model:

$$Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \tag{5.7}$$

where $\{\varepsilon_i\}$ is a sequence of independent and identically distributed random variables with the standard normal distribution. The volatility function has one change in the observed interval and the regression function is known (i.e., $\mu(x) = 0$). Here, we consider the following volatility model:

$$\sigma(X_i) = \begin{cases} 0.1\exp\{2(X_i - 0.5)^2 + 1\} & \text{if } i \leqslant k_0, \\ (0.1 + \tau)\exp\{2(X_i - 0.5)^2 + 1\} & \text{if } i > k_0. \end{cases} \tag{5.8}$$

In model (5.7), we assume that $\mu(x) = 0$ (equivalent to the assumption that $\mu(x)$ is known). We generate 500 series for three lengths ($T = 100, 200$ and 400) from this model, where $\{X_i\}$ is from the uniform distribution on $[0, 1]$. Estimators of change points are calculated for $\tau = 0.0, 0.1, 0.2, 0.5$ at different locations $\theta_0 = 0.3, 0.5, 0.7$ and 0.85, where there is a change in volatility, and $\theta_0 = k_0/T$. Table 1 summarizes the results for test statistic $\tilde{k}(v)$, when $v = 0$, $v = 1/2$ and $v = 1/4$, respectively.

Our simulations have led to some interesting findings. The difference, $\tilde{\theta}(v) - \theta_0$ (i.e., the bias of estimation), decreases rapidly as τ increases. The estimator $\tilde{k}(0)$ works better than both $\tilde{k}(1/2)$ and $\tilde{k}(1/4)$ in a small sample (i.e., here the size of sample is less than 200), since the bias and the variance of the estimator $\tilde{k}^* = \tilde{k}(0)$ are less than those of $\tilde{k}(1/2)$ and $\tilde{k}(1/4)$. Meanwhile, as the size of the sample increases, the bias and the variance of the estimates decrease. On the other hand, the estimator $\tilde{k}(v)$ for $0 \leqslant v \leqslant 1/2$ seems to be consistent since the bias becomes smaller when the sample size increases, just as the asymptotic results predict. The asymptotic sample distributions of $\tilde{k}(v)$ for $0 \leqslant v \leqslant 1/2$ are skewed. In larger samples, the estimator $\tilde{k}(1/2)$ seems to work better than the others as we expected. This is because $\tilde{k}(1/2)$ is the maximum likelihood estimator of k_0 if the errors are independently and normally distributed. This implies that factor $(k/T(1-k/T))^{1/2-v}$ has an effect on the estimator $\tilde{k}(v)$.

Example 5.2. We consider model (5.7) with unknown regression and volatility functions. The error term is the same as in (5.7). Assume that the regression function is $\mu(x) = 1 + 2x^2$, and that the volatility is

$$\sigma(X_i) = \begin{cases} 0.1\exp\{0.2X_i^2\} & \text{if } i \leqslant k_0, \\ (0.1+\tau)\exp\{0.2X_i^2\} & \text{if } i > k_0. \end{cases} \tag{5.9}$$

The volatility has one change point k_0 that takes four different values, i.e. $k_0 = 0.3T, 0.5T, 0.7T$ and $0.85T$. Let $\theta_0 = k_0/T$. τ also takes four different values, i.e., $\tau = 0.0, 0.1, 0.2$ and 0.5.

This example is designed to examine the effect of using the nonparametrically estimated regression function on the estimates of change points.

Table 2 shows almost the same parameter estimates and standard errors as those in Table 1, although the regression function has been estimated. That is, even if the conditional mean $\mu(\cdot)$ is unknown, there is little effect on the significance level and power of the tests of $\tilde{D}_k$ (i.e., $\tilde{V}_k^v$, $v = 0$) and $\tilde{V}_k$ (i.e., $\tilde{V}_k^v$, $v = 1/2$). The simulation result on $\tilde{V}_k^{1/4}$ is omitted to save space. In sum, the estimates of change points are reasonably well attained for the modest values of τ and T and seem to depend on the locations of change points.

Tables 3 and 4 show the simulation results on the empirical sizes and powers of $\tilde{V}_k$ and $\tilde{D}_k$ where the regression and volatility function are estimated from model (5.10). Tables 5 and 6 contain the values of test statistics, $\tilde{V}_k$ and $\tilde{D}_k$, calculated directly as if the regression function and volatility were known from this model. Again, there are little differences between Tables 3 and 5 and between Tables 4 and 6, respectively.

Similarly, we can take the error term as a *dependent* time series, in particular, the AR(1) process. Here, we show the results only when $\{\varepsilon_i\}$ is a sequence of i.i.d. random variables with normal distribution. When $\{\varepsilon_i\}$ is from the AR(1) process, the results are very similar to those with $\{\varepsilon_i\}$ i.i.d. random variables. Thus, we omit the details.

Example 5.3. In order to evaluate the influence of the scatter coefficient (ϕ) on volatility, we study the following model, where the coefficient ϕ takes three different values. This model, for simplicity, assumes that the regression function is known. We consider a case in which there exist change points in three different places, such as $\theta_0 = 0.3, 0.5$ and 0.85. This model is

$$Y_i = \mu(X_i) + \phi\sigma(X_i)\varepsilon_i,$$

where $\mu(X_i) = 0$ and

$$\sigma(X_i) = \begin{cases} 1 + 2 * X_i^2 + \sin^2(2\pi X_i) & \text{if } i \leqslant k_0, \\ (1+\tau)(1 + 2 * X_i^2 + \sin^2(2\pi X_i)) & \text{if } i > k_0 \end{cases} \tag{5.10}$$

Table 2
The estimation of jump points for model (5.7) with different sample sizes T and estimators $\tilde{k}^*$ and $\tilde{k}(1/2)$ $\tilde{k}(1/4)$ when $\mu(x)$ and $\sigma_0(x)$ unknown

T	τ	θ_0			
		0.3	0.5	0.7	0.85
Panel A: Estimates and standard error for $\tilde{k}(0)$					
100	0.1	0.41(0.120)	0.55(0.059)	0.71(0.048)	0.81(0.093)
	0.2	0.39(0.103)	0.53(0.048)	0.71(0.026)	0.84(0.041)
	0.5	0.37(0.091)	0.53(0.042)	0.71(0.019)	0.85(0.024)
200	0.1	0.36(0.076)	0.52(0.035)	0.71(0.019)	0.83(0.061)
	0.2	0.35(0.029)	0.52(0.029)	0.71(0.014)	0.85(0.018)
	0.5	0.34(0.052)	0.51(0.021)	0.70(0.010)	0.85(0.006)
400	0.1	0.33(0.055)	0.51(0.019)	0.70(0.014)	0.84(0.030)
	0.2	0.32(0.035)	0.51(0.014)	0.70(0.076)	0.85(0.008)
	0.5	0.32(0.028)	0.50(0.012)	0.70(0.005)	0.85(0.003)
Panel B: Estimates and standard error for $\tilde{k}(1/2)$					
100	0.1	0.44(0.178)	0.58(0.116)	0.75(0.066)	0.86(0.058)
	0.2	0.39(0.149)	0.56(0.093)	0.74(0.058)	0.86(0.023)
	0.5	0.38(0.131)	0.55(0.080)	0.74(0.055)	0.86(0.025)
200	0.1	0.36(0.107)	0.53(0.065)	0.73(0.045)	0.87(0.033)
	0.2	0.34(0.068)	0.53(0.049)	0.73(0.040)	0.86(0.025)
	0.5	0.32(0.054)	0.52(0.043)	0.72(0.032)	0.85(0.023)
400	0.1	0.32(0.056)	0.52(0.041)	0.72(0.035)	0.86(0.027)
	0.2	0.31(0.034)	0.51(0.027)	0.71(0.019)	0.86(0.021)
	0.5	0.30(0.016)	0.50(0.014)	0.70(0.016)	0.85(0.018)
Panel C: Estimates and standard error for $\tilde{k}(1/4)$					
100	0.1	0.42(0.144)	0.56(0.089)	0.73(0.075)	0.82(0.117)
	0.2	0.39(0.117)	0.54(0.066)	0.73(0.039)	0.86(0.048)
	0.5	0.37(0.105)	0.54(0.056)	0.72(0.032)	0.86(0.017)
200	0.1	0.36(0.078)	0.53(0.055)	0.71(0.036)	0.86(0.017)
	0.2	0.34(0.057)	0.52(0.037)	0.71(0.022)	0.85(0.017)
	0.5	0.33(0.047)	0.52(0.031)	0.71(0.017)	0.85(0.013)
400	0.1	0.32(0.026)	0.52(0.031)	0.71(0.015)	0.86(0.027)
	0.2	0.31(0.014)	0.51(0.017)	0.71(0.013)	0.85(0.020)
	0.5	0.30(0.011)	0.50(0.009)	0.70(0.011)	0.85(0.011)

and $\{\varepsilon_i\}$ is a sequence of i.i.d. normal random variables or AR(1) process:

$$\varepsilon_i = \beta\varepsilon_{i-1} + e_i,$$

in which $\{e_i\}$ again is a sequence of i.i.d. normal random variables. In our simulation, we choose $k_0 = [\theta_0 T]$ in which T is the size of sample (here, $T = 200$), and we take

Table 3
Empirical sizes and powers of statistic $\tilde{V}_k$ to test H_0 versus H_1 in model (5.10) with different sample sizes T, when $\mu(x)$ and $\sigma_0(x)$ are unknown under different αs

T	τ	$k_0 = 0.3T$		$k_0 = 0.5T$		$k_0 = 0.7T$		$k_0 = 0.85T$	
		$\alpha = 5\%$	10%	$\alpha = 5\%$	10%	$\alpha = 5\%$	10%	$\alpha = 5\%$	10%
100	0	2.6%	11.2%	2.6%	11.2%	2.6%	11.2%	2.6%	11.2%
	0.1	32.4%	62.4%	72.4%	91.0%	64.6%	87.6%	38.4%	66.4%
	0.2	43.2%	81.4%	88.4%	97.4%	80.6%	95.8%	57.8%	85.6%
	0.5	54.0%	89.4%	92.2%	99.6%	86.2%	98.2%	65.0%	92.0%
200	0	3.8%	8.4%	3.8%	8.4%	3.8%	8.4%	3.8%	8.4%
	0.1	81.2%	97.4%	98.8%	99.6%	97.2%	99.8%	59.4%	90.0%
	0.2	95.0%	99.2%	99.6%	100%	99.6%	100%	79.0%	98.0%
	0.5	97.2%	99.6%	99.8%	100%	99.6%	100%	86.4%	99.4%
400	0	5.6%	11%	5.6%	11%	5.6%	11%	5.6%	11%
	0.1	99.8%	100%	100%	100%	100%	100%	95.8%	100%
	0.2	100%	100%	100%	100%	100%	100%	99.6%	100%
	0.5	100%	100%	100%	100%	100%	100%	100%	100%

Table 4
Empirical sizes and powers of statistic $\tilde{D}_k$ to test H_0 versus H_1 in model (5.10) with different sample sizes T, when $\mu(x)$ and $\sigma(x)$ are unknown under different αs

T	τ	$k_0 = 0.3T$		$k_0 = 0.5T$		$k_0 = 0.7T$		$k_0 = 0.85T$	
		$\alpha = 5\%$	10%	$\alpha = 5\%$	10%	$\alpha = 5\%$	10%	$\alpha = 5\%$	10%
100	0	6.4%	10.8%	6.4%	10.8%	6.4%	10.8%	6.4%	10.8%
	0.1	7.4%	20.2%	15.8%	42.0%	44.2%	71.4%	74.2%	85.2%
	0.2	7.7%	26.8%	17.8%	56.6%	48.4%	85.8%	92.4%	97.4%
	0.5	10.2%	34.2%	19.0%	64.0%	63.0%	90.0%	97.4%	99.8%
200	0	7.8%	8.6%	7.8%	8.6%	7.8%	8.6%	7.8%	8.6%
	0.1	38.8%	49.0%	72.2%	82.8%	89.8%	94.6%	96.0%	97.0%
	0.2	61.6%	74.6%	90.8%	95.4%	97.8%	99.0%	100%	100%
	0.5	76.0%	86.0%	95.2%	98.0%	97.8%	99.0%	100%	100%
400	0	6.4%	10.2%	6.4%	10.2%	6.4%	10.2%	6.4%	10.2%
	0.1	92.4%	99.2%	99.0%	100%	99.8%	100%	100%	100%
	0.2	99.2%	99.8%	100%	100%	100%	100%	100%	100%
	0.5	99.8%	100%	100%	100%	100%	100%	100%	100%

different values of τ and ϕ, i.e., $\tau = 0.1, 0.5, 0.85$ and $\phi = 0.1, 0.5, 1$. The results in Table 7 are obtained when $\mu(x) = 0$. We also performed the same simulation when $\mu(x) = 1 - 2x^2$ but omitted the results because they are similar to those in Table 7. The empirical powers and sizes of the tests for the known $\mu(x)$ (which are omitted here)

Table 5
Empirical sizes and powers of statistic $\tilde{V}_k$ to test H_0 versus H_1 in model (5.10) with different sample sizes T when $\mu(x)$ and $\sigma(x)$ are known

T	τ	$k_0 = 0.3T$		$k_0 = 0.5T$		$k_0 = 0.7T$		$k_0 = 0.85T$	
		$\alpha = 5\%$	10%	$\alpha = 5\%$	10%	$\alpha = 5\%$	10%	$\alpha = 5\%$	10%
100	0	3.0%	12.0%	3.0%	12.0%	3.0%	12.0%	3.0%	12.0%
	0.1	29.4%	61.8%	75.2%	92.6%	71.6%	92.6%	45.2%	73.8%
	0.2	40.2%	80.6%	86.8%	98.2%	82.4%	97.0%	57.4%	88.0%
	0.5	48.%	87.6%	90.8%	99.2%	85.8%	97.6%	65.4%	90.6%
200	0	3.8%	10.2%	3.8%	10.2%	3.8%	10.2%	3.8%	10.2%
	0.1	81.0%	97.6%	99.2%	100%	98.6%	100%	59.6%	89.2%
	0.2	94.4%	99.2%	99.8%	100%	99.4%	99.8%	79.6%	98.8%
	0.5	97.8%	99.8%	100%	100%	99.8%	100%	85.6%	99.4%
400	0	4.8%	12.4%	4.8%	12.4%	4.8%	12.4%	4.8%	12.4%
	0.1	99.6%	99.8%	100%	100%	100%	100%	95.8%	99.6%
	0.2	99.8%	99.8%	100%	100%	100%	100%	99.6%	100%
	0.5	100%	100%	100%	100%	100%	100%	100%	100%

Table 6
Empirical sizes and powers of statistic $\tilde{D}_k$ to test H_0 versus H_1 in model (5.10) with different sample sizes T when $\mu(x)$ and $\sigma(x)$ are known

T	τ	$k_0 = 0.3T$		$k_0 = 0.5T$		$k_0 = 0.7T$		$k_0 = 0.85T$	
		$\alpha = 5\%$	10%	$\alpha = 5\%$	10%	$\alpha = 5\%$	10%	$\alpha = 5\%$	10%
100	0	6.8%	11.6%	6.8%	11.6%	6.8%	11.6%	6.8%	11.6%
	0.1	5.4%	18.2%	16.2%	46.2%	48.8%	76.0%	81.4%	90.0%
	0.2	5.5%	25.8%	19.8%	57.6%	58.4%	86.2%	94.6%	99.2%
	0.5	5.8%	32.4%	20.8%	64.6%	62.2%	88.8%	97.2%	99.8%
200	0	7.8%	8.6%	7.8%	8.6%	7.8%	8.6%	7.8%	8.6%
	0.1	40.4%	48.8%	78.0%	84.2%	94.2%	97.0%	95.6%	96.6%
	0.2	59.2%	71.0%	90.0%	94.6%	98.0%	99.5%	99.8%	100%
	0.5	75.0%	84.4%	95.6%	97.6%	98.8%	99.6%	100%	100%
400	0	4.8%	12.4%	4.8%	12.4%	4.8%	12.4%	4.8%	12.4%
	0.1	91.4%	98.2%	99.6%	100%	99.8%	100%	100%	100%
	0.2	98.8%	99.8%	100%	100%	100%	100%	100%	100%
	0.5	99.8%	100%	100%	100%	100%	100%	100%	100%

are very similar to those shown in Tables 3 and 4. Hence, we only give the results of the estimators of change points.

It is easy to find from Table 7 that the scale has no impact on the estimates of change points. Even when the unknown regression is replaced by its consistent estimator, the estimators have the same properties, as expected.

Table 7
The estimation for jump points for model (5.11) with different statistics $\tilde{V}_T^v$ $(v=0,1/2,1/4)$ against different scatter coefficients ϕ at $0.1, 0.5, 1$ when $\mu(x)=0$

θ_0	ϕ	v	Jump = 0.1				Jump = 0.2			
			i.i.d error		AR(1) error		i.i.d error		AR(1) error	
			$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.	$\hat{\theta}$	s.e.
0.3	0.1	0	0.364	0.078	0.385	0.095	0.349	0.062	0.369	0.083
	0.1	0.5	0.355	0.109	0.386	0.143	0.336	0.078	0.363	0.012
	0.1	0.25	0.359	0.093	0.385	0.117	0.337	0.054	0.360	0.093
0.3	0.5	0	0.364	0.078	0.385	0.095	0.349	0.062	0.369	0.083
	0.5	0.5	0.355	0.109	0.386	0.143	0.336	0.078	0.363	0.012
	0.5	0.25	0.359	0.093	0.385	0.117	0.337	0.054	0.360	0.093
0.3	1	0	0.364	0.078	0.385	0.095	0.349	0.062	0.369	0.083
	1	0.5	0.355	0.109	0.386	0.143	0.336	0.078	0.363	0.117
	1	0.25	0.359	0.093	0.385	0.117	0.337	0.054	0.360	0.093
0.5	0.1	0	0.524	0.039	0.530	0.045	0.518	0.029	0.523	0.032
	0.1	0.5	0.541	0.077	0.556	0.097	0.526	0.049	0.543	0.078
	0.1	0.25	0.527	0.046	0.540	0.072	0.520	0.033	0.530	0.048
0.5	0.5	0	0.524	0.039	0.530	0.045	0.518	0.029	0.523	0.032
	0.5	0.5	0.541	0.077	0.556	0.097	0.526	0.049	0.543	0.078
	0.5	0.25	0.527	0.046	0.540	0.072	0.520	0.033	0.530	0.048
0.5	1	0	0.524	0.039	0.530	0.045	0.518	0.029	0.523	0.032
	1	0.5	0.541	0.077	0.556	0.097	0.526	0.049	0.543	0.078
	1	0.25	0.527	0.046	0.540	0.072	0.520	0.033	0530	0.048
0.85	0.1	0	0.823	0.063	0.807	0.095	0.847	0.021	0.844	0.040
	0.1	0.5	0.869	0.033	0.864	0.077	0.868	0.024	0.870	0.026
	0.1	0.25	0.853	0.040	0.844	0.072	0.858	0.016	0.859	0.024
0.85	0.5	0	0.823	0.063	0.807	0.095	0.847	0.021	0.844	0.040
	0.5	0.5	0.869	0.033	0.864	0.077	0.868	0.024	0.870	0.026
	0.5	0.25	0.853	0.040	0.844	0.072	0.858	0.016	0.859	0.024
0.85	1	0	0.823	0.063	0.807	0.095	0.847	0.021	0.844	0.040
	1	0.5	0.869	0.033	0.864	0.077	0.868	0.024	0.870	0.026
	1	0.25	0.853	0.040	0.844	0.072	0.858	0.016	0.859	0.024

Example 5.4. Finally, we apply our approach to detect and estimate change points in the volatility of the Hong Kong Hang Sang Index (HSI). It is known that the stock prices in a bull market exhibit different behavior from those in a bear market. Apparently, there exists a nonlinearity for the prices in the bull and bear markets. Thus, we have transformed the original data by first-differencing the logarithm of the data so that the transformed data become approximately stationary; the first difference of the

logarithmic transformation of the data is written as $\{Y_t, t=1,2,\ldots,\}$, and the logarithm of the index values is written as $\{X_t, t=1,2,\ldots,\}$. We describe such nonlinearity with the following model with conditional heteroscedastic variance:

$$Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t, \tag{5.11}$$

where Y_t denotes the first-order difference of the logarithm of the price. No trend in Y_t is discernible, and the sample autocorrelation function is not significantly different from the Kronecker delta function (not reported here). Hence, we can assume that $\mu(x) = 0$. The data of the HSI are chosen from January 2, 1986 to December 31, 1991. The kernel function of Example 5.2 has been chosen for estimation. We find two significant change points in volatility for the HSI; October 14, 1987 (P-value 0.03) and June 5, 1989 (P-value 0.01), which were found also by Wong et al. (2001). The 1987 crash of the stock market was a "global event" which caused a significant jump in the volatility of the Hong Kong stock market. The Tiananmen event that occurred in June 1989 was only a "local event" and also led to a big jump in the volatility of the Hong Kong stock market.

6. Conclusions

In this paper, we have proposed a procedure to estimate the change points of volatility in nonparametric regression models. This estimation method is a hybrid of the LS procedure and nonparametric smoothers. The nonparametric model plays a key role in finance, and nonparametric estimators are very powerful in distinguishing among many models, for example, among short-rate models and derivative pricing models. The properties of the LS method are exploited to both identify and estimate the change points in volatility when the regression and the volatility functions are estimated by a nonparametric method. We have derived the asymptotic distributions of the estimators of change points and demonstrated their consistency when the unknown regression and volatility functions are replaced by their corresponding estimates. This implies that the asymptotic distribution and the consistency of the proposed estimators are not affected even if the regression function and volatility are unknown.

The general LS approach could potentially be applied to a wide spectrum of processes in economics and finance; in particular, when the observations are a sequence of a *dependent* time series. These proposed procedures, dealing with the dependence sequence, have extended most existing procedures, providing many important empirical applications. That is, the assumption of *independence* is not acceptable in many economic and financial models, including adaptive expectations, stock adjustment and price adjustment.

Other relevant applications can be further considered. Our method can easily be extended to multi-dimensional stochastic models and multi-factor models of the term structure, such as the multi-dimensional nonparametric model, which is defined by

$$Y_t = \mu(X_t, X_{t-1}, \ldots, X_{t-p}) + \sigma(X_t, X_{t-1}, \ldots, X_{t-p})\varepsilon_t,$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are multiple variable (p) functions, which are estimated by multivariate kernel methods.

Acknowledgements

We are grateful to Peter Robinson (the editor), Cheng Hsiao (the co-editor), and three anonymous referees for their detailed comments and suggestions which improved our paper greatly. We assume full responsibility for any remaining errors.

Appendix

In all proofs of our propositions, we only show the derivation in cases where $\{\varepsilon_i\}$ is a strong mixing sequence. When $\{\varepsilon_t\}$ is a martingale difference sequence, the proofs are similar to those of the mixing sequence (and thus, the proofs are omitted. Also refer to Bai (1997)). Complete proofs are available upon request. Let $x \in [a_1, a_2]$. To derive the asymptotic properties of our statistics, we need some lemmas, followed by their proofs. Without loss of generality, we assume $X_i \in [a_1, a_2]$ with the density function $f(x)$, where a_1 and a_2 are some constants. If $X_i \notin [a_1, a_2]$, we take a transformation for X_i by $\arctan(X_i)$ and replace the original X_i by $\arctan(X_i)$, which does not have any impact on our proofs below. Obviously, random variable $\arctan(X_i)$ is in $[-\pi/2, \pi/2]$. The following arguments can be extended to a random array $\{X_{ni}, Y_{ni}, n=1,2,\ldots, i=1,2,\ldots\}$ based on Theorem A.1 and the proofs in Masry (1996) and Kim and Cox (1995).

The following assumptions (AS.1)–(AS.8), need to be satisfied to establish the consistency and asymptotic distributions of estimators of change points in volatility when the regression and conditional variance functions are estimated in basic nonparametric models.

Assumptions:

(AS.1) The random sequence $\{(X_i, \varepsilon_i)\}$ satisfies one of the following two alternative conditions:

(a) Let $\mathscr{F}_t = \{X_1, X_2, \ldots, X_t, \varepsilon_1, \ldots, \varepsilon_{t-1}\}$. Assume that $\{\varepsilon_t\}$ is a martingale difference sequence with respect to $\{\mathscr{F}_t\}$ and $\sup_t \mathrm{E}|\varepsilon_t|^{4+\delta} < \infty$ for some $\delta > 0$.

(b) $\{\varepsilon_i\}$ is a strictly stationary and strong mixing sequence with mixing coefficients satisfying

$$\sum_{n=1}^{\infty} (\alpha(n))^{\delta/(2+\delta)} < \infty \tag{A.1}$$

and $\mathrm{E}|\varepsilon|^{4+\delta} < \infty$.

(AS.2) $\mu(x)$ and $\sigma_0(x)$ are continuous with third-order derivatives on $[a_1, a_2]$.

(AS.3) $f(x)$ is a bounded function with $M < f(x) < M'$ for some positive M and M', and has continuous second-order derivatives on (a_1, a_2).

(AS.4) The conditional density functions $f_{X_1|Y_1}(x \mid y)$ and $f_{(X_1,X_l)\,|\,(Y_1,Y_l)}(x_1, x_l|(y_1, y_l))$ are bounded for all $l > 0$.

(AS.5) Let $\mathrm{E}|X|^l < \infty$ and $\mathrm{E}|Y|^l < \infty$ for some large $l > 0$.

(AS.6) $K(\cdot)$ is a symmetric probability density function supported on the interval $[-c_0, c_0]$ with a bounded derivative, and the Fourier transform of $K(\cdot)$ is absolutely integrable.

(AS.7) For $T = 1, 2, \cdots$, and for any in interval $h = h_n \in \mathscr{H}_n$,

$$c_1 n^{-1/5} \leqslant h_n \leqslant c_2 n^{-1/5}$$

for some positive constants c_1 and c_2, and $b \in \mathscr{H}_n$. b satisfies that $b/h \to c_3$ as $T \to \infty$ for some constant $0 < c_3 < \infty$.

(AS.8) $\{(X_i, \varepsilon_i), i = 1, 2, \ldots, \}$ is a strictly stationary and strongly mixing sequence with coefficient $\alpha(n) = \mathcal{O}(c^n)$ for some $0 < c < 1$.

These assumptions are satisfied by most time series models. (AS.1) guarantees that $S_T/T \to \sigma^2$ almost surely and $Var(S_T)/T \to \sigma_w^2$, where σ^2 and σ_w^2 are two positive constants (see Rio, 1995). The weak invariance principles for the sum of the underlying errors can be used. The ordinary assumption of the errors is a sequence of independent and identically distributed random variables, but we add (AS.2) to allow *dependent* sequences of random variables, such as moving average processes. Further, (AS.2) is the essential condition for smoothness of the conditional mean and the conditional variance, which is required in most nonparametric regression models with conditional heteroscedastic variance. The rate of α-mixing in (AS.8) is assumed for simplicity, but in fact the mixing coefficient can be weakened to $\alpha(k) = \mathrm{O}(k^{-\tau})$ for some large $\tau > 0$ (Kim and Cox, 1995). The strong mixing case as in (AS.8) is considered to extend the results to the time series models. For a detailed discussion of these conditions, see Härdle and Tsybakov (1997), Fan and Yao (1998) and Masry (1996). Model (1.1) is set in a general frame, e.g., the data may be a sequence of *dependent* random variables. We suppose that $\{(X_i, \varepsilon_i)\}$ is a sequence of random variables, satisfying Assumption (AS.8), which includes the i.i.d observation case and other mixing cases such as β-mixing or ϕ-mixing sequences. It also includes many time series models.

Remark A.1. The strictly stationary assumption for the residual sequence $\{\varepsilon_i\}$ is used to guarantee the weak invariance principle of the sum of the sequence $\{\varepsilon_i\}$. This assumption can be disregarded if we assume that the weak invariance principle for the sequence $\{\varepsilon_i\}$ holds. The assumption that the sequence $\{\varepsilon_t\}$ is a mixing sequence with the mixing coefficient satisfying (A.1) is often used to derive the asymptotic distribution of statistics of interest. Some similar assumptions for mixing sequences can be found in many studies. See Bai (1994, 1997), Bai and Perron (1998), Chu et al. (1995) and Nunes et al. (1995), for examples.

Remark A.2. Assumption (AS.8) allows for the MA or AR process among the regressors. It is trivially satisfied whenever the regressors are nonstochastic. Obviously, Assumption (AS.8) can be derived from Assumptions (C.1) and (C.2) of Wooldridge and White (1988). Hence, Assumption (AS.8) is very convenient for deriving the asymptotic distribution under the null hypothesis by using the result of Wooldridge and White (1988).

Theorem A.1 (Wooldridge and White, 1988). *Let* $\{\varepsilon_{ni}, i = 1, 2, \ldots, n = 1, 2, \ldots\}$ *be a double array of real-valued random variables. Assume that with fixed* n, $\{\varepsilon_{ni}, i = 1, 2, \ldots, n = 1, 2, \ldots\}$ *is a stationary mixing sequence of random variables with* $\mathrm{E}\varepsilon_{ni} = \mu$ *and* $\mathrm{E}|\varepsilon_{ni}|^{2+\delta} < \infty$ *for some* $\delta > 0$ *and*

$$\sum_{n=1}^{\infty} (\alpha(n))^{\delta/(\delta+2)} < \infty \tag{A.2}$$

and for some $\sigma > 0$

$$Var(S_{nn})/n \to \sigma_w^2, \tag{A.3}$$

then

$$B_{nk} \Rightarrow B \text{ in } D[0,1], \quad as \quad n \to \infty, \tag{A.4}$$

where $k \to \infty$ *as* $n \to \infty$, $S_{nk} = \varepsilon_{n1} + \varepsilon_{n2} + \cdots + \varepsilon_{nk}$, $B_{nk}(t) = (S_{n[kt]} - [kt]\mu)/Var(S_{nk})$ $0 \leqslant t \leqslant 1$, *and* $\{B(t), 0 \leqslant t \leqslant 1\}$ *denotes the standard Wiener process.*

Lemma A.1. *Suppose that assumption* (AS.1) *is satisfied, and* $k_0 = [\theta_0 T]$, *for* $0 < \theta_0 < 1$. *Then for every* $\varepsilon > 0$ *and* $\delta > 0$, *there exists a constant* $T_0 > 0$ *such that when* $T > T_0$,

$$P(|k - k_0| > T\delta) < \varepsilon.$$

It is easy to prove this lemma using a similar argument of Proposition 2 in Bai (1994).

Lemma A.2. *Suppose that assumption* (AS.1) *is satisfied. Then there exists a constant* $0 < B < \infty$, *such that for every* $c > 0$ *and* $m > 0$,

$$P\left(\sup_{T \geqslant k \geqslant m} \frac{1}{m}\left|\sum_{i=1}^{m}(Z_i^2 - \mathrm{E}Z_i^2)\right| > c\right) \leqslant \frac{B}{c^2 m}.$$

This lemma can be derived using similar arguments in Lemma A.7 in Bai and Perron (1998) and the lemma of Mcleish (1975).

Proof of Proposition 2.1. For the sake of simplicity, let $\mathrm{E}\varepsilon_i^2 = 1$. Let $\xi_i = Z_i - 1$, so $\mathrm{E}\xi_i = 0$ under H_0. σ_w is some positive value, such that

$$Var(S_T)/T \to \sigma_w^2 \quad \text{as} \quad n \to \infty. \tag{A.5}$$

Write

$$X_n(t) = \frac{1}{\sigma_w \sqrt{T}} S_T(t) \quad \text{for} \quad 0 \leqslant t \leqslant 1,$$

where

$$S_T(t) = \begin{cases} \xi_1 + \xi_2 + \cdots + \xi_{[(T+1)t]} & \text{if } \ 0 \leqslant t < 1, \\ \xi_1 + \xi_2 + \cdots + \xi_T & \text{if } \ t \geqslant 1. \end{cases}$$

Obviously, $X_T(t)$ is the stochastic element of the probability space $D[0,1]$, and a right continuous function with a left limit. Without loss of generality, let $k = Tt, k = 1, 2 \ldots, T$. Then,

$$X_T(t) - tX_T(1) = \frac{1}{\sigma_w \sqrt{T}} S_T(t) - \frac{t}{\sigma_w \sqrt{T}} S_T(1)$$
$$= \frac{1}{\sigma_w \sqrt{T}} \left(\sum_{i=1}^{k} Z_i^2 - \frac{k}{T} \sum_{i=1}^{k} Z_i^2 \right)$$
$$= \frac{1}{\sigma_w \sqrt{T}} \sum_{i=1}^{k} Z_i^2 D_k = \frac{1}{\sigma_w \sqrt{T}} S_T D_k.$$

Hence it follows that

$$\sqrt{T} V_k^v = - \left(\frac{T}{k(T-k)} \right)^v \sigma_w (X_T(t) - tX_T(1)),$$
$$\sqrt{T} D_k = \frac{T\sigma_w}{S_T} (X_T(t) - tX_T(1)).$$

Thus, Proposition 2.1 follows from Theorem A.1 and (A.5) that

$$X_T(t) - tX_T(1) \to B(t) \quad \text{in} \quad D[0,1]. \qquad \square$$

Proof of Proposition 2.2. The proof of this proposition is based on the argument of Proposition 3 in Bai (1994) along with Lemmas A.1 and A.2. The details are omitted. □

Proof of Corollary 2.1. Write $\hat{\tau}_1^2(v,k) = S_k/k$ and $\hat{\tau}_2^2(v,k) = S_{T-k}/(T-k)$. Then $\hat{\tau}_1^2 = \hat{\tau}_1^2(v, \hat{k}(v)) = S_{\hat{k}(v)}/\hat{k}(v)$, where $S_k = \sum_{i=1}^k Z_i^2$ and $S_{T-k} = \sum_{i=k+1}^T Z_i^2$. When k_0 is known, the LS estimator of τ_1^2 is $\hat{\tau}_1^2(v, k_0)$. Thus, we obtain

$$\begin{aligned}
&T^{1/2}(\hat{\tau}_1^2(v,\hat{k}(v)) - \hat{\tau}_1^2(v,k_0)) \\
&\quad = I(\hat{k}(v) \leqslant k_0) \left(T^{1/2} \frac{k_0 - \hat{k}(v)}{k_0 \hat{k}(v)} \sum_{i=1}^{k_0} (Z_i^2 - \mathrm{E}Z_i^2) \right) \\
&\qquad + I(\hat{k}(v) > k_0) \left(T^{1/2} \frac{k_0 - \hat{k}(v)}{k_0 \hat{k}(v)} \sum_{i=1}^{k_0} (Z_i^2 - \mathrm{E}Z_i^2) - T^{1/2} \frac{1}{\hat{k}(v)} \sum_{i=k_0+1}^{\hat{k}(v)} (Z_i^2 - \mathrm{E}Z_i^2) \right. \\
&\qquad \left. + T^{1/2} \lambda_T \frac{\hat{k}(v) - k_0}{\hat{k}(v)} \right).
\end{aligned}$$

Similarly, we can obtain the expression of $T^{1/2}(\hat{\tau}_2^2(v,\hat{k}(v)) - \hat{\tau}_2^2(v,k_0))$. We can show that $T^{1/2}(\hat{\tau}_1^2(v,\hat{k}(v)), \hat{\tau}_2^2(v,\hat{k}(v))$ has the same distribution as $T^{1/2}(\hat{\tau}_1^2(v,k_0), \hat{\tau}_2^2(v,k_0))$ when T

is large enough, since $k_0 = [T\theta_0]$, $\hat{k}(v) = k_0 + O_P(\lambda_T^2)$ and $T\lambda_T^2 \to \infty$. The limit distribution of $T^{1/2}(\hat{\tau}_1^2(v, k_0), \hat{\tau}_2^2(v, k_0))$ is given by (2.12) in Corollary 2.1. Hence, we complete the proof of Corollary 2.1. □

Proof of Corollary 2.2. This corollary is based on the arguments of Proposition 2.2. □

Now, we give a result of the local stochastic oscillation moduli for $T(V_k^2 - V_{k_0}^2)$ for those $k's$ in the neighborhood of k_0 such that $k = [k_0 + t\lambda_T]$, where $\lambda_T = \tau_2^2 - \tau_1^2$ and t varies in an arbitrary bounded interval. Define

$$\Lambda_T^v(t) = T[(V^v([k_0 + t\lambda_T^{-2}]))^2 - (V^v(k_0))^2], \quad 0 \leqslant v \leqslant 1/2.$$

We can find the limiting process of $\Lambda_T(t)$ and $\Lambda_T^v(t)$ on $|t| \leqslant M$ for every given $M > 0$.

Lemma A.3. (a) *Suppose that assumption* (AS.1) *and condition* (J_1) *are satisfied. If* $\{\varepsilon_i\}$ *is a sequence of strictly stationary random variables, then*

$$T(V_k^2(v) - V_{k_0}^2(v)) \xrightarrow{d} W^*(k - k_0),$$

where $W^*(m)$ *has been defined in Section* 2.2.

(b) *Suppose that assumption* (AS.1) *and condition* (J_2) *are satisfied. Then for every* $M < \infty$ *and* $0 < v \leqslant 1/2$, $\Lambda_T^v(t)$ *converges weakly in* $C[-M, M]$ *to*

$$\Lambda(t) = 2\{\sigma_w B_s(t) - g_v(t)|t|\}.$$

Proof of Lemma A.3. The proof of part (a) of Lemma A.3 is similar to that of Propositions 1 and 2 in Bai (1997). Theorem A.1 is also utilized. The proof of part (b) is similar to that of Theorem 1 in Bai (1994). Lemma A.1 is also used. □

Proof of Proposition 2.3. We only prove part (b). Lemma A.3 implies that $\Lambda_T^v(t) = T(V^2_{[k_0 + t\lambda_T^{-2}]}(v) - V_{k_0}^2)$ converges weakly in $C[-M, M]$ to $\Lambda(t) = \{\sigma_w B_s(t) - g_v(t)|t|\}$. By the continuous mapping theorem and the consequence of Lemma A.3, $\lambda_T^2(\hat{k}(w) - k_0) \xrightarrow{d} \tilde{\eta}_v$, where $\tilde{\eta}_v = \arg\max_t \Lambda(t)$. Since $bB(t) \overset{d}{=} B(b^2 t)$ for every $b \in R$, a change in variable leads to

$$\arg\max_t \Lambda(t) \overset{d}{=} \sigma_w^2 \arg\max_t \{B_s(t) - g_v(t)|t|\},$$

where $a_n \overset{d}{=} b_n$ denotes the quantities, with a_n and b_n having the same distribution.

Hence,

$$\frac{\lambda_T^2(\hat{k}(v) - k_0)}{\sigma_w^2} \xrightarrow{d} \arg\max_t \{B_s(t) - g_v(t)|t|\}.$$

Thus, we complete the proof of Proposition 2.3. □

The following several lemmas (whose detailed proofs are omitted) are necessary to prove Proposition 3.1.

Lemma A.4. *Assume that assumptions* (AS.2)–(AS.8) *are satisfied, and* $\varphi_1(x)$ *and* $\varphi_2(x)$ *are continuous functions on* $[a_1, a_2]$. *Let* $k/T \to c$ *and* $n/T \to c^*$ *for two constants* $0 < c, c^* < 1$, *as* $T \to \infty$. *Then for any* $k \to \infty$, *we have*

$$\frac{1}{nk^{1/2}h}\sum_{i=1}^{k}\frac{\varphi_1(X_i)}{f(X_i)}\varepsilon_i\sum_{j=1}^{n}K_h(X_i - X_j)\varphi_2(X_j)\varepsilon_j = o_P(1). \tag{A.6}$$

Furthermore, assume that $\{(X_i, \varepsilon_i, \eta_i)\}$ *is a strictly stationary and mixing sequence with a mixing coefficient of* $\alpha(k)$, *satisfying assumption* (AS.8). *Let* $E(\varepsilon_i \mid X_i) = 0$, *and* $E(\eta_i|X_i) = 0$. $W(\cdot)$ *and* $K(\cdot)$ *are two kernel functions, while* h *and* b *are the two corresponding bandwidths. Then*

$$\frac{1}{nk^{1/2}hb}\sum_{i=1}^{k}\frac{\varphi_1(X_i)}{f(X_i)}\eta_i\sum_{j=1}^{n}K_h(X_i - X_j)$$
$$\times W_b(X_i - X_j)\varphi_2(X_j)\varepsilon_j = o_P(1). \tag{A.7}$$

Proof. We only prove (A.6) since the proof of (A.7) is similar. Note that we only need to consider the summation in (A.6) over the set of indices constrained by $|i-j| > N$ for some $N(< T)$. In fact, we can take $N = o(Th/(k^{1/2}\log T))$, such that $N \to \infty$ as $T \to \infty$, and

$$\frac{1}{nk^{1/2}h}\sum_{i=1}^{k}\sum_{j=1}^{n}I(|i-j| \leqslant N)K_h(X_i - X_j)\frac{\varphi_1(X_i)\varphi_2(X_j)}{f(X_i)}\varepsilon_i\varepsilon_j$$
$$\leqslant \frac{1}{nk^{1/2}h}\sum_{i=1}^{k}\left|\frac{\varphi_1(X_i)\varepsilon_i}{f(X_i)}\right|\left|\sum_{|j-i|\leqslant N}K_h(X_i - X_j)\varphi_2(X_j)\varepsilon_i\right|$$
$$= O_p\left(\left(\frac{Nk^{1/2}\log T}{Th}\right)^{1/2}\right)$$
$$= o_p(1),$$

where $I(A)$ denotes the indicator function of A. To prove (A.6) in probability, we show that

$$\frac{1}{k^{1/2}nh}\sum_{i=1}^{k}\sum_{j=1}^{n}I(|i-j| > N)K_h(X_i - X_j)\frac{\varphi_1(X_i)\varphi_2(X_j)}{f(X_i)}\varepsilon_i\varepsilon_j = o_p(1). \tag{A.8}$$

The proof is similar to the verification of (4.3) in Kim and Cox (1995) and arguments of Lemma 1 in Xia et al., (1998).

Assume that $\varphi_3(x)$ is a bounded and continuous function on $[a_1, a_2]$. By similar arguments for Lemma A.4 and Lemma 5.3 of Kim and Cox (1995), we can prove that

$$\frac{1}{n^2k^{1/2}bh}\sum_{l=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n}W_b(X_i-X_l)K_h(X_j-X_i)$$
$$\times\varphi_1(X_i)\varphi_2(X_j)\varphi_3(X_l)\eta_i\varepsilon_j=\mathrm{o_P}(1). \qquad \square \tag{A.9}$$

Lemma A.5. *Suppose that assumptions* (AS.2)–(AS.8) *are satisfied and that* $\varphi(x)$ *is a continuous function on* $[a_1,a_2]$. *Let* $k/T\to c$ *and* $n/T\to c^*$ *for two constants* $0<c,c^*<1$, *as* $T\to\infty$. *Then for* $T\to\infty$, *we have*

$$\frac{1}{nk^{1/2}h}\sum_{i=1}^{k}\varphi(X_i)\varepsilon_i\sum_{j=1}^{n}\{K_h(X_i-X_j)$$
$$\times(\mu(X_j)-\mu(X_i))-\eta(X_i,X_i)\}=\mathrm{o_P}(1), \tag{A.10}$$

where $\eta(X_i,X_i)=\int K_h(X_i-u)(\mu(u)-\mu(X_i))f(u)\,\mathrm{d}u$. *As* $T\to\infty$, *we have*

$$\frac{1}{nk^{1/2}h}\sum_{i=1}^{k}\varphi(X_i)\varepsilon_i\sum_{j=1}^{n}\{K_h(x-X_j)$$
$$\times(\mu(X_j)-\mu(X_i))-\eta(x,X_i)\}=\mathrm{o_P}(1), \tag{A.11}$$

uniformly for $x\in[a_1,a_2]$, *where* $\eta(x,X_i)=\int K_h(x-u)(\mu(u)-\mu(X_i))f(u)\,\mathrm{d}u$. *Furthermore, for the sums of two iterated kernels, we have the following results*:

$$\frac{1}{nk^{1/2}hb}\sum_{i=1}^{k}\varphi(X_i)\varepsilon_i\sum_{j=1}^{n}\{K_h(x-X_j)$$
$$\times(W_b(X_i-X_j)(\mu(X_j)-\mu(X_i))-\xi(x,X_i)\}=\mathrm{o_P}(1), \tag{A.12}$$

uniformly for $x\in[a_1,a_2]$, *where* $\xi(x,X_i)=\int K_h(x-u)W_b(X_i-u)(\mu(u)-\mu(X_i))f(u)\,\mathrm{d}u$ *and*

$$\frac{1}{nk^{1/2}hb}\sum_{i=1}^{k}\varphi(X_i)\varepsilon_i\sum_{j=1}^{n}\{K_h(x-X_j)$$
$$(W_b(X_i-X_j)(\mu(X_j)-\mu(x))-\zeta(x,X_i)\}=\mathrm{o_P}(1), \tag{A.13}$$

uniformly for $x\in[a_1,a_2]$, *where* $\zeta(x,X_i)=\int K_h(x-u)W_b(X_i-u)(\mu(u)-\mu(x))f(u)\,\mathrm{d}u$. *In addition, assume that* η_i *is a stationary sequence and independent of* $\{X_i,i=1,2,\dots,n\}$, *then*

$$\frac{1}{nk^{1/2}hb}\sum_{i=1}^{k}\varphi(X_i)\varepsilon_i\sum_{j=1}^{n}\{K_h(x-X_j)\eta_j$$
$$(W_b(X_i-X_j)(\mu(X_j)-\mu(x))-\zeta(x,X_i)\mathrm{E}\eta_1\}=\mathrm{o_P}(1),$$

uniform for $x\in[a_1,a_2]$, *where* $\zeta(x,X_i)=\int K_h(x-u)W_b(X_i-u)(\mu(u)-\mu(x))f(u)\,\mathrm{d}u$.

Proof. The proof is similar to that of Lemma A.4. Also see Lemma A.1 of Xia et al., (1998). □

Lemma A.6. *Suppose that assumptions of* (AS.2)–(AS.8) *are satisfied and that* $k/T \to c$ *for* $0 < c < 1$ *as* $T \to \infty$. *Then*,

$$\frac{1}{(nh)^m k^{1/2}} \sum_{i=1}^{k} \frac{1}{f^2(X_i)} \left\{ \left[\sum_{j=1}^{n} K_h(X_i - X_j)(\mu(X_j) - \mu(X_i)) \right]^m - \frac{1}{h^m} \int \frac{1}{f(x)} \left[\int K_h(x-t)(\mu(t) - \mu(x)) f(t)\,\mathrm{d}t \right]^m \mathrm{d}x \right\} \to 0 \tag{A.14}$$

and

$$\frac{1}{(nh)^m k^{1/2}} \sum_{i=1}^{k} \frac{1}{f^2(X_i)} \left(\sum_{j=1}^{n} K_h(X_i - X_j)\sigma(X_j)\varepsilon_j \right)^m \to 0 \tag{A.15}$$

in probability, where $m = 1$ *or* 2.

Proof. This lemma can be proved by arguments similar to those in (4.4) and (4.16) in Kim and Cox (1995) and to Theorem 5 in Masry (1996), to and using Lemmas A.4 and A.5. □

Lemma A.7. *Assume that assumptions* (AS.2)–(AS.8) *are satisfied. Then*,

$$U_1 = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} [\mu(X_i) - \hat{\mu}(X_i)]^2 \to 0,$$

in probability.

Proof.

$$\begin{aligned}
&\frac{1}{\sqrt{k}} \sum_{i=1}^{k} [\mu(X_i) - \hat{\mu}(X_i)]^2 \\
&\quad = \frac{1}{n^2 k^{1/2} h^2} \sum_{i=1}^{k} \frac{1}{f^2(X_i)} \left[\sum_{j=1}^{n} K_h(X_i - X_j)(\mu(X_j) - \mu(X_i)) \right]^2 \\
&\quad\quad + \frac{2}{n^2 k^{1/2} h^2} \sum_{i=1}^{k} \frac{1}{f^2(X_i)} \sum_{j=1}^{n} K_h(X_i - X_j)(\mu(X_j) - \mu(X_i)) \\
&\quad\quad \times \sum_{l=1}^{k} K_h(X_i - X_l)\sigma(X_l)\varepsilon_l
\end{aligned}$$

$$+\frac{1}{n^2k^{1/2}h^2}\sum_{i=1}^{k}\frac{1}{f^2(X_i)}\left(\sum_{j=1}^{k}K_h(X_i-X_j)\sigma(X_j)\varepsilon_j\right)^2+o_P(1)$$
$$=H_1+H_2+H_3.$$

By Lemma A.5, we obtain that $H_2 \to 0$ in probability. It follows from (A.15) that $H_3 \to 0$ in probability. Hence, we only need to prove that $H_1 \to 0$ in probability. In fact, from Lemma A.6, we need to prove that

$$\frac{k^{1/2}}{h^2}\int\frac{1}{f(x)}\left(\int K_h(x-t)(\mu(t)-\mu(x))f(t)\,\mathrm{d}t\right)^2\,\mathrm{d}x\to 0. \tag{A.16}$$

Under the assumptions of $\mu(\cdot)$ and $f(\cdot)$, we have

$$\frac{k^{1/2}}{h^2}\int\frac{1}{f(x)}\left(\int K_h(x-t)(\mu(t)-\mu(x))f(t)\,\mathrm{d}t\right)^2\Big/ f(x)\,\mathrm{d}x$$
$$=k^{1/2}h^4\left\{\int\frac{(\mu''(x)f(x)+\mu'(x)f'(x))^2}{f(x)}\left(\int t^2K(t)\,\mathrm{d}t\right)^2\,\mathrm{d}x+o(1)\right\}.$$

This implies that (A.16) holds. Therefore, we complete the proof of Lemma A.7. □

Lemma A.8. *Under assumptions* (AS.2)–(AS.8), *for any* $(k_0-k)/T\to c_1$ *for some* $0<c_1<1$ *as* $T\to\infty$, *we have*

$$\frac{1}{\sqrt{k_0-k}}\sum_{i=k+1}^{k_0}(\hat{W}_i^2-W_i^2)\to 0\quad\textit{in probability}. \tag{A.17}$$

For any $(k-k_0)/T\to c_2$ *for some* $0<c_2<1$ *as* $T\to\infty$ *we have*

$$\frac{1}{\sqrt{k-k_0}}\sum_{i=k_0+1}^{k}(\hat{W}_i^2-W_i^2)\to 0\quad\textit{in probability}, \tag{A.18}$$

where $\sum_{i=a}^{b}\Delta_i=0$ *when* $b<a$, *for any* $\Delta_i\neq 0$, *and for all* $k\leqslant k_0$, $k\to\infty$ *and* $k/T\to c$ *for some* $0<c<1$ *as* $T\to\infty$

$$\frac{1}{\sqrt{k}}\sum_{i=1}^{k}(\hat{W}_i^2-W_i^2)\to 0\quad\textit{in probability}. \tag{A.19}$$

Proof. We only provide a sketch of the proof of (A.19) here. We can show that (A.17) and (A.18) hold by similar arguments. Note that

$$\frac{1}{\sqrt{k}}\sum_{i=1}^{k}(\hat{W}_i^2-W_i^2)=2U_2+U_1,$$

where

$$U_1=\frac{1}{\sqrt{k}}\sum_{i=1}^{k}(\hat{W}_i-W_i)^2,\quad U_2=\frac{1}{\sqrt{k}}\sum_{i=1}^{k}W_i(\hat{W}_i-W_i).$$

By some simple algebra, we have

$$U_2 = \frac{1}{\sqrt{T}}\sum_{i=1}^{k} W_i(W_i - \hat{W}_i) = \frac{1}{\sqrt{k}}\sum_{i=1}^{k}(\hat{\mu}(X_i) - \mu(X_i))\sigma_0(X_i)\varepsilon_i.$$

Next, we show that

$$U_2 \to 0 \quad \text{in probability.} \tag{A.20}$$

Basically, we need to prove that some complicated sums of sequence $\{X_i, \varepsilon_i\}$ are asymptotically negligible in probability. The methods to prove these results are similar to those in Kim and Cox (1995) and Masry (1996) (in particular, the proof of Theorems 5 and 6). It can be shown that

$$\begin{aligned} U_2 &= \frac{1}{nk^{1/2}h}\sum_{i=1}^{k}\sum_{j=1}^{n} K_h(X_i - X_j)(Y_j - \mu(X_i))\sigma_0(X_i)\varepsilon_i / f(X_i) \\ &\quad + \frac{1}{nk^{1/2}h}\sum_{i=1}^{k}\sum_{j=1}^{n} K_h(X_i - X_j)(Y_j - \mu(X_i))\sigma_0(X_i)\varepsilon_i\left[\frac{1}{f_n(X_i)} - \frac{1}{f(X_i)}\right] \\ &= I_1 + I_2, \end{aligned} \tag{A.21}$$

where $\hat{f}_n(x) = 1/(nh)\sum_{i=1}^{n} K_h(X_i - x)$. Now, we utilize Lemmas A.4 and A.5 to prove that I_1 and I_2 are negligible in probability. At last, Lemma A.7 shows that $U_1 \to 0$ in probability.

Proof of Proposition 3.1. We only need to prove that the consequence of Proposition 2.2 holds when the regression and volatility functions are unknown, and other results can be obtained using similar arguments. It is sufficient to prove that

$$\sqrt{T}[\tilde{V}^{\nu}(k) - \tilde{V}^{\nu}(k_0) - (\bar{V}^{\nu}(k) - \bar{V}^{\nu}(k_0))] \xrightarrow{P} 0 \tag{A.22}$$

uniformly on $k \in [T\delta, (1-\delta)T]$ for some $0 < \delta < 1$. Write

$$\Delta_k = \frac{1}{T-k} R_{T-k} - \frac{1}{k} R_k.$$

Hence, the corresponding estimator of W_k is $\hat{W}_k$, and $a_k = (k/T(1-k/T))^{1-\nu}$. The left-hand side of Eq. (A.22) can then be rewritten as

$$\begin{aligned} &\sqrt{T}(a_k\hat{\Delta}_k - a_{k_0}\hat{\Delta}_{k_0} - a_k\Delta_k + a_{k_0}\Delta_{k_0}) \\ &= \sqrt{T}a_k(\hat{\Delta}_k - \Delta_k) - \sqrt{T}a_{k_0}(\hat{\Delta}_{k_0} - \Delta_{k_0}). \end{aligned} \tag{A.23}$$

The first term on the right-hand side of Eq. (A.23) is equal to

$$\begin{aligned} &\sqrt{T}a_k\left[\frac{1}{T-k}(\hat{R}_{T-k} - R_{T-k}) - \frac{1}{k}(\hat{R}_k - R_k)\right] \\ &= \sqrt{T}a_k\left[\frac{1}{T-k}\sum_{i=k+1}^{T}(\hat{W}_i^2 - W_i^2) - \frac{1}{k}\sum_{i=1}^{k}(\hat{W}_i^2 - W_i^2)\right] \end{aligned}$$

$$= \left(\frac{T}{T-k}\right)^{1/2} \frac{1}{(T-k)^{1/2}} \sum_{i=k+1}^{T} (\hat{W}_i^2 - W_i^2)$$

$$-a_k \left(\frac{T}{k}\right)^{1/2} \frac{1}{k^{1/2}} \sum_{i=1}^{k} (\hat{W}_i^2 - W_i^2). \tag{A.24}$$

We can prove from Lemma A.8 that the two terms on the right-hand side of Eq. (A.24) are negligible in the sense of probability convergence. Hence, (A.22) holds, and thus the proof of Proposition 3.1 is completed by Proposition 2.4.

References

Antoch, J., Hušková, M., 1995. Change-point problem and bootstrap. Journal of Nonparametric Statistics 5, 123–144.

Antoch, J., Hušková, M., Praskova, Z., 1996. Effect of dependence on statistics for determination of change. Journal of Statistical Planning and Inference 60, 291–310.

Bai, Z., 1994. Least squares estimation of a shift in linear process. Journal of Time Series Analysis 15, 453–472.

Bai, J., 1997. Estimation of a change point in multiple regression models. The Review of Economics and Statistics 79, 551–563.

Bai, J., 2000. Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices. Annals of Economics and Finance 1, 301–336.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

Bhattacharya, P.K., 1987. Maximum likelihood estimation of a change point in the distribution of independent random variables, general multiparameter case. Journal of Multivariate Analysis 23, 183–208.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–328.

Chiou, J.M., Müller, H.G., 1999. Nonparametric quasi-likelihood. The Annals of Statistics 27, 36–64.

Chu, C.J., Hornik, K., Kuan, C.M., 1995. The moving-estimates test for parameter stability. Econometric Theory 11, 699–720.

Csörgő, M., Horváth, L., 1997. Limit Theorems in Change-point Analysis. Wiley, New York.

Delgado, M.A., Hidalgo, J., 2000. Nonparametric inference on structural breaks. Journal of Econometrics 96, 113–144.

Drost, F., Nijman, T., Werker, B., 1998. Estimation and testing in models containing both jumps and conditional heteroscedasticity. Journal of Business and Economic Statistics 16, 237–243.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and its Applications. Chapman & Hall, London.

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in Stochastic regression. Biometrika 85, 645–660.

Härdle, W., Tsybakov, A., 1997. Local polynomial estimators of the volatility function in nonparametric autoregression. Journal of Econometrics 81, 223–242.

Hinkley, D., 1970. Inference about the change point in a sequence of random variables. Biometrika 57, 1–7.

Inclan, C., Tiao, G.C., 1994. Use of cumulative sums of squares for retrospective detection of change of variance. Journal of the American Statistical Association 89, 913–923.

Jorion, P., 1988. On jump processes in the foreign exchange and stock markets. Review of Financial Studies 1, 427–445.

Kim, T.Y., Cox, D.D., 1995. Bandwidth selection in kernel smoothing of time series. Journal of Time Series Analysis 17, 49–63.

Lamoreux, C.G., Lastrapes, W.D., 1990. Persistence in variance, structural change and the GARCH model. Journal of Business and Economic Statistics 48, 225–234.
Masry, E., 1996. Multivariate local polynomial regression for time series: uniform and strong consistency and rates. Journal of Time Series Analysis 17, 571–599.
Masry, E., Tjøstheim, D., 1995. Nonparametric estimation and identification of nonlinear ARCH times series: strong convergence and asymptotic normality. Econometric Theory 11, 258–289.
Mcleish, D.L., 1975. A maximal inequality and dependent strong laws. The Annals of Probability 3, 829–839.
Müller, H.G., 1992. Change-points in nonparametric regression analysis. The Annals of Statistics 24, 1667–1678.
Nelson, D.B., 1987. Conditional heteroscedasticity in asset returns: an approach. Mimeo, Massachusetts Institute of Technology.
Nunes, L.C., Kuan, C.M., Newbold, P., 1995. Spurious break. Econometric Theory 11, 736–749.
Pagan, A.R., Hong, Y.S., 1991. Nonparametric estimation and the risk premium. In: Barnett, W.A., Powell, A.J., Tauchen, G.E. (Eds.), Nonparametric and Semiparametric Methods in Economic Theory and Econometrics. Cambridge University Press, Cambridge, pp. 51–75.
Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267–290.
Pagan, A.R., Ullah, A., 1988. The econometric analysis of models with risk terms. Journal of Applied Econometrics 3, 87–105.
Peligrad, M., Shao, Q., 1995. Estimation of the variance of partial sums for ρ-mixing random variables. Journal of Multivariate Analysis 52, 140–157.
Perron, B., 2001. Jumps in the volatility of financial markets. Working Paper, University of Montreal.
Rio, E., 1995. The functional law of the iterated logarithm for stationary strongly mixing sequences. The Annals of Probability 23, 1188–1203.
Vlaar, P., Palm, F., 1993. The message in weekly exchange rates in the European monetary system: mean revision, conditional heteroscedasticity, and jumps. Journal of Business and Economic Statistics 11, 351–360.
Wong, H., Ip, W., Li, Y., 2001. Detection of jumps by wavelets in a heteroscedastic autoregressive model. Statistics and Probability Letters 52, 365–372.
Wooldridge, J.M., White, H., 1988. Some invariance principles and central limit theorems for dependent heterogeneous processes. Econometric Theory 4, 210–230.
Yao, Y.C., 1987. Approximating the distribution of the ML estimate of the change-point in a sequence of independent R.V.'s. The Annals of Statistics 3, 1321–1328.
Xia, Y., Tong, H., Li, W.K., 1998. Absolute deviation estimation of volatility in nonparametric models. Research Report No. 177, Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong.