



**HAL**  
open science

# Nonparametric simultaneous testing for structural breaks

Jiti Gao, Irène Gijbels, Sébastien van Bellegem

► **To cite this version:**

Jiti Gao, Irène Gijbels, Sébastien van Bellegem. Nonparametric simultaneous testing for structural breaks. *Econometrics*, 2008, 143 (1), pp.123. 10.1016/j.jeconom.2007.08.009 . hal-00501804

**HAL Id: hal-00501804**

**<https://hal.science/hal-00501804>**

Submitted on 12 Jul 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Author's Accepted Manuscript

Nonparametric simultaneous testing for structural breaks

Jiti Gao, Irène Gijbels, Sébastien Van Bellegem

PII: S0304-4076(07)00168-6  
DOI: doi:10.1016/j.jeconom.2007.08.009  
Reference: ECONOM 2978

To appear in: *Journal of Econometrics*

Cite this article as: Jiti Gao, Irène Gijbels and Sébastien Van Bellegem, Nonparametric simultaneous testing for structural breaks, *Journal of Econometrics* (2007), doi:10.1016/j.jeconom.2007.08.009

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/jeconom](http://www.elsevier.com/locate/jeconom)

# Nonparametric Simultaneous Testing for Structural Breaks

BY JITI GAO<sup>1</sup>, IRÈNE GIJBELS<sup>2</sup> AND SÉBASTIEN VAN BELLEGEM<sup>3</sup>

<sup>1</sup>*The University of Western Australia,* <sup>2</sup>*Katholieke Universiteit Leuven and*

<sup>3</sup>*Université catholique de Louvain*

## Abstract

In this paper we consider a regression model with errors that are martingale differences. This modeling includes the regression of both independent and time series data. The aim is to study the appearance of structural breaks in both the mean and the variance functions, assuming that such breaks may occur *simultaneously* in both the functions. We develop nonparametric testing procedures that simultaneously test for structural breaks in the conditional mean and the conditional variance. The asymptotic distribution of an adaptive test statistic is established, as well as its asymptotic consistency and efficiency. Simulations illustrate the performance of the adaptive testing procedure. An application to the analysis of financial time series also demonstrates the usefulness of the proposed adaptive test in practice.

*JEL Classification:* C12; C14; C32.

Keywords and phrases: conditional mean and variance function, nonparametric testing, structural break, threshold model, time series analysis.

## 1. Introduction and Motivation

Existing studies show that most nonparametric inference methods assume a certain degree of smoothness of the function to be investigated. When estimating a regression function based on an independent and identically distributed sample for example, it is often assumed that this unknown function is continuous, or even a few times differentiable. When dealing with discontinuous functions, the classical nonparametric estimation techniques need to be used with care in order to produce an estimate that shows a similar unsmooth behaviour. Related topics are change-point detection, edge detection and image

---

Corresponding author: Irène Gijbels, Department of Mathematics and University Center for Statistics, Katholieke Universiteit Leuven, Box 2400, Celestijnenlaan 200B, B-3001 Leuven (Heverlee), Belgium; E-mail: irene.gijbels@wis.kuleuven.be; Tel:+32-16-322018; Fax: +32-16-327998

reconstruction. See for example, Müller (1992), Carlstein, Müller and Siegmund (1994), Loader (1996), Qiu (1998), Spokoiny (1998), Müller and Stadtmüller (1999), Gonzalez and Woods (2002), Gijbels and Goderniaux (2004) and Qui (2005), among others.

When detecting or testing for change-points in a mean regression function, it is often assumed that the variance function is a continuous function. A closely related study in the literature of nonparametric testing of change-points is given in Grégoire and Hamrouni (2002), which assume that both the marginal and variance functions are continuous when testing for continuity for the mean function at a known location. This assumption however may not be justified, and in fact both the mean and the variance functions could have change-points. In a time series context, models with structural breaks (change-points) in the conditional mean and variance function have already been discussed. For example, Delgado and Hidalgo (2000) assume in their equation (1) and condition A4 that both the conditional mean and conditional variance functions may have the same structural breaks and then propose estimating both the locations and sizes of possible structural breaks in nonparametric regression with stationary time series observations.

In addition, when there is no information or knowledge about the smoothness of either function or both functions, we should consider testing whether the conditional mean and conditional variance functions both have discontinuities. By doing so, one may avoid assuming continuity mistakenly for one of the functions. Of course this also implies that one may assume discontinuities mistakenly for one of the functions. For this case, however, our experience in Chen and Gao (2005) shows that there would only be a slight reduction of the power for using a simultaneous test when a univariate test for either the conditional mean or the conditional variance should be used.

This motivates us to address the important issue of how to detect possible structural breaks in both the conditional mean and variance functions of nonparametric regression models for both independent and dependent data. Although there are numerous nonparametric tests available in the literature, to the best of our knowledge none of them could be directly applicable to such cases where both the mean and variance may have discontinuities. In order to clearly present our idea and methodology, we mainly consider the case where both the mean and the variance function may have structural breaks at a same number of unknown locations  $\tau_1, \dots, \tau_d$ , where  $d$  is the dimension of the covariate vector. As mentioned in the end of Section 2, however, the proposed theory and methodology in Section 2 is applicable to deal with other related cases where structural changes

of the mean and the variance function may occur at different locations. In our detailed construction, we propose simple left-hand and right-hand limit estimators of the mean and the variance functions. Each possible test statistic is based on a suitably weighted version of the sum of squared differences of these estimates for the mean and the variance functions. The asymptotic distribution of the standardized weighted type of each of the proposed test statistics, under the null hypothesis of no change in the mean and in the variance function, is established.

As  $\tau_1, \dots, \tau_d$  do not exist under the null hypothesis, we propose using an integrated version of the standardized weighted test over all possible values of  $\{\tau_i : 1 \leq i \leq d\}$ . This is a nonparametric counterpart of existing forms for the parametric case (Andrews 1993; Andrews and Ploberger 1994; Hansen 2000; among others). More than what has been established in the literature, we are able to show that the asymptotic distribution of the integrated version can still be standard normal (see Theorem 2.1) through using a newly established central limit theorem for quadratic forms of dependent processes. A crucial issue is then the choice of appropriate bandwidths. One way to deal (partially) with this problem is to construct an adaptive version of a test by taking the maximum of the test over a range of bandwidth values. We then establish that the adaptive test is asymptotically consistent not only for the case where both the mean and variance or either of them have structural change at  $\{\tau_i : 1 \leq i \leq d\}$ , but also for the case where both functions may only have a kind of asymptotic structural change. Both the size and power properties of the proposed adaptive test are illustrated via simulated and real data examples.

The paper is organized as follows. In Section 2 we introduce the general regression model, the testing problems and the proposed test statistics. We also provide the asymptotic distribution of the proposed test statistics under the null hypothesis. At the end of Section 2, we explain how our results may be extended to cover some other complicated cases. An adaptive testing procedure, when taking the maximum over a range of bandwidth values and using a Monte Carlo simulation procedure to approximate the critical level of each of the proposed tests, is discussed in Section 3. We establish consistency and efficiency of the adaptive test procedure. Section 4 contains the simulation study and in Section 5 we apply the methods to the analysis of stock indices. Section 6 contains some conclusions and further discussions. Assumptions and mathematical proofs of all theoretical results are given in the appendix.

## 2. Test statistics and asymptotic theory

Consider a nonparametric regression model of the form

$$Y_t = m(X_t) + \sigma(X_t) e_t, \quad t = 1, 2, \dots, T, \quad (1)$$

where  $\{X_t\}$  is a vector of  $d$ -dimensional strictly stationary time series variables of the form  $X_t = (X_{t1}, \dots, X_{td})^\tau$ , both  $m(x)$  and  $\sigma(x) > 0$  are defined over  $\mathbb{R}^d$  and continuous except at a finite number of points where the functions are discontinuous, and  $\{e_t\}$  is a sequence of martingale differences with  $E[e_t|\Omega_{t-1}] = 0$ ,  $E[e_t^2|\Omega_{t-1}] = 1$ ,  $E[e_t^3|\Omega_{t-1}] = 0$  and  $E[e_t^4|\Omega_{t-1}] < \infty$ , where  $\{\Omega_t\}$  is a sequence of  $\sigma$ -fields generated by  $\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$ .

In order to establish our tests, we need to introduce the following notation: for  $j = 1, 2$ ,

$$\begin{aligned} x &= (x_1, \dots, x_{r-1}, x_r, x_{r+1}, \dots, x_d)^\tau, \quad \tau = (\tau_1, \dots, \tau_{r-1}, \tau_r, \tau_{r+1}, \dots, \tau_d)^\tau, \\ m_1(x) &= m(x), \quad m_2(x) = \sigma^2(x), \quad m_{j+}(x) = \lim_{y \downarrow x} m_j(y), \quad m_{j-}(x) = \lim_{y \uparrow x} m_j(y), \\ \beta_j(\tau) &= m_{j+}(\tau) - m_{j-}(\tau) \quad \text{for } 1 \leq j \leq 2. \end{aligned}$$

Following the discussion in Delgado and Hidalgo (2000), we mainly consider the case where discontinuous points may exist at each of the  $d$  regressors for both  $m(x)$  and  $\sigma(x)$ , located at some unknown locations  $\tau_1, \dots, \tau_d$ . In detail, we assume that there are continuous functions  $m_s(x)$  and  $\sigma_s(x)$  such that

$$m(x) = m_s(x) + \beta_1(\tau) \prod_{i=1}^d I(x_i \geq \tau_i) \quad \text{and} \quad \sigma^2(x) = \sigma_s^2(x) + \beta_2(\tau) \prod_{i=1}^d I(x_i \geq \tau_i), \quad (2)$$

where  $I(A)$  is the conventional indicator function.

We assume without loss of generality that  $-1 < \tau_i < 1$  for  $1 \leq i \leq d$ . Except for some extreme cases where discontinuities of  $\nu(x) = E[Y_t^2|X_t = x]$  and  $m^2(x)$  are cancelled out, the conditional mean function  $m(\cdot)$  and conditional variance function  $\sigma^2(x) = \nu(x) - m^2(x)$  should share some discontinuities at  $\tau_1, \dots, \tau_d$ . In other words, it is not unreasonable to assume that  $m(\cdot)$  and  $\sigma^2(\cdot)$  both have some discontinuities at  $\tau_1, \dots, \tau_d$ . In addition, we also assume throughout this paper that the marginal density,  $f(x)$ , of  $\{X_t\}$  is continuous at  $\tau_i$  for  $1 \leq i \leq d$ . Other cases, including the case where  $m(x)$  and  $\sigma^2(x)$  have different discontinuities, will be briefly mentioned at the end of Section 2.

Note that  $\beta_1(\tau)$  and  $\beta_2(\tau)$  are respectively the jump sizes of  $m(\cdot)$  and  $\sigma^2(\cdot)$  at the possible singular points  $\tau_i$ . The main objective of this paper is to test whether such singular points at  $\tau_i$  for  $1 \leq i \leq d$  exist.

This section considers a simultaneous test problem of the form:

$$\mathcal{H}_0 : \beta_1^2(x) + \beta_2^2(x) = 0 \quad \text{for all } x \in \mathbb{R}^d. \quad (3)$$

In Section 3 below, when we study asymptotic consistency of the proposed test of this paper, we will specify a class of local alternatives under  $\mathcal{H}_1$ .

Before we establish our tests, we propose using a modified local linear kernel estimation method for  $\{m_{j\pm}(\cdot) : j = 1, 2\}$  (Fan and Gijbels 1996; Fan and Yao 1998) of the form

$$\hat{m}_{1\pm}(x) = \sum_{t=1}^T \hat{W}_{1\pm}(x, X_t) Y_t \quad \text{and} \quad \hat{m}_{2\pm}(x) = \sum_{t=1}^T \hat{W}_{2\pm}(x, X_t) (Y_t - \hat{m}_{1\pm}(X_t))^2, \quad (4)$$

where

$$\hat{W}_{j\pm}(x, X_t) = \frac{1}{Th_j^d} \frac{S_{2j\pm}(x) - S_{1j\pm}(x) \prod_{i=1}^d \left( \frac{x_i - X_{ti}}{h_j} \right)}{S_{2j\pm}(x) S_{0j\pm}(x) - S_{1j\pm}(x)^2} K_{j\pm} \left( \frac{x - X_t}{h_j} \right), \quad (5)$$

in which  $S_{lj\pm}(x) = \frac{1}{Th_j^d} \sum_{s=1}^T K_{j\pm} \left( \frac{x - X_s}{h_j} \right) \prod_{i=1}^d \left( \frac{x_i - X_{si}}{h_j} \right)^l$  for  $l = 0, 1, 2$  and  $j = 1, 2$ ,  $\{K_{j\pm} : j = 1, 2\}$  are one-sided probability kernel functions and  $\{h_j : j = 1, 2\}$  are bandwidth parameters.

For each  $1 \leq j \leq 2$ , we estimate  $\beta_j(\tau)$  by

$$\hat{\beta}_j(\tau) = \hat{m}_{j+}(\tau) - \hat{m}_{j-}(\tau). \quad (6)$$

A natural test statistic is based on the following form:

$$L_{1T}(\tau) = \hat{\beta}(\tau)^\top \hat{\beta}(\tau) = (\hat{m}_{1+}(\tau) - \hat{m}_{1-}(\tau))^2 + (\hat{m}_{2+}(\tau) - \hat{m}_{2-}(\tau))^2, \quad (7)$$

where  $\hat{\beta}(\tau) = (\hat{\beta}_1(\tau), \hat{\beta}_2(\tau))^\top$ , with the superscript  $\top$  denoting the transpose of a vector or matrix. Before we provide some heuristic argument about the suitability of form (7), we need the following notation:

$$\epsilon_t = Y_t - m_1(X_t) = \sigma(X_t) e_t \quad \text{and} \quad \eta_t = \epsilon_t^2 - m_2(X_t) = m_2(X_t) [e_t^2 - 1]. \quad (8)$$

This is equivalent to

$$Y_t = m_1(X_t) + \epsilon_t \quad \text{and} \quad \epsilon_t^2 = m_2(X_t) + \eta_t. \quad (9)$$

A simple decomposition of (7) implies that the leading term of  $L_{1T}(\tau)$  is

$$M_{1T}(\tau) = \sum_{t=1}^T \sum_{s=1}^T \hat{W}_1(\tau, X_s) \hat{W}_1(\tau, X_t) Y_s Y_t + \sum_{t=1}^T \sum_{s=1}^T \hat{W}_2(\tau, X_s) \hat{W}_2(\tau, X_t) \epsilon_s^2 \epsilon_t^2, \quad (10)$$

where  $\hat{W}_j(\tau, X_t) = \hat{W}_{j+}(\tau, X_t) - \hat{W}_{j-}(\tau, X_t)$  for  $j = 1, 2$ . This yields under  $\mathcal{H}_0$ ,

$$\begin{aligned}\mu_{1T}(\tau) &= E[L_{1T}(\tau)] = (1 + o(1)) \sum_{t=1}^T E \left[ \hat{W}_1^2(\tau, X_t) m_2(X_t) + \mu_4 \hat{W}_2^2(\tau, X_t) m_2^2(X_t) \right], \\ \sigma_{1T}^2(\tau) &= \text{Var}[L_{1T}(\tau)] = 2(1 + o(1)) \sum_{t=1}^T \sum_{s=1}^T E \left[ \hat{W}_1^2(\tau, X_s) m_2(X_s) \hat{W}_1^2(\tau, X_t) m_2(X_t) \right] \\ &\quad + 2(1 + o(1)) \mu_4^2 \sum_{t=1}^T \sum_{s=1}^T E \left[ \hat{W}_2^2(\tau, X_s) m_2^2(X_s) \hat{W}_2^2(\tau, X_t) m_2^2(X_t) \right],\end{aligned}\quad (11)$$

using  $E[e_t^2 | \Omega_{t-1}] = 1$  and  $E[e_t | \Omega_{t-1}] = E[e_t^3 | \Omega_{t-1}] = 0$ , where  $\mu_4 = E[e_t^4] - 1$ .

It is noted that the fact that  $\epsilon_t$  and  $\eta_t$  are uncorrelated (i.e.  $\text{cov}[\epsilon_t, \eta_t] = 0$ ) has been used in (11). This fact also supports that using an additive version of the form (7) is justifiable for testing  $\mathcal{H}_0$ . Obviously, we may replace the direct sum of  $\hat{\beta}_1^2(\cdot) + \hat{\beta}_2^2(\cdot)$  by a suitably weighted version. The case where  $\epsilon_t$  and  $\eta_t$  may also be correlated can be discussed similarly. For this case, we need to modify form (7) to include some cross terms. To present the main idea and method of this paper, however, we will focus on (7) for the uncorrelated case throughout this paper.

In the implementation,  $\mu_{1T}(\tau)$  and  $\sigma_{1T}^2(\tau)$  are estimated respectively by

$$\begin{aligned}\hat{\mu}_{1T}(\tau) &= \sum_{t=1}^T \left[ \hat{W}_{1+}^2(\tau, X_t) \hat{m}_{2+}(X_t) + \hat{W}_{1-}^2(\tau, X_t) \hat{m}_{2-}(X_t) \right] \\ &\quad + \mu_4 \sum_{t=1}^T \left[ \hat{W}_{2+}^2(\tau, X_t) \hat{m}_{2+}^2(X_t) + \hat{W}_{2-}^2(\tau, X_t) \hat{m}_{2-}^2(X_t) \right], \\ \hat{\sigma}_{1T}^2(\tau) &= 2 \sum_{t=1}^T \left[ \hat{W}_{1+}^2(\tau, X_t) \hat{m}_{2+}(X_t) + \hat{W}_{1-}^2(\tau, X_t) \hat{m}_{2-}(X_t) \right] \\ &\quad \times \sum_{s=1}^T \left[ \hat{W}_{1+}^2(\tau, X_s) \hat{m}_{2+}(X_s) + \hat{W}_{1-}^2(\tau, X_s) \hat{m}_{2-}(X_s) \right] \\ &\quad + 2\mu_4^2 \sum_{t=1}^T \left[ \hat{W}_{2+}^2(\tau, X_t) \hat{m}_{2+}^2(X_t) + \hat{W}_{2-}^2(\tau, X_t) \hat{m}_{2-}^2(X_t) \right] \\ &\quad \times \sum_{s=1}^T \left[ \hat{W}_{2+}^2(\tau, X_s) \hat{m}_{2+}^2(X_s) + \hat{W}_{2-}^2(\tau, X_s) \hat{m}_{2-}^2(X_s) \right].\end{aligned}$$

We now define a normalized version of  $L_{1T}(\tau)$  of the form

$$\hat{L}_{1T}(\tau) = \frac{L_{1T}(\tau) - \hat{\mu}_{1T}(\tau)}{\hat{\sigma}_{1T}(\tau)}. \quad (12)$$



As can be seen from equation (3), the vector of the threshold parameters  $\tau$  only appears under the alternative but not under the null. If  $\tau$  is known, we may use  $\hat{L}_{1T}(\tau)$  as a test statistic for testing (3). This location of structural breaks is however normally unknown. In the parametric case, Andrews and Ploberger (1994) propose using a weighted average power function. This paper extends such an idea to a nonparametric setting by developing a test for testing (3). There are two reasons why we are using such an averaging version. The first one is that the test statistic of this form may have some advantages in terms of weighted average power over test statistics of supremum forms as pointed out in Andrews (1993, p.824). The second reason, which is probably more important to our case and different from existing results (Andrews and Ploberger 1994), is that we are able to derive asymptotically normal distributions for the proposed test  $\hat{L}_{wT}$  in (13) below even under discontinuity and integration. In this paper we do not address any optimality properties of the proposed test statistic. Optimality of tests is an open question here. In our simulation studies we also investigated the performance of a supremum test statistic. The choice between these two and other types of test statistics is related somehow to a discussion on the performances of classical goodness-of-fit tests of the Cramér-Von Mises and Kolmogorov-Smirnov type. Comparisons of these classical goodness-of-fit tests can be found in Stephens (1974) and D'Agostino and Stephens (1986), among others.

Let  $A_{12} = \prod_{i=1}^d [a_{i \min}, a_{i \max}]$  with each  $[a_{i \min}, a_{i \max}] \subset ]-1, 1[$ , where  $a_{i \min}$  and  $a_{i \max}$  are assumed to be known, and the interval includes  $\tau_i$ . The main test statistic of this paper is defined by

$$\hat{L}_{wT} = \hat{L}_{wT}(h_1, h_2) = \frac{L_{wT} - \hat{\mu}_{wT}}{\hat{\sigma}_{wT}}, \quad (13)$$

where  $L_{wT} = \int L_{1T}(u)\pi(u)du$ ,  $\hat{\mu}_{wT} = \int \hat{\mu}_{1T}(u)\pi(u)du$  and

$$\begin{aligned} \hat{\sigma}_{wT}^2 &= 2 \sum_{t=1}^T \int \left[ \hat{W}_{1+}^2(u, X_t) \hat{m}_{2+}(X_t) + \hat{W}_{1-}^2(u, X_t) \hat{m}_{2-}(X_t) \right] \pi(u) du \\ &\times \sum_{s=1}^T \int \left[ \hat{W}_{1+}^2(v, X_s) \hat{m}_{2+}(X_s) + \hat{W}_{1-}^2(v, X_s) \hat{m}_{2-}(X_s) \right] \pi(v) dv \\ &+ 2\mu_4^2 \sum_{t=1}^T \int \left[ \hat{W}_{2+}^2(u, X_t) \hat{m}_{2+}^2(X_t) + \hat{W}_{2-}^2(u, X_t) \hat{m}_{2-}^2(X_t) \right] \pi(u) du \\ &\times \sum_{s=1}^T \int \left[ \hat{W}_{2+}^2(v, X_s) \hat{m}_{2+}^2(X_s) + \hat{W}_{2-}^2(v, X_s) \hat{m}_{2-}^2(X_s) \right] \pi(v) dv, \end{aligned}$$

in which  $\pi(\tau)$  is a weight function over  $A_{12}$ . An example of a weight function is a uniform

weight function:

$$\pi(x) = \begin{cases} \prod_{i=1}^d \left( \frac{1}{a_{i \max} - a_{i \min}} \right) & \text{if } x \in A_{12}, \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

Alternatively, we may suggest using a test statistic of the form

$$\hat{L}_{sT} = \hat{L}_{sT}(h_1, h_2) = \sup_{\tau \in A_{12}} \hat{L}_{1T}(\tau) \quad (15)$$

with  $\hat{L}_{1T}(\tau) = \frac{L_{1T}(\tau) - \hat{\mu}_{1T}(\tau)}{\hat{\sigma}_{1T}(\tau)}$ .

We have not established the asymptotic distribution for  $\hat{L}_{sT}$  which converges to a standardized Bessel process. For  $\hat{L}_{wT}$ , however, we are able to establish the following asymptotic normality. For ease of presentation the assumptions are listed in the beginning of the appendix.

**Theorem 2.1.** *Assume that Assumptions A.1–A.4 hold. Then under  $\mathcal{H}_0$*

$$\hat{L}_{wT} \rightarrow_D N(0, 1) \quad \text{as } T \rightarrow \infty. \quad (16)$$

The proof of Theorem 2.1 is postponed to the appendix below.

When the null hypothesis is rejected, we will need to further test

$$\mathcal{H}_{01} : \beta_1(x) = 0 \quad \text{or} \quad \mathcal{H}_{02} : \beta_2(x) = 0 \quad \text{for all } x \in \mathbb{R}^d. \quad (17)$$

Similarly to the construction of  $\hat{L}_{wT}$ , we may establish a univariate test  $\hat{L}_{wjT}$  for  $j = 1, 2$ . Moreover, analogously to Theorem 2.1, we can show that each  $\hat{L}_{wjT}$  is an asymptotically normal test.

We have considered the case where the same number of discontinuities occur in the  $d$  regressors at  $\tau_1, \dots, \tau_d$ . Similarly to (2) of Delgado and Hidalgo (2000), we may consider the case where discontinuities occur in all the regressors at  $\tau_r^{(j)}$  for  $1 \leq r \leq d; 1 \leq j \leq J$  for  $m_1(\cdot)$  and at  $\mu_r^{(k)}$  for  $1 \leq r \leq d; 1 \leq k \leq K$  for  $m_2(\cdot)$ . In this case, model (2) becomes

$$\begin{aligned} m(x) &= m_s(x) + \sum_{j=1}^J \beta_1(\tau^{(j)}) \prod_{r=1}^d I(x_r \geq \tau_r^{(j)}) \quad \text{and} \\ \sigma^2(x) &= \sigma_s^2(x) + \sum_{k=1}^K \beta_2(\mu^{(k)}) \prod_{r=1}^d I(x_r \geq \mu_r^{(k)}), \end{aligned} \quad (18)$$

where  $\tau^{(j)} = (\tau_1^{(j)}, \dots, \tau_d^{(j)})^T$  and  $\mu^{(k)} = (\mu_1^{(k)}, \dots, \mu_d^{(k)})^T$ .

Similarly to the construction of  $\hat{L}_{wT}$ , we can establish a corresponding test to deal with (18). Since presenting the main idea and methodology of ours is the main objective of this paper, we therefore concentrate our discussion on  $\hat{L}_{wT}$  throughout the rest of this paper.

We conclude this section by pointing out that

(i) We should use  $\hat{L}_{wT}$  to test whether the conditional mean and conditional variance functions both have discontinuities at  $\tau_1, \dots, \tau_d$  if we have no information or knowledge about the smoothness of either function or both functions. By doing so, one may avoid assuming continuity mistakenly for one of the functions. Of course this also implies that one may assume discontinuities mistakenly for one of the functions.

(ii) If the null hypothesis  $\mathcal{H}_0$  is rejected or we have some information or knowledge about the smoothness of either function, we should use either  $\hat{L}_{w1T}$  or  $\hat{L}_{w2T}$  to test which function would have discontinuities at  $\tau_1, \dots, \tau_d$ .

(iii) This section considers the case where the conditional mean and variance functions are purely nonparametric. As discussed in Section 5.2 below, it is also interesting in both theory and applications to discuss structural breaks in semiparametric time series regression models.

### 3. An adaptive testing procedure

An important but very difficult issue is the choice of two suitable bandwidths  $h_1$  and  $h_2$  that are optimal for testing purposes. Since we will emphasize on  $\hat{L}_{wT}(h_1, h_2)$  in the examples in Section 4 below, this section proposes an adaptive version of the proposed test  $\hat{L}_{wT}(h_1, h_2)$  and then discusses its asymptotic behaviors. The discussion given in this section also applies to the other tests  $\hat{L}_{wiT}(h_1, h_2)$ , for  $i = 1, 2$ . As for the smooth case (Horowitz and Spokoiny 2001; Gao and King 2004), we propose using

$$\hat{L}_w = \max_{h \in H_T} \hat{L}_{wT}(h_1, h_2), \quad (19)$$

where  $h = (h_1, h_2)$ ,  $H_T = H_{1T} \times H_{2T}$ ,  $H_{iT} = \{h_i = h_{i \max} a_i^k : h_i \geq h_{i \min}, 0 \leq k \leq J_{iT} - 1\}$ , in which  $0 < h_{i \min} < h_{i \max}$ , and  $0 < a_i < 1$  for  $i = 1, 2$ . With  $J_{iT}$  denoting the number of elements of  $H_{iT}$ , we have  $J_{iT} \leq \log_{1/a_i}(h_{i \max}/h_{i \min})$ .

It should be noted that each  $H_{iT}$  is just a set of discrete bandwidth values, starting with the smallest value  $h_{i \min}$  and ending with the largest value  $h_{i \max}$ . The ratio between two consecutive bandwidths in this (increasing) geometric grid of values is  $1/a_i$ .

**Simulation scheme:** We now discuss how to obtain a critical value for  $\hat{L}_w$ . The exact  $\alpha$ -level critical value,  $l_e(\alpha)$  ( $0 < \alpha < 1$ ) is the  $1 - \alpha$  quantile of the exact finite-sample distribution of  $\hat{L}_w$ . Because the distribution of  $\{e_t\}$  is unknown, we cannot evaluate  $l_e(\alpha)$  in practice. We therefore need to approximate the finite-sample distribution of  $\hat{L}_w$  based on a pre-specified distributional structure of  $\{e_t\}$ .

We now propose using the following Monte Carlo simulation procedure to choose a simulated  $\alpha$ -level critical value,  $l_\alpha$ , as follows.

1. For each  $t = 1, 2, \dots, T$ , generate  $Y_t^* = \hat{m}_1(X_t) + \sqrt{\hat{m}_2(X_t)}e_t^*$ , where the original sample  $(X_1, \dots, X_T)$  acts in the resampling as a fixed design, and  $\hat{m}_i(\cdot)$  for  $i = 1, 2$  are either  $\hat{m}_{i+}(\cdot)$  or  $\hat{m}_{i-}(\cdot)$  and  $\{e_t^*\}$  is a sequence of independent and identically distributed random samples drawn from a pre-specified distribution. Use the data set  $\{(X_t, Y_t^*) : t = 1, 2, \dots, T\}$  to compute  $\hat{L}_w$ . Let  $l_\alpha$  be the  $1 - \alpha$  quantile of the distribution of  $\hat{L}_w$  based on  $\{(X_t, Y_t^*) : 1 \leq t \leq T\}$ .
2. Repeat the above step  $B$  times and produce  $B$  versions of  $\hat{L}_w$  denoted by  $\hat{L}_w^*(b)$  for  $b = 1, 2, \dots, B$ . Use the  $B$  values of  $\hat{L}_w^*(b)$  to construct their empirical distribution function, that is,  $F_w^*(u) = \frac{1}{B} \sum_{b=1}^B I(\hat{L}_w^*(b) \leq u)$ . Use the  $1 - \alpha$  quantile of the empirical distribution function,  $l_\alpha^*$ , to approximate the simulated  $\alpha$ -level critical value,  $l_\alpha$ .

It should be noted that both  $l_\alpha$  and  $l_\alpha^*$  depend on  $\{(X_t, Y_t; e_t^*) : 1 \leq t \leq T\}$ . It should also be noted that a similar procedure for parametric threshold testing has already been proposed in Hansen (1996).

In both theory and practice, there are normally two different simulation procedures we could use for finding our critical value. The first is the proposed Monte Carlo method. The second method is to use a bootstrap resampling procedure to generate  $\{e_t^*\}$  (see Hjellvik, Yao and Tjøstheim 1998; Franke, Kreiss and Mammen 2002 for example). In such a case, we will need to treat  $(X_1, \dots, X_T)$  as a fixed sample when we generate resamples of  $\{e_t^*\}$ . As the choice of  $l_\alpha$  is only a small step in this section, we adopt the more applicable Monte Carlo method and then establish the main theoretical results in Theorems 3.1–3.3 below.

To study the power function of  $\hat{L}_w$ , we need to specify the alternative hypothesis with the following notation: for  $u \in (-1, 1)^d$

$$\rho(u) = \beta(u)^\top \beta(u). \quad (20)$$

Since a jump in  $\hat{\beta}_j(\tau)$  is usually observed in a neighborhood of  $\tau$ , we consider negating  $\mathcal{H}_0$  in an asymptotically small interval containing  $\tau$ . This section thus proposes using the following two types of alternatives:

Case A: Assume that there are some  $c_{1\min} > 0$  and  $\epsilon_1 > 0$  such that

$$\mathcal{H}_1 : \rho(u) \geq c_{1\min} \quad (21)$$

uniformly in  $u \in \prod_{i=1}^d (\tau_i - \epsilon_1, \tau_i + \epsilon_1)$ , where  $\epsilon_1 = \left(\frac{\log(\log(T))}{T}\right)^{\frac{1}{2d}}$ .

Case B: Assume that there are some  $c_{2\min} > 0$  and  $\epsilon_2 > 0$  such that

$$\mathcal{H}_1 : \rho(u) \geq c_{2\min} \frac{\log(\log(T))}{T} \quad (22)$$

uniformly in  $u \in \prod_{i=1}^d (\tau_i - \epsilon_2, \tau_i + \epsilon_2)$ , where  $\epsilon_2 = (\log(\log(T)))^{-\frac{1}{4d}}$ .

We now state the following results; their proofs are relegated to the appendix.

**THEOREM 3.1.** *Assume that Assumptions A.1–A.6 hold. Then under  $\mathcal{H}_0$*

$$\lim_{T \rightarrow \infty} P(\hat{L}_w > l_\alpha) = \alpha.$$

**THEOREM 3.2.** *Assume that Assumptions A.1–A.6 hold. Then under  $\mathcal{H}_1$  of (21)*

$$\lim_{T \rightarrow \infty} P(\hat{L}_w > l_\alpha) = 1.$$

**THEOREM 3.3.** *Let Assumptions A.1–A.6 with  $h_{i\max} = c_{i\max} (\log(\log(T)))^{-\frac{1}{2d}}$  for some constant  $c_{i\max} > 0$  ( $i = 1, 2$ ) hold. Then under  $\mathcal{H}_1$  of (22)*

$$\lim_{T \rightarrow \infty} P(\hat{L}_w > l_\alpha) = 1.$$

Theorem 3.1 implies that  $l_\alpha$  is an asymptotically correct  $\alpha$ -level critical value under any model in  $\mathcal{H}_0$ , while Theorem 3.2 establishes the asymptotic consistency of  $\hat{L}_w$  for the case where  $\rho(u) > 0$  uniformly in  $u \in \prod_{i=1}^d (\tau_i - \epsilon_1, \tau_i + \epsilon_1)$ . Theorem 3.2 therefore implies that the adaptive test  $\hat{L}_w$  is asymptotically consistent when each  $\tau_i$  is a discontinuity of either  $m_1(x)$ ,  $m_2(x)$  or both. Theorem 3.3 shows that  $\hat{L}_w$  is still asymptotically consistent when the ‘distance’ between continuity and discontinuity is of the shortest possible rate of  $T^{-\frac{1}{2}}\sqrt{\log\log T}$  under such a nonparametric setting. To the best of our knowledge, this is the first result of this kind in the field of nonparametric testing for structural breaks.

#### 4. Simulation study

In this section we investigate the finite-sample performance of the proposed testing procedures via simulations. We consider two testing models. Model 1 below is a regression case where the observations are independent and identically distributed (i.i.d.): the design points  $X_t$  are uniformly distributed on  $[-1, 1]$  and the errors  $e_t$  are i.i.d.  $N(0, 1)$ . Model 2 is an example in the time series context. In both examples, we assume that there is some structural break at the same location  $\mu$  for both  $m(\cdot)$  and  $\sigma^2(\cdot)$ .

This section focuses only on the case where both the conditional mean and conditional variance functions have a same single change-point with fixed jump size. When there are more regressors and jumps with the jump size in each case being quite small, the issue of whether the testing procedure is sensitive would need to be examined in future research.

**Model 1.** The mean function  $m_1(x) = m(x)$  and the variance function  $m_2(x) = \sigma^2(x)$  are piecewise constant and contain at most one jump at 0:

$$m_1(x) = \begin{cases} \alpha & -1 \leq x < 0 \\ \alpha + \beta & 0 \leq x \leq 1, \end{cases}$$

$$m_2(x) = \begin{cases} \gamma^2 & -1 \leq x < 0 \\ \gamma^2 + \delta & 0 \leq x \leq 1. \end{cases}$$

The size of the possible jump is quantified by the parameters  $\beta$  and  $\delta$ . Note that  $\delta > -\gamma^2$  is needed to ensure that the variance function is strictly positive.

**Model 2:** Consider the time series model  $Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})e_t$ , with

$$m(Y_{t-1}) = \begin{cases} 0.8Y_{t-1} - \lambda Y_{t-1} & \text{if } Y_{t-1} \leq 0 \\ 0.8Y_{t-1} & \text{if } Y_{t-1} > 0 \end{cases}$$

$$\sigma^2(Y_{t-1}) = \begin{cases} 1 + 0.5Y_{t-1}^2 + \eta(0.5 + 0.3Y_{t-1}^2) & \text{if } Y_{t-1} \leq 0 \\ 1 + 0.5Y_{t-1}^2 & \text{if } Y_{t-1} > 0, \end{cases}$$

and  $e_t \sim N(0, 1)$ .

This is an example of a so-called double threshold autoregressive conditional heteroscedastic model. Such models have been discussed by Wong and Li (1997, 2000) among others, and are especially useful for modelling financial time series. It follows from Lemma 3.1 of Masry and Tjøstheim (1995) that Assumption A.1(i) holds when  $m(Y_{t-1})$  and  $\sigma^2(Y_{t-1})$  are of the forms in Model 2.

In the next subsections we use these two models to investigate various aspects of the proposed testing procedures. In all simulations we used a uniform kernel density function supported on  $[-0.5, 0.5]$ . More details about how the simulation study can be done in practice is available from Gao, Gijbels and Van Bellegem (2006).

#### 4.1. Implementation issues and basic distributional properties of the test statistics

We investigate the distribution of the test statistic under the null hypothesis  $\mathcal{H}_0$ . We therefore consider Model 1 with parameters  $\beta = \delta = 0$ ,  $\alpha = 2$  and  $\gamma = 1$ . We illustrate the behaviour of the statistic  $\hat{L}_{1T}(0)$ , defined in (12), and  $\hat{L}_{wT}$ , as defined in (13), with the weight function as in (14) with  $A_{12} = [-0.9, 0.9]$ . Recall that both statistics have an asymptotic standard normal distribution. The statistic  $\hat{L}_{1T}(0)$  is not appropriate for our testing problem, and it is only included for comparison purpose.

We simulate 100 samples of size  $T = 200$  from Model 1, and present the estimated density, based on the obtained 100 values for  $\hat{L}_{1T}(0)$ , using the standard kernel smoothing procedure of the software package R (i.e. the function `density`). The resulting density estimate is plotted in Figure 1. The other densities plotted in the figure show the estimated densities based on samples of size  $T = 600$  and  $T = 1200$  respectively. Note the slow convergence of the statistic  $\hat{L}_{1T}(0)$ . Further investigation of this revealed that the slow convergence could be due to the slow convergence of the estimator of the variance  $\hat{\sigma}_{1T}^2(0)$ . Looking at the right-hand side of Figure 1 we can see that the statistic  $\hat{L}_{wT}$  converges, under the null hypothesis, reasonably fast to  $N(0, 1)$ .

We now investigate, through the same example, the finite-sample power properties of the testing procedures  $\hat{L}_{wT}$  as defined in (13).

We proceed as follows:

1. Fix the sample size  $T$  (in the following,  $T = 200$  or  $600$ ).
2. Compute the empirical distribution function of the test statistics under the null (i.e. when  $\beta = \delta = 0$ ). This computation is based on  $M = 5000$  simulations of length  $T$  from Model 1 under the null.
3. Fix the level of the test :  $\alpha = 5\%$  or  $10\%$ .

4. For a given value of  $\beta \geq 0$  and  $\delta \geq 0$ , we compute the number of rejections of the test. This is based on 500 simulations from Model 1 under the alternative.

Table 1 reports the results of this simulation. We can see that the powers of the tests increase dramatically when the sample size increases from  $T = 200$  to  $T = 600$ . Furthermore, for the small sample size of  $T = 200$ , the power of the test based on  $\hat{L}_{sT} = \sup_{u \in A_{12}} \hat{L}_{1T}(u)$  seems to be better than the power of the test based on  $\hat{L}_{wT}$ . For the medium sample size of  $T = 600$ , both tests perform quite comparable. From the more extensive simulation study that we performed there was no clear evidence for a best test among the two. For brevity we will, in the following finite sample studies and empirical applications, only assess the performance of an adaptive version of  $\hat{L}_{wT}$ .

#### 4.2. Adaptive testing procedure in time series

So far we have not addressed the issue of how to choose the bandwidth parameters. Recall the notation from Section 3. We restrict here to  $H_{1T} = H_{2T}$ . For the simulation study we take  $J_{iT} = 5$ , for  $i = 1, 2$ . Further, we take  $h_{1\max} = h_{2\max} = c_{\max} \sqrt{(\log \log T)^{-1}}$ , with  $c_{\max}$  being a constant to be chosen. This choice of the largest bandwidth in the grid of bandwidth values is motivated by the conditions in Theorem 3.3.

The adaptive testing procedure, including a practical choice for the grid  $H_{1T} = H_{2T}$ ,

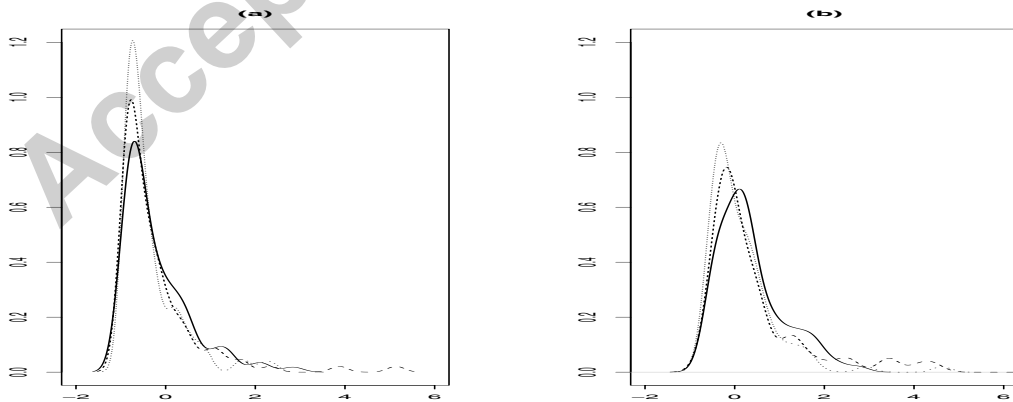


Figure 1: *Estimated density of  $\hat{L}_{1T}(0)$  (left picture) and of  $\hat{L}_{wT}$  (right picture) from 100 simulated samples from Model 1, with  $\beta = \delta = 0$ ,  $\alpha = 2$  and  $\gamma = 1$ , for sample sizes  $T = 200$  (dotted curve),  $T = 600$  (dashed curve), and  $T = 1200$  (solid curve).*



Table 1: *Simulation results on the power of the tests for Model 1 (in percentages) for various alternatives*

Alternative		$T = 200$		$T = 600$			
		$\hat{L}_{sT}$	$\hat{L}_{wT}$	$\hat{L}_{sT}$	$\hat{L}_{wT}$		
(1)	$\beta = 1, \delta = 1$	$\alpha = 5\%$		13.8	8.2	85.6	86.2
		$\alpha = 10\%$		25.4	18.6	94.6	95.4
(2)	$\beta = 2, \delta = 1$	$\alpha = 5\%$		41.2	29.2	98.8	98.6
		$\alpha = 10\%$		52.4	42.2	99.6	100
(3)	$\beta = 2, \delta = 0$	$\alpha = 5\%$		40.0	36.8	92.6	94.4
		$\alpha = 10\%$		48.0	48.6	98.6	96.8
(4)	$\beta = 0, \delta = 2$	$\alpha = 5\%$		39.6	21.8	100	100
		$\alpha = 10\%$		59.8	54.2	100	100
(5)	$\beta = 2, \delta = 2$	$\alpha = 5\%$		52.0	48.8	100	100
		$\alpha = 10\%$		70.0	76.8	100	100

and based on

$$\hat{L}_w = \max_{h_1 \in H_{1T}, h_2 \in H_{2T}} \hat{L}_{wT}(h_1, h_2)$$

reads as follows (for simulated data)

**Step 1 :** *Determining the grids  $H_{1T}$  and  $H_{2T}$  of bandwidths.* Consider  $H_{1T} = H_{2T} = H_T = \{h_k = h_{\max} a^k : h_k \geq h_{\min}, k = 0, 1, \dots, 4\}$ , in which  $0 < h_{\min} < h_{\max} = c_{\max} \sqrt{(\log \log T)^{-1}}$  and  $0 < a < 1$ .

- Simulate  $J$  sets of data  $Z_i^{(j)} = (X_i^{(j)}, Y_i^{(j)})$ ,  $1 \leq i \leq T$  ( $T$  is fixed),  $1 \leq j \leq J$  under  $\mathcal{H}_0$ .
- For each sample  $j$ , we determine the bandwidth  $h^{(j)}$  that minimizes

$$\sum_q \sum_t \left\{ \hat{m}_{1,h}^{-q}(X_t) - Y_t \right\}^2 + \left\{ \hat{m}_{2,h}^{-q}(X_t) - (Y_t - \hat{m}_{1,h}^{-q}(X_t)) \right\}^2,$$

where  $\hat{m}_{i,h}^{-q}$  ( $i = 1, 2$ ) denotes the estimator of  $m_i$  ( $i = 1, 2$ ) based on the sample without  $X_q$  and with bandwidth  $h$ . This selection procedure can be viewed as a cross-validation method.

- Compute the mean of all  $h^{(j)}$  ( $1 \leq j \leq J$ ), and call it  $\bar{h}$ . Fix  $0 < a < 1$  and compute  $h_{\max}$  (or equivalently  $c_{\max}$ ) such that the mean value of the grid  $H_{1T}$  is  $\bar{h}$ .

**Step 2 :** *Estimation of the critical values.* Simulate  $M$  sets of data  $Z_i^{(m)} = (X_i^{(m)}, Y_i^{(m)})$ ,  $1 \leq i \leq T$  ( $T$  is fixed),  $1 \leq j \leq M$  under  $\mathcal{H}_0$ . For each sample, compute  $\hat{L}_w = \hat{L}_w(m)$ . For a given level of the test, the critical value is computed from the empirical distribution function  $F(u) = \frac{1}{M} \sum_{m=1}^M I(\hat{L}_w(m) \leq u)$ .

**Step 3 :** *Estimating the power of the test against  $\mathcal{H}_1$ .* Simulate  $B$  sets of data  $Z_i^{(b)} = (X_i^{(b)}, Y_i^{(b)})$ ,  $1 \leq i \leq T$  ( $T$  is fixed),  $1 \leq j \leq B$  under  $\mathcal{H}_1$ . Perform the test over each sample set, using the critical value derived in step 2, and compute the percentage of rejection.

We now evaluate the performance of the above adaptive testing procedure for Model 2. Model 2 is an example of a double threshold autoregressive conditional heteroscedastic model. We consider the following situations:

- (1)  $\lambda = \eta = 0$ , i.e.  $\mathcal{H}_0$ , an AR – ARCH model;
- (2)  $\lambda = 0.7$  and  $\eta = 0$  i.e. a threshold AR – ARCH, in which the sign of the AR coefficient remains the same after the structural break;
- (3)  $\lambda = 1.6$  and  $\eta = 0$  i.e. a threshold AR – ARCH where the sign of the AR coefficient changes
- (4)  $\lambda = 0.7$  and  $\eta = -1$ , a double threshold AR-ARCH model, with breakpoints that are difficult to detect;
- (5)  $\lambda = 1.6$  and  $\eta = -1$ , a double threshold AR-ARCH model.

The simulation results, for sample sizes  $T = 200$  and  $400$ , are summarized in Table 2 below. The distributional law under  $\mathcal{H}_0$  is based on  $M = 1000$  simulations. Table 2 shows that the adaptive test has little size distortion even when the sample size is as moderate as  $T = 200$ . This is mainly because of the combined employment of a simulated critical value instead of an asymptotic critical value of  $l_{0.05} = 1.645$  as well as an adaptive version of  $\hat{L}_{wT}$  over a set of suitable bandwidths. The computation of the power under the alternatives (1) to (5) is based on  $B = 200$  simulations. In terms of the power

Table 2: Powers (in percentages) of the test for various alternatives for Model 2

Alternative			$T = 200$	$T = 400$
			$\hat{L}_w$	$\hat{L}_w$
(1)	$\lambda = 0, \eta = 0$	$\alpha = 5\%$	4.0	4.0
		$\alpha = 10\%$	7.5	9.0
(2)	$\lambda = 0.7, \eta = 0$	$\alpha = 5\%$	34.5	62.0
		$\alpha = 10\%$	43.0	69.5
(3)	$\lambda = 1.6, \eta = 0$	$\alpha = 5\%$	52.0	73.0
		$\alpha = 10\%$	60.5	79.5
(4)	$\lambda = 0.7, \eta = -1$	$\alpha = 5\%$	58.5	85.0
		$\alpha = 10\%$	62.5	94.5
(5)	$\lambda = 1.6, \eta = -1$	$\alpha = 5\%$	65.5	88.0
		$\alpha = 10\%$	75.5	93.5

performance, the adaptive test procedure seems to work reasonably well and have some good power properties even for moderate sample sizes such as  $T = 400$ .

### 5. Application to the analysis of stock markets

For real data analysis we use the adaptive testing procedure as described in Section 3, in combination with the data-driven way for determining the grid of bandwidth values. Given the observed data  $(Y_t, X_t)$ ,  $t = 1, \dots, T$ , we generate in the empirical bootstrap procedure  $Y_t^* = \hat{m}_1(X_t) + \sqrt{\hat{m}_2(X_t)}e_t^*$ , with  $\hat{m}_1(x)$  and  $\hat{m}_2(x)$  being the (continuous) two-sided local linear kernel estimators with cross-validation bandwidths  $h_1^{CV}$  and  $h_2^{CV}$ , and where  $\{e_t^*\}$  is a sequence of i.i.d. Gaussian errors with mean 0 and variance 1. From the observed data  $(Y_t, X_t)$ ,  $t = 1, \dots, T$ , we compute the test statistic  $\hat{L}_w$  and the corresponding  $p$ -values from the empirical bootstrap distribution  $F_w^*(u)$ .

#### 5.1. Univariate data analysis.

The data analysis focuses on closing values for seven stock indices from July, 1, 1988 to November 14, 2001. These indices are the FTSE 100 (United Kingdom), Dax 30 (Germany), CAC 40 (France), Swiss Market Price (Switzerland), Toronto SE 300 (Canada),

Table 3: Analysis of stock returns:  $p$ -value from the empirical bootstrap distribution  $F_w^*(u)$ . Numbers indicated by \* (respectively \*\*) refer to significance on the 0.1 (respectively 0.05) level

Period	FTSE 100	Dax 30	CAC 40	Swiss	SE 300	Nikkei 225	S&P 500
1	0.83	0.54	0.78	0.06*	0.80	0.75	0.55
2	0.50	0.32	0.19	0.14	0.13	0.01**	0.85
3	0.13	0.90	0.76	0.18	0.15	0.16	0.65
4	0.78	0.44	0.88	0.03**	0.38	0.27	0.73
5	0.37	0.96	0.42	0.89	0.03**	0.08*	0.00**
6	0.79	0.64	0.44	0.76	0.55	0.17	0.57
7	0.95	0.53	0.60	0.99	0.59	0.44	0.84

Nikkei 225 Index (Japan) and the S&P 500 Index (United States). These data were analysed by Chen, Chiang and So (2003) and were taken from *Data Stream International*. For each index  $I_t$ , we computed the daily stock-return series using the conventional definition, that is  $R_t = 100 (\log I_t - \log I_{t-1})$ . Basic statistics on these indices may be found in Table 1 of Chen, Chiang and So (2003).

The aim of this section is to apply our test of continuity on different periods of time for each return. We consider the following seven bi-annual periods of time: PERIOD 1: July, 1, 1988 to June, 29, 1990; PERIOD 2: July, 2, 1990 to June, 30, 1992; PERIOD 3: July, 1, 1992 to June, 30, 1994; PERIOD 4: July, 1, 1994 to June, 28, 1996; PERIOD 5: July, 1, 1996 to June, 30, 1998; PERIOD 6: July, 1, 1998 to June, 30, 2000; PERIOD 7: July, 3, 2000 to November, 14, 2001.

Obtained  $p$ -values from the empirical bootstrap distribution of the  $7 \times 7 = 49$  tests based on  $\hat{L}_w$ . Values that are significant at a 0.1 (respectively 0.05) significance level are indicated by \* (respectively \*\*). A first observation is that the test based on  $\hat{L}_w$  gives some illustrative conclusions. For example, the test rejects the null hypothesis for two periods of the Swiss index. It does run in the same direction for the Japanese and the Canadian indices, which contain several periods where the hypothesis of continuity is rejected. For all periods of the UK, German and French index, this hypothesis is not rejected, meaning that it is likely that there is no change of regime in these markets for the given period.

### 5.2. Intermarket analysis.

As stock markets in different countries operate in different time zones, they have different opening and closing times and it is interesting to study the effect of market closing information flowing from a trading market to another. Some works such as Chen, Chiang and So (2003) have showed that “the causal relationships have been dominated by information running from the US market into the international markets, although a minor feedback is found from the German and Japanese markets.”

Let  $I_t^k$  be the closing price of the market  $k$  at time  $t$ , and  $I_{t-m}^{\text{us}}$  be the closing price of the S&P 500 Index at time  $t - m$ , where  $m$  denotes the delay between the closing of each market (for instance  $m = 5$  hours if  $k$  denotes the UK market). The sequence of observations can then be written as  $\{I_t^k, I_{t+m}^{\text{us}}, I_{t+1}^k, I_{t+1+m}^{\text{us}}, \dots\}$  and, correspondingly, the sequence of returns is  $\{R_t^k, R_{t+m}^{\text{us}}, R_{t+1}^k, R_{t+1+m}^{\text{us}}, \dots\}$ . The causal relationship from the US market to the other markets indicates that the observation of  $\{R_{t+m-1}^{\text{us}}\}$  is a valuable information to explain  $\{R_t^k\}$ . At the same time, nonsynchronous trading is very often responsible for autocorrelation in index returns (see e.g. Lo and MacKinlay 1990). On the basis of these econometric facts, our intermarket analysis will study the relationship between  $X_t = R_{t+m-1}^{\text{us}}$  and  $Y_t = R_t^k - \phi R_{t-1}^k$ , where  $\phi$  is an autoregressive coefficient and  $k$  denotes a market different from the US market.

In the following, we consider the nonparametric regression model (1) between  $X_t$  and  $Y_t$ , and we apply our test of discontinuity on this model. Namely, the alternative of the test can be written

$$R_t^k - \phi R_{t-1}^k = \begin{cases} m_1(R_{t+m-1}^{\text{us}}) + \sigma_1(R_{t+m-1}^{\text{us}})e_t & \text{if } R_{t+m-1}^{\text{us}} \leq \delta \\ m_2(R_{t+m-1}^{\text{us}}) + \sigma_2(R_{t+m-1}^{\text{us}})e_t & \text{if } R_{t+m-1}^{\text{us}} > \delta, \end{cases}$$

where  $\{e_t\}$  is a sequence of Gaussian errors and  $\delta$  is some threshold parameter. The alternative then models two different regimes in response to bad news ( $R_{t+m-1}^{\text{us}} \leq \delta$ ) or good news ( $R_{t+m-1}^{\text{us}} > \delta$ ). In other words, the test allows to check an asymmetrical reaction to the S&P 500 Index. Such asymmetry would mean that a national market behaves differently after negative news from US than after positive news. Note that this alternative is a generalisation of some parametric nonlinear models already proposed in the literature (see e.g. Chen, Chiang and So 2003).

In our experiment, we again divide the data into seven periods of time. Table 4 shows the  $p$ -values from the empirical bootstrap distribution. The test based on  $\hat{L}_w$  gives some informative conclusions, for example, the CAC 40 index: a significant value is shown

Table 4: *Intermarket analysis of stock returns per period: estimation of the autoregressive coefficient ( $\hat{\phi}$ ) (first column) and the empirical  $p$ -value ( $p^*$ ) for the test of asymmetry based on  $\hat{L}_w$  (second column). Numbers indicated by \* (respectively \*\*) refer to significance on the 0.1 (respectively 0.05) level*

	FTSE 100		Dax 30		CAC 40		Swiss		SE 300		Nikkei 225	
	$\hat{\phi}$	$p^*$	$\hat{\phi}$	$p^*$	$\hat{\phi}$	$p^*$	$\hat{\phi}$	$p^*$	$\hat{\phi}$	$p^*$	$\hat{\phi}$	$p^*$
1	0.09	0.86	0.09	0.94	-0.03	0.95	0.07	0.06*	0.05	0.95	0.15	0.86
2	0.03	0.85	0.04	0.88	-0.02	0.35	0.03	0.16	-0.03	0.96	0.32	0.04**
3	0.07	0.74	0.07	0.76	0.07	0.55	0.09	0.52	0.13	0.48	0.26	0.01**
4	0.02	0.15	0.02	0.08**	-0.01	0.95	-0.04	0.38	-0.02	0.54	0.17	0.36
5	0.13	0.56	0.14	0.41	-0.07	0.03**	0.02	0.10*	0.05	0.47	0.19	0.02**
6	0.07	0.81	0.07	0.81	0.07	0.24	0.09	0.16	0.07	0.14	0.08	0.01**
7	-0.02	0.69	-0.02	0.53	-0.01	0.60	-0.04	0.79	0.01	0.62	0.04	0.76

during period 5. This is also the case for period 4 of Dax 30. Our test concludes that the SE 300 index does not reject while the alternative model seems more appropriate to model the Nikkei 225 index. There are also some interesting conclusions from the test, for instance at period 5 for the FTSE 100 and the Dax 30 indices. In these two cases, the empirical  $p$ -value of the test based on  $\hat{L}_w$  is larger than 0.4. Overall, the test statistic based on  $\hat{L}_w$  seems quite illustrative.

## 6. Conclusions

The form considered in (7) is of a very simple nature, and is inspired by the simplest statistics for detecting change-points in a regression function: squared differences of estimated right-hand and left-hand limits. Instead of looking at squared differences one could also look at absolute differences, as is done in the i.i.d. regression case (see e.g. Grégoire and Hamrouni 2002). Also, instead of taking the sum of the squared differences one might consider taking for example, the maximum of these differences, etc. Further research is needed here for identifying the most appropriate test statistics.

An important remaining issue in this nonparametric testing procedure is the choice of the grid of bandwidth values. Although we proposed in Section 4 a data-driven way to choose this grid, we still need to subjectively choose the geometric factor in the grid. The

choice of this factor remains unsolved theoretically. Of course, if computational costs, are of no concern, one could simply choose this factor as close as possible to one to have a very fine grid. Note that in choosing the grid we rely on a cross-validation criterion for the estimation task. This may not be the most appropriate criterion to consider in a testing context, but the important question is what would be a good criterion here.

In Section 5 we illustrated the use of the methods on stock market data. Note that many other economic applications would be worth considering, such as unemployment rate (Koop and Potter 1999) to name but a few.

## Appendix

This appendix lists the necessary assumptions for the establishment and the proof of the main results given in Sections 2 and 3.

### A.1. Assumptions

ASSUMPTION A.1. (i) Assume that the process  $(X_t, Y_t)$  is strictly stationary and  $\alpha$ -mixing with mixing coefficient  $\alpha(t)$  defined by

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty\} \leq C_\alpha \alpha^t$$

for all  $s, t \geq 1$ , where  $0 < C_\alpha < \infty$  and  $0 < \alpha < 1$  are constants, and  $\Omega_i^j$  denotes the  $\sigma$ -field generated by  $\{(X_t, Y_t) : i \leq t \leq j\}$ .

(ii) Assume that  $P(0 < \min_{t \geq 1} \sigma(X_t) \leq \max_{t \geq 1} \sigma(X_t) < \infty) = 1$  and that for all  $t \geq 1$  and  $1 \leq i \leq 4$ ,

$$P(E[e_t^i | \Omega_{t-1}] = \mu_i) = 1,$$

where  $\mu_1 = \mu_3 = 0$ ,  $\mu_2 = 1$  and  $\mu_4$  is a real constant, and  $\Omega_t = \sigma\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$  is a sequence of  $\sigma$ -fields generated by  $\{(X_{s+1}, Y_s) : 1 \leq s \leq t\}$ .

(iii) Let  $\zeta_t = \epsilon_t$  or  $\eta_t$ . In addition,  $E\left[\left|\zeta_{t_1}^{i_1} \zeta_{t_2}^{i_2} \cdot \zeta_{t_l}^{i_l}\right|^{1+\delta_2}\right] < \infty$  for all  $t_1, t_2, \dots, t_l$  and some small constants  $\delta_i > 0$ , where  $2 \leq l \leq 4$  is an integer,  $0 \leq i_j \leq 4$  and  $\sum_{j=1}^l i_j \leq 8$ .

ASSUMPTION A.2. (i) each  $m_i(x)$  has two continuous derivatives at each  $x \in [-1, 1]^d - \{\tau\}$  for some  $\tau \in \mathbb{R}$ .

(ii) Assume that  $m_{i-}^{(j)}(y) = \lim_{x \uparrow y} m_i^{(j)}(x)$  and  $m_{i+}^{(j)}(y) = \lim_{x \downarrow y} m_i^{(j)}(x)$  exist for all  $i = 1, 2$  and  $j = 0, 1, 2$ .

(iii) Let  $f_{s_1, s_2, \dots, s_l}(\cdot)$  be the joint probability density of  $(X_{1+s_1}, \dots, X_{1+s_l})$  ( $1 \leq l \leq 4$ ). Assume that each  $f_{s_1, s_2, \dots, s_l}(x)$  is twice differentiable in  $x \in \mathbb{R}^{ld} = (-\infty, \infty)^{ld}$  for  $1 \leq l \leq 4$ .

Throughout this paper, we will simply use  $f(x_1, \dots, x_l)$  to denote the joint density function of  $(X_{1+s_1}, \dots, X_{1+s_l})$  for simplicity of notation.

ASSUMPTION A.3. (i) For each  $i = 1$  or  $2$ , the kernel function  $K_{i\pm}(x) = \prod_{r=1}^d k_{i\pm}(x_r)$ . Assume that the kernel function  $k_{i-}$  has the support interval  $[-1, 0]$  and satisfies  $k_{i-}(-1) = k_{i-}(0) = 0$  and  $\int_{-1}^0 |u|^j |k_{i+}(u)|^l du < \infty$  for  $i = 1, 2$ ,  $j = 0, 1, 2$  and  $l = 1, 2$ .

(ii) For each  $i$ , the first derivative,  $k_{i-}^{(1)}(\cdot)$ , of  $k_{i-}(\cdot)$  is continuous on  $[-1, 0]$  and  $k_{i-}^{(1)}(0) < 0$ .

(iii) For each  $i = 1$  or  $2$ , the kernel function  $k_{i+}$  defined by  $k_{i+}(x_r) = k_{i-}(-x_r)$  for all  $1 \leq r \leq d$  satisfies the corresponding conditions.

(iv)  $h_i \rightarrow 0$  and  $Th_i^d \rightarrow \infty$  as  $T \rightarrow \infty$  and for  $i = 1, 2$ . In addition, there is an absolute constant  $0 < c_1 < \infty$  such that  $\lim_{T \rightarrow \infty} \frac{h_2}{h_1} = c_1$ .

ASSUMPTION A.4. Assume that  $\pi(\cdot)$  is a bounded probability density function with compact support  $C(\pi)$  in  $\mathbb{R}^d$ . In addition,  $\inf_{\mu \in C(\pi)} \pi(\mu) \geq c_{\min}(\pi) > 0$  for some constant  $c_{\min}(\pi) > 0$ .

ASSUMPTION A.5. (i) For each  $t = 1, 2, \dots, T$ , generate  $Y_t^* = \hat{m}_1(X_t) + \sqrt{\hat{m}_2(X_t)} e_t^*$  under  $\mathcal{H}_0$ , where  $\{e_t^*\}$  is a sequence of independent and identically distributed random samples drawn from a pre-specified distribution with  $E[e_t^*] = E[e_t^{*3}] = 0$ ,  $E[e_t^{*2}] = 1$  and  $E[e_t^{*4}] = \mu_4 < \infty$ , and  $\hat{m}_i(\cdot)$  for  $i = 1, 2$  are defined by either  $\hat{m}_{i+}(\cdot)$  or  $\hat{m}_{i-}(\cdot)$ .

(ii) Let  $m_i^*(x)$  for  $i = 1, 2$  be the corresponding continuous versions of  $m_i(x)$  under  $\mathcal{H}_0$ . Define  $\epsilon_t^* = \sqrt{m_2^*(X_t)} e_t^*$ ,  $\eta_t^* = m_2^*(X_t)[e_t^{*2} - 1]$ , and  $\zeta_t^* = \epsilon_t^*$  or  $\eta_t^*$ . Assume that

$$E \left[ \left| \zeta_{t_1}^{*i_1} \zeta_{t_2}^{*i_2} \cdots \zeta_{t_l}^{*i_l} \right|^{1+\beta_*} \right] < \infty$$

for all  $t_1, t_2, \dots, t_l$  and some small constants  $\alpha_* > 0$  and  $\beta_* > 0$ , where  $2 \leq l \leq 4$  is an integer,  $0 \leq i_j \leq 4$  and  $\sum_{j=1}^l i_j \leq 8$ .

ASSUMPTION A.6. Let  $H_{iT} = \{h_i = h_{i \max} a_i^k : h_i \geq h_{i \min}, k = 0, 1, 2, \dots, J_{iT} - 1\}$ , where  $0 < a_i < 1$ ,  $0 < h_{i \min} = c_{i \min} T^{-\gamma_i} < h_{i \max} = c_{i \max} (\log \log(T))^{-\frac{1}{2d}}$ , in which  $0 < \gamma_i < \frac{1}{3}$  and  $0 < c_{i \min}, c_{i \max} < \infty$  are constants.

Assumptions A.1–A.6 are quite reasonable in this kind of problem. A detailed justification is relegated to Gao, Gijbels and Van Bellegem (2006).

A.2. *A Technical Lemma.* The following lemma is necessary for the proof of the main results stated in Section 2.

Before stating the next lemma, we define the following notation. Let  $\xi_t = (e_t, X_t^T)$ . For



$i = 1, 2$  and  $j = 0, 1, 2$ ,

$$\begin{aligned}
l_{ji\pm} &= \prod_{r=1}^d \left[ \int_{-1}^1 u_r^j k_{i\pm}(u_r) du_r \right], \quad L_{i\pm} = l_{2i\pm} l_{0i\pm} - l_{1i\pm}^2, \\
W_{i\pm}(u) &= \frac{1}{L_{i\pm}} (l_{2i\pm} - l_{1i\pm} u) K_{i\pm}(u), \quad W_i(u) = W_{i+}(u) - W_{i-}(u), \\
a_i(X_s, X_t) &= a_i(X_s, X_t; h) = \frac{1}{Th_1^d} \int W_i \left( \frac{x - X_s}{h_i} \right) W_i \left( \frac{x - X_t}{h_i} \right) \pi(x) dx, \\
\phi_{st} &= \phi(\xi_s, \xi_t) = a_1(X_s, X_t) \epsilon_s \epsilon_t + a_2(X_s, X_t) \eta_s \eta_t, \\
N_{0T}(h) &= \sum_{s=1}^T \sum_{t=1}^T \phi_{st} \quad \text{and} \quad \tilde{L}_{wT}(h) = \frac{N_{0T}(h) - E[N_{0T}(h)]}{\sqrt{\text{Var}[N_{0T}(h)]}}, \tag{A.1}
\end{aligned}$$

where  $h = (h_1, h_2)$  is as defined before and the symbol “ $\int$ ” is used to denote either a univariate or multivariate integral. In the rest of this paper, we may use  $N_{0T}$  as the abbreviation of  $N_{0T}(h)$  when there is no notational confusion.

Recall the definition of  $M_{1T}(x)$  in (10) and the notation in (A.1). Define

$$\begin{aligned}
\bar{M}_{1T}(x) &= \frac{1}{Th_1^d} \sum_{t=1}^T \sum_{s=1}^T W_1 \left( \frac{x - X_s}{h_1} \right) W_1 \left( \frac{x - X_t}{h_1} \right) Y_s Y_t \\
&\quad + \frac{1}{Th_2^d} \sum_{t=1}^T \sum_{s=1}^T W_2 \left( \frac{x - X_s}{h_2} \right) W_2 \left( \frac{x - X_t}{h_2} \right) \epsilon_s^2 \epsilon_t^2, \\
\bar{N}_{1T} &= \int \bar{M}_{1T}(x) \pi(x) dx = \sum_{t=1}^T \sum_{s=1}^T a_1(X_s, X_t) Y_s Y_t + \sum_{t=1}^T \sum_{s=1}^T a_2(X_s, X_t) \epsilon_s^2 \epsilon_t^2, \\
\bar{L}_{wT}(h) &= \frac{\bar{N}_{1T} - E[\bar{N}_{1T}]}{\sqrt{\text{Var}[\bar{N}_{1T}]}}}, \\
\tilde{M}_{1T}(x) &= \frac{1}{Th_1^d} \left( \sum_{t=1}^T W_1 \left( \frac{x - X_t}{h_1} \right) \epsilon_t \right)^2 + \frac{1}{Th_2^d} \left( \sum_{t=1}^T W_2 \left( \frac{x - X_t}{h_1} \right) \eta_t \right)^2, \\
\tilde{M}_{2T}(x) &= \frac{1}{Th_1^d} \left( \sum_{t=1}^T W_1 \left( \frac{x - X_t}{h_1} \right) m_1(X_t) \right)^2 + \frac{1}{Th_2^d} \left( \sum_{t=1}^T W_2 \left( \frac{x - X_t}{h_2} \right) m_2(X_t) \right)^2, \\
\tilde{N}_{1T} &= \int \tilde{M}_{1T}(x) \pi(x) dx \quad \text{and} \quad \tilde{N}_{2T} = \int \tilde{M}_{2T}(x) \pi(x) dx. \tag{A.2}
\end{aligned}$$

Then we have the following lemma.

LEMMA A.1. (i) *Under Assumptions A.1–A.4, we have as  $T \rightarrow \infty$*

$$\max_{h \in H_T} \hat{L}_{wT}(h) = \max_{h \in H_T} \bar{L}_{wT}(h) + o_P(1). \tag{A.3}$$

(ii) *In addition, under  $\mathcal{H}_0$ ,*

$$\max_{h \in H_T} \bar{L}_{wT}(h) = \max_{h \in H_T} \tilde{L}_{wT}(h) + o_P(1). \tag{A.4}$$

(iii) Thus, we have under  $\mathcal{H}_0$

$$\max_{h \in H_T} \hat{L}_{wT}(h) = \max_{h \in H_T} \tilde{L}_{wT}(h) + o_P(1). \quad (\text{A.5})$$

PROOF: As the proof is quite technical, it is relegated to the proof of Lemma A.4 of Gao, Gijbels and Van Bellegem (2006).

### A.3. Proof of Theorem 2.1

PROOF OF THEOREM 2.1: To prove Theorem 2.1, we first show that as  $T \rightarrow \infty$

$$\frac{N_{0T}(h) - \mu}{\sigma_h} \rightarrow N(0, 1), \quad (\text{A.6})$$

where  $\mu = E[N_{0T}(h)]$  and  $\sigma_h^2 = \text{Var}[N_{0T}(h)] = Ch_1^d$  for some constant  $C > 0$ .

To apply Lemma A.1 of Gao and King (2004), let  $\xi_t = (\epsilon_t, \eta_t, X_t)$  and  $\phi(\xi_s, \xi_t) = \phi_{st}$  as defined in (A.1). Let  $M_T$  and  $N_T$  be defined as in the Lemma A.1. We now verify only the following condition listed in the Lemma A.2,

$$\frac{\max\{M_T, N_T\}}{\sigma_h^2} \rightarrow 0 \text{ as } T \rightarrow \infty \quad (\text{A.7})$$

for  $M_{T1}$ ,  $M_{T21}$ ,  $M_{T3}$ ,  $M_{T51}$ ,  $M_{T52}$  and  $M_{T6}$ . The others follow similarly.

For the  $M_T$  part, one justifies only

$$\frac{T^2 M_{T1}^{\frac{1}{1+\delta}}}{\sigma_h^2} \rightarrow 0 \text{ as } T \rightarrow \infty.$$

The others follow similarly.

Without any confusion in this proof, let  $a_{st} = a_1(X_s, X_t)$  and  $\psi_{st} = a_{st}\epsilon_s\epsilon_t$  throughout this proof. It follows that for some  $0 < \delta < 1$  and  $1 \leq i < j < k \leq T$

$$\begin{aligned} E \left[ |\psi_{ik}\psi_{jk}|^{1+\delta} \right] &= E \left[ |\epsilon_i\epsilon_j\epsilon_k^2 a_{ik}a_{jk}|^{1+\delta} \right] \\ &\leq \left\{ E \left[ |\epsilon_i\epsilon_j\epsilon_k^2|^{2(1+\delta)(1+\delta_2)} \right] \right\}^{\frac{1}{2(1+\delta_2)}} \left\{ E \left[ |a_{ij}a_{ik}|^{(1+\delta)(1+\delta_1)} \right] \right\}^{\frac{1}{(1+\delta_1)}} \\ &\leq C_\epsilon \left\{ E \left[ |a_{ij}a_{ik}|^{(1+\delta)(1+\delta_1)} \right] \right\}^{\frac{1}{(1+\delta_1)}}, \end{aligned} \quad (\text{A.8})$$

using Assumption A.1(iii), where  $C_\epsilon$  is a constant.

Since  $0 < \delta_1 < 1$  and  $0 < \delta_2 < 1$  satisfy  $\frac{1}{1+\delta_1} + \frac{1}{2(1+\delta_2)} = 1$  and  $\frac{1+\delta}{3-\delta} < \delta_1 < \frac{1-\delta}{1+\delta}$ , we have that

$$1 < \zeta_1 = (1 + \delta)(1 + \delta_2) < 2 \text{ and } 1 < \zeta_2 = (1 + \delta)(1 + \delta_1) < 2.$$

Similarly, we have

$$\begin{aligned}
a_{ik}a_{jk} &= \frac{1}{T^2 h_1^{2d}} \int \int W_1\left(\frac{x-X_i}{h_1}\right) W_1\left(\frac{x-X_k}{h_1}\right) W_1\left(\frac{y-X_j}{h_1}\right) W_1\left(\frac{y-X_k}{h_1}\right) \pi(x)\pi(y) dx dy \\
&= T^{-2} \int \int W_1(u) W_1\left(u + \frac{X_i-X_k}{h_1}\right) W_1(v) W_1\left(v + \frac{X_j-X_k}{h_1}\right) \\
&\quad \times \pi(X_i+uh)\pi(X_j+vh) dudv \equiv b_{ijk} + \delta_{ijk},
\end{aligned}$$

where  $b_{ijk} = T^{-2}\pi(X_i)\pi(X_j)L_1\left(\frac{X_i-X_k}{h_1}\right)L_1\left(\frac{X_j-X_k}{h_1}\right)$  and  $\delta_{ijk} = a_{ik}a_{jk} - b_{ijk}$  is the remainder term.

For convenience, we use  $\zeta = \zeta_2$ . For the given  $1 < \zeta < 2$  and  $T$  sufficiently large, we can show that

$$\begin{aligned}
M_{T11}^{(1+\delta_1)} &= E |a_{ik}a_{jk}|^\zeta = (1 + o(1))E |b_{ijk}|^\zeta \\
&= T^{-2\zeta} \int \int \int |\pi(u)\pi(v)|^\zeta \left| L_1\left(\frac{u-w}{h}\right) \right|^\zeta \left| L_1\left(\frac{v-w}{h}\right) \right|^\zeta f(u, v, w) dudvdw \\
&= T^{-2\zeta} h_1^{2d} \int \int \int |\pi(z+xh)\pi(z+yh)|^\zeta |L_1(x)L_1(y)|^\zeta f(z+xh, z+yh, z) dx dy dz \\
&= C_p T^{-2\zeta} h_1^{2d} (1 + o(1))
\end{aligned} \tag{A.9}$$

using Assumptions A.2–A.4, where  $C_p$  is a constant and  $f(u, v, w)$  is the joint density function of  $(X_1, X_{1+s_1}, X_{1+s_2})$ .

Thus, as  $T \rightarrow \infty$

$$\frac{T^2 M_{T11}^{1+\delta}}{\sigma_h^2} = C \frac{T^2 (T^{-2\zeta} h_1^{2d})^{1/\zeta}}{h_1^d} = h_1^{\frac{(2-\zeta)d}{\zeta}} \rightarrow 0. \tag{A.10}$$

Hence, (A.8)–(A.10) show that (A.7) holds for the first part of  $M_{T1}$ . The proof for the second part of  $M_{T1}$  follows similarly.

Similarly to (A.9), we have that as  $T \rightarrow \infty$

$$\begin{aligned}
M_{T3} &= E [a_{ik}^2 a_{jk}^2 \epsilon_i^2 \epsilon_j^2 \epsilon_k^4] = (Th_1)^{-4} h_1^{4d} (\mu_4 + 1) \\
&\quad \times E \left[ \pi^2(X_i) \pi^2(X_j) L_1^2\left(\frac{X_i-X_k}{h_1}\right) L_1^2\left(\frac{X_j-X_k}{h_1}\right) m_2(X_i) m_2(X_j) m_2^2(X_k) \right] \\
&= T^{-4} (\mu_4 + 1) \int \int \int \pi^2(x) \pi^2(y) L_1^2\left(\frac{x-z}{h}\right) L_1^2\left(\frac{y-z}{h}\right) m_2(x) m_2(y) m_2^2(z) \\
&\quad \times f(x, y, z) dx dy dz = T^{-4} h_1^{2d} (\mu_4 + 1) \\
&\quad \times \int \int \int \pi^2(uh+w) \pi^2(vh+w) L_1^2(u) L_1^2(v) m_2(uh+w) m_2(vh+w) m_2^2(w) \\
&\quad \times dudvdw = CT^{-4} h_1^{2d},
\end{aligned} \tag{A.11}$$

using Assumptions A.2–A.4, where  $\mu_4 = E[e_t^4] - 1$ .

This implies that as  $T \rightarrow \infty$

$$\frac{T^{3/2}M_{T3}^{\frac{1}{2}}}{\sigma_h^2} = C \frac{T^{3/2}T^{-2}h_1^d}{h_1^d} = CT^{-1/2} \rightarrow 0. \quad (\text{A.12})$$

Thus, (A.11) and (A.12) now show that (A.7) holds for  $M_{T3}$ . It follows from the structure of  $\{a_{ij}\}$  that (A.7) holds automatically for  $M_{T51}$ ,  $M_{T52}$  and  $M_{T6}$ , since  $E[\epsilon_i|X_i] = 0$ .

We now start to prove that (A.7) holds for  $M_{T21}$ . Similarly to (A.8), it follows that for some  $0 < \delta < 1$  and  $1 \leq i < j < k \leq T$

$$\begin{aligned} M_{T21} &= E \left[ |\psi_{ik}\psi_{jk}|^{2(1+\delta)} \right] = E \left[ |\epsilon_i\epsilon_j\epsilon_k^2 a_{ik}a_{jk}|^{2(1+\delta)} \right] \\ &\leq \left\{ E \left[ |\epsilon_i\epsilon_j\epsilon_k^2|^{2(1+\delta)(1+\delta_3)} \right] \right\}^{\frac{1}{1+\delta_3}} \left\{ E \left[ |a_{ij}a_{ik}|^{2(1+\delta)(1+\delta_4)} \right] \right\}^{\frac{1}{(1+\delta_4)}}, \end{aligned}$$

where  $0 < \delta_3 < 1$  and  $0 < \delta_4 < 1$  satisfy  $\frac{1}{1+\delta_3} + \frac{1}{1+\delta_4} = 1$ ,

$$1 < \zeta_3 = (1+\delta)(1+\delta_3) < 2 \quad \text{and} \quad 1 < \zeta_4 = (1+\delta)(1+\delta_4) < 2.$$

Analogously to (A.9) and (A.10), we obtain that as  $T \rightarrow \infty$

$$\frac{T^{3/2}M_{T21}^{\frac{1}{2(1+\delta)}}}{\sigma_h^2} = C \frac{T^{3/2}T^{-2} (h_1^{2d})^{1/(2\zeta_4)}}{h_1^d} = C \frac{1}{T^{1/2}h_1^{(1-\zeta_4^{-1})d}} \rightarrow 0$$

using the fact that  $\lim_{T \rightarrow \infty} Th_1^d = \infty$  and  $(1 - \zeta_4^{-1}) < \frac{1}{2}$ .

This finally completes the proof of (A.7) for  $M_{T21}$  and thus (A.7) holds for the first part of  $\{\phi_{st}\}$ . Similarly, one can show that (A.7) holds for the second part of  $\{\phi_{st}\}$ . Thus, we have shown that (A.6) holds under  $\mathcal{H}_0$ . The proof of Theorem 2.1 therefore follows from (A.6) and Lemma A.1(iii).

#### A.4 More Technical Lemmas

Before stating the necessary lemmas for the proof of the results given in Section 3, we introduce the following notation: Let  $H_T = H_{1T} \times H_{2T}$ ,  $J_T = J_{1T} \cdot J_{2T}$ ,

$$\begin{aligned} M_{1T}^*(x) &= \frac{1}{Th_1^d} \sum_{t=1}^T \sum_{s=1}^T W_1 \left( \frac{x - X_s}{h_1} \right) W_1 \left( \frac{x - X_t}{h_1} \right) Y_s^* Y_t^* \\ &\quad + \frac{1}{Th_2^d} \sum_{t=1}^T \sum_{s=1}^T W_2 \left( \frac{x - X_s}{h_2} \right) W_2 \left( \frac{x - X_t}{h_2} \right) \epsilon_s^{*2} \epsilon_t^{*2}, \end{aligned}$$

$$\begin{aligned} N_{1T}^* &= \int M_{1T}^*(x) \pi(x) dx \\ &= \sum_{t=1}^T \sum_{s=1}^T a_1(X_s, X_t) Y_s^* Y_t^* + \sum_{t=1}^T \sum_{s=1}^T a_2(X_s, X_t) \epsilon_s^{*2} \epsilon_t^{*2}, \end{aligned}$$

$$\begin{aligned}
N_{0T}^* &= \sum_{t=1}^T \sum_{s=1}^T a_1(X_s, X_t) \epsilon_s^* \epsilon_t^* + \sum_{t=1}^T \sum_{s=1}^T a_2(X_s, X_t) \eta_s^* \eta_t^*, \\
L_{1T}^*(h) &= \frac{N_{1T}^* - E[N_{1T}^*]}{\sqrt{\text{Var}[N_{1T}^*]}} \quad \text{and} \quad L_{0T}^*(h) = \frac{N_{0T}^* - E[N_{0T}^*]}{\sqrt{\text{Var}[N_{0T}^*]}}. \tag{A.13}
\end{aligned}$$

We also define  $L_{wT}^*(h)$  as the corresponding version of  $\hat{L}_{wT}(h)$  with  $\{Y_t : t \geq 1\}$  replaced by  $\{Y_t^* : t \geq 1\}$ . We now need the following lemmas to show Theorems 3.1–3.3.

LEMMA A.2. *Suppose that Assumptions A.1–A.6 hold. Then as  $T \rightarrow \infty$*

$$\max_{h \in H_T} L_{wT}^*(h) = \max_{h \in H_T} L_{1T}^*(h) + o_P(1), \tag{A.14}$$

$$\max_{h \in H_T} L_{1T}^*(h) = \max_{h \in H_T} L_{0T}^*(h) + o_P(1) \quad \text{under } \mathcal{H}_0. \tag{A.15}$$

PROOF: The proof of (A.14) is very similar to that of Lemma A.4(i) with the replacement of  $\{Y_t\}$  by  $\{Y_t^*\}$ . In this case, the detailed proof is simpler because  $\{\epsilon_t^*\}$  is now a sequence of independent and identically distributed random samples, and independent of  $\{X_s\}$  for all  $1 \leq s \leq T$ .

To prove (A.15), we need only to follow the proof of Lemma A.4(ii) by noting that under  $\mathcal{H}_0$ ,

$$E[N_{1T}^*] = E[N_{0T}] (1 + o(1)) \quad \text{and} \quad \text{Var}[N_{1T}^*] = \text{Var}[N_{0T}] (1 + o(1)). \tag{A.16}$$

The proof of Lemma A.2 is therefore completed.

LEMMA A.3. *Let Assumptions A.1–A.6 hold. Then  $\max_{h \in H_T} L_{wT}^*(h)$  and  $\max_{h \in H_T} \tilde{L}_{wT}(h)$  have identical asymptotic distributions under  $\mathcal{H}_0$ .*

PROOF: As the proof is extremely technical, it is relegated to the proof of Lemma A.6 of Gao, Gijbels and Van Bellegem (2006).

LEMMA A.4. *Suppose that Assumptions A.1–A.6 hold. Then for any  $x \geq 0$ ,  $h \in H_T$  and sufficiently large  $T$*

$$P(L_{0T}^*(h) > x) \leq \exp\left(-\frac{x^2}{4}\right).$$

Proof: As the proof is straightforward, it is available from the proof of Lemma A.7 of Gao, Gijbels and Van Bellegem (2006).

Before we present the next lemma, let us define, for  $0 < \alpha < 1$ ,  $l_\alpha^*$  to be the  $1 - \alpha$  quantile of  $\max_{h \in H_T} L_{0T}^*(h)$ .

LEMMA A.5. *Suppose that Assumptions A.1–A.6 hold. Then for large enough  $T$*

$$l_\alpha^* \leq 2\sqrt{\log(J_T) - \log(\alpha)}.$$

**Proof:** The proof is similar to that of Lemma 12 of Horowitz and Spokoiny (2001).

LEMMA A.6. *Suppose that Assumptions A.1–A.6 hold, and*

$$\lim_{n \rightarrow \infty} P \left( \frac{\tilde{N}_{2T}}{h_1^{d/2}} \geq C_0 \tilde{l}_\alpha^* \right) = 1 \quad (\text{A.17})$$

for some  $h \in H_T$ , where  $C_0$  is some positive constant and  $\tilde{l}_\alpha^* = \max \left( l_\alpha^*, \sqrt{2 \log(J_T)} + \sqrt{2 \log(J_T)} \right)$ .

Then the adaptive test statistic  $\hat{L}$  defined in (19) satisfies

$$\lim_{T \rightarrow \infty} P(\hat{L}_w > l_\alpha) = 1.$$

**Proof:** The proof is available from that of Lemma A.9 of Gao, Gijbels and Van Bellegem (2006).

#### A.5. Proofs of Theorems 3.1–3.3

**Proof of Theorem 3.1:** The proof follows directly from Lemmas A.1(iii), A.2 and A.3.

**Proof of Theorem 3.2:** It follows from the definition of  $\tilde{N}_{2T}$  in (A.2) that for sufficiently large  $T$ ,

$$\begin{aligned} \tilde{N}_{2T} &= (1 + o_P(1)) \int \cdots \int \rho(u) \pi(u) du \geq (1 + o_P(1)) \int_{\tau_1 - \epsilon}^{\tau_1 + \epsilon} \cdots \int_{\tau_d - \epsilon}^{\tau_d + \epsilon} \rho(u) \pi(u) du \\ &\geq (1 + o_P(1)) \min_{u \in \prod_{i=1}^d (\tau_i - \epsilon, \tau_i + \epsilon)} \rho(u) \cdot c_{\min}(\pi) \cdot 2^d \epsilon^d \end{aligned} \quad (\text{A.18})$$

for  $\epsilon = \epsilon_1$  or  $\epsilon_2$ . Equation (A.17) then follows from (21), (A.18) with  $\epsilon = \epsilon_1$  and the fact that both

$$\lim_{T \rightarrow \infty} \frac{1}{Th_1^d} \tilde{N}_{2T} > 0 \quad \text{and then} \quad \lim_{T \rightarrow \infty} \frac{\tilde{N}_{2T}}{h_1^{d/2}} = \infty \quad (\text{A.19})$$

hold in probability.

**Proof of Theorem 3.3:** Since equation (A.17) follows from (22), (A.18) with  $\epsilon = \epsilon_2$  and the choice of  $h_{1 \max} = c_{1 \max} \sqrt{(\log \log T)^{-\frac{1}{d}}}$ , the proof of Theorem 3.3 is completed.

#### Acknowledgments

The authors thank the Guest Editor and two referees for their helpful comments and suggestions on some closely related references. Thanks from the first author also go to the Australian Research Council for its financial support and The Institute of Statistics at Université catholique de Louvain for hosting part of his sabbatical leave, during which this paper was initialised. The second author gratefully acknowledges support from a grant from the National Science Foundation (FNRS), Belgium, as well as financial support from the Research Fund K.U.Leuven

(GOA/2007/4). The third author is Scientific Researcher at the ‘Fond National de Recherche Scientifique’ (FNRS), Belgium. This work was also supported by the IAP research network nr. P5/24 of the Belgian government.

## References

- Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.
- Andrews, D. W. K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.
- Carlstein, E., Müller, H.-G., David Siegmund, D., 1994. Change-point problems. Papers from the AMS-IMS-SIAM Summer Research Conference held at Mt. Holyoke College, South Hadley, MA, July 11–16, 1992. Edited by Institute of Mathematical Statistics Lecture Notes—Monograph Series, 23. Institute of Mathematical Statistics, Hayward, CA.
- Chen, C. W. S., Chiang, T. C., So, M. K. P., 2003. Asymmetrical reaction to US stock-return news: evidence from major stock market based on a double-threshold model, *Journal of Economics & Business* 55, 487–502.
- Chen, S., Gao, J., 2005. Simultaneous specification tests for the mean and variance structures of time series regression. Available from <http://www.maths.uwa.edu.au/~jiti/cg03.pdf>.
- D’Agostino, R. and Stephens, M.A. (1986). *Goodness-of-fit techniques*. Marcel Dekker, New York.
- Delgado, M. A., Hidalgo, J., 2000. Nonparametric inference on structural breaks. *Journal of Econometrics* 96, 113–144.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85, 645–660.
- Franke, J., Kreiss, J.-P., Mammen, E., 2002. Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* 8, 1–37.
- Gao, J., Gijbels, I., Van Bellegem, S., 2006. Simultaneous testing for structural breaks in nonparametric time series. Available from <http://www.maths.uwa.edu.au/~jiti/ggvb06.pdf>.
- Gao, J., King, M., 2004. Adaptive testing in continuous-time diffusion models. *Econometric Theory* 20, 844–883.
- Gijbels, I., Goderniaux, A.C., 2004. Bandwidth selection for changepoint estimation in nonparametric regression. *Technometrics* 46, 76–86.
- Gonzalez, R.C., Woods, R.E., 2002. *Digital Image Processing (2nd edition)*. Prentice Hall, New York.

- Grégoire, G., Hamrouni, Z., 2002. Two nonparametric tests for change-point problem. *Journal of Nonparametric Statistics* 14, 87–112.
- Hansen, B., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
- Hansen, B.E., 2000. Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.
- Hjellvik, V., Yao, Q., Tjøstheim, D., 1998. Linearity testing using local polynomial approximation. *Journal of Statistical Planning & Inference* 68, 295–321.
- Horowitz, J., Spokoiny, V., 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–632.
- Koop, G., Potter, S. M., 1999. Asymmetries in US unemployment. *Journal of Business & Economic Statistics* 17, 298–313.
- Lo, A.W., MacKinlay, C.A. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics* 45, 181–211.
- Loader, C.R., 1996. Change point estimation using nonparametric regression. *Annals of Statistics* 24, 1667–1678.
- Masry, E., Tjøstheim, D., 1995. Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* 11, 258–289.
- Müller, H.G., 1992. Change-points in nonparametric regression analysis. *Annals of Statistics* 20, 737–761.
- Müller, H.-G., Stadtmüller, U., 1999. Discontinuous versus smooth regression. *Annals of Statistics* 27, 299–337.
- Qiu, P., 1998. Discontinuous regression surfaces fitting. *Annals of Statistics* 26, 2218–2245.
- Qiu, P., 2005. *Image Processing and Jump Regression Analysis*. John Wiley & Sons, New York.
- Spokoiny, V., 1998. Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Annals of Statistics* 26, 1356–1378.
- Stephens, M.A. (1974). EDF statistics for goodness-of-fit and some comparisons. *Journal of the American Statistical Association* 69, 730–737.
- Wong, C. S., Li, W. K., 1997. Testing for threshold autoregression with conditional heteroscedasticity. *Biometrika* 84, 407–418.
- Wong, C. S., Li, W. K., 2000. Testing for double threshold autoregressive conditional heteroscedastic model. *Statistica Sinica* 10, 173–189.