Statistics and Its Interface Volume 1 (2008) 347–366

Testing structural change in time-series nonparametric regression models

Liangjun Su and Zhijie Xiao

We propose a CUSUM type of test for structural change in dynamic nonparametric regression models. It is based upon the cumulative sums of weighted residuals from a single nonparametric regression and complements the conventional parameter instability tests in parametric models. We derive the limiting distributions of the test under both the null hypothesis and sequences of local alternatives. A bootstrap procedure is also proposed and its validity is justified. Finally, simulation experiments are conducted to investigate the finite sample properties of our test.

AMS 2000 subject classifications: Primary 62G10, 62G20; secondary 62G08.
Keywords and phrases: CUSUM test, Structural change, Nonparametric regression, Strong mixing processes, Functional central limit theorem.

1. INTRODUCTION

Since Page (1954), a great deal of research attention has been devoted to the development of tests for parameter instability or, more generally, structural change in statistical models. The problem began with testing for a change in mean in i.i.d. samples, and moved naturally into the time series context. Today, time series with structural breaks are important models in econometrics as economic and financial time series are frequently affected by monetary policy and critical social events that may cause structural change. See, inter alia, Wichern et al. (1976), Perron (1989), Ploberger and Krämer (1992), Andrews (1993), Hidalgo (1995), Lavielle (1999), and Lee and Park (2001).

The greatest amount of research effort on testing for structural changes has been devoted to the parametric linear model:

$$y_t = x_t'\beta + u_t,$$

(or variants of this model) and focuses on the instability of parameters. Most tests are constructed based on a measure of fluctuation in the partial sum of residuals. A procedure that has played an important role in the study of structural change is the CUSUM test proposed by Brown et al. (1975), which is based on the maximum of partial sums of the recursive residuals. Krämer, Ploberger and Alt (1988) extended this test to linear regressions with lagged dependent variables, and Ploberger and Krämer (1992) studied CUSUM tests based on OLS (instead of recursive) residuals.

Linear parametric models provide a parsimonious way in characterizing the relationship among variables, but they also impose restrictions on the regression functional form. From this point of view, nonparametric models allow for a larger class of regression functions and have certain advantages in a variety of applications (see, e.g., Müller, 1992). For this reason, there have been some recent studies on structural change based on nonparametric regression models. Until now, most of the existing tests for structural change in nonparametric regression have focused on sudden, localized changes of a regression function that may not, in fact, be associated with time at all. In particular, Müller (1992) provided a central limit theorem for the estimators of the location and size of the change point, whereas Chu and Wu (1993) proposed a test for the number of jumps in a regression model with fixed design. Loader (1996) studied structural change in a simple nonparametric regression model with fixed design

$$y_i = m(x_i) + u_i, \quad i = 1, 2, \ldots, n,$$

with $x_i = i/n$ and u_i being i.i.d. $N(0,1)$. In the case where u_i is covariance stationary, Kim and Hart (1998) developed an omnibus test for the null hypothesis that the underlying mean is constant for the above model. They considered the alternative where the mean function depends on the design point but keeps fixed over time. Delgado and Hidalgo (2000) proposed estimators of location and size of structural breaks in nonparametric regression models when the regressors are strictly stationary and when lagged dependent variables are present and the break is explained by the regressor "time". Recently, Chen et al. (2005) proposed a hybrid test and estimation procedure for change points in volatility based on the least squares method in nonparametric time series models where there is a scale change in the volatility function at a certain time.

In this paper, we study testing for structural change in time series nonparametric regression models. We propose a CUSUM type of test for structural changes in the regression function in time series framework. As was done in much of the previous literature, the time of the structural change(s) is not specified a priori. The test is based on the cumulative sums of weighted residuals from nonparametric regressions, and has asymptotic pivotal null distribution. In addition, as a companion to the asymptotic test, we propose a bootstrap version of our test to achieve finite sample improvement. The asymptotic validity of the bootstrap test is justified.

There are several important features that distinguish our tests from the existing literature. First, our test is nonparametric. As Kim and Hart (1998) remarked, parametric methods place restrictions on what the data can tell whereas nonparametric techniques rely more on the data set itself. Parametric tests are powerful against certain types of alternatives and perform well in cases of correct specification; but they can also provide misleading conclusions in the case of misspecification. Smoothing-based nonparametric tests work well for a wide class of alternatives and yield good power in a variety of circumstances. As a trade-off, it is well known that nonparametric tests are subject to the curse of dimensionality and usually only have non-trivial power against local alternatives that converge to zero at a rate slower than the parametric $n^{-1/2}$-rate. Fortunately, the proposed CUSUM test in this paper does not suffer much from this problem and has non-trivial power against local alternatives that converge to zero at the parametric $n^{-1/2}$-rate. This is largely due to the effect of averaging because averaging reduces the variance of the original nonparametric estimates, and by choosing the bandwidth appropriately the bias of the nonparametric estimates can be well controlled. Second, our test allows for weak dependence in the data. As a result of using nonparametric regression estimates, the limiting null distributions of our test statistics are free of nuisance parameters. All have limiting distributions associated with the standard Brownian bridge. The asymptotic critical values of our test can easily be tabulated. We demonstrate through simulations that our test works fairly well in finite samples for a wide range of data generating processes. Third, we relax the stationarity of the underlying process, which is often assumed in parametric tests for structural changes. We focus on testing for structural change in the conditional mean process, and allow for flexibility in other aspects of the time series under both the null and alternatives. For example, the conditional variance process can exhibit structural changes under the null and/or alternative.

The rest of the paper is organized as follows. In Section 2, we introduce our hypotheses and test statistics. The asymptotic properties of our test statistics are studied in Section 3. Section 4 provides a bootstrap version of the test. In Section 5, we report the results of Monte Carlo simulations. All proofs are relegated to the appendix.

2. MODEL, HYPOTHESES AND TEST STATISTICS

2.1 The model and hypotheses

Consider the following nonparametric regression model:

$$Y_t = m_t(X_t) + U_t, \quad t = 1, 2, \ldots, n, \tag{2.1}$$

where Y_t is the dependent variable, X_t is a $\mathbb{R}^d$-valued regressor, $m_t(\cdot)$ is an unknown but smooth function, U_t is the random disturbance term satisfying $E(U_t|X_t) = 0$ a.s., and $E(U_t^2|X_t) = \sigma_t^2(X_t)$. We assume that $\{X_t, U_t\}$ is a strong mixing process, but we don't require stationarity. Note that both the regression mean function $m_t(\cdot)$ and the conditional variance function $\sigma_t^2(\cdot)$ may be time-varying.

We are interested in testing whether the conditional mean function is stable over time. Namely, our null hypothesis is

$$H_0 : m_t(X_t) = m(X_t) \quad \text{a.s. for all } t = 1, \ldots, n, \tag{2.2}$$

where $m(\cdot)$ is a smooth function that is not time-dependent. In this case, we will say that there is no structural change or break in the conditional mean process.

The alternative hypothesis can be specified in various ways. The following two types of alternatives are widely used in the literature, and we consider both of them in this paper. The first one is

$$\begin{aligned} H_{1A} : m_t(X_t) &= m(X_t) \quad \text{a.s. for all } t = 1, \ldots, k_0, \\ m_t(X_t) &= m(X_t) + \Delta_n(X_t) \\ &\qquad \text{a.s. for all } t = k_0 + 1, \ldots, n, \end{aligned} \tag{2.3}$$

where $\Delta_n(\cdot)$ is a nonzero function that may depend on the sample size n but not on the time t, and k_0 is an unknown break point. That is, at time $k_0 + 1$ we have a structural change of the conditional mean function. The second one is

$$H_{1B} : m_t(X_t) = m(X_t) + g_n(t/n) \quad \text{a.s. for all } t = 1, \ldots, n, \tag{2.4}$$

where $g_n(\cdot)$ is an arbitrary non-constant function defined on the $[0,1]$ interval. Note that we allow both $\Delta_n(\cdot)$ and $g_n(\cdot)$ to depend on n to facilitate the study of local power properties of our tests.

It is worth mentioning that the above hypotheses do not impose any additional restrictions on the conditional variance process $\left\{\sigma_t^2(X_t)\right\}$, or other aspects of the conditional distribution of Y_t given X_t, or the marginal distribution of X_t. As a matter of fact, we allow for time varying behavior in the conditional variance process and nonstationary distribution of $\{X_t, Y_t\}$ under both the null and alternative hypotheses.

2.2 Test statistics

To proceed, we introduce some notation. First, let $k_0 = \lceil ns_0 \rfloor$ for some $0 < s_0 < 1$, where $\lceil c \rfloor$ denotes the largest integer less than or equal to c. We will call k_0 as the break point and s_0 as the break ratio under H_{1A}.

Next, let $f_t(x)$ denote the marginal density function of X_t evaluated at $x \in \mathbb{R}^d$. Define

$$\overline{f}_n(x) = n^{-1} \sum_{t=1}^{n} f_t(x) \quad \text{and} \quad \overline{f}(x) = \lim_{n \to \infty} \overline{f}_n(x), \tag{2.5}$$

where $\overline{f}(x)$ is regarded as the long run average density of $\{X_1, \ldots, X_n\}$. We estimate $\overline{f}_n(x)$ or $\overline{f}(x)$ by

$$f_{n,h}(x) = n^{-1} \sum_{i=1}^{n} K_h (x - X_i), \tag{2.6}$$

where $K_h(\cdot) = h^{-d} K(\cdot / h)$, $K(\cdot)$ is a symmetric kernel function and $h \equiv h_n$ is a bandwidth parameter. There has been much study on asymptotic properties of kernel smoothers for dependent data under various conditions, including Robinson (1983), Roussas (1988), Liebscher (1996), Bosq (1996), Fan and Yao (2003), and Hansen (2008). Most of these papers deal with stationary processes. A few exceptions include Bosq (1991), Andrews (1995), Sun and Chiang (1997), and Shen and Huang (1998).

Under weak conditions, Andrews (1995) showed $f_{n,h}(x) \xrightarrow{p} \overline{f}(x)$. The Nadaraya-Watson (NW) kernel estimator of $m(x)$ under the null hypothesis is given by

$$m_{n,h}(x) = n^{-1} \sum_{i=1}^{n} Y_i K_h (x - X_i) / f_{n,h}(x). \tag{2.7}$$

It is worth mentioning that $m_{n,h}(x)$ also converges to some non-stochastic object under the alternative provided suitable conditions are met.

Define

$$\widehat{U}_t = Y_t - m_{n,h}(X_t), \quad \text{and} \quad \widehat{V}_t = \widehat{U}_t f_{n,h}(X_t) w(X_t), \tag{2.8}$$

where $w(\cdot)$ is a weight function that has crucial effect on the power of the test. Define the cumulative sums of the weighted residuals $\widehat{V}_t$ as

$$\Gamma_n(s) = \frac{1}{\sqrt{n}\widehat{\sigma}} \sum_{i=1}^{\lceil ns \rfloor} \widehat{V}_i, \tag{2.9}$$

where $\widehat{\sigma} = \{n^{-1} \sum_{i=1}^{n} \widehat{V}_i^2\}^{1/2}$. Assume $\sigma_0^2 \equiv p\lim_{n\to\infty} n^{-1} \times \sum_{i=1}^{n} U_i^2 \overline{f}^2(X_i) w^2(X_i)$ exists. It is easy to show that $\widehat{\sigma} \xrightarrow{p} \sigma_0$ under certain conditions, where σ_0 is the positive square root of σ_0^2.

We shall show that $\Gamma_n(\cdot)$ converges weakly to the standard Brownian bridge $W^0(\cdot)$. Let $l(\cdot)$ be a continuous functional that measures the fluctuation of $\Gamma_n(s)$ around zero. By the continuous mapping theorem,

$$l(\Gamma_n(\cdot)) \xrightarrow{d} l\left(W^0(\cdot)\right), \tag{2.10}$$

where $\xrightarrow{d}$ denotes convergence in distribution. In principle, the choice of l is rich. When we take the classical Kolmogorov-Smirnoff and Cramer-von Mises measures, we obtain the following test statistics

$$KS_n = \sup_{0\le s\le 1} |\Gamma_n(s)| = \max_{1\le j\le n} \left| \frac{1}{\sqrt{n}\widehat{\sigma}} \sum_{i=1}^{j} \widehat{V}_i \right|, \quad \text{and} \tag{2.11}$$

$$CM_n = \int_0^1 \Gamma_n(s)^2 \, ds = \frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{\sqrt{n}\widehat{\sigma}} \sum_{i=1}^{j} \widehat{V}_i \right)^2. \tag{2.12}$$

We will study the limiting distributions of KS_n and CM_n below.

3. ASYMPTOTIC PROPERTIES OF THE TEST

In this section, we study the asymptotic properties of the test under the null hypothesis and sequences of local alternatives.

3.1 Assumptions

For the purpose of asymptotic analysis, we make the following assumptions.

Assumptions

A1. $\{X_t, Y_t\}$ is a strong mixing process with mixing coefficients $\alpha(\tau)$ such that $\sum_{t=1}^{\infty} t^3 \alpha(t)^{\delta/(4+\delta)} \le C < \infty$ for some $\delta > 0$ with $\delta/(4+\delta) \le 1/2$.

A2. $E(U_t | X_t, \ldots, X_1, U_{t-1}, \ldots, U_1) = 0$ and $E(|U_t^{4+\delta}| | X_t) \le c_t(X_t)$ a.s. such that $c_t(X_t)$ is continuous and $E[c_t(X_t) |w(X_t)|^{4+\delta}] < \infty$.

A3. The marginal density $f_t(\cdot)$ of X_t is bounded on its support $\mathcal{X}$. For each $t_1 < \ldots < t_l$, the joint density $f_{t_1,\ldots,t_l}(\cdot)$ of $(X_{t_1}, \ldots, X_{t_l})$ exists and is uniformly bounded on its support, where $l = 2, \ldots, 8$. In addition, the long-run average density $\overline{f}(x) = \lim_{n\to\infty} \overline{f}_n(x)$ exists.

A4. For each $t = 1, 2, \ldots$, $f_t(\cdot) \in \mathcal{G}_r^{\infty}$ and $m(\cdot) \in \mathcal{G}_r^{4+\delta}$ for some integer $r \ge 2$, where $\mathcal{G}_r^{\gamma}$ is a class of functions defined in Definition D.3 in Appendix D.

A5. The kernel function K is symmetric and $K \in \mathcal{K}_r$, where $\mathcal{K}_r$ is defined in Definition D.2 in Appendix D. $\int \left| K^{4+\delta}(u) \right| du < \infty$.

A6. As $n \to \infty$, $nh^{2d} \to \infty$ and $nh^{2r} \to 0$, where r is the same r used in Definition D.2.

Assumption A1 specifies that the serial dependence in the data is strong mixing and it implies that $\alpha(t) = o(t^{-(4+16/\delta)})$. The smaller δ, the faster rate at which $\alpha(t)$ decays to zero. Together with Assumptions A2 and A4, this reflects the trade-off between the degree of dependence and moments of $\{Y_t\}$. Note that Assumption A1 does not require strict stationarity of the process $\{X_t, Y_t\}$. This is important since we allow X_t to include lagged dependent variables. Assumption A2 is typical in time series regressions, it can be relaxed to allow correlation in the error terms at the expense of a more complicated proof. The smoothness condition in Assumptions A3–A4 and the assumptions on the kernel and bandwidth in Assumptions A5–A6 are comparable to the typical assumptions in the nonparametric literature (e.g., Li, 1999). In particular, r represents the order of the kernel K. Nevertheless, we allow the conditional variance function $\sigma_t^2(\cdot)$ and the marginal density $f_t(\cdot)$ of X_t to vary over time. Assumption A6 implies that $r > d$ which is not as restrictive as it appears because of the curse of dimensionality in the nonparametric literature. Nevertheless, it is possible to relax the assumption "$nh^{2d} \to \infty$" to "$nh^d/\log n \to \infty$" with a stronger assumption on the mixing coefficient.

3.2 Asymptotic null distribution

Theorem 3.1. *Under Assumptions A1–A6, and suppose that the weighting function $w(\cdot)$ is uniformly continuous and bounded, then under H_0,*

$$\Gamma_n(\cdot) \Rightarrow W^0(\cdot), \quad \text{as } n \to \infty,$$

where $W^0(\cdot)$ denotes the standard Brownian bridge and $\Rightarrow$ denotes weak convergence in the space $D([0,1])$ with respect to the Skorohod J_1-topology (see Pollard, 1984).

Asymptotic distributions of the Kolmogorov-Smirnoff test (2.11) and the Cramer-von Mises test (2.12) can be immediately obtained by the continuous mapping theorem and Theorem 3.1:

$$KS_n = \sup_{0\le s\le 1} \Gamma_n(s) \xrightarrow{d} \sup_{0\le s\le 1} \left|W^0(s)\right|, \quad \text{and}$$

$$CM_n = \int_0^1 \Gamma_n(s)^2\, ds \xrightarrow{d} \int_0^1 \left|W^0(s)\right|^2 ds.$$

The limiting distributions of the KS_n and CM_n test statistics have the classical Kolmogoroff-Smirnoff and Cramer-von Mises forms, respectively. The critical values for these tests can be easily tabulated via simulations. Alternatively, one can consult Anderson and Darling (1952) for the critical values of the CM_n test statistic (0.3473, 0.4614, and 0.7435 for the 10%, 5% and 1% tests, respectively), and Xiao (1999) for the critical values of the KS_n test statistic (1.128, 1.262, and 1.521 for the 10%, 5% and 1% tests, respectively).

3.3 Asymptotic local power

Now we study the local power of the CUSUM test that is built upon $\Gamma_n(\cdot)$. As mentioned above, once we deviate from the null, several cases can arise. We focus on two scenarios that are most popular in the literature.

We first study the local alternative:

(3.1)
$$H_{1A,n}: E(Y_t|X_t = x) = m(x) + n^{-1/2}\Delta(x)\,1(t \ge k_0 + 1),$$

where $1(\cdot)$ is the indicator function, and $\Delta(x)$ is an arbitrary non-zero function on $\mathcal{X}$. Then we study the second type of local alternative:

$$H_{1B,n}: E(Y_t|X_t = x) = m(x) + n^{-1/2} g(t/n), \tag{3.2}$$

where $g(\cdot)$ is an arbitrary non-constant function defined on the $[0,1]$ interval. When $g(s) = 0$ for $s \le s_0$ and $g(s) = c \ne 0$ for $s > s_0$, (3.2) includes a one-time level shift of the regression function at time $k_0 = s_0 n$ as a special case. This is analogous to the one-time scale change of volatility function in Chen et al. (2005).

We impose the following assumption concerning the alternatives.

Assumption A7. For $\delta(X_t) = \Delta(X_t)$ or 1, $E|\delta(X_t) \times w(X_t)\overline{f}(X_t)|^{1+\epsilon} < \infty$ for some $\epsilon > 0$ and for all t.

Together with Assumption A1, Assumption A7 implies that $n^{-1}\sum_{t=1}^n \delta(X_t) w(X_t)\overline{f}(X_t) - n^{-1}\sum_{t=1}^n E\big[\delta(X_t) \times w(X_t)\overline{f}(X_t)\big] \xrightarrow{a.s.} 0$. See Corollary 3.48 in White (2001).

Theorem 3.2. *Suppose Assumptions A1—A7 hold.*

(i) Under the local alternative $H_{1A,n}$, $\Gamma_n(\cdot) \Rightarrow G_A(\cdot)$, where $G_A(s) = W^0(s) - s(1-s_0)(\mu_1/\sigma)$, and $\mu_1 \equiv \lim_{n\to\infty} n^{-1}\sum_{t=1}^n E\left[\Delta(X_t) w(X_t)\overline{f}(X_t)\right]$.

(ii) Under the local alternative $H_{1B,n}$, $\Gamma_n(\cdot) \Rightarrow G_B(\cdot)$, where $G_B(s) = W^0(s) + (\mu_2/\sigma)(\int_0^s g(v)\,dv - s\int_0^1 g(v)\,dv)$, and $\mu_2 \equiv \lim_{n\to\infty} n^{-1}\sum_{t=1}^n E\left[w(X_t)\overline{f}(X_t)\right]$.

We make some remarks. First, the above theorem says that the KS_n and CM_n tests have non-trivial power in detecting $n^{-1/2}$ local alternatives provided $\mu_1 \ne 0$ or $\mu_2 \ne 0$. Second, the assumption on the weight function $w(x)$ is weak. It does not exclude the case where a weight function only focuses on a certain range of the data. In this case, the test only has power in detecting deviations from the null on the restricted range. Third, the choice of weight function $w(x)$ has important effects on the local power of the tests. For clarity, we now look at the simplest case where the process $\{X_t\}$ is strictly stationary and the local alternative is of the type specified in $H_{1A,n}$. In this case, $\mu_1 = E[\Delta(X_t) w(X_t) f(X_t)]$, where $f(\cdot)$ is the marginal density of X_t. So if $E[\Delta(X_t) f(X_t)] = 0$, say, in the case where X_t is symmetrically distributed around zero and $\Delta(\cdot)$ is an odd function, we need to choose w such that it cannot be an even function in order to detect these kinds of local alternatives. In contrast, for the type of local alternatives specified in $H_{1B,n}$, it suffices to choose $w(x) \equiv 1$. The effect of different choices of w will be studied in our Monte Carlo simulations. Fourth, the fact that the CUSUM test is not consistent against $H_{1A,n}$ if $\mu_1 = 0$ is a nonparametric analog of the parametric case. In the parametric setup, if all structural shifts in the finite dimensional parameters are orthogonal to the average of regressors or the regressors themselves, then the CUSUM test is not consistent. See Ploberger and Krämer (1992, 1996). Fifth, if the local alternative $\Delta(x)$ were known in Theorem 3.2(i), we could derive the optimal choice of weight function given by $w^*(x) = \Delta(x)\overline{f}(x)$ in terms of maximizing the local power. Nevertheless $\Delta(x)$ is typically unknown in practice, so this optimal choice is infeasible. Theoretically, we can follow Andrews (1993) and make some distributional assumption on $\Delta(X_t)$ and maximize certain weighted average of local power. But this is beyond the scope of this paper.

4. A BOOTSTRAP TEST

It is well known that an asymptotic-distribution-based nonparametric test may perform poorly in finite samples. An important alternative approach is to use bootstrap approximation for the distribution of the test statistic. In this section, we propose a bootstrap version of the test discussed above. We stress the fact that the theorems obtained in this paper are based on asymptotic considerations. As Neumann and Paparoditis (2000) noted, in order to get an asymptotically correct estimator of the null distribution of the test statistics, it is not necessary to reproduce the whole dependence structure of the stochastic processes generating the original observations. On the other hand, it is important to impose the null in the resampling scheme. Simple resampling from the empirical distribution of $W_t \equiv (Y_t, X_t')'$ will not impose the null restriction.

The wild bootstrap proposed by Wu (1986) and Liu (1988) is designed to allow heteroskedasticity in the linear regression models. It has been examined in the time series context by Kreiss (1997), and Hafner and Herwartz (2000), among others. Before drawing the bootstrap resamples, we re-center the residuals $\{\widehat{U}_t\}$ to ensure that its sample mean is zero, i.e., we replace $\widehat{U}_t$ by $\overline{\widehat{U}}_t = \widehat{U}_t - \overline{\widehat{U}}$, where $\overline{\widehat{U}} = n^{-1}\sum_{t=1}^{n} \widehat{U}_t$. We then obtain the wild bootstrap residuals by

$$U_t^* = \overline{\widehat{U}}_t \eta_t,$$

where $\{\eta_t\}$ are i.i.d., independent of the process $\{X_t, Y_t\}$, and satisfy the conditions: $E(\eta_t) = 0$, $E(\eta_t^2) = 1$, and $E(\eta_t^4) < \infty$. There are many ways to obtain such a sequence $\{\eta_t\}$. In our simulation, we draw them independently from a distribution with masses $c = (1+\sqrt{5})/(2\sqrt{5})$ and $1-c$ at the points $(1-\sqrt{5})/2$ and $(1+\sqrt{5})/2$, respectively. Consequently, the wild bootstrap draws each U_t^* from a different distribution with mean zero and variance $\overline{\widehat{U}}_t^2$ conditional on the data. We generate the bootstrap resample $\{Y_t^*, X_t^*\}_{t=1}^n$ by[1]

$$Y_t^* = m_{n,h_0}(X_t^*) + U_t^*,$$

where $X_t^* = X_t$. Note that the bandwidth sequence $h_0 \equiv h_{0n}$ used here is different from the bandwidth sequence h that is used to construct the test statistics KS_n and CM_n. See Härdle and Marron (1991) for the explanation why we need different bandwidth choices here.

Based upon the bootstrap resampling data $\{Y_t^*, X_t^*\}_{t=1}^n$, we calculate the bootstrap analogue of $m_{n,h}(x)$ and $\widehat{U}_t$ by

$$m_{n,h}^*(x) = n^{-1}\sum_{i=1}^{n} Y_i^* K_h(x - X_i)/f_{n,h}(x) \quad \text{and}$$
$$\widehat{U}_t^* = Y_t^* - m_{n,h}^*(X_t).$$

Then we can construct the bootstrap version $\{\Gamma_n^*(s)\}$ of the process $\{\Gamma_n(s)\}$:

$$\Gamma_n^*(s) = \frac{1}{\sqrt{n}\widehat{\sigma}^*}\sum_{i=1}^{\lceil ns \rfloor} \widehat{V}_i^*, \tag{4.1}$$

where $\widehat{V}_i^* = \widehat{U}_i^* f_{n,h}(X_i)\, w(X_i)$ and $\widehat{\sigma}^* = \{n^{-1}\sum_{i=1}^n \widehat{V}_i^{*2}\}^{1/2}$. Based upon $\Gamma_n^*(s)$, we can construct the bootstrap version KS_n^* of the test statistic KS_n (similarly for CM_n). We repeat this procedure B times and obtain the sequence $\{KS_{n,j}^*\}_{j=1}^B$. We reject the null when $p^* = B^{-1} \times \sum_{j=1}^B 1(KS_n \leq KS_{n,j}^*)$ is smaller than the given level of significance.

For the validity of the bootstrap method, we need to make some additional assumptions. The following notation is used. Let $\kappa = (\kappa_1, \ldots, \kappa_d)'$ and $\lambda = (\lambda_1, \ldots, \lambda_d)'$ denote d-vector of nonnegative integer constants. For such vectors, define (i) $|\kappa| = \sum_{j=1}^d \kappa_j$, (ii) $\kappa \leq \lambda$ iff $\kappa_j \leq \lambda_j\ \forall j = 1, \ldots, d$, (iii) $\kappa < \lambda$ iff $\kappa \leq \lambda$ and $\kappa_j < \lambda_j$ for some j, (iv) for any function $c(z)$ on $\mathbb{R}^d$, $D^\kappa c(z) = \partial^{|\kappa|}/(\partial z_1^{\kappa_1}, \ldots, \partial z_d^{\kappa_d})(c(z))$, where $z = (z_1, \ldots, z_d)'$, and (v) $z^\kappa = \Pi_{j=1}^d z_j^{\kappa_j}$.

Assumption A8. (i) $\sup_x E(|Y_t|^{4+\delta}\,|X_t = x) f_t(x) \leq \overline{b}_1 < \infty$ for each t, and $\sup_x \|x\|^q E(|Y_t|\,|X_t = x) f_t(x) \leq \overline{b}_2 < \infty$ for each t and some $q \geq d$. There is some $t^* < \infty$ such that for all $t \geq t^* > 1$, $\sup_{x_1, x_t} E(|Y_1 Y_t|\,|X_1 = x_1, X_t = x_t) f_{1t}(x_1, x_t) \leq \overline{b}_3 < \infty$, where $f_{1t}(x_1, x_t)$ denotes the joint density of (X_1, X_t). (ii) Let $\beta = 4 + 16/\delta$. For some $\theta \in (0,1)$, we have $\log n/(n^\theta h_0^d) = o(1)$, and

$$\frac{d}{q} + 3 + 2\theta - \frac{1-\theta}{2}\left(\frac{(2\beta+3)(\delta+2)}{\delta+3} - 2d\right) \leq 0. \tag{4.2}$$

Assumption A9. (i) $f_t(x)$ is continuously differentiable to integral order r on $\mathbb{R}^d$, and $\sup_{n\geq 1}\sup_{x\in\mathbb{R}^d} |D^\kappa \overline{f}_n(x)| < \infty\ \forall\kappa$ with $|\kappa| \leq r$. (ii) $\overline{m}_n(x) \equiv n^{-1}\sum_{t=1}^n m_t(x)$ is continuously differentiable to integral order r on $\mathbb{R}^d$, and $\sup_{n\geq 1}\sup_{x\in\mathbb{R}^d} |D^\kappa[\overline{f}_n(x)\,\overline{m}_n(x)]| < \infty\ \forall\kappa$ with $|\kappa| \leq r$.

Assumption A10. The kernel function $K : \mathbb{R}^d \to \mathbb{R}$ has a compact support $\mathcal{U}$. For any $|\lambda| \leq r$, $\sup_{u\in\mathcal{U}} |D^\lambda K(u)| \leq \overline{c}_1 < \infty$, and $D^\lambda K(u) = 0$ for $\|u\| \geq \overline{c}_2$, and $\|D^\lambda K(u) - D^\lambda K(u')\| \leq \overline{c}_3\|u - u'\|$ for any $u, u' \in \mathbb{R}^d$ and some $\overline{c}_3 < \infty$.

Assumption A8 (i) controls the tail behavior of the conditional expectations $E(|Y_t|^{4+\delta}\,|X_t = x)$, $E(|Y_t|\,|X_t = x)$, and $E(|Y_1 Y_t|\,|X_1 = x_1, X_t = x_t)$. For example, the first one can increase to infinity but at a rate slower than $f_t^{-1}(x)$. Assumption A8 (ii) reflects the trade-off between the mixing coefficient, moments of the process $\{X_t, Y_t\}$, and the bandwidth h_0. For fixed $\theta \in (0,1)$ and $q \geq d$, (4.2) can easily be satisfied by requiring sufficiently small δ. Assumptions A9 and A10 are needed for the rth derivative of $m_{n,h_0}(x)$ to be well behaved. In particular, we allow $m_t(x)$ to depend on t

[1] Note that even if X_t includes lagged dependent variables, say $X_t = Y_{t-1}$, we can generate the bootstrap data $\{Y_t^*, X_t^*\}$ in this way because we don't need to mimic the dependence structure of the process $\{Y_t\}$ by that of $\{Y_t^*\}$. We gratefully thank a referee for making this point to us.

in the proof of the following theorem. But we can relax the compact support of K at the cost of lengthier arguments (see Hansen, 2008).

Theorem 4.1. *Suppose Assumptions A1–A6 and A8–A10 hold. Suppose the weight function $w(x)$ is uniformly continuous and bounded and has support on the set $S_n \equiv \{x \in \mathbb{R}^d : n^{-1}\sum_{t=1}^n f_t(x) \geq d_n\}$ where $d_n = o(1)$. Suppose $nh_0^{d+2r} \to C \in (0,\infty]$, and $nh^{2r}d_n^{-2(1+r)}\log n \to 0$ as $n \to \infty$. Then*

$$\Gamma_n^*(\cdot) \overset{p}{\Rightarrow} W^0(\cdot),$$

where $\overset{p}{\Rightarrow}$ denotes weak convergence in probability as defined by Giné and Zinn (1990).

Note that we have restricted ourselves to the class of weight functions that have support on S_n. We do so because the derivatives of $m_{n,h_0}(x)$ are not well behaved if $n^{-1}\sum_{t=1}^n f_t(x)$ is too small. In practice, we can consider choosing $d_n = O(1/\log n)$ and we find through simulations in the next section that the bootstrap tests are not sensitive to this choice.

Theorem 4.1 shows that the bootstrapped process $\{\Gamma_n^*(\cdot)\}$ also converges weakly to the standard Brownian bridge and thus provides an asymptotic valid approximation to the limit null distribution of the test statistics KS_n and CM_n that are constructed from $\{\Gamma_n(\cdot)\}$. This holds as long as we generate the bootstrap data by imposing the null hypothesis.

It is well known that the optimal bandwidth in minimizing the integrated mean squared errors (IMSE) of estimators $m_{n,h_0}(x)$ of $\overline{m}(x)$ is proportional to $n^{-1/(d+2r)}$. Clearly, we can choose h_0 by the least squares cross validation. Since undersmoothing is required for the estimate $m_{n,h}^*(x)$, we propose a rule of thumb to choose h according to the optimal choice of h_0 in our Monte Carlo – see Section 5 for more discussion on this issue.

5. MONTE CARLO SIMULATIONS

In this section we conduct a small set of Monte Carlo simulations to evaluate the finite sample performance of our tests. We first focus on their finite sample performance under the null and then examine their power properties.

5.1 Finite sample performance of the tests under the null

To examine the finite sample performance of the tests under the null, we generate data from the following data generating processes (DGPs).

DGP s1: $Y_t = X_t^2 + U_t$, where X_t and U_t are i.i.d. $N(0,1)$ and mutually independent.

DGP s2: $Y_t = 0.6Y_{t-1} + \sqrt{0.1 + 0.6Y_{t-1}^2}U_t$, where U_t are i.i.d. $N(0,1)$.

DGP s3: $Y_t = \Phi(Y_{t-1}) + \sqrt{0.1 + 0.4Y_{t-1}^2}\varepsilon_t$ for $t \leq 0.5n$, and $Y_t = \Phi(Y_{t-1}) + \sqrt{0.1 + 0.8Y_{t-1}^2}\varepsilon_t$ for $t > 0.5n$, where $\Phi(\cdot)$ is the standard normal's cumulative distribution function, and ε_t are i.i.d. $N(0,1)$.

DGP s1 specifies an i.i.d. sequence $\{Y_t, X_t\}$. DGP s2 specifies a typical stationary AR(1)-ARCH(1) process $\{Y_t\}$. DGP s3 yields a nonstationary process that has a structural change in the conditional variance but not in the conditional mean. To conduct our test for DGPs s2–s3, we set $X_t = Y_{t-1}$ and throw away the first 500 observations.

To construct the test statistics, let $1_x = 1(f_{n,h}(x) > 0.001/\log n)$. We choose the weighting function to be $w(x) = (\sin(x) + \cos(x))1_x$ unless otherwise specified. Notice that 1_x is required for the bootstrap test to avoid a random denominator issue. For the kernel, we choose the fourth order Epanechnikov kernel $K(u) = \frac{3}{4\sqrt{5}}\left(\frac{15}{8} - \frac{7}{8}u^2\right)\left(1 - \frac{1}{5}u^2\right)1\left(|u| \leq \sqrt{5}\right)$.

We next discuss how to choose the two bandwidth sequences $\{h_0, h\}$. In principle, we can choose the bandwidths to optimize the size and power trade-off. However, this would require higher order expansion of the testing statistics (see, Fan and Linton (2003) and Sun, Phillips and Jin (2008) for related research along this direction). A higher order expansion of our testing statistic is possible but beyond the scope of the current study. We wish to investigate this issue in later research.

Notice that Theorem 4.1 allows us to use the optimal rate of bandwidth for h_0. So we can choose h_0 by the least-squares cross-validation (LSCV). To be specific, denote the following leave-one-out kernel estimator of the long run average density $\overline{f}(x)$ of $\{X_1, \ldots, X_n\}$ by $f_{n,h}^{-t}(X_t) = n^{-1}\sum_{i=1, i\neq t}^n K_h(X_t - X_i)$. Then we can construct the leave-one-out estimate of $\overline{m}_n(X_t)$ by $m_{n,h}^{-t}(X_t) = n^{-1}\sum_{i=1,i\neq t}^n Y_i K_h(X_t - X_i)/f_{n,h}^{-t}(X_t)$, and the LSCV criterion function $CV(h) = \frac{1}{n}\sum_{t=1}^n [Y_t - m_{n,h}^{-t}(X_t)]^2 1(|X_t - \overline{X}| \leq 2\widehat{\sigma}_X)$, where $\overline{X}$ and $\widehat{\sigma}_X$ are the sample mean and sample standard deviation of $\{X_t\}$. We denote the minimizer of the LSCV objective function as h_0. Notice that h_0 is now data dependent and converges to zero at rate $n^{-1/9}$. (Even though we assume non-stochastic bandwidth sequences in Section 3, standard stochastic equicontinuity arguments can be applied to show that stochastic bandwidth sequences are also applicable under suitable conditions.) Since undersmoothing is required for h, we follow Lee (2003, p. 16) to use the rule of thumb: $h = h_0 n^{\frac{1}{9}} n^{-1/\gamma}$, where we shall study the tests for different choices of $\gamma = 7, 6, 5$. See Robinson (1991, p. 448) for very similar devices.

Table 1 reports the finite sample performance of our tests for DGPs s1–s3. To save space, we only report the rejection frequencies for the 5% test. We use 1000 replications for each DGP and 199 bootstrap resamples in each replication. We find that: (a) For different choices of γ, the KS_n and CM_n tests behave similarly. This indicates our tests are robust to

Table 1. Finite sample rejection frequency under the null (nominal level: 0.05)

Sample size n	DGP	Bandwidth: h_0 chosen by LSCV, $h = h_0 n^{1/9} n^{-1/\gamma}$					
		$\gamma = 7$		$\gamma = 6$		$\gamma = 5$	
		KS_n	CM_n	KS_n	CM_n	KS_n	CM_n
100	s1	0.047	0.046	0.054	0.044	0.047	0.048
	s2	0.064	0.046	0.069	0.052	0.052	0.054
	s3	0.057	0.049	0.062	0.063	0.062	0.071
200	s1	0.056	0.053	0.064	0.052	0.061	0.053
	s2	0.067	0.047	0.059	0.054	0.054	0.047
	s3	0.056	0.057	0.062	0.055	0.061	0.061
400	s1	0.046	0.051	0.046	0.047	0.050	0.053
	s2	0.062	0.055	0.056	0.046	0.051	0.051
	s3	0.054	0.052	0.052	0.053	0.060	0.054

different choices of h. (b) For small sample sizes, the tests (KS_n in particular) are oversized for DGP s2 but improves in terms of size as the sample size increases.

5.2 Finite sample performance of the tests under different alternatives

To examine the power performance of the tests, we consider three alternatives:

DGP p1: $Y_t = m_t(X_t) + U_t$, where X_t and U_t are i.i.d. $N(0,1)$ and mutually independent, and

$$m_t(X_t) = \begin{cases} X_t^2, & \text{for } t \le ns_0 \\ X_t^2 + \Delta_0 & \text{for } t > ns_0 \end{cases} . \tag{5.1}$$

DGP p2: $Y_t = m_t(X_t) + U_t$, where X_t and U_t are i.i.d. $N(0,1)$ and mutually independent, and

$$m_t(X_t) = \begin{cases} X_t^2, & \text{for } t \le ns_0 \\ X_t^2 + \Delta_0 X_t & \text{for } t > ns_0 \end{cases} . \tag{5.2}$$

DGP p3: $Y_t = m_t(Y_{t-1}) + \sigma_t(Y_{t-1})\varepsilon_t$, where ε_t are i.i.d. $N(0,1)$, $m_t(Y_{t-1}) = \Phi(Y_{t-1}) + \Delta_0 g(t/n)$, $g(z) = z - \frac{1}{2}z^2$, and

$$\sigma_t(Y_{t-1}) = \begin{cases} \sqrt{0.1 + 0.4Y_{t-1}^2}, & \text{for } t \le 0.5n \\ \sqrt{0.1 + 0.8Y_{t-1}^2} & \text{for } t > 0.5n \end{cases} . \tag{5.3}$$

Note that for DGP p2, $E(\Delta_0 X_t f(X_t)) = 0$, so that the CUSUM test with the unit weight function $(w(x) \equiv 1)$ has no power in detecting such kind of alternatives. For DGP p3, we have structural changes in both the conditional mean and conditional variance process but our interest is still in testing the structural change in the conditional mean process.

We will consider three different break ratios $s_0 = 0.25, 0.5, 0.75$ and examine whether the tests are sensitive to the location of the structural change point. Also, we will consider three different break sizes $\Delta_0 = 0.5, 1, 2$ and check how the test is sensitive to the size of the structural change.

Table 2 reports the rejection frequencies of our tests for DGP p1. To save time, hereafter we use 250 replications for each DGP and 199 bootstrap resamples in each replication. We summarize some main findings from Table 2: (a) The CM_n and KS_n tests behave similarly. (b) As the sample size n or the break size Δ_0 increases, the power of all tests increases. (c) It is easiest to detect a break when it occurs at the break ratio $s_0 = 0.5$.

Table 3 reports the rejection frequencies of our tests for DGP p2 where the test statistics are constructed by us-

Table 2. Finite sample rejection frequencies under DGP p1 (nominal level: 0.05)

Sample size n	Break ratio s_0	Break size Δ_0	Bandwidth: h_0 chosen by LSCV, $h = h_0 n^{1/9} n^{-1/\gamma}$					
			$\gamma = 7$		$\gamma = 6$		$\gamma = 5$	
			KS_n	CM_n	KS_n	CM_n	KS_n	CM_n
100	0.25	0.5	0.164	0.180	0.140	0.168	0.140	0.156
		1	0.612	0.588	0.584	0.588	0.580	0.564
		2	0.996	0.988	0.996	0.980	0.992	0.984
	0.50	0.5	0.296	0.304	0.300	0.288	0.280	0.292
		1	0.912	0.900	0.916	0.884	0.904	0.876
		2	1	1	1	1	1	1
	0.75	0.5	0.196	0.208	0.188	0.168	0.172	0.168
		1	0.668	0.628	0.660	0.608	0.628	0.576
		2	0.992	0.980	0.984	0.972	0.984	0.976
200	0.25	0.5	0.448	0.468	0.440	0.468	0.416	0.436
		1	0.928	0.888	0.936	0.892	0.916	0.892
		2	1	1	1	1	1	1
	0.50	0.5	0.656	0.656	0.668	0.648	0.640	0.628
		1	0.992	0.992	1	0.996	0.996	0.996
		2	1	1	1	1	1	1
	0.75	0.5	0.432	0.440	0.404	0.424	0.392	0.408
		1	0.936	0.912	0.928	0.904	0.920	0.896
		2	1	1	1	1	1	1

Table 3. Finite sample rejection frequencies under DGP p2 (nominal level: 0.05)

Sample size n	Break ratio s_0	Break size Δ_0	Bandwidth: h_0 chosen by LSCV, $h = h_0 n^{1/9} n^{-1/\gamma}$ $\gamma = 7$ KS_n	CM_n	$\gamma = 6$ KS_n	CM_n	$\gamma = 5$ KS_n	CM_n
			$w(x) = (\sin x + \cos x)1_x$					
100	0.25	0.5	0.052	0.052	0.052	0.048	0.056	0.052
		1	0.176	0.160	0.144	0.152	0.144	0.160
		2	0.524	0.460	0.532	0.452	0.520	0.448
	0.50	0.5	0.096	0.104	0.092	0.092	0.088	0.088
		1	0.284	0.312	0.244	0.284	0.272	0.280
		2	0.832	0.796	0.820	0.800	0.820	0.784
200	0.25	0.5	0.184	0.176	0.180	0.180	0.168	0.168
		1	0.408	0.424	0.408	0.424	0.428	0.428
		2	0.852	0.840	0.860	0.848	0.852	0.852
	0.50	0.5	0.240	0.280	0.252	0.268	0.256	0.264
		1	0.632	0.632	0.628	0.632	0.620	0.616
		2	0.952	0.944	0.952	0.948	0.944	0.948
			$w(x) = 1_x$					
100	0.25	0.5	0.068	0.056	0.052	0.052	0.052	0.056
		1	0.060	0.068	0.056	0.060	0.056	0.056
		2	0.072	0.080	0.092	0.092	0.088	0.088
	0.50	0.5	0.064	0.056	0.060	0.056	0.064	0.048
		1	0.060	0.056	0.072	0.064	0.076	0.060
		2	0.100	0.096	0.112	0.104	0.112	0.112
200	0.25	0.5	0.056	0.056	0.068	0.068	0.068	0.072
		1	0.064	0.076	0.064	0.084	0.064	0.076
		2	0.104	0.104	0.096	0.108	0.096	0.108
	0.50	0.5	0.076	0.056	0.072	0.068	0.068	0.068
		1	0.096	0.108	0.072	0.080	0.084	0.084
		2	0.180	0.160	0.180	0.168	0.180	0.152

Table 4. Finite sample rejection frequencies under DGP p3 (nominal level: 0.05)

Sample size n	Break ratio s_0	Break size Δ_0	Bandwidth: h_0 chosen by LSCV, $h = h_0 n^{1/9} n^{-1/\gamma}$ $\gamma = 7$ KS_n	CM_n	$\gamma = 6$ KS_n	CM_n	$\gamma = 5$ KS_n	CM_n
100	0.25	0.5	0.136	0.128	0.140	0.128	0.140	0.144
		1	0.312	0.356	0.336	0.364	0.348	0.396
		2	0.584	0.640	0.676	0.708	0.672	0.712
	0.50	0.5	0.204	0.212	0.216	0.216	0.232	0.208
		1	0.472	0.456	0.508	0.480	0.540	0.492
		2	0.744	0.760	0.780	0.796	0.832	0.836
200	0.25	0.5	0.260	0.280	0.300	0.312	0.300	0.320
		1	0.648	0.648	0.688	0.680	0.692	0.704
		2	0.840	0.868	0.872	0.896	0.908	0.936
	0.50	0.5	0.360	0.364	0.416	0.392	0.424	0.408
		1	0.768	0.764	0.800	0.796	0.840	0.840
		2	0.908	0.912	0.960	0.960	0.972	0.976

ing two weight functions: $w(x) = (\sin(x) + \cos(x))1_x$ and $w(x) \equiv 1_x$. To save space, hereafter we only report the results for the break ratios $s_0 = 0.25, 0.5$ because the case of $s_0 = 0.75$ is similar to that of $s_0 = 0.25$. We find: (a) When we choose the weight function that diverges from the direction where the test has no power in detecting deviations from the null, the KS_n and CM_n tests perform reasonably well. Otherwise, the KS_n and CM_n tests lose their power. (b) The effects of the sample size, break size and break ratio are similar to the case of DGP p1. Table 4 reports the rejection frequencies for DGP p3 where $w(x) = (\sin(x) + \cos(x))1_x$. We find that both the KS_n and CM_n tests work fairly well in detecting the breaks in the conditional mean process in this case too.

APPENDIX

We use C to signify a generic constant whose exact value may vary from case to case and E_i to denote expectation with respect to $v_i \equiv (U_i, X_i')'$. Denote $K_{h,ij} = K_h(X_i - X_j)$. We write $A_n \simeq B_n$ to signify that $A_n = B_n(1 + o_p(1))$ as $n \to \infty$.

A Proof of Theorem 3.1

By definition, $\widehat{V}_i = (U_i + m(X_i) - m_{n,h}(X_i)) f_{n,h}(X_i) w(X_i) = U_i \overline{f}(X_i) w(X_i) + U_i w(X_i)[f_{n,h}(X_i) - \overline{f}(X_i)] - [m_{n,h}(X_i) - m(X_i)] f_{n,h}(X_i) w(X_i)$. Noticing that $\sum_{i=1}^{\lceil ns \rfloor} U_i w(X_i)[f_{n,h}(X_i) - \overline{f}(X_i)] = n^{-1} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} U_i w(X_i) \left[K_{h,ij} - \overline{f}_n(X_i)\right] + \sum_{i=1}^{\lceil ns \rfloor} U_i w(X_i) \left[\overline{f}_n(X_i) - \overline{f}(X_i)\right]$, and that

$$\begin{aligned}
&\sum_{i=1}^{\lceil ns \rfloor} (m_{n,h}(X_i) - m(X_i)) f_{n,h}(X_i) w(X_i) \\
&\quad = s \sum_{j=1}^{n} U_j w(X_j) \overline{f}(X_j) + n^{-1} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} U_j w(X_j) \left[K_{h,ij} - \overline{f}(X_j)\right] \\
&\qquad + \left(\frac{\lceil ns \rfloor}{n} - s\right) \sum_{j=1}^{n} U_j w(X_j) \overline{f}(X_j) + n^{-1} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} U_j (w(X_i) - w(X_j)) K_{h,ij} \\
&\qquad + n^{-1} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} (m(X_j) - m(X_i)) w(X_i) K_{h,ij},
\end{aligned}$$

we have

$$\begin{aligned}
\text{(A.1)} \qquad \widehat{\sigma}\Gamma_n(s) &= \left\{ n^{-1/2} \sum_{i=1}^{\lceil ns \rfloor} U_i w(X_i) \overline{f}(X_i) - s n^{-1/2} \sum_{i=1}^{n} U_i w(X_i) \overline{f}(X_i) \right\} \\
&\quad + n^{-3/2} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} U_i w(X_i) \left[K_{h,ij} - \overline{f}_n(X_i)\right] + n^{-1/2} \sum_{i=1}^{\lceil ns \rfloor} U_i w(X_i) \left[\overline{f}_n(X_i) - \overline{f}(X_i)\right] \\
&\quad - n^{-3/2} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} U_j w(X_j) \left[K_{h,ij} - \overline{f}(X_j)\right] - n^{-1/2} \left(\frac{\lceil ns \rfloor}{n} - s\right) \sum_{j=1}^{n} U_j w(X_j) \overline{f}(X_j) \\
&\quad - n^{-3/2} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} U_j (w(X_i) - w(X_j)) K_{h,ij} - n^{-3/2} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} (m(X_j) - m(X_i)) w(X_i) K_{h,ij} \\
&\equiv A_{n1}(s) + A_{n2}(s) + A_{n3}(s) - A_{n4}(s) - A_{n5}(s) - A_{n6}(s) - A_{n7}(s).
\end{aligned}$$

By the invariance principle for a strong mixing process that is not necessarily stationary (e.g., Herrndorf, 1985), $A_{n1}(\cdot)$ converges weakly to $\sigma_0 W^0(\cdot)$. The conclusion then follows from Lemmata A.1–A.6 below and the fact that $\widehat{\sigma} \xrightarrow{p} \sigma_0$ under the null. □

We prove the following lemmata under the conditions of Theorem 3.1.

Lemma A.1. *$A_{n2}(s) \equiv n^{-3/2} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} U_i w(X_i) \left[K_{h,ij} - \overline{f}_n(X_i)\right] = o_p(1)$ uniformly in s.*

Proof. Recall $v_i = (U_i, X_i')'$. Let $\varphi_1(v_j, v_i) = U_i w(X_i)[K_{h,ij} - E_j K_{h,ij}]$, and $\varphi_2(v_j, v_i) = U_i w(X_i)[E_j K_{h,ij} - f_j(X_i)]$. Then $A_{n2}(s) = n^{-3/2} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} \varphi_1(v_j, v_i) + n^{-3/2} \sum_{i=1}^{\lceil ns \rfloor} \sum_{j=1}^{n} \varphi_2(v_j, v_i) \equiv A_{n21}(s) + A_{n22}(s)$. By Assumption A4 and Lemma D.4, $\sup_{1 \le s \le 1} |A_{n22}(s)| \le n^{-3/2} h^r \sum_{i=1}^{n} \sum_{j=1}^{n} |U_i w(X_i) D_f(X_i)| = O_p(n^{1/2} h^r) = o_p(1)$, where $D_f(\cdot)$ is defined in Lemma D.4; see also the remark after it. To show $\sup_{1 \le s \le 1} |A_{n21}(s)| = o_p(1)$, write

$$\begin{aligned}
A_{n21}(s) &= n^{-3/2} \sum_{i=1}^{\lceil ns \rfloor} \varphi_{1ii} + n^{-3/2} \sum_{1 \le j < i \le \lceil ns \rfloor} \varphi_{1ji} + n^{-3/2} \sum_{1 \le i < j \le n} \varphi_{1ji} - n^{-3/2} \sum_{\lceil ns \rfloor + 1 \le i < j \le n} \varphi_{1ji} \\
&\equiv A_{n21a}(s) + A_{n21b}(s) + A_{n21c} - A_{n21d}(s),
\end{aligned}$$

where $\varphi_{1ji} = \varphi_1(v_j, v_i)$. It suffices to show that $\sup_{0\le s\le 1}|A_{n21a}(s)| = o_p(1)$, $\sup_{0\le s\le 1}|A_{n21b}(s)| = o_p(1)$, and $\sup_{0\le s\le 1}|A_{n21d}(s)| = o_p(1)$. First, by the invariance principle for the strong mixing process $\{X_i, U_i\}$, it is easy to show that uniformly in s,

$$A_{n21a}(s) = n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor} U_i w(X_i)\left[h^{-d}K(0) - \int K(u) f_i(X_i + hu)\,du\right] = O_p\left(n^{-1}h^{-d}\right) = o_p(1).$$

Next, write

(A.2)
$$E[A_{n21b}(s)]^4 = n^{-6}\sum_{1\le i_1<i_2\le\lceil ns\rfloor}\sum_{1\le i_3<i_4\le\lceil ns\rfloor}\sum_{1\le i_5<i_6\le\lceil ns\rfloor}\sum_{1\le i_7<i_8\le\lceil ns\rfloor}\varphi_1(v_{i_1}, v_{i_2})\varphi_1(v_{i_3}, v_{i_4})\varphi_1(v_{i_5}, v_{i_6})\varphi_1(v_{i_7}, v_{i_8}).$$

It is easy to show that the dominating terms in the above summation constitute two cases: (a) $i_1, \ldots, i_8$ are distinct integers; (b) $\{i_1, i_2\}$, $\{i_3, i_4\}$, $\{i_5, i_6\}$ and $\{i_7, i_8\}$ form two identical pairs (e.g., $\{i_1, i_2\} = \{i_3, i_4\}$ and $\{i_5, i_6\} = \{i_7, i_8\}$). We will use $EA_{n21b(l)}$ to denote these two cases ($l = a, b$).

For case (a), let $i_1, \ldots, i_8$ be distinct integers with $1 \le i_j \le \lceil ns\rfloor$. Let $1 \le k_1 < \ldots < k_8 \le \lceil ns\rfloor$ be the permutation of $i_1, \ldots, i_8$ in ascending order and let d_c be the c-th largest difference among $k_{j+1} - k_j$, $j = 1, \ldots, 7$. Define $H(k_1, \ldots, k_8) = \varphi_1(v_{i_1}, v_{i_2})\varphi_1(v_{i_3}, v_{i_4})\varphi_1(v_{i_5}, v_{i_6})\varphi_1(v_{i_7}, v_{i_8})$. For any $1 \le j \le 7$, put $P_0^{(8)}\left(E^{(8)}\right) = P\left((v_{i_1}, \ldots, v_{i_8}) \in E^{(8)}\right)$, and $P_j^{(8)}\left(E^{(j)} \times E^{(8-j)}\right) = P((v_{i_1}, \ldots, v_{i_j}) \in E^{(j)})P((v_{i_{j+1}}, \ldots, v_{i_8}) \in E^{(8-j)})$, where $E^{(j)}$ is a Borel set in $\mathbb{R}^{j(d+1)}$. By Assumptions A2–A3, one can show that for any $0 \le j \le 7$, $\int |H(k_1, \ldots, k_8)|^{1+\delta/4}\,dP_j^{(8)} \le Ch^{-d\delta}$. Applying Lemma D.1 with $\vartheta = \delta/4$,

$$|E[H(k_1, \ldots, k_8)]| \le \begin{cases} Ch^{-4d\delta/(4+\delta)}\alpha^{\frac{\delta}{4+\delta}}(k_2 - k_1) & \text{if } k_2 - k_1 = d_1 \\ Ch^{-4d\delta/(4+\delta)}\alpha^{\frac{\delta}{4+\delta}}(k_8 - k_7) & \text{if } k_8 - k_7 = d_1. \end{cases}$$

Therefore

(A.3)
$$\begin{aligned}\sum_{\substack{1\le k_1<\ldots<k_8\le n\\ k_2-k_1=d_1}}|E[H(k_1, \ldots, k_8)]| &\le Ch^{-4d\delta/(4+\delta)}\sum_{k_1=1}^{n-7}\sum_{k_2=k_1+\max_{j\ge 3}\{k_j-k_{j-1}\}}^{n-6}\sum_{k_3=k_2+1}^{n-5}\cdots\sum_{k_8=k_7+1}^{n}\alpha^{\frac{\delta}{4+\delta}}(k_2 - k_1)\\ &\le Ch^{-4d\delta/(4+\delta)}\sum_{k_1=1}^{n-7}\sum_{k_2=k_1+1}^{n-6}(k_2 - k_1)^6\alpha^{\frac{\delta}{4+\delta}}(k_2 - k_1) \le Cnh^{-4d\delta/(4+\delta)}\sum_{j=1}^{n}j^6\alpha^{\frac{\delta}{4+\delta}}(j).\end{aligned}$$

Similarly, we have

$$\sum_{\substack{1\le k_1<\ldots<k_8\le n\\ k_8-k_7=d_1}}|E[H(k_1, \ldots, k_8)]| \le Cnh^{-4d\delta/(4+\delta)}\sum_{j=1}^{n}j^6\alpha^{\frac{\delta}{4+\delta}}(j), \tag{A.4}$$

$$\sum_{\substack{1\le k_1<\ldots<k_8\le n\\ k_2-k_1=d_2 \text{ or } k_8-k_7=d_2}}|E[H(k_1, \ldots, k_8)]| \le Cn^2h^{-4d\delta/(4+\delta)}\sum_{j=1}^{n}j^5\alpha^{\frac{\delta}{4+\delta}}(j), \tag{A.5}$$

$$\sum_{\substack{1\le k_1<\ldots<k_8\le n\\ k_2-k_1=d_3 \text{ or } k_8-k_7=d_3}}|E[H(k_1, \ldots, k_8)]| \le Cn^3h^{-4d\delta/(4+\delta)}\sum_{j=1}^{n}j^4\alpha^{\frac{\delta}{4+\delta}}(j), \tag{A.6}$$

and for all other subcases ($k_2 - k_1 = d_c$ and $k_8 - k_7 = d_{c'}$ for $c, c' \ge 4$)

$$\sum_{\substack{1\le k_1<\ldots<k_8\le n\\ \text{other subcases}}}|E[H(k_1, \ldots, k_8)]| \le Cn^4h^{-4d\delta/(4+\delta)}\sum_{j=1}^{n}j^3\alpha^{\frac{\delta}{4+\delta}}(j). \tag{A.7}$$

By (A.3)–(A.7), Assumptions A1 and A6, we have

(A.8)
$$EA_{n21b(a)} \leq n^{-6} \sum_{1\leq k_1<\ldots<k_8\leq n} \left|E\left[H\left(k_1,\ldots,k_8\right)\right]\right| \leq Cn^{-2}h^{-4d\delta/(4+\delta)} \sum_{j=1}^{n} j^3 \alpha^{\frac{\delta}{4+\delta}}\left(j\right) = O\big(n^{-2}h^{-4d\delta/(4+\delta)}\big) = o\big(n^{-1}\big).$$

Now for case (b), some calculations show that

$$EA_{n21b(b)} = O\left(n^{-2}h^{-2d}\right) = o\big(n^{-1}\big). \tag{A.9}$$

Hence $E\left[A_{n21b}\left(s\right)\right]^4 = o\big(n^{-1}\big)$ by (A.8)–(A.9) and the remark after (A.2). For any $\epsilon > 0$,

$$P\left(\sup_{0\leq s\leq 1} ||A_{n21b}\left(s\right)|| > \epsilon\right) = \sum_{l=1}^{n} P\left(A_{n21b}\left(l/n\right) > \epsilon\right) \leq \epsilon^{-4} \sum_{l=1}^{n} E\left|A_{n11}\left(l/n\right)\right|^4 = o\left(1\right).$$

It follows that $\sup_{0\leq s\leq 1}\left|A_{n21b}\left(s\right)\right| = o_p\left(1\right)$.

Now, let $\widetilde{\varphi}\left(v_i,v_j\right) = \varphi_1\left(v_j,v_i\right)$ and $\widetilde{v}_i = v_{n-i+1}$ for $1\leq i,j,\leq n$. Then

$$\sup_{1\leq l\leq n}\left|\sum_{l\leq i<j\leq n} \varphi_1\left(v_j,v_i\right)\right| = \sup_{1\leq l\leq n}\left|\sum_{1\leq i<j\leq n-l+1} \varphi_1\left(v_{n-j+1},v_{n-i+1}\right)\right| = \sup_{1\leq l\leq n}\left|\sum_{1\leq i<j\leq l} \widetilde{\varphi}\left(\widetilde{v}_i,\widetilde{v}_j\right)\right|.$$

So we can apply the above method to the $\{\widetilde{v}_i\}$ variable to obtain ${}_{0\leq s\leq 1}\left|A_{n21d}\left(s\right)\right| = o_p\left(1\right)$. Hence $\sup_{0\leq s\leq 1}\left|A_{n21}\left(s\right)\right| = o_p\left(1\right)$. □

Lemma A.2. $A_{n3}\left(s\right) \equiv n^{-1/2}\sum_{i=1}^{\lceil ns\rfloor} U_i w\left(X_i\right)\left[\overline{f}_n\left(X_i\right) - \overline{f}\left(X_i\right)\right] = o_p\left(1\right)$ *uniformly in* s.

Proof. Let $d_i = U_i w\left(X_i\right)\left[\overline{f}_n\left(X_i\right) - \overline{f}\left(X_i\right)\right]$. Recall $v_i = \left(U_i, X_i'\right)'$. Define $\mathcal{F}_l = \sigma\text{-field}\left(v_1,\ldots,v_l\right)$ for $1\leq l\leq n$. Then $\{d_i,\mathcal{F}_i\}_{i=1,\ldots,n}$ is a martingale difference sequence (m.d.s.) by Assumption A2, and $A_{n3}\left(s\right) = n^{-1/2}\sum_{i=1}^{\lceil ns\rfloor} d_i$. By the Kolmogorov's maximal inequality for m.d.s. (see, e.g., Theorem 15.14 in Davidson, 1994), we have for any $\epsilon > 0$,

$$P\left(\sup_{1\leq s\leq 1}\left|A_{n3}\left(s\right)\right| > \epsilon\right) = P\left(\sup_{1\leq l\leq n}\left|\sum_{i=1}^{l} d_i\right| > n^{1/2}\epsilon\right) \leq n^{-1}\epsilon^{-2}\sum_{l=1}^{n}\mathrm{Var}\left(d_i\right).$$

Noting that $E(d_i) = 0$, by the dominated convergence theorem $\mathrm{Var}\left(d_i\right) = E\{U_i^2 w^2\left(X_i\right)[\overline{f}_n\left(X_i\right) - \overline{f}\left(X_i\right)]^2\} = o(1)$. So $P\left(\sup_{1\leq s\leq 1}\left|A_{n3}\left(s\right)\right| > \epsilon\right) = o(1)$, implying that $\sup_{1\leq s\leq 1}\left|A_{n3}\left(s\right)\right| = o_p\left(1\right)$. □

Lemma A.3. $A_{n4}\left(s\right) \equiv n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{n} U_j w\left(X_j\right)\left[K_{h,ij} - \overline{f}\left(X_j\right)\right] = o_p\left(1\right)$ *uniformly in* s.

Proof. Recall $v_i = \left(U_i, X_i'\right)'$. Let $\phi\left(v_i,v_j\right) = U_j w\left(X_j\right)\left[K_{h,ij} - \overline{f}\left(X_j\right)\right]$. Then

$$A_{n4}\left(s\right) = n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{\lceil ns\rfloor}\phi\left(v_i,v_j\right) + n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=\lceil ns\rfloor+1}^{n}\phi\left(v_i,v_j\right) \equiv A_{n4a}\left(s\right) + A_{n4b}\left(s\right).$$

It suffices to show $\sup_{0\leq s\leq 1}\left|A_{n4a}\left(s\right)\right| = o_p\left(1\right)$ and $\sup_{0\leq s\leq 1}\left|A_{n4b}\left(s\right)\right| = o_p\left(1\right)$.

We now write $A_{n4b}\left(s\right)$ as the sum of three U-statistics:

$$A_{n4b}\left(s\right) = n^{-3/2}\left\{\sum_{1\leq i<j\leq n}\phi\left(v_i,v_j\right) - \sum_{1\leq i<j\leq\lceil ns\rfloor}\phi\left(v_i,v_j\right) - \sum_{\lceil ns\rfloor+1\leq i<j\leq n}^{n}\phi\left(v_i,v_j\right)\right\}.$$

We shall prove that for all $\epsilon_n > 0$ and some fixed sufficiently large C,

$$p_1 \equiv P\left(\sup_{1\leq l\leq n}\left|\sum_{1\leq i<j\leq l}\phi\left(v_i,v_j\right)\right| > \epsilon_n\right) \leq C\epsilon_n^{-2}n^2h^{-d}, \quad \text{and} \tag{A.10}$$

$$p_2 \equiv P\left(\sup_{1\leq l\leq n}\left|\sum_{l\leq i<j\leq n}\phi\left(v_i,v_j\right)\right| > \epsilon_n\right) \leq C\epsilon_n^{-2}n^2h^{-d}. \tag{A.11}$$

By construction, $\iint \phi(v_1, v_2)\, dF_{v_1}(v_1)\, dF_{v_2}(v_2) = 0$, where F_{v_i} is the distribution function of v_i. Let $\phi(v) = \int \phi(v, v_2)\, dF_{v_2}(v_2)$ and $\widetilde{\phi}(v) = \int \phi(v_1, v)\, dF_{v_1}(v_1)$. By Assumption A2, $\phi(v) = 0$ and $\widetilde{\phi}(v) = 0$. For the investigation of p_1, define $\mathcal{F}_l = \sigma$-field$(v_1, \ldots, v_l)$ for $1 \le l \le n$. Let $B_1 = 0$, $\Delta_1 = 0$, $B_l = \sum_{1\le i<j\le l} \phi(v_i, v_j)$, and $\Delta_l = B_l - B_{l-1} = \sum_{i=1}^{l-1} \phi(v_i, v_l)$, for $l = 2, \ldots, n$. Then $\{\Delta_l, \mathcal{F}_l\}_{l=1,\ldots,n}$ is an m.d.s. By the Kolmogorov's maximal inequality for m.d.s., we have

$$P\left(\sup_{1\le l\le n}\left|\sum_{1\le i<j\le l}\phi(v_i,v_j)\right| > \epsilon_n\right) = P\left(\sup_{1\le l\le n}\left|\sum_{i=1}^{l}\Delta_i\right| > \epsilon_n\right) \le \epsilon_n^{-2}\sum_{l=1}^{n}\operatorname{Var}(\Delta_l).$$

Noticing that $E[\Delta_l] = 0$, we have

$$\operatorname{Var}(\Delta_l) = \sum_{i=1}^{l-1} E[\phi(v_i, v_l)]^2 + 2\sum_{i=1}^{l-2}\sum_{j=i+1}^{l-1}\operatorname{Cov}(\phi(v_i, v_l), \phi(v_j, v_l)). \tag{A.12}$$

It is easy to show that $E[\phi(v_i, v_l)]^2 \le Ch^{-d}$ for all $i < l \le n$. Applying Lemma D.1 with $\vartheta = \delta$, we have that for all $1 \le i < j < l \le n$,

$$\begin{aligned}
|\operatorname{Cov}(\phi(v_i, v_l), \phi(v_j, v_l))| &= |E\{\phi(v_i, v_l)\phi(v_j, v_l)\}| \\
&\le \left|\int E[\phi(v_i, v)\phi(v_j, v)]\, dF_{v_l}(v)\right| + Ch^{\frac{-2\delta d}{1+\delta}}[\alpha(l-j)]^{\frac{\delta}{1+\delta}} \\
&\le C[\alpha(j-i)]^{\frac{\delta}{1+\delta}} \max_{1\le j\le n}\int \left(E|\phi(v_j, v)|^{2+2\delta}\right)^{1/(1+\delta)} dF_{v_l}(v) + Ch^{\frac{-2\delta d}{1+\delta}}[\alpha(l-j)]^{\frac{\delta}{1+\delta}} \\
&= Ch^{\frac{-2\delta d}{1+\delta}}[\alpha(j-i)]^{\frac{\delta}{1+\delta}} + Ch^{\frac{-2\delta d}{1+\delta}}[\alpha(l-j)]^{\frac{\delta}{1+\delta}}.
\end{aligned}$$

Note that Assumption A1 implies that $\alpha^{\delta/(1+\delta)}(j)$ is summable. Hence

$$\begin{aligned}
\operatorname{Var}(\Delta_{l+1}) &\le C\left(lh^{-d} + 2h^{\frac{-2\delta d}{1+\delta}}\sum_{i=1}^{l-2}\sum_{j=i+1}^{l-1}\left\{[\alpha(j-i)]^{\frac{\delta}{1+\delta}} + [\alpha(l-j)]^{\frac{\delta}{1+\delta}}\right\}\right) \\
&\le C\left(lh^{-d} + 2lh^{\frac{-2\delta d}{1+\delta}}\sum_{j=1}^{l-1}\left\{[\alpha(j-i)]^{\frac{\delta}{1+\delta}} + \alpha(l-j)^{\frac{\delta}{1+\delta}}\right\}\right) \\
&\le C\left(lh^{-d} + Clh^{\frac{-2\delta d}{1+\delta}}\right) \le Clh^{-d},
\end{aligned}$$

and assertion (A.10) follows. Let $\widetilde{\phi}(v_i, v_j) = \phi(v_j, v_i)$ and $\widetilde{v}_i = v_{n-i+1}$ for $1 \le i, j, \le n$. Then

$$\sup_{1\le l\le n}\left|\sum_{l\le i<j\le n}\phi(v_i,v_j)\right| = \sup_{1\le l\le n}\left|\sum_{1\le i<j\le n-l+1}\phi(v_{n-j+1}, v_{n-i+1})\right| = \sup_{1\le l\le n}\left|\sum_{1\le i<j\le l}\widetilde{\phi}(\widetilde{v}_i, \widetilde{v}_j)\right|.$$

So we can apply (A.10) to the $\widetilde{v}_i$ variables to get (A.11).

Now taking $\epsilon_n = n^{3/2}\epsilon$ for some finite $\epsilon > 0$, we obtain

$$p_1 \equiv P\left(\sup_{1\le l\le n}\left|\sum_{1\le i<j\le l}\phi(v_i,v_j)\right| > n^{3/2}\epsilon\right) \le C\epsilon^{-2}n^{-1}h^{-d} \to 0, \quad \text{and}$$

$$p_2 \equiv P\left(\sup_{1\le l\le n}\left|\sum_{l\le i<j\le n}\phi(v_i,v_j)\right| > n^{3/2}\epsilon\right) \le C\epsilon^{-2}n^{-1}h^{-d} \to 0.$$

Consequently, $\sup_{0\le s\le 1}|A_{n4b}(s)| = o_p(1)$.

To show $\sup_{0\le s\le 1}|A_{n4a}(s)| = o_p(1)$, let $\varphi(v_i, v_j) = U_j w(X_j)[K_{h,ij} - f_i(X_j)]$. Then

$$\begin{aligned}
A_{n4a}(s) &= n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{\lceil ns\rfloor}\varphi(v_i, v_j) + n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{\lceil ns\rfloor}U_j w(X_j)\left[f_i(X_j) - \overline{f}(X_j)\right] \\
&= n^{-3/2}\sum_{1\le i<j\le\lceil ns\rfloor}\varphi(v_i, v_j) + n^{-3/2}\sum_{1\le i<j\le\lceil ns\rfloor}\varphi(v_j, v_i) + n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\varphi(v_i, v_i) \\
&\quad + n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{\lceil ns\rfloor}U_j w(X_j)\left[f_i(X_j) - \overline{f}(X_j)\right] \\
&\equiv D_{n3}(s) + D_{n4}(s) + D_{n5}(s) + D_{n6}(s).
\end{aligned}$$

By similar arguments to the above analysis and that of Lemma A.1, we have $\sup_{0\le s\le 1}|D_{nj}(s)| = o_p(1)$ for $j = 3, 4$. We can readily show that

$$D_{n5}(s) = n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}U_i w(X_i)\left[h^{-d}K(0) - f_i(X_i)\right] = O_p\left(n^{-1}h^{-d}\right) = o_p(1) \text{ uniformly in } s.$$

Now write

$$D_{n6}(s) = n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{\lceil ns\rfloor}U_j w(X_j)\left[f_i(X_j) - \overline{f}(X_j)\right] = \frac{\lceil ns\rfloor}{n^{3/2}}\sum_{j=1}^{\lceil ns\rfloor}U_j w(X_j)\left[\overline{f}_{\lceil ns\rfloor}(X_j) - \overline{f}(X_j)\right].$$

Analogously to the proof of by Lemma A.2, we can show that $\sup_{0\le s\le 1}|D_{n6}(s)| = o_p(1)$. Consequently $\sup_{0\le s\le 1}|A_{n4a}(s)| = o_p(1)$. ☐

Lemma A.4. $A_{n5}(s) \equiv n^{-1/2}(\lceil ns\rfloor/n - s)\sum_{j=1}^{n}U_j w(X_j)\overline{f}(X_j) = o_p(1)$ *uniformly in* s.

Proof. The result follows from the fact that $\lceil ns\rfloor/n - s = O(n^{-1})$ uniformly in s and the fact that $n^{-1/2}\sum_{j=1}^{n}U_j w(X_j)\times \overline{f}(X_j) = O_p(1)$. ☐

Lemma A.5. $A_{n6}(s) \equiv n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{n}U_j(w(X_i) - w(X_j))K_{h,ij} = o_p(1)$ *uniformly in* s.

Proof. The proof is analogous to that of Lemma A.3. ☐

Lemma A.6. $A_{n7}(s) \equiv n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{n}(m(X_j) - m(X_i))w(X_i)K_{h,ij} = o_p(1)$ *uniformly in* s.

Proof. Using Lemma D.4, we can show that uniformly in x,

$$\begin{aligned}
&\left|n^{-1}\sum_{j=1}^{n}[m(X_j) - m(x)]w(x)K_h(x - X_j)\right| \\
&\quad \simeq \left|n^{-1}\sum_{j=1}^{n}E\{[m(X_j) - m(x)]w(x)K_h(x - X_j)\}\right| \le h^r D_m(x)|w(x)|.
\end{aligned}$$

Consequently, $\sup_{0\le s\le 1}|A_{n7}(s)| \le n^{-1/2}h^r\sum_{i=1}^{n}D_m(X_i)|w(X_i)| = O_p\left(n^{1/2}h^r\right) = o_p(1)$. ☐

B Proof of Theorem 3.2

Under both $H_{1A,n}$ and $H_{1B,n}$, Lemmata A.1–A.5 also hold true and $\widehat{\sigma} \xrightarrow{p} \sigma_0$. Now

$$\widehat{\sigma}\Gamma_n(s) = A_{n1}(s) + A_{n2}(s) + A_{n3}(s) - A_{n4}(s) - A_{n5}(s) - A_{n6}(s) - \widetilde{A}_{n7}(s),$$

where $A_{ni}(s)$, $i = 1, \ldots, 6$, are as defined in (A.1), and $\widetilde{A}_{n7}(s) = n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{n}(m_j(X_j) - m_i(X_i))w(X_i)K_{h,ij}$. Let $A_{n7a}(s) = n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{n}(m_j(X_i) - m_i(X_i))w(X_i)K_{h,ij}$, and $A_{n7b}(s) = n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{n}(m_j(X_j) - m_j(X_i))\times w(X_i)K_{h,ij}$. Then $\widetilde{A}_{n7}(s) = A_{n7a}(s) + A_{n7b}(s)$. Analogous to the proof of Lemma A.6, we can prove that

$\sup_{0\le s\le 1}|A_{n7b}(s)| = o_p(1)$ under both H_{1A} and H_{1B}. To analyze $A_{n7a}(s)$, we first focus on the case of H_{1A}. Under H_{1A}, $m_j(X_i) - m_i(X_i) = n^{-1/2}\Delta(X_i)\,1(j \ge k_0+1)$. Uniformly in s, we have

$$A_{n7a}(s) = \frac{1}{n^2}\sum_{i=1}^{\lceil ns\rfloor}\Delta(X_i)\,w(X_i)\sum_{j=k_0+1}^{n}K_{h,ij} = \frac{1-s_0}{n}\sum_{i=1}^{\lceil ns\rfloor}\Delta(X_i)\,w(X_i)\,\overline{f}(X_i) + o_p(1) = s(1-s_0)\,\mu_1 + o_p(1),$$

where $\mu_1 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left[\Delta(X_i)\,w(X_i)\,\overline{f}(X_i)\right]$.

We now focus on the case of H_{1B}. Under H_{1B}, $m_j(X_i) - m_i(X_i) = n^{-1/2}(g(j/n) - g(i/n))$. By Lemma 4 of Krämer, Ploberger and Alt (1988), we have that uniformly in s,

$$\begin{aligned}
A_{n7a}(s) &= \frac{1}{n^2}\sum_{i=1}^{\lceil ns\rfloor} w(X_i)\sum_{j=1}^{n}\left(g\left(\frac{j}{n}\right) - g\left(\frac{i}{n}\right)\right)K_{h,ij}\\
&= \frac{1}{n^2}\sum_{j=1}^{n} g\left(\frac{j}{n}\right)\sum_{i=1}^{\lceil ns\rfloor} w(X_i)\,K_{h,ij} - \frac{1}{n^2}\sum_{i=1}^{\lceil ns\rfloor} g\left(\frac{i}{n}\right)w(X_i)\sum_{j=1}^{n}K_{h,ij}\\
&\simeq \frac{s}{n}\sum_{j=1}^{n} g\left(\frac{j}{n}\right)w(X_j)\,\overline{f}(X_j) - \frac{1}{n}\sum_{i=1}^{\lceil ns\rfloor} g\left(\frac{i}{n}\right)w(X_i)\,\overline{f}(X_i) + o_p(1)\\
&\simeq s\mu_2\int_0^1 g(v)\,dv - \mu_2\int_0^s g(v)\,dv = \mu_2\left(s\int_0^1 g(v)\,dv - \int_0^s g(v)\,dv\right),
\end{aligned}$$

where $\mu_2 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left[w(X_i)\,\overline{f}(X_i)\right]$, the third line follows from the fact that $\frac{1}{n}\sum_{j=1}^{n} h^{-d}K((x-X_j)/h) = \overline{f}(x) + o_p(1)$ uniformly in x, and that $\frac{1}{n}\sum_{i=1}^{\lceil ns\rfloor} w(X_i)\,h^{-d}K((x-X_i)/h) \simeq \frac{1}{n}\sum_{i=1}^{\lceil ns\rfloor} E\left[w(X_i)\,h^{-d}K((x-X_i)/h)\right] \to s\,w(x)\,\overline{f}(x)$ uniformly in x and s. □

C Proof of Theorem 4.1

Let P^* denote the probability conditional on the original sample $\mathcal{W} \equiv \{(Y_t, X_t)\}_{t=1}^{n}$ and E^* denote the expectation with respect to P^*. Let $O_{p^*}(1)$ and $o_{p^*}(1)$ denote the probability order in the bootstrap world, e.g., $b_n = o_{p^*}(1)$ if for any $\epsilon > 0$, $P^*(\|b_n\| > \epsilon) = o_p(1)$. Note that $b_n = o_p(1)$ implies that $b_n = o_{p^*}(1)$. Similar conclusions hold when little o is replaced by big O.

By definition, we can write $\widehat{V}_i^* = \left[U_i^* + m_{n,h_0}(X_i) - m_{n,h}^*(X_i)\right]f_{n,h}(X_i)\,w(X_i) = U_i^* f_{n,h}(X_i)\,w(X_i) - n^{-1}\sum_{j=1}^{n} U_j^* K_{h,ij} w(X_i) - \left\{E^*\left[m_{n,h}^*(X_i)\right] - m_{n,h_0}(X_i)\right\}f_{n,h}(X_i)\,w(X_i)$. Hence

$$\begin{aligned}
\widehat{\sigma}^*\Gamma_n^*(s) &= \left\{n^{-1/2}\sum_{i=1}^{\lceil ns\rfloor} U_i^* f_{n,h}(X_i)\,w(X_i) - s n^{-1/2}\sum_{i=1}^{n} U_i^* f_{n,h}(X_i)\,w(X_i)\right\}\\
&\quad - n^{-3/2}\sum_{i=1}^{\lceil ns\rfloor}\sum_{j=1}^{n} U_j^* w(X_j)\left[K_{h,ij} - f_{n,h}(X_j)\right] - n^{-1/2}\left(\frac{\lceil ns\rfloor}{n} - s\right)\sum_{j=1}^{n} U_j^* f_{n,h}(X_j)\,w(X_j)\\
&\quad - n^{-1/2}\sum_{i=1}^{\lceil ns\rfloor}\left\{E^*\left[m_{n,h}^*(X_i)\right] - m_{n,h_0}(X_i)\right\}f_{n,h}(X_i)\,w(X_i)\\
&\equiv A_{n1}^*(s) - A_{n2}^*(s) - A_{n5}^*(s) - R_n^*(s),
\end{aligned} \tag{C.1}$$

where the expressions of $A_{n1}^*(s)$, $A_{n2}^*(s)$, and $A_{n5}^*(s)$ parallel those of $A_{n1}(s)$, $A_{n2}(s)$, and $A_{n5}(s)$ in the proof of Theorem 3.1. It suffices to show that conditional on $\mathcal{W}$,

(i) $A_{n1}^*(\cdot) \overset{p}{\Rightarrow} \sigma^* W^0(\cdot)$,
(ii) $A_{nj}^*(s) = o_{p^*}(1)$ uniformly in s for $j = 2, 5$,
(iii) $R_n^*(s) = o_{p^*}(1)$ uniformly in s,
(iv) $\widehat{\sigma}^{*2} = \sigma^{*2} + o_{p^*}(1)$,

where $\sigma^{*2} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\{\left[\overline{f}(X_i)\,U_i + \delta(X_i)\right]^2 w^2(X_i)\}$, and $\delta(X_i) = \lim_{n\to\infty} n^{-1}\sum_{j=1}^{n}[m_i(X_i) - m_j(X_i)] \times f_j(X_i)$. Clearly, under the null hypothesis or local alternatives, $\delta \equiv 0$. Otherwise, it is not identically zero.

Let $v_i^* = U_i^* f_{n,h}(X_i) w(X_i)$. Conditional on $\mathcal{W}$, $\{v_i^*\}$ is an independent but not identically distributed (i.n.i.d. for abbreviation) sequence. To show (i), it suffices to show that conditional on $\mathcal{W}$, $\{v_i^*\}$ satisfies the Liapounov i.n.i.d. functional central limit theorem, see, e.g., Theorem 7.16 of White (2001). By construction, $E^*(v_i^*) = 0$. $\mathrm{Var}^*(v_i^*) = \overline{\widehat{U}}_i^2 f_{n,h}^2(X_i) w^2(X_i)$ and $E|v_i^*|^4 = E(\eta_i^4) \overline{\widehat{U}}_i^4 f_{n,h}^4(X_i) w^4(X_i)$ are bounded in probability for all i and for sufficiently large n. In addition, noting that $\widehat{U}_i f_{n,h}(X_i) = n^{-1}\sum_{j=1}^n K_{hij}(Y_i - Y_j) = \overline{f}(X_i) U_i + \delta(X_i) + o_p(1)$, we have

$$\begin{aligned}\mathrm{Var}^*\left(n^{-1/2}\sum_{i=1}^n v_i^*\right) &= n^{-1}\sum_{i=1}^n \overline{\widehat{U}}_i^2 f_{n,h}^2(X_i) w^2(X_i) = n^{-1}\sum_{i=1}^n \widehat{U}_i^2 f_{n,h}^2(X_i) w^2(X_i) + o_p(1)\\ &= n^{-1}\sum_{i=1}^n \left[\overline{f}(X_i) U_i + \delta(X_i)\right]^2 w^2(X_i) o_p(1) = \sigma^{*2} + o_p(1).\end{aligned}$$

Hence the conditions of Theorem 7.16 of White (2001) are satisfied, and $A_{n1}^*(\cdot) \overset{p}{\Rightarrow} \sigma^* W^0(\cdot)$.

To show $A_{n2}^*(s) = o_{p^*}(1)$, let $\xi_j^* = (\eta_j, \overline{\widehat{U}}_j, X_j)$ and $\phi(\xi_i^*, \xi_j^*) = \eta_j \overline{\widehat{U}}_j w(X_j)[K_{h,ij} - f_{n,h}(X_j)]$. Then $A_{n2}^*(s) = n^{-3/2}\sum_{i=1}^{\lceil ns \rfloor}\sum_{j=1}^n \phi_{ij}$, where $\phi_{ij} = \phi(\xi_i^*, \xi_j^*)$. Fix $s \in [0,1]$. For any $\epsilon > 0$,

$$\text{(C.2)}\qquad P^*(A_{n2}^*(s) \geq \epsilon) \leq n^{-6}\epsilon^{-4}\sum_{i_1=1}^n\sum_{i_2=1}^{\lceil ns \rfloor}\sum_{i_3=1}^n\sum_{i_4=1}^{\lceil ns \rfloor}\sum_{i_5=1}^n\sum_{i_6=1}^{\lceil ns \rfloor}\sum_{i_7=1}^n\sum_{i_8=1}^{\lceil ns \rfloor} E^*[\phi_{i_1i_2}\phi_{i_3i_4}\phi_{i_5i_6}\phi_{i_7i_8}].$$

Note that $E^*[\phi_{i_1i_2}\phi_{i_3i_4}\phi_{i_5i_6}\phi_{i_7i_8}]$ is nonzero if and only if (a) $i_1 = i_3 = i_5 = i_7$, or (b) $i_1 = i_3$ and $i_5 = i_7$, or $i_1 = i_5$ and $i_3 = i_7$, or $i_1 = i_7$ and $i_3 = i_7$ and these four indices are not all equal. In Case (a), the summation in (C.2) is

$$\begin{aligned}&n^{-6}\epsilon^{-4}\sum_{i_1=1}^n\sum_{i_2=1}^{\lceil ns \rfloor}\sum_{i_4=1}^{\lceil ns \rfloor}\sum_{i_6=1}^{\lceil ns \rfloor}\sum_{i_8=1}^{\lceil ns \rfloor} E^*[\phi_{i_1i_2}\phi_{i_1i_4}\phi_{i_1i_6}\phi_{i_1i_8}]\\ &= n^{-6}\epsilon^{-4}\sum_{i_1=1}^n\sum_{i_2=1}^{\lceil ns \rfloor}\sum_{i_4=1}^{\lceil ns \rfloor}\sum_{i_6=1}^{\lceil ns \rfloor}\sum_{i_8=1}^{\lceil ns \rfloor} \overline{\widehat{U}}_{i_1}^4 w^4(X_{i_1})[K_{h,i_2i_1} - f_{n,h}(X_{i_1})]\\ &\quad \times [K_{h,i_4i_1} - f_{n,h}(X_{i_1})][K_{h,i_6i_1} - f_{n,h}(X_{i_1})][K_{h,i_8i_1} - f_{n,h}(X_{i_1})]\\ &\leq C\left(n^{-1}\left(n^{-1/2}h^{-d/2}\sqrt{\log n} + h^r\right)^4\right) n^{-1}\sum_{i_1=1}^n \overline{\widehat{U}}_{i_1}^4 w^4(X_{i_1})\\ &= O_p\left(n^{-1}\left(n^{-1/2}h^{-d/2}\sqrt{\log n} + h^r\right)^4\right) = o_p(n^{-1}),\end{aligned}$$

where the first inequality follows from the fact that for any $s \in [0,1]$, $\sup_{x\in\mathbb{R}^d}|\frac{1}{\lceil ns \rfloor}\sum_{i=1}^{\lceil ns \rfloor}[K_h(x - X_i) - f_{n,h}(x)]| \leq \sup_{x\in\mathbb{R}^d}|\frac{1}{\lceil ns \rfloor}\sum_{i_2=1}^{\lceil ns \rfloor} K_h(x - X_i) - \overline{f}_n(x)| + \sup_{x\in\mathbb{R}^d}\left|f_{n,h}(x) - \overline{f}_n(x)\right| = O_p(n^{-1/2}h^{-d/2}\sqrt{\log n} + h^r)$ by Lemmas D.4 and D.6. Similarly, in Case (b), the summation in (C.2) is of the order

$$\begin{aligned}&n^{-6}\epsilon^{-4}\sum_{i_1=1}^n\sum_{i_2=1}^{\lceil ns \rfloor}\sum_{i_4=1}^{\lceil ns \rfloor}\sum_{i_5=1}^n\sum_{i_6=1}^{\lceil ns \rfloor}\sum_{i_8=1}^{\lceil ns \rfloor} E^*[\phi_{i_1i_2}\phi_{i_1i_4}\phi_{i_5i_6}\phi_{i_5i_8}]\\ &= n^{-6}\epsilon^{-4}\sum_{i_1=1}^n\sum_{i_2=1}^{l}\sum_{i_4=1}^{l}\sum_{i_5=1}^n\sum_{i_6=1}^{l}\sum_{i_8=1}^{l} \overline{\widehat{U}}_{i_1}^2 w^2(X_{i_1})[K_{h,i_2i_1} - f_{n,h}(X_{i_1})][K_{h,i_4i_1} - f_{n,h}(X_{i_1})]\\ &\quad \times \overline{\widehat{U}}_{i_5}^2 w^2(X_{i_5})[K_{h,i_6i_5} - f_{n,h}(X_{i_5})][K_{h,i_8i_5} - f_{n,h}(X_{i_5})]\\ &= O_p\left(\left(n^{-1/2}h^{-d/2}\sqrt{\log n} + h^r\right)^4\right) = o_p(n^{-1}).\end{aligned}$$

Consequently $P^*\left(\sup_{0\leq s\leq 1}|A_{n2}^*(s)| \geq \epsilon\right) \leq \sum_{1\leq l\leq n} P^*(A_{n2}^*(l/n) \geq \epsilon) = o_p(1)$ and $\sup_{0\leq s\leq 1}|A_{n2}^*(s)| = o_{p^*}(1)$ by the Markov inequality.

It is straightforward to $A_{n5}^*(s) = O_{p^*}\left(n^{-1/2}\right) = o_{p^*}(1)$ uniformly in s.

To show (iii), it suffices to show that $\{E^*[m^*_{n,h}(x)] - m_{n,h_0}(x)\} f_{n,h}(x) w(x) = O_p(h^r d_n^{-(1+r)} \sqrt{\log n})$ uniformly in x since then $\sup_{0 \le s \le 1} |R^*_n(s)| \le n^{-1/2} \sum_{i=1}^n \left| \{E^*[m^*_{n,h}(X_i)] - m_{n,h_0}(X_i)\} f_{n,h}(X_i) w(X_i) \right| = O_p(n^{1/2} h^r d_n^{-(1+r)} \sqrt{\log n}) = o_p(1) = o_{p^*}(1)$. Standard bias calculation in the Nadaraya-Watson estimation shows that uniformly in x,

$$
\begin{aligned}
&\left| \left\{ E^* \left[m^*_{n,h}(x) \right] - m_{n,h_0}(x) \right\} f_{n,h}(x) w(x) \right| \\
&\simeq \left| h^r n^{-1} \sum_{j=1}^n \sum_{l=1}^d \frac{\partial^r m_{n,h_0}(x)}{\partial x_l^r} f_j(x) \int K(u) u_l^r du \; w(x) \right| \\
&\le C h^r \overline{f}_n(x) \sum_{l=1}^d \sup_{x \in \mathbb{R}^d} \left| \frac{\partial^r m_{n,h_0}(x)}{\partial x_l^r} \right| 1 \left(f_{n,h_0}(x) \ge d_n \right) \\
&= h^r \overline{f}_n(x) O_p \left(n^{-1/2} h_0^{-d/2-r} d_n^{-2-r} \sqrt{\log n} + d_n^{-(1+r)} \right) = O_p \left(h^r d_n^{-(1+r)} \sqrt{\log n} \right),
\end{aligned}
$$

by Lemma D.6 and equation (D.8). (Note that the dominating terms in the summation $\sum_{|\lambda|=r} D^\lambda [m_{n,h_0}(x) f_j(x)]$ are $\sum_{|\lambda|=r} D^\lambda m_{n,h_0}(x) \times f_j(x)$.) Now,

$$
\begin{aligned}
\widehat{\sigma}^{*2} &= n^{-1} \sum_{i=1}^n \widehat{V}_i^{*2} = n^{-1} \sum_{i=1}^n \left[U_i^* - \left(m^*_{n,h}(X_i) - m_{n,h_0}(X_i) \right) \right]^2 f^2_{n,h}(X_i) w^2(X_i) \\
&= n^{-1} \sum_{i=1}^n \eta_i^2 \overline{\widehat{U}}_i^2 f^2_{n,h}(X_i) w^2(X_i) + n^{-1} \sum_{i=1}^n \left(m^*_{n,h}(X_i) - m_{n,h_0}(X_i) \right)^2 f^2_{n,h}(X_i) w^2(X_i) \\
&\quad + n^{-1} \sum_{i=1}^n 2 \eta_i \overline{\widehat{U}}_i \left(m^*_{n,h}(X_i) - m_{n,h_0}(X_i) \right) f^2_{n,h}(X_i) w^2(X_i) \equiv T_{n1} + T_{n2} + T_{n3}.
\end{aligned}
$$

It is easy to show that $T_{n1} = n^{-1} \sum_{i=1}^n \widehat{U}_i^2 f^2_{n,h}(X_i) w^2(X_i) + o_p(1) = \sigma^{*2} + o_p(1)$, and $T_{n2} = o_{p^*}(1)$. Hence $T_{n3} = o_{p^*}(1)$ by the Cauchy-Schwartz inequality and $\widehat{\sigma}^{*2} = \sigma^{*2} + o_{p^*}(1)$.

D Some technical lemmas

This appendix presents some technical lemmas that are used in proving the main results.

Lemma D.1. *Let $\{\xi_i\}$ be a strong mixing process with the mixing coefficient $\alpha(i)$. For any integer $p > 1$ and integers $(i_1, \ldots, i_p)$ such that $1 \le i_1 < i_2 < \cdots < i_p$, let g be a Borel function such that $\int |g(x_1, \ldots, x_p)|^{1+\vartheta} \times dF^{(1)}(x_1, \ldots, x_j) dF^{(2)}(x_{j+1}, \ldots, x_p) \le M$ for some $\vartheta > 0$ and $M > 0$, where $F^{(1)} = F_{i_1, \ldots, i_j}$ and $F^{(2)} = F_{i_j+1, \ldots, i_p}$ are the distribution functions of $(\xi_{i_1}, \ldots, \xi_{i_j})$ and $(\xi_{i_{j+1}}, \ldots, \xi_{i_p})$, respectively. Let F denote the distribution function of $(\xi_{i_1}, \ldots, \xi_{i_p})$. Then $|\int g(x_1, \ldots, x_p) dF(x_1, \ldots, x_p) - \int g(x_1, \ldots, x_p) dF^{(1)}(x_1, \ldots, x_j) dF^{(2)}(x_{j+1}, \ldots, x_p)| \le 4M^{1/(1+\vartheta)} \alpha(i_{j+1} - i_j)^{\vartheta/(1+\vartheta)}$.*

Proof. This is Lemma 2.1 of Sun and Chiang (1997). □

The following definitions are adopted from Robinson (1988).

Definition D.2. $\mathcal{K}_r, r \ge 2$, is the class of even functions K that is a product of a univariate kernel function k satisfying $\int_{\mathbb{R}} u^i k(u) du = \delta_{i0}$ $(i = 0, 1, \ldots, r-1)$, $\int_{\mathbb{R}} u^r k(u) du \ne 0$, and $k(u) = O((1 + |u|^{r+1+\epsilon})^{-1})$ for some $\epsilon > 0$, where δ_{ij} is the Kronecker's delta.

Definition D.3. $\mathcal{G}^\gamma_\mu, \gamma > 0, \mu > 0$, is the class of functions $g : \mathbb{R}^d \to \mathbb{R}$ satisfying: g is $(m-1)$-times partially differentiable, for $m - 1 \le \mu \le m$; for some $\rho > 0$, $\sup_{y \in \phi_{z\rho}} |g(y) - g(z)| / |y - z|^\mu \le D_g(z)$ for all z, where $\phi_{z\rho} = \{y : |y - z| < \rho\}$; $Q_g = 0$ when $m = 1$; Q_g is a $(m-1)$th degree homogeneous polynomial in $y - z$ with coefficients the partial derivatives of g at z of orders 1 through $m - 1$ when $m > 1$; and $g(z)$, its partial derivatives of order $m - 1$ and less, and $D_g(z)$, have finite γth moments.

Lemma D.4. *Denote the density function of X_j as f_j. Suppose $K \in \mathcal{K}_r$, $f_j \in \mathcal{G}^\gamma_r$, and $m \in \mathcal{G}^\gamma_r$ where $r \ge 2$ is an integer. Let $x \in \mathbb{R}^d$, and $h \to 0$ as $n \to \infty$. Then*

(i) $\left| E \left[K((X_j - x)/h) - h^d f_j(x) \right] \right| \le h^{d+r} D_{f_j}(x)$, *uniformly in* x,
(ii) $|E\{[m(X_j) - m(x)] K((X_j - x)/h)\}| \le h^{d+r} D_{m,j}(x)$, *uniformly in* x,

where both $D_{f_j}(\cdot)$ and $D_{m,j}(\cdot)$ have finite γth moments.

Proof. See Lemmas 4–5 of Robinson (1988). □

To apply Lemma D.4, we will suppress the dependence of $D_{f_j}(\cdot)$ and $D_{m,j}(\cdot)$ on $j \in \{1, 2, \ldots, n\}$ by assuming that they are dominated respectively by the functions $D_f(\cdot)$ and $D_m(\cdot)$ that have finite γth moments.

Lemma D.5. *Let $\{\xi_t \in \mathbb{R}^q, t = 1, 2, \ldots\}$ be a strong mixing process, not necessarily stationary, with the mixing coefficient $\alpha(t)$ satisfying $\sum_{t=1}^{\infty} \alpha(t) < \infty$. Suppose that $\varsigma_n : \mathbb{R}^q \to \mathbb{R}$ is a measurable function such that $E[\varsigma_n(\xi_t)] = 0$, and $|\varsigma_n(\xi_t)| \leq M_n$ for every $t = 1, 2, \ldots$, then for any $\epsilon > 0$,*

$$P\left(\left|n^{-1}\sum_{i=1}^{n} \varsigma_n(\xi_t)\right| > \epsilon\right) \leq C_0 \exp\left(-\frac{np_n\epsilon^2}{C_1\sigma^2(p_n) + C_2M_np_n(p_n+1)\epsilon}\right) + C_3\sqrt{\frac{M_n}{\epsilon}}\frac{n}{p_n}\alpha(p_n+1),$$

where $1 \leq p_n \leq n/2$, $\sigma^2(p_n) = \sup_{1\leq j\leq 2p_n} \max\left\{\sigma^2_{j,p_n}, \sigma^2_{j,p_n+1}\right\}$, $\sigma^2_{j,p_n} = E\left[\sum_{t=1}^{p_n} \varsigma_n(\xi_{j+t})\right]^2$, and $C_i's$, $i = 0, 1, 2, 3$, are constants that do not depend on n, ϵ, M_n, and p_n.

Proof. See Lemma 5.2 in Shen and Huang (1998). □

Lemma D.6. *Let $\mathcal{K} : \mathbb{R}^d \to \mathbb{R}$ be a kernel-like function with a compact support $\mathcal{U}$ satisfying $\sup_{u\in\mathcal{U}} |\mathcal{K}(u)| \leq \overline{C}_1 < \infty$, $\mathcal{K}(u) = 0$ for $\|u\| \geq \overline{C}_2$, and $|\mathcal{K}(u) - \mathcal{K}(u')| \leq \overline{C}_3 \|u - u'\|$ for any $u, u' \in \mathbb{R}^d$. Suppose that $\{X_i, Y_i\}$ is a strong mixing process satisfying the mixing, moment and tail conditions in Assumptions A1 and A8. Then $\sup_{x\in\mathbb{R}^d} |\Psi(x) - E\Psi(x)| = O_p(n^{-1/2}h_0^{-d/2}\sqrt{\log n})$, where $\Psi(x) = n^{-1}h_0^{-d}\sum_{i=1}^{n} Y_i\mathcal{K}((x - X_i)/h_0)$.*

Remark. Andrews (1995, Lemma A-2) obtains a slower rate of convergence under more general conditions that allow for near-epoch-dependent arrays. Applying Lemma A-2 of Andrews with $\eta = \infty$ (see the remark on p. 569 of Andrews, 1995) gives us $\sup_{x\in\mathbb{R}^d} |\Psi(x) - E\Psi(x)| = O_p\left(n^{-1/2}h_0^{-d}\right)$. Hansen (2008, Theorem 4) proves the result for a stationary strong mixing process. Since we don't assume stationarity of $\{Y_i, X_i\}$, we need to modify the proof of Hansen (2008) by using Lemma D.5.

Proof. The proof follows closely from that of Theorems 2 and 4 of Hansen (2008). We only outline the difference here. Let $c_n = O((\log n)^{1/d} n^{1/(2q)})$ for some $q \geq d$, we first show that

$$\sup_{\|x\|\leq c_n} |\Psi(x) - E\Psi(x)| = O_p\left(n^{-1/2}h_0^{-d/2}\sqrt{\log n}\right). \tag{D.1}$$

Let $a_n = n^{-1/2}h_0^{-d/2}\sqrt{\log n}$ and $\tau_n = a_n^{-1/(3+\delta)}$. Similarly to Hansen (2008), we can show that we can replace Y_i with the truncated process $Y_i 1(|Y_i| \leq \tau_n)$ by Assumption A8. So in the following we simply assume that $|Y_i| \leq \tau_n$. By selecting $N = O\left(c_n^d h_0^{-d} a_n^{-d}\right)$ grid points $x_1, \ldots, x_N$, we can cover the region $\{x : \|x\| \leq c_n\}$ by $A_j = \{x : \|x - x_j\| \leq a_n h_0\}$, $j = 1, \ldots, N$. By the triangle inequality

$$\begin{aligned} a_n^{-1} \sup_{\|x\|\leq c_n} |\Psi(x) - E\Psi(x)| \leq{}& a_n^{-1} \sup_{1\leq j\leq N, x\in A_j} |\Psi(x) - \Psi(x_j)| + a_n^{-1} \sup_{1\leq j\leq N} |\Psi(x_j) - E\Psi(x_j)| \\ &+ a_n^{-1} \sup_{1\leq j\leq N, x\in A_j} |E\Psi(x_j) - E\Psi(x)| \equiv T_{n1} + T_{n2} + T_{n3}. \end{aligned} \tag{D.2}$$

It suffices to show that $T_{ni} = O_p(1)$ for $i = 1, 2, 3$. By the Lipschitz continuity of $\mathcal{K}(\cdot)$ and the fact that $\|x - x_j\| \leq a_n h_0$ for all $x \in A_j$, we have

$$T_{n3} \leq \sup_{1\leq j\leq N} n^{-1}h_0^{-d} \sum_{i=1}^{n} E\left|Y_i\mathcal{K}^*\left(\frac{x_j - X_i}{h_0}\right)\right| = O(1), \tag{D.3}$$

and

$$T_{n1} \leq \sup_{1\leq j\leq N} n^{-1}h_0^{-d} \sum_{i=1}^{n} \left|Y_i\mathcal{K}^*\left(\frac{x_j - X_i}{h_0}\right)\right| = O_p(1), \tag{D.4}$$

where $\mathcal{K}^*(u) = \overline{C}_3 1\left(\|u\| \leq 2\overline{C}_2\right)$ is bounded and integrable.

For any $M>0$,

$$\text{(D.5)}\qquad P\left(a_n^{-1}\sup_{1\le j\le N}|\Psi(x_j)-E\Psi(x_j)|>M\right)\le N\max_{1\le j\le N}P\left(a_n^{-1}|\Psi(x_j)-E\Psi(x_j)|>M\right).$$

Define $\xi_{ni}(x)=Y_i\mathcal{K}((x-X_i)/h_0)-E[Y_i\mathcal{K}((x-X_i)/h_0)]$. Then $\sup_x E\left[\sum_{i=1}^{p_n}\xi_{ni}(x)\right]^2\le Cp_nh_0^d$ and $|\xi_{ni}(x)|\le 2\overline{C}_1\tau_n$. We apply Lemma D.5 with $M_n=2\overline{C}_1\tau_n$, $p_n=a_n^{-1}\tau_n^{-1}$, $\epsilon=Ma_nh_0^d$ and $\sigma^2(p_n)=Cp_nh_0^d$ to obtain

$$\begin{aligned}
\text{(D.6)}\qquad & N\max_{1\le j\le N}P\left(a_n^{-1}|\Psi(x_j)-E\Psi(x_j)|>M\right)\\
&=N\max_{1\le j\le N}P\left(\left|n^{-1}\sum_{i=1}^{n}\xi_{ni}(x_j)\right|>Ma_nh_0^d\right)\\
&\le NC_0\exp\left(-\frac{nM^2a_n^2h_0^{2d}}{(C_1Ch_0^d+2\overline{C}_1C_2Mh_0^d)}\right)+NC_3\sqrt{\frac{2\overline{C}_1\tau_n}{Ma_nh_0^d}}na_n\tau_n\alpha\left((a_n\tau_n)^{-1}\right)\\
&\le NC_0\exp\left(-\frac{M\log n}{(C_1C+2\overline{C}_1C_2)}\right)+c_n^dh_0^{-d}nh_0^{-d/2}O\big(a_n^{\beta+1/2-d}\tau_n^{\beta+3/2}\big)\\
&\le NC_0\exp\left(-\frac{M\log n}{(C_1C+2\overline{C}_1C_2)}\right)+n^{3/2}\big(c_n^dh_0^{-d}\big)\big(n^{-1/2}h_0^{-d/2}\big)O\big(a_n^{\beta+1/2-d}\tau_n^{\beta+3/2}\big)\\
&\le o(n^{(d/2q)+\theta+d(1-\theta)/2-M/\left(C_1C+2\overline{C}_1C_2\right)})+o\big(n^{\tau/2}\big)\\
&=o(1)\ \text{ by Assumption A8(ii) for sufficiently large } M,
\end{aligned}$$

where

$$\tau=\frac{d}{q}+3+2\theta-\frac{1-\theta}{2}\left(\frac{(2\beta+3)(\delta+2)}{\delta+3}-2d\right),$$

and the last inequality follows from the facts that $c_n^dh_0^{-d}=o\left(n^{(d/2q)+\theta}\right)$, $a_n=\left(n^{-1}h_0^{-d}\log n\right)^{1/2}=o\left(n^{-(1-\theta)/2}\right)$, and $\tau_n=a_n^{-1/(\delta+3)}=o\left(n^{(1-\theta)/[2(\delta+3)]}\right)$. Then (D.1) follows from (D.2)–(D.6).

Now define $\widetilde{\Psi}(x)=n^{-1}h_0^{-d}\sum_{i=1}^{n}Y_i\mathcal{K}((x-X_i)/h_0)1\times(\|X_i\|\le c_n/2)$. Noting that $c_n^{-q}=O(a_n)$, by Assumption A8(i) and the conditions on $\mathcal{K}$, we can follow the proof of Theorem 4 of Hansen (2008) to show that $\sup_x|\Psi(x)-E\Psi(x)|=\sup_x|\widetilde{\Psi}(x)-E\widetilde{\Psi}(x)|+O_p(a_n)$. (D.1) implies that $\sup_{\|x\|\le c_n}|\widetilde{\Psi}(x)-E\widetilde{\Psi}(x)|=O_p(a_n)$. So we are left to show that $\sup_{\|x\|>c_n}|\widetilde{\Psi}(x)-E\widetilde{\Psi}(x)|=O_p(a_n)$. Since $c_n\to\infty$, $c_n>2\overline{C}_2$ for sufficiently large n. On the set $\{\|x\|>c_n,\|X_i\|\le c_n/2\}$, $\|x-X_i\|\ge c_n/2>\overline{C}_2$ and thus $\mathcal{K}((x-X_i)/h_0)=0$ as $h_0<1$ for sufficiently large n. Consequently, $\sup_{\|x\|>c_n}|\widetilde{\Psi}(x)|\le n^{-1}h_0^{-d}\sum_{i=1}^{n}|Y_i|\sup_{\|x\|>c_n}\mathcal{K}((x-X_i)/h_0)1(\|X_i\|\le c_n/2)=0$. This completes the proof. □

Lemma D.7. *Under the conditions of Theorem 4.1,* $\sup_{x\in S_n}\left|D^\lambda m_{n,h_0}(x)\right|=O_p(n^{-1/2}h_0^{-d/2-|\lambda|}d_n^{-2-|\lambda|}\sqrt{\log n}+d_n^{-(1+|\lambda|)})$, *where* $S_n=\{x\in\mathbb{R}^d:\overline{f}_n(x)\ge d_n\}$ *and* $|\lambda|\le r$.

Proof. The proof follows closely from that of Theorem 1(b) of Andrews (1995). So we only sketch it. Recall $\overline{f}_n(x)=n^{-1}\sum_{i=1}^{n}f_i(x)$ and $\overline{m}_n(x)=n^{-1}\sum_{i=1}^{n}m_i(x)$. Let $\overline{g}_n(x)=\overline{m}_n(x)\overline{f}_n(x)$ and $g_{n,h_0}(x)=m_{n,h_0}(x)f_{n,h_0}(x)$. If $\overline{f}_n(x)\ne 0$, we have

$$\text{(D.7)}\qquad D^\lambda\overline{m}_n(x)=D^\lambda[\overline{g}_n(x)\overline{f}_n^{-1}(x)]=\sum_{0\le\kappa\le\lambda}C_\kappa D^{\lambda-\kappa}\overline{g}_n(x)D^\kappa[\overline{f}_n^{-1}(x)]$$

for some positive finite constants C_κ.

Let $S_n^*=\{x\in\mathbb{R}^d:f_{n,h_0}(x)\ge d_n\}$ and $\widetilde{S}_n=S_n\cap S_n^*$. By Assumption A9(i), $\sup_{x\in\widetilde{S}_n}\left|D^\kappa\overline{f}_n^{-1}(x)\right|=O_p(d_n^{-(1+|\kappa|)})\forall\kappa\le\lambda$. Noting that $D^\kappa f_{n,h_0}(x)=n^{-1}h_0^{-(|\kappa|+d)}\sum_{i=1}^{n}D^\kappa[K(x-X_i)/h_0]$ we can apply Lemma D.6 with $\mathcal{K}(u)=D^\kappa K(u)$ and $Y_i=1$ to obtain

$$\text{(D.8)} \qquad \sup_{x\in\mathbb{R}^d} \left|D^\kappa f_{n,h_0}(x) - D^\kappa \overline{f}_n(x)\right| \le \sup_{x\in\mathbb{R}^d} \left|D^\kappa f_{n,h_0}(x) - E\left[D^\kappa f_{n,h_0}(x)\right]\right| + \sup_{x\in\mathbb{R}^d} \left|E\left[D^\kappa f_{n,h_0}(x)\right] - D^\kappa \overline{f}_n(x)\right|$$
$$= O_p\left(n^{-1/2} h_0^{-d/2-|\kappa|}\sqrt{\log n}\right) + O\left(h_0^r\right) \ \forall \kappa \le \lambda.$$

Hence

$$\text{(D.9)} \qquad \sup_{x\in\mathbb{R}^d} \left|D^\kappa f_{n,h_0}(x)\right| = O_p\left(n^{-1/2} h_0^{-d/2-|\kappa|}\sqrt{\log n} + 1\right).$$

Some further calculations show that

$$\text{(D.10)} \qquad \sup_{x\in\widetilde{S}_n} \left|D^\kappa [f_{n,h_0}^{-1}(x)] - D^\kappa \overline{f}_n^{-1}(x)\right| = \sum_{j=0}^{|\kappa|}\sum_{|\iota|=j} O_p\left(d_n^{-(2+|\kappa|-j)}\right) \sup_{x\in\mathbb{R}^d}\left|D^\iota f_{n,h_0}(x) - D^\iota \overline{f}_n(x)\right|$$
$$= \sum_{j=0}^{|\kappa|} O_p\left(d_n^{-(2+|\kappa|-j)} n^{-1/2} h_0^{-d/2-j}\sqrt{\log n}\right).$$

Similarly,

$$\text{(D.11)} \qquad \sup_{x\in\mathbb{R}^d} \left|D^{\lambda-\kappa} g_{n,h_0}(x) - D^{\lambda-\kappa}\overline{g}_n(x)\right|$$
$$\le \sup_{x\in\mathbb{R}^d} \left|D^{\lambda-\kappa} g_{n,h_0}(x) - E\left[D^{\lambda-\kappa} g_{n,h_0}(x)\right]\right| + \sup_{x\in\mathbb{R}^d}\left|E\left[D^{\lambda-\kappa} g_{n,h_0}(x)\right] - D^{\lambda-\kappa}\overline{g}_n(x)\right|$$
$$= O_p\left(n^{-1/2} h_0^{-d/2-|\lambda-\kappa|}\sqrt{\log n}\right) + O\left(h_0^r\right) \quad \forall \kappa \le \lambda.$$

Combining (D.7)–(D.11), we can show that $\sup_{x\in\widetilde{S}_n} \left|D^\lambda m_{n,h_0}(x)\right| \le \sup_{x\in\widetilde{S}_n}\left|D^\lambda m_{n,h_0}(x) - D^\lambda \overline{m}_n(x)\right| + \sup_{x\in\widetilde{S}_n}\left|D^\lambda \overline{m}_n(x)\right| = O_p(n^{-1/2}h_0^{-d/2-|\lambda|}d_n^{-2-|\lambda|}\sqrt{\log n} + d_n^{-(1+|\lambda|)})$. The result follows because (D.8) implies that $d_n^{-1}\sup_{x\in\mathbb{R}^d}\left|f_{n,h_0}(x) - \overline{f}_n(x)\right| = O_p(n^{-1/2}h_0^{-d/2}\sqrt{\log n}d_n^{-1} + h_0^r d_n^{-1}) = o_p(1)$ and we can replace $\widetilde{S}_n$ by S_n with probability approaching 1. ☐

ACKNOWLEDGEMENTS

The authors thank Jae-Young Kim, Peter C. B. Phillips, Peter Robinson, Yoon-Jae Whang, and an anonymous referee for their many constructive comments on an early version of the paper. The first author gratefully acknowledges the financial support from the NSFC under grant numbers 70501001 and 70601001.

Received 28 February 2008

REFERENCES

[1] Anderson, T. W., and Darling, D. A. (1952). Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *Annals of Mathematical Statistics* **23** 193–212. MR0050238

[2] Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

[3] Andrews, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory* **11** 560–596. MR1349935

[4] Bosq, D. (1991). Nonparametric prediction for unbounded almost stationary processes. In G. Roussas (ed.) *Nonparametric Functional Estimation and Related Topics*, NATO, 389–403. MR1154341

[5] Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer, New York. MR1441072

[6] Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B* **37** 149–163. MR0378310

[7] Chen, G., Choi, Y. K., and Zhou, Y. (2005). Nonparametric estimation of structural change points in volatility models for time series. *Journal of Econometrics* **126** 79–114. MR2118279

[8] Chu, J. S. and Wu, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. *Annals of Statistics* **21** 1545–1566. MR1241278

[9] Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press. MR1430804

[10] Delgado, M. A. and Hidalgo, J. (2000). Nonparametric inference on structural breaks. *Journal of Econometrics* **96** 113–144. MR1768755

[11] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Models*. Springer-Verlag. MR1964455

[12] Fan, Y. and Linton, O. (2003). Some higher order theory for a consistent nonparametric model specification test. *Journal of Statistical Planning and Inference* **109** 125–154. MR1946644

[13] Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Annals of Probability* **18** 851–869. MR1055437

[14] Hafner, C. M. and Herwartz, H. (2000). Testing for linear autoregressive dynamics under heteroskedasticity. *Econometrics Journal* **3** 177–197. MR1820413

[15] Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24** 726–748. MR2409261

[16] Härdle, W. and J. S. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics* **19** 778–796. MR1105844

[17] Herrndorf, N. (1985). A functional central limit theorem for strongly mixing sequences of random variables. *Z. Wahrscheinlichkeitstheorie verw. Grbiete* **69** 541–550. MR0791910

[18] Hidalgo, J. (1995). A nonparametric conditional moment test for structural stability. *Econometric Theory* **11** 671–698. MR1363629

[19] Kim, J. H. and Hart, J. D. (1998). Tests for change in mean function when the data are dependent. *Journal of Time Series Analysis* **19** 399–424. MR1652191

[20] Krämer, W., Ploberger, W., and Alt, R. (1988). Testing for structural change in dynamic models. *Econometrica* **56** 1355–1369. MR0973984

[21] Kreiss, J. P. (1997). Asymptotic properties of residual bootstrap for autoregressions. *Manuscript*. Institute for Mathematical Statistics, Technical University of Braunschweig, Germany.

[22] Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications* **83** 79–102. MR1705601

[23] Lee, S. (2003). Efficient semiparametric estimation of partially linear quantile regression model. *Econometric Theory* **19** 1–31. MR1965840

[24] Lee, S. and Park, S. (2001). The CUSUM of squares test for scale change in infinite order moving average processes. *Scandinavian Journal of Statistics* **28** 625–644. MR1876504

[25] Li, Q. (1999). Consistent model specification tests for time series econometric models. *Journal of Econometrics* **92** 101–147. MR1706996

[26] Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications* **65** 69–80. MR1422880

[27] Liu, R. Y. (1988). Bootstrap procedure under some non-i.i.d. models. *Annals of Statistics* **16** 1696–1708. MR0964947

[28] Loader, C. R. (1996). Change point estimation using nonparametric regression. *Annals of Statistics* **24** 1667–1678. MR1416655

[29] Müller, H.-G. (1992). Change points in nonparametric regression analysis. *Annals of Statistics* **20** 737–761. MR1165590

[30] Neumann, M. H. and Paparoditis, E. (2000). On bootstrapping L_2-type statistics in density testing. *Statistics & Probability Letters* **50** 137–147. MR1791508

[31] Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **42** 100–115. MR0088850

[32] Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* **57** 1361–1401.

[33] Ploberger, W. and Krämer, W. (1992). The CUSUM test with OLS residuals. *Econometrica* **56** 1355–1369. MR1162619

[34] Ploberger W. and Krämer, W. (1996). A trend resistant test for structural change based on ols residuals. *Journal of Econometrics* **70** 175–186.

[35] Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis* **4** 185–207. MR0732897

[36] Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762

[37] Robinson, P. M. (1991). Consistent nonparametric entropy-based testing. *Review of Economic Studies* **58** 437–453. MR1108130

[38] Roussas, G. (1988). Nonparametric estimation in mixing sequences of random variables. *Journal of Statistical Planning and Inference* **18** 135–149. MR0922204

[39] Shen, J. and Huang, Y.-M. (1998). Nonparametric estimation in a model with a trend. *Statistical Inference for Stochastic Processes* **1** 43–60. MR1670224

[40] Sun, S. and Chiang, C.-Y. (1997). Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables. *Journal of Applied Mathematics and Stochastic Analysis* **10** 3–20. MR1437948

[41] Sun, Y., Phillips, P. C. B., and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* **76** 175–194. MR2374985

[42] White, H. (2001). *Asymptotic Theory for Econometricians*. 2nd edition. Academic Press, San Diego.

[43] Wichern, D. W., Miller, R. B., and Hsu, D. A. (1976). Changes of variance in first-order autoregressive time series models – with an application. *Applied Statistics* **25** 248–256.

[44] Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* **14** 1261–1295. MR0868303

[45] Xiao, Z. (1999). A residual based test for the null hypothesis of cointegration. *Economics Letters* **64** 133–141. MR1705435

Liangjun Su
Guanghua School of Management
Peking University
Beijing 100871, China
E-mail address: lsu@gsm.pku.edu.cn

and

School of Economics
Singapore Management University
90 Stamford Road
Singapore 178903
E-mail address: ljsu@smu.edu.sg

Zhijie Xiao
Department of Economics
Boston College
Chestnut Hill
MA 02467, USA
E-mail address: xiaoz@bc.edu