A Note on Nonparametric Conditional Moment Tests for Structural Stability

Toshio Honda

Abstract

This paper considers the asymptotic theory for the conditional moment tests for structural stability proposed by Hidalgo(1995). The proofs of Hidalgo(1995) are difficult to follow and some implicit assumptions are made in the proofs. In addition the explicit and implicit assumptions are inconsistent. We reconsider the test statistic and give a correct proof of the weak convergence of the test statistic. The result here allows for dynamic models and dependent errors.

Key words: structural stability, conditional moment test, Nadaraya-Watson kernel estimator, Brownian Bridge, uniformly mixing process, exponential inequality

1 Introduction

This paper considers conditional moment tests of structural stability which are proposed by Hidalgo(1995) using the Nadaraya-Watson nonparametric regression estimators. The result here is also closely related to Ploberger and Krämer(1992). It seems rather difficult to specify the maintained hypothesis in the cases where a structural change really exists. So nonparametric approaches might be more useful than parametric approaches in some cases.

Since the pioneering work by Chow(1960) and Quandt(1960), an extensive statistical and econometric literature on change point problems has developed. Chow(1960) considered the problem of structural stability under the alternative of one known time of structural change. But many authors have pointed out that knowledge of the possible time of structural change is the exception rather than the rule. If one estimates the possible time of structural change by inspection to test structural stability, the procedure is invalid from a statistical point of view.
For the cases of unspecified change points, a large number of procedures for testing structural stability have been proposed. For example, Brown et al. (1975) proposed the CUSUM test for stability in the coefficients in a linear regression, which is based on the cumulative sum of recursively computed regression residuals, and Krämer et al. (1988) considered the CUSUM test for dynamic models. Ploberger and Krämer (1992) extended the CUSUM test to the OLS regression residuals. See also Ploberger and Krämer (1996). Some standard textbooks on econometrics, for example, Greene (1993), contain an exposition on the CUSUM tests. Andrews (1993) constructed LR, LM, and Wald type tests of structural stability using GMM and examined the asymptotic properties under the general conditions. Likelihood ratio test statistics for changes in the parameters of the normal population are investigated in Horváth (1993). As for the estimation of the change point, Bai (1995) deals with the estimation of the change point in the regression coefficients by LAD (least absolute deviation) and obtains the asymptotic distribution of the estimator of the change point.

Relevant references on the these subjects, testing of structural stability, estimation of the change points, and so on, are given in the above papers. Lai (1995) contains a brief survey of sequential change point detection procedures. See also Carlstein et al. (1995).

In the rest of this section we mention on the results of Hidalgo (1995) because our results are closely related to them. Let \( \{ (X_t', \epsilon_t)' \} , \ t = 1, 2, \ldots , \) be a \((P+1)\) dimensional stationary stochastic process and we observe \( \{ (Y_t, X_t')' \} \) such that

\[
Y_t = g_t(X_t) + \epsilon_t , \quad t = 1, 2, \ldots , T ,
\]

where \( \epsilon_t \) is an error term with conditional mean 0 with respect to \( X_t \), and \( g_t(x) = g(t/T, x) \). Hidalgo (1995) proposed some nonparametric conditional moment tests of the null hypothesis \( H_0 : g(t/T, x) = g(x) \). Those tests are constructed from

\[
U_\tau = \frac{1}{\sqrt{T}} \sum_{t=1}^{[\tau T]} \tilde{g}_t(X_t) \tilde{f}_t(X_t) \tilde{f}(X_t)(Y_t - \hat{g}(X_t)) ,
\]

where \( 0 \leq \tau \leq 1 , \tilde{g}_t \) and \( \hat{g} \) are the Nadaraya-Watson kernel estimators of \( g_t , \tilde{f}_t \) and \( \tilde{f} \) are the Nadaraya-Watson kernel estimators of \( f \), the density function of \( X_t \). \( \tilde{g}_t \) is the estimator of \( g_t \) under the alternative. As is well known, \( \tilde{f}_t \) and \( \tilde{f} \) are the denominators of \( \tilde{g}_t \) and \( \hat{g} \) respectively.
A Note on Nonparametric Conditional Moment Tests

$U_\tau$ is devised to do without trimming. Trimming is necessary to avoid small denominators. But we think that $\hat{f}$ gives a weight which depends on the precision of $\hat{g}(X_t)$ and the corresponding residual $\hat{\varepsilon}_t$. When $\hat{f}(X_t)$ is negative or small albeit positive, there are not enough observations to estimate $g(X_t)$ and the corresponding $\hat{g}(X_t)$ and $\hat{\varepsilon}_t$ should be discarded. But it is hard to evaluate the effects of trimming mathematically. $U_\tau$ trims the inappropriate estimates in a natural way.

The weak convergence of $U_\tau$ to Brownian bridge is proved in Hidalgo (1995). But the proofs are difficult to follow. The assumptions are complicated and inconsistent. For example, the proof of Lemma 7 implicitly assumes the compact support of the kernel. This is contradictory to Assumption A.7. In addition the proofs implicitly assume independence of the regressors and errors and seem to pay no attention to uniformity in $\tau$.

In section 2 we prove the result on the weak distribution of

$$ S_\tau = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \hat{g}(X_t)\hat{f}(X_t)\hat{f}(X_t)(Y_t - \hat{g}(X_t)), $$

with less restrictive assumptions. Numerical examples of Hidalgo (1995) seem to be calculated using $S_\tau$ instead of $U_\tau$. We carried out simulation studies to compare the conditional moment tests and other tests. The numerical examples are given in section 3.

## 2 Proof of the weak convergence of $S_\tau$

In this section we show that $S_\tau$ in (3) converges weakly to a constant multiple of one-dimensional Brownian Bridge under $H_0$. The proof of Hidalgo (1995) are doubtful, so we present another rather simple proof under consistent assumptions. See Hidalgo (1995) for the ideas of $U_\tau$ and $S_\tau$. Some arguments on the power under local alternatives are also given there.

Now we formally define $S_\tau$. We consider the model (1) under $H_0 : g_0(x) = g(x)$. Letting $K(x)$ be the Kernel for the nonparametric regression estimator and $h$ be the bandwidth, we define $\hat{g}(X_t)\hat{f}(X_t)$ and $\hat{f}(X_t)(Y_t - \hat{g}(X_t))$ by

$$ \hat{g}(X_t)\hat{f}(X_t) = \hat{f}\hat{g}(X_t) $$

and
\begin{align}
\hat{f}(X_t)(Y_t - \hat{g}(X_t)) &= \hat{f}(X_t)Y_t - \hat{g}(X_t),
\end{align}

where
\begin{align}
\hat{f}(X_t) &= \frac{1}{Th^p} \sum_{s=1}^{T} K\left(\frac{X_s - X_t}{h}\right)
\end{align}

and
\begin{align}
\hat{g}(X_t) &= \frac{1}{Th^p} \sum_{s=1}^{T} K\left(\frac{X_s - X_t}{h}\right)Y_t.
\end{align}

Hereafter we write \( K_{st} \) for \( K((X_s - X_t)/h)/(Th^p) \).

The normalized cumulative sum of \( \hat{g}(X_t)\hat{f}(X_t)\hat{f}(X_t)(Y_t - \hat{g}(X_t)) \) is written as
\begin{align}
S_{\tau} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \hat{f}(X_t)\hat{f}(X_t)Y_t - \hat{g}(X_t),
\end{align}

where \([a]\) is the largest integer not exceeding \( a \) and \( \tau \in [0,1] \). We prove the weak convergence of \( S_{\tau} \) to a constant multiple of one-dimensional Brownian bridge in \( D[0,1] \).

Before presenting the theorem and the proof, we state the assumptions. The assumptions and the proof are simpler than those of Hidalgo(1995). Let \( C_i \) be generic positive constants.

**Assumptions**

**A1.** Let \( \{ (X'_t, \epsilon_t)' \} \) be stationary and uniformly mixing with mixing coefficients \( \{ \phi(k) \} \), where \( \phi(k) = O(k^{-a}) (a > 6) \).

**A2.** \( E[\epsilon_t | X_t] = 0 \) a.s., \( E[\epsilon_t^2 | X_t] < C_1 \) a.s., and \( X_t \) has the bounded density function \( f(x) \).

**A3.** \( K(x) \) is symmetric, bounded and continuously differentiable, and there exists a positive integer \( q \) such that \( \int K(u) \prod u_i^s |u| \, du = \delta_0 \), where \( s = \sum_{i=1}^{P} s_i \), \( s_i \geq 0 \), and \( 0 \leq s < q \). Suppose also that \( |K(u)| < C_2 \exp(-C_3|u|) \) and \( \left| \frac{\partial K}{\partial u_i}(u) \right| < C_2 \exp(-C_3|u|) \).

**A4.** \( f, fg, fg^2, f^2g \), and \( (fg)^2 \) belong to \( C^{q+1}(R^p) \) and are positive a.e. on \( R^p \). In addition, the \( (q - 1) \) th derivatives are bounded and the \( q \) th derivatives have the third moment.

**A5.** \( E[|g(X_t)|^3] < \infty \) and \( |g(x)| < C_4(1 + |x|^k) \) \( (k > 0) \).
A6. There exist positive numbers $\gamma_1$, $v_1$, and $r_1$ such that $\gamma_1 > 1/4$, $v_1 > 2r_1$, $E[|X_i|^v] < \infty$, $\frac{1}{T^{1-2\gamma_1}h^P} = O(1)$, and $\frac{T^{1/(1+\sigma)+\gamma_1+\gamma_2}}{Th^P} = O(1)$ for some positive $\gamma_2$. In addition $fg^2$ is bounded.

A7. There exist positive numbers $\gamma_3$, $v_2$, and $r_2$ such that $\gamma_3 > 1/4$, $E[|\epsilon_t|^v] < \infty$, $\frac{1}{T^{(v_2-1)/2-\gamma_3}h^P} = O(1)$, $\frac{1}{T^{1-2\gamma_3}h^P} = O(1)$, and $\frac{T^{1/(1+\sigma)+\gamma_3+\gamma_4}}{Th^P} = O(1)$ for some positive $\gamma_4$.

A8. There exists positive numbers $\gamma_5$ and $M$ such that $E[(f(X_t)\epsilon_t g(X_s))^6] < M$, $E[(f(X_t)\epsilon_t g(X_s))^6] < M$, $E[(f g^2)(X_t)^6] < M$, $E[(f g)(X_t)g(X_s)]^6 < M$, and $E[(f g)(X_t)\epsilon_t^6] < M$.

A9. $\sqrt{T} h^q \to 0$ as $T \to \infty$.

A10. $Th^{4p} \to \infty$ as $T \to \infty$.

The limit distribution of $S\tau$ in $D[0,1]$ under $H_0$ is given in the following theorem. Hereafter $W$ is one-dimensional standard Brownian motion in $D[0,1]$ and $\to$ means weak convergence.

**Theorem 2.1** *Under $H_0$ and Assumptions A1-10,*

$$
\text{if } \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} (f g)(X_t)\epsilon_t \overset{D}{\to} \sqrt{\sigma^2}W(\tau) \text{ in } D[0,1], \text{ then in } D[0,1],
$$

$$
S_{\tau} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} f g(X_t)(f(X_t)Y_t - f g(X_t)) \overset{D}{\to} \sqrt{\sigma^2}(W(\tau) - \tau W(1)).
$$

When \{ $(f^2 g)(X_t)\epsilon_t$ \} is a martingale difference, $\sigma^2$ in (10) can be estimated by

$$
\frac{1}{T} \sum_{t=1}^{T} (f g(X_t)(f(X_t)Y_t - f g(X_t)))^2.
$$

The proof of Theorem 2.1 needs some preparations. We introduce several notations.

$$
E_t\{h(X_s,\epsilon_s, X_t, \epsilon_t)\} = \int h(x,\epsilon, X_t, \epsilon_t)F(dx, dc),
$$

where $F$ is the joint distribution function of $(X'_s,\epsilon_s)'$.
We have from Assumption A4,

\[
\begin{align*}
\hat{f}(X_t) &= f(X_t) + h^q \left\{ \int K(u) \sum_{s_l=q}^{1} \frac{\partial^q f}{\partial x_l^{s_1} \ldots \partial x_P^{s_P}}(X_t) \prod u^{s_i}_i \, du + O(h) \right\} \\
&\quad + \sum_{s=1}^{T} \{ K_{st} - E_t(K_{st}) \} \\
&= f(X_t) + h^q \xi_{1t} + \xi_{2t}
\end{align*}
\]

and

\[
\begin{align*}
\hat{f}(X_t) &= (fg)(X_t) + h^q \left\{ \int K(u) \sum_{s_l=q}^{1} \frac{\partial^q (fg)}{\partial x_l^{s_1} \ldots \partial x_P^{s_P}}(X_t) \prod u^{s_i}_i \, du + O(h) \right\} \\
&\quad + \sum_{s=1}^{T} \{ K_{st}g(X_s) - E_t(K_{st}g(X_s)) \} + \sum_{s=1}^{T} K_{st}\epsilon_s \\
&= (fg)(X_t) + h^q \eta_{1t} + \eta_{2t} + \sum_{s=1}^{T} K_{st}\epsilon_s.
\end{align*}
\]

(13)

By (13) and (14) \( S_t \) is rewritten as

\[
S_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \left\{ (fg)(X_t) + h^q \eta_{1t} + \eta_{2t} + \sum_{s=1}^{T} K_{st}\epsilon_s \right\} \\
\times \left\{ (f(X_t) + h^q \xi_{1t} + \xi_{2t}) (g(x_t) + \epsilon_t) - (fg)(X_t) - h^q \eta_{1t} - \eta_{2t} - \sum_{s=1}^{T} K_{st}\epsilon_s \right\}
\]

\[
= \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \left\{ (fg)(X_t) + h^q \eta_{1t} + \eta_{2t} + \sum_{s=1}^{T} K_{st}\epsilon_s \right\} \\
\times \left\{ f(X_t)\epsilon_t + h^q \xi_{1t}g(X_t) + h^q \xi_{1t}\epsilon_t + \xi_{2t}g(X_t) + \xi_{2t}\epsilon_t - h^q \eta_{1t} - \eta_{2t} - \sum_{s=1}^{T} K_{st}\epsilon_s \right\}
\]

\[
= \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \left\{ (fg)(X_t)\epsilon_t - (fg)(X_t) \sum_{s=1}^{T} K_{st}\epsilon_s \right\} \\
\times \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{[T\tau]} \left\{ h^q \eta_{1}, f(X_t)\epsilon_t + \eta_{2t}f(X_t)\epsilon_t + \left( \sum_{s=1}^{T} K_{st}\epsilon_s \right) f(X_t)\epsilon_t \\
+ h^q (fg)(X_t)\xi_{1t} + h^q \eta_{1t}g(X_t) + h^q \eta_{2l}g(X_t) + \xi_{2t}g(X_t) + \xi_{2t}\epsilon_t \right\} \\
+ h^q (fg)(X_t)\xi_{1t}\epsilon_t + h^q \eta_{1t}\xi_{1t}\epsilon_t + h^q \eta_{2l}\xi_{1t}\epsilon_t + h^q \left( \sum_{s=1}^{T} K_{st}\epsilon_s \right) \xi_{1t}\epsilon_t \\
+ (fg)(X_t)\xi_{2t} + h^q \eta_{1t}g(X_t)\xi_{2t} + g(X_t)\eta_{2t}\xi_{2t} + g(X_t)\left( \sum_{s=1}^{T} K_{st}\epsilon_s \right) \xi_{2t} \\
+ (fg)(X_t)\xi_{2t}\epsilon_t + h^q \eta_{1t}\xi_{2t}\epsilon_t + h^q \eta_{2t}\xi_{2t}\epsilon_t + \left( \sum_{s=1}^{T} K_{st}\epsilon_s \right) \xi_{2t}\epsilon_t \right\}
\]
Hereafter the $i$th term of the second summation of (15) is denoted by $B_i$. We have from Assumption A9,

$$B_1 = o_p(1), \quad B_4 = o_p(1), \quad B_5 = o_p(1), \quad B_8 = o_p(1), \quad B_9 = o_p(1),$$

$$B_{20} = o_p(1), \quad \text{and} \quad B_{21} = o_p(1), \quad \text{uniformly in } \tau. \quad (16)$$

Next we evaluate $\eta_{2t}$, $\xi_{2t}$, and $\sum_{s=1}^{T} K_{st} \varepsilon_s$ using the exponential inequality (Theorem A.1). The proofs of Propositions 2.1 and 2.2 are deferred to Appendix B.

**Proposition 2.1** Suppose that Assumptions A1, A3-4, and A6 hold. Then

$$\xi_{2t} = O\left(\frac{\log T}{T^{\alpha_2}}\right) \quad \text{and} \quad \eta_{2t} = O\left(\frac{\log T}{T^{\alpha_2}}\right), \quad \text{uniformly in } t. \quad (17)$$

**Proposition 2.2** Suppose that Assumptions A1-4 and A6-7 hold. Then

$$\sum_{s=1}^{T} K_{st} \varepsilon_s = \frac{1}{Th^p} \sum_{s=1}^{T} K \left(\frac{X_s - X_t}{h}\right) \varepsilon_s = O_p\left(\frac{\log T}{T^{\alpha_2}}\right) \quad \text{uniformly in } t. \quad (18)$$

Using the results on U-statistics in Yoshihara (1976) and Harel and Puri (1990), we evaluate $B_2$, $B_3$, $B_{12}$, $B_{15}$, and $B_{24}$. If we were able to prove

$$E\left\{\left|K \left(\frac{X_s - X_t}{h}\right) g(X_s) f(X_t) \varepsilon_t\right|^r\right\} = O(h^p), \quad (19)$$

$$E\left\{\left|K \left(\frac{X_s - X_t}{h}\right) \varepsilon_s f(X_t) \varepsilon_t\right|^r\right\} = O(h^p), \quad (20)$$

$$E\left\{\left|K \left(\frac{X_s - X_t}{h}\right) (fg)(X_t) \varepsilon_t\right|^r\right\} = O(h^p), \quad (21)$$

$$E\left\{\left|K \left(\frac{X_s - X_t}{h}\right) (fg)(X_s) \varepsilon_s\right|^r\right\} = O(h^p), \quad (22)$$

$$E\left\{\left|K \left(\frac{X_s - X_t}{h}\right) (fg') (X_t)\right|^r\right\} = O(h^p), \quad (23)$$

and
we could relax Assumption A10. Without some assumptions on the joint distribution \( \{ (X_t', \epsilon_t)' \} \) and \( g \), (19)-(24) would not hold. Lemma B.4 of Hidalgo(1992) is doubtful. But (19)-(24) are true if \( \{ (X_t', \epsilon_t)' \} \) is a Gaussian process and \( fg^2 \) is bounded. We proceed without (19)-(24).

**Proposition 2.3** Suppose that Assumptions A1-3, A8, and A10 hold. Then

\[
B_2 = o_p(1) \quad \text{uniformly in } \tau. \tag{25}
\]

**Proof**

\[
B_2 = \frac{\sqrt{T}}{h^P} \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[ K\left( \frac{X_s - X_t}{h} \right) g(X_s) - E_t \left[ K\left( \frac{X_s - X_t}{h} \right) g(X_s) \right] \right] f(X_t) \epsilon_t. \tag{26}
\]

The summand is rewritten as

\[
K\left( \frac{X_s - X_t}{h} \right) g(X_s) f(X_t) \epsilon_t - E_t \left[ K\left( \frac{X_s - X_t}{h} \right) g(X_s) f(X_t) \epsilon_t \right]
- E_s \left[ K\left( \frac{X_s - X_t}{h} \right) g(X_s) f(X_t) \epsilon_t \right] + E_s \left[ K\left( \frac{X_s - X_t}{h} \right) g(X_s) f(X_t) \epsilon_t \right].
\]

So it is easily seen from the proofs of Lemma 4 of Yoshihara(1976) and Lemma 2.2 of Harel and Puri(1990) that

\[
E\{B_2^4\} = O(1/(T^2 h^{4P})). \tag{27}
\]

Then the proposition is verified. \( \Box \)

The proofs of Propositions 2.4 and 2.5 are similar to that of Proposition 2.3. So omitted.

**Proposition 2.4** Suppose that Assumptions A1-3, A8, and A10 hold. Then

\[
B_3 = o_p(1) \quad \text{uniformly in } \tau. \tag{28}
\]

\( \Box \)

**Proposition 2.5** Suppose that Assumptions A1-3, A8, and A10 hold. Then

\[
B_{16} = o_p(1) \quad \text{uniformly in } \tau. \tag{29}
\]

\( \Box \)
The proof of Proposition 2.6 is deferred to Appendix B.

**Proposition 2.6** Suppose that Assumptions A1-5 and A8-10 hold. Then

\[ B_{12} - B_{24} = o_p(1) \quad \text{uniformly in } \tau. \]  

(30)

Here we evaluate the second term of the first summation in (15).

**Proposition 2.7** Suppose that Assumptions A1-4 and A8-10 hold. Then

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} (f g)(X_t) \sum_{s=1}^{T} K_{st} \epsilon_s = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} (f^2 g)(X_s) \epsilon_s + o_p(1) \quad \text{uniformly in } \tau. \]  

(31)

**Proof**

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} (f g)(X_t) \sum_{s=1}^{T} K_{st} \epsilon_s
\]

\[
= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \left\{ \frac{[Tr]}{T} (f^2 g)(X_s) + \sum_{t=1}^{[Tr]} \left\{ (f g)(X_t) K_{st} - E_s \{(f g)(X_t) K_{st}\} \right\} 
+ \frac{[Tr]}{T} h^q \left\{ \int K(u) \sum_{s_1=0}^{T} \prod_{s_1}^{T} \frac{\partial^q (f^g)}{\partial x_1^{s_1} \cdots \partial x_2^{s_2}} (X_s) \prod_{s_1}^{T} u^{s_1}_t \, du + O(h) \right\} \right\} \epsilon_t
\]

The third term is \(O_p(\sqrt{T} h^q)\). Applying the proofs of Lemma 4 of Yoshihara (1976) and Lemma 2.2 of Harel and Puri (1990) to

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \sum_{s=1}^{T} \{(f g)(X_t) K_{st} \epsilon_s - E_s \{(f g)(X_t) K_{st} \epsilon_s\}\},
\]

we obtain (31). \(\square\)

We prove Theorem 2.1.

**Proof of Theorem 2.1** By Assumption A9 and Propositions 2.1-2, we can show that \(B_6, B_7, B_{10}, B_{11}, B_{14}, B_{15}, B_{17}, B_{18}, B_{19}, B_{22}, B_{23}, B_{25}, B_{26}, B_{27}, B_{28}, B_{29},\) and \(B_{30}\) are \(o_p(1)\) uniformly in \(\tau\). So we have Theorem 2.1 from (15), (16), and Propositions 2.3-7. \(\square\)

We cannot say much about how to choose the bandwidth \(h\). This is an untractable problem. The optimal rate \(O(T^{-1/(2q+p)})\) in estimating \(g(x)\) itself cannot be used here. The effects of bias terms might be more serious. Then \(q = 3\) and \(h = O(T^{-1/3})\) might be recommended in the case of \(P = 1\) even though this violates Assumption A10.
3 Numerical examples

In this section we give some results of simulation studies. We consider 6 tests and 3 models. In the first place we describe test statistics.

**Test 1** Conditional moment test 1. We used the normal kernel,

\[ (1.5 - 0.5x^2) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \]

with \( q = 3 \) and \( h = 0.2 \) in this section. We reject the null hypothesis \( H_0 \) if \( \max_{0 \leq \tau \leq 1} |S_\tau| > 1.3581 \).

**Test 2** Conditional moment test 2, which is essentially the same as Test 1 of Hidalgo (1995). We reject the null hypothesis \( H_0 \) if \( \max_{0.2 \leq \tau \leq 0.8} |S_\tau|^2 / (1 - \tau) > 8.45 \). 8.45 is taken from Andrews (1993).

**Test 3** CUSUM test with \( \hat{f}_\tau \), which is proposed in section 2 of Honda (1996). We reject the null hypothesis \( H_0 \) if \( \max_{0 \leq \tau \leq 1} |\sum_{i=1}^{[T\tau]} \hat{f}_\tau| > 1.3581 \).

**Test 4** CUSUM test with nonparametric regression residuals by the Nadaraya-Watson kernel estimator. We took the trimming parameter \( l \) to be 0.05. We reject the null hypothesis if \( \max_{0 \leq \tau \leq 1} |\sum_{i=1}^{[T\tau]} \hat{e}_\tau| > 1.3581 \).

**Test 5** CUSUM test with OLS residuals. Assuming that \( Y_i = a + bX_i + cX_i^2 + \epsilon_i \), we reject the null hypothesis if the maximum of the modulus of the cumulative sums of OLS residuals exceeds 1.3581.

**Test 6** Max \( F \)-test. Assuming that \( Y_i = a + bX_i + cX_i^2 + \epsilon_i \), we reject the null hypothesis if the OLS \( F \)-test statistics for testing structural change at \( [T\tau] \) \((0.2 \leq \tau \leq 0.8) \) exceed 13.69, which is taken from Andrews (1993).

Next we give models.

**Model 1.1** \( X_i \sim \text{NID}(2,1) \) and \( \epsilon_i \sim \text{NID}(0,1/4) \). \( \{X_i\} \) and \( \{\epsilon_i\} \) are independent.

\[ Y_i = X_i + 0.5X_i^2 + \epsilon_i, \quad i = 1, 2, \ldots, 100. \] (32)
A Note on Nonparametric Conditional Moment Tests

Model 1.2 (32), \( i = 1, 2, \ldots, 50 \), and
\[
Y_i = 0.1 + X_i + 0.5X_i^2 + \epsilon_i, \quad i = 51, 52, \ldots, 100.
\]

Model 1.3 (32), \( i = 1, 2, \ldots, 50 \), and
\[
Y_i = 1.1X_i + 0.5X_i^2 + \epsilon_i, \quad i = 51, 52, \ldots, 100.
\]

Model 2.1 \( X_i \sim \text{NID}(0, 1) \) and \( \epsilon_i \sim \text{NID}(0, 1/4) \). \( \{X_i\} \) and \( \{\epsilon_i\} \) are independent.
\[
Y_i = \frac{\pi}{2} + \arctan(2X_i) + \epsilon_i, \quad i = 1, 2, \ldots, 100.
\] (33)

Model 2.2 (33), \( i = 1, 2, \ldots, 50 \), and
\[
Y_i = \frac{\pi}{2} + 0.1 + \arctan(2X_i) + \epsilon_i, \quad i = 51, 52, \ldots, 100.
\]

Model 2.3 (33), \( i = 1, 2, \ldots, 50 \), and
\[
Y_i = 1.1\left(\frac{\pi}{2} + \arctan(2X_i)\right) + \epsilon_i, \quad i = 51, 52, \ldots, 100.
\]

Model 3.1 \( X_i \sim \text{NID}(1, 1) \) and \( \epsilon_i \sim \text{NID}(0, 1/4) \). \( \{X_i\} \) and \( \{\epsilon_i\} \) are independent.
\[
Y_i = X_i^2 + X_i^3 + \epsilon_i, \quad i = 1, 2, \ldots, 100.
\] (34)

Model 3.2 (34), \( i = 1, 2, \ldots, 50 \), and
\[
Y_i = 0.2 + X_i^2 + X_i^3 + \epsilon_i, \quad i = 51, 52, \ldots, 100.
\]

Model 3.3 (34), \( i = 1, 2, \ldots, 50 \), and
\[
Y_i = 1.1X_i^2 + X_i^3 + \epsilon_i, \quad i = 51, 52, \ldots, 100.
\]

The powers of Tests 1-6 are given in Table 1. The number of iterations is 5000. The calculation was carried out with C programs.

The powers of Tests 1-4 are similar in Models 1 and 2. In Model 3.2, Tests 3 and 4 are superior to Tests 1 and 2. On the other hand Tests 1 and 2 dominate Tests 3 and 4 in Model 3.3. max \( F \) test might give misleading results when the model is misspecified. When the model is correctly specified, Tests 1-4 might be comparable to Tests 5 and 6.
<table>
<thead>
<tr>
<th>Model</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>2.1</th>
<th>2.2</th>
<th>2.3</th>
<th>3.1</th>
<th>3.2</th>
<th>3.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>0.03</td>
<td>0.08</td>
<td>0.28</td>
<td>0.03</td>
<td>0.10</td>
<td>0.23</td>
<td>0.03</td>
<td>0.13</td>
<td>0.21</td>
</tr>
<tr>
<td>Test 2</td>
<td>0.03</td>
<td>0.07</td>
<td>0.24</td>
<td>0.03</td>
<td>0.09</td>
<td>0.20</td>
<td>0.03</td>
<td>0.12</td>
<td>0.19</td>
</tr>
<tr>
<td>Test 3</td>
<td>0.03</td>
<td>0.09</td>
<td>0.28</td>
<td>0.03</td>
<td>0.09</td>
<td>0.19</td>
<td>0.03</td>
<td>0.24</td>
<td>0.15</td>
</tr>
<tr>
<td>Test 4</td>
<td>0.04</td>
<td>0.10</td>
<td>0.31</td>
<td>0.04</td>
<td>0.10</td>
<td>0.21</td>
<td>0.05</td>
<td>0.20</td>
<td>0.15</td>
</tr>
<tr>
<td>Test 5</td>
<td>0.04</td>
<td>0.11</td>
<td>0.35</td>
<td>0.04</td>
<td>0.09</td>
<td>0.18</td>
<td>0.03</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Test 6</td>
<td>0.05</td>
<td>0.09</td>
<td>0.30</td>
<td>0.22</td>
<td>0.25</td>
<td>0.35</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
</tbody>
</table>

4 Concluding remarks

We considered the mathematical validity of the nonparametric conditional moment tests which are proposed in Hidalgo (1995) using the Nadaraya-Watson kernel estimators. We gave a correct proof and the consistent assumptions on the result of the limit distribution $S_r$ defined by (3). The result here allows for the dependence of $\{X_t\}$ and $\{\epsilon_t\}$.

As is well known, the CUSUM tests and Tests 1 and 2 in the previous section are insensitive to some alternatives. It seems difficult to specify the maintained hypothesis in the case where a structural change really exists. When we use these nonparametric test statistics, we may as well use several procedures simultaneously Then we have to note that the actual size might be larger than the nominal one. We will carry out further studies to compare the parametric and nonparametric tests elsewhere.

Appendix A

We reproduce here the exponential inequality of Babu (1989) for convenience. A similar inequality is applied to the kernel estimation problem in Collomb and Härdle (1986).

Let $\{Z_t\}$ be a uniformly mixing process with mixing coefficients $\{\phi_k\}$. In addition we assume that

$$E\{Z_t\} = 0, \quad |Z_t| \leq b \quad \text{a.s., and} \quad \sum \phi_k^{1/2} < \infty. \quad (35)$$

We choose two integer $p (1 \leq p \leq T)$ and $k$ such that $k = [T/(2p)]$ and $k\phi_p \leq M$. Then we have the following theorem.
Theorem A.1 For any positive $\eta$ and $y$, 
\[
P \left( \left| \sum_{i=1}^{T} Z_i \right| > D \right) \leq 12 \exp \left( -\frac{yD}{2} + M_\phi y^2 \sum \text{var}(Z_i) + M \right) + \frac{2T}{p} (yp^\eta),
\]
where $M_\phi$ depends only on $\{\phi_k\}$.

Suppose that $\phi_t = O(t^{-\alpha})$ and $\alpha > 2$. Choosing $p = [T^{1/(1+\alpha)}]$, we have 
\[
k_\phi = O(1) \quad \text{and} \quad \frac{T}{p} (yp^\eta) = O(T^{\alpha/(1+\alpha)}(ybT^{1/(1+\alpha)})^\eta).
\]

Appendix B

The proofs of Proposition 2.1 and 2.2 follow standard arguments in kernel estimation. For example, Müller and Stadtmüller (1987) and Collomb and Härdle (1986). See also Bai (1995).

Proof of Proposition 2.1 We show the case of $\eta_{2t}$. The case of $\xi_{2t}$ is similar and easier.

From the Borel-Cantelli theorem, we have only to prove that uniformly on $\{|x| \leq T^{1/r}\}$, 
\[
\eta_{2x} = \frac{1}{Th^p} \sum_{s=1}^{T^n} \left[ K \left( \frac{X_s - x}{h} \right) g(X_s) - E \left\{ K \left( \frac{X_s - x}{h} \right) g(X_s) \right\} \right] = O \left( \frac{\log T}{T^n} \right).
\]
We evaluate (38) using the exponential inequality. We apply the exponential inequality to (38) with $x$ fixed. The parameters in Appendix A are as follows:
\[
b = O \left( \frac{T^{k/r_1}}{Th^p} \right), \quad \sum \text{var}(Z_i) = O \left( \frac{1}{Th^p} \right), \quad D = LT^{-r_1} \log T, \quad \text{and} \quad y = T^{r_1},
\]
where $L$ will be taken to be large enough later. Then 
\[
Dy = L \log T, \quad \sum \text{var}(Z_i)y^2 = O \left( \frac{1}{T^{1-2r_1}h^p} \right), \quad \text{and} \quad ybp = \frac{T^{k/r_1 + 1/(1+\alpha) + r_1}}{Th^p}.
\]

Next we consider uniformity in $x$. When $|\bar{x} - x| < h^{P+1}/T^{k/r_1 + r_1}$,
\[
\left| K \left( \frac{X_s - x}{h} \right) g(X_s) - K \left( \frac{X_s - \bar{x}}{h} \right) g(X_s) \right| < C_5 T^{-r_1},
\]
where $C_5$ does not depend on $x$. By taking $O(T^{P(k/r_1 + r_1)} / h^{P(P+1)})$ grid points and sufficiently large $L$, we can show (38) $= O(\log T/T^{r_1})$ uniformly on $\{|x| \leq T^{1/r_1}\}$ by the Borel-Cantelli theorem.
**Proof of Proposition 2.2** From the Borel-Cantelli theorem, we have only to prove that uniformly on \(|x| \leq T^{1/r_2}\),

\[
\frac{1}{Th^p} \sum_{s=1}^{T} K\left(\frac{X_s - x}{h}\right) \epsilon_s = O_p\left(\frac{\log T}{T^{\gamma_2}}\right) \tag{42}
\]

By using \(E\{\epsilon_s|X_s\} = 0\), (42) can be rewritten as

\[
\frac{1}{Th^p} \sum_{s=1}^{T} K\left(\frac{X_s - x}{h}\right) (\epsilon_s - E\{\epsilon_s|X_s\}) + \frac{1}{Th^p} \sum_{s=1}^{T} K\left(\frac{X_s - x}{h}\right) E\{\epsilon_s I\{|\epsilon_s| > T^{1/r_2}\}|X_s\}, \tag{43}
\]

where \(\epsilon_s = \epsilon_s I\{|\epsilon_s| \leq T^{1/r_2}\}\). The second term of (43) is \(O_p(T^{-\gamma_2})\) uniformly in \(x\). We evaluate the first term using the exponential inequality. We apply the exponential inequality with \(x\) fixed. The parameters in Appendix A are as follows:

\[
b = O\left(\frac{T^{1/r_2}}{Th^p}\right), \quad \sum \text{var}(Z_t) = O\left(\frac{1}{Th^p}\right), \quad D = LT^{-\gamma_3} \log T, \quad \text{and} \quad y = T^{\gamma_3}, \tag{44}
\]

where \(L\) will be taken to be large enough later. Then

\[
Dy = L \log T, \quad \sum \text{var}(Z_t) y^2 = O\left(\frac{1}{T^{1-2\gamma_3}h^p}\right), \quad \text{and} \quad ybp = \frac{T^{1/r_2 + 1/(1+\alpha)+\gamma_3}}{Th^p}. \tag{45}
\]

The uniformity in \(x\) follows the same argument as Proposition 2.1. \(\square\)

**Proof of Proposition 2.6**

\[
B_{12} - B_{24} = \frac{\sqrt{T}}{h^p} \sum_{t=1}^{T} \sum_{s=1}^{T} (fg^2)(X_t) K\left(\frac{X_s - X_t}{h}\right) - \text{Ex}\{(fg^2)(X_t) K\left(\frac{X_s - X_t}{h}\right)\}
\]

\[
- (fg)(X_t) g(X_s) K\left(\frac{X_s - X_t}{h}\right) + \text{Ex}\{(fg)(X_t) g(X_s) K\left(\frac{X_s - X_t}{h}\right)\}\]

Noting that

\[
\text{Ex}\{(fg^2)(X_t) K\left(\frac{X_s - X_t}{h}\right)\} = h^p ((fg)^2(X_s) + h^q \delta_a(X_s)) \tag{47}
\]

and

\[
\text{Ex}\{(fg)(X_t) g(X_s) K\left(\frac{X_s - X_t}{h}\right)\} = h^p g(X_s) ((f^2)(g)(X_s) + h^q \delta_b(X_s)), \tag{48}
\]

where \(\delta_a(X_s)\) and \(\delta_b(X_s)\) have the third moment, the summand of (46) can be written as

\[
-(fg)(X_t) g(X_s) K\left(\frac{X_s - X_t}{h}\right) + \text{Ex}\{(fg)(X_t) g(X_s) K\left(\frac{X_s - X_t}{h}\right)\} \tag{49}
\]
\[ +E_s\{(fg)(X_t)g(X_s)K\left(\frac{X_s - X_t}{h}\right)\} - EE_s\{(fg)(X_t)g(X_s)K\left(\frac{X_s - X_t}{h}\right)\} \\
+ (fg^2)(X_t)K\left(\frac{X_s - X_t}{h}\right) - E_t\{(fg^2)(X_t)K\left(\frac{X_s - X_t}{h}\right)\} \\
- E_s\{(fg^2)(X_t)K\left(\frac{X_s - X_t}{h}\right)\} + EE_s\{(fg^2)(X_t)K\left(\frac{X_s - X_t}{h}\right)\} + h^{q+1}\delta_c(X_s), \]

where \( E\{|\delta_c(X_s)|\} < \infty \). Applying the proofs of Lemma 4 of Yoshihara (1976) and Lemma 2.2 of Harel and Puri (1990) to (49) and (50), we obtain (30).

\[ \square \]

References


