

Chapter 15

Comparison of Autoregressive Curves Through Partial Sums of Quasi-Residuals

Fang Li

15.1 Introduction

This chapter is concerned with testing the equality of two autoregressive functions against two sided alternatives when observing two independent strictly stationary and ergodic autoregressive times series of order one. More precisely, let $Y_{1,i}, Y_{2,i}, i \in \mathbb{Z} := \{0, \pm 1, \dots\}$, be two observable autoregressive time series such that for some real valued functions μ_1 and μ_2 , and for some positive functions σ_1, σ_2 ,

$$Y_{1,i} = \mu_1(Y_{1,i-1}) + \sigma_1(Y_{1,i-1})\varepsilon_{1,i}, \quad Y_{2,i} = \mu_2(Y_{2,i-1}) + \sigma_2(Y_{2,i-1})\varepsilon_{2,i}. \quad (15.1)$$

The errors $\{\varepsilon_{1,i}, i \in \mathbb{Z}\}$ and $\{\varepsilon_{2,i}, i \in \mathbb{Z}\}$ are assumed to be two independent sequences of independent and identically distributed (i.i.d.) r.v.'s with mean zero and unit variance. Moreover, $\varepsilon_{1,i}, i \geq 1$ are independent of $Y_{1,0}$, and $\varepsilon_{2,i}, i \geq 1$ are independent of $Y_{2,0}$. And the time series are assumed to be stationary and ergodic.

Consider a bounded interval $[a, b]$ of \mathbb{R} . The problem of interest is to test the null hypothesis:

$$H_0 : \mu_1(x) = \mu_2(x), \quad \forall x \in [a, b],$$

against the two sided alternative hypothesis:

$$H_1 : \mu_1(x) \neq \mu_2(x), \quad \text{for some } x \in [a, b], \quad (15.2)$$

based on the data set $Y_{1,0}, Y_{1,1}, \dots, Y_{1,n_1}, Y_{2,0}, Y_{2,1}, \dots, Y_{2,n_2}$.

In hydrology, autoregressive time series are often used to model water reservoirs, see, e.g., Bloomfield (1992). The above testing problem could be applied in comparing the water levels of two rivers.

Few related studies had been conducted under the two sample autoregressive setting. Koul and Li (2005) adapts the covariate matching idea used in regression setting to a one-sided tests for the superiority among two time series. Li (2009)

1990 IMS Subject Classification: Primary 62M10, Secondary 62F03.

F. Li (✉)

Department of Mathematical Sciences, Indiana University Purdue University at Indianapolis, IUPUI, 402 N. Blackford Street, 46202 Indianapolis, IN, USA
e-mail: fli@math.iupui.edu

studied the same testing problem, but the test is based on the difference of two sums of quasi-residuals. This method is also an extension of T_2 in Koul and Schick (1997) from regression setting to autoregressive setting.

The papers that address the above two sided testing problem in regression setting include Hall and Hart (1990); Delgado (1993); Kulasekera (1995) and Scheike (2000). In particular, Delgado (1993) used the absolute difference of the cumulative regression functions for the same problem, assuming common design in the two regression models. Kulasekera (1995) used quasi-residuals to test the difference between two regression curves, under the conditions that do not require common design points or equal sample sizes. The current article adapts Delgado’s idea of using partial sum process and Kulasekera’s idea of using quasi residuals to construct the tests for testing the difference between two autoregressive functions.

Similarly, as in Delgado (1993), let

$$\Delta(t) := \int_a^t (\mu_1(x) - \mu_2(x)) (f_1(x) + f_2(x)) dx, \quad \forall a \leq t \leq b, \quad (15.3)$$

where μ_1, μ_2 are assumed to be continuous on $[a, b]$ and f_1, f_2 are the stationary densities of the two time series $Y_{1,i}$ and $Y_{2,i}$, respectively. We also assume that f_1, f_2 are continuous and positive on $[a, b]$. It is easy to show that $\Delta(t) \equiv 0$ when the null hypothesis holds and $\Delta(t) \neq 0$ for some t under H_a . This suggests to construct tests of H_0 vs. H_a based on some consistent estimators of $\Delta(t)$. One such estimator is obtained as follows.

First, as in Kulasekera (1995), we define quasi-residuals

$$e_{1,i} = Y_{1,i} - \hat{\mu}_2(Y_{1,i-1}), \quad i = 1, \dots, n_1, \quad (15.4)$$

and

$$e_{2,j} = Y_{2,j} - \hat{\mu}_1(Y_{2,j-1}), \quad j = 1, \dots, n_2. \quad (15.5)$$

Here, $\hat{\mu}_1$ and $\hat{\mu}_2$ are appropriate estimators, such as Nadaraya–Watson estimators used in this article, of μ_1 and μ_2 . See Nadaraya (1964) and Watson (1994).

Now, let

$$U_n(t) = \frac{1}{n_1} \sum_{i=1}^{n_1} e_{1,i} 1_{[a \leq Y_{1,i-1} \leq t]} - \frac{1}{n_2} \sum_{j=1}^{n_2} e_{2,j} 1_{[a \leq Y_{2,j-1} \leq t]}, \quad (15.6)$$

where the subscript n , here and through out the chapter, represents the dependence on n_1 and n_2 . With uniformly consistent estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ of μ_1 and μ_2 such as kernel estimates and under some mixing condition on the time series $Y_{1,i}$ and $Y_{2,j}$ such as strongly α -mixing, $U_n(t)$ can be shown to be $U_{1n}(t) + U_{2n}(t) + U_{3n}(t)$ with

$$\begin{aligned} U_{1n}(t) &= \frac{1}{n_1} \sum_{i=1}^{n_1} \sigma_1(Y_{1,i-1}) \varepsilon_{1,i} 1_{[a \leq Y_{1,i-1} \leq t]} \\ &\quad - \frac{1}{n_2} \sum_{j=1}^{n_2} \sigma_2(Y_{2,j-1}) \varepsilon_{2,j} 1_{[a \leq Y_{2,j-1} \leq t]} = o_P(1), \end{aligned}$$

$$\begin{aligned}
 U_{2n}(t) &= \frac{1}{n_1} \sum_{i=1}^{n_1} (\mu_1(Y_{1,i-1}) - \mu_2(Y_{1,i-1})) I_{[a \leq Y_{1,i-1} \leq t]} \\
 &\quad - \frac{1}{n_2} \sum_{j=1}^{n_2} (\mu_2(Y_{2,j-1}) - \mu_1(Y_{2,j-1})) I_{[a \leq Y_{2,j-1} \leq t]} \\
 &= \int_a^t (\mu_1(x) - \mu_2(x)) (f_1(x) + f_2(x)) dx + o_P(1), \\
 U_{3n}(t) &= \frac{1}{n_1} \sum_{i=1}^{n_1} (\mu_2(Y_{1,i-1}) - \hat{\mu}_2(Y_{1,i-1})) I_{[a \leq Y_{1,i-1} \leq t]} \\
 &\quad - \frac{1}{n_2} \sum_{j=1}^{n_2} (\mu_1(Y_{2,j-1}) - \hat{\mu}_1(Y_{2,j-1})) I_{[a \leq Y_{2,j-1} \leq t]} = o_P(1),
 \end{aligned}$$

uniformly for all $t \in [a, b]$. Thus, $U_n(t)$ provides a uniformly consistent estimator of $\Delta(t)$. This suggests to base tests of H_0 on some suitable functions of this process. In this chapter, we shall focus on the Kolmogorov–Smirnov type test based on $\sup_{a \leq t \leq b} |U_n(t)|$.

To determine the large sample distribution of the process $U_n(t)$, one needs to normalize this process suitably. Let

$$\begin{aligned}
 \tau_n^2(t) &= q_1 E \left\{ \sigma_1^2(Y_{1,0}) \left(1 + \frac{f_2(Y_{1,0})}{f_1(Y_{1,0})} \right)^2 1_{[a \leq Y_{1,0} \leq t]} \right\} \\
 &\quad + q_2 E \left\{ \sigma_2^2(Y_{2,0}) \left(1 + \frac{f_1(Y_{2,0})}{f_2(Y_{2,0})} \right)^2 1_{[a \leq Y_{2,0} \leq t]} \right\}, \tag{15.7}
 \end{aligned}$$

where, $q_1 = \frac{N}{n_1} = \frac{n_2}{n_1+n_2}$, $q_2 = \frac{N}{n_2} = \frac{n_1}{n_1+n_2}$ and $N = \frac{n_1 n_2}{n_1+n_2}$.

We consider the following normalized test statistics:

$$T := \sup_{a \leq t \leq b} \left| \frac{N^{1/2} U_n(t)}{\sqrt{\tau_n^2(b)}} \right|. \tag{15.8}$$

In the case σ_i 's and f_i 's are known, the tests of H_0 could be based on T , being significant for its large value. But, usually those functions are unknown which renders T of little use. This suggests to replace τ_n with its estimate $\hat{\tau}_n^2$ which satisfies

$$\frac{\hat{\tau}_n^2(b)}{\tau_n^2(b)} \rightarrow_P 1. \tag{15.9}$$

An example of such estimator $\hat{\tau}_n(t)$ of $\tau_n(t)$ is

$$\begin{aligned}
 \hat{\tau}_n^2(t) &= q_1 \frac{1}{n_1} \sum_{i=1}^{n_1} \left\{ (Y_{1,i} - \tilde{\mu}_1(Y_{1,i-1}))^2 \left(1 + \frac{\hat{f}_2(Y_{1,i-1})}{\hat{f}_1(Y_{1,i-1})} \right)^2 1_{[a \leq Y_{1,i-1} \leq t]} \right\} \\
 &\quad + q_2 \frac{1}{n_2} \sum_{j=1}^{n_2} \left\{ (Y_{2,j} - \tilde{\mu}_2(Y_{2,j-1}))^2 \left(1 + \frac{\hat{f}_1(Y_{2,j-1})}{\hat{f}_2(Y_{2,j-1})} \right)^2 1_{[a \leq Y_{2,j-1} \leq t]} \right\}, \tag{15.10}
 \end{aligned}$$

where, $\tilde{\mu}_i$'s and \hat{f}_i 's are appropriate estimators, such as kernel estimators used in this paper, of μ_i 's and f_i 's. Therefore, the proposed tests will be based on the adaptive version of T , namely

$$\hat{T} := \sup_{a \leq t \leq b} \left| \frac{N^{1/2} U_n(t)}{\sqrt{\hat{\tau}_n^2(b)}} \right| \tag{15.11}$$

We shall study the asymptotic behavior of \hat{T} as the sample sizes n_1 and n_2 tend to infinity. Theorem 2.1 of Sect. 15.2 shows that under H_0 , T weakly converge to supremum of Brownian motion over $[0, 1]$, under some general assumptions and with $\hat{\mu}_1$ and $\hat{\mu}_2$ being Nadaraya–Watson estimators of μ_1 and μ_2 . Then, in Corollary 2.1, under some general assumptions on the estimates $\tilde{\mu}_1$, $\tilde{\mu}_2$ and \hat{f}_1 , \hat{f}_2 , we derive the same asymptotic distributions of \hat{T} under H_0 . Remark 2.2 proves that the power of the test based on \hat{T} converges to 1, at the fixed alternative (15.2) or even at the alternatives that converge to H_0 at a rate lower than $\sqrt{\tau_n^2(b)}$. In Sect. 15.3, we conduct a Monte Carlo simulation study of the finite sample level and power behavior of the proposed test \hat{T} . The simulation results are shown to be consistent with the asymptotic theory at the moderate sample sizes considered. In Sect. 15.4, we study some properties of kernel smoothers and weak convergence of both empirical processes and marked empirical processes. Those studies facilitate the proof of our main results in Sect. 15.2. But, they may also be of interest on their own, hence are formulated and proved in Sect. 15.4. The other proofs are deferred to Sect. 15.5.

15.2 Asymptotic Behavior of T and \hat{T}

This section investigates the asymptotic behavior of T given in (15.8) and the adaptive statistic \hat{T} given in (15.11) under the null hypothesis and the alternatives (15.2). We write P for the underline probability measures and E for the corresponding expectations. In this chapter we consider Nadaraya–Watson estimators $\hat{\mu}_1$, $\hat{\mu}_2$ of μ_1 and μ_2 , i.e.,

$$\hat{\mu}_i(x) = \frac{\sum_{j=1}^{n_i} Y_{i,j} K_{h_i}(Y_{i,j-1} - x)}{\sum_{j=1}^{n_i} K_{h_i}(Y_{i,j-1} - x)}, \quad i = 1, 2, \tag{15.12}$$

where $K_{h_i}(x) = \frac{1}{h_i} K(\frac{x}{h_i})$, with K being a kernel density function on the real line with compact support $[-1, 1]$, $h_1, h_2 > 0$ are the bandwidths. First, we recall the following definition from Bosq (1998):

Definition 2.1 For any real discrete time process $(X_i, i \in \mathbb{Z})$ define the strongly mixing coefficients

$$\alpha(k) := \sup_{t \in \mathbb{Z}} \alpha(\sigma\text{-field}(X_i, i \leq t), \sigma\text{-field}(X_i, i \geq t + k)); \quad k = 1, 2, \dots$$

where, for any two sub σ -fields \mathcal{B} and \mathcal{C} ,

$$\alpha(\mathcal{B}, \mathcal{C}) = \sup_{B \in \mathcal{B}, C \in \mathcal{C}} |P(B \cap C) - P(B)P(C)|.$$

Definition 2.2. The process $(X_i, i \in \mathbb{Z})$ is said to be geometrically strong mixing (GSM) if there exists $c_0 > 0$ and $\rho \in [0, 1)$ such that $\alpha(k) \leq c_0 \rho^k$, for all $k \geq 1$.

The following assumptions are needed in this paper.

- (A.1) The autoregressive functions μ_1, μ_2 are continuous on an open interval containing $[a, b]$ and they have continuous derivatives on $[a, b]$.
- (A.2) The kernel function $K(x)$ is a symmetric Lipschitz-continuous density on \mathbb{R} with compact support $[-1, 1]$.
- (A.3) The bandwidths h_1, h_2 are chosen such that $h_i^2 N^{1-c} \rightarrow \infty$ for some $c > 0$ and $h_i^4 N \rightarrow 0$.
- (A.4) The densities f_1 and f_2 are bounded and their restrictions to $[a, b]$ are positive. Moreover, they have continuous second derivatives over an open interval containing $[a, b]$.
- (A.5) The conditional variance functions σ_1^2 and σ_2^2 are positive on $[a, b]$ and continuous on an open interval containing $[a, b]$.
- (A.6) $Y_{1,i}, Y_{2,i}, i \in \mathbb{Z}$ are GSM processes.
- (A.7) For some $M < \infty$, we have

$$E(\varepsilon_{i,1}^4) \leq M, \quad i = 1, 2.$$

- (A.8) For $i = 1, 2$, the joint densities $g_{i,l}$ between $Y_{i,0}$ and $Y_{i,l}$ for all $l \geq 1$ are uniformly bounded over an open interval \mathcal{I}_0 containing \mathcal{I} , i.e., $\sup_{l \geq 1} \sup_{x,y \in \mathcal{I}_0} g_{i,l}(x, y) < \infty$.
- (A.9) The densities g_1 and g_2 of the innovations $\varepsilon_{1,1}$ and $\varepsilon_{2,1}$ are bounded.

Let $\mathcal{K}(y) = \int_{-1}^y K(t) dt$ be the distribution function corresponding to the kernel density $K(y)$ on $[-1, 1]$ and let

$$\begin{aligned} V_n(t) = & \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \sigma_1(Y_{1,i-1}) \left(1_{[a \leq Y_{1,i-1} \leq t]} + \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \left(\mathcal{K} \left(\frac{t - Y_{1,i-1}}{h_2} \right) \right. \right. \\ & \left. \left. - \mathcal{K} \left(\frac{a - Y_{1,i-1}}{h_2} \right) \right) \right) \\ & - \frac{1}{n_2} \sum_{j=1}^{n_2} \varepsilon_{2,j} \sigma_2(Y_{2,j-1}) \left(1_{[a \leq Y_{2,j-1} \leq t]} + \frac{f_1(Y_{2,j-1})}{f_2(Y_{2,j-1})} \left(\mathcal{K} \left(\frac{t - Y_{2,j-1}}{h_1} \right) \right. \right. \\ & \left. \left. - \mathcal{K} \left(\frac{a - Y_{2,j-1}}{h_1} \right) \right) \right) \end{aligned} \tag{15.13}$$

and

$$\begin{aligned} W_n(t) = & \frac{1}{n_1} \sum_{i=1}^{n_1} (\mu_1(Y_{1,i-1}) - \mu_2(Y_{1,i-1})) 1_{[a \leq Y_{1,i-1} \leq t]} \\ & + \frac{1}{n_2} \sum_{j=1}^{n_2} (\mu_1(Y_{2,j-1}) - \mu_2(Y_{2,j-1})) 1_{[a \leq Y_{2,j-1} \leq t]} \end{aligned} \tag{15.14}$$

We are now ready to state the main result.

Theorem 2.1 *Suppose, the conditions (A.1)–(A.9) hold true. Then, under both null and alternative hypotheses, as $n_1 \wedge n_2 \rightarrow \infty$,*

$$\sup_{a \leq t \leq b} \left| \frac{N^{1/2}}{\sqrt{\tau_n^2(b)}} (U_n(t) - V_n(t) - W_n(t)) \right| = o_P(1). \tag{15.15}$$

Here, U_n is given in (15.6) with $\hat{\mu}_1, \hat{\mu}_2$ of (15.12), and V_n and W_n are given in (15.13) and (15.14). Consequently,

$$\frac{N^{1/2}}{\sqrt{\tau_n^2(b)}} (U_n(t) - W_n(t)) \implies B \circ \varphi(t), \quad \varphi(t) = \lim_{n_1 \wedge n_2 \rightarrow \infty} \frac{\tau_n^2(t)}{\tau_n^2(b)}, \tag{15.16}$$

in the Skorohod space $D[a, b]$, where $B \circ \varphi$ is a continuous Brownian motion on $[a, b]$ with respect to time φ . Therefore, under H_0 , T of (15.8) satisfies

$$T \implies \sup_{0 \leq t \leq 1} |B(t)|,$$

where $B(t)$ is a continuous Brownian motion on \mathbb{R} .

Proof: The proof is given in Sect. 15.5.

Next, we need the following additional assumption to obtain the asymptotic distribution of \hat{T} given in (15.11)

Assumption 2.1 *Let $\tilde{\mu}_i, \hat{f}_i$ be estimators of μ_i and f_i , respectively, satisfying*

$$\sup_{a \leq x \leq b} |\tilde{\mu}_i(x) - \mu_i(x)| = o_P(1), \quad \sup_{a \leq x \leq b} |\hat{f}_i(x) - f_i(x)| = o_P(1), \quad i = 1, 2,$$

under both null and alternative hypotheses.

Corollary 2.1 *Suppose the conditions of Theorem 2.1 hold true. In addition, suppose that there are estimates $\tilde{\mu}_i$ and \hat{f}_i in (15.10) satisfying Assumption 2.1. Then, as $n_1 \wedge n_2 \rightarrow \infty$ and under H_0 , \hat{T} of (15.11) satisfies*

$$\hat{T} \implies \sup_{0 \leq t \leq 1} |B(t)|.$$

Proof: It suffices to prove (15.9). Let

$$O_{1n} = E \left\{ \sigma_1^2(Y_{1,0}) \left(1 + \frac{f_2(Y_{1,0})}{f_1(Y_{1,0})} \right)^2 1_{[a \leq Y_{1,0} \leq b]} \right\}$$

and

$$O_{1n} = \frac{1}{n_1} \sum_{i=1}^{n_1} \left\{ (Y_{1,i} - \tilde{\mu}_1(Y_{1,i-1}))^2 \left(1 + \frac{\hat{f}_2(Y_{1,i-1})}{\hat{f}_1(Y_{1,i-1})} \right)^2 1_{[a \leq Y_{1,i-1} \leq b]} \right\}.$$

By Assumption 2.1, (A.4), (A.5) and Chebyshev’s inequality, it can be derived that

$$O_{1n} = O_1 + o_P(1). \tag{15.17}$$

Similarly, we obtain

$$O_{2n} = O_2 + o_P(1), \tag{15.18}$$

with

$$O_2 = E \left\{ \sigma_2^2(Y_{2,0}) \left(1 + \frac{f_1(Y_{2,0})}{f_2(Y_{2,0})} \right)^2 1_{[a \leq Y_{2,0} \leq b]} \right\}$$

and

$$O_{2n} = \frac{1}{n_2} \sum_{j=1}^{n_2} \left\{ (Y_{2,j} - \tilde{\mu}_2(Y_{2,j-1}))^2 \left(1 + \frac{\hat{f}_1(Y_{2,j-1})}{\hat{f}_2(Y_{2,j-1})} \right)^2 1_{[a \leq Y_{2,j-1} \leq b]} \right\}.$$

From (15.17), (15.18) and the fact that O_1, O_2 are some positive constants, we have

$$\frac{\hat{\tau}_n^2(b) - \hat{\tau}_n^2(a)}{\tau_n^2(b) - \tau_n^2(a)} = \frac{q_1(O_1 + o_P(1)) + q_2(O_2 + o_P(1))}{q_1 O_1 + q_2 O_2} \rightarrow_P 1,$$

which completes the proof of the corollary. □

Remark 2.1 An example of estimates $\tilde{\mu}_i$ and \hat{f}_i satisfying Assumption 2.1 are: $\tilde{\mu}_i = \hat{\mu}_i$ of (15.12) and

$$\hat{f}_i(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} K_{h_i}(Y_{i,j-1} - x), \quad i = 1, 2, \tag{15.19}$$

with h_1, h_2 being appropriate bandwidths that could be different for constructing $\hat{\mu}_i$ in U_n of (15.6). For example, here we can take $h_i = O(n_i^{-1/5})$, See Bosq (1998). But to construct $\hat{\mu}_i$ in U_n of (15.6), we need to choose bandwidths that satisfy (A.3).

Remark 2.2 Testing property of \hat{T} : Under the model (15.1), consider the following alternative that is the same as in (15.2):

$$H_a : \quad \mu_1(x) - \mu_2(x) = \delta(x) \neq 0 \quad \text{for some } x \in [a, b],$$

where δ is continuous on $[a, b]$ since μ_1, μ_2 are continuous.

Theorem 2.1 and its corollary suggest to reject the null hypothesis for large values of \hat{T} given in (15.11) under Assumption 2.1.

Let

$$\hat{T}(t) = \frac{N^{1/2}}{\sqrt{\hat{\tau}_n^2(b)}} U_n(t), \quad \hat{T}_1(t) = \frac{N^{1/2}}{\sqrt{\hat{\tau}_n^2(b)}} (U_n(t) - W_n(t))$$

Then,

$$\hat{T}(t) = \hat{T}_1(t) + h(t), \quad h(t) = \frac{N^{1/2}}{\sqrt{\hat{\tau}_n^2(b)}} W_n(t)$$

By Ergodic theorem,

$$\begin{aligned} W_n &\rightarrow_P \int_a^t \delta(x) dx, \\ h(t) &\sim \frac{N^{1/2}}{\sqrt{\hat{\tau}_n^2(b)}} \int_a^t \delta(x) dx. \end{aligned} \tag{15.20}$$

This, together with $\frac{N^{1/2}}{\sqrt{\hat{\tau}_n^2(b)}} \rightarrow \infty$, (15.9), and the fact that $\int_a^t \delta(x) dx$ is not 0 for some $a \leq t \leq b$ implies,

$$\sup_{a \leq t \leq b} |h(t)| \rightarrow_P \infty. \tag{15.21}$$

Hence, in view of (15.16) and (15.9),

$$\hat{T} = \sup_{a \leq t \leq b} |\hat{T}(t)| = \sup_{a \leq t \leq b} |\hat{T}_1(t) + h(t)| \rightarrow_P \infty. \tag{15.22}$$

This, together with Corollary 2.1 indicates that the test bases on \hat{T} is consistent for H_a .

Note: By using the same arguments as above, we even can claim that under Assumption 2.1, the test based on \hat{T} is consistent for the alternatives converging to the null hypothesis at any rate α_n that is lower than $N^{-1/2}$, since (15.21) is still satisfied when $\delta(x)$ is replaced by $\delta(x)\alpha_n$. Furthermore, under $H_1 : \mu_1(x) - \mu_2(x) = \frac{\sqrt{\hat{\tau}_n^2(b)}}{N^{1/2}}\delta(x)$, $x \in [0, 1]$, the limiting powers of the asymptotic level α tests T is computed as

$$\lim_{n \rightarrow \infty} P(\hat{T}_1 > b_\alpha) = P\left(\sup_{a \leq t \leq b} |B \circ \varphi(t) + g(t)| > b_\alpha\right),$$

where b_α is defined such that

$$P\left(\sup_{0 \leq t \leq 1} |B(t)| > b_\alpha\right) = \alpha.$$

15.3 Simulation

In this section, we investigate the finite sample behavior of the nominal level of the proposed test \hat{T} under H_0 and power of \hat{T} against some nonparametric alternatives. As sample sizes, we choose the moderate sample sizes $n_1 = n_2 = n = 50, 100, 300$,

Table 15.1 The critical values b_α

α	0.05	0.025	0.01
b_α	2.24241	2.49771	2.80705

600, 1,000, and 2,000 with each simulation being repeated for 1,000 times. The data is simulated from model (15.1), where the two autoregressive functions are chosen to be $\mu_2(x) = 1 + x/2$ and $\mu_1(x) = \mu_2(x) + \delta(x)$, and the innovations $\{\varepsilon_{1,i}\}$ and $\{\varepsilon_{2,i}\}$ are taken to be independent standard normal $\mathcal{N}(0, 1)$. We choose $\delta(x) = 0$ corresponding to H_0 and $\delta(x) = 1, 2(x - 3)/x$, and $2 \left(\frac{n_1 n_2}{n_1 + n_2} \right)^{-1/2} = 2 N^{-1/2}$ corresponding to H_a . Note that the second choice of δ is negative for $x < 3$ and positive for $x > 3$ and converge to 0 for $x \rightarrow 3$; the last choice of δ corresponds to the local alternatives with a rate being the same as τ_n of (15.7). For simplicity, the conditional variance functions σ_1 and σ_2 are chosen to be (i) $\sigma_1(x) = \sigma_2(x) = 1$ and (ii) $\sigma_1(x) = \sigma_2(x) = 3/\sqrt{1 + x^2}$. Finally, the interval $[a, b]$ in (15.2) is taken to be $[2, 4]$.

To construct the test statistics \hat{T} of (15.11), we consider Nadaraya–Watson estimators $\hat{\mu}_1, \hat{\mu}_2$ of (15.12) with kernel $K(x) = \frac{3}{4}(1 - x^2)\mathbf{1}(|x| \leq 1)$ and we considered three different bandwidths $h_1 = h_2 = 0.15, 0.2$ and 0.25 . The estimates $\tilde{\mu}_1, \tilde{\mu}_2$ and \hat{f}_1, \hat{f}_2 in $\hat{\tau}_n$ of (15.10) are from Remark 2.1 with $h_i = n_i^{-1/5}, i = 1, 2$. Let b_α satisfy $P(\sup_{0 \leq t \leq 1} |B(t)| > b_\alpha) = \alpha$. Then, the empirical size (power) is computed by the proportion of rejects $\frac{\# \text{ of } \{\hat{T} > b_\alpha\}}{1000}$.

In Table 15.1, we give the critical values b_α obtained from the formula $P(\sup_{0 \leq t \leq 1} |B(t)| < b) = P(|B(1)| < b) + 2 \sum_{i=1}^\infty (-1)^i P((2i - 1)b < B(1) < (2i + 1)b)$ given on page 553 of the book by Resnick (1992).

The simulation programming was done using R. To generate each of the two samples, we first generated $(500 + n)$ error variables from $\mathcal{N}(0, 1)$. Using these errors and model (15.1) with the initial value $Y_{i,0}$ randomly chosen from $\mathcal{N}(0, 1)$, we generated $(501 + n)$ observations. The last $(n + 1)$ observations from the data thus generated are used in carrying out the simulation study.

The results of the simulation study are shown in Table 15.2 below. Three rows correspond each choice of $\delta(x)$ with the first row corresponding to bandwidth 0.15, the second to 0.2 and the third to 0.25. The finite sample level and power behavior of the tests are shown to be quite stable across the various choices of the bandwidth. One sees that for both choices of σ_1 and σ_2 , the empirical sizes of the test are not much different from the nominal levels for most moderate samples sizes, but they are closer to the true levels when the sample size gets larger. The simulated powers under fixed alternative $\delta(x) = 1$ are close to 1 for all moderate sample sizes, even at α -level .025. The simulated powers under fixed alternative $\delta(x) = 2(x - 3)/x$ are seen to increase quickly with n and they are quite large for $n \geq 600$. The simulated powers under local alternative $\delta(x) = 2 N^{-1/2}$ are stable for most moderate sample sizes. In summary, the simulated levels and powers are consistent with the asymptotic theory at most moderate sample sizes considered.

Table 15.2 Proportion of rejections ($\hat{T} > 2.24241 (2.49771)$) at level $\alpha = .05 (.025)$ for (i) $\sigma_1 = 1 = \sigma_2$ and (ii) $\sigma_1(x) = 3/\sqrt{1+x^2} = \sigma_2(x)$. Three rows correspond each choice of $\delta(x)$ with the first row corresponding to bandwidth 0.15, the second to 0.2 and the third to 0.25

α -level	$\delta(x) \setminus n$	(i)						(ii)					
		50	100	300	600	1,000	2,000	50	100	300	600	1,000	2,000
0.05	0	0.085	0.056	0.048	0.048	0.036	0.034	0.070	0.062	0.041	0.038	0.044	0.043
		0.090	0.057	0.039	0.040	0.040	0.044	0.084	0.049	0.040	0.042	0.037	0.032
		0.085	0.050	0.042	0.041	0.035	0.045	0.068	0.057	0.042	0.030	0.049	0.044
1		0.873	0.989	1	1	1	1	0.910	0.993	1	1	1	1
		0.897	0.996	1	1	1	1	0.918	0.995	1	1	1	1
		0.916	0.996	1	1	1	1	0.930	0.995	1	1	1	1
0.025	$2(x - 3)/x$	0.191	0.222	0.474	0.827	0.968	1	0.147	0.194	0.349	0.657	0.888	0.998
		0.219	0.247	0.483	0.827	0.970	1	0.165	0.187	0.356	0.655	0.886	0.998
		0.212	0.221	0.449	0.798	0.973	1	0.146	0.181	0.337	0.671	0.878	0.998
1	$2 N^{-1/2}$	0.335	0.317	0.266	0.228	0.243	0.262	0.335	0.295	0.349	0.244	0.221	0.207
		0.333	0.313	0.245	0.258	0.232	0.195	0.320	0.294	0.227	0.209	0.192	0.222
		0.342	0.278	0.256	0.256	0.243	0.232	0.331	0.256	0.238	0.234	0.218	0.214
0.025	0	0.048	0.031	0.028	0.023	0.020	0.020	0.043	0.040	0.025	0.016	0.024	0.025
		0.061	0.029	0.022	0.015	0.020	0.028	0.039	0.036	0.023	0.020	0.024	0.015
		0.055	0.034	0.022	0.020	0.018	0.021	0.041	0.031	0.021	0.017	0.025	0.019
1		0.824	0.982	1	1	1	1	0.870	0.985	1	1	1	1
		0.841	0.991	1	1	1	1	0.885	0.992	1	1	1	1
		0.874	0.990	1	1	1	1	0.892	0.991	1	1	1	1
0.025	$2(x - 3)/x$	0.141	0.154	0.357	0.730	0.937	1	0.103	0.126	0.248	0.533	0.803	0.991
		0.160	0.163	0.356	0.722	0.924	1	0.119	0.134	0.254	0.522	0.807	0.994
		0.148	0.158	0.327	0.704	0.933	1	0.104	0.121	0.224	0.531	0.801	0.994
1	$2 N^{-1/2}$	0.259	0.227	0.193	0.167	0.161	0.177	0.264	0.214	0.167	0.173	0.155	0.150
		0.255	0.225	0.160	0.185	0.171	0.133	0.254	0.226	0.166	0.155	0.132	0.150
		0.248	0.188	0.173	0.170	0.164	0.172	0.246	0.176	0.171	0.155	0.146	0.149

15.4 Properties of Kernel Smoothers and Weak Convergence of Empirical Processes

In this section, we first study the asymptotic behavior of the following kernel smoothers over $[a, b]$ for $i = 1, 2$:

$$\hat{f}_i(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} K_{h_i}(Y_{i,j-1} - x), \quad \Lambda_{i,n}(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} \sigma_i(Y_{i,j-1}) \varepsilon_{i,j} K_{h_i}(Y_{i,j-1} - x), \quad (15.23)$$

$$\Psi_i(x, y) = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{K_{h_i}(Y_{i,j-1} - x)}{\Psi_i(Y_{i,j-1})} 1_{[Y_{i,j-1} \leq y]}, \quad \Psi_1 = f_2, \quad \Psi_2 = f_1, \quad (15.24)$$

$$\Gamma_1(y) = \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \Psi_2(Y_{1,i-1}, y), \quad \Gamma_2(y) = \frac{1}{n_2} \sum_{i=1}^{n_2} \varepsilon_{2,i} \Psi_1(Y_{2,i-1}, y), \quad (15.25)$$

$$H_{i,j}(y) = \frac{1}{n_i} \sum_{k=1}^{n_i} K_{h_i}(Y_{i,k-1} - y)(Y_{i,k-1} - y)^j, \quad i, j = 1, 2. \quad (15.26)$$

By Lemma 1–4 in Li (2008), we have the following results.

Lemma 4.1 *Suppose conditions (A.2), (A.3) and (A.4)–(A.8) hold. Then,*

$$\sup_{a \leq x \leq b} |\Lambda_{i,n}(x)| = O_P \left(\sqrt{\frac{\log n_i}{n_i h_i}} \right), \quad i = 1, 2,$$

where $\Lambda_{i,n}(x)$ is given in (15.23).

Lemma 4.2 *Suppose conditions (A.2), (A.3), (A.4), (A.6) and (A.8) hold. Then f_i of (15.23) satisfies*

$$\sup_{a \leq x \leq b} |\hat{f}_i(x) - f_i(x)| = O_P \left(\sqrt{\frac{\log n_i}{n_i h_i}} \right) + O(h_i^2), \quad i = 1, 2.$$

Lemma 4.3 *Suppose condition (A.2), (A.3), (A.4), (A.6) and (A.8) hold. Then, $\Psi_i(x, y)$ of (15.24) satisfies*

$$\sup_{\substack{a - h_i \leq x \leq b + h_i \\ a \leq y \leq b}} \text{Var}\{\Psi_i(x, y)\} = O\left(\frac{1}{n_i h_i}\right), \quad i = 1, 2.$$

Lemma 4.4 *Suppose condition (A.2)- (A.4), (A.6) and (A.8) hold. Then $H_{i,j}$ of (15.26) satisfies*

$$H_{i,j}(y) = h_i^j f_i(y) u_j + O_P \left(h_i^{j+1} + h_j^j \sqrt{\frac{\log n_i}{n_i h_i}} \right),$$

$$u_j = \int_{-1}^1 K(u) u^j du, \quad i, j = 1, 2$$

uniformly on $a \leq x \leq b$.

Next, we study the property of some empirical processes. The weak convergence of marked empirical process proved in Theorem 2.2.6 of Koul (2002) implied the following lemma:

Lemma 4.5 *Suppose conditions (A.4), (A.6), (A.7) and (A.9) hold, then for $i = 1, 2$,*

$$\sup_{a \leq y \leq b} \left| \frac{1}{n_i} \sum_{j=1}^{n_i} (|\varepsilon_{i,j}| - E|\varepsilon_{i,j}|) \mathbf{1}_{[Y_{i,j-1} \leq y]} \right| = O_P(n_i^{-1/2}).$$

Next, recall that $\mathcal{K}(y) = \int_{-1}^y K(t) dt$ is the distribution function corresponding to the kernel density $K(y)$ on $[-1, 1]$. We prove the following lemma:

Lemma 4.6 *Suppose conditions (A.2), (A.3), (A.4), (A.6) and (A.8) hold. Then Γ_i of (15.25) satisfies*

$$\sup_{a \leq y \leq b} \left| N^{1/2} \left(\Gamma_1(y) - \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \mathcal{K} \left(\frac{y - Y_{1,i-1}}{h_2} \right) \right) \right| = o_P(1),$$

$$\sup_{a \leq y \leq b} \left| N^{1/2} \left(\Gamma_2(y) - \frac{1}{n_2} \sum_{i=1}^{n_2} \varepsilon_{2,i} \frac{f_1(Y_{2,i-1})}{f_2(Y_{2,i-1})} \mathcal{K} \left(\frac{y - Y_{2,i-1}}{h_1} \right) \right) \right| = o_P(1).$$

Proof: The proof is similar to the proof of Lemma 1 of Li (2008) which is in turn similar to Lemma 6.1 of Fan and Yao (2003). It is sufficient to prove the first equality. Let C denote a generic constant, which can vary from one place to another. Also let

$$N^{1/2} \left(\Gamma_1(y) - \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \mathcal{K} \left(\frac{y - Y_{1,i-1}}{h_2} \right) \right) = A_n(y)$$

Now, decompose $A_n(y)$ into $A_{1,n}(y) + A_{2,n}(y)$ with

$$A_{1,n}(y) = N^{1/2} \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} (\Psi_2(Y_{1,i-1}, y) - E(\Psi_2(Y_{1,i-1}, y))),$$

$$A_{2,n}(y) = N^{1/2} \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \left(E(\Psi_2(Y_{1,i-1}, y)) - \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \mathcal{K} \left(\frac{y - Y_{1,i-1}}{h_2} \right) \right).$$

First, we show $\sup_{a \leq y \leq b} |A_{2,n}(y)| = o_P(1)$. For some $(Y_{1,i-1}^* \in [a - 2h_2, b + 2h_2])$, By Taylor expansion, We have

$$\begin{aligned}
 & A_{2,n}(y) \\
 &= N^{1/2} \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \left(\int_{-1}^{\frac{y-Y_{1,i-1}}{h_2}} K(u) \left(\frac{f_2(Y_{1,i-1} + h_2u)}{f_1(Y_{1,i-1} + h_2u)} - \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \right) du \right) \\
 &= N^{1/2} \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \left(h_2 \int_{-1}^{\frac{y-Y_{1,i-1}}{h_2}} uK(u) \frac{f_2'(Y_{1,i-1})f_1(Y_{1,i-1}) - f_1'(Y_{1,i-1})f_2(Y_{1,i-1})}{f_1(Y_{1,i-1} + h_2u)f_1(Y_{1,i-1})} du \right. \\
 &\quad \left. + \frac{h_2^2}{2} \int_{-1}^{\frac{y-Y_{1,i-1}}{h_2}} u^2 K(u) \frac{f_2''(Y_{1,i-1}^*)f_1(Y_{1,i-1}) - f_1''(Y_{1,i-1}^*)f_2(Y_{1,i-1})}{f_1(Y_{1,i-1} + h_2u)f_1(Y_{1,i-1})} du \right) \\
 &\leq N^{1/2} h_2 \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \int_{-1}^{\frac{y-Y_{1,i-1}}{h_2}} uK(u) \frac{f_2'(Y_{1,i-1})f_1(Y_{1,i-1}) - f_1'(Y_{1,i-1})f_2(Y_{1,i-1})}{f_1(Y_{1,i-1} + h_2u)f_1(Y_{1,i-1})} du \\
 &\quad + N^{1/2} h_2^2 \frac{1}{n_1} \sum_{i=1}^{n_1} |\varepsilon_{1,i}| \cdot C, \quad \text{by (A.2) and (A.4),}
 \end{aligned}$$

uniformly over $[a, b]$,

By a similar argument in proving Lemma 4.1 or Lemma 1 of Li (2008), it can be shown that

$$\begin{aligned}
 & \sup_{a \leq y \leq b} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \int_{-1}^{\frac{y-Y_{1,i-1}}{h_2}} uK(u) \frac{f_2'(Y_{1,i-1})f_1(Y_{1,i-1}) - f_1'(Y_{1,i-1})f_2(Y_{1,i-1})}{f_1(Y_{1,i-1} + h_2u)f_1(Y_{1,i-1})} du \right| \\
 &= O_P \left(\sqrt{\frac{\log n_1}{n_1 h_2}} \right).
 \end{aligned}$$

Also, $N^{1/2} h_2^2 \frac{1}{n_1} \sum_{i=1}^{n_1} |\varepsilon_{1,i}| = O_P(N^{1/2} h_2^2) = o_P(1)$ by (A.3). Hence, by (A.3) we have

$$\begin{aligned}
 \sup_{a \leq y \leq b} |A_{2,n}(y)| &= O_P \left(\sqrt{q_1 h_2 \log n_1} \right) + o_P(1) \\
 &= O_P \left(\sqrt{q_1 h_2 \log \frac{N}{q_1}} \right) + o_P(1) = o_P(1). \tag{15.27}
 \end{aligned}$$

Now, it is left to prove

$$\sup_{a \leq y \leq b} |A_{1,n}(y)| = o_P(1). \tag{15.28}$$

Slightly simpler than the proof of Lemma 1 in Li (2008) and Lemma 6.1 in Fan and Yao (2003), the proof consists of the following two steps:

(a) (Discretization). Partition the interval $[a, b]$ with length L into $M = \lceil (N^{1+c})^{1/2} \rceil$ subintervals $\{I_k\}$ of equal length. Let $\{y_k\}$ be the centers of I_k . Then

$$\sup_{a \leq y \leq b} |A_{1,n}(y)| \leq \max_{1 \leq k \leq M} |A_{1,n}(y_k)| + o_P(1). \tag{15.29}$$

(b) (Maximum deviation for discretized series). For any small ϵ ,

$$P \left(\max_{1 \leq j \leq M} |A_{1,n}(y_j)| > \epsilon \right) \rightarrow 0. \tag{15.30}$$

Let $G_{i,n}(y) = \sqrt{n_i} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} 1_{[Y_{i,j-1} \leq y]} - P(Y_{i,j-1} \leq y) \right)$. The strong approximation theorem for the empirical process of a stationary sequence of strong mixing random variables established in Berkes and Philipp (1997) and in Theorem 4.3 of the monograph edited by Dehling et al (2002) implied

$$\sup_{1 \leq k \leq M} \sup_{y \in I_k} |G_{1,n}(y) - G_{1,n}(y_k)| = o_P(1). \tag{15.31}$$

Now, we prove part (a). First, for any $1 \leq K \leq M$ and all $y \in I_k$, we decompose $A_{1,n}(y) - A_{1,n}(y_k)$ as $D_{1,n}(y) + D_{2,n}(y)$ with

$$D_{1,n}(y) = N^{1/2} \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} (\Psi_2(Y_{1,i-1}, y) - \Psi_2(Y_{1,i-1}, y_k)),$$

$$D_{2,n}(y) = N^{1/2} \frac{1}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} (E(\Psi_2(Y_{1,i-1}, y)) - E(\Psi_2(Y_{1,i-1}, y_k))),$$

Without losing generality, it is sufficient to consider all $y_k \leq y \in I_k$. It is easy to see that

$$\begin{aligned} & |D_{1,n}(y)| \\ & \leq CN^{1/2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} |\varepsilon_{1,i}| 1_{[y_k - h_2 \leq Y_{1,i-1} \leq y + h_2]} \right) \left(\frac{1}{n_2 h_2} \sum_{i=1}^{n_2} 1_{[y_k \leq Y_{2,i-1} \leq y]} \right) \\ & \leq CN^{1/2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} (|\varepsilon_{1,i}| - E|\varepsilon_{1,i}|) 1_{[y_k - h_2 \leq Y_{1,i-1} \leq y + h_2]} + \frac{1}{n_1} \sum_{i=1}^{n_1} 1_{[y_k - h_2 \leq Y_{1,i-1} \leq y + h_2]} \right) \\ & \quad \left(\frac{1}{h_2} \left[\frac{1}{\sqrt{n_2}} (G_{2,n}(y) - G_{2,n}(y_k)) + P(y_k \leq Y_{2,i-1} \leq y) \right] \right) \\ & = C \frac{N^{1/2}}{h_2} \left(O_P \left(\frac{1}{\sqrt{n_1}} \right) + \frac{1}{\sqrt{n_1}} (G_{1,n}(y + h_2) - G_{1,n}(y_k - h_2)) \right. \\ & \quad \left. + P(y_k - h_2 \leq y_{1,i-1} \leq y + h_2) \right) \left(o_P \left(\frac{1}{\sqrt{n_2}} \right) + O_P \left(\frac{1}{M} \right) \right) \text{ by Lemma 4.5, (15.31)} \\ & = C \frac{N^{1/2}}{h_2} \left(O_P \left(\frac{1}{\sqrt{n_1}} \right) + O_P(h_2) \right) \left(o_P \left(\frac{1}{\sqrt{n_2}} \right) + O_P \left(\frac{1}{M} \right) \right) \text{ again by (15.31)} \\ & = o_P(1), \end{aligned}$$

and similarly,

$$\begin{aligned} |D_{2,n}(y)| &\leq CN^{1/2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} |\varepsilon_{1,i}| \mathbf{1}_{[y_k - h_2 \leq Y_{1,i-1} \leq y + h_2]} \right) \left(\frac{1}{h_2} P(y_k \leq Y_{2,i-1} \leq y) \right) \\ &= C \frac{N^{1/2}}{h_2} O_P(h_2) O_P \left(\frac{1}{N^{1/2+c/2}} \right) = o_P(1). \end{aligned}$$

Hence, we have

$$\sup_{1 \leq k \leq M} \sup_{y \in I_j} |A_{1,n}(y) - A_{1,n}(y_k)| = o_P(1).$$

This proves part (a). Next,

$$\begin{aligned} P(\max_{1 \leq k \leq M} |A_{1,n}(y_k)| > \epsilon) &\leq M \max_{1 \leq k \leq M} E(A_{1,n}^2(y_k)) / \epsilon^2 \\ &= N^{1/2+c/2} N \frac{1}{n_1} O \left(\frac{1}{n_2 h_2} \right) \rightarrow 0, \quad \text{by Lemma 4.3 and (A.3).} \end{aligned}$$

This proves part (b) and hence finishes the proof of (15.28) and the Lemma. \square

15.5 Proofs

Here, we shall give the proof of our main result, Theorem 2.1. The lemmas proved in Sect. 15.4 will facilitate the proof of this theorem. As usual, let C be a generic constant. It suffices to prove (15.15) and (15.16). Now consider $N^{1/2}U_n(t)$ for all $a \leq t \leq b$. We decompose $N^{1/2}U_n(t)$ as $B_{1,n}(t) - B_{2,n}(t)$ with

$$\begin{aligned} B_{1,n}(t) &= N^{1/2} \frac{1}{n_1} \sum_{i=1}^{n_1} (Y_{1,i} - \hat{\mu}_2(Y_{1,i-1})) \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]}, \\ B_{2,n}(t) &= N^{1/2} \frac{1}{n_2} \sum_{i=1}^{n_2} (Y_{2,i} - \hat{\mu}_1(Y_{2,i-1})) \mathbf{1}_{[a \leq Y_{2,i-1} \leq t]}. \end{aligned}$$

We first consider $B_{1,n}(t)$. Recall definitions (15.12) and (15.23)–(15.25). By decomposition and simple algebra, we rewrite $B_{1,n}(t)$ as $I(t) - II(t) + III(t)$ with

$$I(t) = \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} (\varepsilon_{1,i} \sigma_1(Y_{1,i-1}) + (\mu_1(Y_{1,i-1}) - \mu_2(Y_{1,i-1}))) \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]} \quad (15.32)$$

$$II(t) = \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \frac{\sum_{j=1}^{n_2} \varepsilon_{2,j} \sigma_2(Y_{2,j-1}) K_{h_2}(Y_{2,j-1} - Y_{1,i-1})}{n_2 f_2(Y_{1,n-1})} \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]} \quad (15.33)$$

$$III(t) = \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \frac{\sum_{j=1}^{n_2} (\mu_2(Y_{1,i-1}) - \mu_2(Y_{2,j-1})) K_{h_2}(Y_{2,j-1} - Y_{1,i-1})}{n_2 \hat{f}_2(Y_{1,i-1})} \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]} \quad (15.34)$$

Now we consider $II(t)$. By decomposition, we rewrite it as $II_1(t) + II_2(t)$ with

$$\begin{aligned}
 II_1(t) &= N^{1/2} \frac{1}{n_2} \sum_{j=1}^{n_2} \varepsilon_{2,j} \sigma_2(Y_{2,j-1}) (\Psi_1(Y_{2,j-1}, t) - \Psi_1(Y_{2,j-1}, a)) \\
 &= N^{1/2} \frac{1}{n_2} \sum_{j=1}^{n_2} \varepsilon_{2,j} \sigma_2(Y_{2,j-1}) \frac{f_1(Y_{2,j-1})}{f_2(Y_{2,j-1})} \left(\mathcal{K} \left(\frac{t - Y_{2,j-1}}{h_1} \right) \right. \\
 &\quad \left. - \mathcal{K} \left(\frac{a - Y_{2,j-1}}{h_1} \right) \right) + o_P(1), \tag{15.35}
 \end{aligned}$$

uniformly on $a \leq t \leq b$ by Lemma 4.6 and its proof, and uniformly on $a \leq t \leq b$,

$$\begin{aligned}
 II_2(t) &= \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \frac{\sum_{j=1}^{n_2} \varepsilon_{2,j} \sigma_2(Y_{2,j-1}) K_{h_2}(Y_{2,j-1} - Y_{1,i-1})}{n_2 f_2(Y_{1,i-1})} \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]} \\
 &\quad \frac{f_2(Y_{1,i-1}) - \hat{f}_2(Y_{1,i-1})}{\hat{f}_2(Y_{1,i-1})} \\
 &\leq C \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \left| \frac{1}{n_2} \sum_{j=1}^{n_2} \varepsilon_{2,j} \sigma_2(Y_{2,j-1}) K_{h_2}(Y_{2,j-1} - Y_{1,i-1}) \right| \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]} \\
 &\quad \cdot \sup_{a \leq Y_{1,i-1} \leq t} \left| \frac{f_2(Y_{1,i-1}) - \hat{f}_2(Y_{1,i-1})}{\hat{f}_2(Y_{1,i-1})} \right| \\
 &= CN^{1/2} O_P \left(\sqrt{\frac{\log n_2}{n_2 h_2}} \right) \cdot \left(O_P \left(\sqrt{\frac{\log n_2}{n_2 h_2}} \right) + O_P(h_2^2) \right), \text{ by Lemma 4.1 and 4.2} \\
 &= o_P(1), \quad \text{by (A.3)}.
 \end{aligned}$$

Next, we consider $III(t)$. Let $\mu_2^{(1)}$ denote the first derivative of μ_2 . Then, by (A.1), uniformly on $a \leq t \leq b$,

$$\begin{aligned}
 III(t) &\leq \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \frac{\mu_2^{(1)}(Y_{1,i-1}) |H_{2,1}(Y_{1,i-1})| + C |H_{2,2}(Y_{1,i-1})|}{\hat{f}_2(Y_{1,i-1})} \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]} \\
 &= O_P \left(N^{1/2} \left(h_2^2 + h_2 \sqrt{\frac{\log n_2}{n_2 h_2}} \right) \right) = o_P(1), \text{ by Lemma 4 and (A.3)} \tag{15.36}
 \end{aligned}$$

Hence, by (15.32) and (15.35)–(15.36), we have uniformly on $a \leq t \leq b$,

$$\begin{aligned} B_{1,n}(t) &= \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} (\varepsilon_{1,i} \sigma_1(Y_{1,i-1}) + (\mu_1(Y_{1,i-1}) - \mu_2(Y_{1,i-1}))) \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]} \\ &\quad - \frac{N^{1/2}}{n_2} \sum_{j=1}^{n_2} \varepsilon_{2,j} \sigma_2(Y_{2,j-1}) \frac{f_1(Y_{2,j-1})}{f_2(Y_{2,j-1})} \left(\mathcal{K} \left(\frac{t - Y_{2,j-1}}{h_1} \right) \right) \\ &\quad - \mathcal{K} \left(\frac{a - Y_{2,j-1}}{h_1} \right) \Big) + o_P(1) \end{aligned} \quad (15.37)$$

Similarly, we have uniformly on $a \leq t \leq b$,

$$\begin{aligned} B_{2,n}(t) &= \frac{N^{1/2}}{n_2} \sum_{j=1}^{n_2} (\varepsilon_{2,j} \sigma_2(Y_{2,j-1}) + (\mu_2(Y_{2,j-1}) - \mu_1(Y_{2,j-1}))) \mathbf{1}_{[a \leq Y_{2,j-1} \leq t]} \\ &\quad - \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \sigma_1(Y_{1,i-1}) \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \left(\mathcal{K} \left(\frac{t - Y_{1,i-1}}{h_2} \right) \right) \\ &\quad - \mathcal{K} \left(\frac{a - Y_{1,i-1}}{h_2} \right) \Big) + o_P(1) \end{aligned} \quad (15.38)$$

By (15.37) and (15.38), we proved (15.15).

Now, we need to prove (15.16). Applying the CLT for martingales [Hall and Heyde (1980), Corollary 3.1], we first could show that the finite-dimensional distributions of $\frac{N^{1/2}}{\tau_n^2(b)} V_n(t)$ tend to the right limit. Then, apply theorem for weak convergence on functional space [Hall and Heyde (1980), Theorem A.2], we need to prove the tightness of $\frac{N^{1/2}}{\tau_n^2(b)} V_n(t)$. It suffices to prove the tightness of

$$\frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \sigma_1(Y_{1,i-1}) \mathbf{1}_{[a \leq Y_{1,i-1} \leq t]}$$

and

$$\frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \sigma_1(Y_{1,i-1}) \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \mathcal{K} \left(\frac{t - Y_{1,i-1}}{h_2} \right).$$

The tightness of the first sequence is implied by the weak convergence of a marked empirical process [Koul and Stute (1999), Lemma3.1].

Since $\mathcal{K} \left(\frac{t - Y_{1,i-1}}{h_2} \right) = 1$ for $Y_{1,i-1} \leq t - h_2$ and $\mathcal{K} \left(\frac{t - Y_{1,i-1}}{h_2} \right) \mathbf{1}_{[t - h_2 \leq Y_{1,i-1} \leq t + h_2]}$ just behaves like $h_2 K_{h_2}(t - Y_{1,i-1})$, the second sequence can be rewritten as

$$\begin{aligned} &\frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \sigma_1(Y_{1,i-1}) \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \mathbf{1}_{[Y_{1,i-1} \leq t - h_2]} \\ &\quad + \frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \sigma_1(Y_{1,i-1}) \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} \mathcal{K} \left(\frac{t - Y_{1,i-1}}{h_2} \right) \mathbf{1}_{[t - h_2 \leq Y_{1,i-1} \leq t + h_2]}, \end{aligned}$$

with second term being $o_p(1)$ uniformly on $a \leq t \leq b$ by a proof similar to that of Lemma 4.1. Again, by the weak convergence of a marked empirical process [Koul and Stute (1999), Lemma 3.1], we could prove the tightness of $\frac{N^{1/2}}{n_1} \sum_{i=1}^{n_1} \varepsilon_{1,i} \sigma_1(Y_{1,i-1}) \frac{f_2(Y_{1,i-1})}{f_1(Y_{1,i-1})} 1_{[Y_{1,i-1} \leq t-h_2]}$. Therefore, we complete the proof of the main theory. \square

Acknowledgements The author is Dr. Hira L. Koul's 22nd Ph.D. student. She graduated in 2004 and would like to thank Hira for his patient guidance and generous support for more than ten years now. The author is also very grateful to Hira for his academic advice and knowledge and many insightful discussions and suggestions. This manuscript is specially dedicated to him on the occasion of his 70th Birthday.

References

- Berkes I, Philipp W (1997) Approximation theorems for independent and weakly dependent random vectors. *Ann Probab* 7:29–54
- Bloomfield P (1992) Trends in global temperatures. *Climatic Change* 21:275–287
- Bosq D (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer, New York
- Delgado MA (1993) Testing the equality of nonparametric regression curves. *Stat Probab Lett* 17:199–204
- Dehling H, Mikosch T, Sørensen M (2002) *Empirical Process Techniques for Dependent Data*. Birkhauser, Boston. MR1958776
- Fan J, Yao Q (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Verlag, New York
- Hall P, Hart JD (1990) Bootstrap test for difference between means in nonparametric regression. *J Amer Stat Assoc* 85:1039–1049
- Hall P, Heyde CC (1980) *Martingale limit theory and its Applications*. Academic Press, Inc., New York
- Koul HL (2002) *Weighted Empirical Processes in Dynamic Nonlinear Models*. Springer, New York
- Koul HL, Li F (2005) Testing for Superiority among two time series. *Statistical Inference for Stochastic Processes*. 8:109–135
- Koul HL, Schick A (1997) Testing the equality of two nonparametric regression curves. *J Stat Plann Inference* 65:293–314
- Kulasekera KB (1995) Comparison of regression curves using quasi-residuals. *J Am Stat Assoc* 90:1085–1093
- Li F (2008) Asymptotic properties of some kernel smoothers. Preprint. pr08-03, Department of Mathematical Sciences. Indiana University Purdue University Indianapolis.
- Li F (2009) Testing for the equality of two autoregressive functions using quasi residual. *Communicat Stat-Theory Meth* 38(9):1404–1421
- Nadaraya EA (1964) On estimating regression. *Theory Probab Appl* 9:141–142
- Resnick SI (1992) *Adventures in Stochastic Processes*. Springer, New York
- Scheike TH (2000) Comparison of non-parametric regression functions through their cumulatives. *Stat Probab Lett* 46:21–32
- Watson GS (1964) Smooth regression analysis. *Sankhyā Ser A* 26:359–372