ELSEVIER
Journal of Econometrics 92 (1999) 101–147
JOURNAL OF Econometrics
www.elsevier.nl/locate/econbase

Consistent model specification tests for time series econometric models

Qi Li*

Department of Economics, Texas A&M University, College Station, TX 77843, U.S.A., and Department of Economics, University of Guelph, Guelph, Ont., Canada, N1G 2W1

Received 1 August 1997; received in revised form 1 September 1998; accepted 8 October 1998

Abstract

In this paper we consider general hypothesis testing problems for nonparametric and semiparametric time-series econometric models. We apply the general methodology to construct a consistent test for omitted variables and a consistent test for a partially linear model. The proposed tests are shown to have asymptotic normal distributions under their respective null hypotheses. We also discuss the problems of testing portfolio conditional mean-variance efficiency and testing a semiparametric single index model. Monte Carlo simulations are conducted to examine the finite sample performances of the nonparametric omitted variable test and the test for a partially linear specification. © 1999 Published by Elsevier Science S.A. All rights reserved.

JEL classification: C12; C14.

Keywords: Consistent tests; Absolutely regular process; Degenerate U-statistics; Kernel estimation; Omitted variables; Partially linear model; Asymptotic normality

1. Introduction

There is a rich literature on constructing consistent model specification tests using nonparametric estimation techniques.[1] For example various test statistics for consistently testing a parametric regression functional form have been

* Tel.: 519 824 4120 x8945; Fax: 519 763 8497; e-mail: qi@css.uoguelph.ca.

[1] Bierens (1982) was the first to give a consistent conditional moment model specification test, see also Bierens and Ploberger (1997) and the references therein. Using nonparametric estimation technique to construct consistent model specification tests was first suggested by Ullah (1985). Robinson (1989) was the first to propose some nonparametric tests for time-series models.

PII: S0304-4076(98)00087-6

proposed by Bierens (1982, 1990), Bierens and Ploberger (1997), Eubank and Spiegelman (1990), Eubank and Hart (1992), Fan and Li (1992), Gozalo (1993), Härdle and Mammen (1993), Hong and White (1995), Horowitz and Härdle (1994), Li (1994), Robinson (1991), Wooldridge (1992), Yatchew (1992), and Zheng (1996), to mention only a few. Testing the insignificance of a subset of regressors (omitted variable test) have been considered by Ait-Sahalia et al. (1994), Fan and Li (1996a), Lavergne and Vuong (1996b) and Lewbel (1993). For consistent testing some semiparametric versus nonparametric regression models, see Ait-Sahalia et al. (1994), Fan and Li (1996a), Linton and Gozalo (1997) and Whang and Andrews (1993). Delgado and Stengos (1994) and Lavergne and Vuong (1996a) considered non-nested hypothesis testing problems. Lewbel (1993, 1995) considered general hypothesis testing problem with independent data. Robinson (1989) considered general hypothesis testing problems for time-series econometric models.

Most of the above-mentioned works deal with independent data. While Bierens and Ploberger (1996) and Fan and Li (1996b) allow for dependent data, both Bierens and Ploberger (1997) and Fan and Li (1996b) only considered the case of testing a parametric null model. Robinson (1989) considered general hypothesis testing problem with time-series data, however, his procedure may not produce *consistent* tests in the sense that there exist alternatives that cannot be detected by Robinson's (1989) testing procedure. Recently, Chen and Fan (1997) modify Robinson's general testing procedure and construct consistent test statistics for time-series models. Their idea is similar to the approach of Bierens (1982) and the asymptotic distributions of their test statistics are non-standard. Hence, they suggest to use the conditional Monte Carlo method of Hansen (1996) or the stationary bootstrap method of Politis and Romano (1994) to approximate the null distribution of their test statistics. While bootstrap methods are quite successful for providing reliable null approximations to test statistics with independent data (e.g., Härdle and Mammen, 1993; Fan and Linton, 1997; Li and Wang, 1998), they are less satisfactory with dependent observations. In this paper we consider the general hypothesis testing problem with time-series data and we establish the asymptotic normality of the proposed test statistics. Thus, our results generalize many testing results including those in Fan and Li (1996a) and Lavergne and Vuong (1996b) to time-series models. We also show that our testing procedure can be applied to a wide range of hypotheses testing problem (with weakly dependent data). The regularity conditions we use are quite weak and they are very similar to the conditions used for independent data cases (e.g., Fan and Li, 1996a). Some simple but important tricks are used in establishing the asymptotic normal distributions of the test statistics.

One leading case of the null hypotheses we consider in this paper is testing the significance of a subset of the regressors (a nonparametric omitted variable test). Ait-Sahalia et al. (1994), Fan and Li (1996a) and Lavergne and Vuong (1996b) all consider such a testing problem with independent data. Recently, Christoffersen and Hahn (1997) applied the test statistics proposed by the above authors to test whether ARCH volatility has additional explanatory power to option pricing given other relevant economic variables. When applying the test of Fan and Li (1996a) and Lavergne and Vuong (1996b) to a time-series model, Christoffersen and Hahn (1997) conjectured that this test is applicable with time-series data. This paper provides a formal proof that the nonparametric significant test proposed by Fan and Li (1996a) and Lavergne and Vuong (1996b) is indeed applicable for weakly dependent data.

The paper is organized as follows. In Section 2 we first describe a general testing procedure with time-series data. In Section 3 we apply the methodology presented in Section 2 to derive the asymptotic distribution of a nonparametric significance test. Section 4 presents a test for a partially linear model. Section 5 discusses the problems of testing portfolio conditional mean-variance efficiency and testing a semiparametric single index model. We examine the finite sample performances of the nonparametric significance test and the test for a partially linear model in Section 6. The proofs of the main results are given in the Appendices A and B. Appendix C contains some technical lemmas that are used in the proofs of Appendices A and B. Throughout the rest of this paper, all the limits are taken as $n \to \infty$. $\sum_t = \sum_{t=1}^n$, $\sum_t\sum_{s\neq t} = \sum_t\sum_{s\neq t,s=1}^n$, $\sum\sum\sum_{s\neq t\neq l} = \sum_t\sum_{s\neq t}\sum_{l\neq t,l\neq s,l=1}^n$, etc.

2. A general framework of kernel-based test

In this paper we consider the general hypothesis testing problem of the form $\mathrm{E}(U|X) = 0$ almost everywhere (a.e.) for some suitably chosen random variables (vectors) U and X. There are many examples that the null hypothesis can be written as $\mathrm{E}(U|X) = 0$ a.e., for example, in the context of testing a parametric regression model, say linearity, $Y = X'\gamma + U$. The null hypothesis of $\mathrm{E}(Y|X) = X'\gamma$ a.e. is equivalent to $\mathrm{E}(U|X) = 0$ a.e. For a kernel based test for parametric functional form, see Zheng (1996). Other examples that the null hypothesis can be written as $\mathrm{E}(U|X) = 0$ a.e. including testing for omitted variables (Ait-Sahalia et al., 1994; Fan and Li, 1996a; Lavergne and Vuong, 1996b; Lewbel, 1993) and testing semiparametric partially linear models (Ait-Sahalia et al., 1994; Yatchew, 1992; Whang and Andrews, 1993; Fan and Li, 1996a), testing a semiparametric index model (Fan and Li, 1996a), testing a rational expectation model (Robinson, 1989), testing conditional symmetry Zheng, 1998a), testing conditional parametric distribution (Andrews, 1997; Zheng, 1998b), and testing portfolio conditional mean-variance efficiency (Wang, 1997; Chen and Fan, 1997).

Even when the null hypothesis can be written as $\mathrm{H}_0 : \mathrm{E}(U|X) = 0$ a.e., consistent testing H_0 can still be done using different distance measures. When the kernel estimation method is used, the distance measure $I = \mathrm{E}[U\mathrm{E}(U|X)f(X)]$ turns out to be a convenient choice. First note that $I = \mathrm{E}\{[\mathrm{E}(U|X)]^2 f(X)\} \geqslant 0$ and the equality holds if and only if H_0 is true. Hence I serves as a proper candidate for consistent testing H_0. For the advantages of using distance measure I to construct kernel-based consistent tests, see Li and Wang (1998) and Hsiao and Li (1997). A kernel-based sample analogue of I is $I_n = (n(n-1)h^d)^{-1}\sum_t\sum_{s\neq t} U_t U_s K((X_t - X_s)/h)$, where $K(\cdot)$ is the kernel function and h is the smoothing parameter.

Often U_t is not observable but can be consistently estimated. When the null models are parametric models, usually U_t can be $\sqrt{n}$-consistently estimated under quite general conditions, e.g., Fan and Li (1996b) and Hsiao and Li (1997). In this paper we will consider the case that the null model contains some nonparametric components, say the null model is a nonparametric or a semiparametric regression model and we will use kernel methods to estimate these unknown regression functions. In such cases, we can only consistently estimate U_t with the nonparametric (kernel estimation) rate which is slower than the parametric rate of $\sqrt{n}$. Thus, the derivation of the asymptotic distributions of the test statistics in this paper is much more complex than the case of testing a parametric null model (e.g., Fan and Li, 1996b; Hsiao and Li, 1997). One leading example we consider in this paper is that the null model is $Y = \mathrm{E}(Y|W) + U$, where W is a proper subset of X. Then the null hypothesis of $\mathrm{E}(U|X) = 0$ a.e. is a nonparametric omitted variable test because under this null hypothesis, $\mathrm{E}(Y|W) = \mathrm{E}(Y|X)$ a.e., the extra regressors in X (but not in W) do not help to explain Y.

Let W (W can be different from X) be the variable that enter the null model nonparametrically and denotes $\hat{f}_{w_t}$ the kernel estimator of $f_w(W_t)$, where $f_w(W_t)$ is the density function of W_t. Also let $\tilde{U}_t$ be a consistent estimator of U_t (under H_0). Then $\hat{f}_{w_t}$ will appear in the denominator of $\tilde{U}_t$, the so-called random denominator problem associated with kernel estimation. In order to avoid the random denominator problem, we choose to use a density weighted version of I (or I_n) as the basis of our test statistic: $J = \mathrm{E}\{Uf_w(W)\mathrm{E}[Uf_w(W)|X]f(X)\} \equiv \mathrm{E}\{\varepsilon \mathrm{E}[\varepsilon|X]f(X)\}$, where $\varepsilon = Uf_w(W)$ and $f(X_t)$ is the density function of X_t (e.g., Fan and Li, 1996a; Lavergne and Vuong, 1996b). A kernel-based sample analogue of J is[2]

$$J_n = \frac{1}{n(n-1)h^d}\sum_t \sum_{s\neq t} U_t f_{w_t} U_s f_{w_s} K_{ts} \equiv \frac{1}{n(n-1)h^d}\sum_t \sum_{s\neq t} \varepsilon_t \varepsilon_s K_{ts}, \tag{1}$$

[2] J_n can be viewed as a conditional moment test with the weight function given by $\mathrm{E}(\varepsilon_t|X_t)f(X_t)$, see Newey (1985) and Tauchen (1985).

where $\varepsilon_t = U_t f_{w_t}$, $f_{w_t} = f_w(W_t)$, $K_{ts} = K((X_t - X_s)/h)$ is the kernel function and $h = h_n$ is the smoothing parameter.

A feasible test statistic is obtained by replacing $\varepsilon_t = U_t f_{w_t}$ by its kernel estimator $\tilde{U}_t \hat{f}_{w_t}$:

$$\hat{J}_n = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \tilde{U}_t \hat{f}_{w_t} \tilde{U}_s \hat{f}_{w_s} K_{ts}, \tag{2}$$

where $\tilde{U}_t$ and $\hat{f}_{w_t}$ are the kernel estimators of U_t and f_{w_t} respectively, their specific definitions (depending on the specific null models) will be given later.

The test statistics (with independent data) considered by Fan and Li (1996a) and Lavergne and Vuong (1996b) all have the form of Eq. (2) with Eq. (1) as the leading term. J_n given in Eq. (1) is a second order degenerate U-statistic. Hall (1984) and De Jong (1987) established the asymptotic normal distribution for a general second order degenerate U-statistics with independent observations. Recently, Fan and Li (1996b) generalize Hall's (1984) result to the weakly dependent data case. Therefore, one can use the result of Fan and Li (1996b) to derive the asymptotic distribution of $\hat{J}_n$ provided one can show that $\hat{J}_n - J_n$ has an order smaller than J_n. We re-state a result from Fan and Li (1996b) in a lemma below for ease of reference.

Lemma 2.1. Let $\mathscr{X}_t = (\varepsilon_t, X'_t)'$ be a strictly stationary process that satisfies the condition (D1) of Appendix A, $\varepsilon_t \in R$ and $X_t \in R^d$, $K(\cdot)$ be the kernel function with h being the smoothing parameter that satisfy the condition (D2) of Appendix A. Define $\sigma^2_\varepsilon(x) = \mathrm{E}[\varepsilon^2_t | X_t = x]$ and $J_n = (n(n-1)h^d)^{-1} \sum_t \sum_{s \neq t} \varepsilon_t \varepsilon_s K((X_t - X_s)/h)$. Then $nh^{d/2} J_n \to \mathrm{N}(0, \sigma^2_0)$ in distribution, where $\sigma^2_0 = 2\mathrm{E}[\sigma^4_\varepsilon(X_t) f(X_t)][\int K^2(u)\,\mathrm{d}u]$ and $f(\cdot)$ is the marginal density function of X_t.

Proof. Lemma 2.1 is a special case of Theorem 2.1 of Fan and Li (1996b).

Note that with $\varepsilon_t = U_t f_{w_t}$, Lemma 2.1 gives the asymptotic normal distribution of J_n defined in Eq. (1). In the remaining part of the paper we will apply the above testing procedure to derive the asymptotic distributions of a nonparametric significance test and of a test for partially linear model (with weakly dependent data). We will also discuss some other hypotheses testing problems that fit the above framework.

3. A nonparametric significance test

In this section we apply the general hypothesis testing procedure of Section 2 to construct a nonparametric significance test. We consider the following nonparametric regression model:

$$Y_t = \mathrm{E}(Y_t|X_t) + U_t, \tag{3}$$

where Y_t is a scalar and $X_t \in R^d$.

Let $X_t = (W_t', Z_t')'$, where W_t is a $q \times 1$ vector ($1 \leqslant q \leqslant d-1$) and Z_t is of dimension $(d-q) \times 1$. Then the null hypothesis that a subset of regressors, Z_t (say), is insignificant for the regression model (3) if $\mathrm{E}(Y_t|X_t) = \mathrm{E}(Y_t|W_t)$ a.e. Let $r(w) = \mathrm{E}(Y_t|W_t = w)$, $U_t = Y_t - r(W_t)$ and $\varepsilon_t = U_t f_w(W_t)$. Then the null hypothesis can be written as

$$\mathrm{H}_0^a\colon \quad \mathrm{E}(\varepsilon|X) = 0 \text{ a.e.}$$

The alternative hypothesis is

$$\mathrm{H}_1^a\colon \quad \mathrm{E}(\varepsilon|X) \neq 0 \text{ on a set with positive measure.}$$

Following the approach of Section 2, we construct our test statistic based on sample analogue of $J = \mathrm{E}\{\varepsilon \mathrm{E}[\varepsilon|X] f(X)\}$. The sample analogue of J is J_n as given in Eq. (1). To obtain a feasible test statistic, note that $U_t = Y_t - \mathrm{E}(Y_t|W_t)$ under H_0^a, we estimate $\varepsilon_t = U_t f_{w_t}$ by $\tilde{U}_t \hat{f}_{w_t} \equiv (Y_t - \hat{Y}_t)\hat{f}_{w_t}$, where

$$\hat{Y}_t = \frac{1}{(n-1)a^q} \sum_{s \neq t} Y_s L_{ts} / \hat{f}_{w_t} \tag{4}$$

is the kernel estimator of $\mathrm{E}(Y_t|W_t)$, $L_{ts} = L((W_t - W_s)/a)$ is the kernel function and a is the smoothing parameter, and

$$\hat{f}_{w_t} = \frac{1}{(n-1)a^q} \sum_{s \neq t} L_{ts} \tag{5}$$

is the kernel estimator of $f_w(W_t)$. Hence, replacing ε_t in J_n of Eq. (2) by $(Y_t - \hat{Y}_t)\hat{f}_{w_t}$, we obtain a feasible test statistic for H_0^a:

$$\hat{J}_n^a = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} (Y_t - \hat{Y}_t)\hat{f}_{w_t}(Y_s - \hat{Y}_s)\hat{f}_{w_s} K_{ts}. \tag{6}$$

To derive the asymptotic distribution of $\hat{J}_n^a$, the following assumptions will be used, where we also use the definitions of Robinson (1988) for the class of kernel functions $\mathscr{K}_\lambda$ and the class of functions $\mathscr{G}_\mu^\alpha$, see Appendix A for details.

(A1) (i) $\{Y_t, X_t'\}_{t=1}^n$ is a strictly stationary absolutely regular process with the mixing coefficient β_τ that satisfies $\beta_\tau = O(\rho^\tau)$ for some $0 < \rho < 1$. (ii) $f_w(\cdot) \in \mathscr{G}_v^\infty$, $r(\cdot) \in \mathscr{G}_v^{4+\eta}$, and $f \in \mathscr{G}_v^\infty$ for some integer $v \geqslant 2$, also f is bounded. (iii) the error $U_t = Y_t - r(W_t)$ is a martingale difference process, $E[|U_t^{4+\eta}|] < \infty$ and $E[|U_{t_1}^{i_1} U_{t_2}^{i_2}, \dots, U_{t_l}^{i_l}|^{1+\xi}] < \infty$ for some arbitrarily small $\eta > 0$ and $\xi > 0$, where $2 \leqslant l \leqslant 4$ is an integer, $0 \leqslant i_j \leqslant 4$ and $\sum_{j=1}^l i_j \leqslant 8$, $\sigma_u^2(x) = E(U_t^2 | X_t = x)$, $\mu_4(x) = E(U_t^4 | X_t = x)$, f_w, f and r all satisfy some Lipschitz conditions: $|m(u+v) - m(u)| \leqslant D(u)\|v\|$, $D(\cdot)$ has finite $(2+\eta')$th moment for some small $\eta' > 0$, where $m(\cdot) = \sigma_u^2(\cdot)$, $\mu_4(\cdot)$, $f_w(\cdot)$, $f(\cdot)$ or $r(\cdot)$. (iv) Let $f_{\tau_1,\dots,\tau_l}(.,\dots,.)$ be the joint probability density function of $(X_1, X_{1+\tau_1}, \dots, X_{1+\tau_l})$ $(1 \leqslant l \leqslant 3)$. Then $f_{\tau_1,\dots,\tau_l}(.,\dots,.)$ is bounded and satisfies a Lipschitz condition: $|f_{\tau_1,\dots,\tau_l}(x_1+u_1, x_2+u_2, \dots, x_l+u_l) - f_{\tau_1,\dots,\tau_l}(x_1, x_2, \dots, x_l)| \leqslant D_{\tau_1,\dots,\tau_l}(x_1, x_2, \dots, x_l)\|u\|$, where $D_{\tau_1,\dots,\tau_l}(.,\dots,.)$ is integrable and satisfies the condition that $\int D_{\tau_1,\dots,\tau_l}(x,\dots,x)\|x\|^{2\xi} < M < \infty$, $\int D_{\tau_1,\dots,\tau_l}(x_1,\dots,x_l) f_{\tau_1,\dots,\tau_l}(x_1,\dots,x_l)\,dx < M < \infty$ for some $\xi > 1$.

(A2) (i) we use product kernel for both $L(\cdot)$ and $K(\cdot)$, let l and k be their corresponding univariate kernel, then $l(\cdot) \in \mathscr{K}_v$, $k(\cdot)$ is non-negative and $k(\cdot) \in \mathscr{K}_2$. (ii) $a \to 0$, $h = O(n^{-\bar{\alpha}})$ for some $0 < \bar{\alpha} < (7/8)d$. (iii) $h^d/a^{2q} \to 0$, $nh^{d/2}a^{2v} \to 0$ (all the limits are taken as $n \to \infty$).

3.1. Some remarks on the regularity conditions

Condition (A1)(i) requires that $\{Y_t, X_t'\}$ to be a stationary absolutely regular process with geometric decay rate. (A1)(ii)–(iv) are mainly some smoothness and moments conditions, these conditions are quite weak in the sense that they are similar to the ones used in Fan and Li (1996a) for independent data case. However, for ARCH or GARCH type error processes as considered in Engle (1982) and Bollerslev (1986), the error term U_t may not have finite fourth moments under some situations. For example, let $U_t | U_{t-1} \sim N(0, \alpha_0 + \alpha_1 U_{t-1}^2)$, Engle (1982) showed that U_t does not have a finite fourth moment if $\alpha_1 > 1/\sqrt{3}$. Thus, assumption (A1)(iii) will be violated in such a case.

(A2) (i) requires $L(\cdot)$ to be a vth ($v \geqslant 2$) order kernel, this condition together with (A1)(ii) ensures that the bias in the kernel estimation (of the null model) is $O(a^v)$. The requirement that k is a non-negative second order kernel function in (A2)(ii) is a quite weak and standard assumption. We emphasize here that the assumption k is *non-negative* plays an important role in simplifying our proofs of the main results. See the discussions below Theorem 3.1 for more details on this. (A2)(ii) and (iii) imply $h \to 0$, $nh^d \to \infty$ and $na^{2q} \to \infty$, (A2)(iii) comes from the fact that the mean square error in the kernel estimation of the null model is of smaller order than $(nh^{d/2})^{-1}$, i.e., $nh^{d/2}(a^{2v} + (na^q)^{-1}) = o(1)$. Our regular conditions are quite weak in the sense that they are very similar to the ones used for independent data case (e.g., Fan and Li, 1996a).

Theorem 3.1. Assume the conditions (A1) and (A2) hold. Then

(i) *Under* H_0^a, $T_n^a \overset{\text{def}}{=} nh^{d/2}\hat{J}_n^a/\hat{\sigma}_a \to N(0,1)$ *in distribution, where* $\hat{\sigma}_a^2 = (2/n(n-1)h^d) \sum_t \sum_{s \neq t} \tilde{U}_t^2 \hat{f}_{w_t}^2 \tilde{U}_s^2 \hat{f}_{w_s}^2 K_{ts}^2$ *is a consistent estimator of* $\sigma_a^2 = 2E[\sigma_u^4(X_t) f_w^2(W_t) f(X_t)] [\int K^2(z)\,dz]$.
(ii) *Under* H_1^a, $\text{Prob}[T_n^a > B_n] \to 1$ *for any non-stochastic sequence* $B_n = o(nh^{d/2})$.

A detailed proof of Theorem 3.1 is given in the Appendix A. The proof involve steps showing that (I) $\hat{J}_n^a - J_n = o_p((nh^{d/2})^{-1})$, and (II) $\hat{\sigma}_a^2 = \sigma_a^2 + o_p(1)$. (I) and (II) imply that $nh^{d/2}\hat{J}_n^a/\hat{\sigma}_a = nh^{d/2}J_n/\sigma_a + o_p(1)$. Hence Theorem 3.1(i) follows from Lemma 2.1.

Here we would also like to comment on some simple but important tricks that are used to greatly simplify the proof of Theorem 3.1. As can be seen from the proof of Appendix A, we need to obtain probability bounds for terms that involve four summations.[3] The second moments of these terms have eight summations and it is extremely tedious to get sharp bounds involving eight summations with mixing data. In the proof we try to bound terms by non-negative terms with a simple structure (i.e., with less summations). Then we only need to compute the expectations of these non-negative terms. For example in the proof of Lemma A.1, using the fact that k is nonnegative together with some other simple tricks, we show that the leading term of J_{n1} (see Appendix A for definition of J_{n1}) is nonnegative and contains only three summations, while in Fan and Li (1996a) they have to compute the second moment of J_{n1}, which involves eight summations (because they did not assume k is nonnegative), in order to evaluate the probability order of J_{n1}. In contrast we only need to compute the expectations of terms with three or four summations (rather than dealing with eight summations). For details see Appendix A.

Theorem 3.1 generalizes the results of Fan and Li (1996a) and Lavergne and Vuong (1996b) of independent observations to the weakly dependent data case. It should be mentioned that Lavergne and Vuong (1996b) also studied the local power property of their proposed test and showed that their test can detect local alternatives that approach the null model at a rate of $O_p(n^{-1/2}h^{-d/4})$. Although I do not study the local power property of the $\hat{J}_n^a$ test in this paper, I conjecture that the local power property of the $\hat{J}_n^a$ test is similar to the independent data case as considered by Lavergne and Vuong (1996b), i.e., the $\hat{J}_n^a$ test can detect local alternatives that approach the null model at a rate of $O_p(n^{-1/2}h^{-d/4})$.

Our Monte Carlo simulations show that the $\hat{J}_n^a$ test has substantial finite sample bias which causes the $\hat{J}_n^a$ test undersized. For the independent data case, Lavergne and Vuong (1996b) suggested a modified test which has smaller finite

[3] Although Denker and Keller (1983) provide bounds for finite order U-statistics with mixing data, their results are not sharp enough to deliver the results we need.

sample bias than the $\hat{J}_n^a$ test. To motivate this new test of Lavergne and Vuong (1996b), we substitute Eqs. (4) and (5) into Eq. (6) to get

$$\hat{J}_n^a = \frac{1}{n(n-1)^3 h^d a^{2q}} \sum_t \sum_{s \neq t} \sum_{l \neq t} \sum_{k \neq s} (Y_t - Y_l)(Y_s - Y_k) L_{tl} L_{sk} K_{ts}. \tag{7}$$

The terms of $l = s$, $l = t$ and $k = t$ may cause finite sample bias for the $\hat{J}_n^a$ test (since these terms contain squares of the error terms). Subtracting these terms from $\hat{J}_n^a$ and replacing $n(n-1)^3$ by $n^{(4)} = n(n-1)(n-2)(n-3)$ lead to a new test (denotes it by V_n^a) with possibly smaller finite sample bias:

$$V_n^a = \frac{1}{n^{(4)}}\{n(n-1)^3 \hat{J}_n^a - n^{(3)} V_{1n}^a - 2n^{(3)} V_{2n}^a\}, \tag{8}$$

where $n^{(3)} \equiv n(n-1)(n-2)$, also

$$V_{1n}^a = \frac{1}{n^{(3)} a^{2q} h^d} \sum\sum_{t \neq s \neq l}\sum (Y_t - Y_l)(Y_s - Y_l) L_{tl} L_{sl} K_{ts} \tag{9}$$

and

$$V_{2n}^a = \frac{1}{n^{(3)} a^{2q} h^d} \sum\sum_{t \neq s \neq l}\sum (Y_t - Y_s)(Y_s - Y_l) L_{ts} L_{sl} K_{ts}. \tag{10}$$

The next corollary shows the V_n^a test has the same asymptotic distribution as the $\hat{J}_n^a$ test.

Corollary 3.2. Under the same conditions as in Theorem 3.1, we have

(i) *Under* H_0^a, $nh^{d/2} V_n^a / \hat{\sigma}_a \to N(0,1)$ *in distribution, where* $\hat{\sigma}_a^2$ *is the same as defined in Theorem 3.1.*
(ii) *Under* H_1^a, $\text{Prob}[nh^{d/2} V_n^a / \hat{\sigma}_a > B_n] \to 1$ *for any non-stochastic sequence* $B_n = o(nh^{d/2})$.

The proof of Corollary 3.2 is given in Appendix A.

Corollary 3.2 shows that the V_n^a test with weakly dependent data has the same asymptotic distribution as the independent data case considered by Lavergne and Vuong (1996b). However, our assumption that $h^d/a^{2q} \to 0$ as $n \to \infty$ (see (A3)(iii)) is stronger than the condition of $h^d/a^q \to 0$ (as $n \to \infty$) used by Lavergne and Vuong (1996b). It might be possible to relax the condition (A2)(iii) to the same condition as used in Lavergne and Vuong (1996b) in establishing Corollary 3.2. But I am unable to provide a short proof for this conjecture.

4. A test for a partially linear model

In this section we show that the result of the nonparametric significance test of Theorem 3.1 can be used to easily derive the asymptotic distribution for testing a partially linear model.

Using the same notation as introduced in Section 3 (i.e., $X=(W',Z')'$), the null hypothesis of a partially linear regression model is (e.g., Engle et al., 1986; Robinson, 1988; Stock, 1989)

$$\mathrm{H}_0^b\!: \quad \mathrm{E}(Y|X) = Z'\gamma_0 + \theta(W) \text{ a.e.}$$

for some $\gamma_0 \in R^{d-q}$ and some smooth function $\theta(\cdot): R^q \to R$.

Given that the null model of a partially linear model also contains a nonparametric component $\theta(W)$, we present the null hypothesis in the following density-weighted form. Define $U = Y - Z'\gamma_0 - \theta(W)$ and $\varepsilon = Uf_w(W)$. Then the null hypothesis can also be written as

$$\mathrm{H}_0^b\!: \quad \mathrm{E}(\varepsilon|X) = 0 \text{ a.e.}$$

The alternative hypothesis is

$$\mathrm{H}_1^b\!: \quad \mathrm{E}(\varepsilon|X) \neq 0 \text{ on a set with positive measure.}$$

As in Section 3, we obtain our test statistic by replacing $\varepsilon_t = U_t f_{w_t}$ in J_n of Eq. (1) by some estimate of it. We use a two-step method as in Robinson (1988) and Fan and Li (1996c) to estimate $U_t f_{w_t}$. First we estimate γ_0 by[4]

$$\hat{\gamma} = S_{(Z-\hat{Z})\hat{f}_w} S_{(Z-\hat{Z})\hat{f}_w,(Y-\hat{Y})\hat{f}_w}, \tag{11}$$

where $S_{A\hat{f}_w,B\hat{f}_w} = n^{-1}\sum_t A_t\hat{f}_{w_t}B_t'\hat{f}_{w_t}$ and $S_{A\hat{f}_w} = S_{A\hat{f}_w,A\hat{f}_w}$. Note that $U_t = Y_t - \mathrm{E}(Y_t|W_t) - (Z_t - \mathrm{E}(Z_t|W_t))'\gamma_0$, therefore we estimate U_t by $\tilde{U}_t = (Y_t - \hat{Y}_t) - (Z_t - \hat{Z}_t)'\hat{\gamma}$, where $\hat{Y}_t$ is given in Eq. (4), $\hat{f}_{w_t}$ is given in Eq. (5) and $\hat{Z}_t = (1/(n-1)a^q)\sum_{s\neq t} Z_s L_{ts}/\hat{f}_{w_t}$ ($\hat{z}_t$ is the kernel estimator of $\mathrm{E}(z_t|w_t)$). The density

[4] As correctly pointed out by a referee, one can use any $\sqrt{n}$-consistent estimator of γ, not necessarily the one given in Eq. (11), the proof of Theorem 4.1 below remains unchanged. Here we choose Eq. (11) because the regularity conditions that ensure $\hat{\gamma} - \gamma_0 = \mathrm{O_p}(n^{-1/2})$ are quite weak, see Theorem 2.1 of Fan and Li (1996c).

weighted error $\varepsilon_t = U_t f_{w_t}$ is estimated by $\tilde{U}_t \hat{f}_{w_t}$. Hence, our test statistic for testing H_0^b is

$$\hat{J}_n^b = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \tilde{U}_t \hat{f}_{w_t} \tilde{U}_s \hat{f}_{w_s} K_{ts}. \tag{12}$$

The asymptotic distribution of $\hat{J}_n^b$ is given in the next theorem.

Theorem 4.1. Let (B1) and (B2) be the same as (A1) and (A2) except that $r(\cdot)$ in (A2) is replaced by $\theta(\cdot)$. Also define $\xi(w) = \mathrm{E}(Z_t | W_t = w)$. Then under conditions (B1) and (B2), and the assumption that $\xi(\cdot) \in \mathcal{G}_v^{4+\delta}$, the following results are true:

(i) *Under* H_0^b, $T_n^b \stackrel{\text{def}}{=} nh^{d/2} I_n^b / \hat{\sigma}_b \to \mathrm{N}(0,1)$ *in distribution, where* $\hat{\sigma}_b^2 = (2/n(n-1)h^d) \sum_t \sum_{s \neq t} \tilde{U}_t^2 \hat{f}_{w_t}^2 \tilde{U}_s^2 \hat{f}_{w_s}^2 K_{ts}^2$ with $\tilde{U}_t = Y_t - \hat{Y}_t - (Z_t - \hat{Z}_t)'\hat{\gamma}$.

(ii) *Under* H_1^b, $\mathrm{Prob}[T_n^b > B_n] \to 1$ *for any non-stochastic sequence* $B_n = o(nh^{d/2})$.

Proof. We only prove (i) here since the proof of (ii) is much easier than that of (i). First,

$$\begin{aligned}
\tilde{U}_t \hat{f}_{w_t} &= [(Y_t - \hat{Y}_t) - (Z_t - \hat{Z}_t)'\gamma_0] \hat{f}_{w_t} - (Z_t - \hat{Z}_t)'(\hat{\gamma} - \gamma_0) \hat{f}_{w_t} \\
&= [(\theta_t - \hat{\theta}_t) + U_t - \hat{U}_t] \hat{f}_{w_t} - (Z_t - \hat{Z}_t)'(\hat{\gamma} - \gamma_0) \hat{f}_{w_t} \\
&\equiv \bar{U}_t \hat{f}_{w_t} - (Z_t - \hat{Z}_t)'(\hat{\gamma} - \gamma_0) \hat{f}_{w_t},
\end{aligned} \tag{13}$$

where $\bar{U}_t = (\theta_t - \hat{\theta}_t) + U_t - \hat{U}_t$, $\theta_t = \theta(W_t)$, $\hat{\theta}_t = (1/(n-1)a^q) \sum_{s \neq t} \theta_s L_{ts} / \hat{f}_{w_t}$ and $\hat{U}_t = (1/(n-1)a^q) \sum_{s \neq t} U_s L_{ts} / \hat{f}_{w_t}$. Substituting Eq. (13) into Eq. (12), we get

$$\begin{aligned}
\hat{J}_n^b &= \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \bar{U}_t \hat{f}_{w_t} \bar{U}_s \hat{f}_{w_s} K_{ts} \\
&\quad - \frac{2}{n(n-1)h^d} \sum_t \sum_{s \neq t} \bar{U}_t \hat{f}_{w_t} (Z_s - \hat{Z}_s)' \hat{f}_{w_s} K_{ts} (\hat{\gamma} - \gamma_0) \\
&\quad + (\hat{\gamma} - \gamma_0)' \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} (Z_t - \hat{Z}_t) \hat{f}_{w_t} (Z_s - \hat{Z}_s)' \hat{f}_{w_s} K_{ts} (\hat{\gamma} - \gamma_0) \\
&\equiv J_{1n} - 2(\hat{\gamma} - \gamma_0) J_{2n} + (\hat{\gamma} - \gamma_0)' J_{3n} (\hat{\gamma} - \gamma_0), \text{ say.}
\end{aligned} \tag{14}$$

Note that J_{1n} can be obtained by replacing r_t by θ_t in $\hat{J}_n^a$ (see Eq. (A.1) of Appendix A). Hence by the result of Theorem 3.1, we know that $nh^{d/2} J_{1n} / \hat{\sigma}_b \to \mathrm{N}(0,1)$ in distribution.

It remains to show the last two terms on the right-hand side of Eq. (14) are both $o_p((nh^{d/2})^{-1})$. Recall that $\xi_t = E(Z_t|W_t)$ and define $\eta_t = Z_t - \xi_t$, we have

$$J_{2n} = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \bar{U}_t \hat{f}_{w_t}(Z_s - \hat{Z}_s)\hat{f}_{w_s} K_{ts}$$

$$= \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} [(\theta_t - \hat{\theta}_t) + U_t - \hat{U}_t]$$

$$\times \hat{f}_{w_t}[(\xi_s - \hat{\xi}_s) + \eta_s - \hat{\eta}_s]\hat{f}_{w_s} K_{ts}.$$

Comparing the above expression of J_{2n} with J_{1n}, one can easily see that all the terms in J_{2n} is $o_p((nh^{d/2})^{-1})$ except $J_{2n,1} \equiv (1/n(n-1)h^d)\sum_t\sum_{s\neq t} U_t \eta_s K_{ts}$. $J_{2n,1} = o_p((nh^d)^{-1/2})$ by Lemma A.1 of Hsiao and Li (1997). Hence, $(\hat{\gamma} - \gamma_0)J_{2n} = O_p(n^{-1/2})o_p((nh^d)^{-1/2}) = o_p((nh^{d/2})^{-1})$ because $\hat{\gamma} - \gamma_0 = O_p(n^{-1/2})$ by Theorem 2.1 of Fan and Li (1996c).

Finally, $J_{3n} = (1/n(n-1)h^d)\sum_t\sum_{s\neq t}[(\xi_t - \hat{\xi}_t) + \eta_t - \hat{\eta}_t]'\hat{f}_{w_t}[(\xi_s - \hat{\xi}_s) + \eta_s - \hat{\eta}_s]\hat{f}_{w_s} K_{ts}$. Compare J_{3n} with J_{1n}, one can easily see that $J_{3n} = (1/n(n-1)h^d) \sum_t\sum_{s\neq t}\eta_t\eta_s K_{ts} + o_p((nh^{d/2})^{-1}) = O_p(1)$. Hence, $(\hat{\gamma} - \gamma_0)' J_{3n}(\hat{\gamma} - \gamma_0) = O_p(n^{-1/2})O_p(1)O_p(n^{-1/2}) = o_p((nh^{d/2})^{-1})$. This finishes the proof of Theorem 4.1(i). □

Similar to the V_n^a test of Section 3. One can also define an asymptotically equivalent (and possibly has less bias in finite samples) test V_n^b as follows.

First note that $\hat{J}_n^b$ can be written as

$$\hat{J}_n^b = \frac{1}{n(n-1)^3 h^d a^{2q}} \sum_t \sum_{s\neq t} \sum_{l \neq t} \sum_{k \neq s} [(Y_t - Y_l) - (Z_t - Z_l)\hat{\gamma}]$$

$$\times [(Y_s - Y_k) - (Z_s - Z_k)\hat{\gamma}] L_{tl} L_{sk} K_{ts}. \tag{15}$$

Removing the terms of $l = k$, $l = s$ and $k = t$ in Eq. (15) and also replacing $n(n-1)^3$ by $n^{(4)} = n(n-1)(n-2)(n-3)$ leads to

$$V_n^b = \frac{1}{n^{(4)}}\{n(n-1)^3 \hat{J}_n^b - n^{(3)} V_{1n}^b - 2n^{(3)} V_{2n}^b\}, \tag{16}$$

where

$$V_{1n}^{b}=\frac{1}{n^{(3)}}\sum\sum\sum_{t\neq s\neq l}[(Y_t-Y_l)-(Z_t-Z_l)\hat{\gamma}]$$

$$\times[(Y_s-Y_l)-(Z_s-Z_l)\hat{\gamma}]L_{tl}L_{sl}K_{ts} \tag{17}$$

with $n^{(3)}=n(n-1)(n-2)$ and

$$V_{2n}^{b}=\frac{1}{n^{(3)}}\sum\sum\sum_{t\neq s\neq l}[(Y_t-Y_s)-(Z_t-Z_s)\hat{\gamma}]$$

$$\times[(Y_s-Y_l)-(Z_s-Z_l)\hat{\gamma}]L_{ts}L_{sl}K_{ts}. \tag{18}$$

From Theorem 4.1, we immediately have the following corollary.

Corollary 4.2. Under the same conditions as in Theorem 4.1, we have

(i) *Under* H_0^b, $nh^{d/2}V_n^b/\hat{\sigma}_b\to N(0,1)$ *in distribution, where* $\hat{\sigma}_b^2$ *is the same as defined in Theorem* 4.1.
(ii) *Under* H_1^b, $\mathrm{Prob}[nh^{d/2}V_n^b/\hat{\sigma}_b>B_n]\to 1$ *for any non-stochastic sequence* $B_n=o(nh^{d/2})$.

The proof of Corollary 4.2 is similar to the proof of Corollary 3.2 except that one needs to cite the result of Theorem 4.1 rather than the result of Theorem 3.1. Therefore the proof for Corollary 4.2 is omitted here.

5. Extensions: Some additional hypotheses testing problems

In this paper we propose a general framework for consistent testing time-series econometric models. We present a general methodology in Section 2 and apply it to construct a consistent test for omitted variables and a consistent test for partially linear model, both with weakly dependent observations. The test statistics are shown to have asymptotic normal distributions under their respective null hypotheses. Using the technical lemmas provided in this paper, one can easily derive asymptotic distributions of other consistent tests for time-series non-parametric or semiparametric econometric models. We give two more examples in this section to illustrate this point but due to space limitation, we will only provide a proof for the first example.

5.1. A test for portfolio conditional mean-variance efficiency

The first example is testing for portfolio conditional mean-variance efficiency as considered by Chen and Fan (1997), see also Gibbons and Ferson (1985), Gibbons et al. (1989), Cochrane (1996) and Wang (1997). Let $r_{m,t+1}$ be the return on the portfolio m in excess of the riskless rate, and r_{t+1} be a $p\times 1$ vector of excess returns of the other assets. The null hypothesis that the portfolio m is conditional mean-variance efficient if

$$\mathrm{E}[r_{t+1}|\mathscr{F}_t] = \mathrm{E}[r_{m,t+1}|\mathscr{F}_t]\,\frac{cov(r_{t+1}, r_{m,t+1}|\mathscr{F}_t)}{var(r_{m,t+1}|\mathscr{F}_t)}, \tag{19}$$

where $\mathscr{F}_t$ is the sigma field generated by all the state variables up to period t. Under Markovian assumptions on the processes r_t and $r_{m,t}$, Eq. (19) is equivalent to

$$\mathrm{E}[r_{t+1}|X_t] = \mathrm{E}[r_{m,t+1}|X_t]\frac{cov(r_{t+1}, r_{m,t+1}|X_t)}{var(r_{m,t+1}|X_t)}, \quad \text{a.e.} \tag{20}$$

for some $d\times 1$ vector X_t. Eq. (20) can also be written as

$$\mathrm{H}_0^c\colon \quad \mathrm{E}\{(\mathrm{E}[r_{m,t+1}^2|X_t] - \mathrm{E}[r_{m,t+1}|X_t]r_{m,t+1})r_{t+1}|X_t\} = 0 \quad \text{a.e.} \tag{21}$$

as considered by Chen and Fan (1997). If we define $U_t = (\mathrm{E}[r_{m,t+1}^2|X_t] - \mathrm{E}[r_{m,t+1}|X_t]\,r_{m,t+1})r_{t+1}$, then H_0^c is just $\mathrm{E}(U_t|X_t) = 0$ a.e. To avoid the random denominator problem in the kernel estimation, we can equivalently test: $\mathrm{E}(\varepsilon_t|X_t) = 0$ a.e. for $\varepsilon_t = U_t f(X_t)$. Note that this testing problem is slightly different from the earlier ones in that (i) U_t is a $p\times 1$ vector rather than a scalar, and (ii) $W_t = X_t$ rather than W_t is a proper subset of X_t. Nevertheless the testing procedure is still the same. Let $\tilde{U}_t$ be the kernel-based estimator of U_t, i.e., $\tilde{U}_t$ is obtained from U_t with $\mathrm{E}[r_{m,t+1}^2|X_t]$ and $\mathrm{E}[r_{m,t+1}|X_t]$ replaced by their corresponding kernel estimators $\hat{\mathrm{E}}[r_{m,t+1}^2|X_t]$ and $\hat{\mathrm{E}}[r_{m,t+1}|X_t]$, respectively, where

$$\hat{\mathrm{E}}[r_{m,t+1}^2|X_t] = \frac{1}{(n-1)a^d}\sum_{l\neq t} r_{m,l+1}^2 L((X_t - X_l)/a)/\tilde{f}_t \tag{22}$$

and

$$\hat{\mathrm{E}}[r_{m,t+1}|X_t] = \frac{1}{(n-1)a^d}\sum_{l\neq t} r_{m,l+1} L((X_t - X_l)/a)/\tilde{f}_t, \tag{23}$$

with $\tilde{f}_t$ being the kernel estimator of $f(X_t)$:

$$\tilde{f}_t = \frac{1}{(n-1)a^d} \sum_{l \neq t} L((X_t - X_l)/a). \tag{24}$$

Then a feasible test statistic for H_0^c is given by a $p \times 1$ vector $\hat{J}_n^c$,

$$\hat{J}_n^c = (\hat{J}_{n1}^c, \hat{J}_{n2}^c, \dots, \hat{J}_{np}^c)' \tag{25}$$

where the ith component of J_n^c is given by

$$\hat{J}_{ni}^c = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \tilde{U}_{it} \tilde{f}_t \tilde{U}_{is} \tilde{f}_s K_{ts}, \quad (i = 1, \dots, p), \tag{26}$$

$\tilde{U}_{it}$ is the ith component of $\tilde{U}_t$ and $K_{ts} = K((X_t - X_s)/h)$.

Define $g_{(1)}(x) = \mathrm{E}(r_{m,t+1}|X_t = x)$ and $g_{(2)}(x) = \mathrm{E}(r_{m,t+1}^2|X_t = x)$. The following assumptions will be used to derive the asymptotic distribution of $\hat{J}_n^c$.

Let (C1) be same as (A1) with the following changes: (i) $W_t = X_t$ rather than W_t is a subvector of X_t, (ii) $r(.) \in \mathcal{G}_v^{4+\eta}$ in (A1) is replaced by $g_{(1)}(\cdot) \in \mathcal{G}_v^{4+\eta}$ and $g_{(2)}(\cdot) \in \mathcal{G}_v^{4+\eta}$. Let (C2) be the same as (A2) except that the condition $h^d/a^{2q} \to 0$ is replaced by $h^d/a^{2d} \to 0$, or equivalently, $h/a^2 \to 0$.

Theorem 5.1. Under (C1) and (C2) as described above. The following results hold:

(i) *Under* H_0^c, $T_n^c \stackrel{\text{def}}{=} n^2 h^d (\hat{J}_n^c)' (\hat{\Omega}_c)^{-1} \hat{J}_n^c \to \chi^2(p)$ *in distribution, where* $\hat{\Omega}_c$ *is a* $p \times p$ *matrix with its* (i,j)*th element given by* $(\hat{\Omega}_c)_{ij} = (2/n(n-1)h^d) \sum_t \sum_{s \neq t} \tilde{U}_{it}^2 \tilde{f}_t^2 \tilde{U}_{js}^2 \tilde{f}_s^2 K_{ts}^2$.
Note that $\hat{\Omega}_c$ *is a consistent estimator of* Ω_c, *where the ijth element of* Ω_c *is* $(\Omega_c)_{ij} = 2\{\mathrm{E}[\sigma_{ij}^4(X_t) f^3(X_t)]\}\{\int K^2(z)\,\mathrm{d}z\}$ *with* $\sigma_{ij}^2(X_t) = \mathrm{E}(U_{it} U_{jt}|X_t)$.

(ii) *If* H_0^c *is false*, $\mathrm{Prob}[T_n^c > B_n] \to 1$ *for any non-stochastic sequence* $B_n = \mathrm{o}(nh^{d/2})$.

The proof of Theorem 5.1 is given in the Appendix B. Here we provide some intuitions as why one should expect that Theorem 5.1(i) is true.

Using similar arguments as in the proof of Theorem 3.1(i), one can show that $\hat{J}_{ni}^c = J_{ni}^c + \mathrm{o_p}((nh^{d/2})^{-1})$, where

$$J_{ni}^c = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} U_{it} f(X_t) U_{is} f(X_s) K_{ts}. \tag{27}$$

Lemma 2.1 implies that $nh^{d/2}J^c_{ni} \to N(0,(\Omega_c)_{ii})$ in distribution under H^c_0, where $(\Omega_c)_{ii}$ is the ith diagonal element of Ω_c as given in Theorem 5.1. Also it is straightforward to show that $cov(nh^{d/2}\hat{J}^c_{ni}, nh^{d/2}\hat{J}^c_{nj}) = (\Omega_c)_{ij} + o(1)$. Then by Cramer–Wold device, one can show that $nh^{d/2}\hat{J}^c_n \to N(0,\Omega_c)$. Theorem 5.1(i) follows because $\hat{\Omega}_c = \Omega_c + o_p(1)$.

As pointed out by a referee, the χ^2 statistic T^c_n given in Theorem 5.1 is a two-sided test, it will reject the null when each of the components of $\hat{J}^c_n$ take large enough negative values, which asymptotically can occur only under the null. Therefore the T^c_n test is less powerful than some properly constructed one-sided test. Gourieroux et al. (1982) provide a general approach on linear model specification testing with inequality constraints and showed that such tests usually have mixed (weighted) χ^2 distributions. The weights of the mixed χ^2 statistics are in general quite complex and some simulations methods may be needed to compute the weights numerically. The approach of Gourieroux et al. (1982) method should be useful in our context in constructing some more powerful one-sided tests against H^c_0. However, the asymptotic analysis of such one-sided tests will be quite complex since our null model contains nonparametric components. Therefore this issue is left for possible future research.

5.2. A test for a semiparametric single index model

In the second extension we consider the problem of testing a semiparametric single index model. The null hypothesis is

$$H^d_0 : E(Y|X) = g(X'\alpha_0) \quad \text{a.e.}$$

for some smooth but unknown function $g(\cdot)$, where α_0 is $d\times 1$ unknown parameter. Let $U_t = Y_t - g(X_t'\alpha_0)$ and $f_v(\cdot)$ be the density function of the univariate variable $V_t = X_t'\alpha_0$. Then H^d_0 is equivalent to $E(U_t f_v(V_t)|X_t) = 0$ a.e., a form of the conditional moment test discussed in Section 2. Denotes $\hat{f}_{v_t} = (na)^{-1}\sum_{i\neq t} L((X_t - X_i)'\hat{\alpha}/a)$, the kernel estimator of $f_v(V_t)$, where $\hat{\alpha}$ is a $\sqrt{n}$-consistent estimator of α_0 under H^c_0 (e.g., Powell et al., 1989). One can estimate U_t by $\tilde{U}_t = Y_t - \hat{E}(Y_t|X_t'\hat{\alpha})$, where $\hat{E}(Y_t|X_t'\hat{\alpha}) = (na)^{-1}\sum_{i\neq t} Y_i L((X_t - X_i)'\hat{\alpha}/a)/\hat{f}_{v_t}$ is the kernel estimator of $E(Y_t|X_t'\alpha_0)$. Then a feasible test statistic for H^d_0 is given by

$$\hat{J}^d_n = \frac{1}{n(n-1)h^d}\sum_t \sum_{s\neq t} \tilde{U}_t \hat{f}_{v_t} \tilde{U}_s \hat{f}_{v_s} K_{ts},$$

where $K_{ts} = K((X_t - X_s)/h)$. Under some regularity conditions similar to those as given in Power et al. (1989), and the conditions of Theorem 4.2 of

Fan and Li (1996a), one should be able to show that the leading term of $\hat{J}_n^d$ is J_n^d given by

$$J_n^d = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} U_t f_{v_t} U_s f_{v_s} K_{ts}, \tag{28}$$

where $f_{v_t} = f_v(V_t)$. Using Lemma 2.1, we immediately have $nh^{d/2} J_n^d / \hat{\sigma}_d \to \mathrm{N}(0,1)$ in distribution under H_0^d, where $\hat{\sigma}_d^2 = (2/n(n-1)h^d) \sum_t \sum_{s \neq t} \tilde{U}_t^2 \hat{f}_{v_t}^2 \tilde{U}_s^2 \hat{f}_{v_s}^2 K_{ts}^2$ is a consistent estimator of $\sigma_d^2 = 2\{\mathrm{E}[\sigma_u^4(X_t) f_v^2(V_t) f(X_t)]\}\{\int K^2(z)\,\mathrm{d}z\}$. This results in $nh^{d/2} \hat{J}_n^d / \hat{\sigma}_d \to \mathrm{N}(0,1)$ in distribution under H_0^d, provided one can show that $\hat{J}_n^d - J_n^d = o_p((nh^{d/2})^{-1})$. While I conjecture that $\hat{J}_n^d - J_n^d = o_p((nh^{d/2})^{-1})$ under some regularity conditions, I am unable to provide a simple (short) proof for this result.

6. Monte Carlo results

In this section we report some Monte Carlo simulation results to examine the finite sample performances of the nonparametric significance tests of $\hat{J}_n^a$ and V_n^a, and the $\hat{J}_n^b$ and V_n^b tests for a partially linear specification.

6.1. The case of the non-parametric significant test

To study the size and power properties of the $\hat{J}_n^a$ and V_n^a tests, we use the following data generating processes (DGP):

$$\text{DGP1: } Y_t = W_t + 0.5W_t^2 + U_t,$$

$$\text{DGP2: } Y_t = W_t + 0.5W_t^2 + \alpha_1 Z_t + \alpha_2 Z_t^2 + U_t,$$

$$\text{DGP3: } Y_t = 0.5Y_{t-1} + U_t,$$

$$\text{DGP4: } Y_t = 0.5Y_{t-1} + \alpha_3 Z_t + \alpha_4 Z_t^2 + U_t,$$

where $W_t = 0.5W_{t-1} + V_t$, $Z_t = 0.5Z_{t-1} + \eta_t$, U_t, V_t and η_t are independent processes and all of them are i.i.d. $\mathrm{N}(0,1)$. DGP1 is the null model (H_0^a) with $\mathrm{E}(Y_t|W_t) = \mathrm{E}(Y_t|W_t, Z_t)$. DGP2 is an alternative model (H_1^a). We consider two different cases for DGP2: case (i), $(\alpha_1, \alpha_2) = (0.5\sqrt{32/12}, 0)$; and case (ii), $(\alpha_1, \alpha_2) = (0, 0.5)$. Under the above choices of α_1 and α_2, we have $var(\alpha_1 Z_t) = var(\alpha_2 Z_t^2)$, so that case (i) and case (ii) have the similar deviations from the null model. Similarly, DGP3 is a null model with

$E(Y_t|Y_{t-1}) = E(Y_t|Y_{t-1}, Z_t)$, and DGP4 is an alternative model and we also consider two different cases for DGP4: case (i), $(\alpha_3, \alpha_4) = (0.5\sqrt{4/3}, 0)$; and case (ii), $(\alpha_3, \alpha_4) = (0, 0.5)$.

We use standard normal kernel functions for both $L(\cdot)$ and $K(\cdot)$ with smoothing parameters chosen via $a_w = w_{sd}n^{-1/5}$, $h_w = cw_{sd}n^{-1/4}$ and $h_z = cz_{sd}n^{-1/4}$ for DGP1 and DGP2; $a_y = y_{-1,sd}n^{-1/5}$, $h_y = cy_{-1,sd}n^{-1/4}$ and $h_z = cz_{sd}n^{-1/4}$ for DGP3 and DGP4, where w_{sd}, z_{sd} and $y_{-1,sd}$ are sample standard deviations of $\{W_t\}_{t=1}^n$, $\{Z_t\}_{t=1}^n$ and $\{Y_{t-1}\}_{t=2}^n$, respectively. The smoothing parameter a is associated with kernel $L(\cdot)$ that is used for estimating the (restricted) null model and h is the smoothing parameter associated with kernel $K(\cdot)$. The above choices of a and h satisfy condition (A2) of Theorem 3.1. To check the sensitivity of our tests with respect to different values of a and h. We fixed the value of a and change h via different values of c: we use $c = 0.25, 0.5, 1, 2$. The number of replications is 2000 for all cases.

Estimated sizes of the $\hat{J}_n^a$ and the V_n^a tests (for DGP1) based on asymptotic one-sided normal critical values are reported in Tables 1 and 2, respectively.

From Table 1 we observe that the estimated sizes for the $\hat{J}_n^a$ test under estimates the nominal sizes for all cases considered. The results does suggest that as n increases, the estimated sizes convergent to their nominal sizes although at a fairly slow rate. The estimated sizes are closer to their nominal sizes for smaller values of c (for the range of c values considered). This result can be explained by the fact that the rate our test converges to a standard normal variate (under H_0^a) is $O_p(nh^{d/2}(a^{2v} + (na^q)^{-1})) = O_p(nh(a^4 + (na)^{-1}))$. Hence, for a fixed value of n and a, a smaller h (i.e., smaller c) will lead to a smaller error in the normal approximation. But this does not mean that one should use a very small value of h in practice. Because too small a h may cause the kernel estimation to be inaccurate and more importantly, under H_1^a, our test diverges to $+\infty$ at the rate of $(nh^{d/2})$, too small a h will make the test not powerful (this is confirmed in our simulations, see Tables 3 and 4).

Table 2 shows that the estimated sizes for the V_n^a test are closer to their nominal values than the $\hat{J}_n^a$ test of Table 1. In contrast to the negative bias of $\hat{J}_n^a$, V_n^a test has positive (finite sample) bias which makes the $\hat{V}_n^a$ oversized for most cases for $0.25 \leqslant c \leqslant 1$. For $c = 2$, V_n^a is undersized mainly because its standard deviation is significantly less than one.

Table 3 gives the estimated powers of the $\hat{J}_n^a$ and the V_n^a tests for DGP2. The results show that for most cases, the power of V_n^a dominates the power of $\hat{J}_n^a$. This is because the $\hat{J}_n^a$ is undersized under H_0^a. Therefore, this will hurt the finite sample power of $\hat{J}_n^a$ under H_1^a. In general, both the $\hat{J}_n^a$ and the V_n^a tests are quite powerful in detecting alternatives of case (i) and case (ii) of DGP2 as they should since our nonparametric tests are consistent tests.

An interesting fact is that for all cases considered in Table 3, and for c values between 0.5 and 2, the higher value is the smoothing parameter h (i.e, higher value of c), the higher are the powers of the $\hat{J}_n^a$ and the V_n^a tests. This result can be

Table 1
Size of $\hat{J}_n^a$ (DGP1)

	c = 0.25				c = 0.5			
n	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
50	0.002	0.032	0.083	– 0.154 (0.981)	0.008	0.034	0.062	– 0.239 (0.944)
100	0.004	0.043	0.094	– 0.102 (0.981)	0.009	0.045	0.073	– 0.220 (0.969)
200	0.007	0.042	0.095	– 0.101 (1.00)	0.011	0.032	0.078	– 0.191 (0.973)
500	0.006	0.043	0.078	– 0.125 (0.986)	0.009	0.036	0.067	– 0.209 (0.953)
1000	0.009	0.043	0.083	– 0.062 (0.986)	0.011	0.036	0.072	– 0.190 (0.966)
	c = 1				c = 2			
	1%	5%	10%	mean (std)	1%	5%	10%	Mean (std)
50	0.007	0.021	0.043	– 0.381 (0.852)	0.003	0.005	0.010	– 0.527 (0.606)
100	0.006	0.024	0.047	– 0.358 (0.855)	0.004	0.015	0.024	– 0.404 (0.678)
200	0.008	0.029	0.052	– 0.309 (0.893)	0.005	0.014	0.025	– 0.361 (0.705)
500	0.007	0.028	0.058	– 0.291 (0.908)	0.005	0.019	0.036	– 0.297 (0.736)
1000	0.009	0.025	0.048	– 0.260 (0.882)	0.006	0.020	0.038	– 0.278 (0.768)

Table 2
Size of V_n^a (DGP1)

	c = 0.25				c = 0.5			
n	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
50	0.002	0.039	0.106	0.046 (0.915)	0.006	0.043	0.094	0.056 (0.871)
100	0.007	0.061	0.117	0.090 (0.961)	0.015	0.063	0.120	0.143 (0.932)
200	0.010	0.059	0.120	0.082 (0.959)	0.017	0.067	0.118	0.117 (0.939)
	c = 1				c = 2			
	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
50	0.010	0.042	0.082	0.114 (0.773)	0.005	0.020	0.048	0.151 (0.577)
100	0.022	0.063	0.114	0.230 (0.855)	0.014	0.038	0.072	0.286 (0.660)
200	0.018	0.060	0.116	0.221 (0.856)	0.013	0.049	0.087	0.264 (0.694)

Table 3
Power for DGP2

Test	n	$c = 0.25$			$c = 0.5$			$c = 1$			$c = 2$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Case (i) of DGP2													
$\hat{J}_n^a$	50	0.027	0.199	0.364	0.227	0.461	0.578	0.581	0.746	0.811	0.779	0.891	0.925
V_n^a	50	0.027	0.224	0.405	0.262	0.518	0.642	0.651	0.804	0.885	0.863	0.941	0.975
$\hat{J}_n^a$	100	0.179	0.457	0.606	0.626	0.811	0.894	0.940	0.979	0.987	0.989	0.995	0.997
V_n^a	100	0.191	0.481	0.635	0.671	0.855	0.917	0.959	0.987	0.992	0.993	0.997	0.998
Case (ii) of DGP2													
$\hat{J}_n^a$	50	0.022	0.174	0.316	0.181	0.391	0.506	0.438	0.613	0.694	0.495	0.663	0.735
V_n^a	50	0.019	0.171	0.326	0.190	0.434	0.560	0.497	0.681	0.773	0.603	0.767	0.847
$\hat{J}_n^a$	100	0.135	0.396	0.535	0.510	0.708	0.782	0.822	0.911	0.936	0.929	0.969	0.979
V_n^a	100	0.142	0.416	0.566	0.555	0.754	0.819	0.873	0.939	0.969	0.963	0.983	0.991

Table 4
Size of $\hat{J}_n^a$ (DGP3)

	$c = 0.25$				$c = 0.5$			
n	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
50	0.003	0.030	0.075	− 0.211 (0.975)	0.004	0.024	0.056	− 0.379 (0.943)
100	0.005	0.038	0.073	− 0.208 (0.999)	0.006	0.029	0.048	− 0.335 (0.950)
200	0.004	0.032	0.064	− 0.201 (0.960)	0.003	0.028	0.052	− 0.345 (0.922)
500	0.005	0.035	0.079	− 0.187 (0.981)	0.006	0.024	0.056	− 0.331 (0.927)
1000	0.009	0.049	0.089	− 0.132 (10.00)	0.007	0.030	0.060	− 0.277 (0.964)
	$c = 1$				$c = 2$			
n	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
50	0.002	0.010	0.027	− 0.583 (0.809)	0.001	0.006	0.011	− 0.716 (0.570)
100	0.005	0.017	0.028	− 0.560 (0.825)	0.002	0.007	0.012	− 0.689 (0.601)
200	0.003	0.013	0.025	− 0.554 (0.821)	0.003	0.008	0.015	− 0.664 (0.654)
500	0.005	0.015	0.032	− 0.534 (0.861)	0.004	0.013	0.021	− 0.658 (0.741)
1000	0.005	0.023	0.040	− 0.487 (0.912)	0.005	0.019	0.030	− 0.604 (0.824)

explained by the fact that our tests diverge to $+\infty$ at the rate of $nh^{d/2}$ under H_1^a. Hence, a higher h (in certain range) will lead to a more powerful test against some fixed alternatives (in finite samples). Another explanation for this result is that the DGP2 contains a low frequency linear (function) deviation from the null model, and it is known that a relative large value of h should be used for low frequency alternatives. But we caution the applied researchers that in practice, h cannot be chosen too large, a very large h will in fact lead to a test that does not have any power because it over smooth the data too much and hence obscure any deviation of the data from the null DGP.

Summarizing the results of Tables 1–3, we observe the followings: (i) for the range of c values we considered, $c = 2$ gives the best power results for both the $\hat{J}_n^a$ and the V_n^a tests but at the same time, $c = 2$ also correspond to the most size distortions. However, the size distortions are at the direction of under size. Therefore, the case of $c = 2$ also give the smallest type I error. In this sense the

$c = 2$ case gives the best results because both tests have the smallest type I and type II errors for $c = 2$ ($c = 2$ is the most undersized under H_0^a and the most powerful one under H_1^a). This result is quite interesting because usually the size-power trade off of a test statistic is that the more powerful case also tends to be the more over-sized case. Of course one cannot draw general conclusions about the $\hat{J}_n^a$ and the V_n^a tests based on the limited Monte Carlo experiments reported above.

The estimated sizes for the $\hat{J}_n^a$ and the V_n^a tests for DGP3 are reported in Tables 4 and 5, respectively.

For the $\hat{J}_n^a$ test, the result is very similar the case of DGP1 as given in Table 1, i.e., the $\hat{J}_n^a$ test under estimates the nominal sizes for all cases considered, the smaller values of c gives better estimated sizes (for $0.5 \leqslant c \leqslant 2$). Also, the estimated sizes seem to convergent to their nominal sizes although at a fairly slow rate.

For the V_n^a test, from Table 5 we observe that the estimated sizes of V_n^a are much better than that of the $\hat{J}_n^a$ test of Table 4. In particular, the biases of V_n^a are fairly small for all cases. The estimated standard deviation of V_n^a decreases as c increases, causing the V_n^a test undersized for large values of c.

Table 6 gives the estimated power of the $\hat{J}_n^a$ and the V_n^a tests against DGP4. Similar to the case of DGP2 (see Table 3), the results of Table 6 show that the power of V_n^a dominates the power of $\hat{J}_n^a$. Both tests are quite powerful in detecting these alternative processes. For the range of c values considered, the larger value of c (or h) leads to a more powerful tests against DGP4.

Table 5
Size of V_n^a: (DGP3)

	$c = 0.25$				$c = 0.5$			
n	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
50	0.004	0.040	0.101	0.019 (0.910)	0.006	0.043	0.089	0.002 (0.878)
100	0.006	0.047	0.097	0.003 (0.952)	0.011	0.043	0.090	0.010 (0.909)
200	0.007	0.039	0.088	– 0.003 (0.923)	0.010	0.046	0.091	0.013 (0.885)
	$c = 1$				$c = 2$			
n	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
50	0.006	0.029	0.068	– 0.006 (0.760)	0.003	0.016	0.028	– 0.021 (0.540)
100	0.011	0.035	0.067	0.010 (0.792)	0.002	0.018	0.036	0.013 (0.579)
200	0.009	0.030	0.069	0.012 (0.787)	0.005	0.022	0.044	0.032 (0.637)

Table 6
Power for DGP4, case (i)

Test	n	$c = 0.25$			$c = 0.5$			$c = 1$			$c = 2$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Case (i) of DGP2													
$\hat{J}_n^a$	50	0.007	0.089	0.193	0.070	0.202	0.294	0.215	0.386	0.473	0.367	0.535	0.610
V_n^a	50	0.012	0.110	0.236	0.088	0.243	0.355	0.290	0.469	0.581	0.490	0.667	0.755
$\hat{J}_n^a$	100	0.037	0.216	0.335	0.206	0.428	0.544	0.568	0.735	0.799	0.817	0.900	0.944
V_n^a	100	0.045	0.237	0.373	0.272	0.499	0.622	0.667	0.816	0.883	0.889	0.958	0.978
Case (ii) of DGP2													
$\hat{J}_n^a$	50	0.024	0.152	0.285	0.155	0.333	0.447	0.354	0.525	0.617	0.389	0.553	0.645
V_n^a	50	0.024	0.149	0.292	0.179	0.368	0.482	0.405	0.613	0.695	0.477	0.663	0.774
$\hat{J}_n^a$	100	0.098	0.290	0.424	0.386	0.606	0.705	0.739	0.850	0.885	0.849	0.919	0.952
V_n^a	100	0.110	0.302	0.455	0.439	0.659	0.756	0.795	0.889	0.938	0.906	0.966	0.978

6.2. Testing the null of a partially linear model

This subsection reports Monte Carlo results to study the finite sample performances of the $\hat{J}_n^b$ and the V_n^b tests. For testing the null of a partially linear model, we use the following data generating processes:

$$\text{DGP5:}\quad Y_t = 1 + Z_t + W_t + U_t,$$

$$\text{DGP6:}\quad Y_t = 1 + Z_t + W_t + \alpha_5(W_t Z_t) + U_t,$$

$$\text{DGP7:}\quad Y_t = 0.5Y_{t-1} + W_t + U_t,$$

$$\text{DGP8:}\quad Y_t = 0.5Y_{t-1} + W_t + \alpha_6(Y_{t-1}W_t) + U_t,$$

where W_t, Z_t and U_t are generated by the same ways as in DGP1–DGP4. DGP5 is an null (H_0^b) of a partially linear model: $Y_t = 1 + Z_t + W_t + U_t \equiv Z_t + \theta(W_t) + U_t$. DGP6 is an alternative model (H_1^b) and we choose $\alpha_5 = 1$. Similarly DGP7 is an null of a partially linear model: $Y_t = 0.5Y_{t-1} + \theta(W_t) + U_t$. DGP8 is an alternative model (H_1^b) and we choose $\alpha_6 = 0.5$.

Again we use standard normal kernel functions and the smoothing parameters are chosen using the same methods as in Section 6.1. In particular, we fix the smoothing parameter a and change the smoothing parameter h via different choices of c ($c = 0.25, 0.5, 1, 2$). The number of replications are 2000 for all cases.

Table 7 reports the estimated sizes for the $\hat{J}_n^b$ and the V_n^b tests for DGP5.

For the $\hat{J}_n^b$ test, the results is in general similar the case of DGP1 (see Table 1). That is, the $\hat{J}_n^b$ test under estimates the nominal sizes and a larger c value corresponds to a larger size distortion. The case of $c = 2$ is even more down sized than the case of $c = 1$, the estimated sizes of $\hat{J}_n^b$ for $c = 2$ is not reported here to save space.

For the V_n^b test, the estimated sizes are much closer to their nominal values than the $\hat{J}_n^b$ test. The biases are fairly small for all cases considered. Similar to the case of DGP1 and DGP3, the standard deviation of V_n^b decreases as c increases causing the V_n^b test downsized for large values of c. The case of $c = 2$ is more undersized than the case of $c = 1$, the estimated sizes of V_n^b for $c = 2$ is not reported here to save space.

Table 8 reports the estimated power of the $\hat{J}_n^b$ and the V_n^b tests against DGP6. We observe that the V_n^b test dominates the $\hat{J}_n^b$ test. Also as expected we observe that the power of both tests increase as n increases. However, for the range of c values considered, the power of these tests are no longer monotone in c. There are a few cases that the powers of both tests are larger for $c = 1$ than for $c = 2$.

Table 7
Size for DGP5

	$c = 0.25$				$c = 0.5$				$c = 1$			
n	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
The $\hat{J}_n^b$ test												
50	0.002	0.027	0.069	− 0.241 (0.973)	0.002	0.017	0.040	− 0.439 (0.913)	0.002	0.008	0.017	− 0.683 (0.758)
100	0.003	0.037	0.079	− 0.190 (0.997)	0.006	0.028	0.051	− 0.345 (0.946)	0.005	0.015	0.024	− 0.576 (0.812)
200	0.007	0.036	0.079	− 0.174 (0.979)	0.005	0.027	0.054	− 0.352 (0.946)	0.003	0.013	0.025	− 0.554 (0.821)
The V_n^b test												
50	0.003	0.037	0.087	− 0.025 (0.907)	0.005	0.032	0.065	− 0.092 (0.842)	0.002	0.018	0.038	− 0.140 (0.683)
100	0.004	0.051	0.105	− 0.011 (0.953)	0.012	0.043	0.089	− 0.005 (0.905)	0.009	0.026	0.054	− 0.026 (0.776)
200	0.007	0.040	0.100	− 0.010 (0.949)	0.008	0.046	0.083	− 0.054 (0.919)	0.007	0.029	0.070	− 0.053 (0.797)

Table 8
Power for DGP6

		$c = 0.25$			$c = 0.5$			$c = 1$			$c = 2$		
Test	n	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{J}_n^b$	50	0.016	0.108	0.221	0.070	0.209	0.301	0.178	0.311	0.407	0.119	0.235	0.331
V_n^b	50	0.018	0.113	0.245	0.090	0.246	0.367	0.220	0.383	0.510	0.181	0.349	0.483
$\hat{J}_n^b$	100	0.071	0.232	0.349	0.261	0.454	0.563	0.538	0.706	0.788	0.635	0.796	0.869
V_n^b	100	0.081	0.263	0.397	0.307	0.531	0.635	0.642	0.788	0.857	0.756	0.886	0.997

Table 9
Size for DGP7

n	$c = 0.25$				$c = 0.5$				$c = 1$			
	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)	1%	5%	10%	Mean (std)
The $\hat{J}_n^b$ test												
50	0.002	0.032	0.077	− 0.242 (0.990)	0.003	0.023	0.048	− 0.422 (0.912)	0.001	0.007	0.016	− 0.688 (0.725)
100	0.004	0.037	0.073	− 0.221 (0.983)	0.004	0.022	0.041	− 0.419 (0.910)	0.002	0.009	0.017	− 0.649 (0.752)
200	0.005	0.034	0.078	− 0.203 (0.974)	0.003	0.025	0.051	− 0.360 (0.945)	0.002	0.010	0.024	− 0.581 (0.812)
The V_n^b test												
50	0.003	0.036	0.092	− 0.022 (0.917)	0.007	0.037	0.073	− 0.052 (0.848)	0.004	0.018	0.039	− 0.114 (0.674)
100	0.004	0.047	0.091	− 0.016 (0.939)	0.008	0.036	0.072	− 0.056 (0.868)	0.004	0.020	0.045	− 0.079 (0.713)
200	0.010	0.040	0.096	− 0.009 (0.950)	0.007	0.045	0.098	− 0.005 (0.930)	0.005	0.035	0.066	− 0.021 (0.796)

Table 10
Power for DGP8

Test	n	$c = 0.25$			$c = 0.5$			$c = 1$			$c = 2$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{J}_n^b$	50	0.102	0.218	0.305	0.182	0.291	0.377	0.237	0.339	0.406	0.145	0.238	0.304
V_n^b	50	0.103	0.223	0.305	0.180	0.290	0.377	0.237	0.339	0.406	0.145	0.238	0.304
$\hat{J}_n^b$	100	0.293	0.418	0.506	0.438	0.557	0.632	0.558	0.668	0.731	0.514	0.653	0.740
V_n^b	100	0.293	0.418	0.506	0.438	0.557	0.632	0.558	0.680	0.760	0.514	0.675	0.756

As we mentioned earlier, a large value of c corresponds to over-smooth the alternative model and hence may lead to a low power test if it over-smooth too much the data (especially for high frequency alternatives).

Finally, Tables 9 and 10 give the estimated sizes and powers of the $\hat{J}_n^b$ and the V_n^b tests for DGP7 and DGP8. The results are similar to the cases of DGP5 and DGP6. In particular, the $\hat{J}_n^b$ test is under-sized while the V_n^b test has much better estimated sizes. The $\hat{J}_n^b$ test is negatively biased while the V_n^b test has fairly small biases. This is why V_n^b is more powerful than $\hat{J}_n^b$. Also from Table 10 we observe that both the $\hat{J}_n^b$ and the V_n^b tests are more powerful for the case of $c = 1$ than the case of $c = 2$, giving more evidence that over-smooth too much the data will lead to low power tests.

Summarizing the limited Monte Carlo simulation results reported above. The $\hat{J}_n$ ($\hat{J}_n^a$ or $\hat{J}_n^b$) test is substantially undersized for all cases considered. The V_n (V_n^a or V_n^b) test gives much better estimated sizes than the $\hat{J}_n$ test.

In general the $\hat{J}_n$ test is less powerful than the V_n test due to the fact that the $\hat{J}_n$ test is biased toward accepting the null (i.e., it is substantially undersized). The estimated powers of both tests are sensitive to the relative smoothing parameter choices (as is often the case with nonparametric kernel estimation methods). For low frequency alternatives, a relatively large smoothing parameter h will lead to a high power test. While for high-frequency alternatives, a relatively small smoothing parameter h should be used (will lead to a high power test). Therefore, how to choose the relative smoothing parameters optimally in the sense that the power of the tests are maximized and at the same time to keep the type I error under control is an important future research topic.

Another research topic that deserves effort is to investigate the possibility of using various parametric and nonparametric bootstrap methods (for dependent data) to approximate the null distributions of the proposed tests. Bootstrap tests may provide better estimated sizes than both the $\hat{J}_n$ and the V_n tests. To my knowledge, even with the independent data, the tests considered in Fan and Li (1996a) and Lavergne and Vuong (1996b) have not been investigated by bootstrap methods. The asymptotic theory established in this paper will be useful to the bootstrap analysis of these tests (for dependent data case). The theoretical justification of bootstrap techniques in our context, and specifically the conditions under which they apply, are left for future research.

Acknowledgements

I would like to thank two referees and Peter Robinson for helpful suggestions that greatly improved the paper. This research is supported by the Social Science and Humanity Research Council of Canada and the Natural Sciences Engineering Research Council of Canada.

Appendix A. Proof of the Theorem 3.1 and Corollary 3.2

Below we first list the conditions (D1) and (D2) that are used in Lemma 2.1 and the two definitions for the class of kernel function $\mathscr{K}_\lambda$ and the class of function of $\mathscr{G}_\mu^\alpha$ (see Robinson (1988)).

Let $\mathscr{Z}_t$ be a strictly stationary process and $\mathscr{M}_s^t(\mathscr{Z})$ denote the sigma algebra generated by $(Z_s, \dots, Z_t)$ for $s \leqslant t$. The process $\mathscr{Z}_t$ is called absolutely regular, if as $\tau \to \infty$,

$$\beta_\tau = \sup_{s \in N} \mathrm{E}\left[\sup_{A \in \mathscr{M}_{s+\tau}^\infty} \{ |P(A|\mathscr{M}_{-\infty}^s) - P(A)| \} \right] \to 0.$$

The following conditions are used for Lemma 2.1.

(D1) (i) *The process* $\mathscr{Z}_t = \{\varepsilon_t, X_t'\}$ $(X_t \in R^d)$ *is strictly stationary and absolutely regular with the mixing coefficient* $\beta_m = \mathrm{O}(\rho^m)$ *for some* $0 < \rho < 1$; *with probability one*, $\mathrm{E}[\varepsilon_t | \mathscr{M}_{-\infty}^t(X), \mathscr{M}_{-\infty}^{t-1}(\varepsilon)] = 0$. (ii) $\mathrm{E}[|\varepsilon_t^{4+\eta}|] < \infty$ *and* $\mathrm{E}[|\varepsilon_{t_1}^{i_1}\varepsilon_{t_2}^{i_2}, \dots, \varepsilon_{t_l}^{i_l}|^{1+\xi}] < \infty$ *for some arbitrarily small* $\eta > 0$ *and* $\xi > 0$, *where* $2 \leqslant l \leqslant 4$ *is an integer*, $0 \leqslant i_j \leqslant 4$ *and* $\sum_{j=1}^l i_j \leqslant 8$. (iii) Let $\sigma_\varepsilon^2(x) = \mathrm{E}(\varepsilon_t^2 | X_t = x)$, $\mu_{\varepsilon 4}(x) = \mathrm{E}(\varepsilon_t^4 | X_t = x)$. $\sigma_\varepsilon^2(x)$ *and* $\mu_{\varepsilon 4}(x)$ *satisfy some Lipschitz conditions*: $|\sigma_\varepsilon^2(u+v) - \sigma_\varepsilon^2(u)| \leqslant D(u)\|v\|$ *and* $|\mu_{\varepsilon 4}(u+v) - \mu_{\varepsilon 4}(u)| \leqslant D(u)\|v\|$ *with* $\mathrm{E}[|D(X)|^{2+\eta'}] < \infty$ *for some small* $\eta' > 0$. (iv) *Let* $f_{\tau_1,\dots,\tau_l}(.,\dots,.)$ *be the joint probability density function of* $(X_1, X_{1+\tau_1}, \dots, X_{1+\tau_l})$ $(1 \leqslant l \leqslant 3)$. *Then* $f_{\tau_1,\dots,\tau_l}(.,\dots,.)$ *exists and satisfies a Lipschitz condition*: $|f_{\tau_1,\dots,\tau_l}(x_1+u_1, x_2+u_2, \dots, x_l+u_l) - f_{\tau_1,\dots,\tau_l}(x_1, x_2, \dots, x_l)| \leqslant D_{\tau_1,\dots,\tau_l}(x_1, x_2, \dots, x_l)\|u\|$, *where* $D_{\tau_1,\dots,\tau_l}(.,\dots,.)$ *is integrable and satisfies the condition that* $\int D_{\tau_1,\dots,\tau_l}(x,\dots,x)\|x\|^{2\xi} < M < \infty$, $\int D_{\tau_1,\dots,\tau_l}(x_1,\dots,x_l) f_{\tau_1,\dots,\tau_l}(x_1,\dots,x_l)\,\mathrm{d}x < M < \infty$ *for some* $\xi > 1$.

(D2) (i) $K(\cdot)$ *is bounded and symmetric with* $\int K(u)\,\mathrm{d}u = 1$ *and* $\int \|u\|^2 K(u)\,\mathrm{d}u < \infty$. (ii) *The smoothing parameter* $h = \mathrm{O}(n^{-\bar{\alpha}})$ *for some* $0 < \bar{\alpha} < (7/8)d$.

The following definitions are adopted from Robinson (1988).

Definition A.1. $\mathscr{K}_\lambda, \lambda \geqslant 1$, is the class of even functions $k: R \to R$ satisfying

$$\int_R u^i k(u)\,\mathrm{d}u = \delta_{i0} \quad (i = 0, 1, \dots, \lambda - 1),$$

$$k(u) = \mathrm{O}((1 + |u|^{\lambda+1+\varepsilon})^{-1}), \quad \text{some } \varepsilon > 0,$$

where δ_{ij} is the Kronecker's delta.

Definition A.2. $\mathscr{G}_\mu^\alpha$, $\alpha > 0$, $\mu > 0$, is the class of functions $g: R^d \to R$ satisfying: g is $(m-1)$-times partially differentiable, for $m - 1 \leqslant \mu \leqslant m$; for some $\rho > 0$,

$\sup_{y\in\phi_{z\rho}}|g(y)-g(z)-Q_g(y,z)|/|y-z|^{\mu}\leqslant D_g(z)$ 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 αth moments.

The remaining parts of this appendix prove Theorem 3.1 and Corollary 3.2. Throughout, the symbol C denotes a generic constant. The notation $A\sim B$ means that A has an order no larger than that of B. We denote $\hat{f}_t=(1/(n-1)h^d)\sum_{s\neq t}K((X_t-X_s)/h)$, the kernel estimator of $f(X_t)$.

Proof of Theorem 3.1. We will only prove Theorem 3.1(i) since the proof of Theorem 3.1(ii) is similar to and in fact much simpler than the proof of Theorem 3.1(i). We often write u_t for U_t and w_t for W_t to save space. Variables with subscript are always random variables even when small letter case is used.

Using $\tilde{U}_t=Y_t-\hat{Y}_t=(r_t-\hat{r}_t)+U_t-\hat{U}_t$, where $r_t=r(W_t)$ and $\hat{r}_t=(na^q)^{-1}\sum_{s\neq t}r_sL_{ts}/\hat{f}_{w_t}$, the following expression for $\hat{J}^a_n$ is immediate from Eq. (6):

$$\hat{J}^a_n=\frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\{(r_t-\hat{r}_t)\hat{f}_{w_t}(r_s-\hat{r}_s)\hat{f}_{w_s}+u_tu_s\hat{f}_{w_t}\hat{f}_{w_s}$$

$$+\hat{u}_t\hat{f}_{w_t}\hat{u}_s\hat{f}_{w_s}+2u_t\hat{f}_{w_t}(r_s-\hat{r}_s)\hat{f}_{w_s}-2\hat{u}_t\hat{f}_{w_t}(r_s-\hat{r}_s)\hat{f}_{w_s}$$

$$-2u_t\hat{f}_{w_t}\hat{u}_s\hat{f}_{w_s}\}K_{ts}\stackrel{\text{def}}{=}J_{n1}+J_{n2}+J_{n3}+2J_{n4}-2J_{n5}-2J_{n6}. \quad \text{(A.1)}$$

We shall complete the proof of Theorem 3.1(i) by showing that $J_{ni}=o_p((nh^{d/2})^{-1})$ for $i=1,3,4,5,6$ and $nh^{d/2}J_{n2}/\hat{\sigma}_a\to N(0,1)$ in distribution. These results are proved in Lemmas A.1 to A.6 below.

Lemma A.1. $J_{n1}=o_p((nh^{d/2})^{-1})$.

Proof. Note that $K(\cdot)$ is a non-negative function and $\hat{f}_t=(1/(n-1)h^d)\sum_{s\neq t}K_{ts}$, we have

$$E|J_{n1}|=E\left|\frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}(r_t-\hat{r}_t)\hat{f}_{w_t}(r_s-\hat{r}_s)\hat{f}_{w_s}K_{ts}\right|$$

$$\leqslant \frac{1}{2(n(n-1)h^d} \sum_t \sum_{s \neq t} \mathrm{E}\{[(r_t - \hat{r}_t)^2 \hat{f}^{\,2}_{w_t} + (r_s - \hat{r}_s)^2 \hat{f}^{\,2}_{w_s}] K_{ts}\}$$

$$= \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \mathrm{E}[(r_t - \hat{r}_t)^2 \hat{f}^{\,2}_{w_t} K_{ts}]$$

$$= n^{-1} \sum_t \mathrm{E}[(r_t - \hat{r}_t)^2 \hat{f}^{\,2}_{w_t} \hat{f}_t] \equiv J_{1b}, \text{ say.}$$

$$J_{1b} = n^{-1} \sum_t \mathrm{E}[(r_t - \hat{r}_t)^2 \hat{f}^{\,2}_{w_t} f_t] + n^{-1} \sum_t \mathrm{E}[(r_t - \hat{r}_t)^2 \hat{f}^{\,2}_{w_t} (\hat{f}_t - f_t)]$$

$$\leqslant Cn^{-1} \sum_t \mathrm{E}[(r_t - \hat{r}_t)^2 \hat{f}^{\,2}_{w_t}] + n^{-1} \sum_t \mathrm{E}[(r_t - \hat{r}_t)^2 \hat{f}^{\,2}_{w_t} (\hat{f}_t - f_t)]$$

$$= \mathrm{O}(a^{2v} + (na^q)^{-1}) = \mathrm{o}((nh^{d/2})^{-1})$$

by Lemmas C.3(i) and C.4(i).

Summarizing the above, we have shown that $\mathrm{E}|J_{n1}| \leqslant J_{1b} = \mathrm{o}((nh^{d/2})^{-1})$. Hence, $J_{n1} = \mathrm{o_p}((nh^{d/2})^{-1})$.

Lemma A.2. (i) *$nh^{d/2} J_{n2} \to \mathrm{N}(0, \sigma_a^2)$ in distribution, where $\sigma_a^2 = 2\mathrm{E}[f(X_1)\sigma^4(X_1) f^4_{w_1}][\int K^2(u)\,\mathrm{d}u]$,*
(ii) $\hat{\sigma}_a^2 = \sigma_a^2 + \mathrm{o_p}(1)$.

Proof of (i).

$$J_{n2} = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} u_t u_s f_{w_t} f_{w_s} K_{ts}$$

$$+ \frac{2}{n(n-1)h^d} \sum_t \sum_{s \neq t} u_t u_s (\hat{f}_{w_t} - f_{w_t}) f_{w_s} K_{ts}$$

$$+ \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} u_t u_s (\hat{f}_{w_t} - f_{w_t})(\hat{f}_{w_s} - f_{w_s}) K_{ts}$$

$$\equiv J_{n21} + 2J_{n22} + J_{n23}, \text{ say}$$

J_{n21} is a second order degenerate U-statistic of the form of Eq. (2) with $\varepsilon_t = u_t f_{w_t}$. It is easy to check that the conditions (A1)–(A2) imply (D1)–(D2). Hence by Lemma 2.1, we have $J_{n21} \to \mathrm{N}(0, \sigma_a^2)$ in distribution.

Next, $J_{n22} = \mathrm{o_p}((nh^{d/2})^{-1})$ by Lemma C.5(ii).

Finally,

$$\mathrm{E}|\,J_{n23}| \leqslant \frac{1}{2n(n-1)h^d} \sum_t \sum_{s \neq t} \mathrm{E}\{u_t^2(\hat{f}_{w_t} - f_{w_t})^2 K_{ts} + u_s^2(\hat{f}_{w_s} - f_{w_s})^2 K_{ts}\}$$

$$= \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \mathrm{E}[u_t^2(\hat{f}_{w_t} - f_{w_t})^2 K_{ts}]$$

$$= n^{-1} \sum_t \mathrm{E}[u_t^2(\hat{f}_{w_t} - f_{w_t})^2 \hat{f}_t] \equiv J_{2b}.$$

$$J_{2b} = n^{-1} \sum_t \mathrm{E}[u_t^2(\hat{f}_{w_t} - f_{w_t})^2 f_t] + n^{-1} \sum_t \mathrm{E}[u_t^2(\hat{f}_{w_t} - f_{w_t})^2(\hat{f}_t - f_t)]$$

$$\leqslant Cn^{-1} \sum_t \mathrm{E}[u_t^2(\hat{f}_{w_t} - f_{w_t})^2] + n^{-1} \sum_t \mathrm{E}[u_t^2(\hat{f}_{w_t} - f_{w_t})^2(\hat{f}_t - f_t)]$$

$$= \mathrm{O}(a^{2v} + (na^q)^{-1})$$

by Lemmas C.3(ii) and C.4(ii).

Hence, $\mathrm{E}|J_{n23}| \leqslant J_{2b} = \mathrm{O}(a^{2v} + (na^q)^{-1}) = \mathrm{o}((nh^{d/2})^{-1})$, which implies $J_{n23} = \mathrm{o_p}((nh^{d/2})^{-1})$.

Proof of (ii). $\hat{\sigma}_a^2 = \sigma_a^2 + \mathrm{o_p}(1)$. The proof for (ii) is similar to (and much easier than) that of (i). Hence, we will provide a sketchy proof here. Using $\tilde{u}_t = u_t + \mathrm{o_p}(1)$, one can show that

$$\hat{\sigma}_a^2 = \frac{2}{n(n-1)h^d} \sum_t \sum_{s \neq t} [\tilde{u}_t \tilde{u}_s]^2 [\hat{f}_{w_t} \hat{f}_{w_s}]^2 K_{ts}$$

$$= \frac{2}{n(n-1)h^d} \sum_t \sum_{s \neq t} u_t^2 u_s^2 [\hat{f}_{w_t} \hat{f}_{w_s}]^2 K_{ts} + \mathrm{o_p}(1)$$

$$= \frac{2}{n(n-1)h^d} \sum_t \sum_{s \neq t} u_t^2 u_s^2 [f_{w_t} f_{w_s}]^2 K_{ts} + \mathrm{o_p}(1) \equiv \bar{\sigma}_a^2 + \mathrm{o_p}(1), \text{ say.}$$

Finally the proof of $\bar{\sigma}_a^2 = \sigma_a^2 + o_p(1)$ follows from the facts that (using Lemma C.1)

$$E(\bar{\sigma}_a^2) = \sigma_a^2 + \frac{1}{n(n-1)h^d} O\left(\sum_t \sum_{s\neq} \beta_m^{\delta/(1+\delta)}\right)$$

$$= \sigma_a^2 + O((nh^d)^{-1}) \quad \text{and} \quad var(\bar{\sigma}_a^2) = o(1).$$

Lemma A.3. $J_{n3} = o_p((nh^{d/2})^{-1})$.

Proof.

$$E|J_{n3}| = E\left|\frac{1}{n(n-1)h^d} \sum_t \sum_{s\neq t} \hat{u}_t \hat{f}_{w_t} \hat{u}_s \hat{f}_{w_s} K_{ts}\right|$$

$$\leqslant \frac{1}{n(n-1)h^d} \sum_t \sum_{s\neq t} E[\hat{u}_t^2 \hat{f}_{w_t}^2 K_{ts}] = n^{-1} \sum_t E[\hat{u}_t^2 \hat{f}_{w_t}^2 \hat{f}_t] \equiv J_{3b}, \text{ say.}$$

$$J_{3b} = n^{-1} \sum_t E[\hat{u}_t^2 \hat{f}_{w_t}^2 f_t] + n^{-1} \sum_t E[\hat{u}_t^2 \hat{f}_{w_t}^2 (\hat{f}_t - f_t)]$$

$$= O((na^q)^{-1} + a^{2v}) = o((nh^{d/2})^{-1})$$

by Lemmas C.3(iii) and C.4(iii), which implies $J_{n3} = o_p((nh^{d/2})^{-1})$.

Lemma A.4. $J_{n4} = o_p((nh^{d/2})^{-1})$.

Proof.

$$J_{n4} = \frac{1}{n(n-1)h^d} \sum_t \sum_{s\neq t} u_t \hat{f}_{w_t} (r_s - \hat{r}_s) \hat{f}_{w_s} K_{ts}$$

$$= \frac{1}{n(n-1)h^d} \sum_t \sum_{s\neq t} u_t f_{w_t} (r_s - \hat{r}_s) \hat{f}_{w_s} K_{ts}$$

$$+ \frac{1}{n(n-1)h^d} \sum_t \sum_{s\neq t} u_t (\hat{f}_{w_t} - f_{w_t})(r_s - \hat{r}_s) \hat{f}_{w_s} K_{ts}$$

$$\equiv J_{n41} + J_{n42}, \text{ say.}$$

$J_{n41} = o_p((nh^{d/2})^{-1})$ by Lemma C.5(i) and

$$E|J_{n42}| \leqslant \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} E\{[(r_s - \hat{r}_s)^2 \hat{f}^2_{w_s} + u_t^2(\hat{f}_{w_t} - f_{w_t})^2]K_{ts}\}$$

$$= J_{1b} + J_{2b} = o((nh^{d/2})^{-1})$$

by the proofs of Lemmas A.1 and A.2, where J_{1b} and J_{2b} are defined in the proofs of Lemmas A.1 and A.2, respectively. Hence, $J_{n42} = o_p((nh^{d/2})^{-1})$.

Lemma A.5. $J_{n5} = o_p((nh^{d/2})^{-1})$.

Proof.

$$E|J_{n5}| = E\left|\frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \hat{u}_t \hat{f}_{w_t}(r_s - \hat{r}_s)\hat{f}_{w_s}\right|$$

$$\leqslant \frac{1}{2n(n-1)h^d} \sum_t \sum_{s \neq t} E\{(r_s - \hat{r}_s)^2 \hat{f}^2_{w_s} K_{ts} + \hat{u}_t^2 \hat{f}^2_{w_t} K_{ts}\}$$

$$= (1/2)n^{-1} \sum_t E\{(r_t - \hat{r}_t)^2 \hat{f}^2_{w_t} \hat{f}_t + \hat{u}_t^2 \hat{f}^2_{w_t} \hat{f}_t\}$$

$$\equiv J_{1b} + J_{3b} = o((nh^{d/2})^{-1}),$$

by the proofs of Lemmas A.1 and A.3. Hence, $J_{n5} = o_p((nh^{d/2})^{-1})$.

Lemma A.6. $J_{n6} = o_p((nh^{d/2})^{-1})$.

Proof.

$$J_{n6} = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} u_t \hat{f}_{w_t} \hat{u}_s \hat{f}_{w_s} K_{ts}$$

$$= \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} u_t f_{w_t} \hat{u}_s \hat{f}_{w_s} K_{ts} + \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} u_t(\hat{f}_{w_t} - f_{w_t})\hat{u}_s \hat{f}_{w_s} K_{ts}$$

$$\equiv J_{n61} + J_{n62}, \text{ say.}$$

$J_{n61} = o_p((nh^{d/2})^{-1})$ by Lemma C.5(iii).

$\mathrm{E}|J_{n62}| \leqslant 1/2n(n-1)h^d \sum_t \sum_{s \neq t} \mathrm{E}\{u_t^2(\hat{f}_{w_t} - f_{w_t})^2 K_{ts} + \hat{u}_s^2 \hat{f}_{w_s} K_{ts}\} = (1/2)n^{-1} \sum_t \mathrm{E} \{u_t^2(\hat{f}_{w_t} - f_{w_t})^2 \hat{f}_t + \hat{u}_t^2 \hat{f}_{w_t} \hat{f}_t\} = J_{2b} + J_{3b} = \mathrm{o}((nh^{d/2})^{-1})$ by the proofs of Lemmas A.2 and A.3. Hence, $J_{n62} = \mathrm{o_p}((nh^{d/2})^{-1})$.

Proof of Corollary 3.2. We will only provide a proof for Corollary 3.2(i) since the proof of Corollary 3.2(ii) is much easier than the proof of Corollary 3.2(i).

From Eq. (8) and given the result of Theorem 3.1(i), it suffices to show that $(n^{(3)}/n^{(4)}) V_{1n}^a = \mathrm{o_p}((nh^{d/2})^{-1})$ and $(n^{(3)}/n^{(4)}) V_{2n}^a = \mathrm{o_p}((nh^{d/2})^{-1})$.

$$\frac{n^{(3)}}{n^{(4)}} V_{1n}^a = \frac{1}{n^{(4)} a^{2q} h^d} \sum \sum \sum_{t \neq s \neq l} (Y_t - Y_l)(Y_s - Y_l) L_{tl} L_{sl} K_{ts}$$

$$= \frac{1}{n^{(4)} a^{2q} h^d} \sum \sum \sum_{t \neq s \neq l} [(r_t - r_l) + u_t - u_l] \; [(r_s - r_l) + u_s - u_l] L_{tl} L_{sl} K_{ts}. \tag{A.2}$$

First we consider the term on the right-hand side of Eq. (A.2) that does not have an error term u. We use $V_{1n,1}^a$ to denote it. $V_{1n,1}^a \equiv (1/n^{(4)} a^{2q} h^d) \sum\sum\sum_{t \neq s \neq l} (r_t - r_l)(r_s - r_l) L_{tl} L_{sl} K_{ts}$. $\mathrm{E}|V_{1n,1}^a| = n^{-4} \mathrm{O}(n^3 a^{2\nu}) = \mathrm{O}(a^{2\nu}) = \mathrm{o}((nh^{d/2})^{-1})$ by assumptions (A1) (ii) and (A2). Hence, $V_{1n,1}^a = \mathrm{O_p}(a^{2\nu}) = \mathrm{o_p}((nh^{d/2})^{-1})$.

Next, we consider the terms with one error term u. One such term is $V_{1n,2}^a \equiv (1/n^{(4)} a^{2q} h^d) \sum\sum\sum_{t \neq s \neq l} (r_t - r_l) u_s L_{tl} L_{sl} K_{ts}$. $V_{1n,2}^a = \mathrm{o_p}((nh^{d/2})^{-1})$ by Lemma C.5(i). Similar arguments show that all the terms with one error term u is of the order of $\mathrm{o_p}((nh^{d/2})^{-1})$.

Finally, we consider the terms with two error terms. Say $V_{1n,3}^a \equiv (1/n^{(4)} a^{2q} h^d) \sum\sum\sum_{t \neq s \neq l} u_t u_s L_{tl} L_{sl} K_{ts}$. $V_{1n,3}^a = \mathrm{o_p}((nh^{d/2})^{-1})$ by Lemma C.5(iii). By the same reasoning one can show that all the other terms (with two error terms) are of the order of $\mathrm{o_p}((nh^{d/2})^{-1})$.

Summarizing the above we have shown that $(n^{(3)}/n^{(4)}) V_{1n}^a = \mathrm{o_p}((nh^{d/2})^{-1})$. Similarly one can show that $(n^{(3)}/n^{(4)}) V_{2n}^a = \mathrm{o_p}((nh^{d/2})^{-1})$. Therefore, $nh^{d/2} V_n^a / \hat{\sigma}_a = nh^{d/2} \hat{J}_n^a / \hat{\sigma}_a + \mathrm{o_p}(1) \to \mathrm{N}(0,1)$ under H_0^a by Theorem 3.1(i). This finishes the proof of Corollary 3.2(i).

Appendix B. Proof of Theorem 5.1

Proof of Theorem 5.1. We will only prove Theorem 5.1(i) since the proof of Theorem 5.1(ii) is similar to, and in fact much simpler than, the proof of Theorem 5.1(i).

We will first prove that $nh^{d/2}J^c_{in} \to \mathrm{N}(0,(\Omega_c)_{ii})$ in distribution.

Let $r_{(i),t}$ denote the ith component of r_t $(i = 1,\dots,p)$ and define the following short-hand notations: $g_{1,t} \equiv \mathrm{E}[r_{m,t+1}|X_t]$, $g_{2,t} \equiv \mathrm{E}[r^2_{m,t+1}|X_t]$, $\widehat{r^2_{m,t}} \equiv \hat{\mathrm{E}}[r^2_{m,t+1}|X_t]$ and $\hat{r}_{m,t} \equiv \hat{\mathrm{E}}[r_{m,t+1}|X_t]$ (see Eqs. (22) and (23)). Then we have

$$\tilde{U}_{it} = (\widehat{r^2_{m,t}} - \hat{r}_{m,t}r_{m,t+1})r_{(i),t}$$

$$= (g_{2,t} - g_{1,t}r_{m,t+1})r_{(i),t} + [(\widehat{r^2_{m,t}} - g_{2,t}) - (\hat{r}_{m,t} - g_{1,t})r_{m,t+1}]r_{(i),t}$$

$$\equiv U_{it} + [(\widehat{r^2_{m,t}} - g_{2,t}) - (\hat{r}_{m,t} - g_{1,t})r_{m,t+1}]r_{(i),t}, \tag{B.1}$$

where $U_{it} = (g_{2,t} - g_{1,t}r_{m,t+1})r_{(i),t}$.

Substituting Eq. (B.1) into Eq. (26) we get

$$\hat{J}^c_{ni} = \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\{U_{it}U_{is} + 2U_{it}[(\widehat{r^2_{m,s}} - g_{2,s}) - (\hat{r}_{m,s} - g_{1,s})r_{m,s+1}]r_{(i),s}$$

$$+ [(\widehat{r^2_{m,t}} - g_{2,t}) - (\hat{r}_{m,t} - g_{1,t})r_{m,t+1}]r_{(i),t}\ \ [(\widehat{r^2_{m,s}} - g_{2,s})$$

$$- (\hat{r}_{m,s} - g_{1,s})r_{m,s+1}]r_{(i),s}\}\tilde{f}_t\tilde{f}_sK_{ts} \equiv J^c_{ni,1} + 2J^c_{ni,2} + J^c_{ni,3}, \tag{B.2}$$

where $J^c_{ni,1} = (1/n(n-1)h^d)\sum_t\sum_{s\neq t}U_{it}U_{is}\tilde{f}_t\tilde{f}_sK_{ts}$ and the definitions of $J^c_{ni,2}$ and $J_{ni,3}$ should be apparent.

First for $J^c_{ni,1}$. Comparing $J^c_{ni,1}$ with the J_{n2} term of Lemma A.2, we immediately know that $nh^{d/2}J^c_{ni,1} \to \mathrm{N}(0,(\Omega_c)_{ii})$ in distribution (by the same proof of Lemma A.2 (i)), where $(\Omega_c)_{ii}$ is the ith diagonal element of Ω_c as defined in Theorem 5.1.

Next, to evaluate the order of $J^c_{ni,2}$. Define $v_{1,t} = r_{m,t+1} - \mathrm{E}[r_{m,t+1}|X_t] \equiv r_{m,t+1} - g_{1,t}$ and $v_{2,t} = r^2_{m,t+1} - \mathrm{E}[r^2_{m,t+1}|X_t] \equiv r^2_{m,t+1} - g_{2,t}$. Also define $\hat{g}_{1,t} = (na^d)^{-1}\sum_{s\neq t}g_{1,s}L_{ts}/\tilde{f}_t$, $\hat{g}_{2,t} = (na^d)^{-1}\sum_{s\neq t}g_{2,s}L_{ts}/\tilde{f}_t$, $\hat{v}_{1,t} = (na^d)^{-1}\sum_{s\neq t}v_{1,s}L_{ts}/\tilde{f}_t$, and $\hat{v}_{2,t} = (na^d)^{-1}\sum_{s\neq t}v_{2,s}L_{ts}/\tilde{f}_t$.

Then obviously we have

$$\hat{r}_{m,t} = \hat{g}_{1,t} + \hat{v}_{1,t} \quad \text{and} \quad \widehat{r^2_{m,t}} = \hat{g}_{2,t} + \hat{v}_{2,t}. \tag{B.3}$$

Using Eq. (B.3), we have

$$J^c_{ni,2} = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} U_{it}\{(\hat{g}_{2,s} - g_{2,s}) + \hat{v}_{2,s} - [(\hat{g}_{1,s} - g_{1,s})$$

$$+ \hat{v}_{1,s}]r_{m,s+1}]\}r_{(i),s}\tilde{f}_t\tilde{f}_s$$

$$= \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} U_{it}\{(\hat{g}_{2,s} - g_{2,s})$$

$$+ \hat{v}_{2,s} - (\hat{g}_{1,s} - g_{1,s})r_{m,s+1} - \hat{v}_{1,s}r_{m,s+1}\}r_{(i),s}\tilde{f}_t\tilde{f}_s$$

$$\equiv D_{1n} + D_{2n} - D_{3n} - D_{4n}.$$

$$D_{1n} = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} U_{it}(\hat{g}_{2,s} - g_{2,s})\tilde{f}_t\tilde{f}_s$$

$$= O_p(a^{2v} + (na^d)^{-1}) = o_p((nh^{d/2})^{-1})$$

by the same proof as Lemma A.4. $D_{2n} = (1/n(n-1)h^d)\sum_t\sum_{s \neq t} U_{it}\hat{v}_{2,s}\tilde{f}_t\tilde{f}_s = O_p(a^{2v} + (na^d)^{-1}) = o_p((nh^{d/2})^{-1})$ by the same proof as Lemma A.6. Similar arguments lead to $D_{3n} = o_p((nh^{d/2})^{-1})$ and $D_{4n} = o_p((nh^{d/2})^{-1})$.

Thus, $J^c_{ni,2} = o_p((nh^{d/2})^{-1})$.

Similarly for $J^c_{ni,3}$, using Eq. (B.3) we have

$$J^c_{ni,3} = \frac{1}{n(n-1)h^d} \sum_t \sum_{s \neq t} \{[(\hat{g}_{2,t} - g_{2,t}) + \hat{v}_{2,t} - (\hat{g}_{1,t} - g_{1,t})r_{m,t+1}$$

$$- \hat{v}_{1,t}r_{m,t+1}][(\hat{g}_{2,s} - g_{2,s}) + \hat{v}_{2,s} - (\hat{g}_{1,s} - g_{1,s})r_{m,s+1}$$

$$- \hat{v}_{1,s}r_{m,s+1}]\}r_{(i),t}r_{(i),s}\tilde{f}_t\tilde{f}_sK_{ts}$$

$$= O_p(a^{2v} + (na^d)^{-1}) = o_p((nh^{d/2})^{-1})$$

by the same proofs as in Lemmas A.1, A.3 and A.5.

Summarizing the above, we have proved that

$$nh^{d/2}J^c_{ni} = nh^{d/2}J^c_{ni,1} + o_p(1) \to N(0, (\Omega_c)_{ii}) \quad \text{in distribution.}$$

Next, let $(\bar{U}_{it}, \bar{X}_t)$ denote an independent process that has the same marginal distribution as (U_{it}, X_t). Also denote $\bar{K}_{ts} = K((\bar{X}_t - \bar{X}_s)/h)$. Then follow the same arguments as in the proof of Theorem 2.1 of Fan and Li (1996b), it is straightforward to show that the covariance between $nh^{d/2}\hat{J}^c_{ni}$ and $nh^{d/2}\hat{J}^c_{nj}$ is

$$cov(nh^{d/2}\hat{J}^c_{ni}, nh^{d/2}\hat{J}^c_{nj}) = \frac{2}{(n-1)^2h^d}\sum_t\sum_{s\neq t} \mathrm{E}[U_{it}U_{jt}U_{is}U_{js}K^2_{ts}] + o(1)$$

$$= \frac{2}{(n-1)^2h^d}\sum_t\sum_{s\neq t} \mathrm{E}[\bar{U}_{it}\bar{U}_{jt}\bar{U}_{is}\bar{U}_{js}\bar{K}^2_{ts}] + o(1)$$

$$= 2\mathrm{E}[\bar{U}_{i1}\bar{U}_{j1}\bar{U}_{i2}\bar{U}_{j2}\bar{K}^2_{12}] + o(1)$$

$$= 2\mathrm{E}[\sigma^2_{ij}(\bar{X}_1)\sigma^2_{ij}(\bar{X}_2)\bar{K}^2_{12}] + o(1)$$

$$= 2\mathrm{E}[\sigma^4_{ij}(X)f(X)]\left[\int K^2(u)\,\mathrm{d}u\right] + o(1) = (\Omega_c)_{ij} + o(1),$$

where $\sigma^2_{ij}(x) = \mathrm{E}[U_{it}U_{jt}|X_t]$.

By the Cramer–Wold device, one can show that for any $c \in R^p$ with $\|c\| = 1$ (here $\|.\|$ is the Euclidean norm), $c'[nh^{d/2}\hat{J}^c_n] \to \mathrm{N}(0, c'\Omega c)$. Therefore, we obtain the desired result that

$$nh^{d/2}\hat{J}^c_n \to \mathrm{N}(0, \Omega_c) \quad \text{in distribution.}$$

Finally $\hat{\Omega}_c - \Omega_c = o_p(1)$ follows similar arguments as in the proof of Lemma A.2(ii). This finishes the proof of Theorem 5.1(i).

Appendix C. Some useful lemmas

This appendix presents some useful lemmas. Throughout this appendix. We will use the tilde notation to denote independent process. For example, $\{\tilde{X}_t\}_{t=1}^n$ is an i.i.d. sequence having the same marginal distribution as $\{X_t\}$. We will use the shorthand notation: $\tilde{r}_t = r(\tilde{W}_t)$ and $\tilde{K}_{t,s} = K((\tilde{X}_t - \tilde{X}_s)/h)$, etc. Also $\mathrm{E}_{t_1}[A(X_{t_1}, X_{t_2})] \equiv \int A(X_{t_1}, x)\,\mathrm{d}F(x)$ and $\mathrm{E}_{t_1,t_2}[B(X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4})] \equiv \iint A(X_{t_1}, X_{t_2}, x, y)\,\mathrm{d}F_{|t_4 - t_3|}(x, y)$, where $F(\cdot)$ is the marginal distribution function for X_t and $F_\tau(.,.)$ is the joint distribution function for $(X_t, X_{t+\tau})$. Like in appendix A, we often use u_t for U_t and w_t for W_t to save space. These should not cause any confusions because variables with subscripts always mean random variables even small letter case is used.

Lemma C.1. Let $\xi_1, \dots, \xi_n$ be random vectors taking values in R^p satisfying an absolute regularity (i.e., β-mixing) condition and denote by β_τ the mixing coefficient (see Appendix A for the definition of β_τ). Let $h(x_1, \dots, x_k)$ be a Borel measurable function such that for some $\delta > 0$,

$$M = \max\left\{ \int_{R^{kp}} |h(x_1, \dots, x_k)|^{1+\delta} \mathrm{d}F(x_1, \dots, x_k), \int\int_{R^{kp}} |h(x_1, \dots, x_k)|^{1+\delta} \mathrm{d}F^{(1)} \right.$$

$$\left. \times (x_1, \dots, x_j) \, \mathrm{d}F^{(2)}(x_{j+1}, \dots, x_k) \right\}$$

exists. Then

$$\left| \int_{R^{kp}} h(x_1, \dots, x_k) \, \mathrm{d}F(x_1, \dots, x_k) - \int_{R^{kp}} h(x_1, \dots, x_k) \, \mathrm{d}F^{(1)}(x_1, \dots, x_j) \, \mathrm{d}F^{(2)} \right.$$

$$\left. (x_{j+1}, \dots, x_k) \right| \leqslant 4M^{1/(1+\delta)} \beta_\tau^{\delta/(1+\delta)},$$

where $\tau = i_{j+1} - i_j$, F, $F^{(1)}$, and $F^{(2)}$ are distribution functions of random vectors $(\xi_{i_1}, \dots, \xi_{i_k})$, $(\xi_{i_1}, \dots, \xi_{i_j})$, and $(\xi_{i_{j+1}}, \dots, \xi_{i_k})$, respectively, and $i_1 < i_2 < \cdots < i_k$.

Proof. This is Lemma 1 in Yoshihara (1976).

Lemma C.2. Let $r(\cdot) \in \mathscr{G}_\alpha^v$, $f_w \in \mathscr{G}_v^\alpha$ and $L(\cdot) \in \mathscr{K}_v$, where $v \geqslant 2$ is an integer. $w \in R^q$, $a \to 0$ as $n \to \infty$. Then

(i) $|\mathrm{E}[L((W-w)/a) - a^q f_w(w)| \leqslant a^{q+v} D_f(w)$, *uniformly in* w,
(ii) $|\mathrm{E}\{[r(W) - r(w)] L((W-w)/a)\}| \leqslant a^{q+v} D_r(w)$, *uniformly in* w,

where both $D_f(\cdot)$ and $D_r(\cdot)$ have finite αth moments.

Proof. (i) and (ii) were proved in Lemmas 4 and 5 of Robinson (1988), respectively.

Lemma C.3. (i) $n^{-1}\sum_t \mathrm{E}[(\hat{r}_t - r_t)^2 \hat{f}_{w_t}^2] = \mathrm{O}((na^q)^{-1} + a^{2v})$.
(ii) $n^{-1}\sum_t \mathrm{E}[(\hat{f}_{w_t} - f_{w_t})^2 \eta_t^2] = \mathrm{O}((na^q)^{-1}) + a^{2v})$, *where* $\eta_t = 1$, *or* $\eta_t = u_t$ *or* $\eta_t = u_t^2$.
(iii) $n^{-1}\sum_t \mathrm{E}[\hat{u}_t^2 \hat{f}_{w_t}^2] = \mathrm{O}((na^q)^{-1})$.

Proof. (i) is proved in the proof of Lemma A.2 of Fan and Li (1996c). Intuitively this result is easy to understand. It says that the average mean square error (MSE) of $(\hat{r}_t - r_t)\hat{f}_{w_t}$ is $O((na^q)^{-1} + a^{2\nu})$. While this is a standard result with independent observations, one can show that the same average MSE convergence rate holds for weakly dependent data.

For (ii), the case $\eta_t = u_t$ is proved in the proof of Lemma A.1 of Fan and Li (1996c). By *exactly* the same proof of Lemma A.1 of Fan and Li (1996c), one can show that (ii) holds when $\eta_t = 1$ or $\eta_t = u_t^2$.

Finally (iii) is proved in the proof of Lemma A.4(i) of Fan and Li (1996c).

Lemma C.4. (i) $A_{1n} \stackrel{\text{def}}{=} n^{-1}\sum_t \mathrm{E}[(r_t - \hat{r}_t)^2 \hat{f}^2_{w_t}(\hat{f}_t - f_t)] = o(a^{2\nu} + (na^q)^{-1}) = o((nh^{d/2})^{-1})$.

(ii) $A_{2n} \stackrel{\text{def}}{=} n^{-1}\sum_t \mathrm{E}[u_t^2(f_{w_t} - \hat{f}_{w_t})^2(\hat{f}_t - f_t)] = o(a^{2\nu} + (na^q)^{-1}) = o((nh^{d/2})^{-1})$.

(iii) $A_{3n} \stackrel{\text{def}}{=} n^{-2}\sum_t \mathrm{E}[\hat{u}_t^2 \hat{f}^2_{w_t}(\hat{f}_t - f_t)] = o(a^{2\nu} + (na^q)^{-1}) = o((nh^{d/2})^{-1})$.

Proof of (i). Let $m = [b\log(n)]$ (the integer part of $b\log(n)$) and b is a large positive constant so that $n^8\beta_m^{\delta/(1+\delta)} = o(1)$ by (A1)(i).

Using $\hat{r}_t\hat{f}_{w_t} = (1/(n-1)a^q)\sum_{i\neq t} r_i L_{it}$, $\hat{f}_{w_t} = (1/(n-1)a^q)\sum_{i\neq t} L_{it}$ and $\hat{f}_t = (1/(n-1)h^d)\sum_{s\neq t} K_{ts}$, we have

$$A_{1n} = (n^3a^{2q})^{-1}\sum_t \sum_{i\neq t} \sum_{j\neq t} \mathrm{E}\{(r_t - r_i)L_{it}(r_t - r_j)L_{jt}[n^{-1}\sum_{s\neq t}(h^{-d}K_{ts} - f_t)]\}.$$

We consider two different cases for A_{1n}: (a) $\min\{|s-t|, |s-i|, |s-j|\} > m$ and (b) $\min\{|s-t|, |s-i|, |s-j|\} \leqslant m$. We use $A_{1n(a)}$ and $A_{1n(b)}$ to denote these two cases. For case (a), denote $K_{x,t} = K((x - X_t)/h)$ and use Lemma C.1, we have

$$\begin{aligned} A_{1n(a)} &\leqslant \frac{1}{n^3a^{2q}} \sum_t \sum_{i\neq t} \sum_{j\neq t} \left| \mathrm{E}\left\{[(r_t - r_i)L_{it}(r_t - r_j)L_{jt}]n^{-1} \right.\right. \\ &\quad \left.\left. \times \sum_{s\neq t} \int (h^{-d}K_{x,t} - f_t)\,\mathrm{d}F(x)\right\}\right| + 4(a^{2q}h^d)^{-1}M_n^{1/(1+\delta)}\beta_m^{\delta/(1+\delta)} \\ &\leqslant Ch^2n^{-1}\sum_t \mathrm{E}\{(r_t - \hat{r}_t)^2\hat{f}^2_{w_t}\} + C(a^{2q}h^d)^{-1}\beta_m^{\delta/(1+\delta)} \\ &= O(h^2)O_p(a^{2\nu} + (na^q)^{-1}) + O(n^3\beta_m^{\delta/(1+\delta)}), \end{aligned}$$

where we used facts that $\int (h^{-d}K_{x,t} - f_t)\,dF(x)\} = O(h^2)$ and $M_n \sim \max_{i \neq t, j \neq t, s \neq t} E[|(r_t - r_i)L_{it}(r_t - r_j)L_{jt}K_{ts}|^{1+\delta} \leqslant C\max_{i \neq t} E[(r_t - r_i)^{2(1+\delta)}] = O(1)$ (M_n is the bound function as defined in Lemma C.1).

Next, for case (b), without loss of generality, we assume $|s - t| \leqslant m$. Hence, for any t, $n^{-1}\sum_{|s-t| \leqslant m}(h^{-d}K_{ts} - f_t) \leqslant Cn^{-1}mh^{-d} = O(m(nh^d)^{-1})$. Thus, using Lemma C.1, we have $A_{1n(b)} \leqslant Cm(nh^d)^{-1}n^{-1}\sum_t E[(r_t - \hat{r}_t)^2 \hat{f}^2_{w_t}] = O(m(nh^d)^{-1})O(a^{2v} + (na^q)^{-1})$ by Lemma C.3(i).

Hence, $A_{1n} = O((h^2 + m(nh^d)^{-1})(a^{2v} + (na^q)^{-1})) + O(n^3\beta_m^{\delta/(1+\delta)}) = o(a^{2v} + (na^q)^{-1})$.

Proof of (ii). The proof of (ii) follows the same steps as the proof of (i) above except that we need to cite Lemma C.3(ii) instead of Lemma C.3(i) in the proof.

Proof of (iii). The proof of (iii) is *exactly* the same as (i) above except that we need to cite Lemma C.3(iii) instead of Lemma C.3(i) in the proof.

Lemma C.5. (i) $B_{1n} \overset{\text{def}}{=} (1/n(n-1)h^d)\sum_t\sum_{s \neq t} u_t f_{w_t}(r_s - \hat{r}_s)\hat{f}_{w_s}K_{ts} = o_p((nh^{d/2})^{-1})$.
(ii) $B_{2n} \overset{\text{def}}{=} (1/n(n-1)h^d)\sum_t\sum_{s \neq t} u_t u_s(\hat{f}_{w_t} - f_{w_t})f_{w_s}K_{ts} = o_p((nh^{d/2})^{-1})$.
(iii) $B_{3n} \overset{\text{def}}{=} (1/n(n-1)h^d)\sum_t\sum_{s \neq t} u_t f_{w_t}\hat{u}_s\hat{f}_{w_s}K_{ts} = o_p((nh^{d/2})^{-1})$.

In the proofs below $m = [b\log(n)]$ as defined in the proof of Lemma C.4(i).

Proof of (i). Writing B_{1n} as $(n^3h^da^q)^{-1}\sum\sum\sum_{t_1, t_2 \neq t_1, t_3 \neq t_1} u_{t_1}f_{w_{t_1}}(r_{t_2} - r_{t_3})L_{t_2,t_3}K_{t_1,t_2}$, its second moment is

$$EB_1 \overset{\text{def}}{=} E[B_{1n}^2] = (n^3h^da^q)^{-2}\sum\sum\sum_{t_1, t_2 \neq t_1, t_3 \neq t_1}\sum\sum\sum_{t_4, t_5 \neq t_4, t_6 \neq t_4}$$

$$\times E[u_{t_1}f_{w_{t_1}}(r_{t_2} - r_{t_3})L_{t_2,t_3}K_{t_1,t_2}u_{t_4}f_{w_{t_4}}(r_{t_5} - r_{t_6})L_{t_5,t_6}K_{t_4,t_5}]$$

We consider four different cases: (a) for all i's, $|t_i - t_j| > m$ for all $j \neq i$; (b) for exactly four different i's, $|t_i - t_j| > m$ for all $j \neq i$; (c) for exactly three different i's, $|t_i - t_j| > m$ for all $j \neq i$; (d) all the other remaining cases. We will use $EB_{1(s)}$ to denote these cases ($s = a, b, c, d$).

Using Lemma C.1, we have

$$EB_{1(a)} \leqslant 0 + Cn^6\beta_m^{\delta/(1+\delta)} = O(n^6\beta_m^{\delta/(1+\delta)}) = O(n^{-2}) = o((n^2h^d)^{-1}).$$

For case (b), we only need to consider the case $|t_1 - t_4| \leqslant m$, since otherwise we will have t_1 or t_4 is at least m periods away from any other indices and by Lemma C.1, we know it is bounded by $O(n^6 \beta_m^{\delta/(1+\delta)})$. For case (b) with $|t_1 - t_4| \leqslant m$, we must have t_i at least m periods away from any other indices for $i = 2, 3, 5, 6$. Hence, use Lemma C.1 four times and let t_{-i} denote all t_j's with $j \neq i$, we get (recall the tilde notation is for independent random variables),

$$
\begin{aligned}
\mathrm{E}B_{1(b)} &\leqslant (n^3 h^d a^q)^{-2} \sum\sum\sum\sum\sum \sum_{|t_i - t_{-i}| > m, i = 2,3,5,6, |t_1 - t_4| \leqslant m} \mathrm{E}\{u_{t_1} u_{t_4} f_{w_{t_1}} f_{w_{t_4}} \\
&\quad \times \mathrm{E}_{t_1}[(\tilde{r}_{t_2} - \tilde{r}_{t_3}) \tilde{L}_{t_2,t_3} \tilde{K}_{t_1,t_2}] \mathrm{E}_{t_4}[(\tilde{r}_{t_5} - \tilde{r}_{t_6}) \tilde{L}_{t_5,t_6} \tilde{K}_{t_4,t_5}]\} \\
&\quad + C n^6 \beta_m^{\delta/(1+\delta)} \leqslant C n^2 a^{2(q+v)} (n^3 h^d a^q)^{-2} \sum \sum_{|t_4 - t_1| \leqslant m} \sum \sum_{t_2 \neq t_5} \\
&\quad \times \mathrm{E}\{|u_{t_1} u_{t_4}| \mathrm{E}_{t_1}[D_r(w_{t_2}) K_{t_1,t_2}] \mathrm{E}_{t_4}[D_r(w_{t_5}) K_{t_4,t_5}]\} \\
&\quad + C n^6 \beta_m^{\delta/(1+\delta)} \leqslant C n^4 a^{2(q+v)} h^{2d} (n^3 h^d a^q)^{-2} \sum \sum_{1 < |t_4 - t_1| \leqslant m} \sum \sum_{t_2 \neq t_5} \\
&\quad \times \mathrm{E}\{|u_{t_1} u_{t_4}| D_r(w_{t_1}) D_r(w_{t_4})\} \\
&\quad + C n^6 \beta_m^{\delta/(1+\delta)} \leqslant O(n^5 m a^{2(q+v)} h^{2d} (n^3 h^d a^q)^{-2}) \\
&\quad + C n^6 \beta_m^{\delta/(1+\delta)} = O(m a^{2v} n^{-1}) + C n^6 \beta_m^{\delta/(1+\delta)} = o((n^2 h^d)^{-1})
\end{aligned}
$$

by Lemma C.2(ii).

For case (c), we only need to consider $|t_1 - t_4| \leqslant m$, $|t_i - t_1| \leqslant m$ (or $|t_i - t_4| \leqslant m$) for exactly one $i \in \{2, 3, 4, 5\}$ since otherwise it will be bounded by $O(n^6 \beta_m^{\delta/(1+\delta)})$ by Lemma C.1. By symmetry we only need to consider $i = 2$ and $i = 3$. First for $i = 2$, using Lemma C.1 three times, we have

$$
\begin{aligned}
\mathrm{E}B_{1(c)} &\leqslant (n^3 h^d a^q)^{-2} \sum\sum\sum\sum\sum \sum_{case\ (c)} \mathrm{E}\{u_{t_1} u_{t_4} f_{w_{t_1}} f_{w_{t_4}} K_{t_1,t_2} \\
&\quad \times \mathrm{E}_{t_2}[(\tilde{r}_{t_2} - \tilde{r}_{t_3}) \tilde{L}_{t_2,t_3}] \mathrm{E}_{t_4}[(\tilde{r}_{t_5} - \tilde{r}_{t_6}) \tilde{L}_{t_5,t_6} \tilde{K}_{t_4,t_5}]\} + C n^6 \beta_m^{\delta/(1+\delta)}
\end{aligned}
$$

$$\leqslant Cn^2(n^3h^da^q)^{-2}\sum\sum\sum_{|t_1-t_4|\leqslant m,|t_2-t_1|\leqslant m}\sum_{t_5\neq t_4} \mathrm{E}\{|u_{t_1}u_{t_4}|K_{t_1,t_2}D_r(w_{t_2})\mathrm{E}_{t_4}$$

$$\times [D_r(w_{t_5})K_{t_4,t_5}]\}a^{2(q+v)} + ,Cn^6\beta_m^{\delta/(1+\delta)}$$

$$\leqslant Cn^3(n^3h^da^q)^{-2}\sum\sum\sum_{|t_1-t_4|\leqslant m,|t_2-t_1|\leqslant m}$$

$$\times \mathrm{E}\{|u_{t_1}u_{t_4}|D_r(w_{t_2})D_r(w_{t_4})]\}a^{2(q+v)}h^d + Cn^6\beta_m^{\delta/(1+\delta)}$$

$$= \mathrm{O}(n^4m^2a^{2(q+v)}h^d(n^3h^da^q)^{-2}) + Cn^6\beta_m^{\delta/(1+\delta)}$$

$$= (n^2h^d)^{-1}\mathrm{O}(m^2a^{2v}) + Cn^6\beta_m^{\delta/(1+\delta)} = \mathrm{o}((n^2h^d)^{-1})$$

by Lemma C.2(ii).

Similarly for $i = 3$, using Lemma C.1 three times, we get

$$\mathrm{E}B_{1(c)} \leqslant (n^3h^da^q)^{-2}\sum\sum\sum\sum\sum\sum_{case\ (c)} \mathrm{E}\{u_{t_1}u_{t_4}f_{w_{t_1}}f_{w_{t_4}}$$

$$\times \mathrm{E}_{t_1,t_3}[(r_{t_2}-r_{t_3})L_{t_2,t_3}K_{t_1,t_2}]\mathrm{E}_{t_4}[(\tilde{r}_{t_5}-\tilde{r}_{t_6})\tilde{L}_{t_5,t_6}\tilde{K}_{t_4,t_5}]\} + Cn^6\beta_m^{\delta/(1+\delta)}$$

$$\leqslant Cn^2(n^3h^da^q)^{-2}\sum\sum\sum_{|t_4-t_1|\leqslant m,|t_3-t_1|\leqslant m}\sum_{t_5\neq t_4} \mathrm{E}\{|u_{t_1}u_{t_4}|(|r_{t_1}|$$

$$+ |r_{t_3}|)\mathrm{E}_{t_4}[D_r(w_{t_5})K_{t_4,t_5}]\}a^{(q+v)}h^d + Cn^6\beta_m^{\delta/(1+\delta)}$$

$$\leqslant Cn^2(n^3h^da^q)^{-2}\sum\sum\sum_{|t_4-t_1|\leqslant m,|t_3-t_1|\leqslant m} \mathrm{E}\{|u_{t_1}u_{t_4}|(|r_{t_1}|$$

$$+ |r_{t_3}|)D_r(w_{t_5})]\}a^{(q+v)}h^{2d} + Cn^6\beta_m^{\delta/(1+\delta)}$$

$$= (n^2a^q)^{-1}\mathrm{O}(m^2a^v) + Cn^6\beta_m^{\delta/(1+\delta)} = \mathrm{o}((n^2a^q)^{-1}) = \mathrm{o}((n^2h^d)^{-1})$$

by Lemma C.2(ii).

Note that case (d) has at most n^3m^3 terms, then using Lemma C.1, it is straightforward to show that

$$\mathrm{E}B_{1(d)} \leqslant Cm^3n^3(n^3h^da^q)^{-2}[\mathrm{O}(a^{q+v}h^d + h^{2d}) + \mathrm{O}(n^6\beta_m^{\delta/(1+\delta)})] = \mathrm{o}((n^2h^d)^{-1}).$$

Proof of (ii). $B_{2n} = (n^3h^da^q)^{-1}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_2}u_{t_1}u_{t_2}(L_{t_1,t_3} - a^qf_{t_1})K_{t_1,t_2}$, its second moment is

$$\mathrm{E}B_2 \stackrel{\mathrm{def}}{=} \mathrm{E}[B_{2n}^2] = (n^3h^da^q)^{-2}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_2}\sum\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4}$$

$$\mathrm{E}\{u_{t_1}u_{t_2}(L_{t_1,t_3} - a^qf_{t_1})K_{t_1,t_2}u_{t_4}u_{t_5}(L_{t_4,t_6} - a^qf_{t_4})K_{t_4,t_5}\}$$

We consider three different cases for $\mathrm{E}B_2$. (a) for at least three different i's, $|t_i - t_j| > m$ for all $j \neq i$; (b) for exactly two different i's, $|t_i - t_j| > m$ for all $j \neq i$; and (c) all the remaining cases.

Note that for case (a) $|t_i - t_j| > m$ for all $j \neq i$ holds for at least one $i \in \{1, 2, 4, 5\}$, hence using Lemma C.1, we have $\mathrm{E}B_{2(a)} \leqslant 0 + Cn^6\beta_m^{\delta/(1+\delta)} = \mathrm{O}(n^6\beta_m^{\delta/(1+\delta)})$.

For case (b) we only need to consider $|t_3 - t_i| > m$ for all $i \neq 3$ and $|t_6 - t_j| > m$ for all $j \neq 6$ since otherwise $\mathrm{E}B_{2(b)}$ will be bounded by $\mathrm{O}(n^6\beta_m^{\delta/(1+\delta)})$ by Lemma C.1. Case (b) have n^4m^2 terms and they correspond to either (i) $|t_1 - t_4| \leqslant m$ and $|t_2 - t_5| \leqslant m$, or (ii) $|t_1 - t_2| \leqslant m$ and $|t_4 - t_5| \leqslant m$. We use $\mathrm{E}B_{2(b(i))}$ and $\mathrm{E}B_{2(b(ii))}$ to denote these two subcases.

Using Lemma C.1 four times, we have

$$\mathrm{E}B_{2(b(i))} \leqslant (n^3h^da^q)^{-2}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_1}\sum\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4}$$

$$\times \mathrm{E}\{u_{t_1}u_{t_4}\mathrm{E}_{t_1,t_4}[u_{t_2}u_{t_5}K_{t_1,t_2}K_{t_4,t_5}][\mathrm{E}_{t_1}(L_{t_1,t_3} - a^qf_{t_1})]$$

$$\times [\mathrm{E}_{t_4}(L_{t_4,t_6} - a^qf_{t_4})]\} \leqslant Cn^4m^2(n^3h^da^q)^{-2}\mathrm{O}(a^{2(q+\nu)}h^d)$$

$$= \mathrm{O}(m^2a^{2\nu}(n^2h^d)^{-1}) = \mathrm{o}((n^2h^d)^{-1})$$

by Lemma C.2(i) and that $\mathrm{E}_{t_1,t_4}[u_{t_2}u_{t_5}K_{t_1,t_2}K_{t_4,t_5}] \leqslant \{\mathrm{E}[u_{t_2}^2u_{t_5}^2]\mathrm{E}_{t_1,t_4}[K_{t_1,t_2}^2K_{t_4,t_5}^2]\}^{1/2} = \mathrm{O}(h^d)$.

Similarly we have, by Lemma C.1,

$$\mathrm{E}B_{2(b(ii))} \leqslant (n^3h^da^q)^{-2}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_1}\sum\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4}$$

$$\times \mathrm{E}\{u_{t_1}u_{t_2}K_{t_1,t_2}[\mathrm{E}_{t_1}(L_{t_1,t_3} - a^qf_{t_1})]\}\mathrm{E}\{u_{t_4}u_{t_5}K_{t_4,t_5}[\mathrm{E}_{t_4}(L_{t_4,t_6} - a^qf_{t_4})]\}$$

$$\leqslant Cn^2(n^3h^da^q)^{-2}\sum\sum_{t_1,t_2\neq t_1}\sum\sum_{t_4,t_5\neq t_4}$$

$$\times \mathrm{E}[|u_{t_1}u_{t_2}|K_{t_1,t_2}D_f(w_{t_1})]\,\mathrm{E}[|u_{t_4}u_{t_5}|K_{t_4,t_5}D_f(w_{t_4})]a^{2(q+v)}$$

$$= Cn^4m^2(n^3h^da^q)^{-2}\mathrm{O}(a^{2(q+v)}h^q) = (n^2h^d)^{-1}\mathrm{O}(m^2a^{2v}) = \mathrm{o}((n^2h^d)^{-1})$$

by Lemma C.2(i) and the fact that $\mathrm{E}[u_{t_1}u_{t_2}K_{t_1,t_2}] \leqslant \{\mathrm{E}[u_{t_1}^2u_{t_2}^2]\,\mathrm{E}[K_{t_1,t_2}^2]\}^{1/2} = \mathrm{O}(h^{d/2})$.

Finally note that for case (c), it has at most n^3m^3 terms and using Lemma C.1, it is easy to show that

$$\mathrm{E}B_{2(c)} \leqslant Cn^3m^3(n^3h^da^q)^{-2}\mathrm{O}(a^{(q+v)}h^d + h^{2d})$$

$$= (n^2h^d)^{-1}m^3\mathrm{O}(a^v(nh^d)^{-1} + (na^{2q})^{-1}) = \mathrm{o}((n^2h^d)^{-1}).$$

Proof of (iii). $B_{3n} = (n^3h^da^q)^{-1}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_1}u_{t_1}f_{w_{t_1}}u_{t_2}L_{t_1,t_2}K_{t_1,t_3}$, its second moment is

$$\mathrm{E}B_3 \overset{\text{def}}{=} \mathrm{E}[B_{3n}^2] \leqslant (n^3h^da^q)^{-2}\Bigg\{\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_1}\sum\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4}$$

$$\times \mathrm{E}\{u_{t_1}u_{t_2}L_{t_1,t_2}K_{t_1,t_3}u_{t_4}u_{t_5}L_{t_4,t_5}K_{t_4,t_6}\Bigg\}.$$

The proof of $\mathrm{E}B_3 = \mathrm{o}((n^2h^d)^{-1})$ is very similar to the proof of $\mathrm{E}B_2 = \mathrm{o}((n^2h^d)^{-1})$. We consider three cases: (a) for at least three different i's, $|t_i - t_j| > m$ for all $j \neq i$; (b) for exactly two different i's, $|t_i - t_j| > m$ for all $j \neq i$; and (c) all the remaining cases.

By Lemma C.1, we have

$$\mathrm{E}B_{3(a)} \leqslant 0 + Cn^6\beta_m^{\delta/(1+\delta)} = \mathrm{O}(n^6\beta_m^{\delta/(1+\delta)}).$$

For case (b), we only need to consider for $i = 3$ and 6, $|t_i - t_j| > m$ for all $j \neq i$. Two subcases are, (i) $|t_1 - t_4| \leqslant m$ and $|t_2 - t_5| \leqslant m$, and (ii) $|t_1 - t_2| \leqslant m$ and $|t_4 - t_5| \leqslant m$. We use $\mathrm{E}B_{3(b(i))}$ and $\mathrm{E}B_{3(b(ii))}$ to denote these two subcases.

Using Lemma C.1 four times, we have

$$\mathrm{E}B_{3(b(i))} \leqslant (n^3h^da^q)^{-2}\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_1}\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4,\ sub-case(i)}\sum$$

$$\times\,\mathrm{E}\{u_{t_1}u_{t_4}\mathrm{E}_{t_1,t_4}[u_{t_2}u_{t_5}L_{t_1,t_2}L_{t_4,t_5}][\mathrm{E}_{t_1}K_{t_1,t_3}][\mathrm{E}_{t_4}K_{t_1,t_6}]\}$$

$$\leqslant Cn^4m^2(n^3h^da^q)^{-2}\mathrm{O}(h^{2d}) = \mathrm{O}((n^2h^d)^{-1}m^2h^d(na^{2q})^{-1})$$

$$= \mathrm{o}((n^2h^d)^{-1})$$

by the facts that $\mathrm{E}_{t_1}[K_{t_1,t_3}] = \mathrm{O}(h^d)$ and $\mathrm{E}_{t_1,t_4}[u_{t_2}u_{t_5}L_{t_1,t_2}L_{t_4,t_5}] \leqslant C\mathrm{E}[|u_{t_2}u_{t_5}|] = \mathrm{O}(1)$.

Similarly we have, by Lemma C.1,

$$\mathrm{E}B_{3(b(ii))} \leqslant (n^3h^da^q)^{-2}\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_1}\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4,\ sub-case(ii)}\sum$$

$$\times\,\mathrm{E}\{u_{t_1}u_{t_2}L_{t_1,t_2}\mathrm{E}_{t_1}[K_{t_1,t_3}]\}\mathrm{E}\{u_{t_4}u_{t_5}L_{t_4,t_5}\mathrm{E}_{t_4}[K_{t_4,t_6}]\}$$

$$\times\,Cn^4m^2(n^3h^da^q)^{-2}\mathrm{O}(h^{2d}a^q) = (n^2h^d)^{-1}\mathrm{O}(m^2h^da^{-q})$$

$$= \mathrm{o}((n^2h^d)^{-1})$$

by the facts that $\mathrm{E}_{t_1}[K_{t_1,t_3}] = \mathrm{O}(h^d)$ and $\mathrm{E}[|u_{t_4}u_{t_5}L_{t_4,t_5}|] \leqslant \{\mathrm{E}[u_{t_4}^2u_{t_5}^2]\,\mathrm{E}[L_{t_4,t_5}^2]\}^{1/2} = \mathrm{O}(a^{q/2})$.

Finally note that for case (c), it has at most n^3m^3 terms and using Lemma C.1, it is easy to show that

$$\mathrm{E}B_{3(c)} \leqslant \mathrm{C}n^3m^3(n^3h^da^q)^{-2}\mathrm{O}(a^{(q+v)}h^d + h^{2d})$$

$$= (n^2h^d)^{-1}m^3\mathrm{O}(a^v(nh^d)^{-1} + (na^{2q})^{-1}) = \mathrm{o}((n^2h^d)^{-1}).$$

References

Ait-Sahalia, Y., Bickel, P.J., Stoker, T.M., 1994. Goodness-of-fit tests for regression using kernel methods. Manuscript, University of Chicago.

Andrews, D.W.K., 1997. A conditional Kolmogorov test. Econometrica 65, 1097–1128.

Bierens, H.J, 1982. Consistent model specification tests. Journal of Econometrics 20, 105–134.

Bierens, H.J, 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443–1458.

Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129–1154.

Bollerslev, T, 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Chen, X., Fan, Y., 1997. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. forthcoming in Journal of Econometrics.

Christoffersen, P., Hahn, J., 1997. Nonparametric testing of ARCH for option pricing. Unpublished manuscript.

Cochrane, J, 1996. A cross-sectional test of an investment-based asset pricing model. Journal of Political Economy 104, 572–621.

DeJong, P, 1987. A central limit theorem for generalized quadratic forms. Probability Theory and Related Fields 75, 261–277.

Delgado, M.A, Stengos, T, 1994. Semiparametric testing of non-nested econometric models. Review of Economic Studies 75, 345–367.

Denker, M., Keller, G., 1983. On U-statistics and v. Mises statistics for weakly dependent processes. Zeitschrift Wahrscheinlichkeitstheorie verw, Gebiete 64, 505–522.

Engle, R.F, 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987–1008.

Engle, R.F, Granger, C.W.J, Rice, J, Weiss, A, 1986. Semiparametric estimation of the relation between weather and electricity sales. Journal of the American Statistical Association 81, 310–320.

Eubank, R, Hart, J, 1992. Testing goodness-of-fit in regression via order selection criteria. The Annals of Statistics 20, 1412–1425.

Eubank, R, Spiegelman, S, 1990. Testing the goodness of fit of a linear model via nonparametric regression techniques. Journal of the American Statistical Association 85, 387–392.

Fan, Y., Li, Q., 1992. The asymptotic expansion for the kernel sum of squared residuals and its applications in hypotheses testing. Manuscript, University of Windsor.

Fan, Y., Li, Q., 1996a. Consistent model specification tests: omitted variables, parametric and semiparametric functional forms. Econometrica 64, 865–890.

Fan, Y., Li, Q., 1996b. Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification tests. Journal of Nonparametric Statistics, forthcoming.

Fan, Y., Li, Q., 1996c. Root-N-consistent estimation of partially linear time series models. Journal of Nonparametric Statistics, forthcoming.

Gibbons, M, Ferson, W, 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. Journal of Financial Economics 14, 217–236.

Gibbons, M, Ross, S, Shanken, J, 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Gourieroux, C, Holly, A, Monfort, A, 1982. Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters. Econometrica 50, 63–80.

Gozalo, P.L, 1993. A consistent model specification test for nonparametric estimation of regression function models. Econometric Theory 9, 451–477.

Hall, P, 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. Journal of Multivariate Analysis 14, 1–16.

Hansen, B.E, 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.

Härdle, W, Mammen, E, 1993. Comparing nonparametric versus parametric regression fits. The Annals of Statistics 21, 1926–1947.

Hong, Y, White, H, 1995. Consistent specification testing via nonparametric series regression. Econometrica 63, 1133–1159.

Horowitz, J.L, Härdle, W, 1994. Testing a parametric model against a semiparametric alternative. Econometric Theory 10, 821–848.

Hsiao, C., Li, Q., 1997. A consistent test for conditional heteroskedasticity in time-series regression models. Manuscript.

Lavergne, P., Vuong, Q., 1996a. Nonparametric selection of regressors: the nonnested case. Econometrica 64, 207–219.

Lavergne, P., Vuong, Q., 1996b. Nonparametric significance testing. Manuscript.

Lewbel, A., 1993. Consistent tests with nonparametric components with an application to Chinese production data. Manuscript, Brandeis University.

Lewbel, A, 1995. Consistent nonparametric testing with an application to testing Slusky symmetry. Journal of Econometrics 67, 379–401.

Li, Q., 1994. Some simple consistent tests for a parametric model versus semiparametric or nonparametric alternatives. Manuscript.

Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression functional form. Journal of Econometrics 87, 145–165.

Linton, O., Gozalo, P.L., 1997. Consistent testing of additive models. Manuscript.

Newey, W.K, 1985. Maximum likelihood specification testing and conditional moment tests. Econometrica 53, 1047–1070.

Politis, D.N, Romano, J.P, 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Powell, J.L, Stock, J.H, Stoker, T.M, 1989. Semiparametric estimation of index coefficients. Econometrica 57 (6), 1403–1430.

Robinson, P.M, 1988. Root-N-consistent semiparametric regression. Econometrica 56 (4), 931–954.

Robinson, P.M, 1989. Hypothesis testing in semiparametric and nonparametric models for econometric time series. Review of Economic Studies 56, 511–534.

Robinson, P.M, 1991. Consistent nonparametric entropy-based testing. Review of Economic Studies 58, 437–453.

Stock, J.H., 1989. Nonparametric policy analysis. Journal of the American Statistical Association 84, 567–575.

Tauchen, G, 1985. Diagnostic testing and evaluation of maximum likelihood models. Journal of Econometrics 30, 415–443.

Ullah, A, 1985. Specification analysis of econometric models. Journal of Quantitative Economics 2, 187–209.

Whang, Y.J, Andrews, D.W.K, 1993. Tests of specification for parametric and semiparametric models. Journal of Econometrics 57, 277–318.

Wooldridge, J, 1992. A test for functional form against nonparametric alternatives. Econometric Theory 8, 452–475.

Yatchew, A.J, 1992. Nonparametric regression tests based on least squares. Econometric Theory 8, 435–451.

Wang, Q., 1997. A nonparametric test of the conditional mean-variance efficiency. Ph.D Thesis. University of Chicago.

Zheng, J.X, 1996. A consistent test of functional form via nonparametric estimation technique. Journal of Econometrics 75, 263–289.

Zheng, J.X., 1998a. Consistent specification testing for conditional symmetry. Econometric Theory 14, 139–149.

Zheng, J.X., 1998b. A specification test of conditional parametric distribution using kernel estimation methods. Manuscript.