A Nonparametric Goodness-of-fit-based Test for Conditional Heteroskedasticity*

Liangjun Su[a], Aman Ullah[b]

[a]*School of Economics, Singapore Management University, ljsu@smu.edu.sg*

[b]*Department of Economics, University of California, Riverside, aman.ullah@ucr.edu*

Oct 18, 2011

Abstract

In this paper we propose a new nonparametric test for conditional heteroskedasticity based on a measure of nonparametric goodness-of-fit (R^2) that is obtained from the local polynomial regression of the residuals from a parametric regression on some covariates. We show that after being appropriately standardized, the nonparametric R^2 is asymptotically normally distributed under the null hypothesis and a sequence of Pitman local alternatives. We also prove the consistency of the test and propose a bootstrap method to obtain the bootstrap p-values. We conduct a small set of simulations and compare our test with some popular parametric and nonparametric tests in the literature.

KEY WORDS: ANOVA; Conditional homoskedasticity; Local polynomial regressions; Nonparametric R^2; Nonparametric test.

JEL Classifications: C12, C14, C22

1 Introduction

Since the 1960s a large literature on testing for heteroskedasticity has developed; see Goldfeld and Quandt (1965), Glejser (1969), Godfrey (1978), Breusch and Pagan (1979), White (1980), Koenker and Bassett (1982), and Newey and Powell (1987), among others. Pagan and Pak (1993) argued that most of these early tests could be regarded as special cases of the conditional moment tests, which unfortunately are not robust against functional misspecification. Hong (1993) also demonstrated the inconsistency of these tests. For this reason, Hong (1993), Hsiao and Li (2001, **HL** hereafter), and Zheng (2006) proposed nonparametric consistent tests for heteroskedasticity. Hong constructed his test by comparing the kernel estimate of the conditional variance with the estimate of the unconditional variance for independently and identically distributed (IID) observations. **HL** borrowed the idea of consistent tests for model specification and constructed their test for heteroskedasticity applicable to time series data. Zheng's test

*We would like to express our sincere thank to the co-editor, Yoon-Jae Whang, and two anonymous referees for their constructive suggestions and comments that have led to a substantial improvement of the paper. We also thank the participants at the 2010 International Symposium on Econometric Theory and Applications (SETA 2010), the 2010 Econometric Society World Congress (ESWC 2010), the Rimini Conference on Economics and Finance (RCEF 2010), the 2011 Summer International Econometrics Symposium at SUFE, Chengdu, and the seminars at West Virginia and McGill universities, all of whom provided valuable suggestions and discussion. The second author acknowledges the financial support from the academic senate, UCR.

works for both parametric and nonparametric regression models but is limited to IID observations. A close look at these three tests indicates that they share the same formula despite different motivations and derivations.

In this paper we propose a new test for conditional homoskedasticity based on a novel measure for nonparametric goodness-of-fit (R^2). Huang and Chen (2008) proposed a measure of goodness-of-fit for local polynomial regressions in the spirit of analysis of variance (ANOVA) decomposition in multiple linear regression models. We believe that this measure can serve as a useful statistic for testing many popular hypotheses in econometrics and statistics by playing a role comparable to the important role that R^2 plays in the parametric setup. It is well-known that many LM-type and residual-based test statistics in the parametric framework can be recast as nR^2 (e.g., Greene, 2000, pp. 156-157, 196-197, 440, 541, 572), where n is the sample size and R^2 is the coefficient of determination from some residual-based auxiliary regressions that are *parametrically* specified. In the case of functional misspecification in these auxiliary regressions, these tests might be inconsistent and thus lead to misleading conclusions. To avoid such misspecification, we propose to adopt *nonparametric* models in place of parametric models in the auxiliary regressions. Then we construct a nonparametric analogue of the parametric residual-based test by applying the nonparametric measure of goodness-of-fit. To stay focused here, we apply the nonparametric R^2 to test for conditional homoskedasticity. After fitting a parametric model for the conditional mean regression, we obtain the residuals whose squares are used in the second-stage auxiliary local polynomial regression. We calculate the nonparametric R^2 from this regression. It becomes small and close to 0 under the null of conditional homoskedasticity and deviates from 0 under the alternative of conditional heteroskedasticity. We show that after being properly standardized, it is asymptotically normally distributed under the null hypothesis and a sequence of Pitman local alternatives that converge to the null at the usual nonparametric rate.

A great advantage of our test is that it works for both local constant regressions and local polynomial regressions. It is well known that the uniform consistency of the local polynomial estimators typically requires compact support for the conditioning variable so that its density is bounded away from zero on the support, whereas the local constant estimator has a boundary bias in this case. For this reason, the two subclasses of estimators have to be addressed separately. Since our nonparametric R^2 test is based on an *integrated* measure of the explained sum of squares, the boundary bias issue of local constant estimators does not pop up in our framework and thus the asymptotic theory for our nonparametric R^2-based tests work for both cases. In the case of univariate regression, we focus on the widely used local constant, local linear, and local quadratic regressions and find that the local constant R^2-based test has higher asymptotic local power than the local linear R^2-based test, which in turn has higher power than the local quadratic R^2-based test when the *same* kernel and bandwidth are used.

We also compare our tests with **HL**'s test for conditional heteroskedasticity. Both tests are residual-based, consistent against all fixed global alternatives, and have nontrivial power against the *same* sequence of Pitman local alternatives if the *same* bandwidth is used. In general the two tests are not comparable because **HL**'s test is based on density-weighted moment conditions so that the density of the conditioning variable enters their asymptotic local power function explicitly whereas our test is based on nonparametric R^2 which is self-normalized and measurement-unit free so that the density function is absent from our asymptotic local power function. To make a fair comparison with **HL**'s test, we consider a density weighted version of our local constant R^2-based test. We find that this test has a larger asymptotic power than **HL**'s test against the *same* sequence of Pitman local alternatives if the *same* bandwidth sequence and kernel function are used in constructing both tests.

The rest of the paper is organized as follows. We state the hypothesis and define the nonparametric

R^2 in Section 2. In Section 3 we study the asymptotic distributions of our test statistic under the null and a sequence of local alternatives and establish its global consistency. In Section 4 we conduct Monte Carlo experiments to evaluate the finite sample performance of our test in comparison with some other tests. We conclude in Section 5. All technical proofs are relegated to the Appendix.

2 Basic Framework

In this section we first introduce the null and alternative hypotheses, then propose a test statistic based on the measure of nonparametric goodness-of-fit.

2.1 Hypotheses

Following **HL**, we consider a nonlinear model of the form

$$Y_{nt} = g\left(Z_{nt}, \theta_0\right) + U_{nt}, \ t = 1, 2, ..., n, \tag{2.1}$$

where $g(\cdot, \cdot)$ is a function of known form, θ_0 is a $d \times 1$ vector of unknown parameters, Z_{nt} is a $k \times 1$ vector of regressors, and U_{nt} is a scalar error term such that $E\left(U_{nt}|Z_{nt}\right) = 0$ almost surely (a.s.). Note that we have written (2.1) using triangular array notation, which will greatly facilitate the study of the local power property of out test. It is also possible to allow g to depend on n. But we feel it is natural to assume that the functional relationship between the dependent and independent variables is invariant to the sample size n. Similar remark holds for the parameter of interest θ_0.

Let $V_{nt} \equiv U_{nt}^2$ and $m_n\left(X_{nt}\right) \equiv E\left(V_{nt}|X_{nt}\right)$ where X_{nt} is a $p \times 1$ vector of variables. The null of interest is that conditional on X_{nt}, U_{nt}'s are homoskedastic, i.e.,

$$\mathbb{H}_0 : P\left[m_n\left(X_{nt}\right) = \sigma_0^2\right] = 1 \text{ for some } \sigma_0^2 > 0. \tag{2.2}$$

The alternative hypothesis is $\mathbb{H}_1 : P\left[m_n\left(X_{nt}\right) = \sigma_0^2\right] < 1$ for all $\sigma_0^2 > 0$. Note that we allow the elements in X_{nt} to be distinct from those in Z_{nt}.[1] To examine the asymptotic local power of our test, we will consider the following sequence of Pitman local alternatives:

$$\mathbb{H}_1(\gamma_n) : m_n\left(x\right) = \sigma_0^2 + \gamma_n \Delta_n\left(x\right), \tag{2.3}$$

where $\gamma_n \to 0$ as $n \to \infty$ and $\Delta_n\left(x\right)$ is a nonconstant continuous function such that $m_n\left(X_{nt}\right) > 0$ a.s.

The consistent tests of Hong (1993), **HL**, and Zheng (2006) are all residual-based tests that rely on the observation that $E\left(V_{nt} - \sigma_0^2|X_{nt}\right) = 0$ a.s. under $\mathbb{H}_0$. Let $\epsilon_{nt} \equiv V_{nt} - \sigma_0^2$, and let $f_n\left(\cdot\right)$ denote the marginal probability density function (PDF) of X_{nt}. It is easy to see that

$$J_n \equiv E\left[\epsilon_{nt} E\left(\epsilon_{nt}|X_{nt}\right) f_n\left(X_{nt}\right)\right] = E\{\left[E\left(\epsilon_{nt}|X_{nt}\right)\right]^2 f_n\left(X_{nt}\right)\} \tag{2.4}$$

is 0 under $\mathbb{H}_0$ and strictly positive otherwise. Let $\hat{\theta}$ denote the nonlinear least squares (NLS) estimator of θ_0 in (2.1). Let $\hat{U}_t \equiv Y_{nt} - g(Z_{nt}, \hat{\theta})$, $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{t=1}^n \hat{V}_t$, and $\hat{\epsilon}_t \equiv \hat{V}_t - \hat{\sigma}_0^2$, where $\hat{V}_t = \hat{U}_t^2$. The above observation motivates **HL** to consider the following test statistic

$$\hat{J}_n = \frac{1}{n^2} \sum_{t=1}^{n} \sum_{s=1, s \neq t}^{n} \hat{\epsilon}_t \hat{\epsilon}_s K_h\left(X_{nt} - X_{ns}\right) \tag{2.5}$$

where $K_h\left(\cdot\right) \equiv K\left(\cdot/h\right)/h$, $K\left(\cdot\right)$ is a symmetric PDF on $\mathbb{R}^p$, and $h \equiv h\left(n\right)$ is a bandwidth sequence.

Below, we propose a new consistent test for $\mathbb{H}_0$ by using R^2 for the nonparametric regression models.

[1] There is no requirement on the relationship between X_{nt} and Z_{nt}. The two random vectors can be identical or totally different. It is also possible for X_{nt} to be a subset of Z_{nt} or the other way around.

2.2 A nonparametric R^2-based test for conditional heteroskedasticity

If ϵ_{nt} were observable, we could consider the nonparametric regression model

$$\epsilon_{nt} = \bar{m}_n(X_{nt}) + \varepsilon_{nt} \tag{2.6}$$

where $\varepsilon_{nt} \equiv V_{nt} - m_n(X_{nt})$ and $\bar{m}_n(X_{nt}) = m_n(X_{nt}) - \sigma_0^2$. Under $\mathbb{H}_0$, $\bar{m}_n(X_{nt}) = 0$ a.s. so that any goodness-of-fit measure for the above nonparametric regression model should be close to 0. This motivates us to propose a test based on Huang and Chen's (2008) nonparametric measure of goodness-of-fit.

A feasible regression model is given by

$$\hat{\epsilon}_t = \bar{m}_n(X_{nt}) + e_{nt} \tag{2.7}$$

where e_{nt} is the new error term in the above regression. Given observations $\{(\hat{\epsilon}_t, X_{nt})\}_{t=1}^n$, the qth-order local-polynomial regression of $\hat{\epsilon}_t$ on X_{nt} is fitted by the weighted least squares (WLS) as follows

$$\min_{\boldsymbol{\beta}} n^{-1} \sum_{t=1}^n \left(\hat{\epsilon}_t - \sum_{0 \le |\mathbf{j}| \le q} \beta_{\mathbf{j}} (X_{nt} - x)^{\mathbf{j}} \right)^2 K_h(X_{nt} - x), \tag{2.8}$$

where we use the notation of Masry (1996): $\mathbf{j} = (j_1, ..., j_p)$, $|\mathbf{j}| = \sum_{i=1}^p j_i$, $x^{\mathbf{j}} = \mathbf{\Pi}_{i=1}^p x_i^{j_i}$, $\sum_{0 \le |\mathbf{j}| \le q} = \sum_{k=0}^q \sum_{\substack{j_1=0 \\ j_1+\cdots+j_p=k}}^k \cdots \sum_{j_p=0}^k$, and $\boldsymbol{\beta}$ is a stack of $\beta_{\mathbf{j}}$ $(0 \le |\mathbf{j}| \le q)$ in the lexicographical order (with highest priority to last position so that $(0, 0, ..., l)$ is the first element in the sequence and $(l, 0, ..., 0)$ is the last element).

Let $\hat{\beta}_{\mathbf{j}}(x; h)$ $(0 \le |\mathbf{j}| \le q)$ denote the solution to the above problem. Based on the normal equations for the above regression and the fact that $\hat{\epsilon}_t$ has zero sample mean, it is easy to verify the following local ANOVA decomposition of the total sum of squares (TSS)

$$TSS(x) = ESS_q(x) + RSS_q(x) \tag{2.9}$$

where $TSS(x) \equiv \sum_{t=1}^n \hat{\epsilon}_t^2 K_{tx}$, $ESS_q(x) \equiv \sum_{t=1}^n [\sum_{0 \le |\mathbf{j}| \le q} \hat{\beta}_{\mathbf{j}}(x; h)(X_{nt} - x)^{\mathbf{j}}]^2 K_{tx}$, $RSS_q(x) \equiv \sum_{t=1}^n [\hat{\epsilon}_t - \sum_{0 \le |\mathbf{j}| \le q} \hat{\beta}_{\mathbf{j}}(x; h)(X_{nt} - x)^{\mathbf{j}}]^2 K_{tx}$, and $K_{tx} \equiv K_h(X_{nt} - x)$. Note that $TSS(x)$ does not depend on q. A global ANOVA decomposition of TSS is given by

$$TSS = ESS_q + RSS_q \tag{2.10}$$

where $TSS \equiv \int_{\mathcal{X}_n} TSS(x)\, dx$, $ESS_q \equiv \int_{\mathcal{X}_n} ESS_q(x)\, dx$, and $RSS_q \equiv \int_{\mathcal{X}_n} RSS_q(x)\, dx$, where $\mathcal{X}_n$ is a compact subset of the support of the PDF $f_n(\cdot)$ of X_{nt}. In particular, if $f_n(\cdot)$ has compact support $\mathbb{X}_n$, then one can take $\mathcal{X}_n$ to be $\mathbb{X}_n$. Then one can define the nonparametric goodness-of-fit (R^2) for the above qth-order local polynomial regression as

$$R_q^2 = 1 - \frac{RSS_q}{TSS} = \frac{ESS_q}{TSS}. \tag{2.11}$$

For more interpretations of R_q^2 and its local version, we refer the readers to Huang and Chen (2008).

Clearly R_q^2 lies between 0 and 1. The smaller the value of R_q^2 is, the worse is the fit. In the extreme case, if no regressors among X_{nt} can explain ϵ_{nt}, we expect a value of R_q^2 close to 0 in any given sample of observations on $\{\hat{\epsilon}_t, X_{nt}\}$. Let $X_{q,tx} \equiv \mu_q(X_{nt} - x)$ denote the stack of $(X_{nt} - x)^{\mathbf{j}}$, $0 \le |\mathbf{j}| \le q$, in the lexicographical order.[2] For example, $X_{q,tx} = 1$ if $q = 0$, and $X_{q,tx} = (1, (X_{nt} - x)')'$ if $q = 1$. Let $\mathbf{X}_{q,x} \equiv (X_{q,1x}, \cdots X_{q,nx})'$, $\mathbf{W}_x \equiv$diag$(K_{1x}, \cdots, K_{nx})$, $\mathbf{H}_{q,x} \equiv \mathbf{W}_x \mathbf{X}_{q,x} \left(\mathbf{X}'_{q,x} \mathbf{W}_x \mathbf{X}_{q,x}\right)^{-1} \mathbf{X}'_{q,x} \mathbf{W}_x$, and $H_q^* \equiv \int_{\mathcal{X}_n} \mathbf{H}_{q,x} dx$. It is easy to verify that $ESS_q = \hat{\mathbf{v}}' M H_q^* M \hat{\mathbf{v}}$, where $\hat{\mathbf{v}} \equiv (\hat{V}_1, \cdots, \hat{V}_n)'$, $M \equiv I_n - L$, and I_n and L denote an $n \times n$ identity matrix and an $n \times n$ matrix with entries $1/n$, respectively. Then

$$R_q^2 = \frac{ESS_q}{TSS} = \frac{\hat{\mathbf{v}}' M H_q^* M \hat{\mathbf{v}}}{TSS}. \tag{2.12}$$

Clearly, if the *same* bandwidth and kernel functions are used in constructing the nonparametric R^2 statistics for different orders of local polynomial regressions, then we observe that $R_{q+1}^2 \geq R_q^2$ for any $q \geq 0$. We will study the asymptotic properties of R_q^2 in the next section.

Remark 1. It is worth mentioning that the above formulation of the R_q^2 statistic works for any finite order of local polynomial regressions including the local constant regressions.[3] In practice, typical choices of q are 0, 1, and 2, which correspond to the local constant, local linear, and local quadratic regressions, respectively. For technical reason, we assume that $\mathcal{X}_n$ is compact and has a compact limit $\mathcal{X}$ as $n \to \infty$ in the sense that vol$(\mathcal{X}_n \backslash \mathcal{X})$+vol$(\mathcal{X} \backslash \mathcal{X}_n) \to 0$ as $n \to \infty$, where, e.g., $\mathcal{X}_n \backslash \mathcal{X}$ denotes the relative complement of $\mathcal{X}$ in $\mathcal{X}_n$. This goes along with the literature on classical local polynomial regressions with $q \geq 1$ because the uniform consistency of the local polynomial estimators requires that the support $\mathcal{X}_n$ should be compact and that $f_n(\cdot)$ should be bounded and bounded away from 0 on $\mathcal{X}_n$. In sharp contrast, one does not want to assume compact support when establishing the uniform consistency of the local constant estimator ($q = 0$) because the latter has the notorious boundary bias issue. Interestingly, the asymptotic theory established in this paper works regardless of whether $q = 0$ or $q \geq 1$. The intuition is that our nonparametric R^2 test is based on an *integrated* measure of the explained sum of squares, the boundary bias issue of local constant estimators does not pop up in our framework.

To proceed, we define some notation. Let $N_{ql} = (l + q - 1)!/(l!(q-1)!)$ be the number of distinct q-tuples $\mathbf{j}$ with $|\mathbf{j}| = l$, $0 \leq l \leq q$. It denotes the number of distinct l-th order partial derivatives of $m_n(x)$ with respect to x. Arrange the N_{ql} q-tuples as a sequence in the lexicographical order, and let ϕ_l^{-1} denote this one-to-one map. For each $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2q$, let $\nu_{\mathbf{j}} = \int_{\mathbb{R}^p} x^{\mathbf{j}} K(x) dx$, and define the $N_q \times N_q$ dimensional matrix $\mathbb{S}_q$ and $N_q \times 1$ vector $\mathbb{B}_q$, where $N_q = \sum_{l=0}^q N_{ql}$, by

$$\mathbb{S}_q = \begin{bmatrix} \mathbb{S}_{q,0,0} & \mathbb{S}_{q,0,1} & \cdots & \mathbb{S}_{q,0,q} \\ \mathbb{S}_{q,1,0} & \mathbb{S}_{q,1,1} & \cdots & \mathbb{S}_{q,1,q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{S}_{q,q,0} & \mathbb{S}_{q,q,1} & \cdots & \mathbb{S}_{q,q,q} \end{bmatrix}, \ \mathbb{B}_q = \begin{bmatrix} \mathbb{S}_{q,0,0} \\ \mathbb{S}_{q,1,0} \\ \vdots \\ \mathbb{S}_{q,q,0} \end{bmatrix}, \tag{2.13}$$

where $\mathbb{S}_{q,i,j}$ are $N_{qi} \times N_{qj}$ dimensional matrices whose (l, r) elements are $\nu_{\phi_i(l)+\phi_j(r)}$.

3 Asymptotic Distributions

In this section we first present assumptions and then study the asymptotic distributions of R_q^2-based tests under $\mathbb{H}_0$ and $\mathbb{H}_1(\gamma_n)$. We also prove the consistency of the test and propose a bootstrap method to obtain bootstrap p-values.

[2] $\mu_q(X_{nt} - x)$ essentially stacks the regressors (not weighted by the kernel weight yet) in the qth local polynomial regression.

[3] Notice that local polynomial regressions typically refer to the case where $q \geq 1$.

3.1 Assumptions

Let $C<\infty$ denote a generic constant whose value may change across lines. Let $W_{nt}\equiv(U_{nt},Z_{nt}',X_{nt}')'$. Following Yoshihara (1992) and Su and White (2010), we will use the mixing coefficients $\alpha_n(j)$, defined by $\alpha_n(j)=\sup_{1\le l\le n-j}\{P(A\cap B)-P(A)P(B)|A\in\sigma(W_{nt}:1\le t\le l),B\in\sigma(W_{nt}:l+j\le t\le n)\}$ if $j\le n-1$, and $\alpha_n(j)=0$ if $j\ge n$. Define the coefficients of strong mixing as $\alpha(j)=\sup_{n\in\mathbb{N}}\alpha_n(j)$ for $j\in\mathbb{N}$ and $\alpha(0)=1$. Our assumptions are as follows.

Assumption A1. The process $\{W_{nt},t=1,...,n;\ n=1,2,...\}$ is a strictly stationary strong mixing process with mixing coefficients $\alpha(s)$ such that $\sum_{s=0}^{\infty}s^4\alpha(s)^{\eta/(1+\eta)}\le C$ for some $\eta>0$ with $\eta/(1+\eta)\le 1/2$, and $\alpha(s)^{(2+\tilde\eta)/[3(4+\tilde\eta)]}=O(s^{-1})$ and $\alpha(s)^{\tilde\eta/(2+\tilde\eta)}=O(s^{-2+\epsilon})$ for some $\tilde\eta\in(0,\eta)$ and sufficiently small $\epsilon>0$.

Assumption A2. (i) $E(\varepsilon_{nt}|X_{nt})=0$ a.s. for all n.

(ii) $\sup_{n\ge1}E[|\varepsilon_{nt}|^{4+\eta}]\le C$, and $\sup_{n\ge1}\sup_{1\le t_1,...,t_l\le n}\max_{i_1,..i_l:\ \Sigma_{j=1}^{l}i_j\le 8}E[|\varepsilon_{nt_1}^{i_1}\varepsilon_{nt_2}^{i_2}\cdots\varepsilon_{nt_l}^{i_l}|^{1+\zeta_1}]\le C$ for some arbitrarily small $\zeta_1>0$, where $2\le l\le 4$.

(iii) Let $v_n^2(x)\equiv E[\varepsilon_{nt}^2|X_{nt}=x]$, and $\mu_{4n}(x)\equiv E[\varepsilon_{nt}^4|X_{nt}=x]$. Both $v_n^2(x)$ and $\mu_{4n}(x)$ are Lipschitz continuous in that $|\vartheta_n(x+\tilde x)-\vartheta_n(x)|\le D_{n,\vartheta}(x)\|\tilde x\|$ and $\sup_{n\ge1}E[|D_{n,\vartheta}(X_n)|^{2+\zeta_2}]\le C$ for $\vartheta_n(\cdot)=v_n^2(\cdot)$ or $\mu_{4n}(\cdot)$ and some arbitrarily small $\zeta_2>0$ where $\|\cdot\|$ denotes the Euclidean norm. $v^2(x)\equiv\lim_{n\to\infty}v_n^2(x)$ exists for each x.

(iv) For each $n=1,2,...$ and $l=1,...,4$ such that $1\le t_1<...<t_l\le n$, the joint PDF $f_{n,t_1,\cdots,t_l}(\cdot,\cdots,\cdot)$ of $(X_{nt_1},\cdots,X_{nt_l})$ for exists, is finite, and is Lipschitz continuous in that $|f_{n,t_1,\cdots,t_l}(x_1+z_1,\cdots,x_l+z_l)-f_{n,t_1,\cdots,t_l}(x_1,\cdots,x_l)|\le D_{n,t_1,\cdots,t_l}(x_1,\cdots,x_l)\|\mathbf{z}\|$, where $\mathbf{z}\equiv(z_1'\cdots,z_l')'$, $\sup_{n\ge1}E|D_{n,t_1,\cdots,t_l}(X_{nt_1},\cdots,X_{nt_l})|\le C$, and $\sup_{n\ge1}\int_{\mathcal{X}_n\times\cdots\times\mathcal{X}_n}D_{n,t_1,\cdots,t_l}(x_1,\cdots,x_l)\|\mathbf{x}\|^{2(1+\eta)}d\mathbf{x}\le C$ with $\mathbf{x}\equiv(x_1',\cdots,x_l')'$. When $l=1$, we use $f_n(\cdot)$ to denote the marginal PDF of X_{nt}. $f_n(\cdot)$ is bounded away from 0 on the nonrandom compact set $\mathcal{X}_n$. $\mathcal{X}_n\to\mathcal{X}$ as $n\to\infty$ and $\mathcal{X}$ is compact.[4]

Assumption A3. (i) $E(U_{nt}|Z_{nt})=0$ a.s. for all n.

(ii) The parameter space Θ of θ is a compact subset of $\mathbb{R}^d$. For $n=1,2,...$, $E[Y_{nt}-g(Z_{nt},\theta)]^2$ is uniquely minimized at θ_0 on Θ.

(iii) The regression function $g(z,\theta)$ is continuously differentiable of order 2 in θ. Let $\triangledown g(z,\theta)\equiv\partial g(z,\theta)/\partial\theta$ and $\triangledown^2 g(z,\theta)\equiv\partial^2 g(z,\theta)/\partial\theta\partial\theta'$. $\triangledown g(z,\cdot)$ and $\triangledown^2 g(z,\cdot)$ are continuous in z and are dominated by functions $G_1(z)$ and $G_2(z)$, respectively. $G_1(z)$ and $G_2(z)$ have finite fourth and second moments, respectively.

(iv) For $n=1,2,...$, $E[\triangledown g(Z_{n1},\theta)\triangledown g(Z_{n1},\theta)']$ is nonsingular for all θ in a small open neighborhood of θ_0.

Assumption A4. (i) For $n=1,2,...$, $\bar m_n(x)$ is Lipschitz continuous in x and has all partial derivatives up to order $q+1$ or 2 if $q=0$.

(ii) For $n=1,2,...$, the $(q+1)$th or 2nd order partial derivatives $D^{\mathbf{k}}\bar m_n(x)$ with $|\mathbf{k}|=q+1$ (if $q\ge1$) or 2 (if $q=0$), are uniformly bounded in $x\in\mathcal{X}_n$, and are Hölder continuous in x : $|D^{\mathbf{k}}\bar m_n(x)-D^{\mathbf{k}}\bar m_n(\tilde x)|\le C||x-\tilde x||$.

Assumption A5. (i) The kernel function $K(\cdot)$ is a continuous, bounded, and symmetric PDF.

(ii) $\|x\|^{(4+\eta)q}K(x)$ is integrable and $\mathbb{S}_q$ defined in (2.13) is nonsingular.

[4] Of course one can allow the compact support $\mathcal{X}_n$ to expand slowly as the sample size n passes to the infinity (see, e.g., Andrews (1995), Hansen (2008) and Li, Lu and Linton (2011)), but this is at the cost of slowing down the rate of uniform convergence. One can also allow $\mathcal{X}_n$ to be random at the cost of more complex arguments; see footnote 9.

(iii) Let $\mathbf{K}_{\mathbf{j}}(x) \equiv x^{\mathbf{j}} K(x)$ for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2q+1$. For some $C_1 < \infty$ and $C_2 < \infty$, either $K(x)$ is compactly supported such that $K(x) = 0$ for $\|x\| > C_1$, and $|\mathbf{K}_{\mathbf{j}}(x) - \mathbf{K}_{\mathbf{j}}(\tilde{x})| \leq C_2 \|x - \tilde{x}\|$ for any $x, \tilde{x} \in \mathbb{R}^p$ and for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2q+1$; or $K(x)$ is differentiable, $\|\bigtriangledown \mathbf{K}_{\mathbf{j}}(x)\| \leq C_1$ and for some $\iota_0 > 1$, $|\bigtriangledown \mathbf{K}_{\mathbf{j}}(x)| \leq C_1 \|x\|^{-\iota_0}$ for all $\|x\| > C_2$ and for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2q+1$.

Assumption A6. As $n \to \infty$, $h \to 0$, $nh^{3p/2} \to \infty$, and $nh^{p+2}/(\log n)^2 \to c \in (0, \infty]$.

Assumption A1 is typical in nonparametric inference with time series observations. Here we only assume that the stochastic process $\{W_{nt}\}$ is strong mixing, which is weaker than absolute regularity assumed in Hsiao and Li (2001). Also the restriction on the mixing rate is weaker than the latter's geometric decay rate. Under $\mathbb{H}_0$ and $\mathbb{H}_1$, $\{W_{nt}\}$ is typically not a triangular array and thus written as $\{W_t\}$ in which case the definition of strong mixing coefficients reduces to the usual one. Under $\mathbb{H}_1(\gamma_n)$, we need resort to the triangular array notation and refer the reader to Yoshihara (1992) for the notion of strictly stationary and strong mixing triangular array processes. Strict stationarity can be relaxed at greater complication of notation. Assumption A2 is needed to apply Gao's (2007) CLT for second order U-statistics with strong mixing data and show that certain terms are asymptotically negligible in Lemma A.1. Assumption A3, together with A1 and A2(ii), ensures that $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ by White and Domowitz (1984). Assumptions A4-A6 are used to obtain the uniform consistency for the local polynomial estimator due to Masry (1996) and Hansen (2008). A4 is automatically satisfied under $\mathbb{H}_0$.

3.2 Asymptotic null distribution

Let $H^*_{q,ts}$ denote the (t,s)th element of H^*_q. Let $B_{qn} \equiv h^{p/2} \sum_{t=1}^n \varepsilon^2_{nt} H^*_{q,tt} / (n^{-1}TSS)$ and $\Omega_q \equiv \int \left[\int K(z) \mu_q(z)' \mathbb{S}_q^{-1} \mu_q(z+x) K(z+x) dz\right]^2 dx \int_{\mathcal{X}} \left[v^2(x)\right]^2 dx$, where all integrations are computed over $\mathbb{R}^p$ unless otherwise indicated, and $\mu_q(x)$ denotes the stack of $x^{\mathbf{j}}$, $0 \leq |\mathbf{j}| \leq q$, in the lexicographical order.

Theorem 3.1 *Suppose Assumptions A1-A3 and A5-A6 hold. Then under* $\mathbb{H}_0$, $\Gamma_{qn} \equiv nh^{p/2} R_q^2 - B_{qn} \xrightarrow{d} N(0, \Omega_q / \sigma_V^4)$ *where* $\sigma_V^2 \equiv \lim_{n \to \infty} E\{[V_{n1} - E(V_{n1})]^2 \int_{\mathcal{X}_n} K_h(X_{n1} - x) dx\}$.

Remark 2. The proof of the above theorem is tedious and is relegated to the Appendix. The idea underlying the proof is very simple. Under the null hypothesis, we first demonstrate that $n^{-1}TSS \cdot \Gamma_{qn} = \bar{\Gamma}_{qn} + o_p(1)$, where $\bar{\Gamma}_{qn} \equiv \frac{2}{n} \sum_{1 \leq t < s \leq n} \varphi_n(\xi_{nt}, \xi_{ns})$, $\xi_{nt} \equiv (X'_{nt}, \varepsilon_{nt})'$, $\varphi_n(\xi_{nt}, \xi_{ns}) \equiv h^{p/2} \varepsilon_{nt} \varepsilon_{ns} \int_{\mathcal{X}_n} K_{tx} X'_{q,tx} D_h^{-1} \bar{S}_{qn}^{-1} D_h^{-1} X_{q,sx} K_{sx} dx$, $D_h \equiv$diag$(1, h\mathbf{1}'_{N_{q1}}, \cdots, h^q \mathbf{1}'_{N_{qq}})$ denotes an $N_q \times N_q$ diagonal matrix with typical elements given by h^s, $s = 0, 1, \cdots, q$, $\bar{S}_{qn}(x) \equiv E[S_{qn}(x)]$, and $S_{qn}(x) \equiv n^{-1} D_h^{-1} \mathbf{X}'_{q,x} \mathbf{W}_x \mathbf{X}_{q,x} D_h^{-1}$. Apparently $\bar{\Gamma}_{qn}$ is a second-order U-statistic with symmetric kernel $\varphi_n(\cdot, \cdot)$. Then we can apply the central limit theorem (CLT) for second-order U-statistics under strong mixing processes and demonstrate that $\bar{\Gamma}_{qn} \xrightarrow{d} N(0, \Omega_q)$. The result then follows by noticing that $n^{-1}TSS = \sigma_V^2 + o_p(1)$ regardless of whether $\mathbb{H}_0$ holds or not.[5]

To implement the test, we require consistent estimates of B_{qn} and Ω_q. We propose to estimate B_{qn} and Ω_q respectively by $\hat{B}_{qn} \equiv h^{p/2} \sum_{t=1}^n \hat{\epsilon}_t^2 H^*_{q,tt} / (n^{-1}TSS)$ and $\hat{\Omega}_{qn} \equiv 2n^{-2} h^p \sum_{s=1}^n \sum_{t \neq s}^n \hat{\epsilon}_t^2 \hat{\epsilon}_s^2 (nH^*_{ts})^2$. Then we define a feasible nonparametric R^2-based test statistic as

$$T_{qn} = \left(nh^{p/2} R_q^2 - \hat{B}_{qn}\right) / \sqrt{\hat{\Omega}_{qn} / (n^{-1}TSS)^2}. \tag{3.1}$$

The following corollary establishes the consistency of $\hat{B}_{qn}$ and $\hat{\Omega}_{qn}$ and the asymptotic distribution of T_{qn} under $\mathbb{H}_0$.

[5] If the support $\mathbb{X}_n$ of $f_n(\cdot)$ is compact and $\mathcal{X}_n = \mathbb{X}_n$, we can readily show that $\sigma_V^2 \equiv \lim_{n \to \infty} \text{Var}(V_{nt})$.

Corollary 3.2 *Suppose A1-A3 and A5-A6 hold. Then under* $\mathbb{H}_0$*,* $\hat{B}_{qn} = B_{qn} + o_p(1)$ *,* $\hat{\Omega}_{qn} = \Omega_q + o_p(1)$ *, and* $T_{qn} \xrightarrow{d} N(0,1)$.

Remark 3. It is worth mentioning that we prove the first two parts of the above corollary under $\mathbb{H}_1(\gamma_n)$ defined in (2.3) with $\gamma_n = n^{-1/2}h^{-p/4}$, which implies that $\hat{B}_{qn}$ and $\hat{\Omega}_{qn}$ are consistent estimates of B_{qn} and Ω_q under both $\mathbb{H}_0$ and $\mathbb{H}_1(n^{-1/2}h^{-p/4})$. The last part of Corollary 3.2 implies that the feasible test statistic T_{qn} is asymptotically pivotal. We can compare T_{qn} with the one-sided critical value z_α, i.e., the $100(1-\alpha)$th percentile from the standard normal distribution. We reject the null when $T_{qn} > z_\alpha$ at the α significance level.

3.3 Asymptotic local power

Let $\Lambda_q \equiv \mathbb{B}_q' \mathbb{S}_q^{-1} \mathbb{B}_q \lim_{n\to\infty} \text{Var}[\Delta_n(X_{n1})]$. The following theorem establishes the local power property of our test for the Pitman local alternatives defined in (2.3).

Theorem 3.3 *Suppose Assumptions A1–A6 hold. Suppose that* $\Delta_n(x)$ *is a continuous function such that* $\lim_{n\to\infty} E\left[\triangle_n^2(X_{n1})\right] < \infty$*. Then the local power of* T_{qn} *satisfies* $P\left[T_{qn} \geq z \mid \mathbb{H}_1(n^{-1/2}h^{-p/4})\right] \to 1 - \Phi(z - \Lambda_q/\sqrt{\Omega_q})$ *as* $n \to \infty$*, where* $\Phi(\cdot)$ *is the cumulative distribution function (CDF) of the standard normal.*

Remark 4. Theorem 3.3 implies that the test has non-trivial asymptotic power against alternatives that converge to the null at the rate $n^{-1/2}h^{-p/4}$. Furthermore, the result in Theorem 3.3 holds for any finite order of local polynomial regressions including the local constant regressions. For the local constant, local linear, and local quadratic regressions (i.e., $q = 0$, 1, and 2), it is straightforward to verify that $\mathbb{B}_q'\mathbb{S}_q^{-1}\mathbb{B}_q = 1$ and hence $\Lambda_q = \lim_{n\to\infty}\text{Var}[\Delta_n(X_{n1})]$ under $\mathbb{H}_1(n^{-1/2}h^{-p/4})$. Other than these three cases, we are unable to determine the exact value of $\mathbb{B}_q'\mathbb{S}_q^{-1}\mathbb{B}_q$. If $q = 3$, by the formula for partitioned inverse and the symmetry of the kernel function $K(\cdot)$, we can show that $\mathbb{B}_q'\mathbb{S}_q^{-1}\mathbb{B}_q = (1-a)^3 + a \leq 1$, where $a \equiv \mathbb{S}_{3,0,2}\mathbb{S}_{3,2,2}^{-1}\mathbb{S}_{3,2,0} \leq 1$ by the Cauchy-Schwarz inequality. This implies that Λ_3 is no bigger than $\lim_{n\to\infty}\text{Var}[\Delta_n(X_{n1})]$ in the case of $q = 3$. Similarly, the formula Ω_q appears too complicated to simplify for general values of q. Therefore we focus on $q = 0$, 1, 2 and restrict our attention to the cases where $p = 1$ and $K(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. Then tedious but straightforward calculations show that

$$\begin{aligned} \Omega_0 &= \frac{1}{2\sqrt{2\pi}}\int_{\mathcal{X}}\left[v^2(x)\right]^2 dx \approx 0.1995\int_{\mathcal{X}}\left[v^2(x)\right]^2 dx, \\ \Omega_1 &= \frac{27}{32\sqrt{2\pi}}\int_{\mathcal{X}}\left[v^2(x)\right]^2 dx \approx 0.3366\int_{\mathcal{X}}\left[v^2(x)\right]^2 dx, \\ \Omega_2 &= \frac{2265}{2048\sqrt{2\pi}}\int_{\mathcal{X}}\left[v^2(x)\right]^2 dx \approx 0.4412\int_{\mathcal{X}}\left[v^2(x)\right]^2 dx. \end{aligned} \tag{3.2}$$

Interestingly, the higher the order of q is, the higher is the value of Ω_q. This, in conjunction with the fact that $\Lambda_0 = \Lambda_1 = \Lambda_2$, indicates that the local constant R^2-based test has higher asymptotic local power than the local linear R^2-based test, which in turn has higher power than the local quadratic R^2-based test. This is in sharp contrast to fact that $R_2^2 \geq R_1^2 \geq R_0^2$ when the *same* kernel and bandwidth are used in constructing the nonparametric R^2 statistics. Consequently, it seems that no benefit can be achieved by using higher order local polynomial regressions as far as asymptotic *local* powers are concerned.

Remark 5. It is worthwhile to compare the asymptotic local power property of our test with that of **HL**'s. For the test statistic defined in (2.5), define its normalized version as

$$HL_n = nh^{p/2}\hat{J}_n/\sqrt{\hat{\Omega}_n} \tag{3.3}$$

where $\hat{\Omega}_n = 2n^{-2}h^p \sum_{s=1}^n \sum_{t\neq s}^n \hat{\epsilon}_t^2 \hat{\epsilon}_s^2 [K_h(X_{nt} - X_{ns})]^2$. Theorem 3.4 in **HL** suggests that under some regularity conditions specified in their paper, HL_n has the following asymptotic local power property:

$$P\left[HL_n \geq z \mid \mathbb{H}_1(n^{-1/2}h^{-p/4})\right] \to 1 - \Phi(z - \Lambda_{HL}/\sqrt{\Omega_{HL}}) \text{ as } n \to \infty,$$

where $\Lambda_{HL} \equiv \lim_{n\to\infty} E\left[\Delta_n^2(X_{n1}) f_n(X_{n1})\right]$ and $\Omega_{HL} \equiv 2\lim_{n\to\infty} E\{\left[v_n^2(X_{n1})\right]^2 f_n(X_{n1})\} \int K(u)^2 du$. Notice that the marginal density $f_n(\cdot)$ of X_{n1} enters the definitions of both Λ_{HL} and Ω_{HL} because **HL**'s test is based on a density weighted moment conditions: $E[\epsilon_{nt} f_n(X_{nt})^{1/2} | X_{nt}] = 0$ a.s. under $\mathbb{H}_0$ [see the definition of J_n in (2.4)]. Clearly, we cannot make a direct comparison between the local power of our test and that of **HL**'s in general. In the special case where X_{nt} is uniformly distributed and $E[\Delta_n(X_{n1})] = 0$, we observe that

$$\Lambda_{HL} = \frac{1}{\text{vol}(\mathcal{X})} \lim_{n\to\infty} \text{Var}[\Delta_n(X_{n1})] \text{ and } \Omega_{HL} \equiv \frac{2}{[\text{vol}(\mathcal{X})]^2} \int K(u)^2 du \int_{\mathcal{X}} \left[v^2(x)\right]^2 dx.$$

If we further restrict our attention to cases where $p = 1$ and $K(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, then $\int K(u)^2 du = \frac{1}{2\sqrt{\pi}} \approx 0.2821$. In view of these and Remark 4, we have

$$\frac{\Lambda_2}{\sqrt{\Omega_2}} < \frac{\Lambda_1}{\sqrt{\Omega_1}} < \frac{\Lambda_{HL}}{\sqrt{\Omega_{HL}}} < \frac{\Lambda_0}{\sqrt{\Omega_0}},$$

which implies that in this case the asymptotic local power of **HL**'s test outperforms that of the local linear or quadratic R^2-based test, but is not as good as that of the local constant R^2-based test.

Remark 6. Motivated by the density-weighting idea in **HL**'s test, we can also consider a weighted version of nonparametric R^2-based test, where the weight is given by the kernel estimator of $f_n(x)$: $\hat{f}(x) = \frac{1}{n}\sum_{t=1}^n K_h(X_{nt} - x)$. We define the density weighted local constant R^2 statistic as[6]

$$R_{0,\hat{f}}^2 \equiv \frac{ESS_{0,\hat{f}}}{TSS_{\hat{f}}} \equiv \frac{\int_{\mathcal{X}_n} ESS_0(x)\hat{f}(x)\,dx}{\int_{\mathcal{X}_n} TSS(x)\hat{f}(x)\,dx},$$

and its normalized version as $T_{0n,\hat{f}} \equiv (nh^{p/2}R_{0,\hat{f}}^2 - \hat{B}_{0n,\hat{f}})/\sqrt{\hat{\Omega}_{0n}/(n^{-1}TSS_{\hat{f}})^2}$, where $\hat{B}_{0n,\hat{f}} \equiv h^{p/2} \sum_{t=1}^n \hat{\epsilon}_t^2 H_{0\hat{f},tt}^*/(n^{-1}TSS_{\hat{f}})$, $\hat{\Omega}_{0n,\hat{f}} \equiv 2n^{-2}h^p \sum_{s=1}^n \sum_{t\neq s}^n \hat{\epsilon}_t^2 \hat{\epsilon}_s^2 (nH_{0\hat{f},ts}^*)^2$, and $H_{0\hat{f},ts}^* \equiv n^{-1}\int_{\mathcal{X}_n} K_{tx}K_{sx}dx$. Then following the proofs of Theorems 3.1 and 3.3 closely, we can readily show that

$$P\left[T_{0n,\hat{f}} \geq z \mid \mathbb{H}_1(n^{-1/2}h^{-p/4})\right] \to 1 - \Phi(z - \Lambda_{0,f}/\sqrt{\Omega_{0,f}}) \text{ as } n \to \infty,$$

where $\Lambda_{0,f} \equiv \lim_{n\to\infty} E\{\text{Var}[\Delta_n(X_{n1})] f(X_{n1})\}$ and $\Omega_{0,f} \equiv 2\lim_{n\to\infty} E\{\left[v^2(X_{n1})\right]^2 f_n(X_{n1})\} \int [\int K(z) K(z+x)dz]^2 dx$. If we choose $K(z)$ as the product of the standard normal kernel, then the test $T_{0n,\hat{f}}$ has a larger asymptotic local power than **HL**'s test provided $E[\Delta_n(X_{n1})] = 0$. In this case, $\frac{\Lambda_{0,f}}{\sqrt{\Omega_{0,f}}} = 2^{p/4}\frac{\Lambda_{HL}}{\sqrt{\Omega_{HL}}}$.

3.4 Consistency

The following theorem establishes the consistency of the test.

Theorem 3.4 *Suppose Assumptions A1–A6 hold. Let* $\bar{\Lambda}_q \equiv \mathbb{B}_q' \mathbb{S}_q^{-1} \mathbb{B}_q \lim_{n\to\infty} Var[m_n(X_{n1})]$. *Then under* $\mathbb{H}_1$, $T_{qn}/\left(nh^{p/2}\right) = \bar{\Lambda}_q/(\sigma_V^2 \sqrt{\bar{\Omega}_q}) + o_p(1)$ *where* $\bar{\Omega}_q$ *is the probability limit of* $\hat{\Omega}_{qn}$ *under* $\mathbb{H}_1$.

[6] Given the good asymptotic local power property of the local constant test, we focus on the case $q = 0$ only.

Remark 7. Theorem 3.4 implies that under $\mathbb{H}_1$, $P(T_{qn} > t_n) \to 1$ as $n \to \infty$ for any sequence $t_n = o\left(nh^{p/2}\right)$, thus establishing the global consistency of the test. Even though we only focus on the case of parametric conditional mean model, we can also allow it to be nonparametrically specified. In this case, we can apply the local polynomial method to estimate the unknown but smooth conditional mean function and apply the resulting nonparametric residuals to conduct the nonparametric R^2 test. Following Su and Ullah (2009), we conjecture that the first-stage nonparametric estimation error only plays asymptotically negligible role in the asymptotic distributions of our nonparametric R^2 test statistic.

3.5 A bootstrap version of the test

Despite the asymptotic pivotal property of many nonparametric tests, early studies have shown that their empirical levels are typically sensitive to the choice of bandwidth, and may be highly distorted in finite samples. Therefore we propose a bootstrap method to obtain the bootstrap approximation to the finite sample distribution of our test statistic under the null. As Neumann and Paparoditis (2000) stressed, in order to get an asymptotically correct estimator of the null distribution of T_{qn}, it is not necessary to reproduce the whole dependence structure of the stochastic processes generating the original sample. Based on this observation, we propose a fixed-regressor bootstrap method in the spirit of Hansen (2000), which is quite different from that of **HL** who tried to mimic the data generating process (DGP) when X_{nt} or Z_{nt} contains lagged dependent variables.

For the ease of exposition we consider a nonlinear regression model $Y_{nt} = g(Z_{nt}, \theta_0) + U_{nt}$, where θ_0 can be estimated consistently via the NLS method. We propose to generate the bootstrap version of our test statistic T_{qn} as follows: 1) Obtain the NLS residuals $\hat{U}_t = Y_{nt} - g(Z_{nt}, \hat{\theta})$, where $\hat{\theta}$ is the NLS estimator of θ_0. 2) For $t = 1, \cdots, n$, obtain the bootstrap error U_{nt}^* by random sampling with replacement from $\{\hat{U}_s - \overline{\hat{U}},\ s = 1, \cdots, n\}$, where $\overline{\hat{U}} \equiv n^{-1} \sum_{s=1}^{n} \hat{U}_s$. Generate the bootstrap analog of Y_{nt} by holding Z_{nt} as fixed: $Y_{nt}^* = g(Z_{nt}, \hat{\theta}) + U_{nt}^*$, $t = 1, \cdots, n$. 3) Regress Y_{nt}^* on Z_{nt} to obtain the NLS estimator $\hat{\theta}^*$ of $\hat{\theta}$. Compute the bootstrap residuals $\hat{U}_t^* = Y_{nt}^* - g(Z_{nt}, \hat{\theta}^*)$. 4) Compute the bootstrap test statistic $T_{qn}^* = (nR_q^{*2} - \hat{B}_{qn}^*)/\sqrt{\hat{\Omega}_{qn}^* / (n^{-1}TSS^*)^2}$, where R_q^{*2}, $\hat{B}_{qn}^*$, $\hat{\Omega}_{qn}^*$ and TSS^* are defined analogously to R_q^2, $\hat{B}_{qn}$, $\hat{\Omega}_{qn}$ and TSS but with $\hat{U}_t$ being replaced by $\hat{U}_t^*$. 5) Repeat Steps 2-4 for B times and index the bootstrap statistics as $\{T_{qn,b}^*\}_{b=1}^{B}$. The bootstrap p-value is calculated by $p^* \equiv B^{-1} \sum_{b=1}^{B} 1(T_{qn,b}^* > T_{qn})$, where $1(\cdot)$ is the usual indicator function.

Several facts are worth mentioning here: (i) Conditionally on the original sample $\mathcal{W}_n \equiv \{(Y_{nt}, Z_{nt}, X_{nt}),$ $t = 1, \cdots, n\}$, the bootstrap replicates U_{nt}^* are IID with mean 0 and variance $\hat{\sigma}^2 \equiv n^{-1} \sum_{s=1}^{n} (\hat{U}_s - \overline{\hat{U}})^2$; (ii) the regressor Z_{nt} (resp. X_{nt}) can contain lags of Y_{nt} $\left(\text{resp. } (Y_{nt}, U_{nt}^2)\right)$, but the above bootstrap procedure does not need to mimic the DGP of either Y_{nt} or U_{nt}^2; (iii) the null hypothesis of conditional homoskedasticity is implicitly imposed in the above procedure.

To show that the bootstrap statistic T_{qn}^* can be used to approximate the asymptotic null distribution of T_{qn}, we rely on the notion of *convergence in distribution in probability* [see, e.g., Li, Hsiao and Zinn (2003, p. 307)], which generalizes the usual convergence in distribution to allow for conditional (i.e. random) distribution functions. We choose to work with the concept convergence in distribution in probability instead of convergence in distribution almost surely (with probability one) because the almost sure result is more difficult to establish given the complicated form of our test statistic. As Li, Hsiao and Zinn (2003) remarked, one can also describe the weak convergence in probability of the bootstrap test statistic using the dual bounded Lipschitz metric on probability measures as in Giné and Zinn (1990, Section 3), but their definition is easier to understand.

The following theorem establishes the validity of the above bootstrap procedure.

Theorem 3.5 *Suppose Assumptions A1-A6 hold. Let* z_α^* *be the* α*-level bootstrap critical value based on* B *bootstrap resamples.*[7] *Then* T_{qn}^* *converges to* $N(0,1)$ *in distribution in probability,* $\lim_{n\to\infty} P\left(T_{qn} \geq z_\alpha^*\right) = \alpha$ *under* $\mathbb{H}_0$, $\lim_{n\to\infty} P\left(T_{qn} \geq z_\alpha^*\right) \to 1-\Phi(z_\alpha - \Lambda_q/\sqrt{\Omega_q})$ *under* $\mathbb{H}_1(n^{-1/2}h^{-p/4})$, *and* $\lim_{n\to\infty} P(T_{qn} \geq z_\alpha^*) = 1$ *under* $\mathbb{H}_1$, *where* z_α *denotes the* $100(1-\alpha)$*th percentile of the standard normal distribution.*

Remark 8. The first two parts of Theorem 3.5 indicate that the bootstrap provides an asymptotic valid approximation to the null limit distribution of T_{qn}. The last two parts imply that the T_{qn} tests based upon the bootstrap critical values are consistent against both the designated local alternatives and all global alternative for which $P\left[E\left(U_{nt}^2|X_{nt}\right) = \sigma_0^2\right] < 1$ for any $\sigma_0^2 \in \mathbb{R}^+$.

4 Simulations

4.1 Data generating processes

We use the following two data generating processes (DGPs) in the level study:
DGP 1: $Y_{nt} = 1 + Z_{nt} + U_{nt}$,
DGP 2: $Y_{nt} = 0.5Y_{n,t-1} + U_{nt}$,
where U_{nt} are IID $N(0,1)$ and Z_{nt} are IID $U(-\sqrt{3},\sqrt{3})$ and independent of Z_{nt} in DGP 1. We choose $X_{nt} = Z_{nt}$ in DGP 1 and $X_{nt} = Z_{nt} = Y_{n,t-1}$ in DGP 2.

The following four DGPs are used in the power study:
DGP 3: $Y_{nt} = 1 + Z_{nt} + \sigma_{nt}\eta_t$,
DGP 4: $Y_{nt} = 1 + Z_{nt} + \sigma_{nt}\eta_t$,
DGP 5: $Y_{nt} = 0.5Y_{n,t-1} + \sigma_{nt}\eta_t$,
DGP 6: $Y_{nt} = 0.5Y_{n,t-1} + \sigma_{nt}\eta_t$,
where Z_{nt} are generated as in DGP 1, η_t are IID $N(0,1)$, $\sigma_{nt}^2 = 0.5 + \gamma_n\left(Z_{nt}-1\right)^2$, $0.2 + \gamma_n e^{Z_{nt}}$, $0.1 + 5e^{-\gamma_n Y_{n,t-1}^2}$, and $0.1 + 4\gamma_n U_{n,t-1}^2$ in DGPs 3, 4, 5 and 6, respectively, and $U_{n,t} = Y_{nt} - 0.5Y_{n,t-1}$. We choose $X_{nt} = Z_{nt}$ in DGPs 3-4, $X_{nt} = Z_{nt} = Y_{n,t-1}$ in DGPs 5, and $X_{nt} = U_{n,t-1}$ in DGP 6. To eliminate the starting-up effect, we throw away the first 200 observations when generating the data in DGPs 2, 5 and 6. In view of the fact that $n^{-1/2}h^{-1/4} \propto n^{-9/20}$ if $h \propto n^{-1/5}$, we set $\gamma_n = n^{-9/20}$ in DGPs 3-6 and study the behavior of various tests under $\mathbb{H}_1\left(\gamma_n\right)$. In addition, DGP 6 specifies an AR-ARCH process where we replace $U_{n,t-1}$ by $\hat{U}_{n,t-1} \equiv Y_{n,t-1} - \hat{\beta}_0 - \hat{\beta}_1 Y_{n,t-2}$ to construct tests for conditional heteroskedasticity, and $(\hat{\beta}_0, \hat{\beta}_1)$ is the OLS coefficient estimator in the linear regression of Y_{nt} on $(1, Y_{n,t-1})$.

4.2 Test statistics, kernel, and bandwidth choice

For each DGP, we regress Y_{nt} on $(1, Z_{nt})$ to obtain the residuals $\hat{U}_t$. Based on $\hat{V}_t \equiv \hat{U}_t^2$, we construct seven test statistics. The first one is the Lagrange multiplier (LM) test that tests $\alpha_1 = 0$ in the following parametric regression $\hat{V}_t = \alpha_0 + \alpha_1 X_{nt}^2 + \zeta_t$, where here and below ζ_t are error terms that may change across regressions. The second one is White's (1980) nR^2 test that tests $\alpha_1 = \alpha_2 = 0$ in the parametric regression $\hat{V}_t = \alpha_0 + \alpha_1 X_{nt} + \alpha_2 X_{nt}^2 + \zeta_t$. The third one is Hsiao and Li's (2001) nonparametric test defined in (3.3). The fourth through the sixth are our nonparametric R_0^2, R_1^2, and R_2^2 tests that are based on the local constant, local linear, and local quadratic regressions, respectively. The last one is our density weighted $R_{0,\hat{f}}^2$ test. To save space, we will use LM, $White$, HL, NR_0^2, NR_1^2, NR_2^2, and $NR_{0,f}^2$ to denote these seven tests in order.

[7] Namely, z_α^* is the $1-\alpha$ quantile of the empirical distribution of $\{T_{qn,b}^*\}_{b=1}^B$.

Implementing the last five nonparametric tests requires the choice of both kernel function and bandwidth sequence. To make a fair comparison between **HL**'s test and our nonparametric R^2 tests for different orders of local polynomials, we need to choose the *same* kernel function and bandwidth sequence in all five nonparametric tests. This rules out the choice of regression-based data-driven bandwidth obtained by using the least squares cross-validation (LSCV) method because different orders of local polynomial regressions would yield different "optimal" bandwidths to minimize the associated LSCV criterion functions. In this paper we apply the standard normal PDF as the kernel function and choose the *common* bandwidth sequence by the "rule of thumb" (ROT): $h = cs_X n^{-1/5}$, where s_X is the sample standard deviation of $\{X_{nt}\}$ and $c = 0.5$, 1 and 1.5. The performance of these nonparametric tests under different values of c suggests their sensitivity to the choice of bandwidth.[8]

4.3 Test results

Tables 1-2 report the simulation results based on 1000 replications. To obtain the simulated p-values, we use 200 bootstrap resamples in each replication for both **HL**'s and our tests. To implement our test, we choose $\mathcal{X}_n = [q_{n,0.01}, q_{n,0.99}]$ where $q_{n,\alpha}$ denotes the αth sample quantile of $\{X_{nt}, t = 1, ..., n\}$.[9] Table 1 reports the empirical rejection frequencies of the tests at the 5% nominal level when the null hypothesis holds true. It shows that the empirical levels of both parametric tests (LM, $White$) and nonparametric tests (HL, NR_0^2, NR_1^2, NR_2^2, $NR_{0,f}^2$) are reasonably well behaved despite the fact that the LM test tends to be undersized. In addition, these nonparametric tests are not very sensitive to the choice of the bandwidths sequence as far as the empirical level is concerned.

Table 1: Finite sample rejection frequency under the null (DGPs 1-2, nomial level: 0.05)

DGP	Tests \ n	50			100			200		
		$c = 0.5$	$c = 1$	$c = 1.5$	$c = 0.5$	$c = 1$	$c = 1.5$	$c = 0.5$	$c = 1$	$c = 1.5$
1	LM	0.036			0.041			0.047		
	$White$	0.051			0.047			0.051		
	HL	0.060	0.055	0.054	0.056	0.043	0.043	0.065	0.056	0.055
	NR_0^2	0.054	0.056	0.064	0.048	0.045	0.044	0.056	0.058	0.058
	NR_1^2	0.053	0.046	0.051	0.056	0.052	0.045	0.065	0.050	0.052
	NR_2^2	0.057	0.052	0.046	0.060	0.046	0.051	0.070	0.052	0.047
	$NR_{0,f}^2$	0.055	0.056	0.066	0.052	0.041	0.042	0.062	0.060	0.056
2	LM	0.037			0.030			0.035		
	$White$	0.035			0.043			0.050		
	HL	0.073	0.065	0.064	0.045	0.047	0.046	0.055	0.066	0.063
	NR_0^2	0.058	0.057	0.057	0.038	0.047	0.052	0.053	0.057	0.055
	NR_1^2	0.062	0.061	0.060	0.032	0.042	0.050	0.045	0.056	0.060
	NR_2^2	0.054	0.062	0.060	0.040	0.036	0.047	0.047	0.062	0.059
	$NR_{0,f}^2$	0.071	0.067	0.061	0.043	0.045	0.053	0.058	0.063	0.057

Note: We set the bandwidth $h = cs_X n^{-1/5}$ for all nonparametric tests. The LM and $White$ tests have nothing to do with the choice of h or c.

[8] We conjecture that one can follow Horowitz and Spokoiny (2001) and Chen and Gao (2007) and prove the rate-optimality for **HL**'s and our nonparametric R^2 tests. If this is the case, then in practice one can choose the bandwidth as in these papers. Alternatively, one can choose the bandwidth by minimizing certain criterion function, but the LSCV-based choice of bandwidth is designed mainly for the estimation problem. In principle, one can develop a data-driven choice of bandwidth for our testing problem, but this is beyond the scope of the paper and we leave it for the future research.

[9] We conjecture that the asymptotic results in the paper continue to hold in this case at the cost of more complex arguments due to the usual $n^{-1/2}$-rate of convergence of the sample quantiles to the population quantiles. See also Appendix B in the supplementary material.

Table 2: Finite sample rejection frequency under the alternative (DGPs 3-6, nominal level: 0.05)

DGP	Tests \ n	50			100			200		
		$c=0.5$	$c=1$	$c=1.5$	$c=0.5$	$c=1$	$c=1.5$	$c=0.5$	$c=1$	$c=1.5$
3	LM	0.094			0.145			0.187		
	$White$	0.351			0.535			0.671		
	HL	0.299	0.374	0.417	0.396	0.508	0.577	0.507	0.632	0.678
	NR_0^2	0.334	0.415	0.448	0.460	0.566	0.618	0.578	0.677	0.713
	NR_1^2	0.246	0.342	0.411	0.343	0.472	0.554	0.456	0.603	0.660
	NR_2^2	0.211	0.286	0.328	0.276	0.387	0.455	0.387	0.504	0.590
	$NR_{0,f}^2$	0.340	0.422	0.442	0.458	0.570	0.608	0.579	0.666	0.713
4	LM	0.113			0.192			0.251		
	$White$	0.453			0.675			0.852		
	HL	0.348	0.450	0.494	0.524	0.645	0.691	0.730	0.803	0.843
	NR_0^2	0.419	0.492	0.533	0.596	0.685	0.721	0.774	0.836	0.869
	NR_1^2	0.307	0.419	0.478	0.459	0.599	0.670	0.667	0.782	0.822
	NR_2^2	0.244	0.350	0.412	0.393	0.511	0.584	0.585	0.725	0.768
	$NR_{0,f}^2$	0.402	0.475	0.515	0.581	0.674	0.721	0.772	0.832	0.862
5	LM	0.424			0.823			0.978		
	$White$	0.180			0.554			0.913		
	HL	0.240	0.278	0.257	0.369	0.490	0.524	0.524	0.671	0.739
	NR_0^2	0.369	0.382	0.203	0.638	0.688	0.584	0.822	0.876	0.847
	NR_1^2	0.300	0.434	0.447	0.524	0.701	0.735	0.712	0.861	0.897
	NR_2^2	0.256	0.376	0.444	0.410	0.616	0.713	0.621	0.803	0.869
	$NR_{0,f}^2$	0.250	0.251	0.165	0.431	0.511	0.465	0.603	0.728	0.741
6	LM	0.475			0.685			0.856		
	$White$	0.500			0.690			0.826		
	HL	0.542	0.578	0.589	0.646	0.706	0.724	0.695	0.781	0.811
	NR_0^2	0.441	0.456	0.441	0.541	0.578	0.573	0.632	0.706	0.724
	NR_1^2	0.382	0.446	0.461	0.444	0.546	0.588	0.497	0.650	0.735
	NR_2^2	0.357	0.399	0.462	0.390	0.488	0.558	0.440	0.576	0.672
	$NR_{0,f}^2$	0.573	0.591	0.569	0.681	0.725	0.713	0.744	0.805	0.817

Table 2 reports the empirical power for the seven tests at the 5% nominal level. We summarize some important findings from Table 2 as follows. First, for all tests, the empirical power increases reasonably fast as the sample size doubles or quadruples. Second, even if the two parametric tests (LM, $White$) are not consistent tests, they tend to have power to detect various deviations from the null despite the fact that their powers may be significantly lower than the nonparametric tests; see, e.g., the LM test in DGPs 3 and 4. Third, consider DGPs 3-4 where the deviation from the null is of *local* nature and at the rate $nh^{-1/4}$ and X_{nt} is *uniformly distributed*. Our discussions in Remarks 4-5 indicate that the asymptotic local powers of the nonparametric tests can be ordered when these tests are constructed by using the *same* bandwidth sequence and kernel function: $NR_0^2 > HL > NR_1^2 > NR_2^2$ and $NR_{0,f}^2 > HL$, where $a > b$ signifies that a outperforms b in terms of the asymptotic local power. In addition, NR_0^2 and $NR_{0,f}^2$ are asymptotically equivalent in terms of local power in this case. Apparently, the results in Table 2 are largely consistent with these theoretical predictions. Fourth, consider DGP 5 where the deviation from the null is also at the rate $nh^{-1/4}$ but X_{nt} is *not* uniformly distributed anymore. This is the case where our theory predicts $NR_{0,f}^2 > HL$ and $NR_0^2 > NR_1^2 > NR_2^2$. Interestingly, for small sample sizes ($n = 50$), these predictions are not necessarily true. But as n increases, we do observe more chances for these predictions to occur. In particular, when $n = 200$, we observe that in terms of empirical power and for all choices of bandwidth, $NR_{0,f}^2$ outperforms HL and NR_1^2 outperforms NR_2^2. NR_0^2 also outperform NR_1^2 in cases where $c = 0.5$ and 1. Fifth, consider DGP 6 where we have an AR-ARCH specification and the two parametric tests specify the correct functional form and are expected to outperform the nonparametric tests. This is verified when $n = 200$. In addition, we observe the general pattern that $NR_{0,f}^2 > HL$ and $NR_0^2 > NR_1^2 > NR_2^2$, as predicted.[10]

5 Concluding Remarks

In this paper we propose a nonparametric R^2-based test for conditional heteroskedasticity which is applicable to both IID and time series observations. We demonstrate that after being suitably normalized, the nonparametric R^2 is asymptotically normally distributed under the null hypothesis and a sequence of Pitman local alternatives and is consistent against all kinds of conditional heteroskedasticity. We also propose a bootstrap method and justify its validity. Simulations demonstrate that our test complements that of **HL** and behaves well in finite samples.

We believe that the nonparametric R^2 is useful in many other aspects. For example, it can be used to test for serial correlation of unknown form among the error terms in both parametric and nonparametric regression models. Also it can be used to test linear or nonlinear restrictions on the derivatives of nonparametric functions. We leave these for future research.

Appendix

A Proof of the Main Results

Recall $D_h \equiv$diag$(1, h\mathbf{1}'_{N_{q1}} \cdots, h^q \mathbf{1}'_{N_{qq}})$, $\hat{\mathbf{v}} \equiv (\hat{V}_1, \cdots, \hat{V}_1)'$, and $K_{ix} \equiv K_h(X_{ni} - x)$. Let $\mathbf{u} \equiv (U_{n1}, \cdots, U_{nn})'$, $\boldsymbol{\varepsilon} \equiv (\varepsilon_{n1}, \cdots, \varepsilon_{nn})'$, $\mathbf{m} \equiv (m_n(X_{n1}), \cdots, m_n(X_{nn}))'$, $\boldsymbol{\Delta} \equiv (\Delta_n(X_{n1}), \cdots, \Delta_n(X_{nn}))'$, $\hat{\mathbf{g}} \equiv (g(Z_{n1}, \hat{\theta}), \cdots, g(Z_{nn}, \hat{\theta}))'$, and $\mathbf{g} \equiv (g(Z_{n1}, \theta_0), \cdots, g(Z_{nn}, \theta_0))'$. Let $\mathbf{0}_n$ and $\mathbf{1}_n$ denote an n-vector of zeros and ones, respectively. In this appendix, we first state a lemma that is used in the proof of the main results in Section 3, and then prove Theorems 3.1-3.5. The proof of the lemma can be found at http://www.mysmu.edu/faculty/ljsu/Publications/hetero_supp.pdf.

Lemma A.1 *Suppose Assumptions A1-A2 and A5-A6 hold. Let* $S_{qn}(x) \equiv n^{-1} D_h^{-1} \mathbf{X}'_{q,x} \mathbf{W}_x \mathbf{X}_{q,x} D_h^{-1}$, $\bar{S}_{qn}(x) \equiv E[S_{qn}(x)]$, $\gamma_n \equiv n^{-1/2} h^{-p/4}$, *and* $\varsigma_{ij} \equiv \Delta_n(X_{ni}) \varepsilon_{nj} + \Delta_n(X_{nj}) \varepsilon_{ni}$. *Then*

(i) $R_{n1} \equiv 2n^{-1} h^{p/2} \sum_{1 \le i < j \le n} \varepsilon_{ni} \varepsilon_{nj} \int_{\mathcal{X}_n} K_{ix} X'_{q,ix} D_h^{-1} \left[S_{qn}^{-1}(x) - \bar{S}_{qn}^{-1}(x)\right] D_h^{-1} X_{q,jx} K_{jx} dx = o_p(1)$,

(ii) $R_{n2} \equiv 2n^{-1} h^{p/2} \gamma_n \sum_{1 \le i < j \le n} \varsigma_{ij} \int_{\mathcal{X}_n} K_{ix} X'_{q,ix} D_h^{-1} \bar{S}_{qn}^{-1}(x) D_h^{-1} X_{q,jx} K_{jx} dx = o_p(1)$,

(iii) $R_{n3} \equiv 2n^{-1} h^{p/2} \gamma_n \sum_{1 \le i < j \le n} \varsigma_{ij} \int_{\mathcal{X}_n} K_{ix} X'_{q,ix} D_h^{-1} \left[S_{qn}^{-1}(x) - \bar{S}_{qn}^{-1}(x)\right] D_h^{-1} X_{q,jx} K_{jx} dx = o_p(1)$,

(iv) $R_{n4} \equiv 2n^{-2} h^{p} \sum_{1 \le i \ne j \le n} \varepsilon_{ni}^2 \varepsilon_{nj}^2 \{\int_{\mathcal{X}_n} K_{ix} X'_{q,ix} D_h^{-1} S_{qn}^{-1}(x) D_h^{-1} X_{q,jx} K_{jx} dx\}^2 = \Omega_q + o_p(1)$.

Proof of Theorem 3.1

Noting that

$$\begin{aligned} \hat{V}_i &= \hat{U}_i^2 = [U_{ni} + (\hat{U}_i - U_{ni})]^2 = U_{ni}^2 + (\hat{U}_i - U_{ni})^2 + 2(\hat{U}_i - U_{ni}) U_{ni} \\ &= m_n(X_{ni}) + \varepsilon_{ni} + [g(Z_{ni}, \hat{\theta}) - g(Z_{ni}, \theta_0)]^2 - 2[g(Z_{ni}, \hat{\theta}) - g(Z_{ni}, \theta_0)] U_{ni}, \end{aligned} \tag{A.1}$$

we have

$$ESS_q = \hat{\mathbf{v}}' M H^* M \hat{\mathbf{v}} = A_1 + A_2 + A_3 + 4A_4 + 2A_5 + 2A_6 - 4A_7 + 2A_8 - 4A_9 - 4A_{10}, \tag{A.2}$$

[10] Even though not reported here, we find that the NR_2^2 test may have larger power against fixed alternatives than the NR_0^2 and NR_1^2 tests.

where $A_1 \equiv \mathbf{m}' M H_q^* M \mathbf{m}$, $A_2 \equiv \boldsymbol{\varepsilon}' M H_q^* M \boldsymbol{\varepsilon}$, $A_3 \equiv [(\hat{\mathbf{g}} - \mathbf{g}) \odot (\hat{\mathbf{g}} - \mathbf{g})]' M H_q^* M [(\hat{\mathbf{g}} - \mathbf{g}) \odot (\hat{\mathbf{g}} - \mathbf{g})]$, $A_4 \equiv [(\hat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}]' M H_q^* M [(\hat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}]$, $A_5 \equiv \mathbf{m}' M H_q^* M \boldsymbol{\varepsilon}$, $A_6 \equiv \mathbf{m}' M H_q^* M [(\hat{\mathbf{g}} - \mathbf{g}) \odot (\hat{\mathbf{g}} - \mathbf{g})]$, $A_7 \equiv \mathbf{m}' M H_q^* M [(\hat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}]$, $A_8 \equiv \boldsymbol{\varepsilon}' M H_q^* M [(\hat{\mathbf{g}} - \mathbf{g}) \odot (\hat{\mathbf{g}} - \mathbf{g})]$, $A_9 \equiv \boldsymbol{\varepsilon}' M H_q^* M [(\hat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}]$, $A_{10} \equiv [(\hat{\mathbf{g}} - \mathbf{g}) \odot (\hat{\mathbf{g}} - \mathbf{g})] M H_q^* M \, [(\hat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}]$, and $\odot$ denotes the Hadamard product. Under $\mathbb{H}_0$, $\mathbf{m} = \sigma_0^2 \mathbf{1}_n$. It follows that $A_s = 0$ for $s = 1, 5, 6,$ and 7 as $M\mathbf{1}_n = \mathbf{0}_n$ and M is symmetric. It suffices to prove the theorem by showing that (i) $\bar{A}_2 \equiv h^{p/2} A_2 - h^{p/2} \sum_{i=1}^n \varepsilon_{ni}^2 H_{q,ii}^* \xrightarrow{d} N(0, \Omega_q)$, (ii) $h^{p/2} A_s = o_p(1)$ for $s = 3, 4, 8, 9, 10$, and (iii) $n^{-1} TSS = \sigma_V^2 + o_p(1)$.

We first show (i). In view of the symmetry of L and H_q^*, we have

$$\bar{A}_2 = \left(h^{p/2} \boldsymbol{\varepsilon}' H_q^* \boldsymbol{\varepsilon} - h^{p/2} \sum_{i=1}^n \varepsilon_{ni}^2 H_{q,ii}^* \right) + h^{p/2} \boldsymbol{\varepsilon}' L H_q^* L \boldsymbol{\varepsilon} - \mathbf{2} h^{p/2} \boldsymbol{\varepsilon}' L H_q^* \boldsymbol{\varepsilon} \equiv \bar{A}_{21} + \bar{A}_{22} - 2\bar{A}_{23}, \text{ say.}$$

By Lemma A.1(i),

$$\begin{aligned} \bar{A}_{21} &= 2n^{-1} h^{p/2} \sum_{1 \le i < j \le n} \varepsilon_{ni} \varepsilon_{nj} \bar{H}_{q,ij}^* + 2h^{p/2} \sum_{1 \le i < j \le n} \varepsilon_{ni} \varepsilon_{nj} \left(H_{q,ij}^* - n^{-1} \bar{H}_{q,ij}^* \right) \\ &= 2n^{-1} h^{p/2} \sum_{1 \le i < j \le n} \varepsilon_{ni} \varepsilon_{nj} \int_{\mathcal{X}_n} K_{ix} X_{q,ix}' D_h^{-1} \bar{S}_{qn}(x)^{-1} D_h^{-1} X_{q,jx} K_{jx} dx + o_p(1) \\ &\equiv A_{21} + o_p(1), \text{ say,} \end{aligned} \tag{A.3}$$

where $\bar{H}_{q,ij}^* \equiv \int_{\mathcal{X}_n} K_{ix} X_{q,ix}' D_h^{-1} \bar{S}_{qn}(x)^{-1} D_h^{-1} X_{q,jx} K_{jx} dx$, $A_{21} \equiv \frac{2}{n} \sum_{1 \le i < j \le n} \varphi_n(\xi_i, \xi_j)$, $\varphi_n(\xi_{ni}, \xi_{nj}) \equiv h^{p/2} \varepsilon_{ni} \varepsilon_{nj} \int_{\mathcal{X}_n} K_{ix} X_{q,ix}' D_h^{-1} \bar{S}_{qn}(x)^{-1} D_h^{-1} X_{q,jx} K_{jx} dx$, and $\xi_{ni} \equiv (X_{ni}', \varepsilon_{ni})'$. Note that A_{21} is a second order degenerate U-statistic. Under Assumptions A1-A2 and A5-A6, one can verify that the conditions of Theorem A.1 in Gao (2007) are satisfied so that a central limit theorem applies to $\bar{A}_{21}$.[11] (The geometric mixing rate in the theorem can be relaxed to our requirement on the mixing rate in Assumption A1.) Its asymptotic variance is given by $\lim_{n \to \infty} 2E_i E_j \left[\varphi_n(\xi_{ni}, \xi_{nj})^2 \right] = \int \left[\int K(z) \mu_q(z)' \mathbb{S}_q^{-1} \mu_q(z + x) K(z + x) dz \right]^2 dx \int_{\mathcal{X}} \left[v^2(\tilde{x}) \right]^2 d\tilde{x} = \Omega_q$, where $\chi_{ij,h} \equiv (X_{ni} - X_{nj})/h$, E_i denotes expectation with respect to ξ_{ni}, and we use the fact that $\bar{S}_{qn}(x) = \mathbb{S}_q^{-1} f_n(x)$ for any x on the interior of the support $\mathbb{X}_n$ of f_n. This, together with (A.3), implies that $\bar{A}_{21} \xrightarrow{d} N(0, \Omega_q)$. Observing that $L = n^{-1} \mathbf{1}_n \mathbf{1}_n'$, by straightforward moment calculations and the Davydov inequality, we have $\bar{A}_{22} = h^{p/2} \left(n^{-1/2} \boldsymbol{\varepsilon}' \mathbf{1}_n \right)^2 \left(n^{-1} \mathbf{1}_n' H_q^* \mathbf{1}_n \right) = h^{p/2} O_p(1) O_p\left(n^{-1} h^{-p} + h^p \right) = O_p\left(n^{-1} h^{-p/2} + h^{p/2} \right) = o_p(1)$ and $\bar{A}_{23} = h^{p/2} \left(n^{-1/2} \boldsymbol{\varepsilon}' \mathbf{1}_n \right) \left(n^{-3/2} \mathbf{1}_n' H_q^* \boldsymbol{\varepsilon} \right) = h^{p/2} O_p(1) O_p(n^{-3/2} h^{-2p} + 1) = O_p\left(n^{-3/2} h^{-3p/2} + h^{p/2} \right) = o_p(1)$. Consequently, $\bar{A}_2 \equiv \bar{A}_{21} + o_p(1) \xrightarrow{d} N(0, \Omega_q)$.

We now show (ii). By the fact that $|\mathrm{tr}(B_1 B_2)| \le \lambda_{\max}(B_1) \mathrm{tr}(B_2)$ for symmetric B_1 and p.s.d. B_2 (e.g., Bernstein, 2005, Fact 8.10.16), the repeated use of the rotation property of the trace operator, and the fact that $\lambda_{\max}(M) = 1$, we can show that $h^{p/2} A_3 \le h^{p/2} [(\hat{\mathbf{g}} - \mathbf{g}) \odot (\hat{\mathbf{g}} - \mathbf{g})]' H_q^* [(\hat{\mathbf{g}} - \mathbf{g}) \odot (\hat{\mathbf{g}} - \mathbf{g})] \equiv \bar{A}_3$ and $h^{p/2} A_4 \le h^{p/2} [(\hat{\mathbf{g}} - \mathbf{g}) \mathbf{u}]' H_q^* [(\hat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}] \equiv \bar{A}_4$. By White and Domowitz (1984), $\hat{\theta} - \theta_0 = O_p\left(n^{-1/2} \right)$ under Assumptions A1, A2(ii) and A3. Noting that the elements of H_q^* are uniformly $O_p\left(n^{-1} h^{-p} \right)$, by Assumption A3(iii) and the Markov inequality we have

$$\bar{A}_3 = h^{p/2} \sum_{i=1}^n \delta_{g,i}^4 H_{q,ii}^* + h^{p/2} \sum_{i=1}^n \delta_{g,i}^2 \sum_{j \ne i}^n H_{q,ij}^* \delta_{g,j}^2$$

[11] Even though Theorem A.1 of Gao (2007) was stated for the strong mixing sequences, a close examination of its proof indicates that it also holds true for our triangular array strong mixing processes. The main reason is that the proof mainly relies on some inequalities for strong mixing processes and the central limit theorem for martingale difference sequences, and both can be applied to triangular array processes.

$$\leq O_p\left(n^{-1}h^{-p/2}\right)\sum_{i=1}^{n}\|G_1(Z_{ni})\|^4\left\|\hat{\theta}-\theta_0\right\|^4+O_p\left(n^{-1}h^{-p/2}\right)\left\{\sum_{i=1}^{n}\|G_1(Z_{ni})\|^2\right\}^2\left\|\hat{\theta}-\theta_0\right\|^4$$
$$= O_p\left(n^{-2}h^{-p/2}\right)+O_p\left(n^{-1}h^{-p/2}\right)=o_p(1),$$

where $\delta_{g,i}\equiv g(Z_{ni},\hat{\theta})-g(Z_{ni},\theta_0)$. Noting that $\delta_{g,i}=(\nabla g(Z_{ni},\theta_0))'(\hat{\theta}-\theta_0)+\frac{1}{2}(\hat{\theta}-\theta_0)'\nabla^2 g(Z_{ni},\tilde{\theta})(\hat{\theta}-\theta_0)$ where $\tilde{\theta}$ lies between $\hat{\theta}$ and θ_0 elementwise, we have

$$\begin{aligned}\bar{A}_4 &= h^{p/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{g,i}U_{ni}H^*_{q,ij}\delta_{g,j}U_{nj}\\ &= h^{p/2}(\hat{\theta}-\theta_0)'\sum_{i=1}^{n}\sum_{j=1}^{n}\nabla g(Z_{ni},\theta_0)U_{ni}U_{nj}H^*_{q,ij}\nabla g(Z_{nj},\theta_0)'(\hat{\theta}-\theta_0)\\ &\quad+\frac{1}{4}h^{p/2}(\hat{\theta}-\theta_0)'\sum_{i=1}^{n}\sum_{j=1}^{n}\nabla^2 g(Z_{ni},\tilde{\theta})(\hat{\theta}-\theta_0)U_{ni}U_{nj}H^*_{q,ij}(\hat{\theta}-\theta_0)'\nabla^2 g(Z_{nj},\tilde{\theta})(\hat{\theta}-\theta_0)\\ &\quad+h^{p/2}(\hat{\theta}-\theta_0)'\sum_{i=1}^{n}\sum_{j=1}^{n}\nabla g(Z_{ni},\theta_0)U_{ni}U_{nj}H^*_{q,ij}(\hat{\theta}-\theta_0)'\nabla^2 g(Z_{nj},\tilde{\theta})(\hat{\theta}-\theta_0)\equiv\bar{A}_{41}+\bar{A}_{42}+\bar{A}_{43}.\end{aligned}$$

For the first term, we have $\bar{A}_{41}=(\hat{\theta}-\theta_0)'(\bar{A}_{41a}+\bar{A}_{41b})(\hat{\theta}-\theta_0)$, where $\bar{A}_{41a}=h^{p/2}\sum_{i=1}^{n}U_{ni}^2\nabla g(Z_{ni},\theta_0)H^*_{q,ii}\nabla g(Z_{ni},\theta_0)'$ and $\bar{A}_{41b}=2h^{p/2}\sum_{1\leq i<j\leq n}U_{ni}U_{nj}\nabla g(Z_{ni},\theta_0)H^*_{q,ij}\nabla g(Z_{nj},\theta_0)$. It is easy to show that $\bar{A}_{41a}=O_p\left(h^{-p/2}\right)$ and $\bar{A}_{41b}=O_p(1)$, implying that $\bar{A}_{41}=O_p\left(n^{-1}h^{-p/2}\right)$. For $\bar{A}_{42}$, we have $\bar{A}_{42}\leq\frac{1}{4}h^{p/2}\|\hat{\theta}-\theta_0\|^4\{\sum_{i=1}^{n}\|G_2(Z_{ni})\|\,\|U_{ni}\|\}^2\max_{1\leq i,j\leq n}|H^*_{q,ij}|=O_p(n^{-1}h^{-p/2})$. Similarly, $\bar{A}_{43}=O_p(n^{-1/2}h^{-p/2})$. It follows that $\bar{A}_4=O_p(n^{-1/2}h^{-p/2})$.

Letting $\boldsymbol{\delta}_g\equiv(\hat{\mathbf{g}}-\mathbf{g})\odot(\hat{\mathbf{g}}-\mathbf{g})$, write $h^{p/2}A_8=h^{p/2}\boldsymbol{\varepsilon}'H_q^*\boldsymbol{\delta}_g+h^{p/2}\boldsymbol{\varepsilon}'LH_q^*\boldsymbol{\delta}_g+h^{p/2}\boldsymbol{\varepsilon}'H_q^*L\boldsymbol{\delta}_g+h^{p/2}\boldsymbol{\varepsilon}'LH_q^*L\boldsymbol{\delta}_g\equiv\bar{A}_{81}+\bar{A}_{82}+\bar{A}_{83}+\bar{A}_{84}$. Further decompose $\bar{A}_{81}$ as follows $\bar{A}_{81}=h^{p/2}\sum_{i=1}^{n}\delta_{g,i}^2\varepsilon_{ni}H^*_{q,ii}+h^{p/2}\sum_{i=1}^{n}\varepsilon_{ni}\sum_{j\neq i}^{n}H^*_{q,ij}\delta_{g,j}^2\equiv\bar{A}_{81a}+\bar{A}_{81b}$. By Taylor expansions and Assumption A3, we can show that $\bar{A}_{81a}=O_p\left(n^{-1}h^{-p/2}\right)$, $\bar{A}_{81b}=O_p\left(n^{-1/2}\right)$, and hence $h^{p/2}\bar{A}_{81}=o_p(1)$. Similarly, we can show that $\bar{A}_{8s}=o_p(1)$ for $s=2,3,4$. Consequently $h^{p/2}A_8=o_p(1)$. By the same token, we can show that $h^{p/2}A_9=o_p(1)$. By the Cauchy-Schwarz inequality, $h^{p/2}A_{10}\leq\{h^{p/2}A_3\}^{1/2}\{h^{p/2}A_4\}^{1/2}=o_p(1)\,o_p(1)=o_p(1)$. This completes the proof of (ii).

We now show (iii). Using (A.1), Assumption A3, and the weak law of large numbers (WLLN) yields $n^{-1}\sum_{i=1}^{n}\hat{V}_i=n^{-1}\sum_{i=1}^{n}\hat{U}_i^2=n^{-1}\sum_{i=1}^{n}U_{ni}^2+O_p\left(n^{-1/2}\right)=E\left(U_{ni}^2\right)+O_p\left(n^{-1/2}\right)$. In view of this, we can readily show that under Assumptions A1, A2, A3, A5 and A6

$$\begin{aligned}n^{-1}TSS &= n^{-1}\sum_{i=1}^{n}(\hat{U}_i^2-\hat{\sigma}_0)^2\int_{\mathcal{X}_n}K_h(X_{ni}-x)\,dx\\ &= n^{-1}\sum_{i=1}^{n}\left[U_{ni}^2-E\left(U_{ni}^2\right)\right]^2\int_{\mathcal{X}_n}K_h(X_{ni}-x)\,dx+O_p\left(n^{-1/2}\right)=\sigma_V^2+O_p\left(n^{-1/2}\right).\end{aligned}$$

Consequently, $n^{-1}TSS=\sigma_V^2+o_p(1)$. Note that this holds regardless of whether $\mathbb{H}_0$, $\mathbb{H}_1(n^{-1/2}h^{-p/4})$, or $\mathbb{H}_1$ holds true. ■

Proof of Corollary 3.2

We prove the result under $\mathbb{H}_1(\gamma_n)$ with $\gamma_n=n^{-1/2}h^{-p/4}$. To show $\hat{B}_{qn}=B_{qn}+o_p(1)$, noticing that $n^{-1}TSS=\sigma_V^2+o_p(1)$ under $\mathbb{H}_1(\gamma_n)$, it suffices to show that $D_{1qn}\equiv n^{-1}TSS(\hat{B}_{qn}-B_{qn})=$

$h^{p/2}\sum_{i=1}^{n}\left(\hat{\epsilon}_i^2-\varepsilon_{ni}^2\right)H_{q,ii}^{*}=o_p\left(1\right).$ Using (A.1) and Assumption A3, under $\mathbb{H}_1(\gamma_n)$ we have

$$n^{-1}\sum_{i=1}^{n}\hat{V}_i=n^{-1}\sum_{i=1}^{n}\hat{U}_i^2=n^{-1}\sum_{i=1}^{n}\left[\sigma_0^2+\gamma_n\Delta_n\left(X_{ni}\right)\right]+O_p(n^{-1/2})=\sigma_0^2+O_p\left(\gamma_n\right). \tag{A.4}$$

In view of this and the fact that $H_{q,ii}^{*}=O\left(n^{-1}h^{-p}\right)$ uniformly in i, we can readily show that under Assumption 3 and $\mathbb{H}_1(\gamma_n)$,

$$\begin{aligned} D_{1qn} &= h^{p/2}\sum_{i=1}^{n}\left\{\left[\gamma_n\Delta_n\left(X_{ni}\right)+\varepsilon_{ni}+\delta_{ni}+O_p\left(\gamma_n\right)\right]^2-\varepsilon_{ni}^2\right\}\left(H_{q,ii}^{*}-n^{-1}\right) \\ &= O_p\left(\left(\gamma_n+n^{-1/2}\right)h^{-p/2}\right)=O_p\left(n^{-1/2}h^{-3p/4}\right)=o_p\left(1\right), \end{aligned}$$

where $\delta_{ni}=[g(Z_{ni},\hat{\theta})-g(Z_{ni},\theta_0)]^2-2[g(Z_{ni},\hat{\theta})-g(Z_{ni},\theta_0)]U_{ni}$.

Now, write $\hat{\Omega}_{qn}-\Omega_q=(\hat{\Omega}_{qn}-\Omega_{qn})+(\Omega_{qn}-\Omega_q)$, where $\Omega_{qn}\equiv 2n^{-2}h^p\sum_{i=1}^{n}\sum_{j\neq i}^{n}\varepsilon_{ni}^2\varepsilon_{nj}^2\left(nH_{ij}^{*}-1\right)^2$. For the second term, we can readily show that $\Omega_{qn}=2n^{-2}h^p\sum_{i=1}^{n}\sum_{j\neq i}^{n}\varepsilon_{ni}^2\varepsilon_{nj}^2(nH_{ij}^{*})^2+o_p\left(1\right)=2R_{n4}+o_p\left(1\right)=\Omega_q+o_p\left(1\right)$ by Lemma A.1(iv). For the first term, we have

$$\begin{aligned} \hat{\Omega}_{qn}-\Omega_{qn} &= 2n^{-2}h^p\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\hat{\epsilon}_i^2\hat{\epsilon}_j^2-\varepsilon_{ni}^2\varepsilon_{nj}^2\right)\left(nH_{ij}^{*}\right)^2 \\ &= 4n^{-2}h^p\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\hat{\epsilon}_i^2-\varepsilon_{ni}^2\right)\varepsilon_{nj}^2\left(nH_{ij}^{*}\right)^2+2n^{-2}h^p\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\hat{\epsilon}_i^2-\varepsilon_{ni}^2\right)\left(\hat{\epsilon}_j^2-\varepsilon_{nj}^2\right)\left(nH_{ij}^{*}\right)^2 \\ &\equiv 4D_{2qn}+2D_{3qn},\ \text{say.} \end{aligned}$$

In view of (A.1) and (A.4), using Assumption A3 and the fact that $nH_{ij}^{*}=O_p\left(h^{-p}\right)$ uniformly in (i,j), we have

$$D_{2qn}=n^{-2}h^p\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\{\left[\gamma_n\Delta_n\left(X_{ni}\right)+\varepsilon_{ni}+\delta_{ni}+O_p\left(\gamma_n\right)\right]^2-\varepsilon_{ni}^2\right\}\varepsilon_{nj}^2\left(nH_{ij}^{*}\right)^2=O_p\left(\gamma_n\right)=o_p\left(1\right).$$

By the same token, $D_{3qn}=o_p\left(1\right).$ It follows that $\hat{\Omega}_{qn}-\Omega_{qn}=o_p\left(1\right)$ and $\hat{\Omega}_{qn}-\Omega_q=o_p\left(1\right).$ This, in conjunction with Theorem 3.1, implies that $T_{qn}\xrightarrow{d}N\left(0,1\right)$ under $\mathbb{H}_0$. ■

Proof of Theorem 3.3

The proof follows closely from that of Theorem 3.1, now keeping the additional terms that do not vanish under $\mathbb{H}_1\left(\gamma_n\right)$ with $\gamma_n=n^{-1/2}h^{-p/4}$. In view of the fact that $\hat{B}_{qn}=B_{qn}+o_p\left(1\right)$ and $\hat{\Omega}_{qn}=\Omega_q$ $+o_p\left(1\right)$ under $\mathbb{H}_1\left(\gamma_n\right)$, it suffices to show that under $\mathbb{H}_1\left(\gamma_n\right)$, (i) $h^{p/2}A_1\xrightarrow{p}\Lambda_q$ and (ii) $h^{p/2}A_s=o_p\left(1\right)$ for $s=5,6,7$, where A's are defined after (A.2).

We first show (i). Write $h^{p/2}A_1=h^{p/2}\mathbf{m}'H_q^{*}\mathbf{m}+h^{p/2}\mathbf{m}'LH_q^{*}L\mathbf{m}-2h^{p/2}\mathbf{m}'LH_q^{*}\mathbf{m}\equiv\bar{A}_{11}+\bar{A}_{12}-2\bar{A}_{13}$. Under $\mathbb{H}_1\left(\gamma_n\right)$, $m_n\left(x\right)=\sigma_0^2+\gamma_n\Delta_n\left(x\right)$ and we have

$$\bar{A}_{11}=n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_n\left(X_{ni}\right)\Delta_n\left(X_{nj}\right)\bar{H}_{q,ij}^{*}+n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_n\left(X_{ni}\right)\Delta_n\left(X_{nj}\right)\left(H_{q,ij}^{*}-n^{-1}\bar{H}_{q,ij}^{*}\right)\equiv\bar{A}_{11a}+\bar{A}_{11b}.$$

It is straightforward to show that $\bar{A}_{11b}=o_p\left(1\right).$ For $\bar{A}_{11a}$, by the Fubini theorem, the WLLN, and Assumptions A1, A2(iv) and A5-A6, we have

$$\begin{aligned} \bar{A}_{11a} &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Delta_n\left(X_{ni}\right)\Delta_n\left(X_{nj}\right)\int_{\mathcal{X}_n}K_{ix}X_{q,ix}'D_h^{-1}\bar{S}_{qn}\left(x\right)^{-1}D_h^{-1}X_{q,jx}K_{jx}dx \\ &= \int_{\mathcal{X}_n}\psi_n\left(x\right)'\mathbb{S}_q^{-1}\psi_n\left(x\right)f_n\left(x\right)^{-1}dx=\mathbb{B}_q'\mathbb{S}_q^{-1}\mathbb{B}_q\int_{\mathcal{X}_n}\Delta_n^2\left(x\right)f_n\left(x\right)dx+o_p\left(1\right) \end{aligned}$$

where $\psi_n(x)' = n^{-1}\sum_{i=1}^n \Delta_n(X_{ni}) K_{ix} X'_{q,ix} D_h^{-1}$. By the same token, we can show that $\bar{A}_{12} = \mathbb{B}'_q \mathbb{S}_q^{-1} \mathbb{B}_q \{E[\Delta_n(X_{ni})]\}^2 + o_p(1)$, and $\bar{A}_{13} = \mathbb{B}'_q \mathbb{S}_q^{-1} \mathbb{B}_q \{E[\Delta_n(X_{ni})]\}^2 + o_p(1)$. Consequently $h^{p/2} A_1 = \Lambda_q + o_p(1)$. This proves (i).

Next, we show (ii). We first write $h^{p/2} A_5 = h^{p/2}\mathbf{m}' H_q^* \boldsymbol{\varepsilon} + h^{p/2}\mathbf{m}' L H_q^* L \boldsymbol{\varepsilon} - h^{p/2}\mathbf{m}' L H_q^* \boldsymbol{\varepsilon} - h^{p/2}\mathbf{m}' H_q^* L \boldsymbol{\varepsilon} \equiv \bar{A}_{51} + \bar{A}_{52} - \bar{A}_{53} - \bar{A}_{54}$. We further decompose $\bar{A}_{51}$ as follows

$$
\begin{aligned}
\bar{A}_{51} &= n^{-1} h^{p/2} \gamma_n \sum_{1 \le i \ne j \le n} \left[\Delta_n(X_{ni}) \varepsilon_{nj} + \Delta_n(X_{nj}) \varepsilon_{ni}\right] \bar{H}^*_{q,ij} \\
&\quad + h^{p/2} \gamma_n \sum_{1 \le i \ne j \le n} \Delta_n(X_{ni}) \varepsilon_{nj} \left(H^*_{q,ij} - n^{-1} \bar{H}^*_{q,ij}\right) + h^{p/2} \gamma_n \sum_{i=1}^n \Delta_n(X_{ni}) \varepsilon_{ni} H^*_{q,ii} \\
&\equiv \bar{A}_{51a} + \bar{A}_{51b} + \bar{A}_{51c}, \text{ say.}
\end{aligned}
$$

By Lemmas A.1(ii) and (iii), $\bar{A}_{51a} = R_{n2} = o_p(1)$ and $\bar{A}_{51b} = R_{n3} = o_p(1)$. For $\bar{A}_{51c}$, we have $|\bar{A}_{51c}| = h^{p/2}\gamma_n \max_i n H^*_{q,ii} \{n^{-1}\sum_{i=1}^n |\Delta_n(X_{ni})| \varepsilon_{ni}\} = O_p(n^{-1/2}\ h^{-3p/4}) = o_p(1)$. Consequently $\bar{A}_{51} = o_p(1)$. Similarly, we can show that $\bar{A}_{5s} = o_p(1)$ for $s = 2, 3, 4$. It follows that $h^{p/2} A_5 = o_p(1)$. In view of the above analysis of A_1 and that of A_3 and A_4 in the proof of Theorem 3.1, the Cauchy-Schwarz inequality yields $h^{p/2} A_6 \le \{h^{p/2} A_1\}^{1/2} \{h^{p/2} A_3\}^{1/2} = O_p(1) o_p(1) = o_p(1)$ and $h^{p/2} A_7 \le \{h^{p/2} A_1\}^{1/2} \{h^{p/2} A_4\}^{1/2} = O_p(1) o_p(1) = o_p(1)$. Consequently, $P(T_{qn} \ge z | \mathbb{H}_1(n^{-1/2} h^{-p/4})) \to 1 - \Phi(z - \Lambda_q/\sqrt{\Omega_q})$. This concludes the proof of the theorem. ■

Proof of Theorem 3.4

The proof follows closely from that of Theorems 3.1 and 3.3. By (A.2) and the proof of Theorem 3.1, $ESS_q = A_1 + 2A_5 + 2A_6 - 4A_7 + o_p(h^{-p/2})$. Following the determination of the probability order of $h^{p/2} A_s$ ($s = 5, 6, 7$) in the proof of Theorem 3.2, we can readily show that $n^{-1} A_s = o_p\left(n^{-1} h^{-p/2}\right) = o_p(1)$ under $\mathbb{H}_1$ for $s = 5, 6, 7$. Under $\mathbb{H}_1$, by the Fubini theorem, the WLLN, and Assumptions A1, A2(iv), and A4-A6, we have

$$
\begin{aligned}
n^{-1} A_1 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \delta_m(X_{ni}) \delta_m(X_{nj}) \int_{\mathcal{X}_n} K_{ix} X'_{q,ix} D_h^{-1} \bar{S}_{qn}^{-1} D_h^{-1} X_{q,jx} K_{jx} dx + o_p(1) \\
&= \int_{\mathcal{X}_n} \bar{\psi}_n(x)' \mathbb{S}_q^{-1} \bar{\psi}_n(x) f_n^{-1}(x) dx + o_p(1) = \int_{\mathcal{X}_n} \delta_m(x)^2 \mathbb{B}'_q \mathbb{S}_q^{-1} \mathbb{B}_q f_n(x) dx + o_p(1) \\
&= \bar{\Lambda}_q + o_p(1),
\end{aligned}
$$

where $\delta_m(X_{ni}) \equiv m_n(X_{ni}) - E[m_n(X_{ni})]$, $\bar{\psi}_n(x)' = \frac{1}{n}\sum_{i=1}^n \delta_m(X_{ni}) K_{ix} X'_{q,ix} D_h^{-1}$. Also, $n^{-1} TSS = \sigma_V^2 + o_p(1)$ under $\mathbb{H}_1$. It follows that $R_q^2 = n^{-1} ESS_q / \left(n^{-1} TSS\right) = \bar{\Lambda}_q / \sigma_V^2 + o_p(1)$. Under $\mathbb{H}_1$, we have $\left(nh^{p/2}\right)^{-1} \hat{B}_{qn} = o_p(1)$ and $\hat{\Omega}_{qn} \xrightarrow{p} \bar{\Omega}_q$. It follows that $\left(nh^{p/2}\right)^{-1} T_{qn} = [R_q^2 - \left(nh^{p/2}\right)^{-1} \hat{B}_{qn}] / \sqrt{\hat{\Omega}_{qn}} = \bar{\Lambda}_q / (\sigma_V^2 \sqrt{\bar{\Omega}_q}) + o_p(1)$. ■

Proof of Theorem 3.5

The idea underlying the proof of the first part of the theorem is simple. In order to show that the bootstrap test statistic T^*_{qn} converges to $N(0,1)$ in distribution in probability, we only need to verify that certain conditions hold in probability, which implies that for any subsequence there is a further subsequence that those conditions hold almost surely.

Let P^* denote the probability conditional on the original sample $\mathcal{W}_n$. Let $E^*(\cdot)$ and $\mathrm{Var}^*(\cdot)$ denote the expectation and variance with respect to P^*. $a_n = o_{p^*}(1)$ denotes that $P^*(|a_n| \ge \epsilon) = o_p(1)$ for any positive $\epsilon > 0$ as $n \to \infty$. The notation $O_{p^*}(1)$ is similarly defined. Let $\varepsilon^*_{ni} \equiv U^{*2}_{ni} - \hat{\sigma}^2$ and $\hat{\epsilon}^*_i \equiv \hat{U}^{*2}_i - \hat{\sigma}^2$, where $\hat{\sigma}^2 \equiv E^*\left(U^{*2}_{ni}\right) = n^{-1}\sum_{i=1}^n (\hat{U}_i - \overline{\hat{U}})^2$ and $\overline{\hat{U}} \equiv n^{-1}\sum_{i=1}^n \hat{U}_i$. Let $\hat{\mathbf{g}}^* \equiv (g(Z_{n1}, \hat{\theta}^*), \cdots, g(Z_{nn}, \hat{\theta}^*))'$,

$\mathbf{u}^* \equiv (U_{n1}^{*2}, \cdots, U_{nn}^{*2})'$, $\boldsymbol{\varepsilon}^* \equiv (\varepsilon_{n1}^*, \cdots, \varepsilon_{nn}^*)'$, and $\hat{\mathbf{v}}^* \equiv (\hat{U}_1^{*2}, \cdots, \hat{U}_n^{*2})'$. Write $\hat{U}_i^{*2} = [U_{ni}^* + (\hat{U}_i^* - U_{ni}^*)]^2 = U_{ni}^{*2} + (\hat{U}_i^* - U_{ni}^*)^2 + 2(\hat{U}_i^* - U_{ni}^*)U_{ni}^* = \hat{\sigma}^2 + \varepsilon_{ni}^* + [g(Z_{ni}, \hat{\theta}^*) - g(Z_{ni}, \hat{\theta})]^2 - 2[g(Z_{ni}, \hat{\theta}^*) - g(Z_{ni}, \hat{\theta})]U_{ni}^*$. By the symmetry of M and the fact that $M\mathbf{1}_n = \mathbf{0}_n$, we have

$$ESS_q^* \equiv n^{-1}\hat{\mathbf{v}}^{*\prime} M H_q^* M \hat{\mathbf{v}}^* = A_2^* + A_3^* + 4A_4^* + 2A_8^* - 4A_9^* - 4A_{10}^*, \qquad \text{(A.5)}$$

where A^*'s are the bootstrap analogue of A's. Let $\hat{V}_i^* \equiv \hat{U}_i^{*2}, \overline{\hat{V}}^* \equiv n^{-1}\sum_{i=1}^n \hat{V}_i^*$, and $TSS^* \equiv \sum_{i=1}^n (\hat{V}_i^* - \overline{\hat{V}}^*)^2 \int_{\mathcal{X}_n} K_h(X_{ni} - x)\,dx$. We can show that $n^{-1}TSS^* = n^{-1}\sum_{i=1}^n E^*(\hat{V}_i^* - \overline{\hat{V}}^*)^2 \int_{\mathcal{X}_n} K_h(X_{ni} - x)dx + o_{p^*}(1) = n^{-1}\sum_{i=1}^n (\hat{V}_i - \overline{\hat{V}})^2 \int_{\mathcal{X}_n} K_h(X_{ni} - x)\,dx + o_{p^*}(1) = \sigma_V^2 + o_{p^*}(1)$. We prove the first part of the theorem by showing that: (i) $[h^{p/2}A_2^* - h^{p/2}\sum_{i=1}^n \varepsilon_{ni}^{*2} H_{q,ii}^* \,/\sqrt{\hat{\Omega}_{qn}^*} \xrightarrow{d} N(0,1)$, (ii) $\hat{B}_{qn}^* = h^{p/2}\sum_{i=1}^n \varepsilon_{ni}^{*2} H_{q,ii}^* + o_p(1)$, and (iii) $h^{p/2}A_s^* = o_p(1)$ for $s = 3, 4, 8, 9, 10$.

We first show (i). Analogously to the proof of (i) in the proof of Theorem 3.1, we have $h^{p/2}A_2^* = 2h^{p/2}\sum_{1\le i<j\le n} \varepsilon_{ni}^*\varepsilon_{nj}^* H_{q,ij}^* + h^{p/2}\sum_{i=1}^n \varepsilon_{ni}^{*2} H_{q,ii}^* \equiv A_{21}^* + A_{22}^*$, say. Let $v^{*2} \equiv \text{Var}^*\left(U_{ni}^{*2}\right)$. Noting that A_{21}^* is a second order degenerate U_n-statistic and ε_{ni}^*'s are independent conditional on the data, we can apply the CLT for second order degenerate U-statistic with independent but nonidentically distributed (INID) observations (e.g., De Jong, 1987) and conclude that conditional on the data, $A_{21}^* \xrightarrow{d} N\left(0, \Omega_q^*\right)$, where $\Omega_q^* \equiv \text{plim}_{n\to\infty} \frac{2h^p}{n^2}\sum_{i=1}^n \sum_{j\ne i}^n E^*[\varepsilon_{ni}^{*2}\varepsilon_{nj}^{*2}(\int_{\mathcal{X}_n} K_{ix} X_{q,ix}' D_h^{-1} S_{qn}^{-1}(x) D_h^{-1} X_{q,jx} K_{jx} dx)^2] = 2v^{*4}\text{vol}(\mathcal{X}) \int [\int K(z)\,\mu_q(z)' \mathbb{S}_q^{-1}\mu_q(z+x)\,K(z+x)\,dz]^2 dx$ and $\text{vol}(\mathcal{X}) \equiv \int_{\mathcal{X}} dx$. (i) follows as one can easily show that $\hat{\Omega}_{qn}^* = \Omega_q^* + o_p(1)$. Next, $\hat{B}_{qn}^* - h^{p/2}\sum_{i=1}^n \varepsilon_{ni}^{*2} H_{q,ii}^* = h^{p/2}\sum_{i=1}^n \left(\hat{\epsilon}_i^{*2} - \varepsilon_{ni}^{*2}\right) H_{q,ii}^* = O_{p^*}\left(n^{-1/2}h^{-p/2}\right) = o_{p^*}(1)$, proving (ii). Noting that $\hat{\theta}^* - \hat{\theta} = O_{p^*}(n^{-1/2})$ under our assumptions, the proof of (iii) is analogous to that of (ii) in the proof of Theorem 3.1 and thus omitted.

Recall z_α^* is the $1-\alpha$ quantile of the empirical distribution of $\{T_{qn,b}^*\}_{b=1}^B$. Let $\bar{z}_\alpha^*$ denote the $1-\alpha$ conditional quantile of T_{qn}^* given $\mathcal{W}_n$, i.e., $P\left(T_{qn}^* \ge \bar{z}_\alpha^* | \mathcal{W}_n\right) = \alpha$. By choosing B sufficiently large, the approximation error of z_α^* to $\bar{z}_\alpha^*$ can be made arbitrarily small and negligible. By the first part of the theorem, $\bar{z}_\alpha^* \to z_\alpha$ in probability. Then in view of Theorem 3.1 and the remark after it, $\lim_{n\to\infty} P(T_{qn} \ge z_\alpha^*) = \lim_{n\to\infty} P(T_{qn} \ge z_\alpha) = \alpha$ under $\mathbb{H}_0$. By Theorem 3.3 and the fact that $\hat{B}_{qn} = B_{qn} + o_p(1)$, $\hat{\Omega}_{qn} = \Omega_q + o_p(1)$, and $n^{-1}TSS = \sigma_V^2 + o_p(1)$ under $\mathbb{H}_1(n^{-1/2}h^{-p/4})$, we have $\lim_{n\to\infty} P(T_{qn} \ge z_\alpha^*) = \lim_{n\to\infty} P(T_{qn} \ge z_\alpha) = 1 - \Phi(z_\alpha - \Lambda_q/\sqrt{\Omega_q})$ under $\mathbb{H}_1(n^{-1/2}h^{-p/4})$. Similarly, in view of Theorem 3.4, we have $\lim_{n\to\infty} P(T_{qn} \ge z_\alpha^*) = \lim_{n\to\infty} P(T_{qn} \ge z_\alpha) = 1$ under $\mathbb{H}_1$. ■

References

Andrews, D. W. K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560-596.

Bernstein, D. S. (2005) *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton.

Breusch, T. S. & A. R. Pagan (1979) A simple test for heteroskedasticity and random variation. *Econometrica* 47, 1287-1294.

Chen, S. X. & J. Gao (2007) An adaptive empirical likelihood test for parametric time series regression models. *Journal of Econometrics* 141, 950-972.

De Jong, P. (1987) A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261-277.

Gao, J. (2007) *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* Chapman & Hall/CRC, New York.

Giné, E. & J. Zinn (1990) Bootstrapping general empirical measures. *Annals of Probability* 18, 851-869.

Glejser, H. (1969) A new test for heteroskedasticity. *Journal of the American Statistical Association* 64, 316–323.

Godfrey, L. G. (1978) Testing for multiplicative heteroskedasticity. *Journal of Econometrics* 8, 227-236.

Goldfeld, S.M., & R. E. Quandt (1965) Some tests for homoskedasticity. *Journal of the American Statistical Association* 60, 539-547.

Greene, W. H. (2000) *Econometric Analysis.* 4th ed., Prentice Hall, New Jersey.

Hansen, B. E. (2000) Testing for structural change in conditional models. *Journal of Econometrics* 97, 93-115.

Hansen, B. E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726-748.

Hong, Y. (1993) Consistent testing for heteroskedasticity of unknown form. *Manuscript*, Dept. of Economics, Cornell Univ.

Horowitz, J. L. & V. G. Spokoiny (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599-631.

Hsiao, C. & Li, Q. (2001) A consistent test for conditional heteroskedasticity in time-series regression models. *Econometric Theory* 17, 188-221.

Huang, L-H. & J. Chen (2008) Analysis of variance, coefficient of determination and F-test for local polynomial regression. *Annals of Statistics* 36, 2085-2109.

Koenker, R. & G. Bassett (1982) Robust tests for heteroskedasticity based on quantiles. *Econometrica* 50, 159-171.

Li, D., Z. Lu & O. Linton (2011) Local linear fitting under near Epoch dependence: uniform consistency with convergence rates. *Econometric Theory*, forthcoming.

Li, Q., C. Hsiao, & J. Zinn (2003) Consistent specification tests for semiparametric/nonparametric models based on series estimation method. *Journal of Econometrics* 112, 295-325.

Masry, E., (1996) Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis* 17, 571-599.

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

Newey, W. & J. Powell (1987) Asymmetric least squares estimation and testing. *Econometrica* 55, 819-847.

Pagan, A. R. & Y. Pak (1993) Testing for heteroskedasticity. In G. S. Maddala, C. R. Rao, & H. D. Vinod (eds.), *Handbook of Statistics*, vol 11, pp. 489-518. Amsterdam: Elsevier Science Publishers.

Su, L. & A. Ullah (2009) Testing conditional uncorrelatedness. *Journal of Business and Economic Statistics* 27, 18-29.

Su, L. & H. White (2010) Testing structural change in partially linear models. *Econometric Theory* 26, 1761-1806.

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

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 452-475.

White, H. & I. Domowitz (1984) Nonlinear regression with dependent observations. *Econometrica* 52, 143-161.

Yoshihara, K. (1992) Limiting behavior of generalized quadratic forms generated by regular sequences III. *Yokohama Mathematical* 40, 1-9.

Zheng, X. (2006) Testing heteroscedasticity in nonlinear and nonparametric regressions with an application to interest rate volatility. *Manuscript*, School of Economics, Shanghai Jiaotong Univ.