10-2009

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

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Aman ULLAH
University of California, Riverside

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*

October 8, 2009

ABSTRACT

In this paper we propose a nonparametric test for conditional heteroskedasticity based on a new measure of nonparametric goodness-of-fit (R^2). In analogy with the ANOVA tools for classical linear regression models, the nonparametric R^2 is obtained for the local polynomial regression of the residuals from a parametric regression on some covariates. It is close to 0 under the null hypothesis of conditional homoskedasticity and stays away from 0 otherwise. Unlike most popular parametric tests in the literature, the new test does not require the correct specification of parametric conditional heteroskedasticity form and thus is able to detect all kinds of conditional heteroskedasticity of unknown form. We show that after being appropriately centered and standardized, the nonparametric R^2 is asymptotically normally distributed under the null hypothesis of conditional homoskedasticity and a sequence of Pitman local alternatives. We also prove the consistency of the test, propose a bootstrap method to obtain the critical values or bootstrap p-values, and justify the validity of the bootstrap method. We conduct a small set of Monte Carlo simulations and compare our test with some popular parametric and nonparametric tests in the literature. Applications to the U.S. real GDP growth rate data indicate that our nonparametric test can reveal certain conditional heteroskedasticity which the parametric tests fail to detect.

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

*The first author gratefully acknowledges the financial support from the NSFC under the grant numbers 70501001 and 70601001. The second author acknowledges the financial support from the academic senate, UCR.

1 Introduction

Since the 1960s there has developed a large literature on testing for heteroskedasticity. Most of the early tests for heteroskedasticity can be classified into three categories: those that are based on the Lagrange multiplier (LM) principle, those that are based on the least squares residuals, and those that are based on quantiles or expectiles. For example, the classical tests of Glejser (1969), Godfrey (1978), and Breusch and Pagan (1979), are among the first category; the popular tests of Goldfeld and Quandt (1965), Bickel (1978), and White (1980) are among the second group, and the robust tests of Koenker and Bassett (1982) and Newey and Powell (1987) belong to the third category. For a survey on early methods of testing heteroskedasticity, see Pagan and Pak (1993).

As shown by Pagan and Pak (1993) most of the early tests can be regarded as special cases of the conditional moment tests that are, unfortunately, not robust against functional misspecification. Hong (1993) also demonstrated that many of these existing tests are not consistent in that they are unable to detect certain forms of heteroskedasticity asymptotically. For this reason, several nonparametric consistent tests for heteroskedasticity have been proposed, including Hong (1993), Hsiao and Li (2001), and Zheng (2006). Based on the comparison between the kernel estimator of the conditional variance of a regression model under the alternative and the estimator of the unconditional variance under the null, Hong (1993) proposed a consistent test for heteroskedasticity when the regressand and regressors are independent and identically distributed (IID). In contrast, Hsiao and Li's (2001) test is motivated by the application of heteroskedasticity test to time series models and the wide use of the ARCH type of models, and it is constructed by using the idea analogous to the consistent tests for model specification. Zheng's (2006) test for heteroskedasticity 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 the use of different approaches in the derivations.

In this paper, we propose a new test for conditional homoskedasticity based on a novel measure for nonparametric goodness-of-fit (R^2). Recently Huang and Chen (2008) have proposed a measure of goodness-of-fit for local polynomial regressions, which is based on the decomposition of the total sum of squares (TSS) into the explained sum of squares (ESS) and the residual sum of squares (RSS). Their definition of nonparametric R^2 is analogous to that of R^2 in multiple linear regression models. We think that this measure serves a useful statistic for testing many popular hypotheses in econometrics and statistics just as the important roles it 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 misspecification for the functional form in these auxiliary regressions, these tests may lead to misleading conclusions. To avoid such misspecification of functional form, we propose to adopt nonparametric models in place of parametric models in the auxiliary regressions. Then we can construct a nonparametric analogue of the parametric residual-based test by applying the nonparametric measure of goodness-of-fit.

In this paper, we focus on the case of testing for conditional homoskedasticity based upon the nonparametric R^2. It is a residual-based test. 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 is small and close to 0 under the null of conditional homoskedasticity and lies far away from 0 under the alternative of conditional heteroskedasticity. We show that after being properly standardized, it is asymptotically normally distributed under the null of conditional homoskedasticity and a sequence of Pitman local alternatives. We also establish the consistency of the test and propose a bootstrap method to obtain the bootstrap p-values. Simulations indicate that our test behaves reasonably well in finite samples.

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. We also establish the global consistency of our test and justify the validity of a bootstrap method. In Section 4 we conduct Monte Carlo experiments to evaluate the finite sample performance of our test in comparison with some other tests and apply them to the U.S. real GDP growth rate data. Section 5 concludes. All technical assumptions and proofs are relegated to the Appendix.

To proceed, we define some notation that will be used throughout the paper. For a matrix A, we denote its Euclidean norm as $\|A\| = [\mathrm{tr}\,(AA')]^{1/2}$, where $\mathrm{tr}(\cdot)$ and prime mean trace and transpose, respectively. For a vector $a \equiv (a_1, ..., a_l)'$, $\mathrm{diag}(a)$ denotes a diagonal matrix with a_i as its ith diagonal element. Let I_l denote an $l \times l$ identity matrix. $\mathbf{0}_l$ and $\mathbf{1}_l$ denote a l-vector of zeros and ones, respectively. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ convergence in distributions.

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 Hsiao and Li (2001), we consider a nonlinear model of the form

$$Y_t = g(Z_t, \theta_0) + U_t \tag{2.1}$$

where $g(\cdot, \cdot)$ is a function of known form, θ_0 is a $d \times 1$ vector of unknown parameters, Z_t is a $k \times 1$ vector of regressors, and U_t is a scalar error term such that $E(U_t|Z_t) = 0$ almost surely (a.s. hereafter). The null of interest is that conditional on X_t, a $p \times 1$ vector of variables, U_t's are homoskedastic, i.e.,

$$H_0 : P\left(E\left(U_t^2|X_t\right) = \sigma_0^2\right) = 1 \text{ where } \sigma_0^2 \equiv E\left(U_t^2\right) > 0. \tag{2.2}$$

The alternative hypothesis is

$$H_1 : P\left(E\left(U_t^2|X_t\right) = \sigma_0^2\right) < 1 \text{ for all } \sigma_0^2 \in \mathbb{R}^+. \tag{2.3}$$

Note that we allow the elements in X_t to be distinct from those in Z_t.

The consistent tests of Hong (1993), Hsiao and Li (2001), and Zheng (2006) are all residual-based tests that rely on the observation that $E\left(U_t^2 - \sigma_0^2|X_t\right) = 0$ a.s. under the null hypothesis. Below, we propose an alternative way to test for the above hypotheses by extending the use of R^2 from parametric regression models to nonparametric regression models and achieve consistency of the test at the same time.

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

Let $V_t \equiv U_t^2$ and $m(X_t) \equiv E(V_t|X_t)$. If V_t were observable, we could consider the nonparametric regression model

$$V_t = m(X_t) + \varepsilon_t \tag{2.4}$$

where $\varepsilon_t \equiv V_t - m(X_t)$. Under the null hypothesis of conditional homoskedasticity, we expect $m(X_t) = \sigma_0^2$ a.s., and thus 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 the nonparametric goodness-of-fit measure that was recently proposed by Huang and Chen (2008).

Let $\widehat{\theta}$ denote the nonlinear least squares (NLS) estimator of θ_0 in (2.1). Let $\widehat{U}_t \equiv Y_t - g\left(Z_t, \widehat{\theta}\right)$ and $\widehat{V}_t \equiv \widehat{U}_t^2$. A feasible regression model is given by

$$\widehat{V}_t = m(X_t) + e_t \tag{2.5}$$

where e_t is the new error term in the above regression. The basic idea of local polynomial fit is: if $m(x)$ is a smooth function of $x \equiv (x_1, \cdots, x_p)'$, for any X_t in a neighborhood of x, we have

$$\begin{aligned} m(X_t) &= m(x) + \sum_{1\le|\mathbf{j}|\le q} \frac{1}{\mathbf{j}!} D^{|\mathbf{j}|} m(x) (X_t - x)^{\mathbf{j}} + o(\|X_t - x\|^q) \\ &\equiv \sum_{0\le|\mathbf{j}|\le q} \beta_{\mathbf{j}}(x) (X_t - x)^{\mathbf{j}} + o(\|X_t - x\|^q). \end{aligned}$$

Here, we use the notation of Masry (1996): $\mathbf{j} = (j_1, ..., j_p)$, $|\mathbf{j}| = \sum_{i=1}^p j_i$, $x^{\mathbf{j}} = \mathbf{\Pi}_{i=1}^p x_i^{j_i}$, $\sum_{0\le|\mathbf{j}|\le q} = \sum_{k=0}^q \sum_{j_1=0}^k \cdots \sum_{j_p=0}^k$ (with $j_1+\cdots+j_p=k$), $D^{|\mathbf{j}|} m(x) = \frac{\partial^{|\mathbf{j}|} m(x)}{\partial^{j_1} x_1 \cdots \partial^{j_p} x_p}$, $\beta_{\mathbf{j}}(x) = \frac{1}{\mathbf{j}!} D^{|\mathbf{j}|} m(x)$, where $\mathbf{j}! \equiv \mathbf{\Pi}_{i=1}^p j_i!$. Thus, given observations $\left\{(\widehat{V}_t, X_t)\right\}_{t=1}^n$, the qth-order local-polynomial regression of $\widehat{V}_t$ on X_t is fitted by the weighted least squares (WLS) as follows

$$\min_{\boldsymbol{\beta}} n^{-1} \sum_{t=1}^n \left(\widehat{V}_t - \sum_{0\le|\mathbf{j}|\le q} \beta_{\mathbf{j}} (X_t - x)^{\mathbf{j}} \right)^2 K_h(X_t - x), \tag{2.6}$$

where $\boldsymbol{\beta}$ 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), $K_h(\cdot) \equiv K(\cdot/h)/h$, $K(\cdot)$ is a symmetric probability density function (PDF) on $\mathbb{R}^p$, and $h \equiv h(n)$ is a bandwidth parameter. Let $\widehat{\beta}_{\mathbf{j}}(x; h)$ $(0 \le |\mathbf{j}| \le q)$ denote the solution to the above problem. Based on the normal equations for the above regression, it is easy to verify the following local ANOVA decomposition of the total sum of squares (TSS)

$$TSS(x) = ESS_q(x) + RSS_q(x) \tag{2.7}$$

where

$$\begin{aligned} TSS(x) &\equiv \sum_{t=1}^n \left(\widehat{V}_t - \overline{\widehat{V}} \right)^2 K_h(X_t - x), \\ ESS_q(x) &\equiv \sum_{t=1}^n \left(\sum_{0\le|\mathbf{j}|\le q} \widehat{\beta}_{\mathbf{j}}(x; h) (X_t - x)^{\mathbf{j}} - \overline{\widehat{V}} \right)^2 K_h(X_t - x), \\ RSS_q(x) &\equiv \sum_{t=1}^n \left(\widehat{V}_t - \sum_{0\le|\mathbf{j}|\le q} \widehat{\beta}_{\mathbf{j}}(x; h) (X_t - x)^{\mathbf{j}} \right)^2 K_h(X_t - x), \end{aligned}$$

and $\overline{\widehat{V}} \equiv n^{-1} \sum_{t=1}^n \widehat{V}_t$. A global ANOVA decomposition of TSS is given by

$$TSS = ESS_q + RSS_q \tag{2.8}$$

where $TSS \equiv \int TSS(x)\, dx = n^{-1} \sum_{t=1}^n \left(\widehat{V}_t - \overline{\widehat{V}} \right)^2$, $ESS_q \equiv \int ESS_q(x)\, dx$, and $RSS_q \equiv \int RSS_q(x)\, dx$. Then one can define the nonparametric goodness-of-fit (R^2) for the above qth-order local polynomial regression as

$$R_q^2 = 1 - \frac{RSS_q}{TSS} = \frac{ESS_q}{TSS}. \tag{2.9}$$

For more interpretations of R_q^2 and its local version, we refer the readers to Huang and Chen (2008). It is worth mentioning that the typical choices of q are 1, 2 and 3. So we will focus on these three cases in the following sections.

Clearly R_q^2 lies between 0 and 1. The smaller value of R_q^2, the worse the fit is. In the extreme case, if no regressors among X_t can explain V_t, we expect a value close to 0 in any given sample of observations on $\left\{\widehat{V}_t, X_t\right\}$. Let $X_{tx} \equiv \mu\left(X_t - x\right)$ denote the stack of $(X_t - x)^{\mathbf{j}}$, $0 \leq |\mathbf{j}| \leq q$, in the lexicographical order. For example, $X_{tx} \equiv (1, (X_t - x)')'$ if $q = 1$. Let $\mathbf{X}_x \equiv (X_{1x}, \cdots X_{nx})'$, $\mathbf{W}_x \equiv \text{diag}(K_h\left(X_1 - x\right), \cdots, K_h\left(X_n - x\right))$, $\mathbf{H}_x \equiv \mathbf{W}_x\mathbf{X}_x\left(\mathbf{X}_x'\mathbf{W}_x\mathbf{X}_x\right)^{-1}\mathbf{X}_x'\mathbf{W}_x$, and $H^* \equiv \int \mathbf{H}_x dx$. It is easy to verify that

$$TSS = \widehat{\mathbf{v}}'M\widehat{\mathbf{v}},\ ESS_q = \widehat{\mathbf{v}}'\left(H^* - L\right)\widehat{\mathbf{v}},\ \text{and}\ RSS_q = \widehat{\mathbf{v}}'\left(I_n - H^*\right)\widehat{\mathbf{v}},$$

where $\widehat{\mathbf{v}} \equiv \left(\widehat{V}_1, \cdots, \widehat{V}_n\right)'$, $M \equiv I_n - L$, and L is an $n \times n$ matrix with entries $1/n$. Then the nonparametric R^2 can be written as

$$R_q^2 = \frac{\widehat{\mathbf{v}}'\left(H^* - L\right)\widehat{\mathbf{v}}}{\widehat{\mathbf{v}}'M\widehat{\mathbf{v}}}. \tag{2.10}$$

We will show that after being approximately centered and scaled, the above nonparametric R^2 is asymptotically normally distributed under the null and a sequence of Pitman local alternatives.

To proceed, we define some notation. Let $N_l = (l + q - 1)!/(l!(q - 1)!)$ be the number of distinct q-tuples $\mathbf{j}$ with $|\mathbf{j}| = l$. It denotes the number of distinct l-th order partial derivatives of $m(x)$ with respect to x. Arrange the N_l 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 $\mu_{\mathbf{j}} = \int_{\mathbb{R}^p} x^{\mathbf{j}} K(x) dx$, and define the $N \times N$ dimensional matrix $\mathbb{S}$ and $N \times 1$ vector $\mathbb{B}$, where $N = \sum_{l=0}^{q} N_l$, by

$$\mathbb{S} = \begin{bmatrix} \mathbb{S}_{0,0} & \mathbb{S}_{0,1} & \cdots & \mathbb{S}_{0,q} \\ \mathbb{S}_{1,0} & \mathbb{S}_{1,1} & \cdots & \mathbb{S}_{1,q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{S}_{q,0} & \mathbb{S}_{q,1} & \cdots & \mathbb{S}_{q,q} \end{bmatrix},\ \mathbb{B} = \begin{bmatrix} \mathbb{S}_{0,0} \\ \mathbb{S}_{1,0} \\ \vdots \\ \mathbb{S}_{q,0} \end{bmatrix}, \tag{2.11}$$

where $\mathbb{S}_{i,j}$ are $N_i \times N_j$ dimensional matrices whose (l, r) elements are $\mu_{\phi_i(l)+\phi_j(r)}$. That is, the elements of $\mathbb{S}$ and $\mathbb{B}$ are simply multivariate moments of the kernel K.

3 Asymptotic Distributions

In this section we first study the asymptotic distributions of nonparametric R^2 under the null hypothesis and a sequence of Pitman local alternatives. We then prove the consistency of the test and propose a bootstrap method to obtain bootstrap p-values.

3.1 Asymptotic null distribution

Let H^*_{ts} denote the (t,s)th element of H^*. Define

$$b_n \equiv h^{p/2}\sum_{t=1}^{n}\varepsilon_t^2\left(H^*_{tt}-n^{-1}\right)/\left(n^{-1}TSS\right),$$

$$\Omega_0 \equiv \int\left[\int K(z)\,\mu(z)'\,\mathbb{S}^{-1}\mu(z+x)\,K(z+x)\,dz\right]^2 dx\int\left[v^2(x)\right]^2 dx/\sigma_V^4$$

where $v^2(x)\equiv E\left(\varepsilon_t^2|X_t=x\right)$, $\mu(\cdot)$ is the function used in the definition of X_{tx} $(\equiv \mu(X_t-x))$, and $\sigma_V^2 \equiv$Var(V_t). Define

$$\Gamma_n \equiv nh^{p/2}R_q^2-b_n.$$

Theorem 3.1 *Suppose Assumptions A1-A6 in the Appendix hold and $p\leq 7$. Then under H_0,*

$$\Gamma_n \xrightarrow{d} N(0,\Omega_0).$$

Remark 1. 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 $n^{-1}TSS\cdot\Gamma_n=\sigma_V^2\overline{\Gamma}_n+o_p(1)$, where $\overline{\Gamma}_n\equiv\frac{2}{n}\sum_{1\leq t<s\leq n}\varphi(\xi_t,\xi_s)$, $\xi_t\equiv(X_t',\varepsilon_t)'$, $\varphi(\xi_t,\xi_s)\equiv h^{p/2}\varepsilon_t\varepsilon_s\left(\int K_{tx}X_{tx}'D_h^{-1}\mathbb{S}^{-1}D_h^{-1}X_{sx}K_{sx}f^{-1}(x)\,dx-1\right)$, and $D_h \equiv$diag$\left(1,h\mathbf{1}_{N_1}',\cdots,h^q\mathbf{1}_{N_q}'\right)$ denotes an $N\times N$ diagonal matrix with typical elements given by h^s, $s=0,1,\cdots,q$. Apparently $\overline{\Gamma}_n$ is a second-order U-statistic with symmetric kernel $\varphi(\cdot,\cdot)$. Then we can apply the central limit theorem (CLT) for second-order U-statistics under some martingale condition and demonstrate that $\overline{\Gamma}_n\xrightarrow{d}N(0,\sigma_V^4\Omega_0)$. The result then follows by noticing that $n^{-1}TSS=\sigma_V^2+o_p(1)$.

To implement the test, we require consistent estimates of both the bias term b_n and the variance Ω_0. Let $\widehat{\varepsilon}_t\equiv\widehat{V}_t-n^{-1}\sum_{s=1}^n\widehat{V}_s$. Define

$$\widehat{b}_n\equiv h^{p/2}\sum_{t=1}^{n}\widehat{\varepsilon}_t^2\left(H^*_{tt}-n^{-1}\right)/\left(n^{-1}TSS\right),\ \text{and}\ \widehat{\Omega}\equiv 2n^{-2}h^p\sum_{s=1}^{n}\sum_{t\neq s}^{n}\widehat{\varepsilon}_t^2\widehat{\varepsilon}_s^2\left(nH^*_{ts}-1\right)^2.$$

It is easy to show that $\widehat{b}_n=b_n+o_p(1)$ and $\widehat{\Omega}=\Omega_0+o_p(1)$ under H_0. We can define a feasible nonparametric R^2-based test statistic as

$$T_n=\left(nh^{p/2}R_q^2-\widehat{b}_n\right)/\sqrt{\widehat{\Omega}}. \tag{3.1}$$

We then compare T_n with the one-sided critical value z_α, i.e., the upper αth percentile from the standard normal distribution. We reject the null when $T_n>z_\alpha$ at the α significance level.

To examine the asymptotic local power of our test, we consider the following sequence of Pitman local alternatives:

$$H_1(\gamma_n): m(x)=\sigma_0^2+\gamma_n\Delta(x), \tag{3.2}$$

where $\gamma_n \to 0$ as $n \to \infty$ and $\Delta(x)$ is a nonconstant continuous function. Define

$$\Delta_0 \equiv \mathbb{B}'\mathbb{S}^{-1}\mathbb{B}E\left[\Delta^2(X_1)\right] - \left[E\left(\Delta(X_1)\right)\right]^2. \tag{3.3}$$

The following theorem establishes the local power property of our test.

Theorem 3.2 *Suppose Assumptions A1–A6 in the Appendix hold and* $p \leq 7$. *Suppose that* $\Delta(x)$ *is a continuous function such that* $E\left[\triangle^2(X_1)\right] < \infty$. *Then the local power of the test* T_n *satisfies* $P(T_n \geq z_\alpha | H_1(n^{-1/2}h^{-p/4})) \to 1 - \Phi(z_\alpha - \Delta_0/\sqrt{\Omega_0})$ *as* $n \to \infty$, *where* $\Phi(\cdot)$ *is the cumulative distribution function (CDF) of the standard normal.*

Remark 2. For the local linear and quadratic regressions ($q = 1$ and 2), it is straightforward to verify that $\mathbb{B}'\mathbb{S}^{-1}\mathbb{B} = 1$ and hence $\Delta_0 = \text{Var}(\Delta(X_1)) \geq c$ for some $c > 0$ under $H_1(n^{-1/2}h^{-p/4})$. If $q = 3$, by the formula for partitioned inverse and the symmetry of the kernel function $K(\cdot)$, we can show that $\mathbb{B}'\mathbb{S}^{-1}\mathbb{B} = (1-a)^3 + a \leq 1$, where $a \equiv \mathbb{S}_{0,2}\mathbb{S}_{2,2}^{-1}\mathbb{S}_{2,0} \leq 1$ by the Cauchy-Schwarz inequality. This implies that Δ_0 is no bigger than $\text{Var}(\Delta(X_1))$ in the case of $q = 3$.

Remark 3. Theorem 3.2 implies that the test has non-trivial asymptotic power against alternatives that diverge from the null at the rate $n^{-1/2}h^{-p/4}$. The power increases with the magnitude of $\Delta_0/\sqrt{\Omega_0}$. Furthermore, by taking a large bandwidth we can make the alternative magnitude (of order γ_n) against which the test has non-trivial power arbitrarily close to the parametric rate $n^{-1/2}$.

The following theorem establishes the consistency of the test for both local linear and quadratic regressions.

Theorem 3.3 *Suppose Assumptions A1–A6 in the Appendix hold and* $p \leq 7$. *Let* $\mu_0 \equiv \mathbb{B}'\mathbb{S}^{-1}\mathbb{B}E\left[m^2(X_1)\right] - \left[E\left(m(X_1)\right)\right]^2$. *Then under* H_1, $T_n/\left(nh^{p/2}\right) = \mu_0/(\sigma_V^2\sqrt{\overline{\Omega}}) + o_p(1)$ *where* $\overline{\Omega}$ *is the probability limit of* $\widehat{\Omega}$ *under* H_1.

Remark 4. Following Remark 2, in the case of local linear and quadratic regressions, $\mu_0 = \text{Var}(m(X_1)) \geq c > 0$ under H_1. This implies that under H_1, $P(T_n > t_n) \to 1$ as $n \to \infty$ for any sequence $t_n = o\left(nh^{p/2}\right)$, thus establishing the global consistency of the test.

Remark 5. 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.

3.2 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_n, 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 Hsiao and Li (2001) who tried to mimic the data generating process (DGP) when X_t or Z_t contains lagged dependent variables.

For the ease of exposition we consider a nonlinear regression model $Y_t = g(Z_t, \theta_0) + U_t$, where θ_0 can be estimated consistently via the nonlinear least squares (NLS) method. We propose to generate the bootstrap version of our test statistic T_n as follows:

1. Obtain the NLS residuals $\widehat{U}_t = Y_t - g(Z_t, \widehat{\theta})$, where $\widehat{\theta}$ is the NLS estimator of θ_0.

2. For $t = 1, \cdots, n$, obtain the bootstrap error U_t^* by random sampling with replacement from $\{\widehat{U}_s - \overline{\widehat{U}},\ s = 1, \cdots, n\}$, where $\overline{\widehat{U}} \equiv n^{-1} \sum_{s=1}^{n} \widehat{U}_s$. Generate the bootstrap analog of Y_t by holding Z_t as fixed: $Y_t^* = g(Z_t, \widehat{\theta}) + U_t^*$, $t = 1, \cdots, n$.

3. Regress Y_t^* on Z_t to obtain the NLS estimator $\widehat{\theta}^*$ of $\widehat{\theta}$. Compute the bootstrap residuals $\widehat{U}_t^* = Y_t^* - g(Z_t, \widehat{\theta}^*)$.

4. Let $\widehat{V}_t^* \equiv \widehat{U}_t^{*2}$. Calculate the nonparametric R^2 (denoted as R_q^{*2}) from the qth order local polynomial regression of $\widehat{V}_t^*$ on X_t. Compute the bootstrap test statistic $T_n^* = \left(nR_q^{*2} - \widehat{b}_n^*\right) / \sqrt{\widehat{\Omega}^*}$, where $\widehat{b}_n^*$ and $\widehat{\Omega}^*$ are defined analogously to $\widehat{b}_n$ and $\widehat{\Omega}$ but with $\widehat{U}_t$ being replaced by $\widehat{U}_t^*$.

5. Repeat Steps 2-4 for B times and index the bootstrap statistics as $\left\{T_{n,b}^*\right\}_{b=1}^{B}$. The bootstrap p-value is calculated by $p^* \equiv B^{-1} \sum_{b=1}^{B} 1\left(T_{n,b}^* > T_n\right)$, where $1(\cdot)$ is the usual indicator function.

Several facts are worth mentioning here: (i) Conditionally on the original sample $\mathcal{W} \equiv \{(Y_t, Z_t, X_t), t = 1, \cdots, n\}$, the bootstrap replicates U_t^* are independent and identically distributed (IID) with mean 0 and variance $n^{-1} \sum_{s=1}^{n} (\widehat{U}_s - \overline{\widehat{U}})^2$; (ii) the regressor Z_t (resp. X_t)

can contain lags of Y_t $\left(\text{resp. } Y_t, U_t^2\right)$, but the above bootstrap procedure does not need to mimic the DGP of either Y_t or U_t^2; (iii) the null hypothesis of conditional homoskedasticity is implicitly imposed in the above procedure.

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

Theorem 3.4 *Suppose Assumptions A1-A6 in the Appendix hold and* $p \leq 7$. *Then* $T_n^* \xrightarrow{d} N(0,1)$ *conditionally on* $\mathcal{W}$, *and* $P\left(T_n > T_n^*\right) \to 1$ *under* H_1.

Remark 6. The first part of Theorem 3.4 indicates that the bootstrap provides an asymptotic valid approximation to the null limit distribution of T_n. This holds as long as we generate the bootstrap data by imposing the null hypothesis. The second part of Theorem 3.4 implies that the test T_n based upon the bootstrap critical value is consistent against every global alternative for which $P\left(E\left(U_t^2|X_t\right) = \sigma_0^2\right) < 1$ for any $\sigma_0^2 \in \mathbb{R}^+$. That is, $T_n \to \infty$ with probability approaching 1 under H_1.

4 Monte Carlo Simulation Study and Applications

In this section, we first conduct Monte Carlo simulations to evaluate the finite sample performance of our test in comparison with other tests and then apply these tests to a real dataset.

4.1 Simulation Study

4.1.1 Data generating processes

We generate data according to six data generating processes (DGPs), among which DGPs 1-2 are used for the level study of our test and DGPs 3-6 are for power study.

We use the following two DGPs in the level study:

DGP 1: $Y_t = 1 + Z_t + U_t$,

DGP 2: $Y_t = 0.5Y_{t-1} + U_t$,

where U_t are IID $N(0,1)$, and Z_t are IID sum of 48 independent random variables each uniformly distributed on [-0.25,0.25]. According to the CLT, we can treat Z's as being nearly standard normal random variables but with compact support [-12, 12]. We choose $X_t = Z_t$ in DGP 1 and $X_t = Z_t = Y_{t-1}$ in DGP 2.

The following four DGPs are used in the power study:

DGP 3: $Y_t = 1 + Z_t + \sigma_t\eta_t$,

DGP 4: $Y_t = 1 + Z_t + \sigma_t\eta_t$,

DGP 5: $Y_t = 0.5Y_{t-1} + \sigma_t\eta_t$,

DGP 6: $Y_t = 0.5Y_{t-1} + \sigma_t \eta_t$,
where Z_t are generated as in DGP 1, η_t are IID $N(0,1)$ in DGPs 3, 4, 6 and are IID sum of 48 independent random variables each uniformly distributed on [-0.25,0.25] in DGP 5, $\sigma_t = \sqrt{((Z_t-1)^2+1)/3}$, $\sqrt{((Z_t^2-3)^2+0.1)/6.1}$, $\sqrt{0.2+e^{-Y_{t-1}^2/2}}$, and $\sqrt{0.1+2/(1+e^{-Y_{t-1}})}$ in DGPs 3, 4, 5, and 6, respectively. We choose $X_t = Z_t$ in DGPs 3-4 and $X_t = Z_t = Y_{t-1}$ in DGPs 5-6. Note that X_t is not compactly supported in DGP 6. In addition, to eliminate the starting-up effect, we throw away the first 200 observations when generating the data in DGPs 2, 5 and 6.

4.1.2 Test statistics, kernel, and bandwidth choice

For each DGP, we regress Y_t on $(1, Z_t)$ and obtain the residuals $\widehat{U}_t$. Based on $\widehat{V}_t \equiv \widehat{U}_t^2$, we construct five test statistics. The first one is the Lagrange multiplier (LM) test that tests $\alpha_1 = 0$ in the following parametric regression

$$\widehat{V}_t = \alpha_0 + \alpha_1 X_t^2 + \zeta_t$$

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 following parametric regression

$$\widehat{V}_t = \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2 + \zeta_t.$$

The third one is Hsiao and Li's (2001) nonparametric $\widehat{I}_n$ test, where $\widehat{I}_n \equiv nh^{1/2} J_n/\sqrt{\Omega_n}$, $J_n \equiv \left(n^2h\right)^{-1} \sum_{t=1}^n \sum_{s\neq t}^n \left(\widehat{V}_t - \widehat{\sigma}_0^2\right)\left(\widehat{V}_s - \widehat{\sigma}_0^2\right) K\left(\frac{X_t - X_s}{h}\right)$, $\Omega_n \equiv 2\left(n^2h\right)^{-1} \sum_{t=1}^n \sum_{s\neq t}^n \left(\widehat{V}_t - \widehat{\sigma}_0^2\right)^2 \times \left(\widehat{V}_s - \widehat{\sigma}_0^2\right)^2 K^2\left(\frac{X_t - X_s}{h}\right)$, and $\widehat{\sigma}_0^2 \equiv n^{-1}\sum_{s=1}^n \widehat{V}_s$. The fourth and fifth are our nonparametric R_1^2 and R_2^2 tests that are based on local linear and local quadratic regressions, respectively.

Implementing the latter three tests requires the choice of both kernel function and bandwidth sequence. In all cases, we choose the standard normal PDF as the kernel function. Since it is difficult to pin down the optimal bandwidth for our test, we follow Horowitz and Spokoiny (2001) and Su and Ullah (2009) and consider a set of different bandwidth values. Like them, we use a geometric grid consisting of the points $h_s = \omega^s s_X h_{\min}$, where s_X is the sample standard deviation of X_t, $s = 0, 1, ..., \mathcal{N} - 1$, $\mathcal{N}$ is the number of grid points, $\omega = (h_{\max}/h_{\min})^{1/(\mathcal{N}-1)}$, $h_{\min} = n^{-1/3.01}$, and $h_{\max} = 4n^{-1/1000}$. Following Horowitz and Spokoiny (2001), we choose $\mathcal{N}$ according to the rule of thumb $\mathcal{N} = \lceil \log n \rfloor + 1$, where $\lceil a \rfloor$ means the integer part of a. For each h_s, we calculate the test statistic in (3.1) and denote it as $T_n(h_s)$. Define

$$SupT_n \equiv \max_{0 \leq s \leq \mathcal{N}-1} T_n(h_s). \tag{4.1}$$

Even though $T_n(h_s)$ is asymptotically distributed as $N(0,1)$ under the null for each s, the distribution of $SupT_n$ is generally unknown. Fortunately, we can use bootstrap approximation introduced in Section 3.3.

4.1.3 Test results

Tables 1-2 report the simulation results. We use 1000 replications for each case. To obtain the simulated p-values, we use 200 bootstrap resamples in each replication for both Hsiao and Li's and our tests. To save space in the tables, we use LM, W, HL, NR_1^2 and NR_2^2 to denote the LM test, White's test, Hisao and Li's test, our nonparametric R^2 test based on local linear regression, and our nonparametric R^2 test based on local quadratic regression, respectively.

Table 1: Finite sample rejection frequency under the null (DGPs 1-2)

DGP	$n\backslash$tests	5%					10%				
		LM	W	HL	NR_1^2	NR_2^2	LM	W	HL	NR_1^2	NR_2^2
1	50	0.036	0.040	0.049	0.055	0.054	0.084	0.075	0.102	0.101	0.103
	100	0.046	0.048	0.059	0.062	0.055	0.092	0.078	0.104	0.111	0.107
	200	0.047	0.060	0.036	0.034	0.041	0.099	0.094	0.088	0.077	0.083
2	50	0.037	0.035	0.068	0.061	0.064	0.068	0.084	0.123	0.124	0.126
	100	0.030	0.043	0.051	0.052	0.044	0.077	0.079	0.085	0.105	0.104
	200	0.032	0.050	0.063	0.058	0.058	0.079	0.079	0.112	0.111	0.111

Table 1 reports the empirical rejection frequencies of the tests at 5% and 10% nominal levels when the null hypothesis holds true. It shows that the empirical levels of both the parametric tests (LM, W) and the nonparametric tests (HL, NR_1^2, NR_2^2) are reasonably well behaved despite the fact that the two parametric tests tend to be undersized.

Table 2 reports the empirical power for the five tests at both 5% and 10% nominal levels. We summarize some important findings from Table 2 as follows:

1. For all tests, the empirical power increases quickly as the sample size doubles or quadruples.

2. In DGP 3, the White test utilizes the correct functional form for the conditional variance and it works best among all five tests for small sample size $(n = 50)$. As sample sizes grow, the White test continues to outperform the LM test and our nonparametric R^2 tests, but not Hsiao and Li's test. This indicates that Hsiao and Li's test is very powerful in detecting quadratic form of conditional heteroskedasticity.

3. As can be seen from DGP 4, when the functional form in the parametric tests is incorrectly specified, the nonparametric tests tend to be more powerful than the parametric tests. See also the LM test in DGP 6 in comparison with the nonparametric tests.

4. When the conditional heteroskedasticity is not of quadratic form, our nonparametric R^2 tests tend to outperform Hsiao and Li's test for the DGPs under investigation. For DGPs 4-5, both NR_1^2 and NR_2^2 are more powerful than HL whereas for DGP 6, NR_1^2 outperforms HL which in turn beats NR_2^2.

5. Unexpectedly, the power performance of the local quadratic regression-based R^2 test is not as good as that of the local linear regression-based R^2 test for DGPs 3 and 5-6, even though for the same bandwidth the nonparametric R^2 for the local quadratic regression is always larger than that for the local linear regression. We conjecture that this is due to the differences in both bias-correction terms and variance terms.

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

DGP	$n\backslash$tests	5%					10%				
		LM	W	HL	NR_1^2	NR_2^2	LM	W	HL	NR_1^2	NR_2^2
3	50	0.296	0.660	0.639	0.407	0.355	0.360	0.779	0.752	0.551	0.480
	100	0.493	0.935	0.949	0.772	0.696	0.584	0.966	0.975	0.871	0.818
	200	0.782	0.998	1.000	0.970	0.947	0.842	0.999	1.000	0.989	0.974
4	50	0.361	0.190	0.255	0.534	0.524	0.523	0.344	0.389	0.643	0.628
	100	0.555	0.435	0.603	0.745	0.760	0.659	0.565	0.753	0.828	0.831
	200	0.616	0.567	0.940	0.918	0.926	0.686	0.669	0.974	0.944	0.951
5	50	0.209	0.087	0.127	0.350	0.321	0.410	0.208	0.227	0.487	0.450
	100	0.563	0.315	0.246	0.593	0.559	0.758	0.529	0.385	0.691	0.667
	200	0.943	0.820	0.508	0.913	0.891	0.976	0.924	0.665	0.946	0.936
6	50	0.078	0.278	0.270	0.296	0.263	0.152	0.419	0.405	0.425	0.383
	100	0.116	0.652	0.556	0.582	0.517	0.183	0.801	0.677	0.698	0.632
	200	0.167	0.967	0.861	0.908	0.849	0.247	0.992	0.920	0.942	0.909

4.2 Application to U.S. real GDP growth rates

We now apply the tests to the study of the growth rates of U.S. real gross domestic product (GDP). We download the data from U.S. Bureau of Economic Analysis at http://www.bea.gov/. We have both annual data (1930 - 2008) and seasonally adjusted quarterly data (1947Q2 -

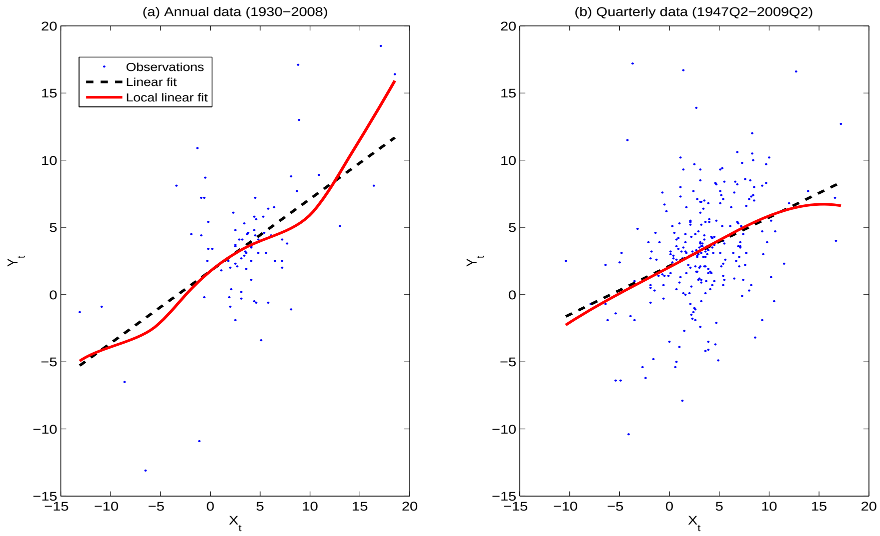

Figure 1: Linear and local linear fits of Y_t on $X_t(\equiv Y_{t-1})$, data are in dots.

2009Q2), both of which are measured at annual percentage rates. For each dataset, let Y_t denote the GDP growth rates at time t and set $X_t \equiv Z_t \equiv Y_{t-1}$. Thus we have $n = 78$ observations for the annual data and $n = 248$ observations for the quarterly data.

Figure 1 plots the Y_t against X_t. Clearly, we can observe positive correlations between Y_t and X_t for both datasets but are not sure whether the relationships are linear or not. For this reason, we consider both parametric linear and nonparametric local liner regressions of Y_t on X_t : $Y_t = \beta_0 + \beta_1 X_t + U_t$, and $Y_t = g(X_t) + U_t$. We use the Gaussian kernel for the local linear regression and the bandwidth is chosen via the least-squares cross validation (LSCV). The linear and local linear fitted lines are also given in Figure 1. Let $\widehat{U}_t$ denote the parametric or nonparametric residuals and set $\widehat{V}_t \equiv \widehat{U}_t^2$. It is not obvious from Figure 1 whether U_t is conditionally homoskedastic or not given X_t. So we now use $\widehat{V}_t$ to conduct tests for conditional homoskedasticity following the implementation procedure as detailed in the simulation subsection. The only exception is that we now use $B = 1000$ bootstrap replications for the three nonparametric tests (HL, NR_1^2, and NR_2^2). In addition, when the first-stage regression is local linear, we generate the bootstrap version of Y_t as $Y_t^* = \widehat{g}_{h_0}(X_t) + U_t^*$, where $\widehat{g}_{h_0}(X_t)$ is the local linear estimate of $g(X_t)$ by using the LSCV bandwidth h_0 and U^*'s are now IID draws from $\{\widehat{U}_s - \overline{\widehat{U}}, s = 1, \cdots, n\}$ with $\widehat{U}_s \equiv Y_s - \widehat{g}_{h_0}(X_s)$.

Table 3 reports the p-values for all tests. For both datasets, no matter whether we use the linear or local linear fits for the first-stage conditional mean model, both LM and W fail to

Table 3: p-values for testing conditional homoskedasticity in U.S. GDP growth rate data

1st-stage regression	Annual data ($n = 78$)					Quarterly data ($n = 248$)				
	LM	W	HL	NR_1^2	NR_2^2	LM	W	HL	NR_1^2	NR_2^2
Linear	0.543	0.831	0.001	0.006	0.003	0.349	0.646	0.019	0.064	0.063
Local linear	0.369	0.668	0	0.045	0.031	0.346	0.641	0.025	0.096	0.082

reject the null hypothesis of conditional homoskedasticity at all conventional significance levels (1%, 5%, 10%). In contrast, for the annual data, both HL and our nonparametric R^2-based tests can reject the null at 5% level; whereas for the quarterly data, HL rejects the null at 5% level and our tests can reject the null at 10% level despite the first stage parametric or nonparametric regressions.

5 Concluding Remarks

In this paper we propose a nonparametric goodness-of-fit-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 of conditional homoskedasticity. Our test has power to detect Pitman local alternatives at the rate $n^{-1/2}h^{-p/4}$ 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 Hsiao and Li (2001) and behaves well in finite samples. Applications to the U.S. real GDP growth rates indicate that both Hsiao and Li test and our test can reveal certain conditional heteroskedasticity which the parametric tests fail to detect.

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, following the LM principle of Breusch and Pagan (1980). Also it can be used to test linear or nonlinear restrictions on the derivatives of nonparametric functions. We leave these for future research.

Appendix: Assumptions and Proofs

Here we give the necessary assumptions for the establishment of the main results in Section 3, along the proofs.

A Assumptions

Let $C < \infty$ denote a generic constant whose value may change across lines. We make the following assumptions on the process $\{U_t, Z_t, X_t\}$, kernel function $K(\cdot)$, and bandwidth sequence h.

Assumption A1. Let $W_t \equiv (U_t, Z_t', X_t')'$. The process $\{W_t\}$ is a strictly stationary strong mixing process with mixing coefficients $\alpha(s)$ such that $\sum_{s=0}^{\infty} s^4 \alpha(s)^{\eta/(4+\eta)} \leq C$ for some $\eta > 0$ with $\eta/(4+\eta) \leq 1/2$. For some some $\gamma \in (0, 1/3]$ such that $n^3 \alpha(n_0)^{\eta/(4+\eta)} = o(1)$, where $n_0 \equiv \lceil n^{\gamma} \rfloor$ is the integer part of n^{γ}.

Assumption A2. (i) $E\left(\varepsilon_t | \mathcal{F}_{-\infty}^{t}(X), \mathcal{F}_{-\infty}^{t-1}(\varepsilon)\right) = 0$ a.s., where for example $\mathcal{F}_{s}^{t}(X)$ denotes the sigma algebra generated by $(X_s, \cdots, X_t)$ for $s < t$.

(ii) $E\left[|\varepsilon_t|^{4+\eta}\right] \leq C$, and $E\left[\left|\varepsilon_{t_1}^{i_1} \varepsilon_{t_2}^{i_2} \cdots \varepsilon_{t_l}^{i_l}\right|^{1+\zeta_1}\right] \leq C$ for some arbitrarily small $\zeta_1 > 0$, where $2 \leq l \leq 4$ and $\sum_{j=1}^{l} i_j \leq 8$.

(iii) Let $v^2(x) \equiv E\left[\varepsilon_t^2 | X_t = x\right]$, and $\mu_4(x) \equiv E\left[\varepsilon_t^4 | X_t = x\right]$. Both $v^2(x)$ and $\mu_4(x)$ are Lipschitz continuous in that $|\vartheta(x + \widetilde{x}) - \vartheta(x)| \leq D_{\vartheta}(x) \|\widetilde{x}\|$ and $E\left[|D_{\vartheta}(X)|^{2+\zeta_2}\right] \leq C$ for $\vartheta(\cdot) = v^2(\cdot)$ or $\mu_4(\cdot)$ and some arbitrarily small $\zeta_2 > 0$, where $\|\cdot\|$ denotes the Euclidean norm.

(iv) The joint probability density function (PDF) $f_{t_1, \cdots, t_l}(\cdot, \cdots, \cdot)$ of $(X_{t_1}, \cdots, X_{t_l})$ for $1 \leq l \leq 4$ exists, is finite, and is Lipschitz continuous in that $|f_{t_1, \cdots, t_l}(x_1 + z_1, \cdots, x_l + z_l) - f_{t_1, \cdots, t_l}(x_1, \cdots, x_l)| \leq D_{t_1, \cdots, t_l}(x_1, \cdots, x_l) \|\mathbf{z}\|$, where $\mathbf{z} \equiv (z_1' \cdots, z_l')'$, $E|D_{t_1, \cdots, t_l}(X_{t_1}, \cdots, X_{t_l})| \leq C$, and $\int D_{t_1, \cdots, t_l}(x_1, \cdots, x_l) \|\mathbf{x}\|^{2(1+\eta)} d\mathbf{x} \leq C$ with $\mathbf{x} \equiv (x_1', \cdots, x_l')'$. When $l = 1$, we use $f(\cdot)$ to denote the marginal PDF of X_t; $f(\cdot)$ is bounded away from 0 on its compact support $\mathcal{X}$.

Assumption A3. (i) $E(U_t | Z_t) = 0$ a.s.

(ii) The parameter space Θ of θ is a compact subset of $\mathbb{R}^d$. $E[Y_t - g(Z_t, \theta)]^2$ is uniquely minimized at θ_0 on Θ.

(iii) The regression function $g(z, \theta)$ is continuously differentiable of order 2 in θ. Let $\bigtriangledown g(z, \theta) \equiv \partial g(z, \theta) / \partial\theta$ and $\bigtriangledown^2 g(z, \theta) \equiv \partial^2 g(z, \theta) / \partial\theta\partial\theta'$. $\bigtriangledown g(z, \cdot)$ and $\bigtriangledown^2 g(z, \cdot)$ 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) $E\left[\bigtriangledown g(Z_1, \theta) \bigtriangledown g(Z_1, \theta)'\right]$ is nonsingular for all θ in a small open neighborhood of θ_0.

Assumption A4. (i)$m(x)$ is Lipschitz continuous in x and has all partial derivatives up to order $q + 1$ if q is odd and $q + 2$ if q is even.

(ii) The $(q + 1)$ or $(q + 2)$th order partial derivatives $D^{\mathbf{k}} m(x)$ with $|\mathbf{k}| = q + 1$ (if q is odd) or $q + 2$ (if q is even), are uniformly bounded in $x \in \mathcal{X}$, and are Hölder continuous in x : $|D^{\mathbf{k}} m(x) - D^{\mathbf{k}} m(\widetilde{x})| \leq C||x - \widetilde{x}||$.

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 $\mathbb{S}$ defined in (2.11) is nonsingular.

(iii) Let $\mathbf{K_j}(x) \equiv x^{\mathbf{j}} K(x)$ 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 $|\mathbf{K_j}(x) - \mathbf{K_j}(\widetilde{x})| \leq C_2 \|x - \widetilde{x}\|$ for any $x, \widetilde{x} \in \mathbb{R}^p$ and for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2q+1$; or $K(x)$ is differentiable, $\|\bigtriangledown \mathbf{K_j}(x)\| \leq C_1$ and for some $\iota_0 > 1$, $|\bigtriangledown \mathbf{K_j}(x)| \leq C_1 \|x\|^{-\iota_0}$ 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 $\{W_t\}$ 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 exponential rate. Assumption A2 is needed to apply Gao's (2007) CLT for second order U-statistic with strong mixing data. The martingale difference assumption that is directly made on ε_t will greatly simplify the proof and the application of the above CLT. Assumption A3, together with A1 and A2(ii), ensures that $\widehat{\theta} - \theta_0 = O_p\left(n^{-1/2}\right)$ by White and Domowitz (1984). Assumptions A4-A5 are used to obtain the uniform consistency for the local polynomial estimator due to Masry (1996) and Hansen (2008). Assumption A6 imposes the conditions on the bandwidth.

B Proof of the Main Results

Recall $D_h \equiv \text{diag}\left(1, h\mathbf{1}'_{N_1} \cdots, h^q \mathbf{1}'_{N_q}\right)$, and $\hat{\mathbf{v}} \equiv \left(\widehat{U}_1^2, \cdots, \widehat{U}_n^2\right)'$. Let $\hat{\mathbf{u}} \equiv \left(\widehat{U}_1, \cdots, \widehat{U}_n\right)'$, $\hat{\mathbf{g}} \equiv \left(g(Z_1, \widehat{\theta}), \cdots, g(Z_n, \widehat{\theta})\right)'$, $\boldsymbol{\varepsilon} \equiv (\varepsilon_1, \cdots, \varepsilon_n)'$, $\mathbf{m} \equiv (m(X_1), \cdots, m(X_n))'$, and $\mathbf{g} \equiv (g(Z_1, \theta_0), \cdots, g(Z_n, \theta_0))'$. It is easy to verify that $H^* \mathbf{1}_n = \mathbf{1}_n$ and $(H^* - L)\mathbf{1}_n = \mathbf{0}_n$. Let $\odot$ and $\otimes$ denote the Hadamard and Kronecker products, respectively.

We first present a technical lemma that is used below.

Lemma B.1 *Let $\{\xi_i, i \geq 1\}$ be a v-dimensional strong mixing process with mixing coefficient $\alpha(\cdot)$. Let $F_{i_1, \ldots, i_m}$, denote the distribution function of $\left(\xi_{i_1}, \ldots, \xi_{i_m}\right)$. For any integer $m > 1$ and integers $(i_1, \ldots, i_m)$ such that $1 \leq i_1 < i_2 < \cdots < i_m$, let θ be a Borel measurable function such that* $\max\{\int |\theta(v_1, \cdots, v_m)|^{1+\widetilde{\eta}} dF_{i_1, \cdots, i_j}(v_1, \cdots, v_j) dF_{i_{j+1}, \cdots, i_m}(v_{j+1}, \cdots, v_m), \int |\theta(v_1, \cdots, v_m)|^{1+\widetilde{\eta}} dF_{i_1, \cdots, i_m}(v_1, \cdots, v_m)\} \leq M_n$ *for some $\widetilde{\eta} > 0$. Then* $|\int \theta(v_1, \cdots, v_m) dF_{i_1, \cdots, i_m}(v_1, \cdots, v_m) - \int \theta(v_1, \cdots, v_m) dF_{i_1, \cdots, i_j}(v_1, \cdots, v_j) dF_{i_{j+1}, \cdots, i_m}(v_{j+1}, \cdots, v_m)| \leq 4 M_n^{1/(1+\widetilde{\eta})} \alpha(i_{j+1} - i_j)^{\widetilde{\eta}/(1+\widetilde{\eta})}$.

Proof. See Lemma 2.1 of Sun and Chiang (1997). ■

Next, we prove a lemma under the conditions of Theorem 3.1.

Lemma B.2 *Let* $S_n(x) \equiv n^{-1} D_h^{-1} \mathbf{X}_x' \mathbf{W}_x \mathbf{X}_x D_h^{-1}$ *and* $S(x) \equiv \mathbb{S} f(x)$. *Then* $R_n \equiv \frac{h^{p/2}}{n} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \int K_{ix} X_{ix}' D_h^{-1} \left[S_n^{-1}(x) - S^{-1}(x) \right] D_h^{-1} X_{jx} K_{jx} dx = o_p(1)$.

Proof. Let $\overline{S}(x) \equiv [S(x) - S_n(x)] S^{-1}(x)$. If the kernel function $K(\cdot)$ is compactly supported, we can verify that under Assumptions A1, A2(iv), and A4-6, the conditions in Corollary 2(ii) of Masry (1996) are all satisfied and conclude that $\sup_{x \in \mathcal{X}} \|S(x) - S_n(x)\| = O_p\left(h + n^{-1/2} h^{-p/2} \sqrt{\log n}\right) = O_p(h)$. In the case where $K(\cdot)$ is not compactly supported, we can apply Theorem 2 of Hansen (2008) to obtain $\sup_{x \in \mathcal{X}} \|S_n(x) - E[S_n(x)]\| = O_p(n^{-1/2} h^{-p/2} \sqrt{\log n}) = O_p(h)$, which implies that $\sup_{x \in \mathcal{X}} ||S(x) - S_n(x)|| = O_p(h)$ by the triangle inequality and the fact that $\sup_{x \in \mathcal{X}} \|E[S_n(x)] - S(x)\| = O(h)$ under Assumptions A2(iv) and A5. In either case, $\sup_{x \in \mathcal{X}} \|\overline{S}(x)\| = O_p(h)$. Now write

$$
\begin{aligned}
R_n &= \frac{2h^{p/2}}{n} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \int K_{ix} X_{ix}' D_h^{-1} S_n^{-1}(x) [S(x) - S_n(x)] S^{-1}(x) D_h^{-1} X_{jx} K_{jx} dx \\
&= \frac{2h^{p/2}}{n} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \int K_{ix} X_{ix}' D_h^{-1} \sum_{l=1}^{\lceil p/2 \rfloor} S^{-1}(x) \overline{S}(x)^l D_h^{-1} X_{jx} K_{jx} dx \\
&\quad + \frac{2h^{p/2}}{n} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \int K_{ix} X_{ix}' D_h^{-1} S_n^{-1}(x) \overline{S}(x)^{\lceil p/2 \rfloor + 1} D_h^{-1} X_{jx} K_{jx} dx \equiv R_{n1} + R_{n2}.
\end{aligned}
$$

We first study R_{n2}. Let $\overline{S}_n(x) \equiv S_n^{-1}(x) \overline{S}(x)^{\lceil p/2 \rfloor + 1}$ and $\varphi_n(x) \equiv \frac{1}{n} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \left(D_h^{-1} X_{jx} \otimes I_{p+1} \right) D_h^{-1} X_{ix} K_{ix} K_{jx}$. Then

$$
\begin{aligned}
R_{n2} &= \int \frac{2h^{p/2}}{n} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \operatorname{tr}\left(\overline{S}_n(x) D_h^{-1} X_{jx} X_{ix}' D_h^{-1} \right) K_{ix} K_{jx} dx \\
&= \int \frac{2h^{p/2} \operatorname{vec}\left(\overline{S}_n(x) \right)'}{n} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \left(D_h^{-1} X_{jx} \otimes I_{p+1} \right) D_h^{-1} X_{ix} K_{ix} K_{jx} dx \\
&= 2h^{p/2} \int \operatorname{vec}\left(\overline{S}_n(x) \right)' \varphi_n(x) dx,
\end{aligned}
$$

Let $b \in \mathbb{R}^{(p+1)^2}$ such that $\|b\| = 1$. Then $E[b' \varphi_n(x)] = 0$ by Assumption A2(i). Write

$$
\begin{aligned}
& E\left[b' \varphi_n(x) \right]^2 \\
=\; & \frac{1}{n^2} \sum_{1 \le i, l < j \le n} E\left[\varepsilon_i \varepsilon_l \varepsilon_j^2 b' \left(D_h^{-1} X_{jx} \otimes I_{p+1} \right) D_h^{-1} X_{ix} \; b' \left(D_h^{-1} X_{jx} \otimes I_{p+1} \right) D_h^{-1} X_{lx} K_{ix} K_{lx} K_{jx}^2 \right] \\
=\; & \frac{2}{n^2} \sum_{1 \le i < l < j \le n} E\left[\varepsilon_i \varepsilon_l \varepsilon_j^2 b' \left(D_h^{-1} X_{jx} \otimes I_{p+1} \right) D_h^{-1} X_{ix} \; b' \left(D_h^{-1} X_{jx} \otimes I_{p+1} \right) D_h^{-1} X_{lx} K_{ix} K_{lx} K_{jx}^2 \right] \\
& + \frac{1}{n^2} \sum_{1 \le i < j \le n}^{n} E\left[\varepsilon_i^2 \varepsilon_j^2 \left\{ b' \left(D_h^{-1} X_{jx} \otimes I_{p+1} \right) D_h^{-1} X_{ix} \right\}^2 K_{ix}^2 K_{jx}^2 \right] \\
\equiv\; & B_{n1} + B_{n2}, \text{ say.}
\end{aligned}
$$

Let $b_{ix} \equiv b' \left(D_h^{-1} X_{jx} \otimes I_{p+1} \right)$ and $\widetilde{X}_{ix} \equiv D_h^{-1} X_{ix}$. Denote $b_{ix,k}$ and $\widetilde{X}_{ix,k}$ as the kth element of b_{ix} and $\widetilde{X}_{ix}$, respectively, where $k = 1, \cdots, N$. Let $\|\xi\|_s \equiv \{E \|\xi\|^s\}^{1/s}$. Then by the Minkowski's inequality, Assumptions A1 and A2(i)-(ii) and (iv), and Lemma B.1 with $\widetilde{\eta} = \eta/4$, we have

$$
\begin{aligned}
\|B_{n1}\| &\leq \frac{2}{n^2} \sum_{1 \leq i < l < j \leq n}^{n} \sum_{k=1}^{N} \sum_{s=1}^{N} \left\| E \left[\varepsilon_i \varepsilon_l \varepsilon_j^2 b_{jx,k}^2 \widetilde{X}_{ix,k} \ \widetilde{X}_{lx,s} D_h^{-1} X_{lx} K_{ix} K_{lx} K_{jx}^2 \right] \right\| \\
&\leq \frac{C}{n^2} \sum_{k=1}^{N} \sum_{s=1}^{N} \sum_{1 \leq i < l < j \leq n}^{n} \alpha \left(j - l \right)^{\eta/(4+\eta)} M_n^{1/(1+\eta/4)} \\
&\leq C h^{-4(1+\eta)p/(4+\eta)} \sum_{\tau=1}^{\infty} \alpha \left(\tau \right)^{\eta/(4+\eta)} = O \left(h^{-2p} \right) \ \text{as } \eta \leq 2
\end{aligned}
$$

where

$$
\begin{aligned}
M_n &= \max_{1 \leq i < l < j \leq n} \max \left\{ E \left| \varepsilon_i \varepsilon_l \varepsilon_j^2 b_{jx,k}^2 \widetilde{X}_{ix,k} \ \widetilde{X}_{lx,s} D_h^{-1} X_{lx} K_{ix} K_{lx} K_{jx}^2 \right|^{1+\eta/4}, \right. \\
&\qquad \left. \int \left| \varepsilon_i \varepsilon_l \varepsilon_j^2 b_{jx,k}^2 \widetilde{X}_{ix,k} \ \widetilde{X}_{lx,s} D_h^{-1} X_{lx} K_{ix} K_{lx} K_{jx}^2 \right|^{1+\eta/4} dF \left(\xi_i, \xi_l \right) dF \left(\xi_j \right) \right\} \\
&= O \left(h^{-(1+\eta)p} \right)
\end{aligned}
$$

$F \left(\xi_j \right)$ and $F \left(\xi_i, \xi_l \right)$ the CDFs of ξ_j and $\left(\xi_i, \xi_l \right)$, respectively, and $\xi_i \equiv \left(\varepsilon_i, X_i' \right)'$. Next, by direct calculation, $B_{n2} = O \left(h^{-2p} \right)$. It follows that $\varphi_n \left(x \right) = O \left(h^{-p} \right)$ and $R_{n2} = O_p \left(h^{(\lceil p/2 \rfloor + 1) - p/2} \right) = o_p \left(1 \right)$ as $\overline{S}_n \left(x \right) = O_p \left(h^{\lceil p/2 \rfloor + 1} \right)$.

For R_{n1}, we focus on the case where $p \leq 3$ since the case of $p > 3$ can be proved similarly but is more tedious. Clearly, if $p = 1$, $R_{n1} = 0$. When $p = 2$ or 3,

$$
\begin{aligned}
R_{n1} &= \frac{2h^{p/2}}{n} \sum_{1 \leq i < j \leq n} \varepsilon_i \varepsilon_j \int K_{ix} X_{ix}' D_h^{-1} S^{-1} \left(x \right) \overline{S} \left(x \right) D_h^{-1} X_{jx} K_{jx} dx \\
&= \frac{2h^{p/2}}{n^2} \sum_{1 \leq i < j \leq n} \sum_{k=1}^{n} \varepsilon_i \varepsilon_j \int K_{ix} X_{ix}' D_h^{-1} S^{-1} \left(x \right) \overline{s}_k \left(x \right) D_h^{-1} X_{jx} K_{jx} dx \\
&= \frac{2h^{p/2}}{n^2} \sum_{1 \leq i < j \leq n} \sum_{k=1}^{n} \varepsilon_i \varepsilon_j \int K_{ix} X_{ix}' D_h^{-1} S^{-1} \left(x \right) \widetilde{s}_k \left(x \right) D_h^{-1} X_{jx} K_{jx} dx \\
&\quad + \frac{2h^{p/2}}{n} \sum_{1 \leq i < j \leq n} \varepsilon_i \varepsilon_j \int K_{ix} X_{ix}' D_h^{-1} S^{-1} \left(x \right) \overline{s} \left(x \right) D_h^{-1} X_{jx} K_{jx} dx \\
&\equiv R_{n11} + R_{n12}, \ \text{say,}
\end{aligned}
$$

where $\overline{s}_k \left(x \right) \equiv [S \left(x \right) - K_{kx} D_h^{-1} X_{kx} X_{kx}' D_h^{-1}] S^{-1} \left(x \right)$, $\widetilde{s}_k \left(x \right) \equiv \{ -K_{kx} D_h^{-1} X_{kx} X_{kx}' D_h^{-1} + E[K_{kx} D_h^{-1} X_{kx} X_{kx}' D_h^{-1} S^{-1} \left(x \right)]\}$, and $\overline{s} \left(x \right) \equiv \left\{ S \left(x \right) - E \left[K_{kx} D_h^{-1} X_{kx} X_{kx}' D_h^{-1} \right] \right\} S^{-1} \left(x \right)$. Noting that

$\overline{s}(x) = O(h)$ uniformly in x, it is straightforward to show that $R_{n12} = O_p(h) = o_p(1)$. Now, write

$$\begin{aligned} R_{n11} &= \frac{2h^{p/2}}{n^2} \sum_{1 \le i < j \le n, k \ne i,j}^{n} \varepsilon_i \varepsilon_j \int K_{ix} X'_{ix} D_h^{-1} S^{-1}(x) \widetilde{s}_k(x) D_h^{-1} X_{jx} K_{jx} dx \\ &\quad + \frac{2h^{p/2}}{n^2} \sum_{1 \le i < j \le n} \varepsilon_i \varepsilon_j \int K_{ix} X'_{ix} D_h^{-1} S^{-1}(x) [\widetilde{s}_i(x) + \widetilde{s}_j(x)] D_h^{-1} X_{jx} K_{jx} dx \\ &\equiv R_{n11,1} + R_{n11,2}, \text{ say.} \end{aligned}$$

Noting that $E[R_{n11,2}] = 0$ and $E\left[R_{n11,2}^2\right] = O\left(n^{-2}h^{-p} + n^{-1}\right)$, we have $R_{n11,2} = o_p(1)$ by the Chebyshev inequality. For $R_{n11,1}$, we have

$$\begin{aligned} R_{n11,1} &= \frac{2h^{p/2}}{n^2} \left\{ \sum_{1 \le k < i < j \le n} + \sum_{1 \le i < k < j \le n} + \sum_{1 \le i < j < k \le n} \right\} \varphi_1(\xi_i, \xi_j, \xi_k) \\ &\equiv D_1 + D_2 + D_3, \text{ say,} \end{aligned}$$

where $\varphi_1(\xi_i, \xi_j, \xi_k) \equiv \varepsilon_i \varepsilon_j \int K_{ix} X'_{ix} D_h^{-1} S^{-1}(x) \widetilde{s}_k(x) D_h^{-1} X_{jx} K_{jx} dx$, and, for example, $D_3 \equiv 2h^{p/2} n^{-2} \sum_{1 \le i < j < k \le n}^{n} \varphi_1(\xi_i, \xi_j, \xi_k)$. We only show that $D_3 = o_p(1)$ because it is simpler to show that $D_1 = o_p(1)$ and $D_2 = o_p(1)$ by noticing that $E(D_l) = 0$ for $l = 1, 2$. We prove $D_3 = o_p(1)$ by showing that $ED_1 \equiv E(D_3) = o(1)$ and $ED_2 \equiv E\left(D_3^2\right) = o(1)$.

Let $\delta \equiv \eta/4$. Let $n_0 \equiv \lceil n^\gamma \rfloor$ for some $\gamma \in (0, 1/3]$ such that $n^3 \alpha(n_0)^{\delta/(1+\delta)} = o(1)$ which is possible by Assumption A1. For ED_1, we write ED_1 as the summation of $E\left[\varphi_1(\xi_i, \xi_j, \xi_k)\right]$ over the indices (i, j, k) corresponding to two cases: (a) $j - i > n_0$ or $k - j > n_0$, (b) $j - i \le n_0$ and $k - j \le n_0$. We use ED_{1s}, $s = a, b$, to denote these two cases. For case (a), without loss of generality, assume that $j - i > n_0$. Note that

$$E\left|\varphi_1(\xi_i, \xi_j, \xi_k)\right|^{1+\delta} \le \mu_{2+2\delta} E\left|\int K_{ix} X'_{ix} D_h^{-1} S^{-1}(x) \widetilde{s}_k(x) D_h^{-1} X_{jx} K_{jx} dx\right|^{1+\delta} \le Ch^{-2p\delta}$$

and similar results hold for $E_i E_{jk} \left|\varphi_1(\xi_i, \xi_j, \xi_k)\right|^{1+\delta}$, where E_i and E_{jk} denote expectations with respect to ξ_i and (ξ_j, ξ_k) respectively by treating them as independent of each other. By Lemma B.1 we have $D_{1a} \le Ch^{p/2 - 2p/(1+\delta)} \sum_{\tau = n_0 + 1}^{n} \alpha(\tau)^{\delta/(1+\delta)} = O(nh^{p/2 - 2p/(1+\delta)} \alpha(n_0)^{\delta/(1+\delta)})$ $= o(1)$. For case (b), the number of terms in the summation is $O\left(nn_0^2\right)$ and each is of order $O(1)$. It follows that $ED_{1b} = O\left(h^{p/2} n_0^2 n^{-1}\right) = o(1)$. Hence $ED_1 = o(1)$. For ED_2, write

$$ED_2 = \frac{4}{n^4} \sum_{1 \le t_1 < t_2 < t_3 \le n} \sum_{1 \le t_4 < t_5 < t_6 \le n} E\left[\varphi_1(\xi_{t_1}, \xi_{t_2}, \xi_{t_3}) \varphi_1(\xi_{t_4}, \xi_{t_5}, \xi_{t_6})\right].$$

We consider two cases: (a) for at least three different i's, $|t_i - t_j| > n_0$ for all $j \ne i$; (b) all the other remaining cases. We use ED_{2s}, $s = a, b$, to denote these cases. In case (a), at least one of t_1, t_3, t_4 and t_6 lies n_0-distance far away from all other indices in $\{t_1, \cdots, t_6\}$ so that we can apply Lemma B.1 to obtain $ED_{2a} = h^p O(n^2 h^{-4p\delta/(1+\delta)} \alpha(n_0)^{\delta/(1+\delta)}) = o(1)$ as $4\delta/(1+\delta) \leq 2$ and $nh^p \to \infty$. In case (b), noting that the total number of terms in the summation is $O(n^3 n_0^3)$, each of which is $O(1)$, we have $ED_{2b} = O(h^p n^{-1} n_0^3) = o(1)$. It follows that $ED_2 = o(1)$ and thus $D_3 = o_p(1)$. Consequently, $R_{n11,1} = o_p(1)$ and $R_{n11} = o_p(1)$. ■

Proof of Theorem 3.1

Noting that $\widehat{U}_i^2 = [U_i + (\widehat{U}_i - U_i)]^2 = U_i^2 + (\widehat{U}_i - U_i)^2 + 2(\widehat{U}_i - U_i)U_i = m(X_i) + \varepsilon_i + [g(Z_i, \widehat{\theta}) - g(Z_i, \theta_0)]^2 - 2[g(Z_i, \widehat{\theta}) - g(Z_i, \theta_0)]U_i$, we have

$$ESS_q = \widehat{\mathbf{v}}' (H^* - L) \widehat{\mathbf{v}} = A_1 + A_2 + A_3 + 4A_4 + 2A_5 + 2A_6 - 4A_7 + 2A_8 - 4A_9 - 4A_{10}, \quad \text{(B.1)}$$

where

$$\begin{aligned}
A_1 &\equiv \mathbf{m}' (H^* - L) \mathbf{m}, \quad A_2 \equiv \boldsymbol{\varepsilon}' (H^* - L) \boldsymbol{\varepsilon}, \\
A_3 &\equiv ((\widehat{\mathbf{g}} - \mathbf{g}) \odot (\widehat{\mathbf{g}} - \mathbf{g}))' (H^* - L) ((\widehat{\mathbf{g}} - \mathbf{g}) \odot (\widehat{\mathbf{g}} - \mathbf{g})), \\
A_4 &\equiv ((\widehat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u})' (H^* - L) ((\widehat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}), \\
A_5 &\equiv \mathbf{m}' (H^* - L) \boldsymbol{\varepsilon}, \\
A_6 &\equiv \mathbf{m}' (H^* - L) ((\widehat{\mathbf{g}} - \mathbf{g}) \odot (\widehat{\mathbf{g}} - \mathbf{g})), \\
A_7 &\equiv \mathbf{m}' (H^* - L) ((\widehat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}), \\
A_8 &\equiv \boldsymbol{\varepsilon}' (H^* - L) ((\widehat{\mathbf{g}} - \mathbf{g}) \odot (\widehat{\mathbf{g}} - \mathbf{g})), \\
A_9 &\equiv \boldsymbol{\varepsilon}' (H^* - L) ((\widehat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}), \\
A_{10} &\equiv ((\widehat{\mathbf{g}} - \mathbf{g}) \odot (\widehat{\mathbf{g}} - \mathbf{g})) (H^* - L) ((\widehat{\mathbf{g}} - \mathbf{g}) \odot \mathbf{u}). \qquad \text{(B.2)}
\end{aligned}$$

Under H_0, $\mathbf{m} = \sigma_0^2 \mathbf{1}_n$. It follows that $A_s = 0$ for s =1, 5, 6, and 7 as $(H^* - L) \mathbf{1}_n = \mathbf{0}_n$ and $H^* - L$ is symmetric. Noting that $n^{-1} TSS = \sigma_V^2 + o_p(1)$, it suffices to prove the theorem by showing that

$$\begin{aligned}
\overline{A}_2 &\equiv h^{p/2} A_2 - h^{p/2} \sum_{i=1}^{n} \varepsilon_i^2 (H_{ii}^* - n^{-1}) \xrightarrow{d} N(0, \Omega_0 \sigma_V^4), \qquad \text{(B.3)} \\
h^{p/2} A_s &= o_p(1) \text{ for } s = 3, 4, 8, 9, 10. \qquad \text{(B.4)}
\end{aligned}$$

We first show (B.3). By Lemma B.2,

$$
\begin{aligned}
\overline{A}_2 &= 2h^{p/2}\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j\left(H_{ij}^* - n^{-1}\right)\\
&= 2h^{p/2}\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j\left(n^{-1}\overline{H}_{ij}^* - n^{-1}\right) + 2h^{p/2}\sum_{1\le i<j\le n}^{n}\varepsilon_i\varepsilon_j\left(H_{ij}^* - n^{-1}\overline{H}_{ij}^*\right)\\
&= \frac{2h^{p/2}}{n}\sum_{1\le i<j\le n}\varepsilon_i\varepsilon_j\left(\int K_{ix}X_{ix}'D_h^{-1}\mathbb{S}^{-1}D_h^{-1}X_{jx}K_{jx}f(x)^{-1}dx - 1\right) + o_p(1)\\
&\equiv \overline{A}_{21} + o_p(1), \qquad\qquad (B.5)
\end{aligned}
$$

where $\overline{H}_{ij}^* \equiv \int K_{ix}X_{ix}'D_h^{-1}\mathbb{S}^{-1}D_h^{-1}X_{jx}K_{jx}f(x)^{-1}dx$, $\overline{A}_{21} \equiv \frac{2}{n}\sum_{1\le i<j\le n}\varphi_n(\xi_i,\xi_j)$, and $\varphi_n(\xi_i, \xi_j) \equiv h^{p/2}\varepsilon_i\varepsilon_j\left(\int K_{ix}X_{ix}'D_h^{-1}\mathbb{S}^{-1}D_h^{-1}X_{jx}K_{jx}f(x)^{-1}dx - 1\right)$. Note that $\overline{A}_{21}$ 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 $\overline{A}_{21}$. [The exponential mixing rate in the theorem can be relaxed to our requirement on the mixing rate in Assumption A1.] Its asymptotic variance is given by

$$
\begin{aligned}
&\lim_{n\to\infty} 2E_iE_j\left[\varphi_n(\xi_i,\xi_j)^2\right]\\
=\ &\lim_{n\to\infty} h^pE_iE_j\left[\varepsilon_i^2\varepsilon_j^2\left(\int K_{ix}X_{ix}'D_h^{-1}\mathbb{S}^{-1}D_h^{-1}X_{jx}K_{jx}f(x)^{-1}dx - 1\right)^2\right]\\
=\ &\lim_{n\to\infty} h^pE_iE_j\left[v^2(X_i)\,v^2(X_j)\left(\int K_{ix}X_{ix}'D_h^{-1}\mathbb{S}^{-1}D_h^{-1}X_{jx}K_{jx}f(x)^{-1}dx\right)^2\right]\\
=\ &\lim_{n\to\infty} h^{-p}E_iE_j\left[v^2(X_i)\,v^2(X_j)\left\{\int K(z)\,\mu(-z)'\,\mathbb{S}^{-1}\mu\left(-\left(z+\chi_{ij,h}\right)\right)K\left(z+\chi_{ij,h}\right)dz\right\}^2 f(X_i)^{-1}\right]\\
=\ &\int\int\left[v^2(\widetilde{x})\right]^2\left[\int K(z)\,\mu(-z)'\,\mathbb{S}^{-1}\mu(-z+x)\,K(z-x)\,dz\right]^2 dx\,d\widetilde{x}\\
=\ &\int\left[\int K(z)\,\mu(z)'\,\mathbb{S}^{-1}\mu(z+x)\,K(z+x)\,dz\right]^2 dx\int\left[v^2(\widetilde{x})\right]^2 d\widetilde{x},
\end{aligned}
$$

where $\chi_{ij,h} \equiv (X_i - X_j)/h$, E_i denotes expectation with respect to ξ_i, and recall $v^2(x) = E\left(\varepsilon_i^2|X_i = x\right)$ and $\mu(\cdot)$ is a function used in the definition of X_{ix} (e.g., $X_{ix} = (1, (X_i - x)')'$ if $q = 1$, i.e., $\mu(z) = (1, z')'$ in this case). That is, $\overline{A}_{21} \xrightarrow{d} N\left(0, \sigma_V^4\Omega_0\right)$. This, together with (B.5), implies that (B.3) follows.

We now show (B.4). By White and Domowitz (1984), $\widehat{\theta} - \theta_0 = O_p\left(n^{-1/2}\right)$ under Assumptions A1, A2(ii) and A3. Noting that the elements of H^* are uniformly $O_p\left(n^{-1}h^{-p}\right)$, by

Assumption A3(iii) we have

$$\begin{aligned}
h^{p/2}A_3 &= h^{p/2}\sum_{i=1}^{n}[g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)]^4\left(H_{ii}^*-n^{-1}\right)\\
&\quad +h^{p/2}\sum_{i=1}^{n}[g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)]^2\sum_{j\neq i}^{n}\left(H_{ij}^*-n^{-1}\right)[g(Z_j,\widehat{\theta})-g(Z_j,\theta_0)]^2\\
&\leq O_p\left(n^{-1}h^{-p/2}\right)\sum_{i=1}^{n}\|G_1(Z_i)\|^4\left\|\widehat{\theta}-\theta_0\right\|^4+O_p\left(n^{-1}h^{-p/2}\right)\left\{\sum_{i=1}^{n}\|G_1(Z_i)\|^2\right\}^2\left\|\widehat{\theta}-\theta_0\right\|^4\\
&= O_p\left(n^{-2}h^{-p/2}\right)+O_p\left(n^{-1}h^{-p/2}\right)=o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
h^{p/2}A_{10} &= h^{p/2}\sum_{i=1}^{n}[g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)]^3U_i\left(H_{ii}^*-n^{-1}\right)\\
&\quad +h^{p/2}\sum_{i=1}^{n}[g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)]^2\sum_{j\neq i}^{n}\left(H_{ij}^*-n^{-1}\right)[g(Z_j,\widehat{\theta})-g(Z_j,\theta_0)]U_j\\
&\leq O_p\left(n^{-1}h^{-p/2}\right)\sum_{i=1}^{n}\|G_1(Z_i)\|^3|U_i|\left\|\widehat{\theta}-\theta_0\right\|^3\\
&\quad +O_p\left(n^{-1}h^{-p/2}\right)\left\{\sum_{i=1}^{n}\|G_1(Z_i)\|^2\right\}\sum_{j=1}^{n}\|G_1(Z_i)\|\,|U_j|\left\|\widehat{\theta}-\theta_0\right\|^3\\
&= O_p\left(n^{-3/2}h^{-p/2}\right)+O_p\left(n^{-1/2}h^{-p/2}\right)=o_p(1).
\end{aligned}$$

For A_4, write

$$\begin{aligned}
h^{p/2}A_4 &\equiv ((\widehat{\mathbf{g}}-\mathbf{g})\odot\mathbf{u})'(H^*-L)((\widehat{\mathbf{g}}-\mathbf{g})\odot\mathbf{u})\\
&= h^{p/2}\sum_{i=1}^{n}[g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)]^2U_i^2\left(H_{ii}^*-n^{-1}\right)\\
&\quad +h^{p/2}\sum_{i=1}^{n}[g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)]U_i\sum_{j\neq i}^{n}\left(H_{ij}^*-n^{-1}\right)[g(Z_j,\widehat{\theta})-g(Z_j,\theta_0)]U_j\\
&\equiv A_{41}+A_{42},\ \text{say.}
\end{aligned}$$

Clearly, $A_{41}\leq O_p\left(n^{-1}h^{-p/2}\right)\sum_{i=1}^{n}\|G_1(Z_i)\|^2|U_i|^2\,||\widehat{\theta}-\theta_0||^2=O_p\left(n^{-1}h^{-p/2}\right)=o_p(1)$. Noting that $g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)=(\nabla g(Z_i,\theta_0))'(\widehat{\theta}-\theta_0)+(\widehat{\theta}-\theta_0)'\nabla^2g(Z_i,\widetilde{\theta})(\widehat{\theta}-\theta_0)$ where $\widetilde{\theta}$ lies between

$\widehat{\theta}$ and θ_0 elementwise, by Assumption A3 we have

$$\begin{aligned}
A_{42} &= 2h^{p/2}(\widehat{\theta}-\theta_0)'\left\{\sum_{1\leq i<j\leq n}^{n}\nabla g\left(Z_i,\theta_0\right)U_iU_j\left(H_{ij}^{*}-n^{-1}\right)\nabla g\left(Z_j,\theta_0\right)'\right\}(\widehat{\theta}-\theta_0) \\
&\quad +\frac{1}{2}h^{p/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\|G_2\left(Z_i\right)\right\|\left\|U_i\right\|\left\|G_2\left(Z_j\right)\right\|\left\|U_j\right\|\left\|\widehat{\theta}-\theta_0\right\|^4\left(H_{ij}^{*}-n^{-1}\right) \\
&= 2h^{p/2}(\widehat{\theta}-\theta_0)'\left\{\sum_{1\leq i<j\leq n}^{n}\nabla g\left(Z_i,\theta_0\right)U_iU_j\left(n^{-1}\overline{H}_{ij}^{*}-n^{-1}\right)\nabla g\left(Z_j,\theta_0\right)'\right\}(\widehat{\theta}-\theta_0) \\
&\quad +2h^{p/2}(\widehat{\theta}-\theta_0)'\left\{\sum_{1\leq i<j\leq n}^{n}\nabla g\left(Z_i,\theta_0\right)U_iU_j\left(H_{ij}^{*}-n^{-1}\overline{H}_{ij}^{*}\right)\nabla g\left(Z_j,\theta_0\right)'\right\}(\widehat{\theta}-\theta_0) \\
&\quad +O_p\left(n^{-1}h^{-p/2}\right).
\end{aligned}$$

The first term on the right-hand side of the last expression is $O_p\left(n^{-1}\right)$ because we can show that $n^{-1}h^{p/2}\sum_{1\leq i<j\leq n}^{n}\nabla g\left(Z_i,\theta_0\right)U_iU_j\left(\overline{H}_{ij}^{*}-1\right)\nabla g\left(Z_j,\theta_0\right)'=O_p\left(1\right)$ under Assumptions A1-A3 and A5-A6 by the second-order U-statistic theory. The second term is $o_p\left(n^{-1}\right)$ by arguments analogous to those used in the proof of Lemma B.2. It follows that $h^{p/2}A_4=o_p\left(1\right).$

For A_8, write

$$\begin{aligned}
h^{p/2}A_8 &= h^{p/2}\sum_{i=1}^{n}[g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)]^2\varepsilon_i\left(H_{ii}^{*}-n^{-1}\right) \\
&\quad +h^{p/2}\sum_{i=1}^{n}\varepsilon_i\sum_{j\neq i}^{n}\left(H_{ij}^{*}-n^{-1}\right)[g(Z_j,\widehat{\theta})-g(Z_j,\theta_0)]^2 \\
&\equiv A_{81}+A_{82},\ \text{say.}
\end{aligned}$$

By Taylor expansions, we can show that $A_{81}=O_p\left(n^{-1}h^{-p/2}\right)=o_p\left(1\right)$ and $A_{82}=O_p\left(n^{-1/2}\right)=o_p\left(1\right).$ Hence $h^{p/2}A_8=o_p\left(1\right).$ Similarly, write

$$\begin{aligned}
h^{p/2}A_9 &= h^{p/2}\sum_{i=1}^{n}[g(Z_i,\widehat{\theta})-g(Z_i,\theta_0)]\varepsilon_iU_i\left(H_{ii}^{*}-n^{-1}\right) \\
&\quad +h^{p/2}\sum_{i=1}^{n}\varepsilon_i\sum_{j\neq i}^{n}\left(H_{ij}^{*}-n^{-1}\right)[g(Z_j,\widehat{\theta})-g(Z_j,\theta_0)]U_j \\
&\equiv A_{91}+A_{92},\ \text{say.}
\end{aligned}$$

Then $A_{91}=h^{p/2}(\widehat{\theta}-\theta_0)'\sum_{i=1}^{n}\nabla g(Z_i,\theta_0)\varepsilon_iU_i\left(H_{ii}^{*}-n^{-1}\right)+O_p\left(n^{-1}h^{-p/2}\right)=O_p\left(n^{-1}\right)+O_p\left(n^{-1}h^{-p/2}\right)=o_p\left(1\right),$ and $A_{92}=h^{p/2}(\widehat{\theta}-\theta_0)'\sum_{i=1}^{n}\sum_{j\neq i}^{n}\nabla g(Z_j,\theta_0)\varepsilon_i\left(H_{ij}^{*}-n^{-1}\right)U_j+O_p\left(n^{-1/2}\right)=O_p\left(n^{-1/2}\right)=o_p\left(1\right).$ This completes the proof of (B.4).■

Proof of Theorem 3.2

The proof follows closely from that of Theorem 3.1, now keeping the additional terms that do not vanish under $H_1\left(n^{-1/2}h^{-p/4}\right)$. It suffices to show that under $H_1\left(n^{-1/2}h^{-p/4}\right)$,

$$h^{p/2}A_1 \xrightarrow{p} \Delta_0, \tag{B.6}$$

$$h^{p/2}A_s = o_p(1), \ \ s = 5,6,7 \tag{B.7}$$

where A_1, A_5, A_6, and A_7 are defined in (B.2). Let $\gamma_n \equiv n^{-1/2}h^{-p/4}$. Under $H_1(\gamma_n)$, $m(x) = \sigma_0^2 + \gamma_n \Delta(x)$ and we have

$$\begin{aligned}
h^{p/2}A_1 &= h^{p/2}\gamma_n^2 \sum_{i=1}^n \sum_{j=1}^n \Delta(X_i)\Delta(X_j)\left(H_{ij}^* - n^{-1}\right) \\
&= n^{-2}\sum_{i=1}^n \sum_{j=1}^n \Delta(X_i)\Delta(X_j)\left(\overline{H}_{ij}^* - 1\right) + n^{-1}\sum_{i=1}^n \sum_{j=1}^n \Delta(X_i)\Delta(X_j)\left(H_{ij}^* - n^{-1}\overline{H}_{ij}^*\right) \\
&\equiv A_{11} + A_{12}, \text{ say,}
\end{aligned}$$

where recall $\overline{H}_{ij}^* = \int K_{ix}X_{ix}' D_h^{-1}\mathbb{S}^{-1} \ D_h^{-1}X_{jx}K_{jx}f(x)^{-1}dx$. It is straightforward to show that $A_{12} = o_p(1)$. For A_{11}, by the Fubini theorem, the weak law of large numbers, and Assumptions A1, A2(iv) and A5-A6, we have

$$\begin{aligned}
& A_{11} \\
=\ & n^{-2}\sum_{i=1}^n \sum_{j=1}^n \Delta(X_i)\Delta(X_j)\left(\int K_{ix}X_{ix}' D_h^{-1}\mathbb{S}^{-1}D_h^{-1}X_{jx}K_{jx}f(x)^{-1}dx - 1\right) \\
=\ & \int \left\{n^{-1}\sum_{i=1}^n \Delta(X_i)K_{ix}X_{ix}'D_h^{-1}\right\}\mathbb{S}^{-1}\left\{n^{-1}\sum_{j=1}^n D_h^{-1}X_{jx}K_{jx}\Delta(X_j)\right\}f(x)^{-1}dx \\
& - \left[n^{-1}\sum_{i=1}^n \Delta(X_i)\right]^2 \\
=\ & \int \Delta^2(x)\mathbb{B}'\mathbb{S}^{-1}\mathbb{B}f(x)\,dx - \left[E(\Delta(X_1))\right]^2 + o_p(1) = \Delta_0 + o_p(1).
\end{aligned}$$

Consequently, $h^{p/2}A_1 \xrightarrow{p} \Delta_0$.

Next, write

$$\begin{aligned}
h^{p/2}A_5 &= n^{-1}h^{p/2}\gamma_n \sum_{1\le i<j\le n} \Delta(X_i)\varepsilon_j\left(\overline{H}_{ij}^* - 1\right) + n^{-1}h^{p/2}\gamma_n \sum_{1\le j<i\le n}\Delta(X_i)\varepsilon_j\left(\overline{H}_{ij}^* - 1\right) \\
&\quad + h^{p/2}\gamma_n \sum_{i=1}^n \sum_{j\ne i}^n \Delta(X_i)\varepsilon_j\left(H_{ij}^* - n^{-1}\overline{H}_{ij}^*\right) + h^{p/2}\gamma_n \sum_{i=1}^n \Delta(X_i)\varepsilon_i\left(H_{ii}^* - n^{-1}\right) \\
&= A_{51} + A_{52} + A_{53} + A_{54}, \text{ say.}
\end{aligned}$$

Using arguments similar to but simpler than those used in the proof of $R_{n11,1} = o_p(1)$ in Lemma B.2, we can show that $A_{51} = o_p(1)$ and $A_{52} = o_p(1)$. Decompose

$$\begin{aligned} A_{53} &= n^{-1}h^{p/2}\gamma_n \sum_{i=1}^{n}\sum_{j\neq i}^{n} \Delta(X_i)\,\varepsilon_j \int K_{ix}X_{ix}' D_h^{-1}\left[S_n^{-1}(x) - S^{-1}(x)\right] D_h^{-1}X_{jx}K_{jx}dx \\ &= n^{-1}h^{p/2}\gamma_n \sum_{i=1}^{n}\Delta(X_i)\int K_{ix}X_{ix}'D_h^{-1}\sum_{l=1}^{\lceil p/4\rfloor} S^{-1}(x)\,\overline{S}(x)^l \sum_{j\neq i}^{n}\varepsilon_j D_h^{-1}X_{jx}K_{jx}dx \\ &\quad + n^{-1}h^{p/2}\gamma_n\sum_{i=1}^{n}\Delta(X_i)\int K_{ix}X_{ix}'D_h^{-1}S^{-1}(x)\,\overline{S}(x)^{\lceil p/4\rfloor+1}\sum_{j\neq i}^{n}\varepsilon_j D_h^{-1}X_{jx}K_{jx}dx \\ &\equiv A_{53a} + A_{53b},\ \text{say,} \end{aligned}$$

where recall $S_n(x) \equiv n^{-1}D_h^{-1}\mathbf{X}_x'\,\mathbf{W}_x\mathbf{X}_x D_h^{-1}$, $S(x) \equiv \mathbb{S}f(x)$ and $\overline{S}(x) \equiv [S(x) - S_n(x)]\,S^{-1}(x)$. If $p = 1, 2, 3$, $A_{53a} = 0$. For $p = 4, 5, 6, 7$, following the proof of R_{n11} in Lemma B.2 we can show that $A_{53a} = o_p(1)$. As to A_{53b}, we have

$$\begin{aligned} &|A_{53b}| \\ \leq\ & h^{p/2}\gamma_n \sup_{x\in\mathcal{X}}\left|\overline{S}(x)^{\lceil p/4\rfloor+1}\right| \sup_{x\in\mathcal{X}}\left\|n^{-1}\sum_{j=1}^{n}\varepsilon_j D_h^{-1}X_{jx}K_{jx}\right\|\left\{\sum_{i=1}^{n}\Delta(X_i)\int K_{ix}\left\|X_{ix}'D_h^{-1}S^{-1}(x)\right\|dx\right\} \\ =\ & h^{p/2}\gamma_n O_p\left(h^{\lceil p/4\rfloor+1}\right)O_p\left(n^{-1/2}h^{-p/2}\sqrt{\log n}\right)O_p(n) = O_p\left(h^{\lceil p/4\rfloor+1-p/4}\sqrt{\log n}\right) = o_p(1). \end{aligned}$$

It follows that $A_{53} = o_p(1)$. Next, $|A_{54}| = h^{p/2}\gamma_n \max_i\left|n^{-1}H_{ii}^* - 1\right|\left\{n^{-1}\sum_{i=1}^{n}|\Delta(X_i)|\,\varepsilon_i\right\} = O_p(n^{-1/2}\ h^{-3p/4}) = o_p(1)$. Consequently $h^{p/2}A_5 = o_p(1)$.

For A_6, we have

$$\begin{aligned} h^{p/2}A_6 &= h^{p/2}\gamma_n\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Delta(X_i)\,[g(Z_j,\widehat{\theta}) - g(Z_j,\theta_0)]^2\left(H_{ij}^* - n^{-1}\right) + \\ &\quad + h^{p/2}\gamma_n\sum_{i=1}^{n}\Delta(X_i)\,[g(Z_i,\widehat{\theta}) - g(Z_i,\theta_0)]^2\left(H_{ii}^* - n^{-1}\right) \\ &\equiv A_{61} + A_{62},\ \text{say.} \end{aligned}$$

By Taylor expansions, it is straightforward to show that $A_{61} = O_p\left(n^{-1/2}h^{p/4}\right) = o_p(1)$ and $A_{62} = O_p\left(n^{-3/2}h^{-3p/4}\right) = o_p(1)$, and hence $h^{p/2}A_6 = o_p(1)$. Now,

$$\begin{aligned} h^{p/2}A_7 &= h^{p/2}\gamma_n\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Delta(X_i)\,U_j[g(Z_j,\widehat{\theta}) - g(Z_j,\theta_0)]\left(H_{ij}^* - n^{-1}\right) + \\ &\quad + h^{p/2}\gamma_n\sum_{i=1}^{n}\Delta(X_i)\,U_i[g(Z_i,\widehat{\theta}) - g(Z_i,\theta_0)]\left(H_{ii}^* - n^{-1}\right) \\ &\equiv A_{71} + A_{72},\ \text{say.} \end{aligned}$$

By Taylor expansions, it is easy to show that $A_{71} = O_p\left(h^{p/4}\right) = o_p(1)$, and $A_{72} = O_p\left(n^{-1}h^{-3p/4}\right) = o_p(1)$, and hence $h^{p/2}A_7 = o_p(1)$.

Consequently, $P(\widehat{T} \geq z_\alpha | H_1(n^{-1/2}h^{-p/4})) \to 1 - \Phi(z_\alpha - \Delta_0/\sigma_0)$. This concludes the proof of the theorem.■

Proof of Theorem 3.3

The proof follows closely from that of Theorems 3.1 and 3.2. By (B.1) and the proof of Theorem 3.1, $ESS_q = A_1 + 2A_5 + 2A_6 - 4A_7 + o_p\left(h^{-p/2}\right)$. 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 H_1 for $s = 5, 6, 7$. Under H_1, by the Fubini theorem, the weak law of large numbers, and Assumptions A1, A2(iv), and A4-A6, we have

$$\begin{aligned} & n^{-1}A_1 \\ = & \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n m(X_i)\, m(X_j)\left\{\int K_{ix}X_{ix}'D_h^{-1}\mathbb{S}^{-1}D_h^{-1}X_{jx}K_{jx}f^{-1}(x)\,dx - 1\right\} + o_p(1) \\ = & \int\left\{\frac{1}{n}\sum_{i=1}^n K_{ix}m(X_i)\, X_{ix}'D_h^{-1}\right\}\mathbb{S}^{-1}\left\{\frac{1}{n}\sum_{j=1}^n D_h^{-1}X_{jx}K_{jx}m(X_j)\right\}f^{-1}(x)\,dx \\ & -\left\{\frac{1}{n}\sum_{i=1}^n m(X_i)\right\}^2 + o_p(1) \\ = & \int m^2(x)\,\mathbb{B}'\mathbb{S}^{-1}\mathbb{B}f(x)\,dx - \{E[m(X_1)]\}^2 + o_p(1) = \mu_0 + o_p(1). \end{aligned}$$

Also, $n^{-1}TSS = \sigma_V^2 + o_p(1)$. It follows that $R_q^2 = n^{-1}ESS_q/\left(n^{-1}TSS\right) = \mu_0/\sigma_V^2 + o_p(1)$.

Under H_1, we have $\left(nh^{p/2}\right)^{-1}\widehat{b}_n = o_p(1)$ and $\widehat{\Omega} \xrightarrow{p} \overline{\Omega}$. It follows that

$$\left(nh^{p/2}\right)^{-1}T_n = \frac{R_q^2 - \left(nh^{p/2}\right)^{-1}\widehat{b}_n}{\sqrt{\widehat{\Omega}}} = \frac{\mu_0}{\sigma_V^2\sqrt{\overline{\Omega}}} + o_p(1).■$$

Proof of Theorem 3.4

Let P^* denote the probability conditional on the original sample $\mathcal{W}$. Let $E^*(\cdot)$ and $\mathrm{Var}^*(\cdot)$ denote the expectation and variance with respect to P^*. $a_n = o_{p^*}(1)$ denotes that $P^*(|a_n| \geq \epsilon) \to 0$ for any positive $\epsilon > 0$ as $n \to \infty$. The notation $O_{p^*}(1)$ is similarly defined. Let $\varepsilon_i^* \equiv U_i^{*2} - \widehat{\sigma}^2$, where $\widehat{\sigma}^2 \equiv E^*\left(U_i^{*2}\right) = n^{-1}\sum_{i=1}^n\left(\widehat{U}_i - \overline{\widehat{U}}\right)^2$ and $\overline{\widehat{U}} \equiv n^{-1}\sum_{i=1}^n\widehat{U}_i$. Let $\widehat{\mathbf{g}}^* \equiv \left(g(Z_1, \widehat{\theta}^*), \cdots, g(Z_n, \widehat{\theta}^*)\right)'$, $\mathbf{u}^* \equiv (U_1^{*2}, \cdots, U_n^{*2})'$, $\boldsymbol{\varepsilon}^* \equiv (\varepsilon_1^*, \cdots, \varepsilon_n^*)'$, and $\widehat{\mathbf{v}}^* \equiv \left(\widehat{U}_1^{*2}, \cdots, \widehat{U}_n^{*2}\right)'$. Write $\widehat{U}_i^{*2} = \left[U_i^* + \left(\widehat{U}_i^* - U_i^*\right)\right]^2 = U_i^{*2} + (\widehat{U}_i^* - U_i^*)^2 + 2\left(\widehat{U}_i^* - U_i^*\right)U_i^*$ $= \widehat{\sigma}^2 + \varepsilon_i^* + [g(Z_i, \widehat{\theta}^*) - g(Z_i, \widehat{\theta})]^2 - 2[g(Z_i, \widehat{\theta}^*) - g(Z_i, \widehat{\theta})]U_i^*$. By the symmetry of $H^* - L$ and the fact that $(H^* - L)\,\mathbf{1}_n = \mathbf{0}_n$, we have

$$ESS_q^* \equiv n^{-1}\widehat{\mathbf{v}}^{*\prime}\,(H^* - L)\,\widehat{\mathbf{v}}^* = A_2^* + A_3^* + 4A_4^* + 2A_8^* - 4A_9^* - 4A_{10}^*, \tag{B.8}$$

where A_s^* are the bootstrap analogues of A_s defined in (B.2) for $s = 2, 3, 4, 8, 9, 10$:

$$\begin{aligned}
A_2^* &\equiv \boldsymbol{\varepsilon}^{*\prime}\,(H^* - L)\,\boldsymbol{\varepsilon}^*, \\
A_3^* &\equiv ((\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}})\odot(\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}}))'\,(H^* - L)\,((\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}})\odot(\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}}))\,, \\
A_4^* &\equiv ((\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}}) \odot \mathbf{u}^*)'\,(H^* - L)\,((\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}}) \odot \mathbf{u}^*)\,, \\
A_8^* &\equiv \boldsymbol{\varepsilon}^{*\prime}\,(H^* - L)\,((\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}})\odot(\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}}))\,, \\
A_9^* &\equiv \boldsymbol{\varepsilon}^{*\prime}\,(H^* - L)\,((\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}}) \odot \mathbf{u}^*)\,, \\
A_{10}^* &\equiv ((\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}})\odot(\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}}))\,(H^* - L)\,((\widehat{\mathbf{g}}^* - \widehat{\mathbf{g}}) \odot \mathbf{u}^*)\,.
\end{aligned} \tag{B.9}$$

Let $\widehat{V}_i^* \equiv \widehat{U}_i^{*2}$, and $TSS^* \equiv \sum_{i=1}^n \left(\widehat{V}_i^* - \overline{\widehat{V}}^*\right)^2$ where $\overline{\widehat{V}}^* \equiv n^{-1}\sum_{i=1}^n \widehat{V}_i^*$. It is straightforward to show that

$$n^{-1}TSS^* = E^*\left(\widehat{V}_i^* - \overline{\widehat{V}}^*\right)^2 + o_{p^*}(1) = n^{-1}\sum_{i=1}^n (\widehat{V}_i - \overline{\widehat{V}})^2 + o_{p^*}(1) = \sigma_V^2 + o_{p^*}(1)\,.$$

We prove the theorem by showing that

$$\left\{h^{p/2}A_2^* - h^{p/2}\sum_{i=1}^n \varepsilon_i^{*2}\left(H_{ii}^* - n^{-1}\right)\right\}/\sqrt{\widehat{\Omega}^*\sigma_V^4} \xrightarrow{d} N(0,1)\,, \tag{B.10}$$

$$\widehat{b}_n^* = h^{p/2}\sum_{i=1}^n \varepsilon_i^{*2}\left(H_{ii}^* - n^{-1}\right) + o_p(1)\,, \text{ and} \tag{B.11}$$

$$h^{p/2}A_s^* = o_p(1) \text{ for } s = 3, 4, 8, 9, 10, \tag{B.12}$$

We first show (B.10). Analogously to the proof of (B.3), we have

$$h^{p/2}A_2^* = 2h^{p/2}\sum_{1\le i<j\le n} \varepsilon_i^*\varepsilon_j^*\left(H_{ij}^* - n^{-1}\right) + h^{p/2}\sum_{i=1}^n \varepsilon_i^{*2}\left(H_{ii}^* - n^{-1}\right) \equiv A_{21}^* + A_{22}^*, \text{ say.}$$

Let $v^{*2} \equiv \mathrm{Var}^*\left(U_i^{*2}\right)$. Noting that A_{21}^* is a second order degenerate U-statistic and ε_i^* 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,

$$A_{21}^* \xrightarrow{d} N\left(0, \Omega^*\sigma_V^4\right)$$

where

$$\begin{aligned}
\Omega^* \sigma_V^4 &\equiv \mathrm{p}\lim_{n\to\infty} \frac{2h^p}{n^2} \sum_{i=1}^n \sum_{j\neq i}^n E^* \left[\varepsilon_i^{*2} \varepsilon_j^{*2} \left(\int K_{ix} X_{ix}' D_h^{-1} S_n^{-1}(x) D_h^{-1} X_{jx} K_{jx} dx - 1 \right)^2 \right] \\
&= \mathrm{p}\lim_{n\to\infty} \frac{2h^p v^{*4}}{n^2} \sum_{i=1}^n \sum_{j\neq i}^n \left(\int K_{ix} X_{ix}' D_h^{-1} S_n^{-1}(x) D_h^{-1} X_{jx} K_{jx} dx - 1 \right)^2 \\
&= \mathrm{p}\lim_{n\to\infty} \frac{2h^p v^{*4}}{n^2} \sum_{i=1}^n \sum_{j\neq i}^n \left(\int K_{ix} X_{ix}' D_h^{-1} \mathbb{S}^{-1} D_h^{-1} X_{jx} K_{jx} f^{-1}(x) dx \right)^2 \\
&= \lim_{n\to\infty} 2h^p v^{*4} E_1 E_2 \left[\left(\int K_{1x} X_{1x}' D_h^{-1} \mathbb{S}^{-1} D_h^{-1} X_{2x} K_{2x} f^{-1}(x) dx \right)^2 \right] \\
&= 2v^{*4} \mathrm{vol}(\mathcal{X}) \int \left[\int K(z) \mu(z)' \mathbb{S}^{-1} \mu(z+x) K(z+x) dz \right]^2 dx,
\end{aligned}$$

where $\mathrm{vol}(\mathcal{X}) \equiv \int_{\mathcal{X}} dx$. (B.10) follows as one can easily show that $\widehat{\Omega}^* = \Omega^* + o_p(1)$. Recall $\varepsilon_i^* \equiv U_i^{*2} - \widehat{\sigma}^2$ and $\widehat{\varepsilon}_i^* \equiv \widehat{U}_i^{*2} - \widehat{\sigma}^2$. Then $\widehat{b}_n^* - h^{p/2} \sum_{i=1}^n \varepsilon_i^{*2} \left(H_{ii}^* - n^{-1} \right) = h^{p/2} \sum_{i=1}^n \left(\widehat{\varepsilon}_i^{*2} - \varepsilon_i^{*2} \right) \left(H_{ii}^* - n^{-1} \right)$ $= O_{p^*}\left(n^{-1/2} h^{-p/2} \right) = o_{p^*}(1)$, proving (B.11). Noting that $\widehat{\theta}^* - \widehat{\theta} = O_{p^*}(n^{-1/2})$ under our assumptions, the proof of (B.12) is analogous to that of (B.4) in the proof of Theorem 3.1 and thus omitted. ■

References

Bickel, P. L., (1978), "Using Residuals Robustly I: Tests for Heteroskedasticity in Time Series Models," *Annals of Statistics*, 6, 266-291.

Breusch, T.S., and Pagan, A. R. (1979), "A Simple Test for Heteroskedasticity and Random Variation," *Econometrica*, 47, 1287-1294.

Breusch, T.S., and Pagan, A. R. (1980), "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47, 239-253.

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.

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., Quandt, R.E. (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., and Spokoiny, V. G. (2001), "An Adaptive, Rate-optimal Test of a Parametric Mean-regression Model against a Nonparametric Alternative," *Econometrica*, 69, 599-631.

Hsiao, C., and Li, Q., (2001), "A Consistent Test for Conditional Heteroskedasticity in Time-series Regression Models," *Econometric Theory*, 17, 188-221.

Huang, L-H., and Chen, J., (2008), "Analysis of Variance, Coefficient of Determination and F-test for Local Polynomial Regression," *Annals of Statistics*, 36, 2085-2109.

Koenker, R., and Bassett, G., (1982), "Robust Tests for Heteroskedasticity Based on Quantiles," *Econometrica*, 50, 159-171.

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., and Paparoditis E. (2000), "On Bootstrapping L_2-type Statistics in Density Testing," *Statistics & Probability Letters*, 50, 137-147.

Newey, W., and Powell J. (1987), "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819-847.

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

Su, L., and Ullah, A. (2009), "Testing Conditional Uncorrelatedness," *Journal of Business and Economic Statistics*, 27, 18-29.

Sun, S., and Chiang, C.Y. (1997), "Limiting Behavior of the Perturbed Empirical Distribution Functions Evaluated at U-statistics for Strongly Mixing Sequences of Random Variables," *Journal of Applied Mathematics and Stochastic Analysis*, 10, 3-20.

White, H. (1980), "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 452-475.

White, H., and Domowitz, I. (1984), "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143-161.

Zheng, X. (2006), "Testing Heteroscedasticity in Nonlinear and Nonparametric Regressions with an Application to Interest Rate Volatility," *Manuscript*, School of Economics, Shanghai Jiaotong Univ.