A NONPARAMETRIC R^2 TEST FOR THE PRESENCE OF RELEVANT VARIABLES[1]

Feng Yao

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
email: feng.yao@mail.wvu.edu
Voice: +1 304 293 7867

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: f.yao@cgiar.org
Voice: + 1 202 862 6488

Aman Ullah

Department of Economics
University of California
Riverside, CA 92521-0427 USA
email: aman.ullah@ucr.edu
Voice:+1 909 787 5685
Fax: + 1 951 827 5685

January, 2013

Abstract. We propose a nonparametric test for the presence of relevant variables based on a measure of nonparametric goodness-of-fit (R^2) in a regression model. It does not require correct specifications of the conditional mean function, thus is able to detect presence of relevant variables of unknown form. Our test statistic is based on an appropriately centered and standardized nonparametric R^2 estimator, which is obtained from a local linear regression. We establish the asymptotic normality of the test statistic under the null hypothesis that relevant variables are not present and a sequence of Pitman local alternatives. We also prove the consistency of the test, and show that the Wild bootstrap/bootstrap method can be used to approximate the null distribution of the test statistic. Under the alternative hypothesis, we establish the asymptotic normality of the nonparametric R^2 estimator at rate $\sqrt{n}$, which facilitates inference using the nonparametric measure of goodness-of-fit. We illustrate the finite sample performance of the tests with a Monte Carlo study and the bootstrap tests perform well relative to other alternatives.

Keywords and phrases: Omitted variables, nonparametric R^2, nonparametric test, local linear regression.

JEL Classifications: C12, C14.

[1]Corresponding author: Feng Yao. We thank the editor and an anonymous referee for comments that improved the paper substantially. Any remaining errors are the authors' responsibility. We are thankful to Liangjun Su and Santosh Mishra for useful discussions on the subject matter of our paper. The second author is also grateful to the Academic Senate, UCR for the financial support.

1 Introduction

Test for the presence of relevant variables or for omitted variables has been of interest in regression analysis since it is often used to support, reject an economic theory or considered for model selection. Conventional tests, including the t and F tests, specify particular parametric forms in the null and alternative hypothesis, are not consistent or do not have good power since the deviation from the null does not always lead to the path dictated in the alternative. Nonparametric tests therefore have appeal to practitioners, given that the test has power in all deviations from the null, see Li and Racine (2007).

Nonparametric test for relevant variables, and for functional form specification, a related topic, has been the focus of many recent papers, see Hart (1997) for a review of the use of nonparametric regression methodology in testing the fit of parametric regression models. Fan and Li (1996), Zheng (1996), Li and Wang (1998), Li (1999), Lavergne and Vuong (2000), Hsiao et al. (2007) and Gu et al. (2007) propose consistent tests of the functional form, and omitted variables with a kernel based test that is the sample analog of a moment condition. Ullah (1985) suggests testing for the correct parametric regression functional form based on the difference between sums of squared residuals. This approach has been taken in Dette (1999), and Fan and Li (2002) to test a parametric functional form specification. Hardle and Mammen (1993) construct a test with the integrated squared difference between the parametric and nonparametric kernel fit, to decide whether the parametric model could be justified. Among others, Azzalini et al. (1989), Azzalini and Bowman (1993) propose using nonparametric kernel regression to check the fit of a parametric model with a pseudo likelihood ratio test. Fan et al. (2001) introduce the generalized likelihood ratio (GLR) tests, which exhibit the Wilks phenomenon and are asymptotically optimal. They can be used to test the goodness-of-fit for a family of parametric models. Hong and Lee (2009) propose a loss function based model specification test, which enjoys the good properties of the GLR test. From a technical perspective, we note that above approaches utilize the fact that the test statistic is a degenerate U-statistic after proper normalization, and converges at a rate faster than $\sqrt{n}$. Different techniques have been used in constructing consistent tests for omitted variables. Racine (1997) proposes a significance test based on nonparametric estimates of partial derivatives, employing pivotal bootstrapping procedures. Hidalgo (1992) uses random weighting and Gozalo (1993) introduces the random search procedure, where the test statistic's distribution is determined by a random term whose order is larger than the degenerating U-statistic. Yatchew (1992) uses sample splitting to circumvent the $\sqrt{n}$-degeneracy problem in a nested situation, and Lavergne and Vuong (1996) treat the non-nested case.

On a related subject, the goodness-of-fit measure such as coefficient of determination or R^2 provides a concise summary of regression model, i.e., the variability of regressand y explained by the variability of regressors. Nonparametric estimation of R^2 has been considered by, among others, Doksum and Samarov (1995), and Martins-Filho and Yao (2006). Recently Huang and Chen (2008) propose a R^2 estimator based on local polynomial regressions. It has a sample ANOVA decomposition that the total sum of squares is equal to the explained sum of squares and the residual sum of squares, facilitating the interpretability of nonparametric R^2 estimations. We think the nonparametric R^2 estimators provide useful statistics for testing many popular hypotheses in econometrics and statistics, and could play an important role just as 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 formulated as nR^2 (Green (2000)), where n is the sample size and R^2 is the coefficient of determination from some residual based and parametrically specified auxiliary regressions. In case the functional form in the auxiliary regressions is misspecified, these tests may lead to misleading conclusions. The nonparametric R^2 estimator allows the functional form to be flexible, thus avoids misspecifications. It provides the basis to construct nonparametric tests, as the analogue of the parametric residual based test. For example, Su and Ullah (2012) propose a nonparametric goodness-of-fit test for the conditional heteroskedasticity.

In this paper, we propose new tests for the presence of continuous relevant variables based on estimators of the nonparametric R^2 of a theoretical ANOVA decomposition or the nonparametric coefficient of determination considered by Doksum and Samarov (1995) in a regression model. Different from Doksum and Samarov (1995) whose focus is on estimation of R^2, where the nonparametric R^2 estimator is constructed with the *leave-one-out* local constant estimator and with a weight function that is equal to zero near the boundary of the support of regressors, we construct the nonparametric R^2 estimators $\hat{R}^2$ for the simple regression and $\hat{R}^2_G$ for the multiple regression with a local linear estimator which is known to possess better boundary properties. In addition, simulation results in Tables 2-4 in Doksum and Samarov (1995) indicate that nonparametric R^2 estimator based on local linear estimators could outperform that based on *leave-one-out* local constant estimators. Furthermore, we include an indicator function in the R^2 estimators such that they are always within $[0, 1]$, while two of the estimators by Doksum and Samarov (1995) may be negative or greater than one with some small probability. Focusing on the estimation of R^2 for $R^2 \in (0, 1)$, their results only imply degenerate normality when $R^2 = 0$ or 1. They mention in their Remark 2.7 the need to study the terms in the expansions to obtain a meaningful distribution convergence result, which is a nontrivial task.

Constructing new tests when $R^2 = 0$ under the null hypothesis to assess significance of explanatory variables is the focus of our paper. First, using the fact that our nonparametric R^2 estimators are small and close to zero under the null that some regressors X are irrelevant, but lie away from zero under the alternative that X are relevant, we develop the test statistic $\hat{T}_n$ based on a properly normalized $\hat{R}^2$. Under the null ($R^2 = 0$) and a sequence of Pitman local alternatives, $\hat{T}_n$ is asymptotically normal at a rate of $nh_n^{\frac{1}{2}}$. Under the global alternative hypothesis ($0 < R^2 < 1$), the asymptotic normality of $\hat{R}^2 - R^2$ is obtained at rate $\sqrt{n}$, thus, the rates of convergence are different in both cases. The result enables us to obtain the consistency of the proposed test. Second, we further propose a Wild bootstrap/bootstrap test and show that it can approximate the null distribution of the test statistic. These two results enable us to propose an asymptotic test as well as a bootstrap test based on two estimators considered by Doksum and Samarov (1995). We obtain their asymptotic properties and compare them via simulations together with above tests. Third, we propose the generalized nonparametric R^2 ($\hat{R}^2_G$) based tests, $\hat{T}_{nG}$, and the bootstrap test $\hat{T}^*_{nG}$ in the multiple regression model, obtain their asymptotic properties, and demonstrate their validity in testing significant variables theoretically and empirically in simulations.

Our test statistic has the following features. We test a nonparametric null that the variables are not present against a nonparametric alternative. Our tests do not use either the randomization or the sample splitting, and deal with the $\sqrt{n}$-degeneracy problem by obtaining the distribution of test statistic directly at rate $nh_n^{\frac{d}{2}}$, where d is the dimension of regressors. The tests are easy to conduct as they are based on local linear regressions, and they can detect sequences of local alternatives that differ from the null at the rate of $(nh_n^{\frac{d}{2}})^{-\frac{1}{2}}$. The test does not require any knowledge of the true likelihood, nor does it require homoskedasticity of the regression errors. When we test for the overall significance, the test is related to the GLR test, but they are numerically different. Under the homoskedasticity assumption, they have the same asymptotic distribution, and the test exhibits the Wilks phenomenon and is asymptotically optimal. Simulation result indicates that our test behaves well in finite sample compared to some alternatives available in the literatures.

The plan of our paper is as follows. We define the R^2 estimators and test statistics in Section 2, state the assumptions and the asymptotic properties of the estimators and tests in Section 3, conduct a Monte Carlo study to illustrate the tests' finite sample performance and compare them with other alternatives in Section 4, and conclude in Section 5. Table 1 is provided in Appendix 1 and the proof of Theorem 6 is relegated to Appendix 2. The statement of three lemmas, the detailed proofs of Theorems 1-5 and remarks 1 and 2 are collected in a separate Appendix (Yao and Ullah (2013)).

2 A nonparametric R^2 test

2.1 Asymptotic nonparametric R^2 tests

Let's consider a nonparametric regression model

$$y_t = m(X_t) + \epsilon_t, t = 1, 2, \cdots, n, \tag{1}$$

where $m(X_t) = E(y_t|X_t), E(\epsilon_t|X_t) = 0, V(\epsilon_t|X_t) = \sigma^2(X_t)$ and $X_t \in \Re^d$.

For the ease of illustration, we start by considering $d = 1$ and whether $X_t \equiv x_t$ is present. If x_t does not show up, then $E(y_t|x_t) = \mu = E(y_t)$. So the null and alternative hypotheses are

$$H_0 : P(E(y_t|x_t) = \mu) = 1, H_1 : P(E(y_t|x_t) = \mu) < 1.$$

Under H_0 that x_t is not present in $m(x_t)$, any goodness of fit measure should be close to zero. Following Doksum and Samarov (1995) to construct the nonparametric R^2 measure based on the theoretical ANOVA decomposition of variance, $V(y) = V(E(y|x)) + E(V(y|x)) = V(m(x)) + E\sigma^2(x)$, the theoretical coefficient of determination is $R^2 = \frac{V(E(y|x))}{V(y)} = \frac{V(m(x))}{V(y)} = 1 - \frac{E(y-m(x))^2}{V(y)}$. We focus on the local linear estimator $\hat{m}(x) = \hat{\alpha}$ for $m(x)$ popularized by Fan (1992) due to its well known desirable properties, where

$$(\hat{\alpha}, \hat{\beta}) = argmax_{\alpha,\beta} \sum_{t=1}^{n} (y_t - \alpha - \beta(x_t - x))^2 K(\frac{x_t - x}{h_n}).$$

$K(\cdot) : \Re \to \Re$ is a kernel function and $0 < h_n \to 0$ as $n \to \infty$ is a bandwidth.

We propose the following nonparametric R^2 estimator,

$$\hat{R}^2 = \left[1 - \frac{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2} \right] I(\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2). \tag{2}$$

$\bar{y}$ is the average of y and $I(\cdot)$ is the indicator function. Note that in general $\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2 \neq \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2 + \frac{1}{n}\sum_{t=1}^{n}(\hat{m}(x_t) - \bar{y})^2$. $\hat{R}^2$ resembles the nonparametric R^2 estimator $\hat{\eta}_1^2$ proposed by Doksum and Samarov (1995). The main differences lie in that we use the local linear estimator which possesses good boundary properties and include the indicator function $I(\cdot)$ such that $\hat{R}^2$ always takes value in $[0, 1]$, while $\hat{\eta}_1^2$, constructed with the *leave-one-out* local constant estimator and with a weight function that is equal to zero near the boundary of the support of the regressors, may be negative or greater than one with some small probability. The smaller the value of $\hat{R}^2$, the worse the fit. In the extreme case that no regressors in x_t can explain y_t, we expect a value close to zero in a given sample of $\{y_t, x_t\}_{t=1}^n$.

We construct the test statistic based on a properly centered and scaled $\hat{R}^2$. Specifically, define the marginal density of x_t at x as $f(x)$. Suppose we know $f(x_t)$, ϵ_t and $\sigma^2(x)$. Define $A_n = \frac{1}{n^3 h_n^2}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ t\neq i}}^{n} K^2(\frac{x_i - x_t}{h_n})\frac{\epsilon_i^2}{f^2(x_t)}$, $A_{1n} = -\frac{2}{n^2 h_n}\sum_{t=1}^{n} K(0)\frac{\epsilon_t^2}{f(x_t)}$, $\sigma_\phi^2 = 2E\frac{\sigma^4(x_t)}{f(x_t)}\int(2K(\psi) - \kappa(\psi))^2 d\psi$, with $\kappa(\psi) = \int K(x)K(\psi + x)dx$ as the convolution of kernel function $K(\cdot)$, and $V_T = \frac{\sigma_\phi^2}{(V(y))^2}$. We construct the infeasible test statistic as

$$T_n = \frac{nh_n^{\frac{1}{2}}\{\hat{R}^2 + I(\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2)\frac{A_{1n} + A_n}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2}\}}{\sqrt{V_T}}.$$

Under H_0 and assumptions in next section, we show in Theorem 1 that T_n asymptotically has a standard normal distribution, which provides the asymptotic theory to construct hypothesis tests. Here A_n and A_{1n} are the "bias" terms used to center $\hat{R}^2$ around zero and $\sigma_\phi^2/(V(y))^2$ are the asymptotic variance of the centered $nh_n^{\frac{1}{2}}\hat{R}^2$. We need to estimate the unknowns in T_n to implement the test. We consider the Rosenblatt (1956) density estimator for $f(x)$ as $\hat{f}(x) = \frac{1}{nh_n}\sum_{t=1}^{n} K(\frac{x_t - x}{h_n})$. Let $\tilde{\epsilon}_t = y_t - \bar{y}$. We note that under H_0, $\tilde{\epsilon}_t$ can estimate ϵ_t at rate $\sqrt{n}$ since $\bar{y}$ is a $\sqrt{n}$ consistent estimator for μ. Define $\hat{A}_n = \frac{1}{n^3 h_n^2}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ t\neq i}}^{n} K^2(\frac{x_i - x_t}{h_n})\frac{\tilde{\epsilon}_i^2}{\hat{f}^2(x_t)}$, $\hat{A}_{1n} = -\frac{2}{n^2 h_n}\sum_{t=1}^{n} K(0)\frac{\tilde{\epsilon}_t^2}{\hat{f}(x_t)}$, $\hat{V}_T = \frac{\hat{\sigma}_\phi^2}{(\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2)^2}$, *and*[1] $\hat{\sigma}_\phi^2 = [\frac{1}{n^2}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ t\neq i}}^{n} K(\frac{x_i - x_t}{h_n})\frac{\tilde{\epsilon}_i^2\tilde{\epsilon}_t^2}{h_n \hat{f}^2(x_t)}](\int 2(2K(\psi) - \kappa(\psi))^2 d\psi)$. We construct the feasible test statistic as

$$\hat{T}_n = \frac{nh_n^{\frac{1}{2}}\{\hat{R}^2 + I(\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2)\frac{\hat{A}_{1n} + \hat{A}_n}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2}\}}{\sqrt{\hat{V}_T}}. \tag{3}$$

[1] An alternative consistent estimator $\tilde{\sigma}_\phi^2 = \frac{2}{n^2}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ t\neq i}}^{n}\frac{\tilde{\epsilon}_i^2\tilde{\epsilon}_t^2}{h_n \hat{f}^2(x_t)}(2K(\frac{x_i - x_t}{h_n}) - \kappa(\frac{x_i - x_t}{h_n}))^2$ can also be used in place of $\hat{\sigma}_\phi^2$.

Inclusion of indicator function $I(\cdot)$ makes sure that the nonparametric R^2 estimate is within the range of zero to one. Since $I(.)$ converges to one in probability as shown in the proof of Theorem 3, we expect the test statistic without the indicator function is equivalent to $\hat{T}_n$ asymptotically in distribution, though numerically the test statistics are different. We use the residual $\tilde{\epsilon}_t$ from the null model to define $\hat{A}_{1n}$ and $\hat{A}_n$ in $\hat{T}_n$. This eliminates the asymptotic "bias" and the test has the right size. Furthermore, $\hat{T}_n$ has a local power as noted later in Theorem 2, and is consistent as shown in Theorem 4.

Doksum and Samarov (1995) have proposed three alternative nonparametric R^2 estimators. By introducing a weight function $w(x)$ supported on a set where the density of x is bounded away from zero, they consider the weighted R^2 measure as $\eta_w^2 = \frac{\int (m(x)-\mu_{y,w})^2 f(x)w(x)dx}{\sigma_{y,w}^2}$, where $\mu_{y,w} = \int w(x)yf(x,y)dxdy$ and $\sigma_{y,w}^2 = \int (y-\mu_{y,w})^2 f(x,y)w(x)dxdy$. We note that η_w^2 is also equal to $corr_w^2(m(x),y)$, the square of the weighted correlation measure between $m(x)$ and y. The first two R^2 estimators are motivated by η_w^2 and constructed as $\hat{\eta}_1^2 = \frac{\frac{1}{n}\sum_{i=1}^{n} w(x_i)[2y_i\tilde{m}(x_i) - \tilde{m}^2(x_i)] - \bar{y}_w^2}{S_y^2}$, and $\hat{\eta}_2^2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(\tilde{m}(x_i)-\bar{m})^2 w(x_i)}{S_y^2}$, where $\tilde{m}(x_i) = \frac{((n-1)h_n)^{-1}\sum_{j\neq i} y_j K(\frac{x_j-x_i}{h_n})}{((n-1)h_n)^{-1}\sum_{j\neq i} K(\frac{x_j-x_i}{h_n})} = \frac{\tilde{g}(x_i)}{\tilde{f}(x_i)}$ is the *leave-one-out* local constant estimator, $\bar{m} = \frac{1}{n}\sum_{i=1}^{n}\tilde{m}(x_i)w(x_i)$, and $S_y^2 = n^{-1}\sum_{i=1}^{n}(y_i-\bar{y}_w)^2 w(x_i)$ for $\bar{y}_w = n^{-1}\sum_{i=1}^{n} y_i w(x_i)$. The third estimator is motivated by $corr_w^2(m(x),y)$ and constructed as $\hat{\eta}_3^2 = \frac{[\frac{1}{n}\sum_{i=1}^{n}(\tilde{m}(x_i)-\bar{m})(y_i-\bar{y}_w)w(x_i)]^2}{\frac{1}{n}\sum_{i=1}^{n}(\tilde{m}(x_i)-\bar{m})^2 w(x_i)s_y^2}$.

Now we extend the test $\hat{T}_n$ in equation (3), and develop two tests based on $\hat{\eta}_1^2$ and $\hat{\eta}_2^2$ for $d=1$. Though the simulation results in Doksum and Samarov (1995) recommend $\hat{\eta}_1^2$ and $\hat{\eta}_3^2$ over $\hat{\eta}_2^2$ when estimating the nonparametric R^2 as $\hat{\eta}_2^2$ is sensitive to the choice of bandwidth, we find that $\hat{\eta}_3^2$ can not be directly used to construct a test statistic as its denominator converges in probability to zero under H_0. Based on *leave-one-out* local constant estimators, they are $\hat{T}_{1n} = \frac{nh^{1/2}[\hat{\eta}_1^2 + (S_y^2)^{-1}\hat{T}_{n0}]}{\sqrt{\hat{\sigma}_{\phi 1}^2/S_y^4}}$ and $\hat{T}_{2n} = \frac{nh^{1/2}[\hat{\eta}_2^2 - (S_y^2)^{-1}\hat{T}_{n0}]}{\sqrt{\hat{\sigma}_{\phi 2}^2/S_y^4}}$, $\hat{T}_{n0} = \frac{1}{n(n-1)^2h_n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} K^2(\frac{x_j-x_i}{h_n})\tilde{\epsilon}_j^2\frac{w(x_i)}{\tilde{f}^2(x_i)}$, $\hat{\sigma}_{\phi 1}^2 = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{\substack{i=1\\ t\neq i}}^{n} K(\frac{x_i-x_t}{h_n})\frac{\tilde{\epsilon}_i^2\tilde{\epsilon}_t^2}{h_n\tilde{f}^2(x_t)}w^2(x_t)2\int (2K(\psi)-\kappa(\psi))^2 d\psi$, $\hat{\sigma}_{\phi 2}^2 = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{\substack{i=1\\ t\neq i}}^{n} K(\frac{x_i-x_t}{h_n})\frac{\tilde{\epsilon}_i^2\tilde{\epsilon}_t^2}{h_n\tilde{f}^2(x_t)}w^2(x_t)2\int \kappa^2(\psi)d\psi$ and $\tilde{\epsilon}_i = y_i - \bar{y}_w$. The tests bear resemblance to $\hat{T}_n$ as they are based on appropriately centered and scaled R^2 estimators. Besides the difference in the R^2 estimators, we notice that there is only one "bias" term in $\hat{T}_{in}$ for $i=1,2$ and they are the same except for the opposite sign. We note that the scaling factors $\hat{\sigma}_{\phi 1}^2$ and $\hat{\sigma}_{\phi 2}^2$ differ only on a constant factor related to the kernel function, while they deviate from $\hat{\sigma}_\phi^2$ in $\hat{T}_n$ further in the residuals and the weight function.

Let us consider a more general regression model

$$y_t = m(X_t) + \epsilon_t, t = 1, 2, \cdots, n, \tag{4}$$

where $X_t' = (x_{1t}, x_{2t})' \in \Re^{d_1+d_2}$, $d = d_1 + d_2$, and $E(\epsilon_t|X_t) = 0$. Under the null hypothesis that x_{2t} are irrelevant, we have $H_{0G}: P(E(y_t|X_t) = E(y_t|x_{1t})) = 1$. Thus,

$$E(y_t - E(y_t|x_{1t}))^2 = E(y_t - E(y_t|X_t))^2 + E[E(y_t|X_t) - E(y_t|x_{1t})]^2,$$

and the last term is equal to zero only when the null hypothesis is true. A generalized version of the coefficient of determination is $R_G^2 = 1 - \frac{E(y_t - E(y_t|X_t))^2}{E(y_t - E(y_t|x_{1t}))^2}$ in equation (4.9) of Doksum and Samarov (1995). They use the estimator $\hat{\eta}_3^2$ to estimate R_G^2, but as we point out before, $\hat{\eta}_3^2$ is not suitable for constructing the test statistic as its denominator converges to zero under H_{0G}. It motivates us to consider the generalized nonparametric R^2 estimator

$$\hat{R}_G^2 = \left[1 - \frac{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2}\right] I(\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2),$$

where $\hat{m}(X_t)$ is the multivariate local linear estimator of $m(X_t) = E(y_t|X_t)$. We estimate the conditional mean $r(x_{1t}) = E(y_t|x_{1t})$ by $\hat{r}(x_{1t}) = \hat{\alpha}_0$, where $\hat{\alpha}_0$ is the local linear estimator constructed from $(\hat{\alpha}_0, \hat{\alpha}_1) = argmax_{\alpha_0,\alpha_1} \sum_{i=1}^{n}(y_i - \alpha_0 - (x_{1i} - x_{1t})\alpha_1)^2 K_1(\frac{x_{1i}-x_{1t}}{h_{1n}})$, in which $K_1(\cdot) : \Re^{d_1} \to \Re$ is a kernel function, and $0 < h_{1n} \to 0$ as $n \to \infty$ is a bandwidth, which is assumed to be the same for all elements in x_{1t}. We can construct the generalized nonparametric R^2 test as

$$\hat{T}_{nG} = \frac{nh_n^{\frac{d}{2}}\{\hat{R}_G^2 + I(\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2)\frac{\hat{A}_{1nG}+\hat{A}_{nG}}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2}\}}{\sqrt{\hat{V}_{TG}}}. \quad (5)$$

Here, we define the multivariate Rosenblatt (1956) density estimator $\hat{f}(X_t) = \frac{1}{nh_n^d}\sum_{i=1}^{n} K(\frac{X_i - X_t}{h_n})$,
$\hat{A}_{nG} = \frac{1}{n^3 h_n^{2d}} \sum_{t=1}^{n}\sum_{i=1, t\neq i}^{n} K^2(\frac{X_i - X_t}{h_n})\frac{\tilde{\epsilon}_i^2}{\hat{f}^2(X_t)}$, $\hat{A}_{1nG} = -\frac{2}{n^2 h_n^d}\sum_{t=1}^{n} K(0)\frac{\tilde{\epsilon}_t^2}{\hat{f}(X_t)}$, $\hat{V}_{TG} = \frac{\hat{\sigma}_{\phi G}^2}{(\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2)^2}$,
$\hat{\sigma}_{\phi G}^2 = [\frac{1}{n^2}\sum_{t=1}^{n}\sum_{i=1, t\neq i}^{n} K(\frac{X_i - X_t}{h_n})\frac{\tilde{\epsilon}_i^2 \tilde{\epsilon}_t^2}{h_n^d \hat{f}^2(X_t)}](\int 2(2K(\psi) - \kappa(\psi))^2 d\psi)$, and $\tilde{\epsilon}_t = y_t - \hat{r}(x_{1t})$.

2.2 Bootstrap tests

The asymptotic distributions of the nonparametric R^2 estimators and tests are provided in the next section. For $d = 1$, one can perform the test for H_0 by comparing the value of $\hat{T}_n$ with its asymptotic critical values. However, many papers have revealed that the asymptotic normal approximation performs poorly in finite sample settings. Specifically, the consistent nonparametric test often suffers from substantial finite sample size distortions, as the distribution of the nonparametric test statistic approaches asymptotically the normal distribution at a slow convergence rate (e.g., Hardle and Mammen (1993), Li and Wang (1998), Fan et al. (2006), Hsiao et al. (2007), and Gu et al. (2007)). Therefore, we provide a Wild bootstrap test as a viable alternative for approximating the finite sample null distribution of the test statistic $\hat{T}_n$. Let $\hat{\epsilon}_t = y_t - \hat{m}(x_t)$ for $t = 1, \cdots, n$. The bootstrap test contains the following steps:

Step 1: generate ϵ_t^* as the wild bootstrap error. For example, ϵ_t^* is generated independently from the two point distribution $\hat{F}_t$ such that $\epsilon_t^* = a\hat{\epsilon}_t$ for $a = \frac{1-\sqrt{5}}{2}$ with probability $p = \frac{\sqrt{5}+1}{2\sqrt{5}}$, and $\epsilon_t^* = b\hat{\epsilon}_t$ for $b = \frac{1+\sqrt{5}}{2}$ with probability $1 - p$. It is called the wild bootstrap error because we use only single residual $\hat{\epsilon}_t$ to estimate the conditional distribution of ϵ_t given x_t by $\hat{F}_t$. It does not mimic the iid structure of $\{x_t, y_t\}_{t=1}^n$. It is easy to verify that $E_{\hat{F}_t}(\epsilon_t^*) = 0$, $E_{\hat{F}_t}(\epsilon_t^*)^2 = \hat{\epsilon}_t^2$, and $E_{\hat{F}_t}(\epsilon_t^*)^3 = \hat{\epsilon}_t^3$.

Step 2: generate y_t^* according to the null model, i.e., $y_t^* = \bar{y} + \epsilon_t^*$, for $t = 1, \cdots, n$. Then use the bootstrap sample $\{x_t, y_t^*\}_{t=1}^n$ to estimate $m(x_t)$ under H_0, which gives $\hat{\mu}^* = \frac{1}{n}\sum_{t=1}^n y_t^*$. We define the bootstrap residual based on H_0 as $\epsilon_{t,0}^* = y_t^* - \hat{\mu}^*$ for $t = 1, \cdots, n$.

Step 3: obtain the nonparametric bootstrap residual as $\epsilon_{t,b}^* = y_t^* - \hat{m}^*(x_t), t = 1, \cdots, n$, where $\hat{m}^*(x_t)$ is the local linear estimate obtained with the bootstrap sample $\{x_t, y_t^*\}_{t=1}^n$.

Step 4: compute the bootstrap test statistic

$$\hat{T}_n^* = \frac{nh_n^{\frac{1}{2}}\{\hat{R}^{2*} + I(\frac{1}{n}\sum_{t=1}^n(\epsilon_{t,0}^*)^2 \geq \frac{1}{n}\sum_{t=1}^n(\epsilon_{t,b}^*)^2)\frac{\hat{A}_{1n}^* + \hat{A}_n^*}{\frac{1}{n}\sum_{t=1}^n(\epsilon_{t,0}^*)^2}\}}{\sqrt{\hat{V}_T^*}}, \ \hat{R}^{2*} = \left[1 - \frac{\frac{1}{n}\sum_{t=1}^n(\epsilon_{t,b}^*)^2}{\frac{1}{n}\sum_{t=1}^n(\epsilon_{t,0}^*)^2}\right] I(\frac{1}{n}\sum_{t=1}^n(\epsilon_{t,0}^*)^2 \geq \frac{1}{n}\sum_{t=1}^n(\epsilon_{t,b}^*)^2),$$

$$\hat{A}_n^* = \frac{1}{n^3h_n^2}\sum_{\substack{t=1\\t\neq i}}^n\sum_{i=1}^n K^2(\frac{x_i - x_t}{h_n})\frac{(\epsilon_{i,0}^*)^2}{\hat{f}^2(x_t)}, \hat{A}_{1n}^* = -\frac{2}{n^2h_n}\sum_{t=1}^n K(0)\frac{(\epsilon_{t,0}^*)^2}{\hat{f}(x_t)}, \hat{V}_T^* = \frac{\hat{\sigma}_\phi^{2*}}{(\frac{1}{n}\sum_{t=1}^n(\epsilon_{t,0}^*)^2)^2},$$

$$and^2\ \hat{\sigma}_\phi^{2*} = [\frac{1}{n^2}\sum_{\substack{t=1\\t\neq i}}^n\sum_{i=1}^n K(\frac{x_i - x_t}{h_n})\frac{(\epsilon_{i,0}^*)^2(\epsilon_{t,0}^*)^2}{h_n\hat{f}^2(x_t)}](\int 2(2K(\psi) - \kappa(\psi))^2 d\psi).$$

Step 5: repeat above four steps B times, and B a large number. Then the original test statistic $\hat{T}_n$ and the B bootstrap test statistics $\hat{T}_n^*$ give us the empirical distribution of the bootstrap statistics, which is then used to approximate the finite sample null distribution of $\hat{T}_n$. The p-value is obtained as the percentage of the number of times that $\hat{T}_n^*$ exceeds $\hat{T}_n$ in the B repetitions.

For the tests based on alternative nonparametric R^2 estimators, we extend the test $\hat{T}_n^*$ and propose the following bootstrap test $\hat{T}_{1n}^*$ based on $\hat{T}_{1n}$ using $\hat{\epsilon}_t = y_t - \tilde{m}(x_t)$ for $t = 1, \cdots, n$.

Step 1: generate ϵ_t^* as in step 1 of the bootstrap test $\hat{T}_n^*$.

Step 2: generate $y_t^* = \bar{y}_w + \epsilon_t^*$ for $t = 1, \cdots, n$. Then use the bootstrap sample $\{x_t, y_t^*\}_{t=1}^n$ to estimate $m(x_t)$ under H_0, which is $\bar{y}_w^* = \frac{1}{n}\sum_{i=1}^n y_i^* w(x_i)$. Define the bootstrap residual based on H_0 as $\epsilon_{t,0}^* = y_t^* - \bar{y}_w^*$.

Step 3: obtain the nonparametric bootstrap residual as $\epsilon_{t,b}^* = y_t^* - \tilde{m}^*(x_t)$ for $t = 1, \cdots, n$, where $\tilde{m}^*(x_t)$ is the *leave-one-out* local constant estimate obtained with the bootstrap sample $\{x_t, y_t^*\}_{t=1}^n$.

Step 4: compute the bootstrap test statistic $\hat{T}_{1n}^* = \frac{nh^{1/2}[\hat{\eta}_1^{*2} + (S_y^{*2})^{-1}\hat{T}_{n0}^*]}{\sqrt{\hat{\sigma}_{\phi 1}^{*2}/S_y^{*4}}}$, with $S_y^{*2} = \frac{1}{n}\sum_{i=1}^n(\epsilon_{i,0}^*)^2 w(x_i)$,

$$\hat{\eta}_1^{*2} = \frac{\frac{1}{n}\sum_{i=1}^n w(x_i)[2y_i^*\tilde{m}^*(x_i) - \tilde{m}^{*2}(x_i)] - \bar{y}_w^{*2}}{S_y^{*2}}, \ \hat{T}_{n0}^* = \frac{1}{n(n-1)^2h_n^2}\sum_{\substack{i=1\\i\neq j}}^n\sum_{j=1}^n K^2(\frac{x_j - x_i}{h_n})(\epsilon_{j,0}^*)^2\frac{w(x_i)}{\hat{f}^2(x_i)}, \text{ and}$$

$$\hat{\sigma}_{\phi 1}^{*2} = \frac{1}{n^2}\sum_{\substack{t=1\\t\neq i}}^n\sum_{i=1}^n K(\frac{x_i - x_t}{h_n})\frac{(\epsilon_{i,0}^*)^2(\epsilon_{t,0}^*)^2}{h_n\hat{f}^2(x_t)}w^2(x_t)2\int(2K(\psi) - \kappa(\psi))^2 d\psi.$$

Step 5: as in step 5 of the bootstrap test $\hat{T}_n^*$, with $\hat{T}_n$ replaced by $\hat{T}_{1n}$.

The bootstrap test $\hat{T}_{2n}^*$ based on $\hat{T}_{2n}$ is defined with steps 1-3 as above, but with

Step 4: compute the bootstrap test statistic $\hat{T}_{2n}^* = \frac{nh^{1/2}[\hat{\eta}_2^{*2} - (S_y^{*2})^{-1}\hat{T}_{n0}^*]}{\sqrt{\hat{\sigma}_{\phi 2}^{*2}/S_y^{*4}}}$, with $\bar{m}^* = \frac{1}{n}\sum_{i=1}^n \tilde{m}^*(x_i)w(x_i)$,

$$\hat{\eta}_2^{*2} = \frac{\frac{1}{n}\sum_{i=1}^n(\tilde{m}^*(x_i) - \bar{m}^*)^2 w(x_i)}{S_y^{*2}}, \ \hat{\sigma}_{\phi 2}^{*2} = \frac{1}{n^2}\sum_{\substack{t=1\\t\neq i}}^n\sum_{i=1}^n K(\frac{x_i - x_t}{h_n})\frac{(\epsilon_{i,0}^*)^2(\epsilon_{t,0}^*)^2}{h_n\hat{f}^2(x_t)}w^2(x_t)2\int\kappa^2(\psi)d\psi.$$

Step 5: as in step 5 of the bootstrap test $\hat{T}_n^*$ with $\hat{T}_n$ replaced by $\hat{T}_{2n}$.

[2] An alternative estimator $\tilde{\sigma}_\phi^{2*} = \frac{2}{n^2}\sum_{\substack{t=1\\t\neq i}}^n\sum_{i=1}^n \frac{(\epsilon_{i,0}^*)^2(\epsilon_{t,0}^*)^2}{h_n\hat{f}^2(x_t)}(2K(\frac{x_i - x_t}{h_n}) - \kappa(\frac{x_i - x_t}{h_n}))^2$ can also be used in place of $\hat{\sigma}_\phi^{2*}$.

For the test in the general regression model, we construct the bootstrap test $\hat{T}^*_{nG}$ following the five steps of $\hat{T}^*_n$, where the univariate x_t is replaced with X_t, with steps 2 and 4 replaced by

Step 2: generate $y^*_t = \hat{r}(x_{1t}) + \epsilon^*_t$ according to the null model. The bootstrap sample is $\{X_t, y^*_t\}^n_{t=1}$. We use the bootstrap sample to estimate $m(X_t) = r(x_{1t})$ under H_{0G} by $\hat{r}^*(x_{1t})$, which is obtained with the local linear estimator by regressing y^*_t on x_{1t}. We define the bootstrap residual based on H_{0G} as $\epsilon^*_{t,0} = y^*_t - \hat{r}^*(x_{1t})$ for $t = 1, \cdots, n$.

Step 4: compute the bootstrap test statistic $\hat{T}^*_{nG} = \frac{nh_n^{\frac{d}{2}}[\hat{R}^{*2}_G + \frac{(\hat{A}^*_{nG} + \hat{A}^*_{1nG})}{\frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,0})^2} I(\cdot)]}{\sqrt{\hat{V}^*_{TG}}}$, for

$\hat{R}^{*2}_G = (1 - \frac{\frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,b})^2}{\frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,0})^2})I(\cdot)$, $\hat{A}^*_{nG} = \frac{1}{n^3h_n^{2d}}\sum_{t=1}^n\sum_{\substack{i=1 \\ t\neq i}}^n \frac{K^2_{it}}{\hat{f}^2(X_t)}(\epsilon^*_{i,0})^2$, $\hat{A}^*_{1nG} = -\frac{2}{n^2h_n^d}K(0)\sum_{t=1}^n \frac{(\epsilon^*_{t,0})^2}{\hat{f}(X_t)}$,

$\hat{V}^*_{TG} = \frac{\hat{\sigma}^{2*}_{\phi G}}{(\frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,0})^2)^2}$, $\hat{\sigma}^{2*}_{\phi G} = \frac{1}{n^2}\sum_{t=1}^n\sum_{\substack{i=1 \\ t\neq i}}^n K_{it}\frac{(\epsilon^*_{i,0})^2(\epsilon^*_{t,0})^2}{h_n^d\hat{f}(X_t)}\int 2(2K(\psi) - \kappa(\psi))^2 d\psi$, $\epsilon^*_{t,b} = y^*_t - \hat{m}^*(X_t)$ and

$I(\cdot) = I(\frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,0})^2 \geq \frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,b})^2)$.

3 Asymptotic properties

We characterize the asymptotic behavior of the test statistics when $d = 1$ with the following assumptions:

A1. $\{x_t, y_t\}^n_{t=1}$ is independently and identically distributed (IID). A2. $0 < V(y) < \infty$.

A3. $E(\epsilon|x) = 0$, $V(\epsilon|x) = \sigma^2(x)$, $\sigma^2(x)$ is continuous at x and $E\sigma^2(x) < \infty$.

A4. Define the marginal density of x by $f(x)$, we have (1) $0 < \underline{B}_f \leq f(x) \leq \bar{B}_f < \infty$ for all $x \in G$, G compact subset of $\Re$. (2) $\forall\, x, x' \in G$, $|f(x) - f(x')| < m_f|x - x'|$ for some $0 < m_f < \infty$. (3) $f(x)$ is uniformly continuous at x, $\forall\, x \in G$.

A5. $0 < \underline{B}_m \leq m(x) \leq \bar{B}_m < \infty$ for all $x \in G$, where $m(x) : \Re \to \Re$ is a measurable twice continuously differentiable function in $\Re$, $|m^{(2)}(x)| < \bar{B}_{2m} < \infty$ for all $x \in G$.

A6. As $n \to \infty$, $nh_n^2 \to \infty$, $nh_n^6 \to 0$.

A7. $K(.) : S \to \Re$ is symmetric density function with compact support $S \subset \Re$ s.t.
(1) $\int xK(x)dx = 0$. (2) $\int x^2K(x)dx = \sigma^2_K < \infty$.
(3) $\forall x \in \Re$, $|K(x)| < B_k < \infty$. (4) $|u^jK(u) - v^jK(v)| \leq c_k|u - v|$, for $j = 0, 1, 2, 3$.

A8. For some $\delta > 0$, $E(|\epsilon|^{2+\delta}|x) < \infty$, $f_{x|\epsilon}(x) < \infty$, $f(x, \epsilon)$ is continuous around x.

A9. (1) $E\sigma^4(x) < \infty$. (2) $E(\epsilon^4_i|x) < \infty$. A10. $ED^4(x) < \infty$, $D(x)$ is a continuous function of x.

We assume the conditional variance $\sigma^2(x)$ to be continuous at x in A3, and $f(x)$ and $m(x)$ to be smooth and bounded in A4 and A5. They enable the use of Taylor expansion. A6 places restriction on the choice of bandwidth, and they are no more restrictive than that used in a nonparametric regression. Specifically, an optimal bandwidth in the kernel regression of order $O(n^{-\frac{1}{5}})$ can be used. A7 requires a bounded symmetric kernel function that satisfies Lipschitz condition. Thus, the popular Epanechnikov kernel can be used. These are commonly used in nonparametric kernel regression (Martins-Filho and Yao (2007)). A8 places additional conditional moments assumption on ϵ, which enables us to obtain the distribution of $\hat{R}^2$ in Theorem 3 with central limit theorem. The null distribution of the test statistics is obtained in Theorem 1 with additional moment assumption in A9. To derive the local power, we need the function $D(x)$ in the local alternative to have fourth moment and to be smooth in A10, which facilitates deriving the asymptotic distribution of tests under local alternatives in Theorem 2.

Theorem 1 *Under H_0 and assumptions A1-A4, A6-A9 we have (a) $T_n \xrightarrow{d} N(0,1)$. (b) $\hat{T}_n \xrightarrow{d} N(0,1)$.*

It shows that asymptotically the unknown items could be replaced with the estimates and $\hat{T}_n$ behaves similarly to T_n. It provides basis for us to conduct hypothesis tests. For example, we can compare $\hat{T}_n$ with the one sided critical value $z_{1-\alpha}$, i.e., the $(1-\alpha)th$ percentile from the standard normal distribution. We reject the null when $\hat{T}_n > z_{1-\alpha}$ at the α significance level.

Next, we examine the asymptotic local power of the test. Define the sequence of Pitman local alternatives as $H_1(l_n) : m(x_t) = \mu + l_n D(x_t)$, where $l_n \to 0$ as $n \to \infty$. $D(x_t)$ is a non-constant continuous function, indicating the deviation of $m(x_t)$ from the constant.

Theorem 2 *Under $H_1(\frac{1}{\sqrt{nh_n^{\frac{1}{2}}}})$ and assumptions A1-A4, A6-A10, we have*

1. $T_n \xrightarrow{d} N(\frac{V(D(x))}{\sqrt{\sigma_\phi^2}}, 1)$. 2. $\hat{T}_n \xrightarrow{d} N(\frac{V(D(x))}{\sqrt{\sigma_\phi^2}}, 1)$.

From Theorem 2, we note that the local power of the test $\hat{T}_n$ satisfies $P(\hat{T}_n \geq z_{1-\alpha} | H_1(\frac{1}{\sqrt{nh_n^{\frac{1}{2}}}})) \to 1 - \Phi(z_{1-\alpha} - \frac{V(D(x))}{\sqrt{\sigma_\phi^2}})$ as $n \to \infty$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. It implies the test has non-trivial asymptotic power against local alternatives that deviate from the null at the rate of $(nh^{\frac{1}{2}})^{-\frac{1}{2}}$. The power increases with the magnitude of $\frac{V(D(x))}{\sqrt{\sigma_\phi^2}}$. By taking a large bandwidth we can make the magnitude of the alternative (of order l_n) against which the test has non-trivial power arbitrarily close to the parametric rate of $n^{-\frac{1}{2}}$. Hong and Lee (2009) and Fan et al. (2001) show that when a local linear smoother is used to estimate $m(\cdot)$ under $H_1(l_n)$ and the bandwidth is of order $n^{-2/9}$, the GLR test can detect local alternatives with rate $l_n = O(n^{-4/9})$, which is optimal according to Lepski and Spokoiny (1999). By Theorem 2, with $h_n = O(n^{-2/9})$, we note $l_n = O(n^{-4/9})$, thus the test $\hat{T}_n$ achieves the optimal convergence rate as well, and it is a powerful nonparametric test procedure. The choice of h_n is consistent with what we assume in A6.

Under fixed alternative H_1 that $m(x_t) \neq \mu$, we obtain the asymptotic normal distribution for $\hat{R}^2$.

Theorem 3 *Under the alternative H_1 and assumptions A1-A8, $\sqrt{n}(\hat{R}^2 - R^2) \xrightarrow{d} N(0, \frac{E(W_t^2)}{V(y)^2})$, where $W_t = \epsilon_t^2 - \frac{E\sigma^2(x_t)}{V(y_t)}(y_t - E(y_t))^2$.*

Note $E(W_t^2)$ is a global measure by the IID assumption A1 and the bias of $\hat{R}^2$ vanishes asymptotically with assumption A6. The result complements Doksum and Samarov (1995) by providing the asymptotic distribution of $\hat{R}^2$ constructed from the local linear estimator, and allows the construction of confidence interval for R^2, which measures the fit of the model. It provide useful information about the type II error of the test at any particular point of the alternative, if the test accepts the null hypothesis. This is particularly important for the application of a goodness-of-fit test, since the acceptance of the null will lead to a subsequent data analysis adapted towards the model under H_0, so it is desirable to estimate the corresponding probability of an error of this procedure at any particular point in the alternative. For example, at significance level α, we reject H_0 if $\hat{T}_n > Z_{1-\alpha}$, or fail to reject H_0 if $\hat{T}_n \leq Z_{1-\alpha}$ for the test statistic $\hat{T}_n$ defined in equation (3). So we fail to reject H_0 when

$$\hat{R}^2 - I(\frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{m}(x_t))^2)\frac{\hat{A}_{1n}+\hat{A}_n}{\frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2} \leq \frac{Z_{1-\alpha}\sqrt{\hat{V}_T}}{nh_n^{\frac{1}{2}}}.$$

Given the result in Theorem 3, for a particular point in $H_1 : E(y_t|x_t) = m(x_t)$ and $R^2 \neq 0$. So the probability of type II error is approximated with $\Phi(\frac{\sqrt{n}}{\sqrt{\frac{E(W_t^2)}{V(y)^2}}}(\frac{Z_{1-\alpha}\sqrt{\hat{V}_T}}{nh_n^{\frac{1}{2}}} - R^2))$, since $\hat{A}_{1n}$ and $\hat{A}_n$ are of order $O_p(\frac{1}{nh_n})$ each. It also helps to establish the global consistency of the test $\hat{T}_n$ in Theorem 4 below.

Theorem 4 *Under H_1, and assumptions A1-A9, we have $P(\hat{T}_n > c_n) \to 1$, for any positive constant $c_n = o(nh_n^{\frac{1}{2}})$. Thus the $\hat{T}_n$ test is consistent.*

Theorem 5 *Assume assumptions A1-A9, we have $\hat{T}_n^* \xrightarrow{d} N(0,1)$ conditionally on $W \equiv \{x_t, y_t\}_{t=1}^n$.*

It indicates the bootstrap provides an asymptotic valid approximation to the null limit distribution of $\hat{T}_n$. Theorem 5 holds regardless of whether H_0 is true. When H_0 is true, the bootstrap procedure will lead asymptotically to correct size of the test, since $\hat{T}_n$ converges in distribution to the same $N(0,1)$ limiting distribution under H_0 as in Theorem 1. When H_0 is false, $\hat{T}_n$ will converge to infinity as shown in the proof of Theorem 4, but asymptotically the bootstrap critical value is still finite for any significance level α different from 0. Thus $P(\hat{T}_n > \hat{T}_n^*) \to 1$ and bootstrap methods is consistent.

Remark 1: We state the asymptotic properties of the alternative tests, where the proof is sketched in Yao and Ullah (2013). (1) With conditions 1-7 in Doksum and Samarov (1995), R1-R4 in Yao and Ullah (2013) and H_0, $\hat{T}_{in} \xrightarrow{d} N(0,1)$ for $i = 1, 2$. (2) With conditions 1-7, R1-R5 and H_1, $P(\hat{T}_{in} > c_n) \to 1$ for any positive constant $c_n = o(nh_n^{1/2})$. Thus, the $\hat{T}_{in}$ tests are consistent for $i = 1, 2$. (3) With conditions 1-7, R1-R5, $\hat{T}_{in}^* \xrightarrow{d} N(0,1)$ conditionally on $W = \{x_t, y_t\}_{t=1}^n$ for $i = 1, 2$.

Now we provide the properties for tests in the more general case ($d > 1$). For a generic function $g(x_{1t})$, we define $g(\cdot) \in C_1^{v_1}$ if $g(x_{1t})$ is $v_1 - 1$ times continuously differentiable, with its $(v_1-1)th$ order derivative uniformly continuous on G_1, and $\sup_{x_{1t}\in G_1} |\frac{\partial^j}{\partial x_{1t}^j} g(x_{1t})| < \infty \ \forall j = 1, \cdots, v_1 - 1$. Here G_1 is a compact subset of $\Re^{d_1}$. Similarly, we denote a generic function $g(X_t) \in C^v$ if $g(X_t)$ is $v-1$ times continuously differentiable, with its $(v-1)th$ order derivative uniformly continuous on G, a compact subset of $\Re^d$ and $\sup_{X_t\in G} |\frac{\partial^j}{\partial X_t^j} g(X_t)| < \infty \ \forall j = 1, \cdots, v-1$. We introduce the following additional assumptions.

- B0. (1) $\{X_t, y_t\}_{t=1}^n$ is IID. $X_t \in \Re^d$ for $d < 8$ and $y_t \in \Re$. (2) $0 < E(y_t - E(y_t|x_{1t}))^2 < \infty$. (3) For $\epsilon_t = y_t - m(X_t)$, $E(\epsilon_t|X_t) = 0$, $V(\epsilon_t|X_t) = \sigma^2(X_t)$, $0 < E(\epsilon_t^2) < \infty$, and $\sigma^2(X_t) \in C^v$.
- B1. (1) Define the marginal density of x_{1t} as $f_1(x_{1t})$, and $0 < \underline{B}_{f_1} \leq f_1(x_{1t}) \leq \bar{B}_{f_1} < \infty$ for all $x_{1t} \in G_1$. $\forall\, x_{1t}, x_{1\tau} \in G_1$, $|f_1(x_{1t}) - f(x_{1\tau})| < m_{f_1}||x_{1t} - x_{1\tau}||$ for some $0 < m_{f_1} < \infty$. $f_1(x_{1t}) \in C_1^{v_1}$.

  (2)Define the marginal density of X_t as $f(X_t)$, and $0 < \underline{B}_f \leq f(X_t) \leq \bar{B}_f < \infty$ for all $X_t \in G$. $\forall\, X_t, X_\tau \in G$, $|f(X_t) - f(X_\tau)| < m_f||X_t - X_\tau||$ for some $0 < m_f < \infty$. $f(X_t) \in C^v$.
- B2. $0 < \underline{B}_r \leq r(x_{1t}) \leq \bar{B}_r < \infty$ for all $x_{1t} \in G_1$, where $r(x_{1t}) : \Re^{d_1} \to \Re$, and $r(x_{1t}) \in C_1^{max\{v,v_1\}}$.
- B3. As $n \to \infty$, (1) $nh_{1n}^{2d_1} \to \infty$, and $nh_n^{\frac{d}{2}} h_{1n}^{2v_1} \to 0$. (2) $\frac{h_n^d}{h_{1n}^{2d_1}} \to 0$. (3) $nh_n^{2d} \to \infty$, and $nh_n^{\frac{d}{2}} h_n^{2v} \to 0$.
- B4. (1) $K_1(.) : S_1 \to \Re$ is kernel function of order v_1, with compact support $S_1 \subset \Re^{d_1}$ such that $|K_1(x)| < B_{k_1} < \infty$. $|u^j K_1(u) - v^j K_1(v)| \leq c_{k_1}||u - v||$, for $j = 0, 1, 2, 3$.

(2) $K(.) : S \to \Re$ is kernel function of order v, with compact support $S \subset \Re^d$ such that $|K(x)| < B_k < \infty$. $|u^j K(u) - v^j K(v)| \leq c_k ||u - v||$, for $j = 0, 1, 2, 3$.

B5. $0 < \underline{B}_m \leq m(X_t) \leq \bar{B}_m < \infty$ for all $X_t \in G$, where $m(X_t) : \Re^d \to \Re$, and $m(X_t)) \in C^v$.

B6. $E(\epsilon_t^4 | X_t) < \infty$, $f_{X|\epsilon}(X) < \infty$, $f(X, \epsilon)$ is continuous around X.

B7. $ED^4(X) < \infty$ and $D(X)$ is a continuous function of X.

B0, B6 and B7 are the multivariate version of assumptions A1-A3 and A8-A10, except that we restrict attention to the case $d < 8$ and we assume $\sigma^2(\cdot)$ to be a smooth function. The conditions in B1, B2 and B5 assume the smoothness and boundedness for the densities of x_{1t}, X_t and the conditional mean functions $r(x_{1t}), m(X_t)$. They allow us to perform Taylor expansions on $r(x_{1t})$ and $m(X_t)$, which shows up in the generalized test statistic. Assumptions B3 and B4 place further requirements on the bandwidth choices and kernel functions, allowing us to use a higher order kernel to control the order of the bias or variance terms introduced in the estimation. Specifically, $B3(1)$ and (2) control the bias and variance terms when estimating $r(x_{1t})$, while $B3(3)$ controls the bias term when estimating $m(X_t)$. If $d > 2d_1$, h_{1n} and h_n can be of the same order, and $v = v_1$ can be used for simplicity. The assumptions are similar in spirit to those imposed in Lavergne and Vuong (2000). As indicated in the asymptotic properties below, the estimation of $r(\cdot)$ does not have impacts on the asymptotic distribution.

Theorem 6 provides the theoretical results for the generalized test statistics. Specifically, (I) gives the asymptotic null distribution of $\hat{T}_{nG}$, (II) characterizes its asymptotic local power, and (III) shows the global consistency of $\hat{T}_{nG}$. The use of the bootstrap test $\hat{T}^*_{nG}$ is justified with result (IV).

Theorem 6

(I) Assuming B0-B4 and B6, under H_{0G}, we have $\hat{T}_{nG} \xrightarrow{d} N(0,1)$.

(II) Assuming B0-B4, B6 and B7, under the Pitman local alternative $H_{1G}(l_n) : m(X_t) = r(x_{1t}) + l_n D(X_t)$, where $l_n = n^{-1/2} h_n^{-d/4}$, we have $\hat{T}_{nG} \xrightarrow{d} N((\sigma^2_{\phi G})^{-1/2} E[D(X_t) - E(D(x_{1t}, x_{2j})|x_{1t})]^2, 1)$, where $\sigma^2_{\phi G} = E\frac{\sigma^4(X_t)}{f(X_t)} 2\int (2K(\psi) - \kappa(\psi))^2 d\psi$.

(III) Assuming B0-B6, under H_{1G} that $E(y_t|X_t) = m(X_t)$, we have $P(\hat{T}_{nG} > c_n) \to 1$ for any positive constant $c_n = o(nh_n^{d/2})$. Thus the $\hat{T}_{nG}$ test is consistent.

(IV) Assuming B0-B6, we have $\hat{T}^*_{nG} \xrightarrow{d} N(0,1)$ conditionally on $W = \{X_t, y_t\}_{t=1}^n$.

Remark 2: With homoskedasticity and bounded support for X, the bootstrap test $\hat{T}^*_{nG}$ can be simply implemented by (i) In step 1, generate ϵ^*_t as the bootstrap error. For example, we resample with replacement from centered $\{\hat{\epsilon}_t\}_{t=1}^n$ to obtain $\{\epsilon^*_t\}_{t=1}^n$. (ii) Follow steps 2-3, using ϵ^*_t defined in (i). (iii) Calculate only $\hat{R}^{*2}_G$ in step 4. (iv) Step 5: repeat steps 1-4 B times, then use the original $\hat{R}^2_G$ and the B bootstrap test statistics $\hat{R}^{*2}_G$ to obtain the empirical distribution of the bootstrap statistics. The finite sample null distribution of $\hat{T}^*_{nG}$ is the same as the empirical distribution in step 5 above. This is due to the fact the other items in $\hat{T}^*_{nG}$ are independent of the DGP characteristics under the null and homoskedasticity, so we do not need to use the Wild bootstrap to preserve the heteroskedasticity structure.

4 Monte Carlo Study

We provide a Monte Carlo study to implement our proposed test statistics and illustrate their finite sample performances relative to several popular nonparametric significance test statistics. We follow Gu et al. (2007) and Lavergne and Vuong (2000) to consider the following data-generating processes:

$$\begin{array}{l} DGP_0 : y_t = 1 + \beta_0 z_t + \beta_1 z_t^3 + u_t, \\ DGP_1 : y_t = 1 + \beta_0 z_t + \beta_1 z_t^3 + \gamma_1 x_t + u_t, \text{ and} \\ DGP_2 : y_t = 1 + \beta_0 z_t + \beta_1 z_t^3 + \gamma_2 sin(2\pi x_t) + u_t. \end{array} \tag{6}$$

z_t and x_t are IID uniform on $[-1, 1]$ and u_t is from a normal $N(0, \sigma^2(x_t))$. DGP_0 corresponds to the null, where x_t is irrelevant. So we investigate the size of tests with DGP_0. We follow Fan and Li (2000) to call DGP_1, DGP_2 a low and a high frequency alternative, respectively, under which we can compare the power of tests. DGP_2 is used in Fan and Li (2000) to demonstrate that a smoothing test can be more powerful than a non-smoothing test against high-frequency alternatives. Since $E(x_t) = E(sin(2\pi x_t)) = 0$, $E(y_t|z_t)$ remains the same across all data generating processes. $\{\gamma_1, \gamma_2\}$ are set to be $\{0.5, 1\}$.

We consider two data generating processes described by (6). The simple regression model is denoted by (S), where we set $\{\beta_0, \beta_1\} = \{0, 0\}$, and $\sigma^2(x) = x^2$. The conditional heteroskedasticity is present and z_t is omitted in (S), and the null and alternative hypothesis correspond to H_0 and H_1, respectively. The multiple regression model is denoted by (M), where $\{\beta_0, \beta_1\} = \{-1, 1\}$, and $\sigma^2(x) = 1$. So z_t is present with homoskedasticity, and the null and alternative are H_{0G} and H_{1G}, respectively.

The implementation of our test statistics requires the choice of bandwidths h_n and h_{1n}. To make a fair comparison, we choose the *same* bandwidth sequence for all tests. Under (S), we select $\hat{h}_n$ with $cR(x_t)n^{-1/3}$, where $R(\cdot)$ is the interquartile range. Under (M), $\hat{h}_{1n}$ is selected as $R(z_t)n^{-1/4+2\delta}$, $\hat{h}_n$ for z_t is $R(z_t)n^{-1/4+\delta}$, and $\hat{h}_n$ for x_t is $cR(x_t)n^{-1/4+\delta}$, where $\delta = 0.01$ is utilized to satisfy the assumption B3. We consider the constant c to be 0.5, 1 and 2 to investigate the sensitivity of results to the smoothing parameter's choice, where we follow Lavergne and Vuong (2000) to use similar bandwidth for the regressor z_t that is common to both the null and alternative. We utilize the Epanechnikov kernel with support $[-\sqrt{5}, \sqrt{5}]$, i.e., $K(u) = \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}u^2)I(|u| \leq \sqrt{5})$ in (S) and the product of the Epanechnikov kernel in (M). The above choices of bandwidth and kernel function satisfy our assumptions A6, A7, B3 and B4.

Under (S), we consider our tests $\hat{T}_n$, $\hat{T}_n^*$, the four alternative tests $\hat{T}_{1n}$, $\hat{T}_{1n}^*$, $\hat{T}_{2n}$, $\hat{T}_{2n}^*$ proposed in section 2, λ_n, $\hat{J}_n$ and $\hat{J}_n^*$. We use the weight function $w(x) = I(\hat{f}(x) \geq 0.01)$ as in Doksum and Samarov (1995) for $\hat{T}_{1n}$, $\hat{T}_{1n}^*$, $\hat{T}_{2n}$ and $\hat{T}_{2n}^*$. λ_n is the GLR test by Fan et al. (2001), motivated with normal error term and constructed as $\lambda_n = \frac{n}{2} ln \frac{RSS_0}{RSS_1} (\approx \frac{n}{2}(\frac{RSS_0}{RSS_1} - 1)$ under the H_0), where $RSS_0 = \sum_{t=1}^{n}(y_t - \bar{y})^2$, and $RSS_1 = \sum_{t=1}^{n}(y_t - \hat{m}(x_t))^2$, also see Ullah (1985). It is somewhat related to our test as they are constructed with the sum of squared residuals (RSS) from H_0 and H_1. One can show in testing overall significance and with homoskedasticity, they have the same asymptotic distribution, but they are always different numerically. Our test $\hat{T}_n$ can be constructed directly without simulations, while the GLR test is generally implemented with simulations. $\hat{J}_n$ and $\hat{J}_n^*$ are based on equations (5) and (9) in Gu et al. (2007), where we modify them so that their first stage estimation can simply be replaced by a sample mean under H_0.

Under (M), we include our test $\hat{T}_{nG}$ and $\hat{T}^*_{nG}$ implemented as in Remark 2,[3] and five alternatives $\hat{J}_n$, $\hat{J}^*_n$, $\hat{J}_{w,n}$, $\hat{J}^*_{w,n}$ and $\hat{T}_{lv}$. The alternative tests are based on H_{0G} such that $E(y_t|z_t, x_t) = E(y_t|z_t)$. Thus let $v_t = y_t - E(y_t|z_t)$, H_{0G} implies $E(v_t|z_t, x_t) = 0$. Note $J = E([E(v_t|z_t, x_t)]^2 f(z_t, x_t)) \geq 0$, and $J = 0$ if and only if H_{0G} is true, where $f(z_t, x_t)$ is the joint density of z_t and x_t. $\hat{J}_n$ and $\hat{J}^*_n$ are considered in Fan and Li (1996), Zheng (1996), and Gu et al. (2007). $\hat{J}_n$ in equation (5) of Gu et al. (2007) is the sample analog of J and replaces the unknown $E(v_t|z_t, x_t)$ and $f(z_t, x_t)$ with the leave-one-out kernel estimates. The bootstrap version $\hat{J}^*_n$ is provided in their equation (9). The density-weighted test statistic based on J has the advantage that the density function does not have to be bounded away from zero. The sample analog version is $\hat{J}_{w,n}$ in their equation (6) and the bootstrap version is $\hat{J}^*_{w,n}$ in equation (12). We follow their simulation to choose the product standard normal kernel, the *rule-of-thumb* bandwidth sequences and multiply the bandwidths for smoothing z_t and x_t by the constant c indicated above to examine the sensitivity of test results. $\hat{T}_{lv}$ by Lavergne and Vuong (2000) page 578 is also based on the term J above and it substantially reduces the bias of the test. We use their equation (2.2) as their asymptotic variance estimator, which is computationally less demanding. We follow their suggestion to choose the product Epanechnikov kernel, the *rule-of-thumb* bandwidth sequences and multiply the bandwidth for x_t the constant c to investigate the sensitivity of test result to the bandwidth's choice.

We consider two sample sizes, 100 with 1000 repetitions, and 200 with 500 repetitions. For all the bootstrap test statistics, the bootstrap repetition times B is fixed to be 399. We summarize the experiment results in terms of empirical levels of rejections for each test statistics at the significance level $\alpha = 0.05$ in Table 1 in Appendix 1. The top two panels are for (S) and the bottom two for (M). The results for DGP_0 correspond to the size of tests, since the null hypothesis is maintained. We provide evidences about the power of tests in DGP_1 with the low frequency alternative and in DGP_2 with high frequency alternative.

When the sample size increases from 100 to 200, there is weak evidence that the size of each test improve towards the designated level, especially under (S), but the power of each test increases significantly in DGP_1 and DGP_2. The observation confirms our results in Theorems 1, 4, 5 and 6 that $\hat{T}_n$, $\hat{T}^*_n$, $\hat{T}_{nG}$ and $\hat{T}^*_{nG}$ are consistent. It is consistent with the results in Gu et al. (2007), Lavergne and Vuong (2000) and Fan et al. (2001) that the other test statistics considered are consistent as well. By examining results for DGP_1 and DGP_2, we find the expected result that it is harder to conduct test in the multiple regression context (M) than the simple regression (S), as the power of test statistics in (M) is smaller than that in (S). The performance of the tests is indeed sensitive to the choice of c in the bandwidth, though the impact seems to be in a nonlinear fashion and differ across different test statistics for the size under DGP_0, consistent with that in above mentioned papers. Being oversized in general, the performance of $\hat{T}_{nG}$ seems to be relatively more sensitive to c in (M), while that of $\hat{T}^*_{nG}$ is fairly robust. There are weak evidence that the power of each test increases with c under GDP_1, but decreases with c under GDP_2. It

[3] Since x and z are independent, then H_{0G} implies $E(y|x) = E[E(y|x,z)|x] = E[E(y|z)|x] = E(y) = \mu$, H_0 and H_1 in section 2.1 can be used, thus $\hat{T}_n$ and $\hat{T}^*_n$ are valid test statistics. In this case, we only need to select the bandwidth parameter for x and perform a single nonparametric regression of y on x to conduct the tests. We use $\hat{T}_{nG}$ and $\hat{T}^*_{nG}$ to provide a fair comparison since all the alternative tests involve regressions with multiple regressors.

is easy for all to reject the null under the high frequency alternative in DGP_2 relative to DGP_1.

To facilitate the comparison in terms of the size, we insert a (1) or (2) on the test's upper right corner to indicate that it is the closest or the second closest to the target significance level. In (S), the best performing tests are $\hat{T}^*_{2n}$, followed by $\hat{T}^*_n$, by $\hat{T}_n$, λ_n, or $\hat{T}^*_{1n}$, then by $\hat{T}_{1n}$. $\hat{J}_n$ and $\hat{T}_{2n}$ are fairly undersized. $\hat{J}^*_n$ improves over $\hat{J}_n$, but not significantly. In (M), our bootstrap test $\hat{T}^*_{nG}$ clearly outperforms the others in terms of being closest to the desired target size. The next best is $\hat{T}_{lv}$, followed by $\hat{J}^*_{w,n}$, and by $\hat{J}^*_n$. The asymptotic tests $\hat{J}_n$ and $\hat{J}_{wn}$ are fairly undersized, while $\hat{T}_{nG}$ is oversized. The observation here is consistent with Gu et al. (2007), which show that in finite sample $\hat{J}^*_n$ and $\hat{J}^*_{w,n}$ substantially improve upon $\hat{J}_n$ and $\hat{J}_{w,n}$. It is also consistent with Lavergne and Vuong (2000) that $\hat{T}_{lv}$ significantly improves the performance over $\hat{J}_n$ and $\hat{J}_{w,n}$. The results indicate that $\hat{T}_{lv}$ competes well with the bootstrap tests $\hat{J}^*_n$ and $\hat{J}^*_{w,n}$. Overall, our proposed bootstrap test statistic $\hat{T}^*_{nG}$ captures the desired target size well compared with $\hat{J}^*_n$, $\hat{J}^*_{w,n}$ and $\hat{T}_{lv}$ in (M), while $\hat{T}^*_n$ and $\hat{T}_n$ are fairly satisfactory when compared with other alternatives in the (S), and the bootstrap tests $\hat{T}^*_{2n}$, $\hat{T}^*_{1n}$ or λ_n are valuable competitors.

Now we compare the power with a low frequency alternative in DGP_1. In (S), all tests exhibit power close to one and the difference is relatively small, with $\hat{J}_n$ showing slightly lower power that others in small sample. In (M), the best tests are frequently $\hat{J}^*_n$, $\hat{J}^*_w$, or $\hat{T}_{nG}$, followed closely by $\hat{T}^*_{nG}$. Their powers are much larger than those of $\hat{T}_{lv}$, $\hat{J}_n$ and $\hat{J}_{w,n}$. With the high frequency alternative in DGP_2, the power of all tests are much closer to one, with exceptions on $\hat{J}_n$ and $\hat{J}_{w,n}$ in (M). In (S), $\hat{T}_{1n}, \hat{T}^*_{1n}$, $\hat{T}_{2n}$ and $\hat{T}^*_{2n}$ seem to be influenced more by the bandwidth. In (M), the relative performance of the tests are similar to what we observe in DGP_1, where the only exception occurs when $c = 2$ and $\hat{T}_{lv}$ performs better.

Based on above observation, we conclude that our proposed bootstrap test statistics $\hat{T}^*_n$ and $\hat{T}^*_{nG}$ perform well in the finite sample study. Their sizes under the null hypothesis are close to the target level. $\hat{T}^*_n$, together with λ_n, $\hat{T}^*_{2n}$ and $\hat{T}^*_{1n}$, exhibit reasonable power in (S), while $\hat{T}^*_{nG}$ and $\hat{T}_{nG}$, together with $\hat{J}^*_n$ and $\hat{J}^*_{w,n}$ demonstrate much larger empirical power than the rest in (M). We found that $\hat{T}_{nG}$'s size performance is relatively sensitive to the choice of bandwidth, so we recommend the bootstrap tests $\hat{T}^*_n$ and $\hat{T}^*_{nG}$ rather than the asymptotic tests. The newly proposed tests $\hat{T}^*_{1n}$ and $\hat{T}^*_{2n}$ show good size and power performances in simple regression, which might deserve further investigation in the general set-up.

5 Conclusion

We propose nonparametric R^2 based tests for the presence of relevant variables in a regression model. Under the null hypothesis that the variables are irrelevant, we establish their asymptotic normality at rate $nh_n^{\frac{d}{2}}$. Our test is consistent against all alternatives and detects local alternatives that deviate from the null at rate $(nh_n^{\frac{d}{2}})^{-\frac{1}{2}}$. We further propose the Wild bootstrap/bootstrap test to approximate the null distribution. The asymptotic normality of the nonparametric R^2 estimator at rate $\sqrt{n}$ is also established under the alternative hypothesis, which facilitate inference with the nonparametric R^2 estimator. We illustrate their finite sample performance in a Monte Carlo study. The bootstrap tests capture the size well, exhibit reasonable power, and provide viable alternatives that complement other tests available.

Appendix 1: Table

Table 1: Empirical levels of rejections with $\alpha = 5\%$,((S), $\sigma^2(x) = x^2$) and ((M), $\sigma^2 = 1$). Size of test statistics for DGP_0(Null). Power of test statistics for DGP_1 (low frequency alternative) and DGP_2 (high frequency alternative).

(S) $n = 100$	DGP_0			DGP_1			DGP_2		
$\sigma^2(x) = x^2$	$c = 0.5$	1	2	0.5	1	2	0.5	1	2
$\hat{T}_n$	.065	.070	$.047^{(1)}$	.957	.982	.994	1	1	.913
$\hat{T}^*_n$	.072	.061	$.054^{(2)}$	.953	.980	.993	1	1	.903
$\hat{J}_n$	.019	.014	.006	.919	.964	.971	1	1	.920
$\hat{J}^*_n$	.042	.019	.005	.952	.966	.965	1	1	.887
λ_n	.062	.060	.039	.951	.981	.993	1	1	.899
$\hat{T}_{1n}$	$.056^{(2)}$	.039	.011	.954	.980	.989	1	1	.300
$\hat{T}^*_{1n}$	.062	$.059^{(2)}$	.035	.956	.987	.997	1	1	.410
$\hat{T}_{2n}$	.034	.020	.005	.970	.980	.960	1	1	.175
$\hat{T}^*_{2n}$	$.054^{(1)}$	$.049^{(1)}$	.032	.982	.991	.994	1	1	.527
(S) $n = 200$	DGP_0			DGP_1			DGP_2		
$\sigma^2(x) = x^2$	$c = 0.5$	1	2	0.5	1	2	0.5	1	2
$\hat{T}_n$	.062	$.058^{(2)}$	.082	1	.998	1	1	1	1
$\hat{T}^*_n$	$.048^{(2)}$	$.048^{(1)}$	.076	1	.998	1	1	1	1
$\hat{J}_n$	.022	.020	.020	1	.998	1	1	1	1
$\hat{J}^*_n$	.030	.022	.020	1	.998	1	1	1	1
λ_n	$.052^{(2)}$	$.048^{(1)}$	$.072^{(2)}$	1	.998	1	1	1	1
$\hat{T}_{1n}$	.042	.030	$.028^{(2)}$	1	.998	1	1	1	.980
$\hat{T}^*_{1n}$	.054	$.052^{(1)}$	.074	1	.998	1	1	1	.988
$\hat{T}_{2n}$	.028	.024	.016	1	.998	1	1	1	.896
$\hat{T}^*_{2n}$	$.050^{(1)}$	$.048^{(1)}$	$.064^{(1)}$	1	1	1	1	1	.996

(M) $n = 100$	DGP_0			DGP_1			DGP_2		
$\sigma^2 = 1$	$c = 0.5$	1	2	0.5	1	2	0.5	1	2
$\hat{T}_{nG}$	.027	.120	.077	.193	.623	.696	.910	.820	.409
$\hat{T}^*_{nG}$	.073	$.053^{(1)}$	$.051^{(1)}$	.383	.489	.606	.986	.698	.321
$\hat{T}_{lv}$	$.035^{(1)}$	$.040^{(2)}$	.025	.203	.308	.319	.935	.945	.337
$\hat{J}_n$	.002	.008	.000	.202	.278	.133	.955	.425	.018
$\hat{J}^*_n$	$.068^{(2)}$	.070	$.060^{(2)}$	.540	.628	.696	.996	.899	.361
$\hat{J}_{w,n}$	.006	.008	.000	.207	.276	.146	.948	.436	.023
$\hat{J}^*_{w,n}$	$.065^{(1)}$	.063	.064	.519	.588	.678	.996	.887	.342
(M) $n = 200$	DGP_0			DGP_1			DGP_2		
$\sigma^2 = 1$	$c = 0.5$	1	2	0.5	1	2	0.5	1	2
$\hat{T}_{nG}$	.134	.144	.098	.760	.906	.908	1	1	.780
$\hat{T}^*_{nG}$	$.050^{(1)}$	$.050^{(1)}$	$.042^{(1)}$	.652	.820	.860	1	.998	.660
$\hat{T}_{lv}$	$.036^{(2)}$	$.028^{(2)}$	$.024^{(2)}$	.420	.566	.622	1	1	.982
$\hat{J}_n$	.006	.002	.002	.568	.688	.544	1	.996	.140
$\hat{J}^*_n$	.072	$.072^{(2)}$	.082	.774	.890	.914	1	1	.746
$\hat{J}_{w,n}$	.004	.004	.002	.560	.650	.544	1	.996	.146
$\hat{J}^*_{w,n}$	$.064^{(2)}$	$.072^{(2)}$	.080	.768	.880	.900	1	1	.724

Appendix 2

Below we outline the proof of Theorem 6 only. The proof of Theorems 1-5 and remarks 1 and 2 are provided in a separate Appendix (Yao and Ullah (2013)), which is attached.

Theorem 6: *Proof.*

(I) We observe that under H_{0G}, $y_t = r(x_{1t}) + \epsilon_t$ and $E(\epsilon_t|x_{1t}) = 0$, thus
$\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - \frac{2}{n}\sum_{t=1}^{n}(\hat{m}(X_t) - r(x_{1t}))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(\hat{m}(X_t) - r(x_{1t}))^2$.
Since $\hat{m}(X_t)$ is the local linear estimator, for 0_d a $d \times 1$ vector of zeros, we can write
$\hat{m}(X_t) - r(x_{1t}) = \frac{1}{nh_n^d}\sum_{i=1}^{n}(1, 0_d')S_n^{-1}(X_t)(1, (\frac{X_i - X_t}{h_n}))'K(\frac{X_i - X_t}{h_n})y_{i*}$, $S_n(X_t) = \begin{pmatrix} s_{0n}(X_t) & s_{1n}(X_t) \\ s_{1n}'(X_t) & s_{2n}(X_t) \end{pmatrix}$,
$s_{jn}(X_t) = \frac{1}{nh_n^d}\sum_{i=1}^{n}K(\frac{X_i - X_t}{h_n})(\frac{X_i - X_t}{h_n})^j$ for $j = 0, 1$, $s_{2n}(X_t) = \frac{1}{nh_n^d}\sum_{i=1}^{n}K(\frac{X_i - X_t}{h_n})(\frac{X_i - X_t}{h_n})'(\frac{X_i - X_t}{h_n})$, $y_{i*} = y_i - r(x_{1t}) - (X_i - X_t)[r^{(1)'}(x_{1t}), 0_{d_2}']' = \epsilon_i + (1/2)(x_{1i} - x_{1t})r^{(2)}(x_{1it})(x_{1i} - x_{1t})'$, and $x_{1it} = \lambda_i x_{1i} + (1 - \lambda_i)x_{1t}$ for $\lambda_i \in (0, 1)$. Define $I_1(X_t) = \frac{1}{nh_n^d f(X_t)}\sum_{i=1}^{n}K(\frac{X_i - X_t}{h_n})\epsilon_i(1 + o_p(1))$ and

$I_2(X_t) = \frac{1}{2nh_n^d f(X_t)}\sum_{i=1}^{n}K(\frac{X_i - X_t}{h_n})(x_{1i} - x_{1t})r^{(2)}(x_{1it})(x_{1i} - x_{1t})'(1 + o_p(1))$, we follow Theorem 1 (a)'s proof step (2) to obtain $\hat{m}(X_t) - r(x_{1t}) = I_1(X_t) + I_2(X_t)$ and
$\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - \frac{2}{n}\sum_{t=1}^{n}(I_1(X_t) + I_2(X_t))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(I_1(X_t) + I_2(X_t))^2$.

We show in sequence the following results

(i) $\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - \frac{2}{n}\sum_{t=1}^{n}I_1(X_t)\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}I_1^2(X_t) + o_p((nh_n^{d/2})^{-1})$, which follows from
(1)$-\frac{2}{n}\sum_{t=1}^{n}I_2(X_t)\epsilon_t = o_p((nh_n^{d/2})^{-1})$.(2)$\frac{1}{n}\sum_{t=1}^{n}I_2^2(X_t) = o_p((nh_n^{d/2})^{-1})$.(3)$\frac{1}{n}\sum_{t=1}^{n}I_1(X_t)I_2(X_t) = o_p((nh_n^{d/2})^{-1})$.

(ii) $\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - \frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - r(x_{1t}))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - r(x_{1t}))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 + o_p((nh_n^{d/2})^{-1})$.

(iii) $\hat{A}_{nG} - A_{nG} = o_p((nh_n^{\frac{d}{2}})^{-1})$, $\hat{A}_{1nG} - A_{1nG} = o_p((nh_n^{\frac{d}{2}})^{-1})$, and $\hat{V}_{TG} - V_{TG} = o_p(1)$.

Let $A_{nG} = \frac{1}{n^3 h_n^{2d}}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ t\neq i}}^{n}K^2(\frac{X_i - X_t}{h_n})\frac{\epsilon_i^2}{f^2(X_t)}$, $A_{1nG} = -\frac{2}{n^2 h_n^d}\sum_{t=1}^{n}K(0)\frac{\epsilon_t^2}{f(X_t)}$, $A_{2nG} = -\frac{2}{n^2 h_n^d}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ t\neq i}}^{n}K(\frac{X_i - X_t}{h_n})\frac{\epsilon_i\epsilon_t}{f(X_t)}$,
$V_{TG} = \frac{\sigma_{\phi G}^2}{(E\epsilon_t^2)^2}$ for $\sigma_{\phi G}^2 = E\frac{\sigma^4(X_t)}{f(X_t)}2\int(2K(\psi) - \kappa(\psi))^2 d\psi$, and for $K_{ij} = K(\frac{X_i - X_j}{h_n})$,
$A_{3nG} = \frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ t\neq i}}^{n}[\frac{\epsilon_i\epsilon_t}{h_n^d}E(\frac{1}{h_n^d f^2(X_j)}K_{ij}K_{tj}|X_t, X_i) + \frac{\epsilon_t\epsilon_i}{h_n^d}E(\frac{1}{h_n^d f^2(X_j)}K_{tj}K_{ij}|X_i, X_t)]$.

We can write $\frac{2}{n}\sum_{t=1}^{n}I_1(X_t)\epsilon_t = \frac{2}{n^2}\sum_{t=1}^{n}\sum_{i=1}^{n}\frac{1}{h_n^d f(X_t)}K_{it}\epsilon_i\epsilon_t(1 + o_p(1)) = -(A_{1nG} + A_{2nG})(1 + 0_p(1))$, and we can follow Theorem 1 (a)'s proof step (2)(ii) to obtain
$\frac{1}{n}\sum_{t=1}^{n}I_1^2(X_t) = \frac{1}{n^3 h_n^{2d}}\sum_{t=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}K_{it}K_{jt}\frac{\epsilon_i\epsilon_j}{f^2(X_t)}(1 + o_p(1)) = (A_{nG} + A_{3nG})(1 + o_p(1))$.

Furthermore, we follow Theorem 1 (a)'s proof step (3) to obtain

$nh_n^{d/2}(A_{2nG} + A_{3nG}) \xrightarrow{d} N(0, \sigma_{\phi G}^2)$. So results (i) and (ii) above give

$nh_n^{d/2}(\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 - \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2 + (A_{nG} + A_{1nG})(1 + o_p(1))) \xrightarrow{d} N(0, \sigma_{\phi G}^2)$. Since (ii) implies $\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 \xrightarrow{p} E\epsilon_t^2 > 0$, $I(\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2) \xrightarrow{p} 1$, we obtain

$nh_n^{d/2}[\hat{R}_G^2 + I(\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2)\frac{A_{nG} + A_{1nG}}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2}] \xrightarrow{d} N(0, V_{TG})$. This result and (iii) give the desired the claim in (I).

We show the claims (i)-(iii) below.

(i) (1) Define $\psi_{nti} = \frac{1}{h_n^d f(X_t)} K(\frac{X_i - X_t}{h_n})(x_{1i} - x_{1t}) r^{(2)}(x_{1it})(x_{1i} - x_{1t})' \epsilon_t$, and we can write $-\frac{2}{n}\sum_{t=1}^{n} I_2(X_t)\epsilon_t = -\frac{1}{n^2}\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}[\underbrace{\psi_{nti} + \psi_{nit}}_{\phi_{nti}}](1+o_p(1)) = -u_n(1+o_p(1))$, where u_n is a U-statistic. Since $E(\epsilon_t|X_t) = 0$, we apply Lemma 1 to obtain $u_n = \frac{1}{n}\sum_{t=1}^{n} E(\psi_{nti}|W_t) + O_p(n^{-1}(E\phi_{nti}^2)^{1/2})$ for $W_t = (X_t, \epsilon_t)$. Since $E\phi_{nti}^2 = O(h_n^4 h_n^{-d})$, with assumptions B1, B2 and B4, for $\psi = (\psi_1, \psi_2)$ and $\lambda \in (0,1)$,

$$E[E^2(\psi_{nti}|W_t)] = E[\frac{\epsilon_t h_n^2}{f(X_t)} \int K(\psi)\psi_1 r^{(2)}(x_{1t} + \lambda h_n \psi_1)\psi_1' f(x_{1t} + h_n\psi_1, x_{2t} + h_n\psi_2) d\psi]^2 = O(h_n^{2v}).$$

So $-\frac{2}{n}\sum_{t=1}^{n} I_2(X_t)\epsilon_t = O_p(n^{-1/2}h_n^v) + o_p((nh_n^{d/2})^{-1})$. With assumption B3(3), we have the claimed result.

(2) Let $\psi_{ntij} = \frac{1}{h_n^{2d} f^2(X_t)} K(\frac{X_i - X_t}{h_n})(x_{1i} - x_{1t}) r^{(2)}(x_{1it})(x_{1i} - x_{1t})' K(\frac{X_j - X_t}{h_n})(x_{1j} - x_{1t}) r^{(2)}(x_{1jt})(x_{1j} - x_{1t})'$,

When $t \neq i \neq j$, $\frac{1}{n}\sum_{t=1}^{n} I_2^2(X_t) = \frac{1}{2n^3}\sum\sum\sum_{1=t<i<j=n} \phi_{ntij}$ is a U-statistic with $\phi_{ntij} = \psi_{ntij} + \psi_{nitj} + \psi_{njit}$. We apply Lemma 3 here and use its notations.

First, $\sigma_{3n}^2 = E\phi_{ntij}^2 = O(h_n^{8-2d})$, thus $H_n^{(3)} = O_p((n^{-3}h_n^{8-2d})^{1/2}) = o_p(n^{-1})$. Similarly, $\sigma_{2n}^2 \leq CE[E^2(\phi_{ntij}|W_i, W_j)] = O(h_n^{8-d})$. Thus, $H_n^{(2)} = O_p((n^{-2}h_n^{8-d})^{1/2}) = o_p((nh_n^{d/2})^{-1})$ with assumption $B3(3)$. So we have $\frac{1}{n}\sum_{t=1}^{n} I_2^2(X_t) = \frac{1}{2}(\theta_n + \frac{3}{n}[\sum_{t=1}^{n}\psi_{1n}(X_t) - n\theta_n]) + o_p((nh_n^{d/2})^{-1})$.

Second, note $\psi_{1n}(X_t) = E(\psi_{ntij}|X_t) + E(\psi_{nitj}|X_t) + E(\psi_{njit}|X_t)$. With assumptions B1-B4, we have $\frac{1}{n}\sum_{t=1}^{n} I_2^2(X_t) = O(h_n^{2v}) + o_p((nh_n^{d/2})^{-1}) = o_p((nh_n^{d/2})^{-1})$.

When $t \neq i = j$, we can show $\frac{1}{n}\sum_{t=1}^{n} I_2^2(X_t) = O_p(n^{-1}h_n^{4-d}) = o_p((nh_n^{d/2})^{-1})$ if $d < 8$.

(3) Repeated applications of Lemmas 1 and 3 together with assumptions B1-B4 give $\frac{1}{n}\sum_{t=1}^{n} I_1(X_t)I_2(X_t) = O_p(n^{-1/2}h_n^v) + o_p((nh_n^{d/2})^{-1}) = o_p((nh_n^{d/2})^{-1})$.

(ii) We note from the proof of Theorem 3(2),

$$\begin{aligned}\hat{r}(x_{1t}) - r(x_{1t}) = \; & \frac{1}{nh_{1n}^{d_1} f_1(x_{1t})}\sum_{i=1}^{n} K_1(\tfrac{x_{1i} - x_{1t}}{h_{1n}})\epsilon_i \\ & + \frac{1}{2nh_{1n}^{d_1} f_1(x_{1t})}\sum_{i=1}^{n} K_1(\tfrac{x_{1i} - x_{1t}}{h_{1n}})(x_{1i} - x_{1t}) r^{(2)}(x_{1it})(x_{1i} - x_{1t})' + w_n(x_{1t}) \\ = \; & r_1(x_{1t}) + r_2(x_{1t}) + w_n(x_{1t})\end{aligned}$$

where x_{1it} lie between x_{1i} and x_{1t}, $w_n(x_{1t})$ is of smaller order than $r_1(x_{1t})$ and $r_2(x_{1t})$, so we only focus on analyzing terms involving $r_1(x_{1t})$ and $r_2(x_{1t})$.

(1) $\frac{1}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - r(x_{1t}))^2 \leq C[\frac{1}{n}\sum_{t=1}^{n} r_1^2(x_{1t}) + \frac{1}{n}\sum_{t=1}^{n} r_2^2(x_{1t}) + \frac{1}{n}\sum_{t=1}^{n} w_n^2(x_{1t})]$.

$\frac{1}{n}\sum_{t=1}^{n} r_1^2(x_{1t}) = \frac{1}{n^3}\sum_{t=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{h_{1n}^{2d_1} f_1^2(x_{1t})} K_1(\frac{x_{1i} - x_{1t}}{h_{1n}}) K_1(\frac{x_{1j} - x_{1t}}{h_{1n}})\epsilon_i\epsilon_j$. We apply Lemmas 1, 2 to obtain $\frac{1}{n}\sum_{t=1}^{n} r_1^2(x_{1t}) = O_p(n^{-1}h_{1n}^{-d_1/2}) + O_p((nh_{1n}^{d_1})^{-1}) = o_p((nh_n^{d/2})^{-1})$, using B_3 that $h_n^{d/2}/h_{1n}^{d_1} = o(1)$.

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} r_2^2(x_{1t}) = \; & \frac{1}{4n^3}\sum_{t=1}^{n}\sum_{\substack{i=1\\i\neq t}}^{n}\sum_{\substack{j=1\\j\neq t}}^{n}\frac{1}{h_{1n}^{2d_1} f_1^2(x_{1t})} K_1(\tfrac{x_{1i} - x_{1t}}{h_{1n}}) K_1(\tfrac{x_{1j} - x_{1t}}{h_{1n}})(x_{1i} - x_{1t}) r^2(x_{1it})(x_{1i} - x_{1t})' \\ & \times (x_{1j} - x_{1t}) r^2(x_{1jt})(x_{1j} - x_{1t})'.\end{aligned}$$

Expressing above as an U-statistic, applying Lemmas 1 and 3, and using the assumptions B1 and B2 that $f_1(\cdot)$ and $r(\cdot)$ are $C_1^{v_1}$, we obtain $\frac{1}{n}\sum_{t=1}^{n} r_2^2(x_{1t}) = O_p(h_{1n}^{2v_1}) + o_p((nh_{1n}^{d_1/2})^{-1}) + O_p((nh_{1n}^{d_1})^{-1}h_{1n}^4)$.

So in all, we obtain $\frac{1}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - r(x_{1t}))^2 = o_p(n^{-1}h_n^{-d/2})$ using assumption B3.

(2) $-\frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-r(x_{1t}))\epsilon_t = -2[\frac{1}{n}\sum_{t=1}^{n} r_1(x_{1t})\epsilon_t + \frac{1}{n}\sum_{t=1}^{n} r_2(x_{1t})\epsilon_t + \frac{1}{n}\sum_{t=1}^{n} w_n(x_{1t})\epsilon_t]$.

Given the results in Theorem 1, we obtain

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} r_1(x_{1t})\epsilon_t &= \frac{1}{n^2}\sum_{\substack{t=1\\t\neq i}}^{n}\sum_{i=1}^{n}\frac{1}{h_{1n}^{d_1}f_1(x_{1t})}K_1(\tfrac{x_{1i}-x_{1t}}{h_{1n}})\epsilon_i\epsilon_t + \frac{1}{n^2}\sum_{t=1}^{n}\frac{1}{h_{1n}^{d_1}f_1(x_{1t})}K_1(0)\epsilon_t^2 \\
&= O_p(n^{-1}h_{1n}^{-d_1/2}) + O_p((nh_{1n}^{d_1})^{-1}) = o_p(n^{-1}h_n^{-d/2}) \text{ using assumption B3.} \\
\frac{1}{n}\sum_{t=1}^{n} r_2(x_{1t})\epsilon_t &= \frac{1}{2n^2}\sum_{\substack{t=1\\t\neq i}}^{n}\sum_{i=1}^{n}\frac{1}{h_{1n}^{d_1}f_1(x_{1t})}K_1(\tfrac{x_{1i}-x_{1t}}{h_{1n}})(x_{1i}-x_{1t})r^2(x_{1it})(x_{1i}-x_{1t})'\epsilon_t \\
&= O_p(n^{-1/2}h_{1n}^{v_1}) + o_p((nh_{1n}^{d_1/2})^{-1}) \text{ again with assumption B3.}
\end{aligned}$$

So $-\frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-r(x_{1t}))\epsilon_t = o_p(n^{-1}h_n^{-d/2}))$. (1) and (2) imply the claim in (ii).

(iii) Given that $\tilde{\epsilon}_t - \epsilon_t = r(x_{1t}) - \hat{r}(x_{1t}) = O_p((\frac{lnn}{nh_{1n}^{d_1}})^{1/2} + h_{1n}^{v_1}) = O_p(L_{1nG})$ uniformly over $x_{1t}\in G_1$, we have $nh_n^{d/2}\frac{L_{1nG}}{nh_n^d} = o_p(1)$ with assumption B3 on the bandwidths. We follow the proof in Theorem 1(b) to have the claimed result.

(II) Under the Pitman local alternative $H_{1G}(l_n)$, $l_n = n^{-1/2}h_n^{-d/4}$, $m(X_t) = r(x_{1t}) + l_nD(X_t)$.

$$\begin{aligned}
y_{il*} &= y_i - r(x_{1t}) - l_nD(X_t) - (x_{1i}-x_{1t})r^{(1)}(x_{1t}) \\
&= \epsilon_i + \underbrace{\frac{1}{2}(x_{1i}-x_{1t})r^{(2)}(x_{1it})(x_{1i}-x_{1t})'}_{r_{it}^*} + l_n(D(X_i)-D(X_t)), \text{ then}
\end{aligned}$$

$$\hat{m}(X_t)-(r(x_{1t})+l_nD(X_t)) = \frac{1}{nh_n^d}\sum_{i=1}^{n}(1,0_d')S_n^{-1}(X_t)(1,(\tfrac{X_i-X_t}{h_n}))'K(\tfrac{X_i-X_t}{h_n})y_{il*} = I_1(X_t)+I_2(X_t)+I_3(X_t).$$

$I_1(X_t)$ and $I_2(X_t)$ are defined in (I), and $I_3(X_t) = \frac{l_n}{nh_n^d f(X_t)}\sum_{i=1}^{n}K(\frac{X_i-X_t}{h_n})(D(X_i)-D(X_t))(1+o_p(1))$.

Following the proof of part (I), the claim of (II) follows from the three results below.

(i) $\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{m}(X_t))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - \frac{2}{n}\sum_{t=1}^{n}I_1(X_t)\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}I_1^2(X_t) + o_p((nh_n^{d/2})^{-1})$.

(ii) $\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{r}(x_{it}))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 + l_n^2E[D(X_t)-E(D(x_{1t},x_{2j})|x_{1t})]^2 + o_p((nh_n^{d/2})^{-1})$.

(iii) $\hat{A}_{nG} - A_{nG} = o_p((nh_n^{\frac{d}{2}})^{-1})$, $\hat{A}_{1nG} - A_{1nG} = o_p((nh_n^{\frac{d}{2}})^{-1})$, and $\hat{V}_{TG} - V_{TG} = o_p(1)$.

$$\begin{aligned}
\text{(i)}\quad & \frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{m}(X_t))^2 \\
=\ & \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 + \frac{2}{n}\sum_{t=1}^{n}(r(x_{1t})+l_nD(X_t)-\hat{m}(X_t))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(r(x_{1t})+l_nD(X_t)-\hat{m}(X_t))^2 \\
=\ & \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - \frac{2}{n}\sum_{t=1}^{n}I_1(X_t)\epsilon_t - \frac{2}{n}\sum_{t=1}^{n}I_2(X_t)\epsilon_t - \frac{2}{n}\sum_{t=1}^{n}I_3(X_t)\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}I_1^2(X_t) + \frac{1}{n}\sum_{t=1}^{n}I_2^2(X_t) \\
& + \frac{1}{n}\sum_{t=1}^{n}I_3^2(X_t) + \frac{2}{n}\sum_{t=1}^{n}I_1(X_t)I_2(X_t) + \frac{2}{n}\sum_{t=1}^{n}I_1(X_t)I_3(X_t) + \frac{2}{n}\sum_{t=1}^{n}I_2(X_t)I_3(X_t).
\end{aligned}$$

From Part (I)(i) above, we have $-\frac{2}{n}\sum_{t=1}^{n}I_2(X_t)\epsilon_t = o_p((nh_n^{d/2})^{-1})$, $\frac{1}{n}\sum_{t=1}^{n}I_2^2(X_t) = o_p((nh_n^{d/2})^{-1})$, and $\frac{2}{n}\sum_{t=1}^{n}I_1(X_t)I_2(X_t) = o_p((nh_n^{d/2})^{-1})$. We follow Theorem 2's proof in part 1 to obtain $\frac{2}{n}\sum_{t=1}^{n}I_3(X_t)\epsilon_t = o_p(n^{-1/2}l_n) = o_p((nh_n^{d/2})^{-1})$, $\frac{1}{n}\sum_{t=1}^{n}I_3^2(X_t) = o_p((nh_n^{d/2})^{-1})$ and $\frac{2}{n}\sum_{t=1}^{n}I_1(X_t)I_3(X_t) = o_p((nh_n^{d/2})^{-1})$. So we only need to show $\frac{2}{n}\sum_{t=1}^{n}I_2(X_t)I_3(X_t) = o_p((nh_n^{d/2})^{-1})$.

$\frac{1}{n}\sum_{t=1}^{n}I_2(X_t)I_3(X_t) = \frac{l_n}{2}\frac{1}{n^3}\sum_{t=1}^{n}\sum_{\substack{i=1\\i\neq t}}^{n}\sum_{\substack{j=1\\j\neq t}}^{n}\frac{1}{h_n^{2d}f^2(X_t)}K(\frac{X_i-X_t}{h_n})K(\frac{X_j-X_t}{h_n})r_{it}^*(D(X_j)-D(X_t))(1+o_p(1))$. By

Lemma 3 and assumptions $B_1 - B_4$, $\frac{1}{n}\sum_{t=1}^{n} I_1(X_t)I_3(X_t) = O_p(l_n(h_n^v + n^{-1/2})) = o_p((nh_n^{d/2})^{-1})$.

(ii) Using the local linear estimator, we define $r_3(X_t) = \frac{l_n}{nh_{1n}^{d_1} f_1(x_{1t})}\sum_{i=1}^{n} K_1(\frac{x_{1i}-x_{1t}}{h_{1n}})(D(X_i) - D(X_t))$, then $\hat{r}(x_{1t}) - r(x_{1t}) - l_n D(X_t) = r_1(x_{1t}) + r_2(x_{1t}) + r_3(X_t) + w_n(x_{1t})$, where $w_n(x_{1t})$ is of smaller order.

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{it}))^2 = & \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 + \frac{2}{n}\sum_{t=1}^{n}(r(x_{1t}) + l_n D(X_t) - \hat{r}(x_{1t}))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(r(x_{1t}) + l_n D(X_t) - \hat{r}(x_{1t}))^2 \\
= & \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - [\frac{2}{n}\sum_{t=1}^{n} r_1(x_{1t})\epsilon_t + \frac{2}{n}\sum_{t=1}^{n} r_2(x_{1t})\epsilon_t + \frac{2}{n}\sum_{t=1}^{n} r_3(X_t)\epsilon_t + \frac{2}{n}\sum_{t=1}^{n} w_n(x_{1t})\epsilon_t] \\
& +[\frac{1}{n}\sum_{t=1}^{n} r_1^2(x_{1t}) + \frac{1}{n}\sum_{t=1}^{n} r_2^2(x_{1t}) + \frac{1}{n}\sum_{t=1}^{n} r_3^2(X_t) + \frac{2}{n}\sum_{t=1}^{n} r_1(x_{1t})r_2(x_{1t}) \\
& + \frac{2}{n}\sum_{t=1}^{n} r_1(x_{1t})r_3(X_t) + \frac{2}{n}\sum_{t=1}^{n} r_2(x_{1t})r_3(X_t)](1 + o_p(1)).
\end{aligned}$$

In (I), $\frac{2}{n}\sum_{t=1}^{n} r_1(x_{1t})\epsilon_t$ and $\frac{2}{n}\sum_{t=1}^{n} r_2(x_{1t})\epsilon_t$ are $o_p((nh_n^{d/2})^{-1})$. With assumption B3 and Lemma 1,

$$\begin{aligned}
\frac{2}{n}\sum_{t=1}^{n} r_3(X_t)\epsilon_t = & \frac{2l_n}{n^2}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ i\neq t}}^{n} \frac{1}{h_{1n}^{d_1} f_1(x_{1t})} K_1(\frac{x_{1i}-x_{1t}}{h_{1n}})(D(X_i) - D(X_t))\epsilon_t \\
= & o_p(n^{-1/2} l_n) + o_p((nh_n^{d/2})^{-1}) = o_p((nh_n^{d/2})^{-1}).
\end{aligned}$$

In (I), we have $\frac{1}{n}\sum_{t=1}^{n} r_1^2(x_{1t})$ and $\frac{1}{n}\sum_{t=1}^{n} r_2^2(x_{1t})$ are $o_p((nh_n^{d/2})^{-1})$, so $\frac{2}{n}\sum_{t=1}^{n} r_1(x_{1t})r_2(x_{1t}) = o_p((nh_n^{d/2})^{-1})$.

$$\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n} r_3^2(x_{1t}) = \frac{l_n^2}{n^3}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ i\neq t}}^{n}\sum_{\substack{j=1 \\ j\neq t}}^{n} \frac{1}{h_{1n}^{2d_1} f_1^2(x_{1t})} K_1(\frac{x_{1i}-x_{1t}}{h_{1n}})K_1(\frac{x_{1j}-x_{1t}}{h_{1n}})(D(X_i) - D(X_t))(D(X_j) - D(X_t)) \\
= & l_n^2 E[D(X_t) - E(D(x_{1t}, x_{2j})|x_{1t})]^2 + o_p((nh_n^{d/2})^{-1}), \text{ by Lemma 3.} \\
& \frac{1}{n}\sum_{t=1}^{n} r_1(x_{1t})r_3(x_{1t}) = \frac{l_n}{n^3}\sum_{t=1}^{n}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq t}}^{n} \frac{1}{h_{1n}^{2d_1} f_1^2(x_{1t})} K_1(\frac{x_{1i}-x_{1t}}{h_{1n}})K_1(\frac{x_{1j}-x_{1t}}{h_{1n}})\epsilon_i(D(X_j) - D(X_t)) , \\
= & o_p(l_n(n^{-1/2} + (nh_{1n}^{d_1/2})^{-1})) + o_p((nh_{1n}^{d_1/2})^{-1}) = o_p((nh_n^{d/2})^{-1}) \text{ with assumption B3.} \\
& \frac{1}{n}\sum_{t=1}^{n} r_2(x_{1t})r_3(x_{1t}) = \frac{l_n}{2n^3}\sum_{t=1}^{n}\sum_{\substack{i=1 \\ i\neq t}}^{n}\sum_{\substack{j=1 \\ j\neq t}}^{n} \frac{1}{h_{1n}^{2d_1} f_1^2(x_{1t})} K_1(\frac{x_{1i}-x_{1t}}{h_{1n}})K_1(\frac{x_{1j}-x_{1t}}{h_{1n}})r_{it}^*(D(X_j) - D(X_t)) \\
= & o_p((nh_n^{d/2})^{-1}) \text{with similar arguments. Above results imply the claim in (ii).}
\end{aligned}$$

(iii) Under $H_{1G}(l_n)$, $\tilde{\epsilon}_t - \epsilon_t = -(\hat{r}(x_{1t}) - r(x_{1t}) - l_n D(X_t)) = -(r_1(x_{1t}) + r_2(x_{1t}) + r_3(X_t) + w_n(x_{1t})) = O_p(L_{1nG}) + r_3(X_t)$ uniformly over $X_{1t} \in G_1$, and $r_3(X_t) = O_p(l_n)$ uniformly over $X_t \in G$. Since $nh_n^{d/2}\frac{l_n}{nh_n^d} = (nh_n^{3d/2})^{-1/2} = o(1)$, we follow (I), Theorems 1(b) and 2 to obtain the claimed result.

(III). We note under H_{1G}, $y_t = m(X_t) + \epsilon_t$. Thus,

$$\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - \frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - m(X_t))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - m(X_t))^2.$$

$-\frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - m(X_t))\epsilon_t = -\frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - r(x_{1t}))\epsilon_t - \frac{2}{n}\sum_{t=1}^{n}(r(x_{1t}) - m(X_t))\epsilon_t = o_p(1)$ by (I)(ii).

$$\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - m(X_t))^2 \\
= & \frac{1}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - r(x_{1t}))^2 + \frac{1}{n}\sum_{t=1}^{n}(r(x_{1t}) - m(X_t))^2 + \frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t}) - r(x_{1t}))(r(x_{1t}) - m(X_t))^2 \\
= & E(r(x_{1t}) - m(X_t))^2 + o_p(1) \text{ by (I)(ii) and } E(r(x_{1t}) - m(X_t))^2 < \infty \text{ by assumption B0.}
\end{aligned}$$

So we have (1) $\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 = E(\epsilon_t^2) + E(r(x_{1t}) - m(X_t))^2 = E(y_t - E(y_t|x_{1t}))^2 + o_p(1)$.

$$\begin{aligned}
(2)\ \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{m}(X_t))^2 = & \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2 - \frac{2}{n}\sum_{t=1}^{n}(\hat{m}(X_t) - m(X_t))\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(\hat{m}(X_t) - m(X_t))^2 \\
= & E(y_t - E(y_t|X_t))^2 + o_p(1),
\end{aligned}$$

following the arguments in Theorem 3, which is valid under assumptions B0-B6. (1) and (2) implies

$I(\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{r}(x_{1t}))^2 \geq \frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{m}(X_t))^2) \equiv I(\cdot) \xrightarrow{p} 1.$

Recall the definition of $\hat{R}_G^2$ in $\hat{T}_{nG}$ in (5) as $\hat{R}_G^2 = [1-\frac{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{m}(X_t))^2}{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{r}(x_{1t}))^2}]I(\cdot)$, so we obtain $\hat{R}_G^2 \xrightarrow{p} R_G^2$, for $0<R_G^2<\infty$ under assumption B0. Since $\hat{A}_{nG}=O_p((nh_n^d)^{-1})$, $\hat{A}_{1nG}=O_p((nh_n^d)^{-1})$ and $\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{r}(x_{1t}))^2 \xrightarrow{p} E(y_t-E(y_t|x_{1t}))^2>0$, we have $\hat{T}_{nG} = \frac{nh_n^{\frac{d}{2}}\{\hat{R}_G^2+I(\cdot)\frac{\hat{A}_{1nG}+\hat{A}_{nG}}{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{r}(x_{1t}))^2}\}}{\sqrt{\hat{V}_{TG}}} = \frac{nh_n^{\frac{d}{2}}\{R_G^2+o_p(1)\}}{\sqrt{\hat{V}_{TG}}} > c_n = o_p(nh_n^{d/2}))$ if $\hat{V}_{TG} \xrightarrow{p} c$ for $0<c<\infty$.

Recall $\hat{V}_{TG} = \frac{\hat{\sigma}_{\phi G}^2}{(\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{r}(x_{1t}))^2)^2}$. Given result (1) above, we only need to show that

$\hat{\sigma}_{\phi G}^2/[\int 2(2K(\psi)-\kappa(\psi))^2 d\psi] = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{\substack{i=1\\t\neq i}}^{n}\frac{1}{h_n^d \hat{f}^2(X_t)}K(\frac{X_i-X_t}{h_n})\tilde{\epsilon}_i^2\tilde{\epsilon}_t^2 \xrightarrow{p} c_1$ for $0<c_1<\infty$.

$\tilde{\epsilon}_t-\epsilon_t = m(X_t)-\hat{r}(x_{1t}) = m(X_t)-r(x_{1t})-(\hat{r}(x_{1t})-r(x_{1t}))$. $\hat{r}(x_{1t})-r(x_{1t}) = O_p(L_{1nG})$ uniformly over $x_{1t}\in G_1$, and $m(X_t)-r(x_{1t})\neq 0$. These results imply

$\frac{1}{n^2}\sum_{t=1}^{n}\sum_{\substack{i=1\\t\neq i}}^{n}\frac{1}{h_n^d\hat{f}^2(X_t)}K(\frac{X_i-X_t}{h_n})\tilde{\epsilon}_i^2\tilde{\epsilon}_t^2 \xrightarrow{p} E(\frac{\sigma^4(X_t)}{f(X_t)}) + E(\frac{(m(X_t)-r(x_{1t}))^4}{f(X_t)}) + 2E(\frac{(m(X_t)-r(x_{1t}))^2\sigma^2(X_t)}{f(X_t)}) = c_1,$

where $0<c_1<\infty$ follows from our assumptions B0 and B1. So we obtain the claim in (III).

(IV). We point out that more complicated expressions show up in the test statistics $\hat{T}_{nG}^*$. Since $y_i^*=\hat{r}(x_{1i})+\epsilon_i^*$, and denote $K_{1it}=K_1(\frac{x_{1i}-x_{1t}}{h_{1n}})$,

$$\begin{aligned}
\hat{m}^*(X_t) &= (1,0_d')\frac{1}{nh_n^d}\sum_{i=1}^{n}S_n^{-1}(X_t)(1,(\frac{X_i-X_t}{h_n}))'K_{it}y_i^* \\
&= [\frac{1}{nh_n^d f(X_t)}\sum_{t=1}^{n}K_{it}\epsilon_i^* + \frac{1}{nh_n^d f(X_t)}\sum_{t=1}^{n}K_{it}\hat{r}(x_{1i})](1+o_p(1)) = [m_1^*(X_t)+m_2^*(X_t)](1+o_p(1)).
\end{aligned}$$

$$\begin{aligned}
\hat{r}^*(x_{1t}) &= \frac{1}{nh_{1n}^{d_1}f_1(x_{1t})}\sum_{i=1}^{n}K_{1it}\epsilon_i^* + \frac{1}{nh_{1n}^{d_1}f_1(x_{1t})}\sum_{i=1}^{n}K_{1it}\hat{r}(x_{1i}) + w_n^*(x_{1t}) \\
&= r_1^*(x_{1t})+r_2^*(x_{1t})+w_n^*(x_{1t}) \text{ where } w_n^*(x_{1t}) \text{ is of smaller order.}
\end{aligned}$$

We first obtain the following results.

$$\begin{aligned}
(1)\ \frac{1}{n}\sum_{t=1}^{n}(\epsilon_{t,0}^*)^2 &= \frac{1}{n}\sum_{t=1}^{n}(y_t^*-\hat{r}^*(x_{1t}))^2 = \frac{1}{n}\sum_{t=1}^{n}(\epsilon_t^*+\hat{r}(x_{1t})-\hat{r}^*(x_{1t}))^2 \\
&= [\frac{1}{n}\sum_{t=1}^{n}(\epsilon_t^*)^2 + \frac{1}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-r_2^*(x_{1t}))^2 + \frac{1}{n}\sum_{t=1}^{n}(r_1^*(x_{1t}))^2 + \frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-r_2^*(x_{1t}))\epsilon_t^* \\
&\quad -\frac{2}{n}\sum_{t=1}^{n}r_1^*(x_{1t})\epsilon_t^* - \frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-r_2^*(x_{1t}))r_1^*(x_{1t})](1+o_p(1)) \\
&= [\frac{1}{n}\sum_{t=1}^{n}(\epsilon_t^*)^2 + o_p((nh_n^{d/2})^{-1})](1+o_p(1)).
\end{aligned}$$

$$\begin{aligned}
(2)\ \frac{1}{n}\sum_{t=1}^{n}(\epsilon_{t,b}^*)^2 &= \frac{1}{n}\sum_{t=1}^{n}(y_t^*-\hat{m}^*(X_t))^2 = \frac{1}{n}\sum_{t=1}^{n}(\epsilon_t^*+\hat{r}(x_{1t})-\hat{m}^*(X_t))^2 \\
&= [\frac{1}{n}\sum_{t=1}^{n}(\epsilon_t^*)^2 + \frac{1}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m_2^*(X_t))^2 + \frac{1}{n}\sum_{t=1}^{n}(m_1^*(X_t))^2 + \frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m_2^*(X_t))\epsilon_t^* \\
&\quad -\frac{2}{n}\sum_{t=1}^{n}m_1^*(X_t)\epsilon_t^* - \frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m_2^*(X_t))m_1^*(X_t)](1+o_p(1)) \\
&= \frac{1}{n}\sum_{t=1}^{n}(\epsilon_t^*)^2 + (A_{nG}^*+A_{1nG}^*+A_{2nG}^*+A_{3nG}^*)(1+o_p(1)) + o_p((nh_n^{d/2})^{-1}),
\end{aligned}$$

where $A_{nG}^* = \frac{1}{n^3h_n^{2d}}\sum_{t=1}^{n}\sum_{\substack{i=1\\t\neq i}}^{n}\frac{K_{it}^2}{f^2(X_t)}(\epsilon_i^*)^2$, $A_{1nG}^* = -\frac{2}{n^2h_n^d}K(0)\sum_{t=1}^{n}\frac{(\epsilon_t^*)^2}{f(X_t)}$, $A_{2nG}^* = -\frac{2}{n^2h_n^d}\sum_{t=1}^{n}\sum_{\substack{i=1\\t\neq i}}^{n}\frac{K_{it}}{f(X_t)}\epsilon_t^*\epsilon_i^*$, and

$A_{3nG}^* = \frac{1}{n^3h_n^{2d}}\sum_{t=1}^{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\t\neq i\neq j}}^{n}\frac{K_{it}K_{jt}}{f^2(X_t)}\epsilon_i^*\epsilon_j^* = \frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{\substack{i=1\\t<i}}^{n}[\frac{\epsilon_i^*\epsilon_t^*}{h_n^{2d}}E(\frac{K_{ij}K_{tj}}{f^2(X_j)}|X_t,X_i)+\frac{\epsilon_t^*\epsilon_i^*}{h_n^{2d}}E(\frac{K_{tj}K_{ij}}{f^2(X_j)}|X_i,X_t)]+o_p((nh_n^{\frac{d}{2}})^{-1}).$

(3) We obtain $(A^*_{2nG}+A^*_{3nG})(S^*_n)^{-1} \xrightarrow{d} N(0,1)$ conditioning on W, where $\phi_n(Z^*_t, Z^*_i) = \frac{1}{nh_n^d}[-\frac{2}{f(X_t)}K_{it}\epsilon^*_i\epsilon^*_t - \frac{2}{f(X_i)}K_{ti}\epsilon^*_t\epsilon^*_i + \frac{\epsilon^*_i\epsilon^*_t}{h_n^{2d}}E(\frac{K_{ij}K_{tj}}{f^2(X_j)}|X_t, X_i) + \frac{\epsilon^*_t\epsilon^*_i}{h_n^{2d}}E(\frac{K_{tj}K_{ij}}{f^2(X_j)}|X_i, X_t)]$, and $(S^*_n)^2 = E((\sum_{\substack{t=1\\t<i}}^n\sum_{i=1}^n \phi_n(Z^*_t, Z^*_i))^2|W) = n^{-2}h_n^{-d}\sigma^2_{\phi G} + o_p(n^{-2}h_n^{-d})$, where $\sigma^2_{\phi G}$ is defined in (I).

Results (1)-(3) imply $\frac{nh_n^{\frac{d}{2}}}{nh_n^{\frac{d}{2}}S^*_n}[\frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,0})^2 - \frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,b})^2 + (A^*_{nG} + A^*_{1nG})(1+o_p(1))] \xrightarrow{d} N(0,1)$.

Given (1), $E(\frac{1}{n}\sum_{t=1}^n(\epsilon^*_t)^2|W) = \frac{1}{n}\sum_{t=1}^n\hat{\epsilon}^2_t \xrightarrow{p} E\epsilon^2_t > 0$, and $I(\frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,0})^2 \geq \frac{1}{n}\sum_{t=1}^n(\epsilon^*_{t,b})^2) \equiv I(\cdot) \xrightarrow{p} 1$, we have $\frac{nh_n^{\frac{d}{2}}}{nh_n^{\frac{d}{2}}\frac{S^*_n}{E\epsilon^2_t}}[\hat{R}^{*2}_G + \frac{(A^*_{nG}+A^*_{1nG})}{\frac{1}{n}\sum_{t=1}^n(\epsilon^*_t)^2}I(\cdot)(1+o_p(1))] \xrightarrow{d} N(0,1)$.

Given the definition of $\hat{T}^*_{nG}$, we only need to show

(4) $\hat{A}^*_{nG} - A^*_{nG} = o_p((nh_n^{\frac{d}{2}})^{-1})$. $\hat{A}^*_{1nG} - A^*_{1nG} = o_p((nh_n^{\frac{d}{2}})^{-1})$ and $\hat{V}^*_{TG} - [nh_n^{\frac{d}{2}}\frac{S^*_n}{E\epsilon^2_t}]^2 = o_p(1)$.

The claim in (IV) follows from (1)-(4) above. (3) is obtained as in Theorem 5's proof (4). So we only sketch the key results in (1), (2) and (4) below.

(1) (i) $\frac{1}{n}\sum_{t=1}^n r^*_1(x_{1t})\epsilon^*_t = \frac{1}{n^2}\sum_{t=1}^n\sum_{i=1}^n\frac{K_{1it}\epsilon^*_i\epsilon^*_t}{h_{1n}^{d_1}f_1(x_{1t})} = o_p((nh_n^{d/2})^{-1})$. Since for $t \neq i$, $V(\frac{1}{n}\sum_{t=1}^n r^*_1(x_{1t})\epsilon^*_t|W) = \frac{1}{n^4}\sum_{t=1}^n\sum_{i=1}^n\frac{K^2_{1it}\hat{\epsilon}^2_i\hat{\epsilon}^2_t}{h_{1n}^{2d_1}f^2_1(x_{1t})} = O_p((nh_{1n}^{d_1/2})^{-2})$, so $\frac{1}{n}\sum_{t=1}^n r^*_1(x_{1t})\epsilon^*_t = O_p((nh_{1n}^{d_1/2})^{-1})$. When $t = i$, $\frac{1}{n}\sum_{t=1}^n r^*_1(x_{1t})\epsilon^*_t = O_p((nh_{1n}^{d_1})^{-1})$. In both cases, $\frac{1}{n}\sum_{t=1}^n r^*_1(x_{1t})\epsilon^*_t = o_p((nh_n^{d/2})^{-1})$ using assumption B3.

(ii) $\frac{1}{n}\sum_{t=1}^n(r^*_1(x_{1t}))^2 = \frac{1}{n^3}\sum_{t=1}^n\sum_{i=1}^n\sum_{j=1}^n\frac{K_{1it}K_{1jt}}{h_{1n}^{2d_1}f^2_1(x_{1t})}\epsilon^*_i\epsilon^*_j = o_p((nh_n^{d/2})^{-1})$. When t, i, and j are different, we show that $V(\frac{1}{n}\sum_{t=1}^n(r^*_1(x_{1t}))^2|W) = \frac{1}{n^6}\sum_{t=1}^n\sum_{t'=1}^n\sum_{i=1}^n\sum_{j=1}^n\frac{K_{1it}K_{1jt}K_{1it'}K_{1jt'}}{h_{1n}^{4d_1}f^2_1(x_{1t})f^2_1(x_{1t'})}\hat{\epsilon}^2_i\hat{\epsilon}^2_j = O_p(n^{-2})$, so $\frac{1}{n}\sum_{t=1}^n(r^*_1(x_{1t}))^2 = O_p(n^{-1})$. When some of the indices are the same, we can show $\frac{1}{n}\sum_{t=1}^n(r^*_1(x_{1t}))^2 = o_p((nh_n^{d/2})^{-1})$.

(iii) $\frac{2}{n}\sum_{t=1}^n(\hat{r}(x_{1t}) - r^*_2(x_{1t}))\epsilon^*_t = o_p((nh_n^{d/2})^{-1})$. $V(\frac{2}{n}\sum_{t=1}^n(\hat{r}(x_{1t}) - r^*_2(x_{1t}))\epsilon^*_t|W) = \frac{4}{n^2}\sum_{t=1}^n(\hat{r}(x_{1t}) - r^*_2(x_{1t}))^2\hat{\epsilon}^2_t$, whose order of magnitude in probability is the same as that of $\frac{4}{n^2}\sum_{t=1}^n(\hat{r}(x_{1t}) - r^*_2(x_{1t}))^2\epsilon^2_t$, which is less than $\frac{c}{n^2}\sum_{t=1}^n(\hat{r}(x_{1t})-r(x_{1t}))^2\epsilon^2_t + \frac{c}{n^2}\sum_{t=1}^n(r(x_{1t})-r^*_2(x_{1t}))^2\epsilon^2_t$. The first term is $O_p(n^{-2}h_{1n}^{-d_1/2}) + O_p(n^{-1}h_{1n}^{2v_1})$ following (I)(ii). $r^*_2(x_{1t}) - r(x_{1t}) = \frac{1}{nh_{1n}^{d_1}f_1(x_{1t})}\sum_{t=1}^n K_{1it}(\hat{r}(x_{1i}) - r(x_{1i}) + r(x_{1t}) - r(x_{1t})) = O_p(L_{1nG})$ uniformly over $x_{1t} \in G_1$, so the second term is $O_p(n^{-1}L^2_{1nG})$. $V(\frac{2}{n}\sum_{t=1}^n(\hat{r}(x_{1t})-r^*_2(x_{1t}))\epsilon^*_t|W) = O_p(n^{-1}(nh_{1n}^{d_1/2})^{-1}) + O_p(n^{-1}h_{1n}^{2v_1}) + O_p(n^{-1}L^2_{1nG})$, and by Markov's inequality, $\frac{2}{n}\sum_{t=1}^n(\hat{r}(x_{1t})-r^*_2(x_{1t}))\epsilon^*_t = O_p(n^{-1}h_{1n}^{-d_1/4}) + O_p(n^{-1/2}h_{1n}^{v_1}) + O_p(n^{-1/2}L_{1nG}) = o_p(n^{-1}h_n^{-d/2})$ by assumption B3.

(iv) $\frac{1}{n}\sum_{t=1}^n(\hat{r}(x_{1t}) - r^*_2(x_{1t}))^2 \leq c[\frac{1}{n}\sum_{t=1}^n(\hat{r}(x_{1t}) - r(x_{1t}))^2 + \frac{1}{n}\sum_{t=1}^n(r(x_{1t}) - r^*_2(x_{1t}))^2 = O_p((nh_{1n}^{d_1/2})^{-1}) + O_p(h_{1n}^{2v_1}) + O_p(L^2_{1nG}) = o_p(n^{-1}h_n^{-d/2})$ with similar arguments.

(v) Finally, $\frac{2}{n}\sum_{t=1}^n(\hat{r}(x_{1t})-r^*_2(x_{1t}))r^*_1(x_{1t}) \leq c[\frac{1}{n}\sum_{t=1}^n(\hat{r}(x_{1t})-r^*_2(x_{1t}))^2 + \frac{1}{n}\sum_{t=1}^n(r^*_1(x_{1t}))^2] = o_p(n^{-1}h_n^{-d/2})$. Above five results imply the claim in (1).

(2) (i) We expand the sums to obtain that $-\frac{2}{n}\sum_{t=1}^n(m^*_1(X_t))\epsilon^*_t = -\frac{2}{n}\sum_{t=1}^n\sum_{i=1}^n\frac{K_{it}}{f(X_t)}\epsilon^*_i\epsilon^*_t(1+o_p(1)) = (A^*_{1nG} + A^*_{2nG})(1+o_p(1))$. Similarly we obtain

(ii) $\frac{1}{n}\sum_{t=1}^n(m^*_1(X_t))^2 = \frac{1}{n^3h_n^{2d}}\sum_{t=1}^n\sum_{i=1}^n\sum_{j=1}^n\frac{K_{it}K_{jt}}{f^2(X_t)}\epsilon^*_i\epsilon^*_j(1+o_p(1)) = (A^*_{nG} + A^*_{3nG})(1+o_p(1)) + o_p((nh_n^{d/2})^{-1})$.

(iii) $\frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m_2^*(X_t))\epsilon_t^* = o_p((nh_n^{d/2})^{-1})$. We obtain $V(\frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m_2^*(X_t))\epsilon_t^*|W) = \frac{4}{n^2}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m_2^*(X_t))^2\hat{\epsilon}_t^2$ whose order of magnitude is the same as that of $\frac{1}{n^2}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m_2^*(X_t))^2\epsilon_t^2 \leq c[\frac{1}{n^2}\sum_{t=1}^{n}(m_2^*(X_t)-r(x_{1t}))^2\epsilon_t^2+\frac{1}{n^2}\sum_{t=1}^{n}(\hat{r}(x_{1t})-r(x_{1t}))^2\epsilon_t^2]$. The second term is $O_p(n^{-2}h_{1n}^{-d_1/2})+O_p(n^{-1}h_{1n}^{2v_1})$ following (I)(ii).

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n}(m_2^*(X_t)-r(x_{1t}))^2\epsilon_t^2 \\
= \;& \frac{1}{n^2}\sum_{t=1}^{n}[O_p(L_{1nG})+\frac{1}{nh_n^d f(X_t)}\sum_{i=1}^{n}K_{it}(r(x_{1i})-r(x_{1t})-(x_{1i}-x_{1t})r^{(1)}(x_{1t}))]^2\epsilon_t^2(1+o_p(1)) \; . \\
= \;& O_p(n^{-1}L_{1nG}^2)+O_p(\frac{1}{n^2}\sum_{t=1}^{n}I_2^2(X_t)\epsilon_t^2) = O_p(n^{-1}L_{1nG}^2)+o_p(n^{-1}(nh_n^{d/2})^{-1}).
\end{aligned}$$

So $\frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m_2^*(X_t))\epsilon_t^* = o_p((nh_{1n}^{d_1/4})^{-1})+O_p(n^{-1/2}h_{1n}^{v_1})+O_p(n^{-1/2}L_{1nG})+o_p(n^{-1}h_n^{-d/4})$, which implies the claim in (ii) with assumption B3.

(iv) $\frac{1}{n}\sum_{t=1}^{n}(m_2^*(X_t)-r(x_{1t}))^2 = O_p(L_{1nG}^2)+O_p(\frac{1}{n}\sum_{t=1}^{n}I_2^2(X_t)) = o_p((nh_n^{d/2})^{-1})$.

$$\begin{aligned}
&\text{(v)} \; -\frac{2}{n}\sum_{t=1}^{n}(\hat{r}(x_{1t})-m^*(X_t))m_1^*(X_t) \\
= \;& -\frac{2}{n}\sum_{t=1}^{n}[\hat{r}(x_{1t})-r(x_{1t})-\frac{1}{nh_n^d f(X_t)}\sum_{i=1}^{n}K_{it}(\hat{r}(x_{1i})-r(x_{1i})+r(x_{1i})-r(x_{1t})-(x_{1i}-x_{1t})r^{(1)}(x_{1t})]m_1^*(X_t) \\
= \;& \{-\frac{2}{n}\sum_{t=1}^{n}[\hat{r}(x_{1t})-r(x_{1t})]\frac{1}{nh_n^d f(X_t)}\sum_{i=1}^{n}K_{it}\epsilon_i^*+\frac{2}{n}\sum_{t=1}^{n}[\frac{1}{nh_n^d f(X_t)}\sum_{i=1}^{n}K_{it}(\hat{r}(x_{1i})-r(x_{1i}))]\frac{1}{nh_n^d f(X_t)}\sum_{i=1}^{n}K_{it}\epsilon_i^* \\
& +\frac{2}{n}\sum_{t=1}^{n}I_2(X_t)\frac{1}{nh_n^d f(X_t)}\sum_{i=1}^{n}K_{it}\epsilon_i^*\}(1+o_p(1)) \\
= \;& o_p((nh_n^{d/2})^{-1}) \text{ with similar arguments by bounding the conditional variance.}
\end{aligned}$$

Results in (i)-(v) above give the claim in (2).

(4) We observe that $\epsilon_{t,0}^* = y_t^*-\hat{r}^*(x_{1t}) = \epsilon_t^*-(\hat{r}^*(x_{1t})-\hat{r}(x_{1t}))$. Since $\hat{r}^*(x_{1t})-\hat{r}(x_{1t}) = r_1^*(x_{1t})+r_2^*(x_{1t})-r(x_{1t})+r(x_{1t})-\hat{r}(x_{1t}) = r_1^*(x_{1t})+O_p(L_{1nG})$ uniformly over $x_{1t}\in G_1$, and recall the definition of $r_1^*(x_{1t})$, we write $\epsilon_{t,0}^* = \epsilon_t^*-\frac{1}{nh_{1n}^{d_1}f_1(x_{1t})}\sum_{i=1}^{n}K_{1it}\epsilon_i^*(1+o_p(1))$. By following the proof in Theorem 5 (5), we obtain the claimed results.

References

Azzalini, A., Bowman, A. W., 1993. On the Use of Nonparametric Regression for Checking Linear Relationships. Journal of the Royal Statistical Society. Series B (Methodological) 55, 549–557.

Azzalini, A., Bowman, A. W., Härdle, W., 1989. On the use of nonparametric regression for model checking. Biometrika 76, 1–11.

Dette, H., 1999. A consistent test for the functional form of a regression based on a difference of variance estimators. Annals of Statistics 27, 1012–1040.

Doksum, K., Samarov, A., 1995. Nonparametric estimation of global functionals and a meaure of the explanatory power of covariates in regression. Annals of Statistics 23, 1443–1473.

Fan, J., 1992. Design-adaptive nonparametric regression. Journal of the American Satistical Association 87, 998–1004.

Fan, J., Zhang, C., Zhang, J., 2001. Generalized Likelihood Ratio Statistics and Wilks Phenomenon. The Annals of Statistics 29 (1), 153–193.

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

Fan, Y., Li, Q., 2000. Consistent model specification tests: kernel-based tests versus Bierens' ICM test. Econometric Theory 16, 1016–1041.

Fan, Y., Li, Q., 2002. A consistent model specification test based on the kernel sum of squares of residuals. Econometric Reviews 21, 337–352.

Fan, Y., Li, Q., Min, I., 2006. A Nonparametric Bootstrap Test of Conditional Distributions. Econometric Theory 22, 587–613.

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

Green, W. H., 2000. Econometric analysis, 4th Edition. Prentice Hall, New Jersey.

Gu, J., Li, D., Liu, D., 2007. Bootstrap non-parametric significance test. Journal of Nonparametric Statistics 19, 215–230.

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

Hart, J. D., 1997. Nonparametric smoothing and lack-of-fit test. Springer, New York.

Hidalgo, J., 1992. A general nonparametric misspecification test. Manuscript, London School of Economics.

Hong, Y., Lee, Y.-J., 2009. A loss function approach to model specification testing and its relative efficiency to the GLR test. Manuscript, Cornell University.

Hsiao, C., Li, Q., Racine, J., 2007. A consistent model specification test with mixed discrete and continuous data. Journal of Econometrics 140, 802–826.

Huang, L.-S., Chen, J., 2008. Analysis of variance, coefficient of determination and F -test for local polynomial regression. The Annals of Statistics 36, 2085–2109.

Lavergne, P., Vuong, Q., 2000. Nonparametric significance testing. Econometric Theory 16, 576–601.

Lavergne, P., Vuong, Q. H., 1996. Nonparametric selection of regressors. Econometrica 64, 207–219.

Lepski, O., Spokoiny, V. G., 1999. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. Bernoulli 5, 333–358.

Li, Q., 1999. Consistent model specification tests for time series econometric models. Journal of Econometrics 92, 101–147.

Li, Q., Racine, J., 2007. Nonparametric econometrics: theory and practice. Princeton University Press, Princeton, NJ.

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

Martins-Filho, C., Yao, F., 2006. A Note on the use of V and U statistics in nonparametric models of regression. Annals of the Institute of Statistical Mathematics 58, 389–406.

Martins-Filho, C., Yao, F., 2007. Nonparametric frontier estimation via local linear regression. Journal of Econometrics 141, 283–319.

Racine, J., 1997. Consistent significance testing for nonparametric regression. Journal of Business and Economic Statistics 15, 369–378.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics 27, 832–837.

Su, L., Ullah, A., 2012. A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. Forthcoming in Econometric Theory.

Ullah, A., 1985. Specification analysis of econometric models. Journal of Quantitative Economics 1, 187–209.

Yao, F., Ullah, A., 2013. Appendix to a nonparametric R-square test for the presence of relevant variables. Economics Department, West Virginia University.
URL `http://community.wvu.edu/ fy006/hp/npr2app.pdf`

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

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