The Canadian Journal of Statistics
Vol. 39, No. 1, 2011, Pages 108–125
La revue canadienne de statistique

A note on testing the regression functions via nonparametric smoothing

Weixing SONG* and Juan DU

Department of Statistics, Manhattan, KS, 66502, USA

Key words and phrases: Consistent test; convolution; kernel estimation; lack-of-fit test; local alternatives

MSC 2010: Primary 62G08; secondary 62G10.

Abstract: The authors present a consistent lack-of-fit test in nonlinear regression models. The proposed procedure possesses some nice properties of Zheng's test such as the consistency, the ability to detect any local alternatives approaching the null at rates slower than the parametric rate. What's more, for a predetermined kernel function, the proposed test is more powerful than Zheng's test and the validity of these findings is confirmed by the simulation studies and a real data example. In addition, the authors find out a close connection between the choices of normal kernel functions and the bandwidths. *The Canadian Journal of Statistics* 39: 108–125; 2011 © 2011 Statistical Society of Canada

Résumé: Les auteurs présentent un test cohérent pour vérifier l'adéquation de modèles de régression non-linéaires. La procédure proposée a quelques des bonnes propriétés du test de Zheng telles quela cohérence et la possibilité de détecter n'importe quelles hypothèses alternatives approchantl'hypothèse nulle à des taux plus lents que le taux paramétrique. De plus, pour un noyau prédéterminé,le test proposé est plus puissant que le test de Zheng et la validité de ces résultats est confirméepar des études de simulation et avec de vraies données. Les auteurs ont trouvé un lien étroit entrele choix des noyaux gaussiens et les largeurs de fenêtre. *La revue canadienne de statistique* 39: 108–125; 2011 © 2011 Société statistique du Canada

1. INTRODUCTION

The relationship between a random variable Y and a random vector X is often investigated via a regression model $Y = \mu(X) + \varepsilon$, where $\mu(X) = E(Y/X)$, ε accounts for the random error. Parametric statistical inference about this model often assumes the regression function $\mu(\cdot)$ has a parametric form. However, if the assumption deviates far from the true one, then the statistical inference results might be misleading or even disastrous. Therefore, it is desirable to develop some diagnostic tools to check if the assumed regression form really fits the data. This is the so called lack-of-fit test or specification test.

There has been growing interest in the lack-of-fit testing procedures. A detailed introduction on the classic lack-of-fit test techniques can be found in the excellent monograph by Hart (1997) and the references therein. Other works include Hausman (1978), Ruud (1984), Newey (1985a,b), Tauchen (1985), White (1982, 1987), and Bierens (1990). But most of the tests are not consistent for some general alternative hypotheses. For example, the reduction method works well when one has a particular, linear alternative hypothesis and the null hypothesis is nested within the alternative. But in an example given on page 126 of Hart (1997), the reduction method essentially has no power. Although some tests are consistent, they usually suffer from the high cost in computation, like Bieren's (1990) test.

* *Author to whom correspondence may be addressed.*
E-mail: weixing@ksu.edu

To deal with the inconsistency problem, various nonparametric methods constructed through smoothing techniques are proposed. Most of these tests are based on certain distances between a nonparametric fit and a parametric fit under the null hypothesis. See Lee (1988), Yatchew (1992), Eubank & Spiegelman (1990), and Härdle & Mammen (1993). Recently Horowitz & Spokoiny (2001) and Guerre & Lavergne (2005) propose two consistent lack-of-fit testing procedures based on nonparametric smoothing technique. Combining the idea of the conditional moment test and the methodology of the nonparametric smoothing, Zheng (1996) proposed a consistent lack-of-fit test for the regression models. He claims that the test is more powerful than some existing ones because of the faster divergence rate to infinity under the alternative. Moreover the computation of the test is much simpler. Recently, Wang & Zhou (2007) applied Zheng's method to check the adequacy of the variance function in a class of heteroscedastic regression models.

To state the motivation of our research presented in this paper, let's briefly introduce the idea sitting behind Zheng's test. Suppose the observations $\{(X_i, Y_i), i = 1, 2, \ldots, n\}$ are obtained from the population (X, Y) such that $E|Y| < \infty$, where X is an $d \times 1$ vector and Y is a scalar. Let $g(x)$ denote the conditional expectation $E(Y|X=x)$. Zheng (1996) considers the following hypothesis:

$$H_0 : P[E(Y|X) = m(X;\theta_0)] = 1 \quad \text{for some } \theta_0 \in \Theta$$

versus

$$H_1 : P[E(Y|X) = m(X;\theta)] < 1 \quad \text{for all } \theta \in \Theta, \tag{1}$$

where $m(x; \theta)$ is a parametric function characterized by the parameter θ, and the parameter space Θ is assumed to be compact. Denote $\varepsilon_i = Y_i - m(X_i;\theta_0)$ and let $f(\cdot)$ be the density function of X, where θ_0 is the true value of θ under H_0. Then under H_0, $\mathrm{E}(\varepsilon_i|X_i) = 0$ and $E(\varepsilon_i E(\varepsilon_i|X_i) f(X_i)) = 0$. While under H_1, $E(\varepsilon_i|X_i) = g(X_i) - m(X_i;\theta_0)$, and

$$E[\varepsilon_i E(\varepsilon_i|X_i) f(X_i)] = E[E^2(\varepsilon_i|X_i) f(X_i)] = E[g(X_i) - m(X_i;\theta_0)]^2 f(X_i) > 0. \tag{2}$$

Zheng's test is based on the quantity

$$n^{-1} \sum_{i=1}^{n} \varepsilon_i E(\varepsilon_i|X_i) f(X_i) \tag{3}$$

which is a sample analogue of $E[\varepsilon_i E(\varepsilon_i|X_i) f(X_i)]$. Replacing $E(\varepsilon_i|X_i)$, $f(X_i)$ with the leave-one-out Nadaraya–Watson kernel estimates, as follows

$$\hat{E}(\varepsilon_i|X_i) = \frac{1}{(n-1)\hat{f}(X_i)} \sum_{j \neq i} \frac{1}{h^d} K_h(i, j) e_j, \quad \hat{f}(X_i) = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{h^d} K_h(i, j), \tag{4}$$

respectively, where $e_i = Y_i - m(X_i;\hat{\theta}_n)$, $i = 1, 2, \ldots, n$, $\hat{\theta}_n$ is any $\sqrt{n}$-consistent estimator of $\theta_0 = \arg\min_{\theta \in \Theta} E[Y - m(X;\theta)]^2$, and $K_h(i, j) = K((X_i - X_j)/h)$. Zheng's test is then constructed from the quantity $Z_n = (n(n-1))^{-1} \sum_{i \neq j} h^{-d} K_h(i, j) e_i e_j$. An interesting question is: why not use the empirical version of the second term in (2) to build the test statistic? An attractive feature of the empirical version of $E(E^2(\varepsilon|X) f(X))$ is that the variance of this empirical version will be less than that of (3) used in Zheng (1996), which is derived from the following fact

$$E(E^4(\varepsilon|X) f^2(X)) \leq EE(\varepsilon^2|X) E^2(\varepsilon|X) f^2(X) = E\varepsilon^2 E^2(\varepsilon|X) f^2(X) \tag{5}$$

by applying Cauchy–Schwarz inequality. So if a new test is constructed based on the standardized sample analogue of the second term in (2), comparing to Zheng's test which uses the standardized sample analogue of the first term in (2) as the test statistic, we will find that these two test statistics might have closed numerators based on the first equality in (2), while the new test statistic has a smaller denominator than Zheng's test statistic. This implies that the new test might be more powerful than Zheng's test. Although that the variance of the population version (5) is smaller than that of (3) does not necessarily imply their empirical counterparts posses the same relationship, in particular, after replacing all unknown quantities with the estimators, but it is intuitively appealing to investigate the actual performance of the new test. Comparing with Zheng's test, the new test statistic is relatively complicated, in particular, the appearance of the kernel estimator of $f(x)$ in the denominator needs some extra conditions to avoid the possible asymptotic negligibility at the boundary points and the possible numeric instability when $f(x)$ is small. In real applications, if we are not sure whether or not these conditions hold for $f(x)$, then special attention should be paid when employing the proposed method. But except that, the new test shares the same advantages as Zheng's test. Another important fact revealed in the current work, to our surprise, is the inherent connection between the selection of smoothing parameter and the choice of kernel functions.

The paper will be organized as follows. Various model assumptions, the testing procedure and the main results will be presented in Section 2; Section 3 considers the consistency and local power of the proposed test; Section 4 contains some simulation studies to illustrate the finite sample behaviour of our test and a read data example is provided in Section 5; All the proofs will be postponed to Appendix Section.

2. TEST STATISTICS

Using the leave-one-out estimators in (4), the sample analogue of $E[(E(\varepsilon_i|X_i))^2 f(X_i)]$ is given by

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{(n-1)h^d}\sum_{j\neq i}K_h(i,j)e_j\right]^2\hat{f}^{-1}(X_i). \tag{6}$$

By expanding the square term, it can be written as

$$\frac{1}{n(n-1)^2h^{2d}}\sum_{i=1}^{n}\left[\sum_{j\neq i}\sum_{k\neq i}K_h(i,j)K_h(i,k)e_je_k\right]\hat{f}^{-1}(X_i). \tag{7}$$

Similar to the leave-one-out technique in (4), we drop all the terms with $k=j$ from the third sum in (7), accordingly, change one $1/(n-1)$ into $1/(n-2)$. That is, we will consider the following statistic

$$S_n=\frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i=1}^{n}\left[\sum_{j\neq i}\sum_{k\neq i,j}K_h(i,j)K_h(i,k)e_je_k\right]\hat{f}^{-1}(X_i).$$

Before we give the asymptotic distribution of S_n under the null hypothesis, various regularity conditions are listed below.

1. C1: The design variable X has a compact support $\mathcal{I}$ and $\min_{x\in\mathcal{I}} f(x)\geq c$, where c is a positive constant.
2. C2: The kernel function K is nonnegative, bounded, continuous and symmetric about 0, such that $\int K(u)du=1$.

3. C3: The bandwidth h is chosen so that $h \to 0$ and $nh^{2d} \to \infty$.
4. C4: The density function $f(x)$ of X and its first-order derivatives are uniformly bounded. $E(Y^4|X = x)$ is continuously differentiable and bounded by a measurable function $b(x)$ such that $Eb^2(X) < \infty$.
5. C5: The parameter space Θ is a compact and convex subset of $\mathbb{R}^p$, $p \geq 1$. $m(x, \theta)$ is a Borel measurable function on $\mathbb{R}^d$ for each θ and a twice continuously differentiable real function on Θ for each $x \in \mathbb{R}^d$. Moreover
$$E\left[\sup_{\theta\in\Theta} |m^2(X;\, \theta)|\right] < \infty,$$
$$E\left[\sup_{\theta\in\Theta} \left\|\frac{\partial m(X;\theta)}{\partial\theta}\frac{\partial m(X;\theta)}{\partial\theta'}\right\|\right] < \infty,$$
$$E\left[\sup_{\theta\in\Theta} \left\|(Y-m(X;\, \theta))^2\frac{\partial m(X;\theta)}{\partial\theta}\frac{\partial m(X;\theta)}{\partial\theta'}\right\|\right] < \infty,$$
$$E\left[\sup_{\theta\in\Theta} \left\|(Y-m(X;\, \theta))^2\frac{\partial^2 m(X;\theta)}{\partial\theta\partial\theta'}\right\|\right] < \infty;$$
6. C6: $E[(Y-m(X;\theta))^2]$ takes a unique minimum at $\theta_0 \in \Theta$ under H_0, where θ_0 is an interior point of Θ;
7. C7: The matrix $E(\partial m(X;\theta_0)/\partial\theta)(\partial m(X;\theta_0)/\partial\theta)'$ is nonsingular.
8. C8: $\tau^2(x) = E(\varepsilon^2|X = x)$ is continuous.

Condition (C1) is a typical restriction that avoids a nonparametric estimator of $f(x)$ from vanishing near the boundary of the design space; Condition (C2) is the same as the assumption (5) in Zheng (1996). Note that the boundedness of K implies $\int K^2(u)\mathrm{d}u < \infty$. Condition (C3) is stronger than the corresponding assumption in Zheng (1996), where $nh^d \to \infty$ is assumed. Conditions (C4)–(C7) are the same as in Zheng (1996), which are essentially the ones used in Bierens (1990, Appendix A) to ensure the consistency and asymptotic normality of nonlinear least squares estimators even in the presence of model misspecification.

In the sequel, we shall assume that for all $x \in \mathcal{I}$, $\hat{f}(x) \geq c$. Otherwise, one can simply replace $\hat{f}(x)$ with $\hat{f}(x) \vee c$, the maximum of $\hat{f}(x)$ and c, all the arguments still apply.

The asymptotic distribution of S_n under the null hypothesis is given in the following theorem,

Theorem 2.1. *Suppose the conditions (C1)–(C8) hold. Then under the null hypothesis,* $nh^{d/2}S_n \Rightarrow N(0, \sigma^2)$, *where the asymptotic variance* σ^2 *has the following form*

$$\sigma^2 = 2\int\left[\int K(u+v)K(v)\mathrm{d}v\right]^2\mathrm{d}u\int[\tau^2(x)]^2 f^2(x)\mathrm{d}x, \tag{8}$$

and $\tau^2(x) = E(\varepsilon^2|X = x)$.

Let $H(u) = \int K(u+v)K(v)\mathrm{d}v$, which is actually the convolution of K. Then σ^2 in (8) can be consistently estimated by

$$\hat{\sigma}^2 = \frac{2}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}H^2\left(\frac{X_i-X_j}{h}\right)e_i^2 e_j^2.$$

Hence, by Theorem 2.1, the test that rejects H_0 whenever $T_n = nh^{d/2}|S_n|/\hat{\sigma} > z_{\alpha/2}$ will be of the asymptotic size α, where z_α is the upper $(1-\alpha)100$th percentile of the standard normal distribution.

The above result is similar to that in Zheng (1996) except for the first integration in σ^2. The integration in Zheng's result is $\int K^2(v)\mathrm{d}v$. Note that H is the convolution of K, by Cauchy–Schwarz inequality, one can easily show that $\int H^2(v)\mathrm{d}v \leq \int K^2(v)\mathrm{d}v$. That is, our test has a smaller asymptotic variance than that of Zheng's (1996) test.

Although the motivation of the current research is to construct a more precise test by modifying Zheng's procedure, it turns out that there are some interesting connections with other lack-of-fit tests. Koul & Ni (2004) proposed a class of minimum distance tests for fitting a parametric regression model to a regression function. The tests are based on the following statistic

$$U_n = \int \left[\frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)e_i\right]^2 \frac{\mathrm{d}G(x)}{\hat{f}_w^2(x)}, \tag{9}$$

where $\hat{f}_w(x)$ is a kernel estimator of the density function f with bandwidth w, possibly different from h. G is a weighting σ-finite measure. If we choose $w = h$, $\mathrm{d}G(x) = \hat{f}_h(x)\mathrm{d}F_n(x)$ in (9), where $F_n(x)$ is the empirical cumulative distribution function of X_i's, then after a slight and obvious modification, U_n is simply (6). Moreover, Remark 5.2 in Koul & Ni (2004) indicates that in some simulation studies, if $\mathrm{d}G(x)$ is chosen to be $\hat{f}_w^2(x)\mathrm{d}x$, then their test outperforms some other tests against the selected alternatives. Notice that the close connection between $\hat{f}^2(x)\mathrm{d}x$ and $\hat{f}(x)\mathrm{d}F_n(x)$, one can expect that our test will work well in the same set up. While our test is not a member of the above class of minimum distance test in that the weighting measure G used in Koul & Ni (2004) is required to possess a continuous Lebesgue density.The proof of Theorem 2.1, which is postponed to Section 5, shows that

$$nh^{d/2}S_n = \frac{1}{(n-1)h^{d/2}}\sum_{j\neq k} H\left(\frac{X_j - X_k}{h}\right)\varepsilon_j\varepsilon_k + o_p(1) := V_n + o_p(1). \tag{10}$$

This also gives an interesting connection between our test and Zheng's test: Our test is asymptotically equivalent to Zheng's test with the kernel function K replaced with the convolution H of K, nevertheless, our test is more powerful. A refree of Zheng's paper pointed out Zheng's test statistic could be viewed as an alternative to an earlier, more complicated (but asymptotically equivalent) statistic. We are not sure whether the complicated statistic he mentioned is the same as ours or not at this point. See the footnote on page 267 of Zheng (1996).

If one wants to construct a test based on V_n in (10) with the random errors ε_i's replaced by the residual e_i's, denoted it as $\hat{V}_n$, that is, we will rejects H_0 whenever

$$R_n = nh^{d/2}|\hat{V}_n|/\hat{\sigma} > z_{\alpha/2}, \tag{11}$$

then the conditions needed for the asymptotic theory can be greatly simplified. For example, (C1) can be removed, and (C3) can be changed to $nh^d \to \infty$.

3. CONSISTENCY AND LOCAL POWER OF THE TEST

For a fixed alternative, a reasonable test should be consistent, that is, the test should have power that tends to 1 as the sample size goes to ∞. In this section, we shall investigate the consistency of the proposed test. The following theorem is obtained,

Theorem 3.1. *Suppose the conditions (C1)–(C8) hold. Then under the alternative hypothesis* H_1 *in* (1), $S_n \to E[g(X) - m(X;\theta_0)]^2 f(X)$ *in probability, and*

$$\hat{\sigma}^2 \to 2\int \left[\int K(u+v)K(v)\mathrm{d}v\right]^2 \mathrm{d}u \int [\tau^2(x) + (g(X) - m(X;\theta_0))]^2 f^2(x)\mathrm{d}x \tag{12}$$

in probability.

The consistency of the test is thus implied by the positiveness of $E[g(X)-m(X;\theta_0)]^2 f(X)$ and the right hand side of (12). Compare with the expression (3.14) in Theorem 2 of Zheng (1996), our result only differs in the denominator, like the null case. This implies our test will be more powerful than Zheng's test for fixed alternatives.

Sometimes, only considering consistency of the test is not enough, since this property does not tell anything about the power when the sample size is small and moderate. In this case, the performance of the test is often investigated by deriving the asymptotic probability that the test rejects the null hypothesis against a sequence of local alternatives given as follows:

$$H_{\text{Loc}} : P[E(Y|X) = m(X, \theta_0) + \delta_n l(X)] = 1, \text{ for some measurable function } l(x), \tag{13}$$

where δ_n is a sequence of real numbers that converges to 0 as $n \to \infty$. If δ_n is chosen to be $(nh^{d/2})^{-1/2}$, then we can obtain the following result,

Theorem 3.2. *Suppose the conditions (C1)–(C8) hold. Then under the local alternative hypothesis H_{Loc},*

$$nh^{d/2} S_n/\hat{\sigma} \Rightarrow N(\mu, 1),$$

where $\mu = El^2(X)f(X)/\sigma$, and σ is defined in (8).

The above result implies that the proposed test has a nontrivial asymptotic power for alternatives in the neighborhood of the null hypothesis. As one referee pointed out, the class of the alternative models considered here is small in that (13) excludes the local alternative models of the form $E(Y|X) = m(X, \theta_0) + \delta_n l_n(X)$, where $l_n(x)$ is a sequence of smooth functions, while the latter enables us to develop tests that have good power against a larger class of local alternative models. See Horowitz & Spokoiny (2001) for more discussion on this issue.

4. NUMERICAL SIMULATION

This section will present two simulation studies. The first simulation uses the test statisticT_n based on Theorem 2.1, and the second simulation uses the asymptotically equivalent test R_n in (11). The significance level is chosen to be 0.05 for all simulations. For each scenario and various sample sizes, we repeated the test procedure 1000 times and the empirical size and power are computed by using $\#\{|T_n| \geq 1.96\}/1000$ and $\#\{|R_n| \geq 1.96\}/1000$, respectively.

Simulation 1: We generate data from the following three models:

1. Model 0: $Y = \theta_0 + \theta_1 X + \varepsilon$,
2. Model 1: $Y = \theta_0 + \theta_1 X + X \sin(X) + \varepsilon$,
3. Model 2: $Y = \theta_0 + \theta_1 X - X \sin(X) + \varepsilon$,

where $X \sim$ Uniform$[-1,1]$, $\varepsilon \sim N(0, 1)$. The null model being fitted is $m(x;\theta) = \theta_0 + \theta_1 x$. The true parameters are chosen to be $\theta_0 = 1$, $\theta_1 = 2$. The kernel function is chosen to be standard normal, and the bandwidth is set to be $n^{-2/5}$. Data from Model 0 are used to study the empirical level, while from models 1–2 are used to study the empirical power of the test. Note that Model 1 is a convex alternative and Model 2 is a concave alternative when $x \in [-1, 1]$. We also did a simulation study when ε has a uniform distribution on $[-1,1]$. Since the simulation results are similar for both cases, only results from normal random errors are reported here.

The simulation result in Table 1 shows that the empirical levels are all less than the nominal levels in all the chosen cases, hence the proposed test is conservative for all chosen sample sizes. This is very common for nonparametric smoothing tests. It has a small power against the alternative models when sample sizes are small, but the power improves with increasing sample size. One might use the bootstrap method to implement the test. Since bootstrap often provides more accurate approximation to the distribution of the test statistic than the asymptotic normal

TABLE 1: Results for simulation 1.

	50	100	200	300	500	800
Model 0	0.001	0.001	0.002	0.009	0.012	0.013
Model 1	0.045	0.097	0.712	0.859	0.975	1.000
Model 2	0.040	0.086	0.692	0.859	0.977	1.000

TABLE 2: $a = 0.5$.

	100	200	300	400	500	600	700
Model 0	0.032	0.049	0.046	0.036	0.054	0.058	0.042
	0.046	0.041	0.049	0.034	0.059	0.060	0.041
Model 1	0.514	0.838	0.960	0.994	0.996	1.000	1.000
	0.766	0.965	0.995	1.000	1.000	1.000	1.000
Model 2	0.159	0.274	0.434	0.530	0.617	0.736	0.814
	0.233	0.458	0.639	0.767	0.846	0.925	0.969
Model 3	0.857	0.993	1.000	1.000	1.000	1.000	1.000
	0.971	1.000	1.000	1.000	1.000	1.000	1.000

theory does when the sample size is small or moderate, see, for example, Härdle & Mammen (1993).

We also conduct two simulation studies when X follows uniform distribution over $[-10, 10]$ or standard normal distribution $N(0, 1)$, although the latter does not satisfies the condition (C1), the possible instability incurred by the small values of the density function seems no significant influence on the performance of the proposed test.

Simulation 2: The following models are used for the simulation:

1. Model 0: $Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \varepsilon$,
2. Model 1: $Y = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_1 X_2 + \varepsilon$,
3. Model 2: $Y = (\theta_0 + \theta_1 X_1 + \theta_2 X_2)^{1/3} + \varepsilon$,
4. Model 3: $Y = (\theta_0 + \theta_1 X_1 + \theta_2 X_2)^{5/3} + \varepsilon$,

where $X \sim$ Normal (0, 1), $\varepsilon \sim N(0, 1)$. The null model being fitted is $m(x; \theta) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$. All parameters are chosen to be 1. The kernel function is chosen to be standard normal and the bandwidth is set to be $an^{-2/5}$. In the simulation, we choose $a = 0.5$, 1, 1.5, and 2 to see the effect of bandwidth on the power. Data from Model 0 are used to study the empirical level, while from models 1–3 are used to study the empirical power of the test. In the simulation, we also choose $X \sim$ Uniform$[-1, 1]$. Since the simulation results are similar for both cases, only results from normal design are reported here. Note that this simulation setup is the same as the one used in Zheng (1996). For comparison, we did some simulation using Zheng's method as well. In Tables 2–5, for each model, the first row is from Zheng's test, and second row is from ours.

The simulation studies show that our test outperforms Zheng's test in all cases. One interesting finding from above simulation is that the finite sample powers are not stable for different choices of a values. Moreover, we can see that the bigger the value of a, the larger the power. The simulation study conducted in Zheng (1996) also shows this trend. A partial explanation for this

TABLE 3: $a = 1$.

	100	200	300	400	500	600	700
Model 0	0.045	0.029	0.042	0.040	0.054	0.054	0.039
	0.043	0.039	0.033	0.040	0.048	0.046	0.043
Model 1	0.904	0.994	1.000	1.000	1.000	1.000	1.000
	0.965	0.999	1.000	1.000	1.000	1.000	1.000
Model 2	0.340	0.663	0.836	0.937	0.961	0.991	0.996
	0.437	0.820	0.943	0.992	0.997	0.999	1.000
Model 3	0.995	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 4: $a = 1.5$.

	100	200	300	400	500	600	700
Model 0	0.039	0.039	0.032	0.042	0.047	0.043	0.044
	0.029	0.029	0.030	0.036	0.038	0.041	0.046
Model 1	0.972	1.000	1.000	1.000	1.000	1.000	1.000
	0.990	1.000	1.000	1.000	1.000	1.000	1.000
Model 2	0.453	0.834	0.950	0.993	0.997	0.999	1.000
	0.549	0.915	0.985	0.997	1.000	1.000	1.000
Model 3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 5: $a = 1.5$.

	100	200	300	400	500	600	700
Model 0	0.030	0.029	0.032	0.038	0.037	0.039	0.044
	0.014	0.015	0.028	0.032	0.024	0.034	0.036
Model 1	0.987	1.000	1.000	1.000	1.000	1.000	1.000
	0.996	1.000	1.000	1.000	1.000	1.000	1.000
Model 2	0.541	0.908	0.983	0.997	0.999	1.000	1.000
	0.569	0.946	0.994	1.000	1.000	1.000	1.000
Model 3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000

phenomenon can be provided based on the theory developed in Section 2. In fact, for a normal kernel, a can be viewed as its standard deviation. Note that the convolution of normal densities is still a normal density, so increasing the value of a, say from $a = 1$ to $a = \sqrt{2}$, is equivalent to replacing a standard normal kernel with the convolution of two standard normal kernels, which, according to our theory, will decrease the asymptotic variance of the test statistic, and leads to

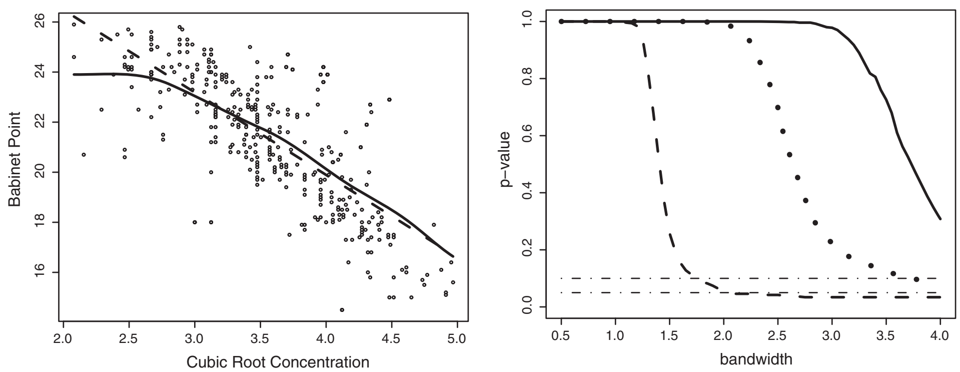

FIGURE 1: Scatter plot and significance trace plot.

a more powerful test. Finally, we point out that the selection of a good or an optimal bandwidth for the proposed test has its own theoretical and practical importance, which deserves a particular study in the future. In fact, the selection of good or optimal bandwidth in nonparametric kernel testing setup has been a focus in many existing literatures. See Horowitz & Spokoiny (2001), Gao & Gilbels (2008), and Zhu (2005) for more details. King, Hart & Wehrly (1991) suggested to use the significance trace plot to circumvent the bandwidth selection dilemma in testing problem. The significance trace plot are obtained by plotting P-values corresponding to several different bandwidths. Based on the significance trace plot, a final decision can be made if all P-values are less than or greater than the prescribed level of significance, otherwise personal judgement will be involved.

5. REAL DATA EXAMPLE

Now we apply the test procedure developed in this paper to the Babinet data set borrowed from Cleveland (1993). The Babinet data set consists of 355 observations from an experiment conducted at the University of Seville, see Bellver (1987), on the scattering sunlight in the atmosphere. There are two variables are of interest. The response variable is the so-called Babinet point which is the scattering angle at which the polarization of sunlight vanishes, and the predictor is the cubic root of particulate concentration in the air. The scatter plot of the Babinet points versus the cubic root of concentration is shown in the left panel of Figure 1, with a fitted linear regression line (dashed line) and a fitted kernel regression curve (solid line, using standard normal kernel, and bandwidth 0.7) being superimposed.

The simple linear regression reports $R^2 = 0.55$, which suggest a rough linear relationship between the Babinet point and the cubic root concentration, but the kernel smoother seems to indicate some curvilinear structure between these two variables. Suppose that we test the hypothesis that the regression function is a straight line using the test procedure developed in the paper. For comparison purpose, the significance trace plots of all three methods are shown in the right panel of Figure 1. To obtain the significance trace plot, for each bandwidth, we use 500 bootstrap samples of size 200 to obtain the true empirical distribution of the test statistic. In fact, to get the true empirical distribution of the test statistics, we use $\hat{e}_i = Y_i - \hat{m}_b(X_i)$, where $\hat{m}_b(x)$ is a kernel estimator of the regression function of the Babinet point versus the cubic root of concentration with standard normal kernel and bandwidth b. Since the simulation results are similar for all the chosen b values from 0.5 to 4, we only report the result for $b = 0.7$. Then the empirical P-values from each method and for each chosen h are obtained by comparing the test statistic calculated from all the data and the bootstrap empirical distribution. The significance trace plot is obtained by plotting the P-values against h values from 0.5 to 4. The solid line is for Zheng's test, the dotted line for the test based on R_n, and the dashed line is from the test based on T_n. It can be seen from the significance trace plot, that for all chosen h values, Zheng's test attempts to accept the hypothesis for both significance levels 0.05 and 0.1, the test based on R_n will accept the hypothesis for a majority of h-values except for very large bandwidths and significance level 0.1. However, the test based on T_n will accept the hypothesis for smaller bandwidths and reject the hypothesis for moderate and large bandwidths, even for significance level 0.01. Given the obvious nonlinear structure shown in the kernel smooth curve, it is evident that the test based on T_n and R_n are more powerful than Zheng's, and among these three, test based on T_n is the most powerful one.

APPENDIX

Proof of Theorem 2.1. *In the following, we will derive the asymptotic distribution of S_n under the null hypothesis. Note that $e_i = Y_i - m(X_i;\hat{\theta}_n) = \varepsilon_i - [m(X_i;\hat{\theta}_n) - m(X_i;\theta_0)] := \varepsilon_i - \Delta m_i$, S_n can be written as the sum of the following three terms*

$$S_{n1} = \frac{1}{n(n-1)(n-2)h^{2d}} \sum_{i=1}^{n} \left[\sum_{j\neq i} \sum_{k\neq i,j} K_h(i,j)K_h(i,k)\varepsilon_j\varepsilon_k \right] \hat{f}^{-1}(X_i),$$

$$S_{n2} = -\frac{2}{n(n-1)(n-2)h^{2d}} \sum_{i=1}^{n} \left[\sum_{j\neq i} \sum_{k\neq i,j} K_h(i,j)K_h(i,k)\varepsilon_j\Delta m_k \right] \hat{f}^{-1}(X_i),$$

$$S_{n3} = \frac{1}{n(n-1)(n-2)h^{2d}} \sum_{i=1}^{n} \left[\sum_{j\neq i} \sum_{k\neq i,j} K_h(i,j)K_h(i,k)\Delta m_j\Delta m_k \right] \hat{f}^{-1}(X_i).$$

We shall write $\tilde{S}_{nj}$ for S_{nj} with $\hat{f}(X_i)$ replaced with $f(X_i)$, for $j = 1, 2, 3$. Note that

$$E[K_h(i,j)K_h(i,k)f^{-1}(X_i)|X_j, X_k] = h^d \int K(u + (X_k - X_j)/h)K(u)\mathrm{d}u := h^d H_h(k,j),$$

then $nh^{d/2}\tilde{S}_{n1}$ is the sum of the following two terms

$$A_{n1} = \frac{1}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{k\neq j} \left[\frac{1}{n-2} \sum_{i\neq j,k} K_h(i,j)K_h(i,k)f^{-1}(X_i) - h^d H_h(j,k) \right] \varepsilon_j\varepsilon_k,$$

$$A_{n2} = \frac{1}{(n-1)h^{d/2}} \sum_{j=1}^{n} \sum_{k\neq j} H_h(j,k)\varepsilon_j\varepsilon_k.$$

Let

$$G_h(j,k) = \frac{1}{n-2} \sum_{i\neq j,k} K_h(i,j)K_h(i,k)f^{-1}(X_i) - h^d H_h(j,k).$$

Note that $G_h(j,k) = G_h(k,j)$, then

$$EA_{n1}^2 = \frac{2}{(n-1)^2h^{3d}}\sum_{j\neq k} EG_h^2(j,k)\varepsilon_j^2\varepsilon_k^2 = \frac{2n}{(n-1)h^{3d}} EG_h^2(1,2)\tau^2(X_1)\tau^2(X_2).$$

While $EG_h^2(1,2)\tau^2(X_1)\tau^2(X_2)$ equals to

$$E\left[\frac{1}{n-2}\sum_{i\neq 1,2} K_h(i,1)K_h(i,2)f^{-1}(X_i) - h^d H_h(1,2)\right]^2 \tau^2(X_1)\tau^2(X_2).$$

Conditioning on (X_1, X_2), $G_h(1,2)$ is a sum of i.i.d. centered random variables. Therefore,

$$EG_h^2(1,2)\tau^2(X_1)\tau^2(X_2)$$

$$\leq \frac{1}{n-2}E[K_h(1,3)K_h(2,3)f^{-1}(X_3) - H_h(1,2)h^d]^2\tau^2(X_1)\tau^2(X_2)$$

$$\leq \frac{1}{n-2}EK_h^2(1,3)K_h^2(2,3)f^{-2}(X_3)\tau^2(X_1)\tau^2(X_2)$$

$$= \frac{h^{2d}}{n-2}\int\int\int K^2(u)K^2(v)\tau^2(x-uh)\tau^2(x-vh)f^{-1}(x)f(x-uh)f(x-vh)\mathrm{d}x\mathrm{d}u\mathrm{d}v$$

$$= O\left(\frac{h^{2d}}{n-2}\right).$$

Therefore, $EA_{n1}^2 = O(1/(nh^d))$, which implies $nh^{d/2}\tilde{S}_{n1} = A_{n2} + o_p(1)$, if we assume $nh^d \to \infty$. Next, we shall show that $nh^{d/2}\tilde{S}_{n2} = o_p(1)$. For this purpose, note that

$$\Delta m_k = (\hat{\theta}_n - \theta_0)'\dot{m}(X_k;\theta_0) + (\hat{\theta}_n - \theta_0)'\ddot{m}(X_k;\bar{\theta}_n)(\hat{\theta}_n - \theta_0),$$

where $\bar{\theta}_n$ is some value between θ_0 and $\hat{\theta}_n$. So, $nh^{d/2}\tilde{S}_{n2}$ can be written as the sum of the following two terms,

$$B_{n1} = \frac{2(\hat{\theta}_n - \theta_0)'}{(n-1)(n-2)h^{3d/2}}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\neq i,j} K_h(i,j)K_h(i,k)f^{-1}(X_i)\varepsilon_j\dot{m}(X_k;\theta_0),$$

$$B_{n2} = \frac{2(\hat{\theta}_n - \theta_0)'}{(n-1)(n-2)h^{3d/2}}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\neq i,j} K_h(i,j)K_h(i,k)f^{-1}(X_i)\varepsilon_j\ddot{m}(X_k;\bar{\theta}_n)(\hat{\theta}_n - \theta_0).$$

By adding and subtracting $h^d H_h(k, j)$, B_{n1} can be written as the sum $B_{n11} + B_{n12}$, where

$$B_{n11} = \frac{2(\hat{\theta}_n - \theta_0)'}{(n-1)h^{3d/2}} \sum_{j \neq k} \left[\frac{1}{n-2} \sum_{i \neq j,k} K_h(i, j) K_h(i, k) f^{-1}(X_i) - h^d H_h(j, k) \right] \varepsilon_j \dot{m}(X_k; \theta_0),$$

$$B_{n12} = \frac{2(\hat{\theta}_n - \theta_0)'}{(n-1)h^{d/2}} \sum_{j \neq k} H_h(j, k) \varepsilon_j \dot{m}(X_k; \theta_0).$$

Let $\dot{m}_l(X_k; \theta_0)$ denote the l-th element of the $p \times 1$ vector $\dot{m}(X_k; \theta_0)$. Note that

$$E \left[\frac{1}{(n-1)h^{3d/2}} \sum_{j \neq k} \left[\frac{1}{n-2} \sum_{i \neq j,k} \frac{K_h(i, j) K_h(i, k)}{f(X_i)} - h^d H_h(j, k) \right] \varepsilon_j \dot{m}_p(X_k; \theta_0) \right]^2$$

$$= \frac{n}{(n-1)^2 h^{3d}} E \left[\sum_{k=2}^{n} \left(\frac{1}{n-2} \sum_{i \neq 1,k} \frac{K_h(i, 1) K_h(i, k)}{f(X_i)} - h^d H_h(1, k) \right) \dot{m}_p(X_k; \theta_0) \right]^2 \tau^2(X_1)$$

$$= \frac{n(n-1)}{(n-1)^2 h^{3d}} E \left(\frac{1}{n-2} \sum_{i \neq 1,2} \frac{K_h(i,1) K_h(i,2)}{f(X_i)} - h^d H_h(1, 2) \right)^2 \dot{m}_p^2(X_2; \theta_0) \tau^2(X_1)$$
$$+ \frac{n(n-1)(n-2)}{(n-1)^2 h^{3d}} E \left(\frac{1}{n-2} \sum_{i \neq 1,2} \frac{K_h(i,1) K_h(i,2)}{f(X_i)} - h^d H_h(1, 2) \right) \dot{m}_p(X_2; \theta_0)$$
$$\left(\frac{1}{n-2} \sum_{i \neq 1,3} \frac{K_h(i,1) K_h(i,3)}{f(X_i)} - h^d H_h(1, 3) \right) \dot{m}_p(X_3; \theta_0) \tau^2(X_1).$$

Similar to show $EG_h^2(1, 2)\tau^2(X_1)\tau^2(X_2) = O(h^{2d}/n)$, one can show that the expectation in the first term on the right is $O(h^{2d}/(n-2))$, therefore, the first term on the right is of $O(1/(nh^d))$; while the second term on the right, after a lengthy but trivial argument, is $O(1)$. Hence $B_{n11} = O_p(1/\sqrt{n}) O_p(1) = o_p(1)$.

Now, let's consider B_{n12}. According to Lemma 3.3b in Zheng (1996)

$$\frac{1}{n(n-1)h^d} \sum_{j=1}^{n} \sum_{k \neq j} H \left(\frac{X_k - X_j}{h} \right) \varepsilon_j \dot{m}(X_k; \theta_0) = O_p \left(\frac{1}{\sqrt{n}} \right).$$

Therefore,

$$B_{n12} = O_p \left(\frac{1}{\sqrt{n}} \right) O_p(\sqrt{n}) h^d \frac{1}{h^{d/2}} = O_p(h^{d/2}) = o_p(1).$$

Hence, $B_{n1} = o_p(1)$. To show that $B_{n2} = o_p(1)$, note that $|B_{n2}|$ is bounded above by

$$\|\hat{\theta}_n - \theta_0\|^2 \left\| \frac{2}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{k \neq j} \left[\frac{1}{n-2} \sum_{i \neq j,k} \frac{K_h(i, j) K_h(i, k)}{f(X_i)} \right] \varepsilon_j \ddot{m}(X_k; \bar{\theta}_n) \right\|.$$

The expectation of the second term is further bounded above by

$$\frac{2n(n-1)}{(n-1)h^{3d/2}} E\left[\frac{K_h(1,2)K_h(1,3)}{f(X_1)} E(|\varepsilon_2||X_2)\|\dot{m}(X_3;\bar{\theta}_n)\|\right] = \frac{2n}{h^{3d/2}} O(h^{2d}) = O(nh^{d/2}).$$

Therefore,

$$B_{n2} = O_p(1/n)O_p(nh^{d/2}) = O_p(h^{d/2}) = o_p(1).$$

The above results, then, imply $nh^{d/2}\tilde{S}_{n2} = o_p(1)$.

The proof of $nh^{d/2}\tilde{S}_{n3} = o_p(1)$ is similar to that of showing $B_{n2} = o_p(1)$. In fact, by Taylor expansion, $nh^{d/2}\tilde{S}_{n3}$ can be written as

$$\frac{1}{(n-1)h^{3d/2}} \sum_{j=1}^{n}\sum_{k\neq j}\left[\frac{(\hat{\theta}_n-\theta_0)'}{n-2}\sum_{i\neq j,k}\frac{K_h(i,j)K_h(i,k)}{f(X_i)}\dot{m}(X_j;\bar{\theta}_n)\dot{m}'(X_j;\bar{\theta}_n)\right](\hat{\theta}_n-\theta_0),$$

where $\bar{\theta}_n$ is some value between θ_0 and $\hat{\theta}_n$. An expectation argument, together with the condition imposed on the first derivative of m, readily leads to the desired result.

Note that the above results are obtained by replacing $\hat{f}(x)$ with $f(x)$. Now we need to take care of this modification. Denote the following term as C_n

$$\frac{nh^{d/2}}{n(n-1)(n-2)h^{2d}} \sum_{i\neq j\neq k}\frac{K_h(i,j)K_h(i,k)}{f(X_i)}\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right]\varepsilon_j\varepsilon_k = \frac{h^{d/2}}{n-1}\sum_{j\neq k} M_n(X_j,X_k)\varepsilon_j\varepsilon_k,$$

where

$$M_n(X_j,X_k) = \frac{1}{n-2}\sum_{i\neq j,k}\frac{K_h(i,j)K_h(i,k)}{h^{2d}f(X_i)}\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right].$$

The symmetry of $M_n(X_j, X_k)$ in its arguments implies

$$EC_n^2 = \frac{4h^d}{(n-1)^2}\sum_{j\neq k} EM_n^2(X_j,X_k)\tau^2(X_j)\tau^2(X_k) = O(h^d)EM_n^2(X_1,X_2)\tau^2(X_1)\tau^2(X_2).$$

One can show that

$$EM_n^2(X_1,X_2)\tau^2(X_1)\tau^2(X_2) \leq E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{f(X_3)}{\hat{f}(X_3)}-1\right]^2\tau^2(X_1)\tau^2(X_2). \quad (14)$$

From condition C1, one can see that the last expectation has the same order as

$$E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}[f(X_3)-\hat{f}(X_3)]^2\tau^2(X_1)\tau^2(X_2) \quad (15)$$

which has the order of $O(1/n^2h^{4d}) + O(1) + O(1/nh^{3d})$. In fact, from

$$\begin{aligned}\hat{f}(X_3)-f(X_3) = \quad & \frac{1}{nh^d}\sum_{j=4}^{n}K\left(\frac{X_j-X_3}{h}\right)-\frac{1}{h^d}E\left[K\left(\frac{X_4-X_3}{h}\right)|X_3\right] \\ & +\frac{1}{h^d}E\left[K\left(\frac{X_4-X_3}{h}\right)|X_3\right]-f(X_3) \\ & +\frac{1}{nh^d}K\left(\frac{X_1-X_3}{h}\right)+\frac{1}{nh^d}K\left(\frac{X_2-X_3}{h}\right)+\frac{1}{nh^d}K(0),\end{aligned}$$

(15) can be bounded above by the sum of the following three terms

$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{nh^d}\sum_{j=4}^{n}K\left(\frac{X_j-X_3}{h}\right)-\frac{1}{h^d}E\left[K\left(\frac{X_4-X_3}{h}\right)|X_3\right]\right]\tau^2(X_1)\tau^2(X_3),$$

$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{h^d}E\left[K\left(\frac{X_4-X_3}{h}\right)|X_3\right]-f(X_3)\right]\tau^2(X_1)\tau^2(X_3),$$

$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{nh^d}K\left(\frac{X_1-X_3}{h}\right)+\frac{1}{nh^d}K\left(\frac{X_2-X_3}{h}\right)+\frac{1}{nh^d}K(0)\right]\tau^2(X_1)\tau^2(X_3).$$

By changing variables when calculating above expectations, one can show that the first term has the order $O(1/nh^{3d})$, the second $O(1)$, and the third $O(1/n^2h^{4d})$.Therefore,

$$C_n^2=O\left(\frac{1}{n^2h^{3d}}\right)+O_p(h^d)+O_p\left(\frac{1}{nh^{2d}}\right).$$

Condition (C3) implies $C_n=o_p(1)$.

Now let us consider

$$D_n=\frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j\neq k}\frac{K_h(i,j)K_h(i,k)}{f(X_i)}\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right]\Delta m_k\varepsilon_j.$$

Using the notation $M_n(X_j,X_k)$, we have $nh^{d/2}D_n=(h^{d/2})/(n-1)\sum_{j\neq k}M_n(X_j,X_k)\Delta m_k\varepsilon_j$. Since

$$\Delta m_k=(\hat{\theta}_n-\theta_0)'\dot{m}(X_k,\theta_0)+(\hat{\theta}_n-\bar{\theta}_n)'\ddot{m}(X_k,\theta_0)(\hat{\theta}_n-\theta_0),$$

therefore, $nh^{d/2}D_n$ can be written as the sum of two terms

$$D_{n1}=(\hat{\theta}_n-\theta_0)'\frac{h^{d/2}}{n-1}\sum_{j\neq k}M_n(X_j,X_k)\dot{m}(X_k,\theta_0)\varepsilon_j=(\hat{\theta}_n-\theta_0)'\tilde{D}_{n1}$$

$$D_{n2}=(\hat{\theta}_n-\theta_0)'\frac{h^{d/2}}{n-1}\sum_{j\neq k}M_n(X_j,X_k)\ddot{m}(X_k,\bar{\theta}_n)\varepsilon_j(\hat{\theta}_n-\theta_0)=(\hat{\theta}_n-\theta_0)'\tilde{D}_{n2}(\hat{\theta}_n-\theta_0).$$

Notice that

$$\begin{aligned}E\tilde{D}_{n1}^2 &= \tfrac{h^d}{(n-1)^2}E\left[\sum_{j=1}^{n}\left(\sum_{k\neq j}M_n(X_j,X_k)\dot{m}(X_k,\theta_0)\right)\varepsilon_j\right]^2\\ &= nh^dE\left[\tfrac{1}{n-1}\sum_{k\neq 1}M_n(X_1,X_k)\dot{m}(X_k,\theta_0)\right]^2\tau^2(X_1)\\ &\leq \tfrac{nh^d}{n-1}\sum_{k\neq 1}EM_n^2(X_1,X_k)\dot{m}^2(X_k,\theta_0)\tau^2(X_1)=nh^dEM_n^2(X_1,X_2)\dot{m}^2(X_2,\theta_0)\tau^2(X_1).\end{aligned}$$

Similar to the arguments in (14), one can show that

$$h^d E M_n^2(X_1, X_2)\dot{m}^2(X_2, \theta_0) = o_p(1).$$

Therefore, the square-root consistency of $\hat{\theta}_n$ implies $D_{n1} = o_p(1)$. From the boundedness of $\ddot{m}(x, \theta)$, we have

$$E\tilde{D}_{n2}^2 = E\left[\frac{h^{d/2}}{n-1}\sum_{j\neq k} M_n(X_j, X_k)\ddot{m}(X_k, \bar{\theta}_n)\varepsilon_j\right]^2 \leq O(n^2h^d)EM_n^2(X_1, X_2)\tau^2(X_1).$$

Similar to the proof of (14), one can show that $EM_n^2(X_1, X_2)\tau^2(X_1)$ has the order $O(1/n^2h^{4d}) + O(1) + O(1/nh^{3d})$. Then from the square-root consistency of $\hat{\theta}_n$, we have

$$D_{n2} = O_p(1/n)O_p(n)\sqrt{O\left(\frac{1}{n^2h^{3d}}\right) + O_p(h^d) + O_p\left(\frac{1}{nh^{2d}}\right)} = o_p(1).$$

Similarly, one can show that

$$\frac{nh^{d/2}}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j\neq k}\frac{K_h(i, j)K_h(i, k)}{f(X_i)}\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right]\Delta m_k\Delta m_j = o_p(1).$$

In summary, we finally get

$$nh^{d/2}S_n = \frac{1}{(n-1)h^{d/2}}\sum_{j\neq k} H\left(\frac{X_j-X_k}{h}\right)\varepsilon_j\varepsilon_k + o_p(1)$$

From Lemma 3.3 a in Zheng (1996), we get $nh^{d/2}S_n \Rightarrow N(0, \sigma^2)$. where σ^2 is defined in (8).

Proof of Theorem 3.1. *Under H_a, we can rewrite S_n as the sum of the following four terms*

$$W_n = \frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j\neq k}\frac{K_h(i, j)K_h(i, k)}{\hat{f}(X_i)}\varepsilon_i\varepsilon_k,$$

$$R_{n1} = \frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j\neq k}\frac{K_h(i, j)K_h(i, k)}{\hat{f}(X_i)}\varepsilon_i[g(X_k)-m(X_k, \hat{\theta}_n)],$$

$$R_{n2} = \frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j\neq k}\frac{K_h(i, j)K_h(i, k)}{\hat{f}(X_i)}[g(X_j)-m(X_j, \hat{\theta}_n)]\varepsilon_k,$$

$$R_{n3} = \frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j\neq k}\frac{K_h(i, j)K_h(i, k)}{\hat{f}(X_i)}[g(X_j)-m(X_j, \hat{\theta}_n)][g(X_k)-m(X_k, \hat{\theta}_n)].$$

From the discussion in the null case, $W_n = o_p(1)$.

Adding and subtracting $m(X_k, \theta_0)$ from $g(X_k)$, R_{n1} can be written as

$$R_{n1} = \frac{1}{n(n-1)(n-2)h^{2d}} \sum_{i\neq j\neq k} \frac{K_h(i,j)K_h(i,k)}{\hat{f}(X_i)} \varepsilon_i [g(X_k) - m(X_k, \theta_0)] + \frac{1}{n(n-1)(n-2)h^{2d}} \sum_{i\neq j\neq k} \frac{K_h(i,j)K_h(i,k)}{\hat{f}(X_i)} \varepsilon_i [m(X_k, \theta_0) - m(X_k, \hat{\theta}_n)].$$

Note that the second term on the right hand side is same as S_{n2} in the previous section, except for some constant, so it is $O_p(1)$. As before, rewrite the first term on the right hand side as

$$\frac{1}{n(n-1)(n-2)h^{2d}} \sum_{i\neq j\neq k} \frac{K_h(i,j)K_h(i,k)}{f(X_i)} \varepsilon_i [g(X_k) - m(X_k, \theta_0)] + \frac{1}{n(n-1)(n-2)h^{2d}} \sum_{i\neq j\neq k} \frac{K_h(i,j)K_h(i,k)}{f(X_i)} \left[\frac{f(X_i)}{\hat{f}(X_i)} - 1\right] \varepsilon_i [g(X_k) - m(X_k, \theta_0)].$$

By investigating the second moments, one can show that the above two terms are all of the order $O_p(1)$. Similarly, one can show that $R_{n2} = o_p(1)$.

Finally, let's look at R_{n3}. Adding and subtracting $m(X_j, \theta_0)$, $m(X_k, \theta_0)$ from $m(X_j, \hat{\theta}_n)$, $m(X_k, \hat{\theta}_n)$, respectively, one can rewrite R_{n3} as the sum of the following four terms

$$R_{n31} = c_n \sum_{i\neq j\neq k} \frac{K_h(i,j)K_h(i,k)}{\hat{f}(X_i)} [g(X_j) - m(X_j, \theta_0)][g(X_k) - m(X_k, \theta_0)],$$

$$R_{n32} = c_n \sum_{i\neq j\neq k} \frac{K_h(i,j)K_h(i,k)}{\hat{f}(X_i)} [g(X_j) - m(X_j, \theta_0)][m(X_k, \theta_0) - m(X_k, \hat{\theta}_n)],$$

$$R_{n33} = c_n \sum_{i\neq j\neq k} \frac{K_h(i,j)K_h(i,k)}{\hat{f}(X_i)} [m(X_j, \theta_0) - m(X_j, \hat{\theta}_n)][g(X_k) - m(X_k, \theta_0)],$$

$$R_{n34} = c_n \sum_{i\neq j\neq k} \frac{K_h(i,j)K_h(i,k)}{\hat{f}(X_i)} [m(X_j, \theta_0) - m(X_j, \hat{\theta}_n)][m(X_k, \theta_0) - m(X_k, \hat{\theta}_n)],$$

where $c_n = 1/[n(n-1)(n-2)h^{2d}]$. Replacing $\hat{f}(X_i)$ with $f(X_i)$ in R_{n31} and denoting the resulting quantity as $\tilde{R}_{n31}$, one has

$$\tilde{R}_{n31} = \frac{1}{n(n-1)} \sum_{j\neq k} H_n(X_j, X_k) L(X_j) L(X_k),$$

where

$$H_n(X_j, X_k) = \frac{1}{n-2} \sum_{i\neq j,k} \frac{K_h(i,j)K_h(i,k)}{h^{2d} f(X_i)}, \quad L(X) = g(X) - m(X, \theta_0).$$

Adding and subtracting $E(K_h(i,j)K_h(i,k)/h^{2d}f(X_i)|X_j,X_k)$ from $H_n(X_j,X_k)$, $\tilde{R}_{n31}$ can be further written as the sum $\tilde{Q}_{n1}+\tilde{Q}_{n2}$, where

$$\tilde{Q}_{n1}=\frac{1}{n(n-1)}\sum_{j\neq k}\left[H_n(X_j,X_k)-E\left(\frac{K_h(i,j)K_h(i,k)}{h^{2d}f(X_i)}|X_j,X_k\right)\right]L(X_j)L(X_k),$$

$$\tilde{Q}_{n2}=\frac{1}{n(n-1)}\sum_{j\neq k}E\left(\frac{K_h(i,j)K_h(i,k)}{h^{2d}f(X_i)}|X_j,X_k\right)L(X_j)L(X_k).$$

Note that

$$E\left(\frac{K_h(i,j)K_h(i,k)}{h^{2d}f(X_i)}\middle|X_j,X_k\right)=\frac{1}{h^d}H\left(\frac{X_j-X_k}{h}\right),$$

therefore, similar to the proof of Lemma 3.4 in Zheng (1996), one can show that

$$\tilde{Q}_{n2}\to E(L(X))^2f(X)=E[g(X)-m(X,\theta_0)]^2f(X)$$

in probability, and $\tilde{Q}_{n1}=o_p(1)$. Furthermore, by the $\sqrt{n}$-consistency of $\hat{\theta}_n$ to θ_0, we can also show that $R_{n3j}=o_p(1)$ for $j=2,3,4$. Therefore, $R_{n3}=E[g(X)-m(X,\theta_0)]^2f(X)+o_p(1)$, which in turn implies $S_n=E[g(X)-m(X,\theta_0)]^2f(X)+o_p(1)$.

Proof of Theorem 3.2. *Since most arguments in the proof are similar to those in the proofs of Theorem 2.1 and 3.1, the details are omitted here for the sake of brevity.*

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the editors and two referees for their helpful comments which improved the presentation of the paper.

BIBLIOGRAPHY

C. Bellver (1987). Influence of particulate pollution on the positions of neutral points in the sky in Seville (Spain). *Atomspheric Environment*, 21, 699–702.

H. J. Bierens (1990). A consistent conditional moment test for functional form. *Econometrica*, 58, 1443–1458.

W. S. Cleveland (1993). "*Visualizing Data,*" Hobart Press, Summit, NJ.

R. L. Eubank & C. H. Spiegelman (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *JASA*, 85, 387–392.

J. Gao & I. Gilbels (2008). Bandwidth selection in nonparametric kernel testing. *JASA*, 103, 1584–1594.

E. Guerre & P. Lavergne (2005). Data-driven rate-optimal specification testing in regression models, *Annals of Statistics*, 33, 840–870.

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

J. D. Hart (1997). "*Nonparametric Smoothing and Lack-of-Fit Tests,*" Springer–Verlag, New York.

J. A. Hausman (1978). Specification testing in econometrics. *Econometrica*, 46, 1251–1271.

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

E. King, J. D. Hart, & T. E. Wehrly (1991). Testing the equality of two regression curves using linear smoothers. *Statistics Probability Letters*, 12, 239–247.

H. L. Koul & P. Ni (2004). Minimum distance regression model checking. *Journal of Statistical Planning Inference*, 119, 109–141.

B. J. Lee (1988). A nonparametric model specification test using a kernel regression method. PhD dissertation, University of Wisconsin, Madison, WI.

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

W. K. Newey (1985b). Generalized method of moments specification testing. *Journal of Econometrics*, 29, 229–256.

P. A. Ruud (1984). Tests of specification in econometrics. *Econometrics Reviews*, 3, 211–242.

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

L. Wang & X. H. Zhou (2007). Assessing the adequacy of variance function in heteroscedastic regression models. *Biometrics*, 63, 1218–1225.

H. White (1982). Maximum likelihood estimation of misspecified models. *Econometrics*, 50, 1–25.

H. White (1987). Specification Testing in Dynamic Models. "*Advances in Econometrics,*" T Bewley, editor. Cambridge University Press, New York.

A. J. Yatchew (1992). Nonparametric regression tests based on least square. *Econometric Theory*, 8, 435–451.

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

L. X. Zhu (2005). "*Nonparametric Monte Carlo Tests and Their Applications,*" Springer, New York.

Received 14 June 2010
Accepted 9 November 2010