The Korean Communications
in Statistics Vol. 8, No. 1, 2001
pp. 1-14

On the Goodness-of-fit Test in Regression Using the Difference Between Nonparametric and Parametric Fits.

Chang Kon Hong[1] and Jae Seon Joo[2]

Abstract

This paper discusses choosing the weight function of the Härdle and Mammen statistic in nonparametric goodness-of-fit test for regression curve. For this purpose, we modify the Härdle and Mammen statistic and derive its asymptotic distribution. Some results on the test statistic from the wild bootstrapped sample are also obtained. Through Monte Carlo experiment, we check the validity of these results. Finally, we study the powers of the test and compare with those of Härdle and Mammen test through the simulation.

Keywords : Goodness-of fit test, Härdle and Mammen test, Power comparison, Wild bootstrap

1. Introduction

When one fits a parametric regression model, it is important to check the appropriateness of the postulated model. A standard parametric approach is the general linear test approach, which is F-test. This parametric approach is efficient in detecting the lack-of-fit in certain specified directions, but inconsistent against the other alternatives. To overcome this difficulty, nonparametric regression approach has been considered by many authors.

Suppose that we have independent and identically distributed($i.i.d.$) n observations $(X_1, Y_1), \cdots, (X_n, Y_n)$, which follow the regression model,

$$Y_j = m(X_j) + \varepsilon_j, \quad j = 1, \cdots, n ,$$

where the regression function m is defined as $m(x) = E(Y_i \mid X_i = x)$, $\varepsilon_j^{'}s$ are error terms such that $E(\varepsilon_j \mid X_j) = 0$ and $\varepsilon_j^{'}$s are conditionally independent given $X_1, \cdots, X_n$. Note that

1) Associate Professor, Department of Statistics, Pusan National University, Pusan 609-735, Korea
E-mail : ckhong@hyowon.cc.pusan.ac.kr
2) Department of Statistics, Pusan National University, Pusan 609-735, Korea

we do not necessarily assume that ε_i's are conditionally identically distributed. So the above model contains the case of conditional heteroscedasticity. We wish to test that $H_0 : m \in \{m_\theta : \theta \in \Omega\}$ against $H_1 : m$ is a smooth function. Here, $m_\theta(x) = \sum_{i=1}^{k} \theta_i g_i(x)$ with θ_i's unknown parameters and g_i's known functions of x.

Nonparametric tests usually use the residuals obtained by fitting the null (parametric) model and derive the test statistic by regressing the residuals nonparametrically. Yanagimoto and Yanagimoto(1987), Cox et al.(1988) and Buckley(1991) proposed nonparametric test statistics for random smooth function. In the case of fixed smooth regression function, Eubank and Hart(1992) used a data-driven smoothing parameter as the test statistic based on series estimator. Härdle and Mammen(1993) suggested the squared L^2-distance between the parametric null model fit and the nonparametric fit as the test statistic. Other related works are those of Härdle and Marron(1988), Munson and Jernigan(1989), Eubank and Spiegelman(1990), and Eubank and LaRiccia(1992).

In this paper, we will basically consider the method of Härdle and Mammen(1993). Härdle and Mammen(1993) used the squared (weighted) L^2-distance between the parametric fit and the nonparametric kernel type fit as the test statistic. But, they didn't mention about the choice of the weight function. In their simulation they used the constant weight. It is intuitive to choose the weight function in such a way that more weight should be given to the values of x which have high probability (density). This is achieved if we use the design density $f(x)$ as the weight function. However, $f(x)$ is usually unknown and must be estimated from the data. Another way to achieve the idea of 'more weight to the high probability density points x' is to use the average squared distance between the parametric and the nonparametric fits at the design points $X_1, \cdots, X_n$.

In section 2, We show that this modified statistic has the same asymptotic property as the Härdle and Mammen test statistic (in the following, on we will denote this by H-M statistic) with $f(x)$ as the weight function. We also obtain the asymptotic results about the wild bootstrapped critical values, which are similar to those in Härdle and Mammen(1993). In section 3, some simulation results are given. Finally, we remarks some conclusions in section 4.

2. Modified H-M test statistic and its asymptotic distribution

Let $m_{\hat{\theta}}$ be the parametric regression estimator under H_0 and $\widehat{m_h}$ be a kernel estimator of m with bandwidth h and kernel K. Here, we consider the Nadaraya-Watson kernel estimator $\widehat{m}_h$, that is,

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x-X_i)Y_i}{\sum_{i=1}^{n} K_h(x-X_i)},$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$. The question is whether the difference between $m_{\widehat{\theta}}(\cdot)$ and $\widehat{m}_h$ can be explained as the fluctuations within the null (parametric) model. Härdle and Mammen(1993) measure this difference by

$$n\sqrt{h}\int(\widehat{m}_h(x) - \mathbf{K}_{h,n} m_{\widehat{\theta}}(x))^2 w(x)dx ,$$

where $\mathbf{K}_{h,n} q(\cdot)$ is the (random) smoothing operator

$$\mathbf{K}_{h,n} q(x) = \frac{\sum_{i=1}^{n} K_h(x-X_i)q(X_i)}{\sum_{i=1}^{n} K_h(x-X_i)}$$

and $w(\cdot)$ is the weight function. Instead of using $m_{\widehat{\theta}}(\cdot)$, they used $\mathbf{K}_{h,n} m_{\widehat{\theta}}(\cdot)$. This is because $E(\widehat{m}_h(\cdot)|X_1,\cdots,X_n) = \mathbf{K}_{h,n} m(\cdot)$. As mentioned in the introduction, they didn't consider the problem of choosing the weight function. The intuitive choice of $w(\cdot)$ is the design density $f(\cdot)$. This choice of the weight function is intuitively appealing since with this weight function high probability density points have more weights in measuring the distance. However, $f(\cdot)$ is usually unknown. Another way to give high weight to high probability density points is to modify the measure of distance. Let

$$T_n = \sqrt{h}\sum_{i=1}^{n}(\widehat{m}_h(X_i) - \mathbf{K}_{h,n} m_{\widehat{\theta}}(X_i))^2.$$

T_n may serve as a test statistic for testing $H_0 : m(\cdot) \in \{m_\theta(\cdot) = \theta' g(\cdot) ; \theta \in R^k\}$, where $\theta = (\theta_1, \cdots, \theta_k)'$ and $g(\cdot) = (g_1(\cdot), \cdots, g_k(\cdot))'$.

To derive the asymptotic power of the test, we assume that the alternative regression function is $m(x) = m_{\theta_0}(x) + c_n \Delta_n(x)$ for certain sequences c_n and Δ_n. Here θ_0 and Δ_n may be chosen as

$$\theta_0 = \underset{\theta}{\arg\min} \int (m(x) - m_\theta(x))^2 f(x)dx$$

and

$$\Delta_n(\cdot) = \frac{1}{c_n}(m(\cdot) - m_{\theta_0}(\cdot))$$

Note that, with these choice of θ_0 and Δ_n, , Δ_n is orthogonal to $\{m_\theta ; \theta \in \Omega\}$ in the sense that

$$\int \Delta_n(x) g_i(x) f(x)\, dx = 0, \qquad i = 1, \cdots, k.$$

Under H_0, the least squares estimator $\hat{\theta}$ of θ is given by

$$\hat{\theta} = \underset{\theta}{\arg\min} \sum_{i=1}^{n} (Y_i - m_\theta(x))^2 .$$

Using the orthogonality of Δ_n to $\{m_\theta \,;\, \theta \in \Omega\}$, it can be shown that

$$\hat{\theta} = \theta_0 + \frac{1}{n} \sum_{i=1}^{n} u(X_i)\varepsilon_i + O_p\left(\frac{c_n}{\sqrt{n}}\right)$$

and

$$m_{\hat{\theta}}(\cdot) - m_{\theta_0}(\cdot) = \frac{1}{n}\left(\sum_{i=1}^{n} u(X_i)\varepsilon_i\right)^T g(\cdot) + \left(O_p\left(\frac{c_n}{\sqrt{n}}\right)\right)^T g(\cdot),$$

where $u(\cdot) = \left(\int g(x) g(x)^T f(x)\, dx\right)^{-1} g(\cdot)$.

For the derivation of the asymptotic properties, some assumptions are needed on the stochastic nature of the observations, the parametric estimator of the regression function, the kernel K and the bandwidth h. We assume the same assumptions as in Härdle and Mammen(1993). For the details of these assumptions, see Härdle and Mammen(1993).

Below, the Mallow's distance $d(\mu, \nu)$ between the probability measures μ and ν is given by

$$d(\mu, \nu) = \inf_{X, Y} \{ E|X - Y|^2 \wedge 1 \,;\, \pounds(X) = \mu,\ \pounds(Y) = \nu \},$$

where $\pounds(T)$ is the law of T. It is well known that the convergence in this metric is equivalent to the weak convergence.

Theorem 1. Under certain conditions , we have

$$d\left(\pounds(T_n),\ N\left(b_h + \int (K_h * \Delta_n(x))^2 f(x)\, dx,\ V\right)\right) \to 0 ,$$

where

$$b_h = h^{-1/2} K * K(0) \int \sigma^2(x)\, dx,$$

$$V = 2h \int \sigma^2(x) \sigma^2(y) (K_h * K_h(x - y))^2 dx\, dy.$$

Here $*$ is the convolution operator.

Proof of Theorem 1 is given in the Appendix.

As was pointed out in Härdle and Mammen(1993), the convergence rate is so slow that it is more appropriate not to use the above asymptotic result to find the critical values. The suggested alternative way is the bootstrap method. Three possible ways of bootstrapping are

(1) naive resampling method,

(2) the adjusted residuals bootstrap, and

(3) the wild bootstrap.

Let T_n^* be the statistic T_n computed from the bootstrapped data. For the first two bootstrap methods, the conditional distribution $£^*(T_n^*)=£(T_n^* \mid X_i, Y_i, i=1,\cdots,n)$ has different asymptotic distribution from $£(T_n)$ even under the null hypothesis (see Härdle and Mammen (1993)). Therefore, we concentrate on the third method, wild bootstrap. The idea of wild bootstrap lies in that the bootstrapped data $\{X_i^*, Y_i^*\}_{i=1,\cdots,n}$ is constructed so that

$$E^*(Y_i^* \mid X_i^*)=m_{\hat{\theta}}(X_i^*)$$

(for references, see Wu(1986), Beran(1986), Liu(1988) and Mammen(1993)). Let $T^{*,W}$ be the T_n obtained from this bootstrapped data. Generate B bootstrap data sets and calculate $T^{*,W}$'s. The empirical $(1-\alpha)$-th quantile $\hat{t}_\alpha^W$ gives the approximate $(1-\alpha)$-th quantile of $£^*(T^{*,W})$. Reject H_0 if $T_n > \hat{t}_\alpha^W$. The following theorem shows that the test procedure really works.

Theorem 2. Let $\hat{\theta}^*$ be the parametric estimator based on the bootstrap sample. Then under the same conditions as in Theorem 1 and the following assumption

(P1) $m_{\theta^*}(\cdot)-m_{\hat{\theta}}(\cdot)=\frac{1}{n}\sum_{i=1}^n g(x)^T u(X_i)\varepsilon_i^*+o_p((n\log n)^{-\frac{1}{2}})$

it holds that

$$d(£^*(T^{*,W}), N(b_h, V)) \to 0 ,$$

where b_h and V are the same as defined in Theorem 1.

Condition (P1) is fulfilled under standard regularity conditions. The proof of Theorem 2 is similar to that of Theorem 1. The conditions of De Jong(1987) that $\sup_i \varepsilon_i^2 = O_p(\log n)$ and $\mathrm{E}\,|\varepsilon_i|^8$ is bounded (uniformly in i and n) can be checked using the condition (A5).

3. Simulation results

We check the validity of the asymptotic results through Monte Carlo experiment. For this purpose, we investigate four different cases: in the following $m_\theta(x)$ is the parametric null model and $m(x)$ is the true regression curve.

i) $m_\theta(x)$ and $m(x)$ are linear, $\sigma(x)=\sigma$

ii) $m_\theta(x)$ and $m(x)$ are linear, $\sigma(x)=(1+x^2)\sigma$

iii) $m_\theta(x)$ and $m(x)$ are quadratic, $\sigma(x)=\sigma$

iv) $m_\theta(x)$ and $m(x)$ are quadratic, $\sigma(x)=(1+x^2)\sigma$.

We generate X_i, $i=1,\cdots,100$ from uniform(0,1) and ε_i, $i=1,\cdots,100$ from $N(0,\sigma^2(x))$ and let $Y_i=m(X_i)+\varepsilon_i$. We use the quartic kernel function $K(u)=\frac{15}{16}(1-u^2)^2 I(|u|\le 1)$. $M=1000$ samples are drawn for each case. One out of these 1000 samples is randomly chosen and the wild bootstrap resampling is performed $B=100$ times from this sample. As was done in Härdle and Mammen(1993), we obtain the Monte Carlo density of T_n and the kernel density of $T^{*,W}$ from one bootstrap sample. The results are similar to those of Härdle and Mammen(1993): In all of the four cases the Monte Carlo densities and the kernel densities from one bootstrap sample are quite close but the asymptotic normal densities of theorem 1 are quite different. We omit the details.

We also study the power of our bootstrap test. The simulations are done for 2 different cases:

i) $m_\theta(x)=\theta_1+\theta_2 x$, $m(x)=1+x+\beta x e^{-2x}$.

ii) $m_\theta(x)=\theta_1+\theta_2 x+\theta_3 x^2$, $m(x)=2x-x^2+\beta(x-\frac{1}{4})(x-\frac{1}{2})(x-\frac{3}{4})$.

$M=100$ samples of size n=100 are drawn from each true model and the wild bootstrap resampling is performed $B=100$ times from only one of these samples. Uniform(0,1) is used as the design density. The critical values (of significance level $\alpha=0.05$) are obtained from the kernel densities from the bootstrapped samples. Tables 1, 2, 3 and 4 show the Monte Carlo estimates of power for different values of β and bandwidth h. The estimated power is the number of the rejections divided by $M=100$.

Power comparison between T_n and H-M is also performed. To compare the powers, we generate the samples using Beta(2,5) as the design density. Here we choose $H_0: m_\theta(x)=\theta_1+\theta_2 x$ against true regression function

i) $m(x)=1+x+\beta x e^{-2x}$

ii) $m(x)=1+x+\beta\sin(2\pi x)$.

For H-M test, the constant weight function, $w(x)=1$, is used, as was done in Härdle and

Mammen(1993). The critical values are chosen so that the level α equals to 0.05. In table 5 and table 6, the underlined rows show the right powers. For both cases T_n has better powers than H-M.

Table 1. Monte Carlo estimates of the power
$m_\theta(x) = \theta_1 + \theta_2 x$, $m(x) = 1 + x + \beta x e^{-2x}$,
$x \sim$ uniform(0, 1), ($\sigma(x) = 0.1$), level $\alpha = 0.05$

h β	0.0	0.5	1.0	2.0
0.10	0.11	0.27	0.80	1.00
0.20	0.09	0.29	0.84	1.00
0.23	0.08	0.28	0.81	1.00
0.25	0.05	0.27	0.81	1.00
0.30	0.04	0.29	0.76	1.00

Table 2. Monte Carlo estimates of the power
$m_\theta(x) = \theta_1 + \theta_2 x + \theta_3 x^2$, $m(x) = 2x - x^2 + \beta(x - \frac{1}{4})(x - \frac{1}{2})(x - \frac{3}{4})$,
$x \sim$ uniform(0, 1), ($\sigma(x) = 0.1$), level $\alpha = 0.05$

h β	0.0	0.5	1.0	2.0
0.10	0.10	0.15	0.45	0.85
0.20	0.05	0.07	0.34	0.82
0.22	0.04	0.07	0.37	0.81
0.30	0.02	0.03	0.27	0.48

Table 3. Monte Carlo estimates of the power
$m_\theta(x) = \theta_1 + \theta_2 x$, $m(x) = 1 + x + \beta x e^{-2x}$,
$x \sim$ uniform(0, 1), ($\sigma(x) = 0.1(1 + x^2)$), level $\alpha = 0.05$

h β	0.0	0.5	1.0	2.0
0.10	0.09	0.19	0.49	0.99
0.20	0.05	0.14	0.52	0.98
0.22	0.04	0.15	0.49	0.98
0.30	0.04	0.14	0.48	0.98

Table 4. Monte Carlo estimates of the power

$m_\theta(x) = \theta_1 + \theta_2 x + \theta_3 x^2,\quad m(x) = 2x - x^2 + \beta(x - \frac{1}{4})(x - \frac{1}{2})(x - \frac{3}{4}),$

$x \sim$ uniform(0, 1), ($\sigma(x) = 0.1(1+x^2)$), level $\alpha = 0.05$

h \ β	0.0	0.5	1.0	2.0
0.10	0.08	0.14	0.35	0.70
0.20	0.06	0.08	0.27	0.68
0.22	0.05	0.06	0.26	0.66
0.30	0.02	0.06	0.22	0.63

Table 5. Power comparison between H-M and T_n

$m_\theta(x) = \theta_1 + \theta_2 x,\quad m(x) = 1 + x + \beta x e^{-2x},$

$x \sim$ beta(2,5), ($\sigma(x) = 0.1$), level $\alpha = 0.05$

h	β	0.0	0.5	1.0	2.0
0.10	HM	0.19	0.37	0.57	0.87
	T_n	0.16	0.24	0.48	0.98
0.20	HM	0.06	0.16	0.42	0.77
	T_n	0.06	0.13	0.44	0.93
0.22	HM	0.05	0.11	0.37	0.73
	T_n	0.05	0.12	0.42	0.92
0.30	HM	0.04	0.08	0.27	0.63
	T_n	0.03	0.05	0.29	0.86

Table 6. Power comparison between H-M and T_n

$m_\theta(x) = \theta_1 + \theta_2 x,\quad m(x) = 1 + x + \beta\sin(2\pi x),$

$x \sim$ beta(2,5), ($\sigma(x) = 0.1$)

h	β	0.0	0.5	1.0	2.0
0.10	HM	0.21	0.95	0.98	1.00
	T_n	0.17	0.96	0.99	1.00
0.20	HM	0.11	0.93	0.94	0.94
	T_n	0.12	0.92	0.97	0.98
0.25	HM	0.06	0.83	0.86	0.85
	T_n	0.05	0.86	0.95	0.95
0.30	HM	0.05	0.72	0.74	0.78
	T_n	0.01	0.78	0.90	0.93

4. Conclusion

In this paper, we show that our test statistic T_n has asymptotic normal distribution with mean and variance given in theorem 1, which is the same as that of H-M statistic with $f(x)$ as the weight function $w(x)$. However, the convergence rate to normal distribution is so slow that we use the wild bootstrap to find the critical values. The simulation results show that wild bootstrap estimates the distribution of T_n quite well. Most of the simulation results are similar to H-M, but T_n has better power than H-M test statistic when design density is not uniform and the parametric null model is linear.

Appendix

Proof of Theorem 1.

It is known that

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = f(x) + O_p(n^{-2/5}(\log n)^{1/2}) \qquad \text{(uniformly in } x)$$

$$\hat{m}_h(x) = m(x) + O_p(n^{-2/5}(\log n)^{1/2}). \qquad \text{(uniformly in } x)$$

Using the above equations, it can be shown that

$$\begin{aligned} T_n &= \sqrt{h}\sum_{i=1}^{n}(\hat{m}_h(X_i) - \pounds_{h,n} m_{\hat{\theta}}(X_i))^2 \\ &= \sqrt{h}\sum_{i=1}^{n}(\hat{m}_h(X_i) - \pounds_{h,n} m_{\hat{\theta}}(X_i))^2 \left(\frac{\hat{f}(X_i)}{f(X_i)}\right)^2 + o_p(1) \\ &= \sqrt{h}\sum_{i=1}^{n}\left(\frac{\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j)(m(X_j) + \varepsilon_j - m_{\hat{\theta}}(X_j))}{f(X_i)}\right)^2 + o_p(1). \end{aligned}$$

Since

$$m(X_j) = m_{\theta_0}(X_j) + n^{-1/2}h^{-1/4}\Delta_n(X_j),$$

T_n can be written as

$$T_n = \sqrt{h}\sum_{i=1}^{n}(U_{n,1}(X_i) + U_{n,2}(X_i) + U_{n,3}(X_i))^2 + o_p(1),$$

where

$$U_{n,1}(X_i) = \frac{\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j) n^{-1/2} h^{-1/4} \Delta_n(X_j)}{f(X_i)},$$

$$U_{n,2}(X_i) = \frac{\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j)\varepsilon_j}{f(X_i)},$$

$$U_{n,3}(X_i) = \frac{-\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j)\frac{1}{n}\sum_{l=1}^{n} g(X_j)^T u(X_l)\varepsilon_l}{f(X_i)}.$$

Now,

$$\begin{aligned}
E\,&\sqrt{h}\sum_{i=1}^{n} U_{n,3}^2(X_i) \\
&= n\sqrt{h}E\,U_{n,3}^2(X_1) \\
&= n\sqrt{h}\cdot\frac{1}{n^4}E\left(\frac{1}{f(X_1)}\sum_{j=1}^{n} K_h(X_1 - X_j)\sum_{l=1}^{n} g(X_j)^T u(X_l)\varepsilon_l\right)^2 \\
&= n^{-3}\sqrt{h}E\left(\frac{1}{f(X_1)}\sum_{l=1}^{n}\sigma^2(X_l)\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} K_h(X_1 - X_{j_1})K_h(X_1 - X_{j_2})\right. \\
&\qquad \left.\times\; g(X_{j_1})^T u(X_l) g(X_{j_2})^T u(X_l)\right) \\
&= n^{-3}\sqrt{h}\cdot O(n^3) \\
&= \sqrt{h} \to 0 \quad \text{as} \quad n\to\infty.
\end{aligned}$$

Similarly it can be shown that

$$E\sqrt{h}\left(\sum_{i=1}^{n} U_{n,l_1}(X_i)\cdot\sum_{j=1}^{n} U_{n,l_2}(X_j)\right)^2 \to 0, \quad \text{as } n\to\infty$$

$$\text{for all } 1\le l_1 < l_2 \le 3.$$

The above results imply

$$\begin{aligned}
T_n &= E\sqrt{h}\left(\sum_{i=1}^{n} U_{n,1}(X_i)^2 + \sum_{j=1}^{n} U_{n,2}(X_j)^2\right) + o_p(1) \\
&= T_{n,1} + T_{n,2} + T_{n,3} + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
T_{n,1} &= \sqrt{h}\sum_{i=1}^{n} U_{n,1}^2(X_i), \\
T_{n,2} &= \sqrt{h}\sum_{i=1}^{n}\frac{\frac{1}{n^2}\sum_{j=1}^{n} K_h^2(X_i - X_j)\varepsilon_j^2}{f^2(X_i)}, \\
T_{n,3} &= \sqrt{h}\sum_{i=1}^{n}\left(\frac{\frac{1}{n^2}\sum_{j\ne l} K_h(X_i - X_j)K_h(X_i - X_l)\varepsilon_j\varepsilon_l}{f^2(X_i)}\right).
\end{aligned}$$

It is sufficient to show that

(*i*) $T_{n,1} = \int (K_h * \Delta_n(x))^2 f(x)dx + o_p(1)$,

(*ii*) $T_{n,2} = b_h + o_p(1)$ and

(*iii*) $d(\pounds(T_{n,3}), N(0, V)) \to 0$.

Proof of (*i*)

$$
\begin{aligned}
E(T_{n,1}) &= \sqrt{h}\cdot nEU_{n,1}^2(X_1) \\
&= n^{-2}E\left(\frac{\sum_{j=1}^{n} K_h(X_1-X_j)\Delta_n(X_j)}{f(X_1)}\right)^2 \\
&= n^{-2}E\left(\frac{K_h(0)\Delta_n(X_1)\sum_{l=1}^{n} K_h(X_1-X_l)\Delta_n(X_l)}{f^2(X_1)}\right. \\
&\qquad + \frac{\sum_{j=2}^{n} K_h^2(X_1-X_j)\Delta_n^2(X_j)}{f^2(X_1)} \\
&\qquad \left. + \frac{2\sum_{j=2}^{n}\sum_{l>j} K_h(X_1-X_j)K_h(X_1-X_l)\Delta_n(X_j)\Delta_n(X_l)}{f^2(X_1)}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= n^{-2}\left\{E\left(\frac{K_h^2(0)\Delta_n^2(X_1)}{f^2(X_1)}\right)\right. \\
&\qquad +(n-1)E\left(\frac{K_h(0)\Delta_n(X_1)K_h(X_1-X_2)\Delta_n^2(X_2)}{f^2(X_1)}\right) \\
&\qquad +(n-1)E\left(\frac{K_h^2(X_1-X_2)\Delta_n^2(X_2)}{f^2(X_1)}\right) \\
&\qquad \left. +2\binom{n-1}{2}E\left(\frac{K_h(X_1-X_2)K_h(X_1-X_3)\Delta_n(X_2)\Delta_n(X_3)}{f^2(X_1)}\right)\right\} \\
&= n^{-2}((n-1)(n-2)A_n + O(h^{-2}) + O(nh^{-1})\ ,
\end{aligned}
$$

where

$$
\begin{aligned}
A_n &= E\left(\frac{K_h(X_1-X_2)K_h(X_1-X_3)\Delta_n(X_2)\Delta_n(X_3)}{f^2(X_1)}\right) \\
&= \int\int\int \frac{K_h(x_1-x_2)K_h(x_1-x_3)\Delta_n(x_2)\Delta_n(x_3)}{f^2(x_1)} \\
&\qquad \times f(x_1)f(x_2)f(x_3)dx_1dx_2dx_3 \\
&= \int\int\int \frac{K_h(u_2)K_h(u_3)\Delta_n(x_1-u_2)\Delta_n(x_1-u_3)}{f(x_1)} \\
&\qquad \times f(x_1-u_2)f(x_1-u_3)\,du_2du_3dx_1 \\
&= \int (K_h*\Delta_n(x))^2 f(x)dx + O(h).
\end{aligned}
$$

This shows that

$$
E(T_{n,1}) = \int (K_h*\Delta_n(x))^2 f(x)dx + o(1). \tag{A.1}
$$

Similarly it can be shown that

$$Var(T_{n,1} | X_1,\cdots,X_n) = o_p(1). \qquad (A.2)$$

Equations (A.1) and (A.2) together imply

$$T_{n,1} = \int (K_h * \Delta_n(x))^2 f(x) dx + o_p(1).$$

Proof of (*ii*)

$$\begin{aligned}
&E(T_{n,2}) \\
&= n^{-1}\sqrt{h} E\left(\frac{\sum_{j=1}^{n} K_h^2(X_1 - X_j)\varepsilon_j^2}{f^2(X_1)} \right) \\
&= n^{-1}\sqrt{h} E\left(\frac{K_h^2(0)\sigma^2(X_1)}{f^2(X_1)} \right) + (n-1) E\left(\frac{K_h^2(X_1 - X_2)\sigma^2(X_2)}{f^2(X_1)} \right) \\
&= \sqrt{h} \cdot n^{-1}(n-1) \int\int \frac{K_h^2(x_1 - x_2)\sigma^2(x_2)}{f^2(x_1)} f(x_1) f(x_2) dx_1 dx_2 + O\left(\frac{\sqrt{h}}{n} h^{-2} \right) \\
&= \sqrt{h} \cdot n^{-1}(n-1) \left(\int K_h^2(u) du \cdot \int \sigma^2(x) dx + O(1) \right) + O(n^{-1} h^{-3/2}) \\
&= h^{-1/2}(K * K(0)) \int \sigma^2(x) dx + o(1).
\end{aligned}$$

Now

$$\begin{aligned}
&Var(T_{n,2} | X_1,\cdots,X_n) \\
&= Var\left(\sqrt{h} \sum_{i=1}^{n} \frac{1}{n^2} \frac{\sum_{j=1}^{n} K_h^2(X_i - X_j)\varepsilon^{2_i}}{f(X_i)} | X_1,\cdots,X_n \right) \\
&= \sqrt{h} \cdot n^{-4} \sum_{i=1}^{n} \left(\sum_{i=1}^{n} \frac{K_h^2(X_i - X_j)}{f^2(X_i)} \right)^2 Var(\varepsilon_j^2 | X_1,\cdots,X_n) \\
&\le h \cdot n^{-4} \sum_{j=1}^{n} \left(\frac{1}{h^2} \cdot n \cdot M_1 \right)^2 M_2(X_1,\cdots,X_n).
\end{aligned}$$

for some M_1 and $M_2(X_1, \cdots, X_n)$. Therefore, $Var(T_{n,2} | X_1,\cdots,X_n) = O_p(n^{-1}h^{-3}) = o_p(1)$.

This proves $T_{n,2} = b_h + o_p(1)$.

Proof of *(iii)*

Put

$$W_{jln} = \begin{cases} \sqrt{h} \sum_{i=1}^{n} \frac{1}{n^2} \frac{K_h^2(X_i - X_j) K_h^2(X_i - X_l) \varepsilon_j \varepsilon_l}{f^2(X_i)} &, \ j \neq l \\ 0 &, \ j = l \end{cases}$$

Then $T_{n,3} = \sum_j \sum_l W_{jln}$.

We will show that

$$Var(T_{n,3}) - V \rightarrow 0 \quad as\ n \rightarrow \infty. \tag{A.3}$$

$$\max_{1 \le j \le n} \sum_{l=1}^{n} Var(W_{jln}) / Var(T_{n,3}) \rightarrow 0. \tag{A.4}$$

$$ET_{n,3}^4 / Var(T_{n,3})^2 \rightarrow 3. \tag{A.5}$$

Then by Theorem 2.1 in De Jong(1987),

$$d(\pounds(T_{n,3}), N(0, V)) \rightarrow 0.$$

Proof of (A.3)

$$\begin{aligned} Var(T_{n,3}) &= E(\sum_j \sum_l W_{jln})^2 \\ &= 4 \sum_{j<l} E(W_{jln}^2) \\ &= 4\binom{n}{2} E(W_{23n}^2). \end{aligned}$$

By tedious calculation, it can be shown that

$$E(W_{23n}^2) = n^{-2}\Big(h \int \int (K_h * K_h(x-y))^2 \sigma^2(x) \sigma^2(y)\, dx\, dy + o(1)\Big).$$

Therefore,

$$\begin{aligned} Var(T_{n,3}) &= 2n(n-1)n^{-2}(V + o(1)) \\ &= V + o(1). \end{aligned}$$

Proof of (A.4)

In the proof of (A.3), we have shown that

$$Var(W_{jln}) = O(n^{-2} h^{-1}).$$

which implies (A.4).

Proof of (A.5)

Similarly as in the calculation of $E(W_{23n}^2)$, the followings can be shown

$$E\ W_{12n} W_{34n} W_{41n} = o(n^{-4})$$

$$E\ W_{12n} W_{23n}^2 W_{31n} = o(n^{-3})$$

$$E\ W_{12n}^4 = o(n^{-2}).$$

Equation (A.5) is proved by the above equations and the following equation

$$\begin{aligned} E\ T_{n,3}^4 &= 12 \sum_{i \ne j} \sum_{k \ne l} E(W_{ijl}^2 W_{kln}^2) + 8 \sum_{i \ne j} E(W_{ijn}^4) \\ &+ 192 \sum_{i \ne j} \sum_{j \ne k} \sum_{k \ne i} E(W_{ijn} W_{jkn}^2 W_{kin}) \\ &+ 48 \sum_{i \ne j} \sum_{j \ne k} \sum_{k \ne l} \sum_{l \ne i} E(W_{ijn} W_{jkn} W_{kln} W_{lin}). \end{aligned}$$

References

[1] Beran,R. (1986). Comment on ''Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu, *The Annals of Statistics*, 14, 1295-1298.

[2] Buckley, M. J. (1992). Detecting a smooth signal : Optimality of cusum based procedures, *The Annals of Statistics*, 12, 101-118.

[3] Cox, D., Koh, E., Washable. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models, *The Annals of Statistics*, 16, 113-119

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

[5] Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria, *The Annals of Statistics,* 20, 1412-1425.

[6] Eubank, R. L and LaRiccia, V. N. (1990). Asymptotic comparison of Cramér-von Mises and nonparametric function estimation techniques for testing goodness-of-fit, *The Annals of Statistics*, 16, 113-119.

[7] Eubank, R. L. and Spiegelman, C. (1990). Testing the goodness-of-fit of linear models via nonparametric regression techniques, *Journal of the American Statistical Association*, 85, 387-392.

[8] Härdle, W. (1990). *Applied Nonparametric Regression* Econometric Society Monograph Series 19, Cambridge Univ. Press.

[9] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *The Annals of Statistics*, 21, 1926-1947.

[10] Härdle, W. and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimal?, *Journal of the American Statistical Association*, 83, 86-95.

[11] Liu, R. (1988). Bootstrap procedures under some non i.i.d. models, *The Annals of Statistics*, 16, 1696-1708.

[12] Mammen, E. (1992). *When does bootstrap work : Asymptotic Results and Simulations*, Lecture notes in statist, 77 Springer, Berlin.

[13] Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models, *The Annals of Statistics*, 21, 255-285.

[14] Munson, P. J. and Jernigan, R. W. (1989). A cubic spline extension of the Durbin-Watson test, *Biometrika,* 76, 39-47.

[15] Yanagimoto, T and Yanagimoto, M.(1987). The use of marginal likelihood for a diagnostic test for the goodness of fit of the simple linear regression model, *Technometrics*, 29, 95-101.

[16] Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion), *The Annals of Statistics*, 29, 1261-1350.