J Syst Sci Complex (20XX) XX: 1–5

Supplementary Material for

Checking Adequacy of Variance Function in Nonparametric Regression with Unknown Mean Function

LIANG Jia · SONG Weixing · SHI Jianhong

DOI:
Received: x x 20xx / Revised: x x 20xx

Abstract This note contains the proof of Lemma 9 in the paper.

To prove Theorem 4, we need the following lemma from [15] which is reproduced here for the sake of completeness.

Lemma 7 *Let* $\widetilde{X}_i$, $1 \le i \le n$, *be i.i.d. random vectors, and define*

$$U_n := \sum_{1\le i<j\le n} H_n(\widetilde{X}_i, \widetilde{X}_j), \quad G_n(x,y) = EH_n(\widetilde{X}_1, x)H_n(\widetilde{X}_1, y),$$

where H_n *is a sequence of measurable functions symmetric under permutation, with*

$$E(H_n(\widetilde{X}_1, \widetilde{X}_2)|\widetilde{X}_1) = 0 \;\; a.s., \quad EH_n^2(\widetilde{X}_1, \widetilde{X}_2) < \infty$$

for each $n \ge 1$. *If*

$$\frac{[EG_n^2(\widetilde{X}_1, \widetilde{X}_2) + n^{-1}EH_n^4(\widetilde{X}_1, \widetilde{X}_2)]}{[EH_n^2(\widetilde{X}_1, \widetilde{X}_2)]^2} \to 0,$$

then U_n *is asymptotically normal with mean 0 and variance* $\lim_{n\to\infty} n^2 EH_n^2(\widetilde{X}_1, \widetilde{X}_2)/2$.

LIANG Jia
Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, 38105, USA. Email: Jia.Liang@STJUDE.ORG
SONG Weixing
Department of Statistics, Kansas State University, Manhattan, KS, 66503, USA. Email: weixing@ksu.edu
SHI Jianhong
College of Mathematics and Computer Science, Shanxi Normal University, Taiyan, 030031, China. Email: shijh700717@qq.com

*Jianhong Shi's research is supported by the National Natural Science Foundation of China (Grant No. 12071267).
⋄

Denote

$$\widetilde{T}_n(\theta_0) = \int_{\mathcal{C}} [n^{-1} \sum_{i=1}^n K_{hi}(x)[(Y_i - m(X_i))^2 - v_0(X_i)] f^2(X_i)]^2 d\varphi(x),$$

$$\widetilde{C}_n = \int_{\mathcal{C}} \frac{1}{n^2} \sum_{i=1}^n K_{hi}^2(x) \xi^2(X_i) f^4(X_i) d\varphi(x).$$

Theorem 4 can be proved by combining the results from the following lemmas.

Lemma 8 *Under H_0, if (e1), (e2), (e4), (f1), (g), (k), and (h1), (h2), (w) hold, then*

$$nh^{d/2}(\widetilde{T}_n(\theta_0) - \widetilde{C}_n) \xrightarrow{d} N(0, \Gamma).$$

Proof: First, by subtracting $m(X_i)$, $f(X_i)$, $1/f^6(x)$ from $\widehat{m}(X_i)$, $\widehat{f}_h(X_i)$, $1/\widehat{f}_w^6(x)$, respectively, we can show that $nh^{d/2}T_n(\theta_0) = nh^{d/2}\widetilde{T}_n(\theta_0) + o_p(1)$, and $\widetilde{T}_n(\theta_0)$ can be further decomposed as $\widetilde{T}_n(\theta_0) = \widetilde{C}_n + M_n$, where

$$M_n = \int_{\mathcal{C}} \frac{1}{n^2} \sum_{i \neq j}^n K_{hi}(x) K_{hj}(x) \xi(X_i) \xi(X_j) f^2(X_i) f^2(X_j) d\varphi(x).$$

To apply Lemma 7, let $\widetilde{X}_i = (X_i^T, \xi(X_i))^T$,

$$H_n(\widetilde{X}_i, \widetilde{X}_j) := \frac{h^{d/2}}{n} \int K_{hi}(x) K_{hj}(x) \xi(X_i) \xi(X_j) f^2(X_i) f^2(X_j) d\varphi(x),$$

and $G_n(t, s) = EH_n(\widetilde{X}_1, t) H_n(\widetilde{X}_1, s)$. It is easy to see that $nh^{d/2}M_n/2 = \sum_{i<j} H_n(\widetilde{X}_i, \widetilde{X}_j)$. After some tedious algebra, we can show that

$$\frac{EG^2(\widetilde{X}_1, \widetilde{X}_2)}{[EH_n^2(\widetilde{X}_1, \widetilde{X}_2)]^2} = o(1), \quad \frac{EH^4(\widetilde{X}_1, \widetilde{X}_2)}{n[EH_n^2(\widetilde{X}_1, \widetilde{X}_2)]^2} = o(1),$$

and finally, after some algebra, we have

$$\frac{1}{2} n^2 EH_n^2(\widetilde{X}_1, \widetilde{X}_2) \to \frac{1}{2} \int \frac{\tau^4(x) g^2(x)}{f^2(x)} dx \int \left(\int K(u) K(v+u) du \right)^2 dv = \frac{\Gamma}{4},$$

by the bounded convergence theorem, and the continuity of $\tau^2(x)$, $f(x)$ and $g(x)$. This concludes the proof of the theorem.

Lemma 9 *Under H_0, if (e1), (e2), (f1), (f2), (k), (v3)-(v5), (h1), (h2) and (w) hold, then $nh^{d/2}|T_n(\widehat{\theta}_n) - T_n(\theta_0)| = o_p(1)$, $nh^{d/2}|T_n(\theta_0) - \widetilde{T}_n(\theta_0)| = o_p(1)$, $nh^{d/2}(\widehat{C}_n - \widetilde{C}_n) = o_p(1)$ and $\widehat{\Gamma}_n - \Gamma = o_p(1)$*

Proof: First we write $T_n(\theta_0) - T_n(\widehat{\theta}_n)$ as $2Q_1 - Q_2$, where

$$Q_1 = \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(x)((Y_i - \widehat{m}(X_i))^2 - v_0(X_i)) \widehat{f}_h^2(X_i) \right] \cdot \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(x)(v_0(X_i) - v(X_i, \widehat{\theta}_n)) \widehat{f}_h^2(X_i) \right] d\widehat{\varphi}_w(x)$$

$$Q_2 = \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(x)(v_0(X_i) - v(X_i, \widehat{\theta}_n)) \widehat{f}_h^2(X_i) \right]^2 d\widehat{\varphi}_w(x).$$

Thus it suffices to show that $nh^{d/2}Q_1 = o_p(1)$, and $nh^{d/2}Q_2 = o_p(1)$. Note that Q_1 can be written as the sum $Q_1 = Q_{11} + Q_{12}$, where

$$Q_{11} = \int U_n(x)\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)d_{ni}\right]\widehat{f}_h^2(X_i)d\widehat{\varphi}_w(x),$$
$$Q_{12} = u_n^T \int U_n(x)\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}_0(X_i)\right]\widehat{f}_h^2(X_i)d\widehat{\varphi}_w(x).$$

By Theorem 2, $nh^dT_n(\theta_0) = O_p(1)$, and the $\sqrt{n}$-consistency of $\widehat{\theta}_n$, we can show that

$$nh^{d/2}|Q_{11}| \le \max_i \frac{|d_{ni}|}{||u_n||}(nh^d)^{1/2}||u_n|| \int \widehat{f}_h^6(x)d\widehat{\varphi}_w(x)\sqrt{n}\int U_n^2(x)d\widehat{\varphi}_w(x) = o_p(1).$$

Similarly, we can show $nh^{d/2}Q_{12} = o_p(1)$, and $nh^{d/2}Q_2 = o_p(1)$. This concludes the proof of $nh^{d/2}|T_n(\widehat{\theta}_n) - T_n(\theta_0)| = o_p(1)$.

Using the second part of Theorem 2.2 in [22], together with the assumption (w), we have

$$\begin{aligned} nh^{d/2}|T_n(\theta_0) - \widetilde{T}_n(\theta_0)| &\le nh^{d/2}\int_{\mathcal{C}} U_n^2(x)d\varphi(x)\cdot \sup_{x\in\mathcal{C}}|f^6(x)/\widehat{f}_w^6(x) - 1| \\ &= nh^{d/2}O_p((nh^d)^{-1})O_p((\log_k n)(\log n/n)^{d/(d+4)}) = o_p(1), \end{aligned}$$

which implies $nh^{d/2}|T_n(\theta_0) - \widetilde{T}_n(\theta_0)| = o_p(1)$.

By the routine addition and subtraction technique, we can write $\widehat{C}_n$ as the sum of $\widetilde{C}_n$ and other eleven terms. We can show that those eleven terms, multiplied by $nh^{d/2}$ have the order $o_p(1)$.

Since we have already known that $\Gamma_n \to \Gamma$, so if suffices to show that $\widehat{\Gamma}_n - \widetilde{\Gamma}_n = o_p(1)$, $\widetilde{\Gamma}_n - \Gamma_n = o_p(1)$, where,

$$\widetilde{\Gamma}_n = \frac{h^d}{n^2}\sum_{i\ne j}\left(\int_{\mathcal{C}} K_{hi}(x)K_{hj}(x)\xi(X_i)\xi(X_j)f_h^2(X_i)f_h^2(X_j)f^{-6}(x)dG(x)\right)^2.$$

To show that $\widehat{\Gamma}_n - \widetilde{\Gamma}_n = o_p(1)$, note that

$$\begin{aligned} \widehat{\Gamma}_n &= \frac{h^d}{n^2}\sum_{i\ne j}\left(\int_{\mathcal{C}} K_{hi}(x)K_{hj}(x)\widehat{\xi}(X_i)\widehat{\xi}(X_j)\widehat{f}_h^2(X_i)\widehat{f}_h^2(X_j)\widehat{f}_w^{-6}(x)dG(x)\right)^2 \\ &= \frac{h^d}{n^2}\sum_{i\ne j}\Bigg(\int_{\mathcal{C}} K_{hi}(x)K_{hj}(x)((Y_i - m(X_i) - \Delta_{im})^2 - v_0(X_i) + v_0(X_i) - v(X_i,\widehat{\theta}_n))\cdot \\ &\qquad ((Y_j - m(X_j) - \Delta_{jm})^2 - v_0(X_j) + v_0(X_j) - v(X_j,\widehat{\theta}_n))\cdot \\ &\qquad (\Delta_{if^2} + f^2(X_i))(\Delta_{if^2} + f^2(X_j))(\widehat{f}_w^{-6}(x) - f^{-6}(x) + f^{-6}(x))dG(x)\Bigg)^2. \end{aligned}$$

After expanding the square terms, $\widehat{\Gamma}_n$ can be written as the sum of $\widetilde{\Gamma}_n$ and other terms which include at least one term from Δ_{im}^2, Δ_{vi} and Δ_{if^2}. Hence, except for $\widetilde{\Gamma}_n$, it is easy to see that all other terms are $o_p(1)$. Finally,

$$\begin{aligned} E\widetilde{\Gamma}_n &= \frac{h^d}{n^2}\sum_{i\neq j} E\left(\int_{\mathcal{C}} K_{hi}(x)K_{hj}(x)\xi(X_i)\xi(X_j)f_h^2(X_i)f_h^2(X_j)f^{-6}(x)dG(x)\right)^2 \\ &= \frac{(n-1)h^d}{n}\int\int (E(K_{h1}(x)K_{h1}(y)\xi^2(X_1)f^4(X_1)))^2 d\varphi(x)d\varphi(y) = \Gamma_n \end{aligned}$$

Therefore, we have

$$\begin{aligned} E(\widetilde{\Gamma}_n - \Gamma_n)^2 &\leq \sum_{i\neq j} EH_n^4(\widetilde{X}_i, \widetilde{X}_j) + c\sum_{i\neq j\neq k} EH_n^2(\widetilde{X}_i, \widetilde{X}_j)H_n^2(\widetilde{X}_j, \widetilde{X}_k) \\ &\leq (n^2 + cn^3)EH^4(\widetilde{X}_1, \widetilde{X}_2) = O(1/nh^d). \end{aligned}$$

This concludes the proof of $\widehat{\Gamma}_n - \Gamma = o_p(1)$.

References

[1] Dette, H., Neumeyer, N., & Van Keilegom, I. (2007). A new test for the parametric form of the variance function in nonparametric regression *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**(5), 903–917.

[2] Samarakoon, N., & Song, W. (2011). Minimum distance conditional variance function checking in heteroscedastic regression models. *Journal of Multivariate Analysis*, **102**(3), 579–600.

[3] Zheng, X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, **75**(2), 263–289.

[4] Song, W., & Du, J. (2011). A note on testing the regression functions via nonparametric smoothing. *The Canadian Journal of Statistics*, **39**(1), 108–125.

[5] Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Annals of Statistics*, **5**(3), 445–463.

[6] Pardo-Fern$r\acute{m}$andez, J. C., & Jim$r\acute{m}$enez-Gamero, M. D. (2019). A model specification test for the variance function in nonparametric regression. *AStA Advances in Statistical Analysis*, **103**(3), 387–410.

[7] Dette, H., et al. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Annals of Statistics*, **27**(3), 1012–1040.

[8] Koul, H. L., & Ni, P. (2004). Minimum distance regression model checking. *Journal of Statistical Planning and Inference*, **119**(1), 109–141.

[9] Dette, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *Journal of Statistical Planning and Inference*, **103**(1–2), 311–329.

[10] Zheng, X. (2009). Testing heteroscedasticity in nonlinear and nonparametric regressions. *Canadian Journal of Statistics*, **37**, 282–300.

[11] Dette, H., Munk, A., & Wagner, T. (1998). Estimating the variance in nonparametric regression—what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **60**(4), 751–764.

[12] Dette, H., & Hetzler, B. (2009). A simple test for the parametric form of the variance function in nonparametric regression. *Annals of the Institute of Statistical Mathematics*, **61**(4), 861.

[13] Zhu, L., Fujikoshi, Y., & Naito, K. (2001). Heteroscedasticity checks for regression models. *Science in China Series A: Mathematics*, **44**(10), 1236–1252.

[14] Samarakoon, N., & Song, W. (2012). Empirical smoothing lack-of-fit tests for variance function. *Journal of Statistical Planning and Inference*, **142**(5), 1128–1140.

[15] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, **14**(1), 1–16.

[16] Mack, Y., & Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **61**(3), 405–415.

[17] Gonz$r\acute{m}$alez-Manteiga, W., & Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST*, **22**(3), 361–411.

[18] Gao, J., & Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, **103**(484), 1584–1594.

[19] Zhu, X., Chen, F., Guo, X., & Zhu, L. (2016). Heteroscedasticity testing for regression models: A dimension reduction-based model adaptive approach. *Computational Statistics & Data Analysis*, **103**, 263–283.

[20] Neumeyer, N., & Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, **101**(5), 1067–1078.

[21] Einmahl, J. H. J., & Van Keilegom, I. (2008). Specification tests in nonparametric regression. *Journal of Econometrics*, **143**(1), 88–102.

[22] Bosq, D. (2012). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer.

[23] Hart, J. (1997). *Nonparametric Smoothing and Lack-of-fit tests*. Springer.

[24] Zheng, X. (2009). Testing heteroscedasticity in nonlinear and nonparametric regressions. *Canadian Journal of Statistics*, **37**(2), 282–300.

[25] Frankel, I. G., Liang, J., Fitzsimmons, E. J., & Newmark, G. L. (2020). Evaluating police-caused crashes. In *99th Annual Meeting of the Transportation Research Board, Washington, D.C.*.