Statistica Sinica **26** (2016), 137-155
doi:http://dx.doi.org/10.5705/ss.2014.112

NONPARAMETRIC TESTING IN REGRESSION MODELS WITH WILCOXON-TYPE GENERALIZED LIKELIHOOD RATIO

Long Feng[1], Zhaojun Wang[1], Chunming Zhang[2] and Changliang Zou[1]

[1]*Nankai University and* [2]*University of Wisconsin, Madison*

Abstract: The generalized likelihood ratio (GLR) statistic (Fan, Zhang, and Zhang (2001)) offered a generally applicable method for testing nonparametric hypotheses about nonparametric functions, but its efficiency is adversely affected by outlying observations and heavy-tailed distributions. Here a robust testing procedure is developed under the framework of the GLR by incorporating a Wilcoxon-type artificial likelihood function, and adopting the associated local smoothers. Under some useful hypotheses, the proposed test statistic is asymptotically normal and free of nuisance parameters and covariate designs. Its asymptotic relative efficiency with respect to the least squares-based GLR method is closely related to that of the signed-rank Wilcoxon test in comparison with the t-test. Simulation results are consistent with the asymptotic analysis.

Key words and phrases: Asymptotic relative efficiency, bootstrap, lack-of-fit test, local polynomial regression, local Walsh-average regression, model specification.

1. Introduction

Over the last two decades, nonparametric modeling techniques have developed rapidly due to the reduction of modeling biases of traditional parametric methods. This raises such important inference questions as whether a parametric family adequately fits a given data set. Here we choose the varying coefficient model for our investigation because it arises in many statistical problems. Suppose $\{(Y_i, \boldsymbol{X}_i, U_i)\}_{i=1}^n$ is a random sample from the varying-coefficient model,

$$Y = \alpha(U) + \boldsymbol{A}(U)^T\boldsymbol{X} + \varepsilon \tag{1.1}$$

with $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^T$ and $\boldsymbol{A}(U) = (a_1(U), \ldots, a_p(U))^T$. Such nonparametric inferences or testings, as the problem of parametric null against nonparametric alternative hypothesis, or model checking for partial linear models, are included as a special cases of hypothesis testing problems under this model. A widely used null hypothesis testing problem is

$$H_{0a} : (\alpha, \boldsymbol{A}^T) = (\alpha_0, \boldsymbol{A}_0^T) \quad \text{versus} \quad H_{1a} : (\alpha, \boldsymbol{A}^T) \neq (\alpha_0, \boldsymbol{A}_0^T), \tag{1.2}$$

where α_0 and $\boldsymbol{A}_0$ are two known functions. An intuitive approach is based on generalizations of the Kolmogorov-Smirnov or Cramer-von Mises statistics to measure the distance between the estimators under the null and alternative models; see Härdle and Mammen (1993), Neumeyer and Van Keilegom (2010), and the references therein. However, it is difficult to find an optimal measure for such type of statistics. Zheng (1996) proposed a consistent test of functional nonlinear regression models by combining the methodology of the conditional moment test (Bierens (1990)) and nonparametric estimation techniques. See Zhang and Dette (2004) for a power comparison of some types of nonparametric regression tests.

In an important work, Fan, Zhang, and Zhang (2001) proposed the generalized likelihood ratio (GLR) test statistic

$$\lambda_n^{\mathrm{GLR}}(\alpha_0, \boldsymbol{A}_0) = l_n^{\mathrm{GLR}}(H_{1a}) - l_n^{\mathrm{GLR}}(H_{0a}) = \frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1} \approx \frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}, \tag{1.3}$$

where $\mathrm{RSS}_0 = \sum_{k=1}^n (Y_k - \alpha_0(U_k) - \boldsymbol{A}_0(U_k)^T \boldsymbol{X}_k)^2$, $\mathrm{RSS}_1 = \sum_{k=1}^n (Y_k - \widehat{\alpha}(U_k) - \widehat{\boldsymbol{A}}(U_k)^T \boldsymbol{X}_k)^2$, and $(\widehat{\alpha}(\cdot), \widehat{\boldsymbol{A}}(\cdot))$ are the local linear estimators under the alternative model. This test is shown to possess the Wilks phenomenon and to be asymptotically optimal in certain sense; it has become a commonly used methodology for constructing nonparametric testing in regression models. See Fan and Jiang (2007) for an overview of the idea of GLR inference in different nonparametric models. Although the GLR test is asymptotically distribution-free, the normal likelihood function and the corresponding local least-squares polynomial estimators (Fan and Gijbels (1996)) are employed. Accordingly, its statistical properties could potentially be (highly) affected when the errors are far from normal or the data contain some outliers.

We develop a robust test under the framework of the GLR. It relaxes the usually strong distributional assumption associated with the least-squares-based GLR by adopting a Wilcoxon-type dispersion function (Hettmansperger and McKean (2010)) and the corresponding local smoothers. We establish that the Wilcoxon-type GLR preserves the Wilks phenomenon without the need to assume a normal likelihood. Under the null hypothesis the test statistic is asymptotically normal and free of nuisance parameters. Under certain conditions, its Pitman asymptotic relative efficiency (ARE) with respect to the GLR test is established. This ARE is closely related to that of the signed-rank Wilcoxon test in comparison with the t-test. Thus, it outperforms the least-squares-based GLR with heavier-tailed data in the sense that asymptotically it can yield substantially larger power. A simulation study was conducted to compare it with some other available procedures in the literature. When the errors deviate from normality, our tests are more powerful than the least-square-based or moment-based methods. Even when the errors are normally distributed, our procedure does not lose much.

Considerable efforts have been devoted to construct some robust nonparametric polynomial smoothers. See Fan, Hu, and Truong (1994), Welsh (1996), Kai, Li, and Zou (2010) and Feng, Zou, and Wang (2012). However, there is very few work on robust inference. Wang and Qu (2007) robustified Zheng's (1996) test based on the centered asymptotic rank transformation of the residuals from a robust fit under the null hypothesis. To our best knowledge in the literature of nonparametric model checking, there is no corresponding test in which robust local smoothers are considered.

2. Methodology

2.1. Test statistic and its null limiting distribution

Our development is based on a new artificial likelihood function. The weighted rank-based L_1 norm is often used in the development of robust statistical procedures (Hettmansperger and McKean (2010)). $\|\boldsymbol{\varepsilon}\|_W = (\sqrt{12}/(n+1)) \sum_{i=1}^{n} r_i|\varepsilon_i|$, where r_i denotes the rank of $|\varepsilon_i|$ among $|\varepsilon_1|, \ldots, |\varepsilon_n|$. It is equivalent to

$$\frac{\sqrt{12}}{n+1}\sum\sum_{i<j}\left|\frac{\varepsilon_i - \varepsilon_j}{2}\right| + \frac{\sqrt{12}}{n+1}\sum\sum_{i\leq j}\left|\frac{\varepsilon_i + \varepsilon_j}{2}\right| \equiv W_n(\boldsymbol{\varepsilon}) + R_n(\boldsymbol{\varepsilon}).$$

In particular, in the univariate location problem, given a sample of observations $\mathbf{x} = \{x_1, \ldots, x_n\}$, the famous Hodges-Lehmann estimate is the solution to the minimization problem $\arg\min_\theta \|\mathbf{x} - \theta\|_W$, which can be further reduced to $\arg\min_\theta R_n(\mathbf{x} - \theta)$. Hence, we define an artificial likelihood function

$$L_R(\mathbf{x};\theta) \propto \exp\left\{-\frac{R_n(\mathbf{x}-\theta)}{\tau}\right\}, \tag{2.1}$$

where $\tau = \{\sqrt{12}\int g^2(x)dx\}^{-1}$ is the so-called Wilcoxon parameter with $g(x)$ being the density function of ε. The motivation for using (2.1) as an artificial-likelihood is two-fold: it can be shown that under the null hypothesis $\theta = \theta_0$, the test statistic $2[\log\{L_R(\mathbf{x};\theta)\} - \log\{L_R(\mathbf{x};\theta_0)\}]$ is asymptotically χ^2 distributed, as for the likelihood ratio test; the loss function $R_n(\mathbf{x} - \theta)$ is analogous to that in the least-squares procedure except that the Euclidean norm is substituted by the rank-based L_1 norm.

Remark 1. To test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ in the linear regression model $Y_i = \boldsymbol{X}_i\boldsymbol{\beta}+\varepsilon_i$, $i = 1, 2, \ldots, n$, McKean and Hettmansperger (1976) considered a counterpart of (2.1),

$$L_W(\beta) \propto \exp\left\{-\frac{W_n(\beta)}{\tau}\right\}. \tag{2.2}$$

They demonstrated that the $L_W(\beta)$-based test methods are powerful and efficient when the error distribution has heavier tails and possesses satisfactory robustness in general. However, (2.2) is not applicable in the estimation of nonparametric regression functions because the intercept term does not affect the value of $L_W(\beta)$. If we assume that the error distribution possesses certain symmetric properties (see Condition (A4$'$)), using (2.1) instead would yield robust estimation as illustrated later; see Feng, Zou, and Wang (2012), Shang, Zou, and Wang (2012) for more details.

We consider a more general case than the symmetric assumption. We assume $P(\varepsilon_1 + \varepsilon_2 \geq 0) = 1/2$, and re-define $\tau = \varphi^{1/2}\left(\int g(-t)g(t)dt\right)^{-1}$, where $\varphi = \int(1/2 - G(-t))^2 dG(t)$. We can show that (2.1) is also an artificial likelihood function under this general case. Thus, according to (2.1), the Wilcoxon-type generalized likelihood under H_{0a} is

$$l_n(H_{0a}) = -\frac{\varphi^{-1/2}\tau^{-1}}{n+1}\sum\sum_{i\leq j}|\widehat{\varepsilon}_i^a + \widehat{\varepsilon}_j^a|,$$

where the residual under the null hypothesis is $\widehat{\varepsilon}_i^a = Y_i - \alpha_0(U_i) - \boldsymbol{A}_0(U_i)^T\boldsymbol{X}_i$. Similarly, the likelihood under H_{1a} is

$$l_n(H_{1a}) = -\frac{\varphi^{-1/2}\tau^{-1}}{n+1}\sum\sum_{i\leq j}|\widehat{\varepsilon}_i + \widehat{\varepsilon}_j|,$$

where $\widehat{\varepsilon}_i$ is the residual under the alternative hypothesis. Thus, in order to construct a GLR test, we need to use nonparametric smoothing estimators under the alternative hypothesis since (1.2) is fully nonparametric.

Let $e_i = Y_i - \boldsymbol{\beta}^T\boldsymbol{V}_i(u_0) - \boldsymbol{\gamma}^T\boldsymbol{Z}_i(u_0)$, $\boldsymbol{V}_i(u_0) = (1, (U_i - u_0)/h)^T$, $\boldsymbol{Z}_i(u_0) = (\boldsymbol{X}_i^T, (U_i - u_0)/h\boldsymbol{X}_i^T)^T$, $\boldsymbol{\beta} = (a, hb)^T$, and $\boldsymbol{\gamma} = (\boldsymbol{A}^*, h\boldsymbol{B}^T)^T$, where $a, b \in \mathbb{R}$, $\boldsymbol{A}^*$ and $\boldsymbol{B}$ are vectors of p-dimensions. For each given u_0, (2.1) leads to a local log-likelihood objective function at the given point u_0,

$$l(\boldsymbol{\beta},\boldsymbol{\gamma}) = -\frac{\varphi^{-1/2}\tau^{-1}}{n+1}\sum\sum_{i\leq j}|e_i + e_j|K_h(U_i - u_0)K_h(U_j - u_0), \tag{2.3}$$

where $K_h(\cdot) = K(\cdot/h)/h$ with K a symmetric kernel function and h a bandwidth. Take $\boldsymbol{\beta}(u_0) = (\alpha(u_0), h\alpha'(u_0))^T$ and $\boldsymbol{\gamma}(u_0) = (\boldsymbol{A}(u_0)^T, h\boldsymbol{A}'(u_0)^T)^T$. Then, the local maximum likelihood estimator, denoted by $(\widehat{\boldsymbol{\beta}}(u_0), \widehat{\boldsymbol{\gamma}}(u_0))$, is defined as $\arg\max l(\boldsymbol{\beta},\boldsymbol{\gamma})$ and the corresponding estimator of $(\alpha(u_0), \boldsymbol{A}(u_0))$ is denoted by $(\widehat{\alpha}(u_0), \widehat{\boldsymbol{A}}(u_0))$; it is termed the local Walsh average estimator by Feng, Zou, and Wang (2012) and Shang, Zou, and Wang (2012). Consequently, the residuals under H_{1a} are given by $\widehat{\varepsilon}_i = Y_i - \widehat{\alpha}(U_i) - \widehat{\boldsymbol{A}}(U_i)^T\boldsymbol{X}_i$.

The Wilcoxon-type GLR test statistic (WGLR) can be defined by

$$\lambda_{na} = l_n(H_{1a}) - l_n(H_{0a}) = \frac{\varphi^{-1/2}\tau^{-1}}{n+1}\sum_{i\leq j}\sum(|\hat{\varepsilon}_i^a + \hat{\varepsilon}_j^a| - |\hat{\varepsilon}_i + \hat{\varepsilon}_j|). \tag{2.4}$$

To establish the asymptotic distribution of λ_{na} under H_{0a}, we need some conditions.

(A1) The marginal density $f(u)$ of U is Lipschitz continuous and bounded away from 0. U has a bounded support Ω.

(A2) $\alpha(u)$ and $\boldsymbol{A}(u)$ have continuous second derivatives.

(A3) The function $K(t)$ is symmetric and bounded. The functions $t^3K(t)$ and $t^3K'(t)$ are bounded and $\int t^4K(t)dt < \infty$.

(A4) The random error ε has a distribution $G(\cdot)$ and finite Fisher information, $\int g^{-1}(x)[g'(x)]^2dx < \infty$, where $g(\cdot)$ is the density function of ε.

(A4′) $P(\varepsilon_1 + \varepsilon_2 \geq 0) = 1/2$.

(A5) $\boldsymbol{X}$ is bounded and $E(\boldsymbol{X}_i|U_i = u) = 0$. The $p \times p$ matrix $\boldsymbol{\Sigma}(u) = \begin{pmatrix} 1 & 0 \\ 0 & E(\boldsymbol{X}\boldsymbol{X}^T|U=u)\end{pmatrix}$ is invertible for each $u \in \Omega$. $\boldsymbol{\Sigma}(u)^{-1}$ and $\boldsymbol{\Sigma}(u)$ are both Lipschitz continuous.

(A6) The bandwidth $h \to 0, nh^{3/2} \to \infty$ and $nh^{9/2} \to 0$.

These conditions are similar to those in Fan, Zhang, and Zhang (2001). In particular, (A5) can be relaxed by using the method in Lemma 7 in Zhang and Gijbels (2003). Thus, the assumption that $\boldsymbol{X}$ is bounded can be replaced by $E\{\exp(c_0||\boldsymbol{X}||) < \infty\}$ for some positive constant c_0. Condition (A4′) is imposed for identifying the intercept term. This is analogous to assuming that $E(\varepsilon) = 0$ and Median$(\varepsilon) = 0$ for least-squares and least-absolute-deviations, respectively, in simple linear regression. Condition (A4′) is a more general than symmetric assumption. For a continuous variable ε which is asymmetric, there always exists a constant η so that $\varepsilon + \eta$ satisfies (A4′). Assuming $E(\alpha_0(U)) = 0$ or $\alpha_0(U)$ is not of interest in the hypotheses (as the cases given later), so it is not required. Otherwise, we may use either local Walsh average estimator along with (A4′) or least-squares or least-absolute-deviation estimators to identify the intercept term.

Theorem 1. *Suppose* (A1)−(A6) *and* (A4′) *hold. Then, under* H_{0a}, $\sigma_{na}^{-1}(\lambda_{na} - \mu_{na}) \xrightarrow{d} N(0,1)$, *where*

$$\mu_{na} = \frac{(p+1)|\Omega|}{h}\Big\{K(0) - \frac{1}{2}\int K^2(t)dt\Big\},$$
$$\sigma_{na}^2 = \frac{2(p+1)|\Omega|}{h}\int\Big\{K(t) - \frac{1}{2}K*K(t)\Big\}^2 dt,$$

and $K*K$ denotes the convolution of K.

Remark 2. By using a scale constant

$$r_K = \frac{K(0) - (1/2)\int K^2(t)dt}{\int (K(t) - (1/2)K*K(t))^2 dt},$$

we can see $r_K\lambda_{na} \sim \chi^2_{r_K\mu_{na}}$, where $\sim$ means approximation in a generalized sense (see Fan, Zhang, and Zhang (2001)). Hence, for the simple null hypothesis (1.2), the asymptotic distribution of λ_{n1} under H_0 is free of nuisance parameters and the Wilks phenomenon holds.

Remark 3. For simplicity, we use local smoothing techniques for estimating smooth functions. We believe any appropriate smoother would also work well. Zhang (2004) shows the equivalence of GLR-type regression tests based on spline and local polynomial smoothers. We expect that the asymptotic normality would be valid if spline smoothers were used in the construction of WGLR. We make no attempt to provide formal analysis but we think that such studies deserve future research.

2.2. Asymptotic power study

The power of the proposed test under contiguous alternatives of the form,

$$H'_{1a} : (\alpha(u), \boldsymbol{A}(u)^T)^T = (\alpha_0(u), \boldsymbol{A}_0(u)^T)^T + (nh)^{-1/2}\boldsymbol{G}(u)$$

can be approximated by using the next theorem, where $\boldsymbol{G}(u) = (g_1(u), \ldots, g_{(p+1)}(u))^T$ is a bounded function that has bounded second derivative. Take $\boldsymbol{W}_i = (1, \boldsymbol{X}_i^T)^T$ and $\mathbf{W} = (\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n)$.

Theorem 2. *Suppose* (A1)$-$(A6) *and* (A4$'$) *hold. Under* H'_{1a}, $(\lambda_{na} - \mu_{na} - d_{2na})/\sigma^*_{na} \xrightarrow{d} N(0,1)$, *where* $d_{2na} = (2h)^{-1}E\{\tau^{-2}\boldsymbol{G}(U)^T\mathbf{W}\mathbf{W}^T\boldsymbol{G}(U)\}$, *and* $\sigma^*_{na} = (\sigma^2_{na} + h^{-1}E\{\tau^{-2}\boldsymbol{G}(U)^T\mathbf{W}\mathbf{W}^T\boldsymbol{G}(U)\})^{1/2}$.

Given a false alarm rate α under the contiguous alternative specified in Theorem 2, the power of WGLR can be expressed as

$$\beta_{\text{WGLR}} = \Phi\left(-\frac{1}{\sqrt{1 + n\sigma_{na}^{-2}\tau^{-2}B(G)}}z_\alpha + \frac{0.5n\tau^{-2}B(G)}{\sqrt{\sigma^2_{na} + n\tau^{-2}B(G)}}\right),$$

where Φ is the distribution function of $N(0,1)$, z_α is the upper α quantile of $N(0,1)$ and $B(G)=(nh)^{-1}E(\boldsymbol{G}(U)^T\mathbf{W}\mathbf{W}^T\boldsymbol{G}(U))$.

According to Theorem 6 in Fan, Zhang, and Zhang (2001), we can show that the power of GLR, with the bandwidth h' under the same contiguous alternative, is

$$\beta_{\mathrm{GLR}} = \Phi\bigg(-\frac{1}{\sqrt{1+nh'/h\sigma_{na}^{-2}\sigma^{-2}B(G)}}z_\alpha+\frac{0.5n\sigma^{-2}B(G)}{\sqrt{h\sigma_{na}^2/h'+n\sigma^{-2}B(G)}}\bigg).$$

Comparing WGLR and GLR under general settings turns out to be difficult, and we consider three representative cases.

(1) If $nB(G)\gg\sigma_{na}^2$, then

$$\beta_{\mathrm{WGLR}}\approx\Phi\left(0.5\sqrt{n\tau^{-2}B(G)}\right) \text{ and } \beta_{GLR}\approx\Phi\left(0.5\sqrt{n\sigma^{-2}B(G)}\right).$$

Thus, the Pitman asymptotic relative efficiency of WGLR with respect to the GLR test is approximately σ^2/τ^2, $\mathrm{ARE(WGLR,GLR)}\approx\sigma^2/\tau^2$.

(2) If $nB(G)\ll\sigma_{na}^2$, then

$$\beta_{\mathrm{WGLR}}\approx\Phi\left(-z_\alpha+0.5n\sigma_{na}^{-1}\tau^{-2}B(G)\right),$$
$$\beta_{\mathrm{GLR}}\approx\Phi\bigg(-z_\alpha+0.5\sqrt{\frac{h'}{h}}n\sigma_{na}^{-1}\sigma^{-2}B(G)\bigg).$$

Accordingly, $\mathrm{ARE(WGLR,GLR)}\approx\tau^{-2}\sigma^2\sqrt{h/h'}$. Different bandwidth choices yield different AREs, and if $h'/h=\sigma^4/\tau^4$, $\mathrm{ARE(WGLR,GLR)}\approx 1$.

(3) If $h'=h$, then

$$\beta_{\mathrm{GLR}} = \Phi\bigg(-\frac{1}{\sqrt{1+n\sigma_{na}^{-2}\sigma^{-2}B(G)}}z_\alpha+\frac{0.5n\sigma^{-2}B(G)}{\sqrt{\sigma_{na}^2+n\sigma^{-2}B(G)}}\bigg).$$

And $\mathrm{ARE(WGLR,GLR)}=\sigma^2/\tau^2$.

The ARE σ^2/τ^2 is essentially the same as the ARE of the signed-rank Wilcoxon test in comparison with the t-test under symmetric error and it has a lower bound 0.864 (Hodges and Lehmann (1963)). The ARE is as high as 0.955 for the normal error distribution, and can be significantly higher than one for many heavier-tailed distributions (Hettmansperger and McKean (2010)). For instance, it is 1.5 for the double exponential distribution, and 1.9 for the t-distribution with three degrees of freedom.

2.3. Composite null hypothesis

We consider the case where null hypotheses depend on nuisance functions. We show that the asymptotic null distribution of our proposed WGLR test statistic is independent of nuisance functions. Write

$$\boldsymbol{A}_0(u) = \begin{pmatrix} \boldsymbol{A}_{10}(u) \\ \boldsymbol{A}_{20}(u) \end{pmatrix}, \ \boldsymbol{A}(u) = \begin{pmatrix} \boldsymbol{A}_1(u) \\ \boldsymbol{A}_2(u) \end{pmatrix}, \ \boldsymbol{X}_k = \begin{pmatrix} \boldsymbol{X}_k^{(1)} \\ \boldsymbol{X}_k^{(2)} \end{pmatrix},$$

where $\boldsymbol{A}_{10}(u), \boldsymbol{A}_1(u)$, and $\boldsymbol{X}_k^{(1)}$ are $p_1(< p)$-dimensional. First, we test the hypothesis

$$H_{0b} : \alpha = \alpha_0, \boldsymbol{A}_1 = \boldsymbol{A}_{10} \ \text{ versus } \ H_{1b} : \alpha \neq \alpha_0 \ \text{ or } \ \boldsymbol{A}_1 \neq \boldsymbol{A}_{10}, \tag{2.5}$$

with $\boldsymbol{A}_2(\cdot)$ completely unknown. Following the same derivation, the logarithm of the generalized likelihood ratio statistic is

$$\lambda_{nb} = l_n(H_{1b}) - l_n(H_{0b}) = \frac{\varphi^{-1/2}\tau^{-1}}{n+1} \sum\sum_{i \le j} (|\widehat{\varepsilon}_i^b + \widehat{\varepsilon}_j^b| - |\widehat{\varepsilon}_i + \widehat{\varepsilon}_j|), \tag{2.6}$$

where $\widehat{\varepsilon}_k^b = Y_k - \alpha_0(U_k) - \boldsymbol{A}_{10}(U_k)^T \boldsymbol{X}_k^{(1)} - \widehat{\boldsymbol{A}}_2^b(U_k)^T \boldsymbol{X}_k^{(2)}$ and $\widehat{\boldsymbol{A}}_2^b(U_k)$ is the local linear Walsh-average estimator at U_k when $(\alpha_0, \boldsymbol{A}_{10})$ is given. Let μ_{nb} and σ_{nb} be the same as μ_{na} and σ_{na}, except replacing p by p_1.

Theorem 3. *Suppose* (A1)$-$(A6) *and* (A4$'$) *hold. Then, under* H_{0b}, $\sigma_{nb}^{-1}(\lambda_{nb} - \mu_{nb}) \xrightarrow{d} N(0,1)$.

Next, we consider the hypothesis

$$H_{0c} : \boldsymbol{A}_2 = \boldsymbol{A}_{20} \ \text{ versus } \ H_{1c} : \boldsymbol{A}_2 \neq \boldsymbol{A}_{20}, \tag{2.7}$$

when both $\alpha(\cdot)$ and $\boldsymbol{A}_1(\cdot)$ are unknown. We show that the assumption of the symmetric error distribution is not required in this situation. We define an "asymmetric" constant c which satisfies $P(\varepsilon_i - c + \varepsilon_j - c \ge 0) = 1/2$, and $\delta = \int g(2c - t)g(t)dt$, $\psi = \int \{1/2 - G(2c - \varepsilon)\}^2 dG(\varepsilon)$. Here $c = 0$, $\varphi = \psi$ and $\tau^2 = \psi\delta^{-2}$ if the error is symmetrically distributed. When not symmetric, the artificial likelihood function (2.1) is modified to

$$L_R(\mathbf{x}; \theta) \propto \exp\{-\psi^{-1/2}\delta R_n(\mathbf{x} - \theta)\}. \tag{2.8}$$

Then, the generalized artificial log-likelihood under H_{0c} and H_{1c} is

$$l_n(H_{0c}) = -\frac{\psi^{-1}\delta}{n+1} \sum\sum_{i \le j} |\widehat{\varepsilon}_i^c + \widehat{\varepsilon}_j^c|, \quad l_n(H_{1c}) = -\frac{\psi^{-1}\delta}{n+1} \sum\sum_{i \le j} |\widehat{\varepsilon}_i + \widehat{\varepsilon}_j|,$$

respectively, where $\widehat{\varepsilon}_i^c = Y_i - \widehat{\alpha}^c(U_i) - \widehat{\boldsymbol{A}}_1^c(U_i)^T \boldsymbol{X}_i^{(1)} - \boldsymbol{A}_{20}(U_i)^T \boldsymbol{X}_i^{(2)}$ and $(\widehat{\alpha}^c(U_i), \widehat{\boldsymbol{A}}_1^c(U_i)^T)^T$ is the local linear Walsh-average estimator at U_i when $\boldsymbol{A}_{20}$ is given. Now, the WGLR statistic is

$$\lambda_{nc} = l_n(H_{1c}) - l_n(H_{0c}) = \frac{\psi^{-1}\delta}{n+1} \sum\sum_{i \le j} (|\widehat{\varepsilon}_i^c + \widehat{\varepsilon}_j^c| - |\widehat{\varepsilon}_i + \widehat{\varepsilon}_j|). \tag{2.9}$$

Let μ_{nc} and σ_{nc} be the same as μ_{na} and σ_{na}, except replacing $(p+1)$ by p_2.

Theorem 4. *Suppose* (A1)−(A6) *hold. Then, under* H_{0u}, $\sigma_{nc}^{-1}(\lambda_{nc} - \mu_{nc}) \xrightarrow{d} N(0,1)$.

Since $\widehat{\boldsymbol{A}}(\cdot)$ is consistent even without the symmetric assumption, (A4′) is not needed; see Shang, Zou, and Wang (2012). It seems that $\psi^{-1/2}\delta$ is necessary for constructing λ_{nc}, however, using the implementation suggested in the next subsection can circumvent this difficulty. Moreover, based on this theorem, we can readily extend our WGLR to model diagnostics under a wide range of distributions. See more details in Section 2.5.

2.4. Implementation

In specification testing problem the rate of convergence of the distribution of the test statistic is usually rather slow; see, e.g., Hall and Hart (1990), Zhang (2003), and Fan and Zhang (2004). For this reason, we propose the application of a resampling procedure based on the bootstrap (see Härdle and Mammen (1993)). The finite sample properties of the resulting tests are then investigated by means of a simulation study. For the hypothesis (1.2) for example, we obtain the bootstrap sample

$$Y_i^* = \alpha_0(U_i) + \boldsymbol{A}_0(U_i)^T\boldsymbol{X}_i + \varepsilon_i^*,$$

where $\{\varepsilon_i^*\}_{i=1}^n$ is a sample drawn from $\{\widehat{\varepsilon}_i\}_{i=1}^n$. A bootstrap test statistic λ_{na}^* is built from the bootstrap sample $\{(\boldsymbol{X}_i, U_i, Y_i^*)\}_{i=1}^n$, as was the original test statistic in (2.4). When this procedure is repeated many times, the bootstrap critical value z_α^* is the empirical $1-\alpha$ quantile of the bootstrap test statistic. Then the null hypothesis H_{0a} is rejected if $\lambda_{na} \geq z_\alpha^*$.

Theorem 5. *If* (A1)−(A6) *and* (A4′) *hold, then*

$$\sup_{z\in\mathbb{R}} |P(T_h^* \leq z|\{\boldsymbol{X}_i, U_i, Y_i, i=1,\ldots,n\}) - P(N(0,1) \leq z)| \xrightarrow{p} 0,$$

where $T_h^* = \sigma_{na}^{-1}(\lambda_{na}^* - \mu_{na})$.

Here the Wilcoxon parameter τ and the constant φ do not matter in the bootstrap procedure because both λ_{na}^* and λ_{na} are using the same τ and φ. Thus, in practice, we do not need to estimate them in calculating the test statistics if we use the bootstrap method to compute p-values rather than using the asymptotic distribution.

Similarly, for the hypothesis (2.5) or (2.7), the bootstrap sample can be generated by

$$\begin{aligned} Y_i^* =& \alpha_0(U_i) + \boldsymbol{A}_{10}(U_i)^T\boldsymbol{X}_i^{(1)} + \widehat{\boldsymbol{A}}_2^b(U_i)^T\boldsymbol{X}_i^{(2)} + \varepsilon_i^*, \\ Y_i^* =& \widehat{\alpha}^c(U_i) + \widehat{\boldsymbol{A}}_1^c(U_i)^T\boldsymbol{X}_i^{(1)} + \boldsymbol{A}_{20}(U_i)^T\boldsymbol{X}_i^{(2)} + \varepsilon_i^*, \end{aligned}$$

respectively. The effectiveness of this bootstrap method is studied by simulation in Section 3.1.

2.5. Model diagnostics

Consider the composite null hypothesis testing problem

$$H_{0d}:(\alpha,\boldsymbol{A}^T)\in\mathcal{A}_0 \quad \text{versus} \quad H_{1d}:(\alpha,\boldsymbol{A}^T)\notin\mathcal{A}_0, \tag{2.10}$$

where $\mathcal{A}_0$ is a set of functions. The WGLR test statistic can be constructed as (2.9) and as in the proof of Theorem 4, its asymptotic representation can be accordingly derived.

2.5.1. Testing linearity

Consider the nonparametric regression model $y=m(x)+\varepsilon$, and the testing linearity problem (Zheng (1996); Wang and Qu (2007))

$$H_{0g}:m(x)=\alpha_0+\alpha_1 x \quad \text{versus} \quad H_{1g}:m(x)\neq\alpha_0+\alpha_1 x, \tag{2.11}$$

where α_0,α_1 are unknown parameters. The WGLR test statistic of (2.11) is

$$\lambda_{ng}=\frac{\psi^{-1}\delta}{n+1}\sum\sum_{i\leq j}(|\widehat{\varepsilon}_i^g+\widehat{\varepsilon}_j^g|-|\widehat{\varepsilon}_i+\widehat{\varepsilon}_j|),$$

where $\widehat{\varepsilon}_i^g=Y_i-\widehat{\alpha}_0-\widehat{\alpha}_1X_i$, $\widehat{\varepsilon}_i=Y_i-\widehat{m}(X_i)$, $\widehat{m}(\cdot)$ is the local linear Walsh-average estimator for $m(\cdot)$ (Feng, Zou, and Wang (2012)), and

$$(\widehat{\alpha}_0,\widehat{\alpha}_1)=\arg\min\frac{2}{n(n+1)}\sum\sum_{i\leq j}|Y_i+Y_j-2\alpha_0-\alpha_1(X_i+X_j)|.$$

Corollary 1. *Suppose* (A1)–(A6) *hold. Then, under* H_{0g}, $\sigma_{ng}^{-1}(\lambda_{ng}-\mu_{ng})\xrightarrow{d} N(0,1)$, *where* μ_{ng} *and* σ_{ng} *are the same as* μ_{na}, *and* σ_{na}, *except* $p=0$.

When the parametric function $m_{\boldsymbol{\theta}}(x)$ in the null hypothesis in (2.11) is not linear/polynomial (with parameter $\boldsymbol{\theta}$), a local linear/polynomial fit results in a biased estimate under the null hypothesis. This bias problem can be significantly attenuated by the bias-correction advocated by Fan and Zhang (2004) and Fan and Jiang (2005). We reparameterize the unknown functions as $m^*(x)=m(x)-m_{\widehat{\boldsymbol{\theta}}}(x)$ and then change the null hypothesis into $H_{0g}':m^*(x)=0$. This also applies as well to the proposed procedure.

2.5.2. Testing homogeneity

A natural question in applying (1.1) is whether the coefficient functions are really varying (Fan, Zhang, and Zhang (2001)). This amounts to testing

$$\begin{aligned}&H_{0h}:\alpha(U)=\beta_0, a_i(U)=\beta_i, \quad i=1,\ldots,p \quad \text{versus}\\ &H_{1h}:\alpha(U)\neq\beta_0 \ \text{ or } \ a_1(U)\neq\beta_1 \ \text{ or } \ a_2(U)\neq\beta_2 \ \ldots \text{ or } \ a_p(U)\neq\beta_p,\end{aligned} \tag{2.12}$$

where $\beta_i, i = 0, \ldots, p$ are some unknown parameters. We define the residuals $\widehat{\varepsilon}_i^h = Y_i - \widehat{\beta}_0 - \boldsymbol{X}_i(\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^T$, where $(\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)$ are the solutions of

$$\arg\min \frac{2}{n(n+1)} \sum\sum_{i \leq j} |Y_i + Y_j - 2\beta_0 - (\boldsymbol{X}_i + \boldsymbol{X}_j)(\beta_1, \ldots, \beta_p)^T|.$$

The WGLR test statistic for (2.12) is

$$\lambda_{nh} = \frac{\psi^{-1}\delta}{n+1} \sum\sum_{i \leq j} (|\widehat{\varepsilon}_i^h + \widehat{\varepsilon}_j^h| - |\widehat{\varepsilon}_i + \widehat{\varepsilon}_j|),$$

where $\widehat{\varepsilon}_i$ is exactly the same as that defined in (2.4).

Corollary 2. *Suppose* (A1)$-$(A6) *hold. Then, under* H_{0h}, $\sigma_{na}^{-1}(\lambda_{nh} - \mu_{na}) \xrightarrow{d} N(0, 1)$.

Corollaries 1 and 2 do not need (A4$'$) is not needed because in them the intercept term is a nuisance parameter and would be "profiled" out, as in the construction of λ_{nc}. Due to the slow convergence rate of the distribution of test statistic, we also adopt the bootstrap method in Section 2.4 to control the empirical size. For hypotheses (2.11) and (2.12), we generate the bootstrap sample by

$$Y_i^* = \widehat{\alpha}_0 + \widehat{\alpha}_1 X_i + \widehat{\varepsilon}_i^*,$$
$$Y_i^* = \widehat{\beta}_0 + \boldsymbol{X}_i(\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^T + \widehat{\varepsilon}_i^*,$$

respectively, where $\widehat{\varepsilon}_i^*$ is a bootstrap sample drawn from the corresponding nonparametric residuals $\{\widehat{\varepsilon}_i\}_{i=1}^n$. As Theorem 5, we can establish the convergence of these bootstrap approximations.

2.6. Heteroscedasticity case

We extend our WGLR test to the heteroscedasticity case. Here we only consider the problem (1.2) with the varying-coefficient model

$$Y = \alpha(U) + \boldsymbol{A}(U)^T \boldsymbol{X} + \varrho(U)\varepsilon,$$

where $\varrho(U)$ is a positive continuous function and $\int_{u \in \Omega} \varrho^2(u) du = 1$. Let $\rho(u) = \varrho(u)\tau$ and then

$$\frac{Y_i - \alpha(U_i) - \boldsymbol{A}(U_i)^T \boldsymbol{X}_i}{\rho(U_i)} = \frac{\varrho(U_i)\varepsilon_i}{\rho(U_i)} = \tau^{-1}\varepsilon_i.$$

Thus, the WGLR test statistic in the heteroscedasticity case can be defined as

$$\lambda_{na} = l_n(H_{1a}) - l_n(H_{0a}) = \frac{\varphi^{-1/2}}{n+1} \sum\sum_{i \leq j} \left(\left| \frac{\widehat{\varepsilon}_i^a}{\widehat{\rho}(U_i)} + \frac{\widehat{\varepsilon}_j^a}{\widehat{\rho}(U_j)} \right| - \left| \frac{\widehat{\varepsilon}_i}{\widehat{\rho}(U_i)} + \frac{\widehat{\varepsilon}_j}{\widehat{\rho}(U_j)} \right| \right),$$

where $\widehat{\rho}(u)$ is a ratio-consistent estimator of $\rho(u)$. We suggest the use of

$$\widehat{\rho^{-1}}(u) = \frac{\varphi^{-1/2}}{2n^2 t_n \widehat{f}^2(u)} \sum_{i=1}^{n} \sum_{j=1}^{n} I(|\widehat{\varepsilon}_i + \widehat{\varepsilon}_j| \le t_n) K_h(U_i - u) K_h(U_j - u), \quad (2.13)$$

where $\widehat{f}(u) = n^{-1} \sum_{i=1}^{n} K_h(U_i - u)$. To establish the asymptotic distribution of λ_{na}, we need additional assumptions.

(A7) The conditional variance function $\varrho(u)$ is continuous for all $u \in \Omega$.

(A8) The bandwidth t_n satisfies $t_n \to 0$ and $nh^2 t_n \to \infty$.

Proposition 1. *If* (A1)$-$(A8) *and* (A4$'$) *hold, we have* $\widehat{\rho}(u) \xrightarrow{p} \rho(u)$, $u \in \Omega$.

Re-define d_{2na}, σ_{na}^* in Theorem 2 by replacing τ^2 by $\rho^2(U_i)$.

Theorem 6. *If* (A1)$-$(A8) *and* (A4$'$) *hold,*

(i) *under* H_{0a}, *we have* $(\lambda_{na} - \mu_{na})/\sigma_{na} \xrightarrow{d} N(0,1)$;

(ii) *under* H_{1a}, *we have* $(\lambda_{na} - \mu_{na} - d_{2na})/\sigma_{na}^* \xrightarrow{d} N(0,1)$.

Theorem 6 shows that the Wilks phenomenon still holds for λ_{na} in the heteroscedasticity case. The null distribution of λ_{na} is also independent of the variance function $\varrho(u)$. Here we also adopt the bootstrap method to determine the critical values. In this case, the bootstrap sample is generated by

$$Y_i^* = \widehat{\alpha}(U_i) + \widehat{\boldsymbol{A}}(U_i)^T \boldsymbol{X}_i + \widehat{\rho}(U_i)\varepsilon_i^*,$$

where $\{\varepsilon_i^*\}_{i=1}^n$ is a bootstrap sample drawn from $\{\widehat{e}_i\}_{i=1}^n$, $\widehat{e}_i = (Y_i - \widehat{\alpha}(U_i) - \widehat{\boldsymbol{A}}(U_i)^T \boldsymbol{X}_i)/\widehat{\rho}(U_i)$.

3. Simulation Study

Throughout this section, for each experiment we run 1,000 replications with bootstrap resampling number set to 400. We take nominal level 5% in each case. Some results are provided in an on-line supplemental file.

3.1. Testing a simple hypothesis

Consider the simple nonparametric regression model $Y_i = m(X_i) + \varepsilon$, where X_i are uniformly distributed with the testing problem

$$H_0 : m(x) = 0 \quad \text{versus} \quad H_1 : m(x) \neq 0. \quad (3.1)$$

Three symmetric distributions of errors are considered: (a) $\varepsilon_i \sim N(0,1)$; (b) $\varepsilon_i \sim t(3)$; (c) $\varepsilon_i \sim T(0.05, 10)$. Here $t(3)$ and $T(0.05; 10)$ denote the standardized

Student-t distribution with four degrees of freedom and the standardized Tukey's contaminated normal model, respectively. The empirical sizes of our WGLR test are summarized in Table 1 for sample sizes, $n = 25, 50$, and 100, with the smoothing parameters $h = 0.12, 0.15, 0.18$, and 0.21. From Table 1, we observe that all the sample sizes are close to the specified nominal level in most cases.

Next, we compare the power of WGLR with that of other tests. The GLR test suggested by Fan, Zhang, and Zhang (2001) is a natural benchmark. In addition, Zheng's (1996) and Wang and Qu's (2007) tests are included in this comparison. Hong and Lee (2013) proposed a test based on loss functions, which measure discrepancies between the null and nonparametric alternative models. Their test, abbreviated as LOSS, is asymptotically more powerful than the GLR test in terms of Pitman's efficiency criterion. Pitman's AREs of LOSS with respect to GLR are reported in Table A.1. The bandwidth h is of the order of $n^{-2/9}$ as in Hong and Lee (2013). The ARE values in this table are different from those in Table 1 of Hong and Lee (2013); there are some errors in their calculations. We observe that the WGLR is asymptotically more powerful than LOSS for the heavy-tailed distributions. To gain more insight, we include LOSS in this finite-sample comparison. The two parameters in the linex function are chosen as 1 because Hong and Lee (2013) demonstrated that the power of LOSS is not sensitive to the choice of parameters.

For a fair comparison, we performed a size-corrected power comparison in the sense that the actual critical values are found through simulations so that all the five tests had accurate sizes in each case. In order to study the power of each test, we considered different function forms

$$\text{(I)} \quad m(x) = 0.3; \ \text{(II)}\ m(x) = 0.7x; \ \text{(III)}\ m(x) = 0.7\cos(3\pi x);$$
$$\text{(IV)}\ m(x) = 0.7\sin(2\pi x); \ \text{(V)}\ m(x) = \frac{(\sin(3\pi x) + x)}{2}; \ \text{(VI)}\ m(x) = \frac{\exp(x)}{4}.$$

We chose the Gaussian kernel and fixed the bandwidth $h = 0.15$ for all the tests. To avoid the effect of degrees of local polynomial fit, the local linear smoother was employed in both GLR and LOSS. The simulated power with different errors and sample sizes $n = 25, 50$ is displayed in Table 2. We observe that the WGLR test generally performs better than the least square-based methods, the GLR, LOSS and Zheng's methods, when the error deviates from the normal. Even in the normal cases, the WGLR has similar performance to those tests. The WQ test cannot detect the constant shift of the function, as we have mentioned before (the results for Scenario I). Under Scenarios (II), (V) and (VI), the WQ test is generally outperformed by the other three tests; the others have asymptotic power depending on $E[m(x)]^2$, where the asymptotic power of WQ test is a function of $E\{m(x) - E[m(x)]\}^2$ (Theorem 3 in Wang and Qu (2007)).

Table 1. Simulated level (%) of test on testing (3.1) with homoscedastic error.

	$n=25$			$n=50$			$n=100$		
h	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
0.12	8.2	7.8	6.0	7.4	6.5	5.3	4.3	5.1	4.7
0.15	7.9	6.2	5.3	6.3	7.0	6.8	5.2	6.1	4.7
0.18	5.3	5.3	4.3	6.0	5.3	5.7	4.9	5.8	4.8
0.21	6.0	5.7	4.0	5.5	5.8	6.2	5.3	4.6	5.7

Table 2. Empirical power (%) of tests on testing (3.1) with homoscedastic error. WQ, Zheng and LOSS stand for the tests of Wang and Qu (2007), Zheng (1996), and Hong and Lee (2013), respectively.

		$n=25$					$n=50$				
	Models	WGLR	GLR	WQ	Zheng	LOSS	WGLR	GLR	WQ	Zheng	LOSS
$N(0,1)$	(I)	24	19	5.6	25	28	46	35	4.9	45	47
	(II)	38	28	14	35	45	67	58	23	65	68
	(III)	27	28	26	15	20	52	58	53	33	34
	(IV)	34	36	53	41	31	67	76	87	77	65
	(V)	45	37	23	38	52	78	71	44	70	77
	(VI)	47	36	8.9	43	54	79	68	11	75	80
$t(3)$	(I)	45	25	6.3	34	32	74	42	7.2	53	50
	(II)	60	41	20	45	52	92	61	43	73	75
	(III)	48	40	35	23	24	87	69	83	47	46
	(IV)	58	51	67	59	47	92	78	98	84	76
	(V)	71	47	35	52	59	96	79	73	78	80
	(VI)	75	47	09	60	60	98	73	20	83	84
$T(0.05,10)$	(I)	82	40	6.3	60	53	100	55	7.1	68	60
	(II)	95	54	45	71	66	100	67	75	80	73
	(III)	88	53	74	46	44	100	70	98	69	52
	(IV)	96	62	97	74	65	100	76	100	85	74
	(V)	97	60	68	70	71	100	76	97	83	82
	(VI)	98	60	20	78	73	100	72	37	84	84

In Appendix D of the supplemental file, we report on a simulation study under the same settings for the heteroscedasticity case. The comparison conclusions are similar to the findings above.

3.2. Testing linearity

Consider the linear model

$$Y_i = 1 + 2X_i + \varepsilon_i, i = 1, \ldots, n,$$

where the covariate X_i's are uniformly distributed on $(0,1)$. Four distributions of errors were considered: (a) $N(0,1)$; (b) $t(3)$; (c) $LN(0,1)$; (d) $T(0.05,10)$, where $LN(0, 1)$ denotes the standardized log-normal distribution. The empirical sizes of the WGLR test are summarized in Table A.4 for $n = 60$, and 100, and smoothing parameters $h = 0.06, 0.09, 0.12$, and 0.15. The empirical sizes are slightly larger than the nominal level when the bandwidth h is small, but are quite close to the nominal level when the bandwidth h gets larger. In general, we observe a reasonable approximation of the level by the bootstrap procedure in all cases. The levels of test are insensitive to the distribution of errors.

Consider the alternative hypotheses

(I) Logarithm Alternative. $Y_i = 1 - X_i + \theta \log(X_i) + \varepsilon_i$;

(II) Square Alternative. $Y_i = 1 - X_i + \theta\sqrt{X_i} + \varepsilon_i$.

The size-corrected power curves against θ for these alternative models with $n = 60$ and $h = 0.09$ are displayed in Figure 1 and Figure A.1 in Appendix E, respectively. Our proposed test generally has better efficiency than the GLR, LOSS, and Zheng's tests for the non-normal cases. The WQ test has some advantage oven WGLR for the asymmetric error (lognormal), while the WGLR performs better in the three symmetric cases.

3.3. Testing homogeneity

Consider the varying-coefficient model:

$$Y_i = X_{1i}\beta_1(U_i) + X_{2i}\beta_2(U_i) + X_{3i}\beta_3(U_i) + \varepsilon_i,$$

where U_i is uniform, the $\boldsymbol{X}_i = (X_{1i}, X_{2i}, X_{3i})$ are the multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = (0.5^{|i-j|})_{1\le i,j\le 3}$, and

$$\beta_1(U_i) = 1 + \theta U_i^2, \quad \beta_2(U_i) = 1.5, \quad \beta_3(U_i) = -0.5 + \theta\sqrt{U_i}.$$

Table A.5 reports the empirical sizes ($\theta = 0$) of WGLR with the same settings as in Table A.4. Figures 2 and Figure A.2 report the simulated size-corrected power curves of the WGLR and GLR tests for different errors with $n = 60$, and 100, respectively. The performance of bootstrap procedure in testing homogeneity is similar to that in testing linearity. The actual level can approximately attain the nominal level in most cases; there is some deviation when the bandwidth is very small. The simulated level is insensitive to the distribution of errors, which again demonstrates the robustness of the proposed procedure. Regarding power, the WGLR test performs better than the GLR test for the non-normal distributions, as expected.

These results suggest that the WGLR test is quite robust and efficient in nonparametric testing.

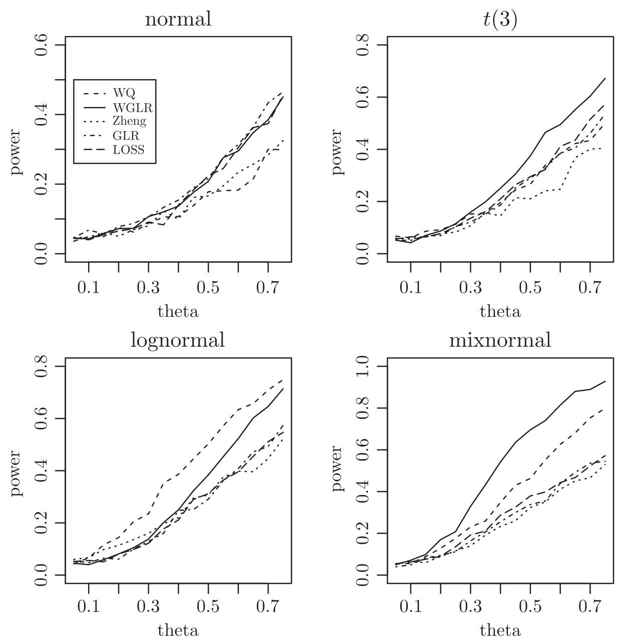

Figure 1. Simulated power curves of logarithm alternative on testing linearity. The legend in the first plot is applicable for all the others.

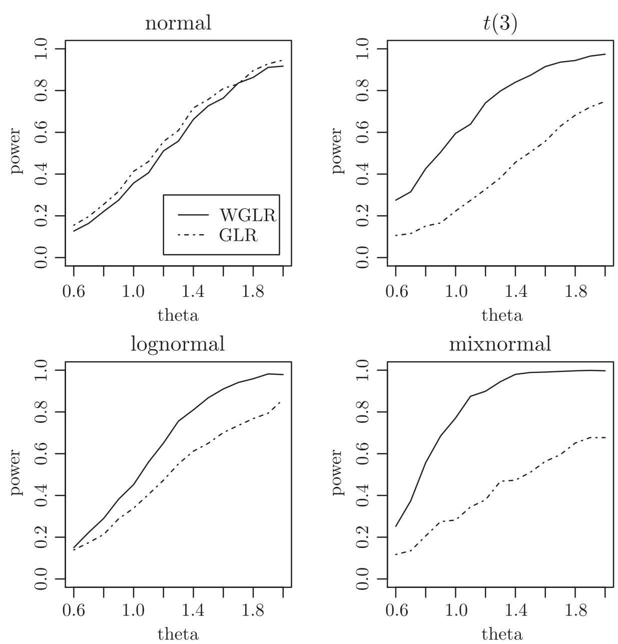

Figure 2. Simulated power curves on testing homogeneity with $n = 60, h = 0.09$.

Table 3. Test results for plasma beta-carotene level data.

Covariate	GLR	WGLR
dietary beta-carotene	0.000 (+)	0.000 (+)
Sex	0.265 (−)	0.048 (+)
Quetelet index	0.001 (+)	0.000 (+)
Calories	0.649 (−)	0.000 (+)
Fat	0.053 (−)	0.017 (+)
Fiber	0.037 (+)	0.000 (+)
Alcohol	0.593 (−)	0.689 (−)
Cholesterol	0.507 (−)	0.002 (+)
Smoking status (1=never)	0.261 (−)	0.000 (+)
Smoking status (2=former)	0.148 (−)	0.016 (+)
Vitamin use (1=yes, fairly often)	0.001 (+)	0.000 (+)
Vitamin use (2=yes, not often)	0.346 (−)	0.078 (−)

Note: "+" means rejection and"–" means acceptance at %5 level.

4. A Data Application

We applied the proposed methodology to the plasma beta-carotene level data set collected by a cross-sectional study (Nierenberg et al. (1989)). The data is available from the StatLib database via the link `lib.stat.cmu.edu/datasets/Plasma_Retinol`. Of interest are the relationships between the plasma beta-carotene level (Y) and the covariates ($\boldsymbol{X}$) listed in Table 3. We fit the data by using the varying coefficient model with U being "*Age*". The covariates "*smoking status*" and "*vitamin use*" are categorical and are thus replaced with dummy variables. The covariates, Y and U are standardized. Figure A.3 shows the normal QQ-plot of residuals obtained by using local linear Walsh-average estimation (Shang, Zou, and Wang (2012)). This figure clearly indicates that the errros are not normal. Then, we applied GLR and our WGLR to test whether each varying coefficient function is zero, respectively.

We adopted the Epanechnikov kernel and the bandwidth was $h = 2.38 \times \mathrm{sd}(U) \times n^{-2/9}$. Table 3 reports the test results at the significant level 0.05. Both tests found that the varying coefficient functions of *dietary beta-carotene, Quetelet index, Fiber, Vitamin use* (*1=yes, fairly often*) are not zero. However, our WGLR test also suggests that *Sex, Calories, Fat, Cholesterol* and *Smoking status* are important variables. To confirm whether the selected covariates are truly relevant, we plotted the estimators of each function and their 95% pointwise confidence intervals in Figure A.4. We found that the confidence intervals of those covariates selected by WGLR do not completely cover 0. This result may reflect, to a degree, that our WGLR would be more powerful and robust than GLR for the heavy-tailed distributions.

5. Discussion

The data-driven bandwidth methods that are well suited for producing visually smooth estimates of the underlying curves may not in general be appropriate for the testing problem. Some efforts have been devoted to construct "semi-data-driven" nonparametric methods, such as Horowitz and Spokoiny (2001). An ongoing effort of the authors is to develop a method integrating a "data-driven" adaptive smoothing parameter selection method to make the test nearly optimal in a certain sense. We can extend the WGLR test to some other popular nonparametric models, such as additive and single-index models.

Acknowledgement

The authors thank the Editor, an associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements. Feng, Wang and Zou were supported by the NNSF of China Grants 11431006, 11131002, 11371202 and the Foundation for the Author of National Excellent Doctoral Dissertation of PR China 201232. Zhang's research is supported by the US NSF grants DMS–1106586 and DMS–1308872, and Wisconsin Alumni Research Foundation.

Supplementary Materials

This file provides the proofs of the asymptotic theorems mentioned in the paper, and some additional simulation results.

References

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

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

Fan, J., Hu, T. and Truong, Y. (1994). Robust nonparametric function estimation. *Scand. J. Statist.* **21**, 433-446.

Fan, J. and Jiang, J. (2005). Nonparametric inference for additive models. *J. Amer. Statist. Assoc.* **100**, 890-907.

Fan, J. and Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests. *Test* **16**, 409-444.

Fan, J. and Zhang, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *Ann. Statist.* **5**, 1858-1907.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153-193.

Feng, L., Zou, C. and Wang, Z. (2012). Local Walsh-average regression. *J. Multivariate Anal.* **106**, 36-48.

Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.* **85**, 1039-1049.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926-1947.

Hettmansperger, T. P. and McKean, J. W. (2010). *Robust Nonparametric Statistical Methods.* Arnold, London.

Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34**, 598-611.

Hong, Y. M. and Lee, Y. J. (2013). A loss function approach to model specification testing and its relative efficiency. *Ann. Statist.* **41**, 1166-1203.

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

Kai, B., Li, R. and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *J. Roy. Statist. Soc. B* **72**, 49-69.

McKean, J. W. and Hettmansperger, T. P. (1976). Tests of hypotheses of the general linear models based on ranks. *Comm. Statist. Theory Methods* **5**, 693-709.

Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.* **101**, 1067-1078.

Nierenberg, D., Stukel, T., Baron, J., Dain, B. and Greenberg, E. (1989). Determinants of plasma levels of beta-carotene and retinol. *Amer. J. Epidemiology* **130**, 511-521.

Shang, S., Zou, C. and Wang, Z. (2012). Local Walsh-average regression for semiparametric varying-coefficient models. *Statist. Probab. Lett.* **82**, 1815-1822.

Wang, L. and Qu, A. (2007). Robust test in regression models with omnibus alternative and bounded influence. *J. Amer. Statist. Assoc.* **102**, 347-358.

Welsh, A. H. (1996). Robust estimation of smooth regression and spread functions and their derivatives. *Statist. Sinica* **6**, 347-366.

Zhang, C. (2003). Calibrating the degrees of freedom for automatic data smoothing and effective curve checking. *J. Amer. Statist. Assoc.* **98**, 609-628.

Zhang, C. (2004). Assessing the equivalence of nonparametric regression tests based on spline and local polynomial smoothers. *J. Statist. Plann. Inference* **126**, 73-95.

Zhang, C. and Dette, H. (2004). A power comparison between nonparametric regression tests. *Statist. Probab. Lett.* **66**, 289-301.

Zhang, J. and Gijbels, I. (2003). Sieve empirical likelihood and extensions of the generalized least squares. *Scand. J. Statist.* **30**, 1-24.

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

Institute of Statistics and LPMC, Nankai University, Tianjin, 300071, China.

E-mail: flnankai@126.com

Institute of Statistics and LPMC, Nankai University, Tianjin, 300071, China.

E-mail: zjwang@nankai.edu.cn

Department of Statistics, University of Wisconsin–Madison, Wisconsin, 53706, USA.

E-mail: cmzhang@stat.wisc.edu

Institute of Statistics and LPMC, Nankai University, Tianjin, 300071, China.

E-mail: nk.chlzou@gmail.com

(Received March 2014; accepted January 2015)

Supplementary Material to "NONPARAMETRIC TESTING IN REGRESSION MODELS WITH WILCOXON-TYPE GENERALIZED LIKELIHOOD RATIO"

Long Feng[1], Zhaojun Wang[1], Chunming Zhang[2], Changliang Zou[1],

[1] *Nankai University and* [2] *University of Wisconsin–Madison*

This supplemental file contains some more discussions on the proposed WGLR test, all the technical proofs and some additional simulation results. The following materials are included:

- Appendix A: Some discussions on two modifications of WGLR tests
- Appendix B: Proof of Theorem 6
- Appendix C: Proofs of other theorems
- Appendix D: Simulation results in heteroscedasticity cases
- Appendix E: Some additional simulation results in Section 3
- Appendix F: Some additional figures in Section 4

Appendix

Appendix A: Some discussions on two modifications of WGLR tests

In constructing λ_{na}, both the "likelihood" function and local smoother are of Wilcoxon-type and come from (2.2). If we replace the local linear estimators in (1.3) by the local linear Walsh-average estimators and denote the resulting test statistic as ω_{na}, i.e.,

$$\omega_{na} = \frac{n}{2}\left(\log\sum_{i=1}^{n}\widehat{\varepsilon}_i^{a2} - \log\sum_{i=1}^{n}\widehat{\varepsilon}_i^{2}\right) \tag{A.1}$$

Similar to Theorems 1-2, we can establish the asymptotic normality of ω_{na}.

Theorem 7 *(i) Suppose the conditions in Theorem 2 hold. Under H_0, we have*

$$(\omega_{na} - \mu_{\omega na})/\sigma_{\omega na} \xrightarrow{d} N(0,1),$$

where

$$\begin{aligned}\mu_{\omega na} =& \frac{1}{h}(p+1)|\Omega|\left(K(0)\sigma^{-2}\zeta^2 - \frac{1}{2}\int K^2(t)dt\right), \zeta^2 = \varphi^{-1/2}\tau\int xG(-x)dG(x),\\ \sigma^2_{\omega na} =& 2h^{-1}\sigma^{-2}\tau^2(p+1)|\Omega|\Bigg(\int K^2(t)dt - \sigma^{-2}\zeta^2\int K(t)K*K(t)dt\\ &+\frac{1}{4}\sigma^{-2}\tau^2\int(K*K)^2(t)dt\Bigg);\end{aligned}$$

(ii) Suppose the conditions in Theorem 2 hold. under H'_{na}, we have

$$[\omega_{na} - \mu_{\omega na} - \sigma^{-2}\tau^2 d_{2na}]/\sigma^*_{\omega na} \xrightarrow{d} N(0,1),$$

where $\sigma^{*2}_{\omega na} = \sigma^2_{\omega na} + \sigma^{-2}h^{-1}E[\boldsymbol{G}(U)^T\mathbf{W}\mathbf{W}^T\boldsymbol{G}(U)]$.

This theorem implies that the power of ω_{na} is

$$\beta_{\omega_{na}} = \Phi\left(-\frac{\sigma_{\omega na}}{\sqrt{\sigma^2_{\omega na} + n\sigma^{-2}B(G)}}z_\alpha + \frac{2^{-1}n\sigma^{-2}B(G)}{\sqrt{\sigma^2_{\omega na} + n\sigma^{-2}B(G)}}\right).$$

It is difficult to calculate the ARE of ω_{na} with respect to the GLR test under the general cases. For convenience, we choose the same bandwidth for ω_{na} and GLR and then

$$\begin{aligned}&\mathrm{ARE}(\omega_{na}, \mathrm{GLR})\\ &= \frac{\sigma\tau^{-1}\left(\int\left\{K(t) - \frac{1}{2}K*K(t)\right\}^2 dt\right)^{1/2}}{\left(\int K^2(t)dt - \sigma^{-2}\zeta^2\int K(t)K*K(t)dt + \frac{1}{4}\sigma^{-2}\tau^2\int(K*K)^2(t)dt\right)^{1/2}}.\end{aligned}$$

Table A.1 shows the ARE of ω_{na}, LOSS (Hong and Lee 2013) and WGLR with respect to GLR for a number of distributions and kernel functions. We observe that ARE(ω_{na}, GLR)'s are similar for different kernels and generally much smaller than ARE(WGLR, GLR) for heavy-tailed distribution. To a certain extent, ω_{na} can be viewed as some compromise between the GLR and WGLR tests. Moreover, if the local linear Walsh-average estimators in (2.4) are replaced by the local linear estimators, similar results to those in Table A.1 can be obtained.

Table A.1: The asymptotic efficiency comparisons of ω_{na} and WGLR. $t(d)$: student's t-distribution with d degrees of freedom. $T(\rho, \sigma)$:Tukey contaminated normal with CDF $F(x) = (1-\rho)\Phi(x) + \rho\Phi(x/\sigma)$ where $\rho \in [0, 1]$ is the contamination proportion.

	Epanechnikov	Biweight	Triweight	Gaussian	
		ARE(LOSS, GLR)			
	1.46	1.46	1.46	1.49	
Errors		ARE(ω_{na}, GLR)			ARE(WGLR, GLR)
$N(0,1)$	0.99	0.99	0.99	0.99	0.96
$t(3)$	1.16	1.16	1.16	1.16	1.90
$t(4)$	1.07	1.07	1.07	1.07	1.40
$T(0.05, 10)$	1.59	1.58	1.58	1.59	4.77
$T(0.10, 10)$	1.88	1.87	1.87	1.89	7.19

Appendix B: Proof of Theorem 6

Here we only provide the proof of Theorem 6. Obviously, Theorem 1 and 2 are the special cases of Theorem 6 with $\varrho^2(x) = 1$. Let $r_n = 1/\sqrt{nh}$. For ease of illustration, we need some notations:

$$\xi_i = \varphi^{-1/2}\rho(U_i)\{G(-\varepsilon_i) - 1/2\},\ w_0 = \int\int t^2(s+t)^2K(t)K(s+t)dtds,$$

$$R_{n10} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i(\alpha_0''(U_i) + \boldsymbol{A}_0''(U_i)^T\boldsymbol{X}_i)\int t^2K(t)dt(1+O(h)+O(n^{-1/2})),$$

$$R_{n20} = \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\alpha_0''(U_i)w_0.$$

$$R_{n30} = \frac{1}{8}E[\alpha_0''(U_i)^2 + \boldsymbol{A}_0''(U_i)^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{A}_0''(U_i)]w_0(1+O(n^{-1/2})),$$

$$\boldsymbol{\alpha}_n(u_0) = r_n^2\boldsymbol{\Gamma}(u_0)^{-1}\sum_{i=1}^n \eta_i\boldsymbol{W}_iK((U_i-u_0)/h), \boldsymbol{c} = (c, \boldsymbol{0}_{1\times p})^T, \boldsymbol{\Gamma}(u) = \boldsymbol{\Sigma}(u)f(u),$$

$$\boldsymbol{R}_n(u_0) = r_n^2\sum_{i=1}^n\boldsymbol{\Gamma}(u_0)^{-1}\Big(\alpha(U_i) - \boldsymbol{\beta}(u_0)^T\boldsymbol{V}_i(u_0) + \boldsymbol{A}(U_i)^T\boldsymbol{X}_i - \boldsymbol{\gamma}(u_0)^T\boldsymbol{Z}_i(u_0)\Big)\boldsymbol{W}_iK((U_i-u_0)/h),$$

$$R_{1n} = \sum_{k=1}^n \xi_k\boldsymbol{R}_n(U_k)^T\mathbf{W}_k/\rho^2(U_k),\ R_{2n} = \sum_{k=1}^n \boldsymbol{\alpha}_n(U_k)^T\mathbf{W}_k\mathbf{W}_k^T\boldsymbol{R}_n(U_k)/\rho^2(U_k),$$

$$R_{3n} = \frac{1}{2}\sum_{k=1}^n \boldsymbol{R}_n(U_k)^T\mathbf{W}_k\mathbf{W}_k^T\boldsymbol{R}_n(U_k)/\rho^2(U_k).$$

Lemma 1 *Let $\widehat{\boldsymbol{A}}$ be the local linear Walsh-average estimator. Then, under conditions (A1)-(A5), uniformly for $u_0 \in \Omega$,*

$$(\widehat{\alpha}(u_0), \widehat{\boldsymbol{A}}(u_0)^T)^T - (\alpha(u_0), \boldsymbol{A}(u_0)^T)^T = \boldsymbol{c} + (\boldsymbol{\alpha}_n(u_0) + \boldsymbol{R}_n(u_0))(1+o_p(1)),$$

and under condition (A4′), $c = 0$ and $\eta_i = \xi_i$, $i = 1, \ldots, n$.

Proof. From Shang *et al.* (2012), we can easily obtain the result. □

Proof of Theorem 6 By Proposition 1, the estimators $\widehat{\rho}_k(x)$ is consistent. Note that under H_{0a}, $\widehat{\varepsilon}_i^a = \varepsilon_i$. Thus, by Slutsky's theorem, we only need to consider the asymptotic property of

$$\begin{aligned}\lambda_{na} &= \frac{\varphi^{-1/2}}{n+1}\sum\sum_{i\le j}(|\tau^{-1}\varepsilon_i + \tau^{-1}\varepsilon_j| - |\rho^{-1}(U_i)\widehat{\varepsilon}_i + \rho^{-1}(U_j)\widehat{\varepsilon}_j|)\\ &= \frac{\varphi^{-1/2}}{n+1}\sum\sum_{i\le j}\Big(|\tau^{-1}\varepsilon_i + \tau^{-1}\varepsilon_j| - |\tau^{-1}\varepsilon_i + \tau^{-1}\varepsilon_j + \phi_{ij})|\Big),\end{aligned}$$

where $\phi_{ij} = \rho^{-1}(U_i)(\alpha(U_i)-\widehat{\alpha}(U_i)+\boldsymbol{A}(U_i)^T\boldsymbol{X}_i-\widehat{\boldsymbol{A}}(U_i)^T\boldsymbol{X}_i)+\rho^{-1}(U_j)(\alpha(U_j)-\widehat{\alpha}(U_j)+\boldsymbol{A}(U_j)^T\boldsymbol{X}_j-\widehat{\boldsymbol{A}}(U_j)^T\boldsymbol{X}_j)$. By using the identity

$$|z|-|z-y| = y\text{sgn}(z)+2(z-y)\{I(0<z<y)-I(y<z<0)\}$$

which holds for $z\neq 0$, we have

$$\begin{aligned}\lambda_{na} =&\Bigg(-\frac{\varphi^{-1/2}}{n+1}\sum\sum_{i\leq j}\phi_{ij}\text{sgn}(\varepsilon_i+\varepsilon_j)+\frac{2\varphi^{-1/2}\tau^{-1}}{n+1}\sum\sum_{i\leq j}(\tau^{-1}(\varepsilon_i+\varepsilon_j)+\phi_{ij})\\ &\qquad\times\Big\{I(0<\tau^{-1}(\varepsilon_i+\varepsilon_j)<-\phi_{ij}))-I(-\phi_{ij}<\tau^{-1}(\varepsilon_i+\varepsilon_j)<0)\Big\}\Bigg)\\ \doteq& A_h-B_h.\end{aligned}$$

Firstly, we consider A_h. By Lemma 1, we have

$$\begin{aligned}A_h =&\Bigg(-\frac{\varphi^{-1/2}}{n+1}\sum\sum_{i\leq j}\left(\boldsymbol{R}_n(U_i)^T\boldsymbol{W}_i/\rho(U_i)+\boldsymbol{R}_n(U_j)^T\boldsymbol{W}_j/\rho(U_j)\right)\text{sgn}(\varepsilon_i+\varepsilon_j)\\ &-\frac{\varphi^{-1/2}}{n+1}\sum\sum_{i\leq j}\left(\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i/\rho(U_i)+\boldsymbol{\alpha}_n(U_j)^T\boldsymbol{W}_j/\rho(U_j)\right)\text{sgn}(\varepsilon_i+\varepsilon_j)\Bigg)(1+o_p(1))\\ \doteq&(C_h+D_h)(1+o_p(1)).\end{aligned}$$

Firstly, note that D_h

$$\begin{aligned}&\frac{\varphi^{-1/2}}{n+1}\sum\sum_{i\leq j}\left(\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i/\rho(U_i)+\boldsymbol{\alpha}_n(U_j)^T\boldsymbol{W}_j/\rho(U_j)\right)\text{sgn}(\varepsilon_i+\varepsilon_j)\\ &=\sum_{i=1}^n\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i/\rho(U_i)\left(\sum\sum_{j\neq i}\frac{\varphi^{-1/2}}{n+1}\text{sgn}(\varepsilon_i+\varepsilon_j)\right)+\frac{\varphi^{-1/2}}{n+1}\sum_{i=1}^n\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\text{sgn}(\varepsilon_i)/\rho(U_i)\\ &=\sum_{i=1}^n\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\xi_i/\rho^2(U_i)+o_p(h^{-1/2}),\end{aligned}$$

where the last equality holds because of the facts that

$$\begin{aligned}&\sum\sum_{j\neq i}\frac{\varphi^{-1/2}}{n+1}\text{sgn}(\varepsilon_i+\varepsilon_j)=\varphi^{-1/2}(G(\varepsilon_i-0.5))+o_p(n^{-1/2}),\\ &E\left(\frac{\varphi^{-1/2}}{n+1}\sum_{i=1}^n\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\text{sgn}(\varepsilon_i)/\rho(U_i)\right)=O\left(\frac{1}{nh}\right)=o(h^{-1/2}),\\ &\text{var}\left(\frac{\varphi^{-1/2}}{n+1}\sum_{i=1}^n\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\text{sgn}(\varepsilon_i)/\rho(U_i)\right)=O\left(\frac{1}{n}+\frac{1}{n^2h}\right)=o(h^{-1}).\end{aligned}$$

Similarly, we can prove that

$$
\begin{aligned}
&\frac{\varphi^{-1/2}}{n+1}\sum\sum_{i\le j}\left(\boldsymbol{R}_n(U_i)^T\boldsymbol{W}_i/\rho(U_i)+\boldsymbol{R}_n(U_j)^T\boldsymbol{W}_j/\rho(U_j))\right)\mathrm{sgn}(\varepsilon_i+\varepsilon_j)\\
=&\sum_{i=1}^n\boldsymbol{R}_n(U_i)^T\boldsymbol{W}_i\xi_i/\rho^2(U_i)+o_p(h^{-1/2}).
\end{aligned}
$$

Thus,

$$
A_h=\sum_{i=1}^n\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\xi_i/\rho^2(U_i)+\sum_{i=1}^n\boldsymbol{R}_n(U_i)^T\boldsymbol{W}_i\xi_i/\rho^2(U_i)+o_p(h^{-1/2}).
$$

Next, we consider B_h which can be written as

$$
\begin{aligned}
B_h=&-\frac{2\varphi^{-1/2}}{n+1}\sum\sum_{i\le j}(\tau^{-1}(\varepsilon_i+\varepsilon_j)+\phi_{ij})I(0<\tau^{-1}(\varepsilon_i+\varepsilon_j)<-\phi_{ij}))\\
&+\frac{2\varphi^{-1/2}}{n+1}\sum\sum_{i\le j}(\tau^{-1}(\varepsilon_i+\varepsilon_j)+\phi_{ij})I(-\phi_{ij}<\tau^{-1}(\varepsilon_i+\varepsilon_j)<0)\\
\doteq&E_h+F_h.
\end{aligned}
$$

On the set $\{\phi_{ij}<0\}$ and conditional on $\{\boldsymbol{X}_i,U_i\}$,

$$
\begin{aligned}
E(E_h)=&-\frac{2\varphi^{-1/2}}{n+1}\sum\sum_{i\le j}\int\int_{-y}^{-\phi_{ij}-y}(\tau^{-1}(x+y)+\phi_{ij})g(x)g(y)dxdy\\
=&\frac{2\varphi^{-1/2}}{n+1}\sum\sum_{i\le j}\int\frac{1}{2}\tau^2\phi_{ij}^2g^2(y)dy+O(n^{-1/2}h^{-3/2})\\
=&\frac{1}{n+1}\sum\sum_{i\le j}\phi_{ij}^2+o(h^{-1/2})
\end{aligned}
$$

and $\mathrm{var}(E_h) = O(n^{-1}h^{-2}) = o(h^{-1})$. Similarly, on the set $\{\phi_{ij} > 0\}$, $E(F_h) = \frac{1}{n+1}\sum\sum_{i\le j}\phi_{ij}^2 + o(h^{-1/2})$ and $\mathrm{var}(E_h) = O(n^{-1}h^{-2}) = o(h^{-1})$. Thus,

$$
\begin{aligned}
B_h =& \frac{1}{n+1}\sum\sum_{i\le j}\Big(\left((\boldsymbol{\alpha}_n(U_i) + \boldsymbol{R}_n(U_i))^T \boldsymbol{W}_i\right)/\rho(U_i) \\
&+ ((\boldsymbol{\alpha}_n(U_j) + \boldsymbol{R}_n(U_j))^T \boldsymbol{W}_j)/\rho(U_j)\Big)^2 + o_p(h^{-1/2}) \\
=& \frac{1}{2}\sum_{i=1}^n (\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i)^2/\rho^2(U_i) + \frac{1}{2}\sum_{i=1}^n (\boldsymbol{R}_n(U_i)^T\boldsymbol{W}_i)^2/\rho^2(U_i) \\
&+ \sum_{i=1}^n \boldsymbol{R}_n(U_i)^T\boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{\alpha}_n(U_i)/\rho^2(U_i) \\
&+ \frac{1}{n+1}\sum\sum_{j\ne i}\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\boldsymbol{\alpha}_n(U_j)^T\boldsymbol{W}_j/\rho(U_i)/\rho(U_j) \\
&+ \frac{1}{n+1}\sum\sum_{j\ne i}\boldsymbol{R}_n(U_i)^T\mathbf{W}_i R_n(U_j)^T\boldsymbol{W}_j/\rho(U_i)/\rho(U_j) \\
&+ \frac{1}{n+1}\sum\sum_{j\ne i}\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\boldsymbol{R}_n(U_j)^T\boldsymbol{W}_j/\rho(U_i)/\rho(U_j) + o_p(h^{-1/2}).
\end{aligned}
$$

After calculating the expectation and variance of the last three sums, we can prove that

$$
\begin{aligned}
&\frac{1}{n+1}\sum\sum_{j\ne i}\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\boldsymbol{\alpha}_n(U_j)^T\boldsymbol{W}_j/\rho(U_i)/\rho(U_j) = O_p(1 + (nh)^{-1/2} + (nh)^{-1}), \\
&\frac{1}{n+1}\sum\sum_{j\ne i}\boldsymbol{R}_n(U_i)^T\mathbf{W}_i R_n(U_j)^T\boldsymbol{W}_j/\rho(U_i)/\rho(U_j) = O_p(h^4 + n^{-1/2}h^{7/2} + n^{-1}h^3), \\
&\frac{1}{n+1}\sum\sum_{j\ne i}\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\boldsymbol{R}_n(U_j)^T\boldsymbol{W}_j/\rho(U_i)/\rho(U_j) = O_p(h^2 + n^{-1/2}h^{3/2} + n^{-1}h),
\end{aligned}
$$

and accordingly,

$$
\begin{aligned}
B_h =& \frac{1}{2}\sum_{i=1}^n (\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i)^2/\rho^2(U_i) + \frac{1}{2}\sum_{i=1}^n (\boldsymbol{R}_n(U_i)^T\boldsymbol{W}_i)^2/\rho^2(U_i) \\
&+ \sum_{i=1}^n \boldsymbol{R}_n(U_i)^T\boldsymbol{W}_i\boldsymbol{W}_i^T\boldsymbol{\alpha}_n(U_i)/\rho^2(U_i) + o_p(h^{-1/2}).
\end{aligned}
$$

This leads to

$$
\begin{aligned}
\lambda_{na} =& \sum_{i=1}^n \boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\xi_i/\rho^2(U_i) - \frac{1}{2}\sum_{i=1}^n (\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i)^2/\rho^2(U_i) \\
&+ R_{n1} - R_{n2} - R_{n3} + o_p(h^{-1/2}).
\end{aligned}
$$

Taking the same procedure as Lemma 7.2 in Fan *et al.* (2001), we can show that

$$\begin{aligned}
R_{n1} &= n^{1/2}h^2 R_{n10} + O(n^{-1/2}h),\\
R_{n2} &= n^{1/2}h^2 R_{n20} + O(n^{-1/2}h),\\
R_{n3} &= nh^4 R_{n30} + O(h^3).
\end{aligned}$$

Also, similar to Lemma 7.4 in Fan *et al.* (2001), it can be verified that

$$\begin{aligned}
&\sum_{i=1}^{n}\boldsymbol{\alpha}_n(U_i)^T\boldsymbol{W}_i\xi_i/\rho^2(U_i) = \frac{1}{h}(p+1)K(0)Ef(U)^{-1}\\
&\quad + \frac{1}{n}\sum_{j\neq i}\sum\rho^{-1}(U_i)\rho^{-1}(U_j)\xi_i\xi_j\boldsymbol{W}_i^T\boldsymbol{\Gamma}(U_j)^{-1}\boldsymbol{W}_jK_n(U_i-U_j) + o_p(h^{-1/2}),\\
&\sum_{i=1}^{n}(\boldsymbol{\alpha}_n(U_i)\boldsymbol{W}_i^T)^2/\rho^2(U_i) = \frac{1}{h}(p+1)Ef(U)^{-1}\int K^2(t)dt\\
&\quad + \frac{2}{nh}\sum_{i<j}\sum\rho^{-1}(U_i)\rho^{-1}(U_j)\xi_i\xi_j\boldsymbol{W}_i^T\boldsymbol{\Gamma}^{-1}(U_i)K*K((U_i-U_j)/h)\boldsymbol{W}_j + o_p(h^{-1/2}).
\end{aligned}$$

Thus, $\lambda_{na} = \mu_{na} - d_{1na} + W(n)h^{-1/2}/2 + o_p(h^{-1/2})$, where $d_{1na} = \tau^{-2}[nh^4R_{30} - n^{1/2}h^2(R_{n10} - R_{n20})] = O_p(nh^4 + n^{1/2}h^2) = o_p(h^{-1/2})$ and

$$\begin{aligned}
W(n) &= \frac{\sqrt{h}}{n}\sum_{j\neq i}\sum\rho^{-1}(U_i)\rho^{-1}(U_j)\xi_i\xi_j[2K_h(U_i-U_j) - K_h*K_h(U_i-U_j)]\boldsymbol{W}_i^T\boldsymbol{\Gamma}(U_j)^{-1}\boldsymbol{W}_j\\
&= \frac{\sqrt{h}}{n}\sum_{j\neq i}\sum\zeta_i\zeta_j[2K_h(U_i-U_j) - K_h*K_h(U_i-U_j)]\boldsymbol{W}_i^T\boldsymbol{\Gamma}(U_j)^{-1}\boldsymbol{W}_j
\end{aligned}$$

where $\zeta_i = \rho^{-1}(U_i)\xi_i = \varphi^{-1/2}\tau(G(\varepsilon_i) - 0.5)$. It remains to show that

$$W(n) \xrightarrow{d} N(0, v)$$

with $v = 2||2K - K*K||_2^2(p+1)Ef(U)^{-1}$. Similar to Fan *et al.* (2001), by applying Theorem 2 in De Jong (1987), we can easily obtain the result.

Under H'_{1a} and by similar arguments as above, it can be checked that

$$\begin{aligned}
\lambda_{na} =&\mu_{na} + d_{2na} - W(n)h^{-1/2}/2\\
&- \sum_{i=1}^{n}h^{-1/2}\boldsymbol{G}^T(U_i)\boldsymbol{W}_i\xi_i/\rho^2(U_i) + o_p(h^{-1/2}),
\end{aligned}$$

Then we can obtain the assertion. □

Appendix C: Proofs of other theorems

Proof of Proposition 1 By Fan and Gijbels (1996), we can easily show that $\widehat{f}(u) = f(u)(1+o_p(1))$. Thus, we only need to show that

$$\widetilde{\rho}^{-1}(u) = \frac{\varphi^{-1/2}}{2n(n-1)t_n f^2(u)} \sum_{i=1}^{n}\sum_{j=1}^{n} I(|\widehat{\varepsilon}_i + \widehat{\varepsilon}_j| \le t_n) K_h(U_i - u) K_h(U_j - u)$$

is a ratio-consistent estimator of $\rho^{-1}(u)$.

$$\begin{aligned}\widetilde{\rho}^{-1}(u) =& \frac{\varphi^{-1/2}}{2n^2 t_n f^2(u)} \sum_{i=1}^{n}\sum_{j=1}^{n} I(|\varrho(U_i)\varepsilon_i + \varrho(U_j)\varepsilon_j| \le t_n) K_h(U_i - u) K_h(U_j - u)\\ &+ \frac{\varphi^{-1/2}}{2n^2 t_n f^2(u)} \sum_{i=1}^{n}\sum_{j=1}^{n} (I(|\widehat{\varepsilon}_i + \widehat{\varepsilon}_j| \le t_n) - I(|\varrho(U_i)\varepsilon_i - \varrho(U_j)\varepsilon_j| \le t_n))\\ &\times K_h(U_i - u) K_h(U_j - u)\\ &\doteq U_{n1} + U_{n2}\end{aligned}$$

Clearly, $U_{n1} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} W_n(i,j)$ is of the form of U-statistic since U_{n1} is symmetric in this arguments. Note that

$$\begin{aligned}E(W_n^2(i,j)) =& \frac{3}{t_n^2 f^4(u)} E(I(|\varrho(U_i)\varepsilon_i + \varrho(U_j)\varepsilon_j| \le t_n) K_h^2(U_i - u) K_h^2(U_j - u))\\ =& \frac{3}{t_n^2 h^4 f^4(u)} E\Bigg(\int \left(G\left(\frac{t_n}{\varrho(U_i)} - \frac{\varrho(U_j)}{\varrho(U_i)}\epsilon \right) - G\left(-\frac{t_n}{\varrho(U_i)} - \frac{\varrho(U_j)}{\varrho(U_i)}\epsilon \right) \right)\\ &\times g(\epsilon) d\epsilon K^2\left(\frac{U_i - u}{h}\right) K^2\left(\frac{U_j - u}{h}\right)\Bigg)\\ =& \frac{\sqrt{3} v_0^2}{t_n \tau \varrho(u) h^2 f^2(u)} (1+o(1)) = O(t_n^{-1} h^{-2}) = o(n)\end{aligned}$$

where the antepenultimate equality is followed by a simple calculation similar to Parzen (1962). Thus, $U_{n1} = E(W_n(i,j)) + o_p(1)$. Similarly,

$$\begin{aligned}E(W_n(i,j)) =& \frac{\varphi^{-1/2}}{2t_n f^2(u)} E(I(|\varrho(U_i)\varepsilon_i + \varrho(U_j)\varepsilon_j| \le t_n) K_h(U_i - u) K_h(U_j - u))\\ =& \frac{\varphi^{-1/2}}{2t_n h^2 f^2(u)} E\Bigg(\int \left(G\left(\frac{t_n}{\varrho(U_i)} - \frac{\varrho(U_j)}{\varrho(U_i)}\epsilon \right) - G\left(-\frac{t_n}{\varrho(U_i)} - \frac{\varrho(U_j)}{\varrho(U_i)}\epsilon \right) \right)\\ &\times g(\epsilon) d\epsilon K\left(\frac{U_i - u}{h}\right) K\left(\frac{U_j - u}{h}\right)\Bigg)\\ =& \frac{1}{\varrho(u)\tau}(1+o(1))\end{aligned}$$

Thus, $U_{n1} = \rho^{-1}(u) + o_p(1)$. Similarly, we can show that $U_{n2} = O(h^2 + (nh)^{-1/2}) = o(1)$ by Lemma 1. Thus, $\widehat{\rho}^{-1}(u)$ is a ratio-consistent estimator of $\rho^{-1}(u)$. □

Proof of Theorem 7 Taking the same procedure as Fan et al. (2001), under H_0, we have

$$\begin{aligned}\omega_{na} =& \sum_{i=1}^n \boldsymbol{\alpha}_n(U_i)^T \boldsymbol{W}_i \varepsilon_i/\sigma^2 - \frac{1}{2}\sum_{i=1}^n (\boldsymbol{\alpha}_n(U_i)^T \boldsymbol{W}_i)^2/\sigma^2 \\ &+ \sum_{i=1}^n \boldsymbol{R}_n(U_i)^T \boldsymbol{W}_i \varepsilon_i/\sigma^2 - R_{n2}/\sigma^2 - R_{n3}/\sigma^2 + o_p(h^{-1/2}).\end{aligned}$$

Also, we can verify that

$$\begin{aligned}\sum_{i=1}^n \boldsymbol{\alpha}_n(U_i)^T \boldsymbol{W}_i \varepsilon_i =& \frac{1}{h}(p+1)K(0)\zeta^2 Ef(U)^{-1} \\ &+ \frac{1}{n}\sum\sum_{j\neq i} \varepsilon_i \xi_j \boldsymbol{W}_i^T \boldsymbol{\Gamma}(U_j)^{-1} \boldsymbol{W}_j K_n(U_i - U_j) + o_p(h^{-1/2}), \\ \sum_{i=1}^n \boldsymbol{R}_n(U_i)^T \boldsymbol{W}_i =& n^{1/2} h^2 R_{\omega n10} + O(n^{-1/2}h).\end{aligned}$$

Thus,

$$\omega_{na} = \mu_{\omega na} + W_\omega(n) h^{-1/2}/2 + o_p(h^{-1/2}),$$

where

$$\begin{aligned}W_\omega(n) =& \frac{2h^{1/2}}{n}\sum\sum_{j\neq i} \sigma^{-2}\varepsilon_i \xi_j \boldsymbol{W}_i^T \boldsymbol{\Gamma}(U_j)^{-1} \boldsymbol{W}_j K_n(U_i - U_j) \\ &- \frac{1}{nh^{1/2}}\sum\sum_{j\neq i} \sigma^{-2}\xi_i \xi_j \boldsymbol{W}_i^T \boldsymbol{\Gamma}^{-1}(U_i) K * K((U_i - U_j)/h) \boldsymbol{W}_j.\end{aligned}$$

Applying the martingale central limit theorem (Hall and Heyde 1980), we can verify that

$$W_\omega(n) \xrightarrow{d} N(0, \varsigma),$$

where

$$\begin{aligned}\varsigma =& 2(p+1)Ef(U)^{-1}\bigg(4\sigma^{-2}\tau^2 \int K^2(t)dt - 4\sigma^{-4}\zeta^2\tau^2 \int K(t)K*K(t)dt \\ &+ \sigma^{-4}\tau^4 \int (K*K)^2(t)dt\bigg).\end{aligned}$$

Under $H_{1a}^{'}$ and by the similar arguments as above, it can be verified that

$$\begin{aligned}\omega_{na} =& \mu_{\omega na} + \sigma^{-2}\tau^2 d_{2na} - W_\omega(n) h^{-1/2}/2 \\ &- \sum_{i=1}^n \sqrt{n}\boldsymbol{G}_n^T(U_i) \boldsymbol{W}_i \varepsilon_i/\sigma^2 + o_p(h^{-1/2}),\end{aligned}$$

from which the assertion follows immediately.

Proof of Theorem 3 Denote

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22}, \end{pmatrix}, \ \mathbf{\Gamma}_{1,2} = \mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{12}\mathbf{\Gamma}_{22}^{-1}\mathbf{\Gamma}_{21},$$

where $\mathbf{\Gamma}_{11}, \mathbf{\Gamma}_{12}, \mathbf{\Gamma}_{21}, \mathbf{\Gamma}_{22}$ are $(p_1+1)\times(p_1+1), (p_1+1)\times p_2, p_2\times(p_1+1), p_2\times p_2$ matrices and $p_2 = p - p_1$. Taking the same procedure as for $\widehat{\boldsymbol{A}}$, we have

$$\begin{aligned}\widehat{\boldsymbol{A}}_2^b(u_0) - \boldsymbol{A}_2(u_0) =& r_n^2\mathbf{\Gamma}_{22}^{-1}(u_0)\sum_{i=1}^n \Big(\xi_i + \boldsymbol{A}_2(U_i)^T\boldsymbol{X}_i^{(2)} - \bar{\eta}_2(u_0, \boldsymbol{X}_i^{(2)}, U_i)\Big) \\ & \times \boldsymbol{X}_i^{(2)}K((U_i - u_0)/h)(1 + o_p(1)),\end{aligned}$$

where $\bar{\eta}_2(u_0, \boldsymbol{X}_i^{(2)}, U_i) = \boldsymbol{A}_2(u_0)^T\boldsymbol{X}_i^{(2)} + \boldsymbol{A}_2'(u_0)^T\boldsymbol{X}_i^{(2)}(U_i - u_0)$. Note that $\lambda_{nb} = \lambda_{na} - \lambda_{nb}'$ where

$$\lambda_{nb}' = \frac{\varphi^{-1/2}\tau^{-1}}{n+1}\sum\sum_{i\le j}\left(|\widehat{\varepsilon}_i^a + \widehat{\varepsilon}_j^a| - |\widehat{\varepsilon}_i^b + \widehat{\varepsilon}_j^b|\right).$$

Similar to the proof of Theorem 6, under H_{0b}, we have

$$\begin{aligned}\lambda_{nb}'\tau^2 =& r_n^2\sum_{i=1}^n\sum_{j=1}^n \xi_i\xi_j{\boldsymbol{X}_i^{(2)}}^T\mathbf{\Gamma}_{22}^{-1}(U_i)\boldsymbol{X}_j^{(2)}K((U_i - U_j)/h) \\ & - \frac{1}{2}r_n^4\sum_{i=1}^n\left(\sum_{j=1}^n \xi_jK((U_i - U_j)/h){\boldsymbol{X}_i^{(2)}}^T\right)\mathbf{\Gamma}_{22}^{-1}(U_i)\boldsymbol{X}_i^{(2)}{\boldsymbol{X}_i^{(2)}}^T\mathbf{\Gamma}_{22}^{-1}(U_i) \\ & \times\left(\sum_{j=1}^n \xi_jK((U_i - U_j)/h)\boldsymbol{X}_i^{(2)}\right) + o_p(h^{-1/2}),\end{aligned}$$

Consequently,

$$\begin{aligned}& -\lambda_{nb}\tau^2 \\ & = -r_n^2\sum_{i=1}^n\sum_{j=1}^n \xi_i\xi_j\left(\boldsymbol{W}_j^{(1)} - \mathbf{\Gamma}_{12}(U_i)\mathbf{\Gamma}_{22}^{-1}(U_i)\boldsymbol{X}_j^{(2)}\right)^T\mathbf{\Gamma}_{1,2}^{-1}(U_i) \\ & \quad \times\left(\boldsymbol{W}_i^{(1)} - \mathbf{\Gamma}_{12}(U_i)\mathbf{\Gamma}_{22}^{-1}(U_i)\boldsymbol{X}_j^{(2)}\right)K((U_i - U_j)/h) \\ & \quad + \frac{r_n^4}{2}\sum_{i=1}^n\sum_{j=1}^n \xi_i\xi_j\sum_{k=1}^n(\boldsymbol{W}_i^{(1)} - \mathbf{\Gamma}_{12}(U_k)\mathbf{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_i^{(2)})^T\mathbf{\Gamma}_{1,2}^{-1}(U_k) \\ & \quad \times(\boldsymbol{W}_k^{(1)} - \mathbf{\Gamma}_{12}(U_k)\mathbf{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_k^{(2)})(\boldsymbol{W}_k^{(1)} - \mathbf{\Gamma}_{12}(U_k)\mathbf{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_k^{(2)})^T \\ & \quad \times\mathbf{\Gamma}_{1,2}^{-1}(U_k)(\boldsymbol{W}_j^{(1)} - \mathbf{\Gamma}_{12}(U_k)\mathbf{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_j^{(2)}) \\ & \quad + R_{n4} + R_{n5} + o_p(h^{-1/2}),\end{aligned}$$

where $\boldsymbol{W}_i^{(1)} = (1, \boldsymbol{X}_i^{(1)^T})^T$ and

$$\begin{aligned}
R_{n4} =& \frac{r_n^4}{2} \sum_{i=1}^n \sum_{j=1}^n \xi_i \xi_j \sum_{k=1}^n (\boldsymbol{W}_i^{(1)} - \boldsymbol{\Gamma}_{12}(U_k)\boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_i^{(2)})^T \boldsymbol{\Gamma}_{1,2}^{-1}(U_k) \\
& \times (\boldsymbol{W}_k^{(1)} - \boldsymbol{\Gamma}_{12}(U_k)\boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_k^{(2)})\boldsymbol{X}_k^{(2)^T} \boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_j^{(2)} \\
& \times K((U_i - U_k)/h)K((U_j - U_k)/h), \\
R_{n5} =& \frac{r_n^4}{2} \sum_{i=1}^n \sum_{j=1}^n \xi_i \xi_j \sum_{k=1}^n (\boldsymbol{W}_j^{(1)} - \boldsymbol{\Gamma}_{12}(U_k)\boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_j^{(2)})^T \boldsymbol{\Gamma}_{1,2}^{-1}(U_k) \\
& \times (\boldsymbol{W}_k^{(1)} - \boldsymbol{\Gamma}_{12}(U_k)\boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_k^{(2)})\boldsymbol{X}_k^{(2)^T} \boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_j^{(2)} \\
& \times K((U_i - U_k)/h)K((U_j - U_k)/h).
\end{aligned}$$

After some tedious calculation, as $nh^{3/2} \to \infty$, $E(R_{n4}^2) = O(n^{-2}h^{-4}) = o(h^{-1})$ and thus $R_{n4} = o_p(h^{-1/2})$. Similarly, we can show $R_{n5} = o_p(h^{-1/2})$. As a consequence,

$$\begin{aligned}
& -\lambda_{nb}\tau^2 \\
& = -r_n^2 \sum_{i=1}^n \sum_{j=1}^n \xi_i \xi_j (\boldsymbol{W}_j^{(1)} - \boldsymbol{\Gamma}_{12}(U_i)\boldsymbol{\Gamma}_{22}^{-1}(U_i)\boldsymbol{X}_j^{(2)})^T \boldsymbol{\Gamma}_{1,2}^{-1}(U_i) \\
& \quad \times (\boldsymbol{W}_i^{(1)} - \boldsymbol{\Gamma}_{12}(U_i)\boldsymbol{\Gamma}_{22}^{-1}(U_i)\boldsymbol{X}_j^{(2)})K((U_i - U_j)/h) \\
& \quad + \frac{r_n^4}{2} \sum_{i=1}^n \sum_{j=1}^n \xi_i \xi_j \sum_{k=1}^n (\boldsymbol{W}_i^{(1)} - \boldsymbol{\Gamma}_{12}(U_k)\boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_i^{(2)})^T \boldsymbol{\Gamma}_{1,2}^{-1}(U_k) \\
& \quad \times (\boldsymbol{W}_k^{(1)} - \boldsymbol{\Gamma}_{12}(U_k)\boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_k^{(2)})(\boldsymbol{W}_k^{(1)} - \boldsymbol{\Gamma}_{12}(U_k)\boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_k^{(2)})^T \\
& \quad \times \boldsymbol{\Gamma}_{1,2}^{-1}(U_k)(\boldsymbol{W}_j^{(1)} - \boldsymbol{\Gamma}_{12}(U_k)\boldsymbol{\Gamma}_{22}^{-1}(U_k)\boldsymbol{X}_j^{(2)}) + o_p(h^{-1/2})
\end{aligned}$$

The remaining proof follows the same lines as those in the proof of Theorem 6. □

Proof of Theorem 4 Let $\eta_i = \delta^{-1}(G(2c - \varepsilon_i) - 1/2)$ and $\boldsymbol{\Gamma}_{2,1} = \boldsymbol{\Gamma}_{22} - \boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{12}$. Analogously to the arguments for $\widehat{\boldsymbol{A}}$, we get

$$(\widehat{\alpha}^c(u_0), \widehat{\boldsymbol{A}}_1^c(u_0)^T)^T - (\alpha(u_0), \boldsymbol{A}_1(u_0)^T)^T = (c, \boldsymbol{0}_{1\times p_1})^T + (\widetilde{\alpha}_n(u_0) + \widetilde{R}_n(u_0))(1 + o_p(1)),$$

where

$$\begin{aligned}
\widetilde{\alpha}_n(u_0) &= r_n^2 \begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{\Gamma}_{11} \end{pmatrix}^{-1} \sum_{i=1}^n \eta_i (1, \boldsymbol{X}_i^{(1)^T})^T K_h((U_i - u_0)/h), \\
\widetilde{R}_n(u_0) &= r_n^2 \begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{\Gamma}_{11} \end{pmatrix}^{-1} \sum_{i=1}^n \Big(\alpha(U_i) - \boldsymbol{\beta}(u_0)^T \boldsymbol{V}_i(u_0) + \boldsymbol{A}_1(U_i)^T \boldsymbol{X}_i^{(1)} \\
& \qquad - \boldsymbol{A}_1(u_0)^T \boldsymbol{X}_i^{(1)} - \boldsymbol{A}_1''(u_0)^T \boldsymbol{X}_i^{(1)} \Big)(1, \boldsymbol{X}_i^{(1)^T})^T K((U_i - u_0)/h).
\end{aligned}$$

Define $\bar{\phi}_{ij} = \alpha(U_i) - \widehat{\alpha}^c(U_i) + \boldsymbol{A}(U_i)^T\boldsymbol{X}_i - \widehat{\boldsymbol{A}}^c(U_i)^T\boldsymbol{X}_i + \alpha(U_j) - \widehat{\alpha}^c(U_j) + \boldsymbol{A}(U_j)^T\boldsymbol{X}_j - \widehat{\boldsymbol{A}}^c(U_j)^T\boldsymbol{X}_j$ and $\boldsymbol{W}_i^{(1)} = (1, \boldsymbol{X}_i^{(1)^T})^T$. Thus,

$$\begin{aligned}
\lambda_{n2u} =& \frac{\psi^{-1}\delta}{n+1}\sum\sum_{i\le j}(|\widehat{\varepsilon}_i^c + \widehat{\varepsilon}_j^c| - |\widehat{\varepsilon}_i + \widehat{\varepsilon}_j|)\\
=& \frac{\psi^{-1}\delta}{n+1}\sum\sum_{i\le j}(|\varepsilon_i + \varepsilon_j - 2c + \bar{\phi}_{ij}| - |\varepsilon_i + \varepsilon_j - 2c + \phi_{ij}|)\\
=& \frac{\psi^{-1}\delta}{n+1}\sum\sum_{i\le j}(|\varepsilon_i + \varepsilon_j - 2c + \bar{\phi}_{ij}| - |\varepsilon_i + \varepsilon_j - 2c|)\\
&- \frac{\psi^{-1}\delta}{n+1}\sum\sum_{i\le j}(|\varepsilon_i + \varepsilon_j - 2c + \phi_{ij}| - |\varepsilon_i + \varepsilon_j - 2c|)\\
\doteq& G_h - H_h.
\end{aligned}$$

Taking the same procedure as in the proof of Theorem 6, we can show that

$$\begin{aligned}
G_h =& \psi^{-1}\delta^2\sum_{i=1}^n(\widetilde{\boldsymbol{\alpha}}_n(U_i) + \widetilde{\boldsymbol{R}}_n(U_i))^T\boldsymbol{W}_i^{(1)}\eta_i + ((\widetilde{\boldsymbol{\alpha}}_n(U_i) + \widetilde{\boldsymbol{R}}_n(U_i))^T\boldsymbol{W}_i^{(1)})^2 + o_p(h^{-1/2}),\\
H_h =& \psi^{-1}\delta^2\sum_{i=1}^n(\boldsymbol{\alpha}_n(U_i) + \boldsymbol{R}_n(U_i))^T\boldsymbol{W}_i\eta_i + ((\boldsymbol{\alpha}_n(U_i) + \boldsymbol{R}_n(U_i))^T\boldsymbol{W}_i)^2 + o_p(h^{-1/2}).
\end{aligned}$$

Finally, similar to the proof of Theorem 3, we can obtain the result. □

Proof of Theorem 5 Denote $\xi_i^* = \sum\sum_{j\neq i}\frac{\varphi^{-1/2}\tau}{2(n+1)}\mathrm{sgn}(\varepsilon_i^* + \varepsilon_j^*)$. We will show that $E^*(\xi_i^*) = o_p(1)$ and $E^*(\xi_i^{*2}) = \tau^2(1 + o_p(1))$ where E^* denotes the conditional expectation given $\{\boldsymbol{X}_i, U_i, Y_i\}_{i=1}^n$.

$$\begin{aligned}
E^*(\xi_i^*) =& E^*\left(\sum\sum_{j\neq i}\frac{\varphi^{-1/2}\tau}{2(n+1)}\mathrm{sgn}(\varepsilon_i^* + \varepsilon_j^*)\right) = \frac{\varphi^{-1/2}n\tau}{2(n+1)}E^*(\mathrm{sgn}(\varepsilon_i^* + \varepsilon_j^*))\\
=& \frac{\varphi^{-1/2}\tau}{2n(n+1)}\sum_{i=1}^n\sum_{j=1}^n\mathrm{sgn}(\widehat{\varepsilon}_i + \widehat{\varepsilon}_j) = \frac{\varphi^{-1/2}\tau}{2n(n+1)}\sum_{1\le i,j\le n}\mathrm{sgn}(\varepsilon_i + \varepsilon_j + \phi_{ij})\\
=& \frac{\varphi^{-1/2}\tau}{2n(n+1)}\sum_{i=1}^n\sum_{j=1}^n[\mathrm{sgn}(\varepsilon_i + \varepsilon_j + \phi_{ij}) - \mathrm{sgn}(\varepsilon_i + \varepsilon_j)]\\
&+ \frac{\varphi^{-1/2}\tau}{2n(n+1)}\sum_{i=1}^n\sum_{j=1}^n\mathrm{sgn}(\varepsilon_i + \varepsilon_j)\\
\doteq& I_{1h} + I_{2h},
\end{aligned}$$

where ϕ_{ij} is defined in the proof of Theorem 6. Taking the similar procedure as for dealing with B_h, we obtain that $E(I_{1h}^2) = O((nh)^{-1})$ and $E(I_{2h}^2) = O(n^{-1})$. Thus,

$E^*(\xi_i^*) = o_p((nh)^{-1/2} + n^{-1/2}) = o_p(1)$. Next, we consider the second moment.

$$
\begin{aligned}
E^*(\xi_i^{*2}) &= E^* \left(\sum\sum_{j\neq i} \frac{\varphi^{-1/2}\tau}{2(n+1)} \mathrm{sgn}(\varepsilon_i^* + \varepsilon_j^*) \right)^2 \\
=& \frac{\varphi^{-1}\tau^2}{4n(n+1)^2} \sum_{i=1}^n \sum_{j=1}^n (\mathrm{sgn}(\varepsilon_i + \varepsilon_j + \phi_{ij}) \mathrm{sgn}(\varepsilon_i + \varepsilon_l + \phi_{il})) \\
=& \frac{\varphi^{-1}\tau^2}{4n(n+1)^2} \sum_{i=1}^n \sum_{j=1}^n (\mathrm{sgn}(\varepsilon_i + \varepsilon_j) \mathrm{sgn}(\varepsilon_i + \varepsilon_l)) \\
&+ \frac{\varphi^{-1}\tau^2}{4n(n+1)^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n ([\mathrm{sgn}(\varepsilon_i + \varepsilon_j + \phi_{ij}) - \mathrm{sgn}(\varepsilon_i + \varepsilon_j)] \mathrm{sgn}(\varepsilon_i + \varepsilon_l + \phi_{il})) \\
&+ \frac{\varphi^{-1}\tau^2}{4n(n+1)^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n (\mathrm{sgn}(\varepsilon_i + \varepsilon_j) [\mathrm{sgn}(\varepsilon_i + \varepsilon_l + \phi_{il}) - \mathrm{sgn}(\varepsilon_i + \varepsilon_l)]) \\
\doteq& J_{1h} + J_{2h} + J_{3h}.
\end{aligned}
$$

Using the similar arguments as for B_h, it can be verified that $E(J_{1h}) = \tau^2$, $\mathrm{var}(J_{1h}) = O(n^{-1})$, $E(J_{2h}^2) = O((nh)^{-1})$ and $E(J_{3h}^2) = O((nh)^{-1})$. Hence, $E^*(\xi_i^{*2}) = \tau^2 + O_p((nh)^{-1/2} + n^{-1/2})$. With these results, Theorem 5 can be established by mimicking the proof of Theorem 6. □

Proof of Corollaries 1-2 The proof of these two results are similar and thus we only elaborate on the first one. We decompose this problem as the following two simple null hypothesis

$$
H_{0g1} : (\alpha_0, \alpha_1) = (\beta_0 + c, \beta_1) \quad \text{versus} \quad H_{1g1} : m(x) = \alpha_0 + \alpha_1 x \tag{A.2}
$$

and

$$
H_{0g2} : (\alpha_0, \alpha_1) = (\beta_0 + c, \beta_1) \quad \text{versus} \quad H_{1g2} : m(x) \neq \alpha_0 + \alpha_1 x, \tag{A.3}
$$

where β_0, β_1 are the true value of parameters. The WGLR test statistics for the hypotheses (A.2) and (A.3) are denoted as λ_{ng2} and λ_{ng2}, respectively. It can be easily seen that $\lambda_{ng} = \lambda_{ng2} - \lambda_{ng1}$. According to Theorem 3, we have $\sigma_{ng}^{-1}(\lambda_{ng2} - \mu_{ng}) \xrightarrow{d} N(0,1)$. Furthermore, by Theorem 3.6.1 in Hettmansperger and McKean (2010), we have $\lambda_{ng1} = O_p(1) = o_p(h^{-1/2})$, from which the result immediately follows. □

References

De Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory Related Field* **75**, 261–277.

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, New York: Academic Press.

Kou, H. L., Sievers, G. L. and McKean, J. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scandinavian Journal of Statistics* **14**, 131–141.

Appendix D: Simulation results in heteroscedasticity cases

In this subsection, we conduct a simulation study in the heteroscedasticity case. All the settings are the same as the above subsection except that the variance function is

$$\varrho^2(u) = \frac{e^u}{\int_0^1 e^t dt}.$$

Similar to Koul *et al.*(1987), we choose the bandwidth t_n in (2.14) as $t_n = h\gamma_\alpha$ where γ_α is the α-th quantile of the empirical distribution function of $\{|\widehat{\varepsilon}_i - \widehat{\varepsilon}_j|\}_{1\leq i<j\leq n}$. Here we choose $\alpha = 0.8$.

Tables A.2 and A.3 report the simulated level of our test and power comparison with other tests, respectively. The simulated results are similar to the homoscedasticity case. We can control the empirical sizes in most cases. Under the normal cases, WGLR performs a litter worse than GLR, Zheng and LOSS tests. However, under the non-normal cases, WGLR is significantly powerful than the other tests. And WGLR still performs better than WQ in the model (I) and (VI). Thus, our WGLR procedure is also robust in the heteroscedasticity case.

Table A.2: Simulated level (%) of test on testing (3.1) with heteroscedastic error

	$n = 25$			$n = 50$			$n = 100$		
h	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
0.12	7.5	7.0	7.1	5.4	4.5	5.8	5.6	5.9	4.8
0.15	7.2	6.7	7.3	5.7	4.9	4.8	5.3	5.2	4.9
0.18	6.3	5.1	5.3	4.3	4.0	4.7	4.1	5.4	5.3
0.21	4.5	5.0	4.2	5.5	4.4	3.8	4.2	4.1	5.7

Table A.3: Empirical power (%) of tests on testing (3.1) with heteroscedastic error.

		$n=25$					$n=50$				
	Models	WGLR	GLR	WQ	Zheng	LOSS	WGLR	GLR	WQ	Zheng	LOSS
$N(0,1)$	(I)	13	13	4.8	25	23	37	27	4.0	43	39
	(II)	17	22	13	33	37	45	46	19	66	61
	(III)	11	24	20	17	14	40	51	47	41	26
	(IV)	16	32	48	46	21	50	65	82	77	49
	(V)	22	30	18	40	41	58	62	37	74	70
	(VI)	28	28	5.8	45	42	62	57	09	76	71
$t(3)$	(I)	40	16	5.5	37	27	75	39	4.6	50	56
	(II)	49	33	19	50	46	84	64	39	68	75
	(III)	43	32	35	23	20	87	68	83	41	47
	(IV)	55	43	71	57	35	90	77	97	79	68
	(V)	61	42	35	51	56	92	77	70	77	80
	(VI)	65	41	09	62	53	96	72	18	79	81
$T(0.05,10)$	(I)	66	35	5.6	57	44	96	54	5.5	65	56
	(II)	78	50	44	72	60	99	70	71	77	73
	(III)	69	48	72	38	31	99	69	99	67	53
	(IV)	81	57	96	71	52	99	74	100	87	65
	(V)	86	60	69	69	66	100	74	95	80	79
	(VI)	89	58	19	74	65	100	76	32	83	80

Appendix E: Some additional simulation results in Section 3

Table A.4: Simulated level (%) of test on testing linearity

	$n=60$				$n=100$			
h	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
0.06	7.2	6.8	7.0	7.3	6.2	6.4	6.5	6.6
0.09	6.4	6.0	6.3	6.6	6.1	5.8	6.2	5.6
0.12	5.3	5.2	4.7	4.6	5.8	5.5	4.4	4.9
0.15	5.0	5.4	4.5	5.3	5.1	4.7	5.3	5.0

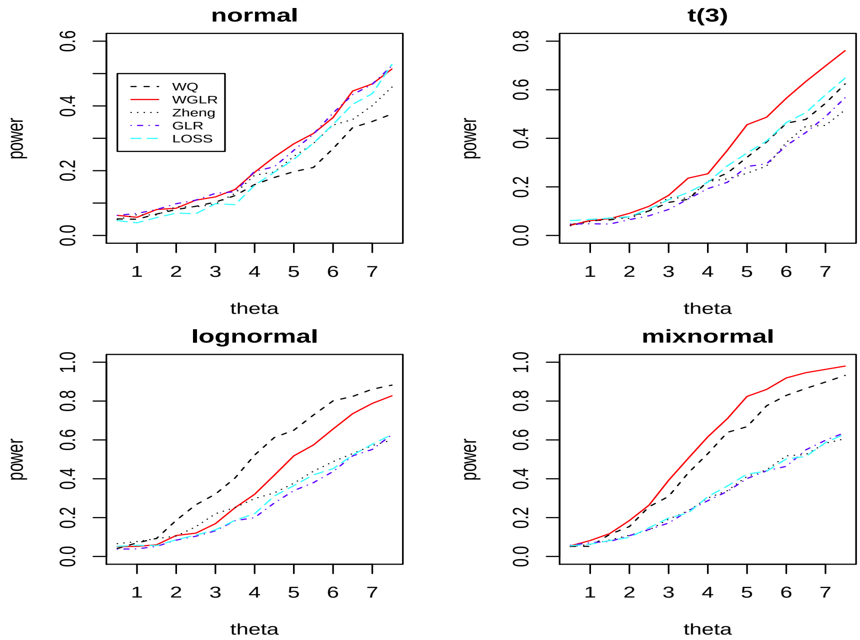

Figure A.1: Simulated power curves of square alternative on testing linearity.

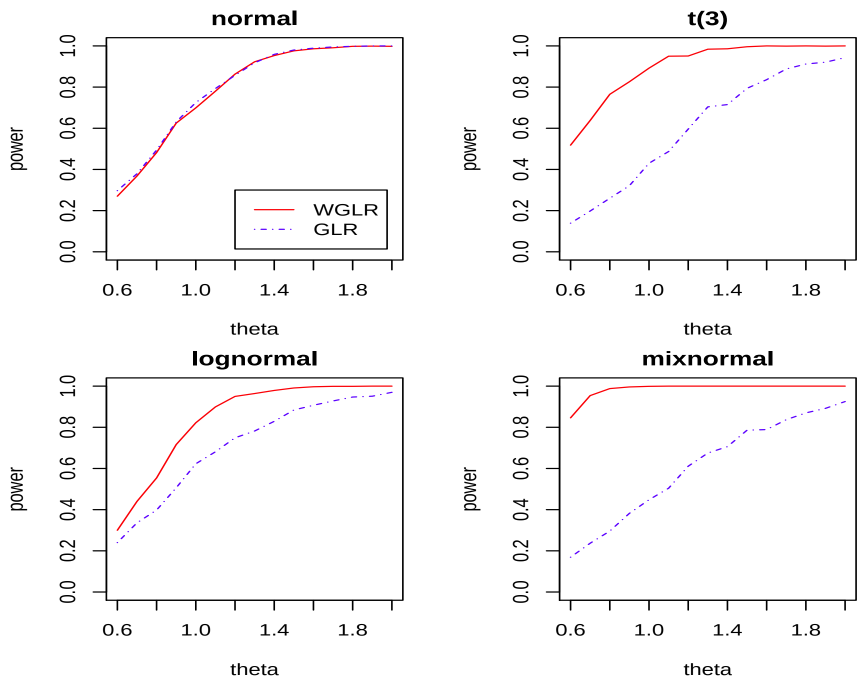

Figure A.2: Simulated power curves on testing homogeneity with $n = 100, h = 0.09$.

Table A.5: Simulated level (%) of test on testing homogeneity

	$n = 60$				$n = 100$			
h	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
0.06	6.9	7.3	7.2	7.3	6.2	6.3	6.0	6.9
0.09	5.9	6.2	5.5	6.8	6.3	5.7	6.2	5.9
0.12	5.3	5.8	5.7	4.4	5.8	5.8	4.7	5.7
0.15	4.4	5.7	4.9	5.9	5.0	4.2	5.6	5.2

Appendix F: Some additional figures in Section 4

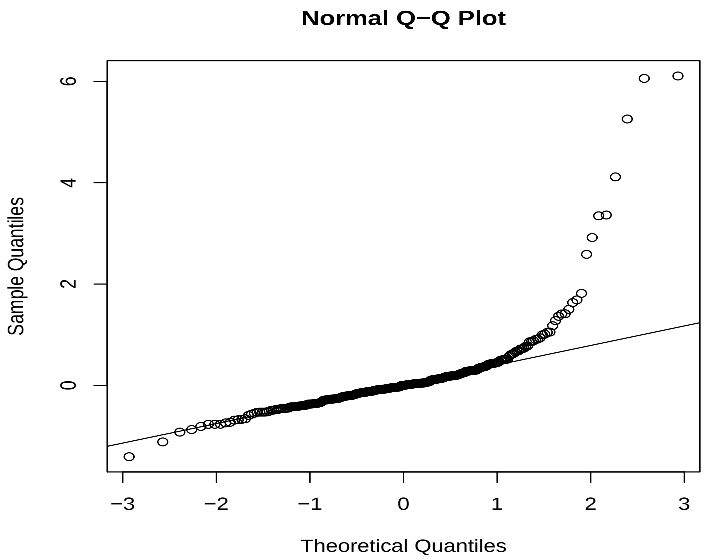

Figure A.3: The normal QQ-plot for the residuals

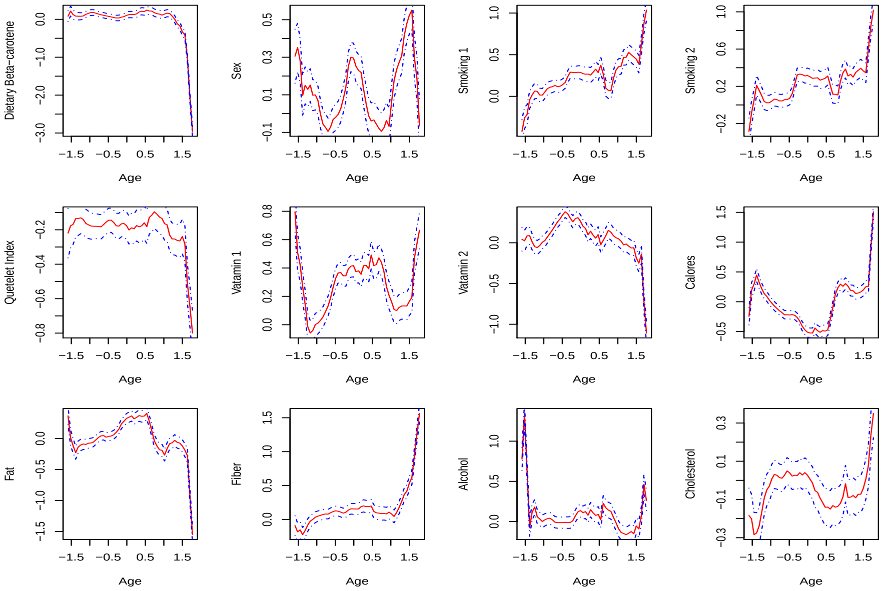

Figure A.4: Fitted coefficient functions and corresponding pointwise 95% confidence interval.