Submission

Journal de la Société Française de Statistique

Bickel-Rosenblatt type goodness-of-fit tests in linear errors in variables model

Titre: Alternative title in french

Hira L. Koul[1] and Weixing Song[2]

Abstract: This paper discusses a class of Bickel-Rosenblatt type goodness-of-fit tests for fitting a parametric family to the regression error density function in linear errors-in-variables models. These tests are based on a class of L_2 distances between a kernel density estimator of the residual and an estimator of its expectation under null hypothesis. The paper investigates asymptotic normality of the null distribution of the proposed test statistics. Asymptotic power of these tests under certain fixed and local alternatives is also considered, and an optimal test within the class is identified. A parametric bootstrap algorithm is proposed to implement the proposed test procedure when the sample size is small or moderate. A finite sample simulation study shows very desirable finite sample behavior of the proposed inference procedures.

Résumé : Abstract in the alternative language.

Keywords: L_2 distance, optimal power, bootstrap approximation
AMS 2000 subject classifications: Primary 62G08, secondary 62G10

1. Introduction

Statistical inference in classical regression models often assumes both response variable and possibly multidimensional predictors are fully observable. But, as is evidenced in the monographs of Fuller (1987) [5] and Carroll, Rupert and Stefanski (1995) [2], in numerous studies of practical importance predictors are often unobservable. Instead, one observes some surrogates for predictors. These models are often called errors-in-variables models or measurement errors models.

Extensive research has been devoted to the estimation of the underlying parameters, both Euclidean and infinite dimensional, in these models. Recent years have seen an increasing research activity in the study of lack-of-fit testing of a parametric regression model in the presence of measurement errors in the predictors. Relatively, little published literature exists for checking the appropriateness of the distributional assumptions on regression errors and/or measurement errors in error-prone predictors. Focus of this paper is to make an attempt at partly filling this void.

More precisely, consider the linear errors-in-variables regression model

$$Y = \alpha + \beta' X + \varepsilon, \qquad Z = X + u, \tag{1.1}$$

[1] Michigan State University.
E-mail: `koul@stt.msu.edu`
[2] Kansas State University.
E-mail: `weixing@ksu.edu`

where Y is the response variable and X is a d-dimensional vector of unobserved predictors. The variables X, u and ε are assumed to be mutually independent. For model identifiability, as is typically the case in these models, we assume density of the measurement error vector u is known. The problem of interest is to develop some goodness-of-fit tests for checking the appropriateness of a specified family of densities of the regression error ε.

Accordingly, let q be a known positive integer, Θ be a compact subset of $\mathbb{R}^q$, $\mathscr{F} := \{f(x,\theta); x \in \mathbb{R}, \theta \in \Theta\}$ be a parametric family of densities with mean 0 on $\mathbb{R}$ and let f denote density of ε. Consider the problem of testing the hypothesis

$$\begin{aligned} &H_0 : f(x) = f(x,\theta), \qquad \text{for some } \theta \in \Theta \text{ and all } x \in \mathbb{R}^d \text{ vs.} \\ &H_1 : H_0 \text{ is not true.} \end{aligned}$$

Goodness-of-fit testing has long been an important research area in statistics. In the case of completely observable data, beginning with Pearson in 1900, the most commonly used goodness-of-fit test statistic is a chi-square statistic. Pearson χ^2 test was originally designed for fitting a finitely supported discrete distribution, but after discretization, this and other related tests can also be used for checking continuous distribution. However, it is well known that the power of these procedures is generally low, see, e.g., D'Agostino and Stephens (1986) [4]. Other well known goodness-of-fit tests are based on certain distances between empirical distribution function and the parametric family of distributions being fitted. Kolmogorov-Smirnov and Cramér-von Mises tests are examples of this methodology. Asymptotic null distributions of these statistics in the case of fitting a parametric family of distributions is often unknown.

In the one sample set up, Bickel and Rosenblatt (1973) [1] proposed to use L_2 distance between a kernel density estimator and its null expected value for fitting a given density. Unlike the tests based on residual empirical processes, in the case of fitting an error density up to an unknown location parameter, asymptotic null distribution of an analog of this statistic based on residuals is the same as if the location parameter were known. In other words, not knowing the nuisance location parameter has no effect on asymptotic level of the test based on the analog of this statistic. Lee and Na (2002) [10], Bachmann and Dette (2005) [3], and Koul and Mimoto (2010) [7] observed that this fact continues to hold for the analog of this statistic when fitting an error density based on residuals in autoregressive and generalized autoregressive conditionally heteroscedastic time series models. In all of these works data is completely observable. To the best of our knowledge to date, this methodology has not been develop for testing of H_0 in the model (1.1).

In this paper we shall construct a class of Bickel-Rosenblatt type (BR-type) goodness-of-fit tests of H_0 in errors-in-variables regression model (1.1). The paper is organized as follows. The class of test statistics and the needed regularity assumptions are described in section 2. Asymptotic normality under H_0 of the test statistics is stated in section 3. Results about consistency of the proposed tests against a fixed alternative and their asymptotic powers against a class of nonparametric local alternatives are stated in section 4 where we also discuss the choice of an optimal test within the class considered that maximizes this asymptotic power. Section 5 reports findings of some simulation studies and a bootstrap approximation to the asymptotic null distribution of the proposed BR-type tests. The proofs of the results stated in sections 3 and 4 appear in the last section.

2. Test Statistics and Assumptions

In this section we shall describe the proposed test statistics and needed assumptions for their asymptotic normality. Let α_0, β_0 be the true values of the regression coefficient in (1.1) and θ_0 be the true value of θ under H_0. Plug in $X = Z - u$ in there to obtain

$$Y = \alpha_0 + \beta_0' Z + \xi, \qquad \xi = \varepsilon - \beta_0' u.$$

With g denoting the density of u, assumed to be known, the density of ξ is

$$h(v) := \int f(v + \beta_0' u) g(u) du.$$

Under H_0, the density of ξ is $h(v; \beta_0, \theta_0)$, where

$$h(v; \beta, \theta) = \int f(v + \beta' u, \theta) g(u) du, \quad v \in \mathbb{R}, \beta \in \mathbb{R}^d, \theta \in \Theta.$$

By the independence of ε and u, characteristic function of $\xi = \varepsilon + \beta_0' u$ is the product of the characteristic functions of ε and $-\beta_0' u$. Because the characteristic function of u, hence $-\beta_0' u$, is known, this implies that the characteristic functions of ξ and ε can be uniquely determined from each other. Therefore, there is a one-to-one map between the densities of ε and ξ. Consequently, testing for H_0 is equivalent to testing for the hypothesis

$$\mathscr{H}_0 : h(v) = h(v; \beta_0, \theta_0), \quad \text{for some } \theta_0 \in \Theta, \text{ and for all } v \in \mathbb{R}.$$

The given data consists of n i.i.d. observations $(Z_i, Y_i), 1 \le i \le n$, from the model (1.1). Let $\hat{\alpha}_n$, $\hat{\beta}_n$ be any $\sqrt{n}$-consistent estimators of α_0, β_0, respectively. Let $\xi_i := Y_i - \alpha_0 - \beta_0' Z_i$, $\hat{\xi}_i = Y_i - \hat{\alpha}_n - \hat{\beta}_n' Z_i$, K be a density kernel function, b denote the bandwidth, $K_b(\cdot) := b^{-1} K(\cdot / b)$, and let

$$\begin{aligned}
h_n(v; \alpha, \beta) &:= \frac{1}{n} \sum_{i=1}^n K_b\big(v - Y_i + \alpha + \beta' Z_i\big), \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}^d, \\
h_n(v) &:= h_n(v; \alpha_0, \beta_0) = \frac{1}{n} \sum_{i=1}^n K_b\big(v - \xi_i\big), \\
\hat{h}_n(v) &:= h_n(v; \hat{\alpha}_n, \hat{\beta}_n) = \frac{1}{n} \sum_{i=1}^n K_b\big(v - \hat{\xi}_i\big), \quad v \in \mathbb{R}.
\end{aligned}$$

For known α_0, β_0, $h_n(v)$ is an estimator of $h(v)$. But since they are rarely known, a genuine estimator of $h(v)$ is provided by the kernel density estimator $\hat{h}_n(v)$.

Next, define

$$h_b(v; \beta, \theta) := \int K_b(v - u) h(u; \beta, \theta) du, \quad \beta \in \mathbb{R}^d, \theta \in \mathbb{R}^q.$$

With E_0 denote the expectation under H_0,

$$E_0(h_n(v)) = h_b(v; \beta_0, \theta_0), \quad v \in \mathbb{R}.$$

Let W be a nondecreasing real valued function inducing a σ-finite measure on $\mathbb{R}$ and set

$$T_n(\alpha,\beta,\theta) \;:=\; \int \big[h_n(v;\alpha,\beta)-E_0 h_n(v;\alpha,\beta)\big]^2 dW(v),\;\; \beta\in\mathbb{R}^d,\, \theta\in\mathbb{R}^q.$$

Clearly,

$$\begin{aligned} T_n(\alpha_0,\beta_0,\theta_0) &= \int [h_n(v;\alpha_0,\beta_0)-E_0(h_n(v))]^2 dW(v) \\ &= \int [h_n(v;\alpha_0,\beta_0)-h_b(v;\beta_0,\theta_0)]^2 dW(v). \end{aligned}$$

Let $\hat\theta_n$ be any $\sqrt{n}$-consistent estimator of θ_0 under the null hypothesis. The proposed class of BR-type goodness-of-fit test statistics, one for each W, for testing for H_0 is

$$T_n(\hat\alpha_n,\hat\beta_n,\hat\theta_n)=\int \big[\hat h_n(v)-h_b(v;\hat\beta_n,\hat\theta_n)\big]^2 dW(v). \tag{2.1}$$

A way to construct an estimator of θ_0 is to use minimum distance (MD) method. For any preliminary estimators of α_0 and β_0, one can estimate θ by

$$\hat\theta_n=\arg\min_{\theta\in\Theta} T_n(\hat\alpha_n,\hat\beta_n,\theta). \tag{2.2}$$

We can show that under some regularity conditions, the MD estimator $\hat\theta_n$ is $\sqrt{n}$-consistent and asymptotically normal. In fact, the proof is similar to the arguments used in Koul and Ni (2004) [8] and Koul and Song (2009) [9]. We do not pursue the proof here.

As for the preliminary estimators of α_0 and β_0, we use the bias-corrected estimators. Let S_{ZZ} and S_{ZY} denote the sample covariance matrices of Z, and of Z and Y, respectively, and let $\Sigma_u := E(uu')$, which is assumed to be known. The respective bias-corrected estimators $\hat\alpha_n=\bar Y-\bar Z'\hat\beta_n$ and $\hat\beta_n=(S_{ZZ}-\Sigma_u)^{-1}S_{ZY}$ of α_0 and β_0 have the following asymptotic expansion. With $a_n:=\hat\alpha_n-\alpha_0$, $d_n:=\hat\beta_n-\beta_0$,

$$\sqrt{n}\begin{pmatrix} a_n\\ d_n\end{pmatrix}=\frac{1}{\sqrt{n}}\sum_{i=1}^n \eta_i+o_p(1),$$

where η_i's are i.i.d. with $E\eta=0$, $\mathrm{Cov}(\eta)=\Sigma_{\alpha\beta}>0$, and $E\|\eta\|^{2+\delta}<\infty$, for some $\delta>0$. Indeed,

$$\begin{aligned} &\sqrt{n}\begin{pmatrix} a_n\\ d_n\end{pmatrix} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} \varepsilon_i-u_i'\beta_0-\mu_X\Sigma_X^{-1}(Z_i-\mu_X)(\varepsilon_i-u_i'\beta_0)-\mu_X'\Sigma_X^{-1}\Sigma_u\beta_0,\\ \Sigma_X^{-1}(Z_i-\mu_X)(\varepsilon_i-u_i'\beta_0)+\Sigma_X^{-1}\Sigma_u\beta_0\end{pmatrix}+o_p(1), \end{aligned} \tag{2.3}$$

where $\mu_X=EX$, $\Sigma_X=\mathrm{Cov}(X)$. Therefore, $\hat\alpha_n,\hat\beta_n$ are $\sqrt{n}$-consistent and asymptotically normal, even if the regression error distribution is misspecified. Certainly there are many other estimators of α_0 and β_0 having similar properties, but for the sake of convenience we shall use the above estimators in this paper.

The following is a list of regularity conditions needed for deriving the following asymptotic results for $T_n(\hat\alpha_n,\hat\beta_n,\hat\theta_n)$. Throughout, for any smooth function $\gamma(v;\beta,\theta)$ of the three variables

v, β, θ, $\dot{\gamma}_x(v; \beta, \theta)$ denotes the vector of first order derivatives of γ with respect to the variable $x = v, \beta$, or θ.

ASSUMPTIONS:

About the kernel function K:

(k1). Density kernel K is four times differentiable, with the ith derivative $K^{(i)}$ bounded for $i = 1, 2, 4$.

(k2). $\int |K^{(1)}(u)| du + \int K^{1+\delta/2}(u) du < \infty$, for some $\delta > 0$.

(k3). $\int u K^{(1)}(u) du \neq 0$, $\int u^i K^{(j)}(u) du = 0$, for all $i = 0, j = 1, 2, 3$; $i = 1, j = 2, 3$.

For example, standard normal density function satisfies all the conditions (k1)-(k3).

For the bandwidth b:

(b). $\quad b \to 0,\, nb^{7/2} \to \infty$.

About the weighting measure W:

(w). The weighting measure W has a compact support $\mathscr{C}$ in $\mathbb{R}$.

For the design variable and measurement error:

(d). $E\|X\|^4 + E\|u\|^4 < \infty$.

For the density function h:

(h1). For all β and θ, $h(x; \beta, \theta)$ is a.s. continuous in $x(W)$.

(h2). For any consistent estimators $\hat{\beta}_n$, $\hat{\theta}_n$ of β_0, θ_0, uniformly in $v \in \mathscr{C}$,

$$\begin{aligned} h_b(v; \hat{\beta}_n, \hat{\theta}_n) - h_b(v; \beta_0, \theta_0) - d_n' \dot{h}_{b\beta}(v; \beta_0, \theta_0) - \Delta_n' \dot{h}_{b\theta}(v; \beta_0, \theta_0) \\ = O_p(\|\hat{\beta}_n - \beta_0\|^2 + \|\hat{\theta}_n - \theta_0\|^2), \end{aligned}$$

and $\dot{h}_{b\beta}(v; \beta_0, \theta_0)$, $\dot{h}_{b\theta}(v; \beta_0, \theta_0)$ are continuous function of v.

(h3). The integrating measure W has a Lebesgue density w such that for some $\delta > 0$,

$$\int \|\dot{h}_\theta(v; \beta_0, \theta_0)\|^{2+\delta} h(v; \beta_0, \theta_0) w^2(v) dv < \infty,$$
$$\int \|\dot{h}_\beta(v; \beta_0, \theta_0)\|^{2+\delta} h(v; \beta_0, \theta_0) w^2(v) dv < \infty.$$

For the sake of brevity, in the sequel, we let

$$h_0(v) := h(v; \beta_0, \theta_0), \quad \dot{h}_\theta(v) := \dot{h}(v; \beta_0, \theta_0),$$
$$a_n := \hat{\alpha}_n - \alpha_0, \quad d_n := \hat{\beta}_n - \beta_0, \quad \Delta_n := \hat{\theta}_n - \theta_0.$$

Moreover, $\psi_{bi}(v) := \psi\big((v - \xi_i)/b\big)/b$, for any function ψ defined on the support of K.

3. Asymptotic null distribution of the test statistics

In this section we shall give asymptotic null distribution of the proposed test statistics. To proceed further we need to define

$$\hat{C}_n = \frac{1}{n^2} \sum_{i=1}^n \int K_b^2(v - \hat{\xi}_i) dW(v), \tag{3.1}$$

$$\hat{\Gamma}_n = 2 \int \hat{h}_n^2(x) w^2(x) dx \int \Big[\int K(v) K(u+v) dv \Big]^2 du.$$

Theorem 3.1. *Under the conditions of (k1)-(k3), (b), (w), (d), (h1)-(h4), and under H_0,*

$$nb^{1/2}\hat{\Gamma}_n^{-1}(T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n)-\hat{C}_n)\to_D N(0,1).$$

Let Φ denote the distribution function of a standardized normal r.v. and for an $0<\alpha<1$, z_α be such that $\Phi(z_\alpha)=1-\alpha$, and let

$$\mathscr{T}_n := nb^{1/2}\hat{\Gamma}_n^{-1}\big(T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n)-\hat{C}_n\big).$$

A consequence of the above theorem is that the test that rejects H_0 whenever $|\mathscr{T}_n|>z_{\alpha/2}$ is of the asymptotic size α. The proof of the above theorem appears in the last section.

As a matter of fact, the tests based on $\mathscr{T}_n$ are nonparametric smoothing tests. For nonparametric smoothing tests, it is well known that the Monte Carlo simulation method and the bootstrap method often provide more accurate approximation to the sampling distribution of the test statistic than the asymptotic normal theory does. Thus, we propose the following parametric bootstrap algorithm:

Step 1: Use the full data set $(Y_i,Z_i), i=1,2,\ldots,n$ to estimate $\hat{\alpha}_n$, $\hat{\beta}_n$ and $\hat{\theta}_n$.

Step 2: Draw independent sample of size B, $\hat{\xi}_j^*$, $j=1,2,\ldots,B$, from the density $h(v;\hat{\theta}_n,\hat{\beta}_n)$, and calculate

$$T_B^* = \int \Big[\frac{1}{B}\sum_{j=1}^{B} K_{b^*}\Big(v-\hat{\xi}_j^*\Big) - h_{b^*}(v;\hat{\beta}_n,\hat{\theta}_n)\Big]^2 dW(v),$$

where b^* satisfies the assumptions $b^*\to 0$, $Bb^{*7/2}\to\infty$ as $B\to\infty$.

Step 3: Repeat Step 2 R times, denote the resulting T_B^*'s as $T_{B,1}^*, T_{B,2}^*,\cdots,T_{B,R}^*$. Sort the absolute values of these R simulated T_B^* in increasing order. For the preassigned significance level α, find the $(1-\alpha/2)100\%$ percentile, denoted as $t_{B,\alpha}^*$.

Step 4: Calculate $T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n)$ using the original full data set. If $|T_n|\ge t_{B,\alpha}^*$, reject the null hypothesis; otherwise, accept the null hypothesis.

Sometimes, in Step 2, it is not easy to draw independent sample from $h(v;\hat{\beta}_n,\hat{\theta}_n)$, but it is easy to draw sample from $f(\varepsilon;\hat{\theta}_n)$, and $g(u)$. Note that h is the density function of $\varepsilon+\beta' u$, so one can draw independent samples ε_j^*, $j=1,2,\ldots,B$ from $f(\varepsilon;\hat{\theta}_n)$, and draw independent samples u_j^*, $j=1,2,\ldots,B$ from $g(u)$, then $\varepsilon_j^*+\hat{\beta}_n' u_j^*, j=1,2,\ldots,B$ can be considered as a sample of size B from $h(v;\hat{\beta}_n,\hat{\theta}_n)$.

4. Consistency and asymptotic power against local alternatives

In this section, we discuss consistency of the $\mathscr{T}_n$-test against a class of fixed alternatives, derive its asymptotic power against sequences of local nonparametric alternatives, and provide an optimal W that maximizes this power among the proposed tests.

4.1. Consistency

We shall show that, under some regularity conditions, the above $\mathscr{T}_n$-test is consistent against certain fixed alternatives. Let $f_a\notin\mathscr{F}$ be a density on $\mathbb{R}$ with mean zero and finite second moment, and consider the alternative hypothesis $H_a : f(v) = f_a(v)$, for all $v \in \mathbb{R}$. Under H_a, density of ξ is $h_a(v;\beta) = \int f_a(v+u'\beta)g(u)du$. We shall assume that $\hat{\theta}_n$ converges to a value θ_a in probability under H_a. In fact, if we define

$$\theta_a = \arg\min_{\theta\in\Theta} \int [h_a(v;\beta_0) - h(v;\theta,\beta_0)]^2 dW(v),$$

one can show that the MD estimator defined in (2.2) converges to θ_a in probability. The proof is omitted for the sake of brevity. Let

$$h_{ab}(v;\beta) = \int K_b(v-u)h_a(u,\beta)du,$$

and $\dot{h}_{ab,\beta}(v;\beta), \dot{h}_{a,\beta}(v;\beta)$ denote the derivatives of h_{ab} and h_a with respect to β, respectively. Assume that

(h1$'$). For all β, $h_a(v;\beta)$ is a.s. continuous in $x(W)$.

(h2$'$). $\dot{h}_{ab,\beta}(v;\beta)$ is continuous. Under H_a,

$$h_{ab}(v;\hat{\beta}_n) - h_{ab}(v;\beta_0) - (\hat{\beta}_n - \beta_0)'\dot{h}_{ab,\beta}(v;\beta_0) = O_p(\|\hat{\beta}_n - \beta_0\|^2).$$

(h3$'$). The integrating measure W has a Lebesgue density w such that for some $\delta > 0$,

$$\int \|\dot{h}_{a,\beta}(v;\beta_0)\|^{2+\delta} h_a(v;\beta_0)w^2(v)dv < \infty.$$

The following theorem states the consistency of the $\mathscr{T}_n$-test. Its proof is given in the last section.

Theorem 4.1. *Under the conditions of (k1)-(k3), (b), (w), (d), (h1$'$)-(h3$'$), H_a, and the additional assumption that*

$$\int [h_a(v;\beta_0) - h(v;\theta_a,\beta_0)]^2 dW(v) > 0,$$

we have

$$nb^{1/2}\hat{\Gamma}_n^{-1/2}|T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n) - \hat{C}_n| \to_p \infty.$$

Consequently, the above $\mathscr{T}_n$-test is consistent against H_a.

4.2. Asymptotic power at local alternatives

Here we shall study asymptotic power of the proposed $\mathscr{T}_n$-test against some local alternatives and the choice of an optimal W that maximizes this power against these alternatives. Accordingly, let φ be a known continuous density on $\mathbb{R}$ with mean 0 and positive variance σ_φ^2, and let $\delta_n := 1/\sqrt{nb^{1/2}}$. Consider the sequence of local alternatives:

$$H_{\ell oc} : f(v) = (1-\delta_n)f(v,\theta_0) + \delta_n\varphi(v), \quad v \in \mathbb{R}. \tag{4.1}$$

Under $H_{\ell oc}$, we shall assume that that $\sqrt{n}\Delta_n$ has the same asymptotic normal distribution as under the null hypothesis. In fact, one can show that the MD estimator $\hat{\theta}_n$ satisfies this assumption.

The following theorem gives asymptotic power of the $\mathscr{T}_n$-test against the local alternative $H_{\ell oc}$. Its proof also appears in the last section. Let

$$\begin{aligned} D(v;\beta,\theta) &:= \int [f(v+u'\beta,\theta)-\varphi(v+u'\beta)]g(u)du, \quad \beta\in\mathbb{R},\theta\in\Theta, \\ D(v) &:= D(v;\beta_0,\theta_0), \qquad K_*(v):=\int K(u)K(v+u)du, \quad v\in\mathbb{R}, \\ c &:= 2\int K_*^2(v)dv, \qquad \Gamma:=c\int h_0^2(v)w^2(v)dv. \end{aligned}$$

Theorem 4.2. *Under the conditions of Theorem 3.1, and under $H_{\ell oc}$,*

$$nb^{1/2}\hat\Gamma_n^{-1/2}(T_n(\hat\alpha_n,\hat\beta_n,\hat\theta_n)-\hat C_n)\to_D N\big(\Gamma^{-1/2}\int D^2(v)dW(v),1\big).$$

Remark 4.1. *Optimal W*. From this theorem we conclude that the asymptotic power of the asymptotic level α $\mathscr{T}_n$-test is

$$1-\Phi\Big(z_{\alpha/2}-\Gamma^{-1/2}\int D^2(v)w(v)dv\Big).$$

Clearly, the w that will maximize this power is the one that maximizes

$$\psi(w):=\Gamma^{-1/2}\int D^2(v)w(v)dv.$$

But,

$$\psi(w) \quad = \quad \frac{\int D^2(v)w(v)dv}{\sqrt{c\int h^2(v)w^2(v)dv}}\le c^{-1/2}\Big(\int\frac{D^4(v)}{h^2(v)}dv\Big)^{1/2},$$

with equality if, and only if, $D^2(v)/h^2(v)\propto w(v)$, for all v. Since $\psi(aw)=\psi(w)$ for all $a>0$, we may take optimal w to be

$$w(v)=\frac{D^2(v)}{h^2(v)}=\Big(\frac{\int[f(v+\beta_0'u,\theta_0)-\varphi(v+\beta_0'u)]g(u)du}{\int f(v+\beta_0'u,\theta_0)g(u)du}\Big)^2.$$

Clearly this w is unknown because of β_0 and θ_0, but one can estimate it by w_n, the analog of w where these parameters are replaced by $\hat\beta_n$ and $\hat\theta_n$.

5. Simulation Studies

To assess the finite sample performance of the BR-type test $\mathscr{T}_n$, we conduct some simulation studies in this section. The null hypothesis H_0 is chosen to be $\varepsilon\sim N(0,\sigma_\varepsilon^2)$, so the unknown parameter θ in the distribution of ε is σ_ε^2. Nine alternative hypotheses are considered: Double exponential distribution with mean 0 and variance 1, Cauchy distribution with location parameter 0 and scale parameter 1, Logistic distribution with location parameter 0 and scale parameter 1, t-distribution with degrees of freedom 3,5 and 10, two-component normal mixture models $0.5N(c,\sigma_\varepsilon^2)+0.5N(-c,\sigma_\varepsilon^2)$ with $c=0.5,0.75$ and 1.

In the simulation, we generate the data from model (1.1) with $\alpha = 1, \beta = 1$, $\sigma_\varepsilon^2 = \sigma_u^2 = 0.5^2$, $X \sim N(0,1)$, $u \sim N(0,\sigma_u^2)$. The weight measure W is taken to be a uniform distribution on a large closed interval so that computationally the integration over a finite interval is nearly same as the integration over the whole real line, the kernel function K is chosen to be standard normal density function, and the bandwidth is chosen to be $b = n^{-0.27}$ based on the condition (b). For each scenario, we repeat the testing procedure 500 times, and the empirical level and power are calculated from $\#\{|nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n) - \hat{C}_n)| \geq z_{\alpha/2}\}/500$. Here, $\hat{\alpha}_n$, $\hat{\beta}_n$ are bias-corrected estimators, $\hat{\theta}_n = \hat{\sigma}_\varepsilon^2 = \hat{s}_\xi^2 - \hat{\beta}_n^2\sigma_u^2$, with $\hat{s}_\xi^2$ is the sample variance of $\xi_i = Y_i - \hat{\alpha}_n - \hat{\beta}_n Z_i$, $i = 1,2,\ldots,n$. In our simulation, the significance level α is 0.05, and the sample sizes are chosen to be 100 and 200.

For comparison, in addition to the BR-type test $\mathscr{T}_n$, we also conduct simulation studies using Kolmogorov-Smirnov (KS) test in which the normality of ξ_i is checked, and Bootstrap BR-type test using the algorithm provided in Section 3 with $B = n$, $b^* = B^{-0.27}$, and $R = 200$. The simulation results are present in the following table.

Model	KS Test		BR-type Test		Bootstrap BR-type Test	
	100	200	100	200	100	200
$N(0,\sigma_\varepsilon^2)$	0.000	0.000	0.002	0.000	0.054	0.052
Logistic$(0,1)$	0.002	0.002	0.012	0.030	0.038	0.048
Cauchy$(0,1)$	0.996	1.000	0.990	0.994	0.998	1.000
Double Exponential $(0,1)$	0.024	0.088	0.150	0.386	0.194	0.452
$t(3)$	0.130	0.360	0.268	0.594	0.344	0.692
$t(5)$	0.014	0.028	0.054	0.130	0.094	0.134
$t(10)$	0.002	0.000	0.006	0.012	0.020	0.016
$0.5N(0.5,\sigma_\varepsilon^2)+0.5N(-0.5,\sigma_\varepsilon^2)$	0.000	0.000	0.002	0.004	0.044	0.026
$0.5N(0.75,\sigma_\varepsilon^2)+0.5N(-0.75,\sigma_\varepsilon^2)$	0.002	0.002	0.026	0.074	0.044	0.136
$0.5N(1,\sigma_\varepsilon^2)+0.5N(-1,\sigma_\varepsilon^2)$	0.010	0.100	0.316	0.724	0.338	0.744

From the simulation simulation, one can see that both the KS test and the BR-type test are conservative. It is also evident that the BR-type test is more powerful than KS test for almost all chosen scenarios, while the simulation result from Bootstrap algorithm is more desirable.

6. Proofs

This section contains the proofs of some of the previously stated results.

Proof of Theorem 3.1. Adding and subtracting $h_n(v)$, $h_b(v;\beta_0,\theta_0)$, the statistic (2.1) can be written as the sum of the following six terms:

$$
\begin{aligned}
T_{n1} &= nb^{1/2}\int[\hat{h}_n(v) - h_n(v)]^2 dW(v), \\
T_{n2} &= nb^{1/2}\int[h_n(v) - h_b(v;\beta_0,\theta_0)]^2 dW(v), \\
T_{n3} &= nb^{1/2}\int[h_b(v;\hat{\beta}_n,\hat{\theta}_n) - h_b(v;\beta_0,\theta_0)]^2 dW(v),
\end{aligned}
$$

$$\begin{aligned}
T_{n4} &= 2nb^{1/2}\int[\hat{h}_n(v)-h_n(v)][h_n(v)-h_b(v;\beta_0,\theta_0)]dW(v),\\
T_{n5} &= 2nb^{1/2}\int[\hat{h}_n(v)-h_n(v)][h_b(v;\beta_0,\theta_0)-h_b(v;\hat{\beta}_n,\hat{\theta}_n)]dW(v),\\
T_{n6} &= 2nb^{1/2}\int[h_n(v)-h_b(v;\beta_0,\theta_0)][h_b(v;\beta_0,\theta_0)-h_b(v;\hat{\beta}_n,\hat{\theta}_n)]dW(v).
\end{aligned}$$

We claim $T_{n1}=o_p(1)$. To see this, by Taylor expansion, write T_{n1} as the sum of the following ten terms.

$$\begin{aligned}
T_{n11} &= nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)K_{bi}^{(1)}(v)\Big]^2dW(v),\\
T_{n12} &= nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^2K_{bi}^{(2)}(v)\Big]^2dW(v),\\
T_{n13} &= nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^3K_{bi}^{(3)}(v)\Big]^2dW(v),\\
T_{n14} &= nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^4K^{(4)}(\tilde{\xi}_i)\Big]^2dW(v),\\
T_{n15} &= -nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)K_{bi}^{(1)}(v)\Big]\\
&\qquad\times\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^2K_{bi}^{(2)}(v)\Big]dW(v),\\
T_{n16} &= \frac{1}{3}nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)K_{bi}^{(1)}(v)\Big]\\
&\qquad\times\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^3K_{bi}^{(3)}(v)\Big]dW(v),\\
T_{n17} &= -\frac{1}{12}nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)K_{bi}^{(1)}(v)\Big]\\
&\qquad\times\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^4K^{(4)}(\tilde{\xi}_i)\Big]dW(v),\\
T_{n18} &= -\frac{1}{6}nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^2K_{bi}^{(2)}(v)\Big]\\
&\qquad\times\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^3K_{bi}^{(3)}(v)\Big]dW(v),\\
T_{n19} &= \frac{1}{24}nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^2K_{bi}^{(2)}(v)\Big]\\
&\qquad\times\Big[\frac{1}{n}\sum_{i=1}^n\Big(\frac{a_n}{b}+\frac{d_n'Z_i}{b}\Big)^4K^{(4)}(\tilde{\xi}_i)\Big]dW(v),
\end{aligned}$$

$$T_{n110} = -\frac{1}{72} nb^{1/2} \int \Big[\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{a_n}{b}+\frac{d_n' Z_i}{b}\Big)^3 K_{bi}^{(3)}(v)\Big] \times \Big[\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{a_n}{b}+\frac{d_n' Z_i}{b}\Big)^4 K^{(4)}(\tilde{\xi}_i)\Big] dW(v).$$

For the sake of simplicity, we shall assume $d = 1$. The argument for the case of $d > 1$ is straightforward. Let's consider T_{n11}. It is bounded above by the sum

$$2nb^{1/2}\Big\{\frac{a_n^2}{b^2}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{bi}^{(1)}(v)\Big]^2 dW(v)+\frac{d_n^2}{b^2}\int\Big[\frac{1}{n}\sum_{i=1}^{n}Z_iK_{bi}^{(1)}(v)\Big]^2 dW(v)\Big\}$$
$$= nb^{1/2}O_p(1/(nb^2))\,O_p(1/(nb)+b^2) = o_p(1).$$

Similarly T_{n12} is bounded above by

$$2nb^{1/2}\Big\{\frac{a_n^4}{b^4}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{bi}^{(2)}(v)\Big]^2 dW(v)+\frac{d_n^4}{b^4}\int\Big[\frac{1}{n}\sum_{i=1}^{n}Z_iK_{bi}^{(2)}(v)\Big]^2 dW(v)\Big\}$$
$$= nb^{1/2}O_p(1/(n^2b^4))\,O_p(1/(nb)+b^4) = o_p(1).$$

Similarly, one can show that $T_{n13} = nb^{1/2}O_p(1/(n^3b^6))\,O_p(1/(nb)+b^4) = o_p(1)$, and $T_{n14} = nb^{1/2}O_p(1/(n^4b^8))\,O_p(1/b^2) = o_p(1)$. By similar arguments and the Cauchy-Schwarz inequality, we can show that $T_{1nk} = o_p(1)$ for $k = 5,\ldots,10$. Therefore $T_{n1} = o_p(1)$.

Now, consider T_{n3}. By (h4) and the $\sqrt{n}$-consistency of $\hat{\theta}_n$ and $\hat{\beta}_n$, $T_{n3} = nb^{1/2}O_p(n^{-1}) = o_p(1)$. By the Cauchy-Schwarz inequality, T_{n4} is bounded above by $2T_{n1}^{1/2}T_{n2}^{1/2}$. We shall show later that $T_{n2} = O_p(1)$, so $T_{n4} = o_p(1)$. Similarly, one can show that $T_{n5} = o_p(1)$.

Next, consider T_{n6}. Note that

$$h_b(v;\hat{\beta}_n,\hat{\theta}_n) - h_b(v;\beta_0,\theta_0) = \Delta_n'\dot{h}_{b\theta}(v;\beta_0,\theta_0) + d_n'\dot{h}_{b\beta}(v;\beta_0,\theta_0) + O_p(n^{-1}).$$

Hence, T_{n6} can be written as the sum of the following three terms.

$$T_{n61} = -2nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{bi}(v)-EK_{b1}(v)\Big]\dot{h}_{b\theta}'(v;\beta_0,\theta_0)dW(v)\Delta_n,$$
$$T_{n62} = -2nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{bi}(v)-EK_{b1}(v)\Big]\dot{h}_{b\beta}'(v;\beta_0,\theta_0)dW(v)d_n,$$
$$T_{n63} = 2nb^{1/2}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{bi}(v)-EK_{b1}(v)\Big]dW(v)\,O_p(n^{-1}).$$

We can show that

$$\sqrt{n}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{bi}(v)-EK_{b1}(v)\Big]\dot{h}_{b\theta}(v;\beta_0,\theta_0)dW(v) = O_p(1). \tag{6.1}$$

In fact, denote $s_{ni} = \int\Big[K_{bi}(v)-EK_{b1}(v)\Big]\dot{h}_{b\theta}'(v)dW(v)$, then

$$\sqrt{n}\int\Big[\frac{1}{n}\sum_{i=1}^{n}K_{bi}(v)-EK_{b1}(v)\Big]\dot{h}_{b\theta}'(v)dW(v) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_{ni}.$$

In the following, we shall show the asymptotic normality of the above entity. For convenience, we shall give the proof here only for the case $q=1$, i.e., when $\dot{h}_{b\theta}(v)$ is one dimensional. For multidimensional case the result can be prove by using linear combination o its components, and applying the same argument. By the Lindeberg-Feller CLT, it suffices to verify that for all $\lambda>0$, Es_{n1}^2 converges to some positive number, and

$$Es_{n1}^2 I[|s_{n1}|\geq \sqrt{n}\lambda]\to 0.$$

To show this, we have

$$Es_{n1}^2=\mathrm{Var}(s_{n1})=E\left[\int K_{b1}(v)\dot{h}_{b\theta}(v)dW(v)\right]^2-\left[\int EK_{b1}(v)\dot{h}_{b\theta}(v)dW(v)\right]^2. \quad (6.2)$$

The first term on the right hand side of (6.2) equals

$$\begin{aligned} & E\iint K_{b1}(x)\dot{h}_{b\theta}(x)K_{b1}(y)\dot{h}_{b\theta}(y)dW(x)dW(y) \\ =\; & \iiint K_b(x-u)\dot{h}_{b\theta}(x;\beta_0,\theta_0)K_b(y-u)\dot{h}_{b\theta}(y)h(u;\beta_0,\theta_0)dW(x)dW(y) \\ =\; & \iiint K(x)K(y)\dot{h}_{b\theta}(u+xb)\dot{h}_{b\theta}(u+yb)h(u)w(u+xb)w(u+yb)dudxdy \\ \to\; & \iiint K(x)K(y)\dot{h}_{\theta}^2(u)h(u)w(u)w(u)dudxdy=\int[\dot{h}_\theta(u)w(u)]^2h(u)du. \end{aligned}$$

Similarly, $\int EK_{b1}(v)\dot{h}_{b\theta}(v)dW(v)\to\int h(v)\dot{h}_\theta(v)w(v)dv$. Hence,

$$\begin{aligned} \mathrm{Var}(s_{n1}) &\to \int[\dot{h}_\theta(u)w(u)]^2h(u)du-\left[\int h(v)\dot{h}_\theta(v)w(v)dv\right]^2 \\ &= \mathrm{Var}(\dot{h}_\theta(\xi)w(\xi)). \end{aligned}$$

or, in the multidimensional case, $\mathrm{Cov}(s_{n1})\to\mathrm{Cov}(\dot{h}_\theta(\xi)w(\xi))$. To verify the Lindeberg-Feller condition, note that for any $\delta>0$, the LHS is bounded above by $\lambda^{-\delta}n^{-\delta/2}Es_{n1}^{2+\delta}$. But,

$$\begin{aligned} E|s_{n1}|^{2+\delta} &= E\left|\int[K_{b1}(v)-EK_{b1}(v)]\dot{h}_{b\theta}(v)dW(v)\right|^{2+\delta} \\ &\leq 2^{1+\delta}\Big(E\left|\int K_{b1}(v)\dot{h}_{b\theta}(v)dW(v)\right|^{2+\delta} \\ &\qquad +\left|\int EK_{b1}(v)\dot{h}_{b\theta}(v)dW(v)\right|^{2+\delta}\Big). \end{aligned}$$

The second term on the RHS is $O(1)$ while the first term, by Hölder's inequality, conditions (k2), (h3), is bounded above by

$$E\Big[\int[K_{b1}(v)]^{1+\delta/2}|\dot{h}_{b\theta}(v)|^{1+\delta/2}dW(v)\Big]^2=O(b^{-\delta}).$$

Consequently, $n^{-\delta/2}E|s_{n1}|^{2+\delta}=O(n^{-\delta/2}b^{-\delta})=O\big((nb^2)^{-\delta/2}\big)=o(1)$. This fact verifies the Lindeberg-Feller condition here. Therefore,

$$\sqrt{n}\int\Big[\frac{1}{n}\sum_{i=1}^n K_{bi}(v)-EK_{b1}(v)\Big]\dot{h}'_{b\theta}(v)dW(v)=O_p(1)$$

and this implies $T_{n61} = o_p(1)$. Similarly, one can show that $T_{n62} = o_p(1)$ and $T_{n63} = o_p(1)$.

Putting all of the above facts together yields

$$nb^{1/2}T_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\theta}_n) = T_{n2} + o_p(1).$$

Now, define

$$\begin{aligned} C_n &:= \frac{1}{n^2}\sum_{i=1}^{n}\int \Big[K_{bi}(v) - EK_{b1}(v)\Big]^2 dW(v), \qquad (6.3)\\ H_n(\xi_i,\xi_j) &:= \frac{b^{1/2}}{n}\int [K_{bi}(v) - EK_{b1}(v)][K_{bj}(v) - EK_{b1}(v)]dW(v). \end{aligned}$$

Then,

$$\begin{aligned} T_{n2} &= nb^{1/2}C_n + \frac{2b^{1/2}}{n}\sum_{1\le i<j\le n}\int [K_{bi}(v) - EK_{b1}(v)][K_{bj}(v) - EK_{b1}(v)]dW(v)\\ &= nb^{1/2}C_n + 2\sum_{1\le i<j\le n} H_n(\xi_i,\xi_j). \end{aligned}$$

To proceed further, we need to recall Theorem 1 in Hall (1984) [6] which is reproduced here for the sake of completeness as

Lemma 6.1. *Let U_i, $1 \le i \le n$, be i.i.d. random vectors, and*

$$V_n := \sum_{1\le i<j\le n} H_n(U_i,U_j), \quad G_n(u,v) = EH_n(U_1,u)H_n(U_1,v),$$

where H_n is a sequence of measurable functions symmetric under permutation, with $EH_n^2(U_1,U_2) < \infty$, and $E(H_n(U_1,U_2)|U_1) = 0$, a.s., for each $n \ge 1$. If, additionally,

$$\frac{EG_n^2(U_1,U_2) + n^{-1}EH_n^4(U_1,U_2)}{[EH_n^2(U_1,U_2)]^2} \to 0,$$

then V_n is asymptotically normally distributed with mean 0 and variance $n^2EH_n^2(U_1,U_2)/2$.

Apply the above lemma to $U_i = \xi_i$ and H_n defined in (6.3). Obviously, this $H_n(\xi_i,\xi_j)$ is symmetric and $E[H_n(\xi_i,\xi_j)|\xi_i] = 0$. Moreover,

$$EH_n^2(\xi_1,\xi_2) = \frac{b}{n^2}\iint \Big(E[K_{b1}(x) - EK_{b1}(x)][K_{b1}(y) - EK_{b1}(y)]\Big)^2 dW(x)dW(y).$$

By the change of variable formula,

$$\begin{aligned} &E[K_{b1}(x) - EK_{b1}(x)][K_{b1}(y) - EK_{b1}(y)]\\ &= \int [K_b(x-u) - EK_b(x-\xi)][K_b(y-u) - EK_b(y-\xi)]h(u)du\\ &= \int K_b(x-u)K_b(y-u)h(u)du - EK_b(x-\xi)K_b(y-\xi)\\ &= \int K(v)\frac{1}{b}K\Big(\frac{y-x}{b} + v\Big)h(x-bv)dv - h_0(v)h(y) + O(b^2). \end{aligned}$$

Therefore, by changing variable again, $EH_n^2(\xi_1,\xi_2)$ is the sum of

$$\frac{1}{n^2}\iint\Big[\int K(v)K(u+v)h(x-bv)dv - bh(v)h(x+bu)\Big]^2 dW(x)dW(y)$$

and another term of the order $O(b^6)$, which together imply

$$\begin{aligned}\frac{n^2}{2}EH_n^2(\xi_1,\xi_2) &\to \frac{1}{2}\iint\Big[\int K(v)K(u+v)dvh(v)\Big]^2 w^2(x)dxdu \qquad (6.4)\\ &= \frac{1}{2}\int\Big[\int K(v)K(u+v)dv\Big]^2 du\int h^2(x)w^2(x)dx.\end{aligned}$$

Now consider $G_n(x,y) = EH_n(\xi_1,x)H_n(\xi_1,y)$. Note that

$$\begin{aligned}&G_n(x,y)\\ &= \frac{b}{n^2}\iint E\{[K_b(v-\xi_1)-EK_{b1}(v)][K_b(u-\xi_1)-EK_{b1}(u)]\}\\ &\quad\times[K_b(v-x)-EK_b(v-\xi)][K_b(u-y)-EK_b(u-\xi)]dW(u)dW(v).\end{aligned}$$

By the change of variables formula,

$$\begin{aligned}&E[K_b(v-\xi)-EK_b(v-\xi)][K_b(u-\xi)-EK_b(u-\xi)]\\ &= \int K_b(v-x)K_b(u-x)h(x)dx - EK_b(v-\xi)K_b(u-\xi)\\ &= \frac{1}{b}\int K(x)K\Big(\frac{u-v}{b}+x\Big)h(v-xb)dx - h_0(v)h(u)+O(b^2).\end{aligned}$$

Using this, direct calculations show that

$$EG_n^2(\xi_1,\xi_2) = O(b/n^4).$$

Similarly, expanding the 4th power and using change of variables formula, one verifies

$$\begin{aligned}EH_n^4(\xi_1,\xi_2) &= \frac{b^2}{n^4}E\Big[\int[K_b(v-\xi_1)-EK_b(v-\xi)][K_b(v-\xi_2)-EK_b(v-\xi)]dW(v)\Big]^4\\ &= O(1/(n^4b)).\end{aligned}$$

From (6.4), we know that $EH_n^2(\xi_1,\xi_2)=O(n^{-2})$. Therefore,

$$\frac{EG_n^2(\xi_1,\xi_2)}{[EH_n^2(\xi_1,\xi_2)]^2} = \frac{O(b/n^4)}{O(1/n^4)} = O(b) = o(1),$$

$$\frac{n^{-1}EH_n^4(\xi_1,\xi_2)}{[EH_n^2(\xi_1,\xi_2)]^2} = \frac{n^{-1}O(1/(n^4b))}{O(1/n^4)} = O(1/(nb)) = o(1).$$

This verifies the applicability of Lemma 6.1. In view of (6.4), we conclude

$$nb^{1/2}(T_n(\hat\alpha_n,\hat\beta_n,\hat\theta_n)-C_n)\to_D N(0,\Gamma), \qquad (6.5)$$

where

$$\Gamma := 2\int h^2(v)w^2(v)dv \int \big(K_*(u)\big)^2 du, \quad K_*(u) := \int K(v)K(u+v)dv.$$

Direct derivations verify that $\hat{\Gamma}_n$ of (3.1) is a consistent estimator of Γ. We shall now show $\hat{C}_n$ of (3.1) is an $nb^{1/2}$-consistent estimator of C_n. Let

$$\tilde{C}_n := \frac{1}{n^2}\sum_{i=1}^n \int K_{bi}^2(v)dW(v).$$

Note that

$$C_n = \tilde{C}_n + \frac{1}{n}\int [EK_{b1}(v)]^2 dW(v) - \frac{2}{n^2}\sum_{i=1}^n \int K_{bi}(v)EK_{b1}(v)dW(v).$$

But,

$$nb^{1/2}\frac{1}{n}\int [EK_{b1}(v)]^2 dW(v) = O(b^{1/2}) = o(1),$$

and

$$\begin{aligned} nb^{1/2}\frac{1}{n^2}\sum_{i=1}^n \int K_{bi}(v)EK_{b1}(v)dW(v) &= 2b^{1/2}\int h_n(v)EK_{b1}(v)dW(v) \\ &= O_p(b^{1/2}) = o_p(1). \end{aligned}$$

Hence, $nb^{1/2}C_n = nb^{1/2}\tilde{C}_n + o_p(1)$. We claim

$$nb^{1/2}(\hat{C}_n - \tilde{C}_n) = o_p(1). \tag{6.6}$$

For this purpose, note that $\hat{C}_n$ can be written as the sum of $\tilde{C}_n$ and the following two terms

$$\begin{aligned} C_{n1} &= \frac{1}{n^2}\sum_{i=1}^n \int [K_b(v-\hat{\xi}_i) - K_b(v-\xi_i)]^2 dW(v), \\ C_{n2} &= \frac{2}{n^2}\sum_{i=1}^n \int [K_b(v-\hat{\xi}_i) - K_b(v-\xi_i)]K_b(v-\xi_i)dW(v). \end{aligned}$$

By Taylor expansion, with $\tilde{\xi}_i$ between $v-\hat{\xi}_i$ and $v-\xi_i$,

$$\begin{aligned} nb^{1/2}C_{n1} &= \frac{b^{1/2}}{n}\sum_{i=1}^n \int \Big[\Big(\frac{a_n}{b} + \frac{d_n' Z_i}{b}\Big)\frac{1}{b}K^{(1)}(\tilde{\xi}_i)\Big]^2 dW(v) \\ &\le \frac{2a_n^2}{b^{7/2}}\frac{1}{n}\sum_{i=1}^n \int [K^{(1)}(\tilde{\xi}_i)]^2 dW(v) + \frac{2d_n^2}{b^{7/2}}\frac{1}{n}\sum_{i=1}^n \int [Z_i K^{(1)}(\tilde{\xi}_i)]^2 dW(v) \\ &= O_p\Big(\frac{1}{nb^{7/2}}\Big) = o_p(1), \end{aligned}$$

and

$$\begin{aligned} nb^{1/2}C_{n2} &= -\frac{2b^{1/2}}{n}\sum_{i=1}^n \int \Big[\Big(\frac{a_n}{b} + \frac{d_n' Z_i}{b}\Big)\frac{1}{b}K^{(1)}(\tilde{\xi}_i)\Big]K_{bi}(v)dW(v) \\ &= -\frac{2a_n}{b^{3/2}}\frac{1}{n}\sum_{i=1}^n \int K^{(1)}(\tilde{\xi}_i)K_{bi}(v)dW(v) - \frac{2d_n'}{b^{3/2}}\frac{1}{n}\sum_{i=1}^n \int Z_i K^{(1)}(\tilde{\xi}_i)K_{bi}(v)dW(v) \\ &= O_p\Big(\frac{1}{\sqrt{n}b^{3/2}}\Big) = o_p(1). \end{aligned}$$

Hence, $nb^{1/2}(\hat{C}_n - \tilde{C}_n) = o_p(1)$. This implies $nb^{1/2}(T_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\theta}_n) - \hat{C}_n) \to_D N(0,\Gamma)$, thereby completing the proof of Theorem 3.1.

Proof of Theorem 4.1: Let $h_{ab}(v,\beta) = \int K_b(v-u)h(u;\beta)du$. Add and subtract $h_{ab}(v;\hat{\beta}_n)$ from $\hat{h}_n(v) - h_b(v;\hat{\beta}_n,\hat{\theta}_n)$ and expand the quadratic term to write $T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n)$ as the sum of the following three terms:

$$
\begin{aligned}
T_{n1} &= \int [\hat{h}_n(v) - h_{ab}(v;\hat{\beta}_n)]^2 dW(v),\\
T_{n2} &= \int [h_{ab}(v;\hat{\beta}_n) - h_b(v;\hat{\beta}_n,\hat{\theta}_n)]^2 dW(v),\\
T_{n3} &= 2\int [\hat{h}_n(v) - h_{ab}(v;\hat{\beta}_n)][h_{ab}(v;\hat{\beta}_n) - h_b(v;\hat{\beta}_n,\hat{\theta}_n)] dW(v).
\end{aligned}
$$

One can show that

$$nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_{n1} - \hat{C}_n) \to_D N(0,1), \tag{6.7}$$

The proof is similar to that of Theorem 3.1. Note that now

$$\hat{\Gamma}_n \to_p 2\int h_a^2(v;\beta_0)w^2(v)dv \int [K_*(u)]^2 du.$$

Also,

$$nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{n2} \;=\; nb^{1/2}\Gamma^{-1/2}\int [h_a(v;\beta_0) - h(v;\theta_a,\beta_0)]^2 dW(v) + o_p(nb^{1/2}). \tag{6.8}$$

By the Cauchy-Schwarz inequality and the elementary inequality $(a+c)^{1/2} \le a^{1/2} + c^{1/2}$ for $a \ge 0, c \ge 0$, one can show that $nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{n3}$ is bounded above by

$$2nb^{1/2}\hat{\Gamma}_n^{-1/2}|T_{n1} - \hat{C}_n|^{1/2}T_{n2}^{1/2} + 2nb^{1/2}\hat{\Gamma}_n^{-1/2}\hat{C}_n^{1/2}T_{n2}^{1/2}.$$

From (6.7), one can see that the first term is $o_p(nb^{1/2})$. Note that $\hat{C}_n \to 0$ in probability, in fact, one can show that $\hat{C}_n = O_p(1/(nb))$. So the second term is also $o_p(nb^{1/2})$. Therefore,

$$
\begin{aligned}
&nb^{1/2}\hat{\Gamma}_n^{-1/2}|T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n) - \hat{C}_n|\\
&\quad= nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_{n1} - \hat{C}_n) + nb^{1/2}\hat{\Gamma}_n^{-1/2}\int [h_a(v;\beta_0) - h(v;\theta_a,\beta_0)]^2 dW(v) + o_p(nb^{1/2}).
\end{aligned}
$$

Hence the theorem.

Proof of Theorem 4.2: Let

$$D_b(v;\beta,\theta) = \int K_b(v-u)D(u;\beta,\theta)du, \quad D_b(v) = \int K_b(v-u)D(u;\beta_0,\theta_0)du,$$

$$\tilde{h}_{nb}(v;\beta,\theta) := h_b(v;\beta,\theta) - \delta_n D_b(v;\beta,\theta), \quad \tilde{h}_{nb}(v) := \tilde{h}_{nb}(v;\beta_0,\theta_0).$$

Adding and subtracting $\delta_n D_b(v;\hat{\beta}_n,\hat{\theta}_n)$ from $\hat{h}_n(v) - h_b(v;\hat{\beta}_n,\hat{\theta}_n)$ in the integrand of $T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n)$, one can rewrite it as the sum of the following three terms

$$
\begin{aligned}
T_{n1} &= \int [\hat{h}_n(v) - \tilde{h}_{nb}(v;\hat{\beta}_n,\hat{\theta}_n)]^2 dW(v),\\
T_{n2} &= \frac{1}{nb^{1/2}}\int D_b^2(v;\hat{\beta}_n,\hat{\theta}_n) dW(v),\\
T_{n3} &= -\frac{2}{\sqrt{nb^{1/2}}}\int [\hat{h}_n(v) - \tilde{h}_{nb}(v;\hat{\beta}_n,\hat{\theta}_n)] D_b(v;\hat{\beta}_n,\hat{\theta}_n) dW(v).
\end{aligned}
$$

Adding and subtracting $h_n(v)$, $\tilde{h}_{nb}(v)$, T_{n1} can be written as the sum of the following six terms:

$$\begin{aligned}
T_{n11} &= \int [\hat{h}_n(v) - h_n(v)]^2 dW(v), \\
T_{n12} &= \int [h_n(v) - \tilde{h}_{nb}(v)]^2 dW(v), \\
T_{n13} &= \int [\tilde{h}_{nb}(v;\hat{\beta}_n,\hat{\theta}_n) - \tilde{h}_{nb}(v)]^2 dW(v), \\
T_{n14} &= 2\int [\hat{h}_n(v) - h_n(v)][h_n(v) - \tilde{h}_{nb}(v)] dW(v), \\
T_{n15} &= -2\int [\hat{h}_n(v) - h_n(v)][\tilde{h}_{nb}(v) - \tilde{h}_{nb}(v;\hat{\beta}_n,\hat{\theta}_n)] dW(v), \\
T_{n16} &= -2\int [h_n(v) - \tilde{h}_{nb}(v)][\tilde{h}_{nb}(v) - \tilde{h}_{nb}(v;\hat{\beta}_n,\hat{\theta}_n)] dW(v).
\end{aligned}$$

One can show that $nb^{1/2}T_{n1j} = o_p(1)$ for $j = 1,3,4,5,6$. The details are similar to those of the proof of Theorem 3.1, only differences are stated here. For example, by assuming $d = 1$ and Taylor expansion, T_{n11} can be written as the sum of ten terms, one of these terms is

$$\int \Big[\frac{1}{nb}\sum_{i=1}^{n}\Big(\frac{a_n}{b} + \frac{d_n' Z_i}{b}\Big) K_{bi}^{(1)}(v)\Big]^2 dW(v)$$

which is bounded above by the sum

$$\frac{2a_n^2}{b^2}\int \Big[\frac{1}{nb}\sum_{i=1}^{n} K_{bi}^{(1)}(v)\Big]^2 dW(v) + \frac{2d_n^2}{b^2}\int \Big[\frac{1}{nb}\sum_{i=1}^{n} Z_i K_{bi}^{(1)}(v)\Big]^2 dW(v).$$

While the integral in the first term is bounded above by

$$2\int \Big(\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{b}K_{bi}^{(1)}(v) - EK_{b1}^{(1)}(v)\Big]\Big)^2 dW(v) + 2\int \Big[EK_{b1}^{(1)}(v)\Big]^2 dW(v).$$

But,

$$E\Big(\frac{1}{n}\sum_{i=1}^{n}\big[\frac{1}{b}K_{bi}^{(1)}(v) - EK_{b1}^{(1)}(v)\big]\Big)^2 \le \frac{1}{n}E\big[K_{b1}^{(1)}(v)\big]^2.$$

Moreover,

$$\begin{aligned}
E\Big[K_{b1}^{(1)}(v)\Big]^2 &= \int \Big[\frac{1}{b}K^{(1)}\Big(\frac{v-u}{b}\Big)\Big]^2 h_0(u)du - \delta_n \int \Big[\frac{1}{b}K^{(1)}\Big(\frac{v-u}{b}\Big)\Big]^2 D(u)du \\
&= O(b^{-1}), \\
EK_{b1}^{(1)}(v) &= \int \Big[\frac{1}{b}K^{(1)}\Big(\frac{v-u}{b}\Big)\Big] h_0(u)du - \delta_n \int \Big[\frac{1}{b}K^{(1)}\Big(\frac{v-u}{b}\Big)\Big] D(u)du \\
&= O(b).
\end{aligned}$$

Therefore,

$$\frac{2a_n^2}{b^2}\int \Big[\frac{1}{nb}\sum_{i=1}^{n} K_{bi}^{(1)}(v)\Big]^2 dW(v) = \frac{1}{nb^2} O_p\Big[\frac{1}{nb} + b^2\Big].$$

Similarly, one can show that

$$\frac{2d_n^2}{b^2}\int\Big[\frac{1}{nb}\sum_{i=1}^n Z_iK_{bi}^{(1)}(v)\Big]^2dW(v)=\frac{1}{nb^2}O_p\Big[\frac{1}{nb}+b^2\Big],$$

which implies $nb^{1/2}T_{n11}=o_p(1)$.

Suppose $D_b(v;\beta,\theta)$ also satisfies condition (h4), then $nb^{1/2}T_{n13}=o_p(1)$ follows from (h2), $\sqrt{n}$-consistency of $\hat\theta_n,\hat\beta_n$, and the following

$$T_{n13}\le 2\int[h_b(v;\hat\beta_n,\hat\theta_n)-h_b(v)]^2dW(v)+\frac{2}{nb^{1/2}}\int[D_b(v;\hat\theta_n,\hat\beta_n)-D_b(v)]^2dW(v).$$

In the following, we shall prove

$$nb^{1/2}T_{n3}=-2\sqrt{nb^{1/2}}\int[\hat h_n(v)-\tilde h_{nb}(v;\hat\beta_n,\hat\theta_n)]D_b(v;\hat\beta_n,\hat\theta_n)dW(v)=o_p(1).$$

Adding and subtracting $h_n(v),\tilde h_{nb}(v)$ from $\hat h_n(v)$, and $D_b(v)$ from $D_b(v;\hat\theta_n,\hat\beta_n)$, $nb^{1/2}T_{n3}$ can be written as the sum of the following six terms:

$$\begin{aligned}
I_{n1} &= 2\sqrt{nb^{1/2}}\int[\hat h_n(v)-h_n(v)]\,D_b(v)dW(v),\\
I_{n2} &= 2\sqrt{nb^{1/2}}\int[h_n(v)-\tilde h_{nb}(v)]\,D_b(v)dW(v),\\
I_{n3} &= 2\sqrt{nb^{1/2}}\int[\tilde h_{nb}(v)-\tilde h_{nb}(v;\hat\beta_n,\hat\theta_n)]\,D_b(v)dW(v),\\
I_{n4} &= 2\sqrt{nb^{1/2}}\int[\hat h_n(v)-h_n(v)]\,[D_b(v;\hat\beta_n,\hat\theta_n)-D_b(v)]dW(v),\\
I_{n5} &= 2\sqrt{nb^{1/2}}\int[h_n(v)-\tilde h_{nb}(v)]\,[D_b(v;\hat\beta_n,\hat\theta_n)-D_b(v)]dW(v),\\
I_{n6} &= 2\sqrt{nb^{1/2}}\int[\tilde h_{nb}(v)-\tilde h_{nb}(v;\hat\beta_n,\hat\theta_n)]\,[D_b(v;\hat\beta_n,\hat\theta_n)-D_b(v)]dW(v).
\end{aligned}$$

By the Cauchy-Schwarz inequality, I_{n1} is bounded above by

$$2\sqrt{nb^{1/2}}\Big(\int[\hat h_n(v)-h_n(v)]^2dW(v)\Big)^{1/2}\Big(\int D_b^2(v;\beta_0,\theta_0)dW(v)\Big)^2.$$

The previous discussion on T_{n11} and the square integrability of $D(v)$ with respect to W imply

$$I_{n1}=\sqrt{nb^{1/2}}\Big(\frac{1}{nb^2}O_p\Big[\frac{1}{nb}+b^2\Big]\Big)^{1/2}=o_p(1).$$

Similar to the proof of (6.1), we have $I_{n2}=o_p(1)$. By the $\sqrt{n}$-consistency of $\hat\beta_n$ and $\hat\theta_n$, one can also show that $I_{n3}=o_p(1)$. Finally, by Cauchy-Schwarz inequality, one can show that $I_{nj}=o_p(1)$ for $j=4,5,6$. Thus, $nb^{1/2}T_{n3}=o_p(1)$.

Using consistency of $\hat\beta_n$ and $\hat\theta_n$, one verifies $nb^{1/2}T_{n2}\to\int D^2(v)dW(v)$. We can also show that $\hat\Gamma_n$ and $\hat C_n$ have the same asymptotic properties as the ones in Section 4. The proofs are also similar, hence omitted here for the sake of brevity. This concludes the proof of Theorem 4.2.

Acknowledgement. Research of the first author was supported in part by the NSF DMS grant 0704130.

References

[1] P. J. Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimates. *Ann. Statist.*, 1:1071–1095, 1973.

[2] R.J. Carroll, D. Ruppert, and L.A. Stefanski. *Measurement Error in Nonlinear Models*. Chapman & Hall/CRC, Boca Raton, 1995.

[3] Bachmann D. and Dette H. A note on the bickel-rosenblatt test in autoregressive time series. *Statist. Probab. Lett.*, 74:221–234, 2005.

[4] R. B. D'Agostino and M. A. Stephens. *Goodness-of-fit techniques*. Marcel Dekker, Inc., New York, 1986.

[5] W.A. Fuller. *Measurement Error Models*. Wiley, New York, 1987.

[6] P.J. Hall. Central limit theorem for integrated square error of multivariate nonparametric density esitmators. *J. Multivariate Anal.*, 14:1–16, 1984.

[7] H.L. Koul and N. Mimoto. A goodness-of-fit test for garch innovation density. 2010. To appear in *Metrika*.

[8] H.L. Koul and P. Ni. Minimum distance regression model checking. *J. Stat. Plan. Inference*, 119:109–141, 2004.

[9] H.L. Koul and W. Song. Minimum distance regression model checking with berkson error. *Ann. Statist.*, 37:132–156, 2009.

[10] S. Lee and S. Na. On the bickel-rosenblatt test for first-order autoregressive models. *Statist. Probab. Lett.*, 56:23–35, 2002.