

## GOODNESS-OF-FIT TESTING OF ERROR DISTRIBUTION IN LINEAR MEASUREMENT ERROR MODELS<sup>1</sup>

BY HIRA L. KOUL   WEIXING SONG   XIAOQING ZHU

*Michigan State University and Kansas State University*

This paper investigates a class of goodness-of-fit tests for fitting an error density in linear regression models with measurement error in covariates. Each test statistic is the integrated square difference between the deconvolution kernel density estimator of the regression model error density and a smoothed version of the null error density, an analog of the so called Bickel and Rosenblatt test statistic. The asymptotic null distributions of the proposed test statistics are derived for both the ordinary smooth and super smooth cases. The asymptotic power behavior of the proposed tests against a fixed alternative and a class of local nonparametric alternatives for both cases is also described. The finite sample performance of the proposed test is evaluated by a simulation study. The simulation study shows some superiority of the proposed test over some other tests. Finally, a real data is used to illustrate the proposed test.

**1. Introduction.** The problem of fitting an error distribution in regression models has been well studied when covariates are fully observed, see, e.g., Loynes (1980), Koul (2002), Khamaldze and Koul (2004, 2009) and the references therein. However, there are examples where covariates are not observable as is demonstrated in the monographs of Fuller (1987), Cheng and Van Ness (1999) and Carroll, Rupert, Stefanski, and Crainiceanu (2006). In these situations one observes a surrogate that provides a measurement, up to an error, of the given covariate. The regression models where covariates are measured with error are often called errors-in-variables or measurement error regression models. Relatively little is known about fitting an error distribution in these models. This paper provides a class of tests for this testing problem based on deconvoluted density estimators of the error density.

Let  $p \geq 1$  be a given dimension of the covariate vector  $X$ . In a multiple linear regression model with measurement error in  $X$  one observes the response variable  $Y$  and a surrogate  $p$ -vector  $Z$  obeying the relations

$$(1.1) \quad Y = \alpha + \beta'X + \varepsilon, \quad Z = X + u,$$

---

*AMS 2000 subject classifications:* Primary 62G10; secondary 62G20.

<sup>1</sup>Research partly supported by the NSF DMS grant 1205271, 1205276.

*Corresponding author:* Hira L. Koul.

*Key words and phrases.* Deconvolution density estimators.  $L_2$ -distance tests.

for some  $\alpha \in \mathbb{R}$ ,  $\beta \in \mathbb{R}^p$ , where the  $p$ -vector  $u$  is the measurement error in  $X$ . Here, for an  $x \in \mathbb{R}^p$ ,  $x'$  denotes its transpose. The variables  $\varepsilon$ ,  $u$  and  $X$  are assumed to be mutually independent, with  $E\varepsilon = 0$  and  $Eu = 0$ . And for the model identifiability reasons, we assume the density  $g$  of the measurement error  $u$  to be known.

Let  $f$  denote the density of  $\varepsilon$ , and  $f_0$  be a known density with zero mean. The problem of interest in this paper is to test the hypothesis

$$(1.2) \quad H_0 : f = f_0 \quad v.s \quad H_1 : f \neq f_0,$$

based on a random sample  $(Y_i, Z_i)$ ,  $1 \leq i \leq n$  from the joint distribution of  $(Y, Z)$  obeying the model (1.1). The case of fitting a parametric family of densities to  $f$  is discussed in Remark 2.1 below. The case of unknown  $g$  is briefly discussed in Remark 2.2.

A motivation for considering the above testing problem is as follows. There are situations where knowing the distribution of  $\varepsilon$  helps to improve the efficiency of the estimators of  $\alpha$  and  $\beta$  in the above measurement error model. Commonly used estimators of  $\alpha$  and  $\beta$  are the so-called bias-corrected estimators:  $\hat{\beta} = (S_{ZZ} - \Sigma_u)^{-1}S_{ZY}$ ,  $\hat{\alpha} = \bar{Y} - \hat{\beta}'\bar{Z}$ , where  $S_{ZZ}$  is the sample covariance matrix of  $Z$ ,  $S_{ZY}$  is the sample covariance matrix of  $Z$  and  $Y$ ,  $\bar{Y}$  and  $\bar{Z}$  are the sample means of  $Y$ 's and  $Z$ 's, respectively, and  $\Sigma_u$  is the covariance matrix of  $u$ . These estimators are consistent and asymptotically normal even when the actual distribution of  $\varepsilon$  is misspecified within the class of distribution with finite variance. But in addition to the normality assumption on  $X$  and  $u$ , if we can further have normality of  $\varepsilon$ , then certainly we can construct more efficient estimators of  $\alpha$  and  $\beta$  than these bias-corrected estimators. For example, if  $p = 1$ ,  $X \sim N(1, 1)$ ,  $u \sim N(0, 1)$ , and  $\varepsilon \sim N(0, 1)$ , then, from Fuller (1987), another possible estimator for  $\beta$  is

$$\tilde{\beta} = [S_{YY} - S_{ZZ} + \sqrt{(S_{YY} - S_{ZZ})^2 + 4S_{ZY}^2}]/2S_{ZY}.$$

A simulation study shows that the mean squared error of  $\tilde{\beta}$  is smaller than that of the bias-corrected estimator  $\hat{\beta}$ , in particular, when the sample size is small. Therefore, by taking the distributional information of the regression error into account, it is possible to construct more efficient estimators of the underlying parameters in the model (1.1).

Note that if in (1.1),  $\beta = 0$ , then  $Y$  bears no relation with  $X$  and hence whether  $X$  is observable or not is irrelevant for making inference about  $f$ . In particular any goodness-of-fit test based on  $Y_i$ ,  $1 \leq i \leq n$ , useful for fitting a density up to an unknown location parameter may be used to test the above hypotheses. Thus, from now onwards we shall assume  $\beta \neq 0$  in this paper.

Since we observe  $Z$  instead of  $X$ , we shall rewrite the model (1.1) as

$$Y = \alpha + \beta'Z + e, \quad e = \varepsilon - \beta'u.$$

Because  $u$  and  $\varepsilon$  are independent, the density of  $e$  is  $h(v) = \int f(v + \beta'u)g(u)du$ ,  $v \in \mathbb{R}$ . Let  $h_0(v) = \int f_0(v + \beta'u)g(u)du$ ,  $v \in \mathbb{R}$ . As argued in Koul and Song (2012), there is a one-to-one map between the densities of  $\varepsilon$  and  $e$ . Hence, testing for  $H_0$  is equivalent to testing for

$$(1.3) \quad \mathcal{H}_0 : h = h_0, \quad \text{vs.} \quad \mathcal{H}_1 : h \neq h_0.$$

In the one sample i.i.d. set up, Bickel and Rosenblatt (1973) goodness-of-fit test for fitting a known density is based on an  $L_2$  distance between a kernel density estimator and its null expected value. This test, when adapted to fitting an error density up to an unknown location parameter, where the density estimator would be based on the estimated residuals, has the property that its asymptotic null distribution is not affected by not knowing the location parameter. 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), Bachmann and Dette (2005), and Koul and Mimoto (2012) observed that this fact continues to hold for the analog of this test statistic when fitting an error density based on residuals in autoregressive and generalized autoregressive conditionally heteroscedastic time series models. This type of property makes these  $L_2$ -distance type tests more desirable, compared to the tests based on residual empirical processes, because the asymptotic null distribution of the standardized residual empirical process depends on the estimators of the underlying nuisance parameters in these models in a complicated fashion, which renders it be unknown. In all of these works all data are completely observable.

In the above measurement error model, Koul and Song (2012) proposed analogous class of tests for the testing problem (1.3) based on kernel density estimators of  $h$  obtained from the residuals  $Y_i - \hat{\alpha} - \hat{\beta}'Z_i$ ,  $1 \leq i \leq n$ , where  $\hat{\alpha}, \hat{\beta}$  are some  $n^{1/2}$ -consistent estimators of  $\alpha, \beta$ , under  $H_0$ . Since this test is about the hypothesis pertaining to the density of  $e$  rather than that of  $\varepsilon$ , it is a direct testing procedure. As pointed out in Holzmann, Bissantz and Munk (2007), when discussing the goodness-of-fit tests for the density of the latent variable  $X$ , the direct method generally is inferior to the indirect ones in which the test statistic is based on the deconvolution estimate of the density of  $X$ . This motivates us to search for similar indirect tests tailored to our set up. Interestingly, when testing signals for a functional regression model with Gaussian errors, Laurent, Loubes and Marteau (2011) showed that the indirect testing procedures might be better than the direct ones.

There is a vast literature on the deconvolution estimators of the density of  $X$  in the measurement error model (1.1), as is evidenced in the papers of Carroll and Hall (1988), Stefanski and Carroll (1990), Fan (1991), van Es and Uh (2004), and Delaigle and Hall (2006). In addition to Holzman et al. (2007), the goodness-of-fit testing problem pertaining to the density function of  $X$  has been also studied by several authors including Butucea (2004), Holzman and Boysen (2006), and Loubes and Marteau (2014). All of these authors use analogs of the above  $L_2$ -distance type tests based either on the deconvoluted estimator of density of  $X$  or on a density estimator of  $Z$  density. None of them address the above problem of testing (1.2) or (1.3) pertaining to the error density in the above measurement error model (1.1).

Assume for the time being that  $\alpha, \beta$  in (1.1) are known. Since we observe  $Y$  and  $Z$ , we can construct a density estimator of density  $h$  of  $e := Y - \alpha - \beta'Z = \varepsilon - \beta'u$ , which is also an estimator of the convolution of the density  $f$  of  $\varepsilon$  with the known density of  $\beta'u$ . From this we obtain a deconvolution density estimator of  $f$ , which we shall use to construct tests of  $H_0$ .

Let  $\Phi_\gamma$  denote the characteristic function of a density  $\gamma$ . Let  $K$  be a kernel function having characteristic function  $\Phi_K$ . Let  $e_j := Y_j - \alpha - \beta'Z_j$ ,  $1 \leq j \leq n$  and  $b_n$  be a bandwidth sequence of positive numbers. We shall often write  $b$  for  $b_n$ . Then a kernel density estimator of  $h$ , when  $\alpha, \beta$  are known, is

$$h_n(x, \alpha, \beta) = \frac{1}{nb} \sum_{j=1}^n K\left(\frac{x - e_j}{b}\right).$$

Clearly, the characteristic function of  $h_n$  is  $\Phi_K(bt)\Psi_n(t)$ , where  $\Psi_n(t) = n^{-1} \sum_{j=1}^n e^{ite_j}$ . Assume that  $\Phi_g \neq 0$ . Then  $\Phi_f(t) = \Phi_h(t)/\Phi_g(-\beta t)$ . Replacing  $\Phi_h(t)$  with  $\Phi_K(bt)\Psi_n(t)$ , and by the inversion formula,

$$f_n(x, \alpha, \beta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi_K(bt) \frac{\Psi_n(t)}{\Phi_g(-\beta t)} dt,$$

is a deconvolution estimate of  $f$  when  $\alpha$  and  $\beta$  are known. But, in practice  $\alpha, \beta$  are seldom known. Let  $\hat{\alpha}, \hat{\beta}$  be estimators of  $\alpha, \beta$ , respectively. Then the corresponding deconvolution estimator of  $f$  is  $\hat{f}_n(x) := f_n(x, \hat{\alpha}, \hat{\beta})$ . The proposed class of tests, one for each  $K$  and  $b$ , of  $H_0$  is to be based on

$$\hat{T}_n = \int_{\mathbb{R}} (\hat{f}_n(x) - K_b * f_0(x))^2 dx,$$

where for any functions  $K$  and  $\gamma$ ,  $K_b(x) = b^{-1}K(x/b)$ ,  $K * \gamma(x) := \int K(x - y)\gamma(y)dy$ .

It is well known that the convergence rate of the deconvolution density estimators depends sensitively on the tail behaviour of the characteristic

function  $\Phi_g$  of the underlying measurement error. There are two general cases: one is the ordinary smooth case, where  $|\Phi_g(t)|$  is of polynomial order  $|t|^{-\kappa}$ , for some  $\kappa > 0$ , as  $|t| \rightarrow \infty$ ; the other is the super smooth case, where  $|\Phi_g(t)|$  is of the order  $|t|^{\lambda_0} e^{-|t|^\lambda/\nu}$ , for some  $\lambda_0 \in \mathbb{R}$ ,  $\lambda > 0$  and  $\nu > 0$ , as  $|t| \rightarrow \infty$ . In this paper, we obtain asymptotic distributions of  $\hat{T}_n$  under  $H_0$  in both the ordinary smooth and super smooth cases in section 2. The consistency against a fixed alternative, the asymptotic power against a class of local nonparametric alternatives and against a fixed alternative for both cases is described in section 3.

The findings of a finite sample simulation that compares the empirical power of a member of the proposed class of tests with that of Koul and Song (2012) test based on  $h_n(\cdot, \hat{\alpha}, \hat{\beta})$  are presented in section 4. The comparison is made for the five choices of the measurement error variance  $\sigma_u^2$  and three alternatives. In the super smooth case, while the empirical levels of both tests are conservative, the proposed test dominates the other test in terms of the empirical power at almost all chosen alternatives for all choices of  $\sigma_u^2 \geq 0.25$ . For more details see section 4. A supplement [Koul, Song and Zhu (2017)] to this paper describes the findings of a simulation that compares these two tests with the Kolmogorov–Smirnov and Cramér–von Mises tests based on the empirical d.f. of  $\{Y_j - \hat{\alpha} - \hat{\beta}'Z_j, 1 \leq j \leq n\}$ . It is observed that the  $\hat{T}_n$  test dominates these two tests for the larger chosen values of  $\sigma_u^2$  at most of the chosen alternatives and for a larger sample size.

In this paper,  $\mathcal{N}(\mu, \sigma^2)$  denotes the normal distribution with mean  $\mu$  and variance  $\sigma^2$ ,  $\rightarrow_d$  ( $\rightarrow_p$ ) denotes the convergence in distribution (probability), the range of integration in all the integrals is  $\mathbb{R}$ , and all limits are taken as  $n \rightarrow \infty$ , unless specified otherwise.

**2. Asymptotic Null Distribution.** This section discusses the asymptotic null distribution  $\hat{T}_n$  for the ordinary smooth and super smooth cases.

*2.1. Ordinary smooth case.* Here we shall first derive the limiting null distribution of  $\hat{T}_n$  for the ordinary smooth case. To begin with we state the needed assumptions.

- (A):  $\Phi_g$  satisfies  $\Phi_g(t) \neq 0$ , for all  $t \in \mathbb{R}^p$ , and  $|\Phi_g(t)| \approx \|t\|^{-\kappa}$ , for a  $\kappa > 0$ , i.e. there are  $c, C > 0$  such that  $c\|t\|^{-\kappa} \leq |\Phi_g(t)| \leq C\|t\|^{-\kappa}$ , for all sufficiently large  $\|t\|$ .
- (B):  $\Phi_f$  satisfies  $|\Phi_f(t)| = O(\|t\|^{-r})$ , for some  $r > 1$ , as  $\|t\| \rightarrow \infty$ .
- (C):  $\Phi_K$  is symmetric around 0 and compactly supported on  $[-1, 1]$ .
- (D):  $E\{\|X\|^4 + |\varepsilon|^4 + \|u\|^4\} < \infty$ .

Next, define  $\psi(\beta, s, t) := \Phi_g(\beta t + \beta s)\Phi_f(t + s)$ , and let

$$T_n(\alpha, \beta) := \int (f_n(x, \alpha, \beta) - K_b * f_0(x))^2 dx, \quad C_{M,b} := \int \frac{|\Phi_K(tb)|^2}{|\Phi_g(\beta t)|^2} dt,$$

$$C_{V,b} := \int \int \frac{|\Phi_K(tb)|^2 |\Phi_K(sb)|^2}{|\Phi_g(\beta t)|^2 |\Phi_g(\beta s)|^2} |\psi(\beta, s, t)|^2 ds dt.$$

Using Theorem 1 of Holzman et al. (2007) one can derive the following result. Suppose  $H_0$  and the assumptions (A)–(C) hold and  $b \rightarrow 0$ ,  $nb \rightarrow \infty$ . Then

$$(2.1) \quad C_{M,b} \approx b^{-(2\kappa+1)}, \quad C_{V,b} \approx b^{-(4\kappa+1)},$$

$$(2.2) \quad nC_{V,b}^{-1/2} \left( T_n(\alpha, \beta) - C_{M,b}/(2\pi n) \right) \rightarrow_d \mathcal{N}(0, 1/2\pi^2).$$

Note that  $\hat{T}_n = T_n(\hat{\alpha}, \hat{\beta})$ . Thus we need the above results to hold with  $\alpha, \beta$  replaced by  $\hat{\alpha}$  and  $\hat{\beta}$ , respectively. Accordingly, write  $\hat{C}_{M,b}$ ,  $\hat{C}_{V,b}$  and  $\hat{\Psi}_n(t)$  for  $C_{M,b}$ ,  $C_{V,b}$  and  $\Psi_n(t)$ , when  $\alpha, \beta$  are replaced by  $\hat{\alpha}, \hat{\beta}$ , respectively. We are now ready to state the following theorem, which provides yet another example where the asymptotic null distributions of these  $L_2$ -distance statistics are not affected by not knowing the nuisance parameters  $\alpha, \beta$ .

**THEOREM 2.1.** *Suppose  $H_0$  holds, assumptions (A), (B) with  $r > 3/2$ , (C) and (D) hold, and that*

$$(2.3) \quad n^{1/2} \{ |\hat{\alpha} - \alpha| + \|\hat{\beta} - \beta\| \} = O_p(1).$$

*In addition, suppose  $b \rightarrow 0$ , and  $nb^{\max\{2\kappa+3, 3.5\}} \rightarrow \infty$ , with  $\kappa$  as in (A). Then  $\hat{C}_{M,b} \approx b^{-(2\kappa+1)}$ ,  $\hat{C}_{V,b} \approx b^{-(4\kappa+1)}$ , and*

$$(2.4) \quad n\hat{C}_{V,b}^{-1/2} \left( \hat{T}_n - \frac{\hat{C}_{M,b}}{2\pi n} \right) \rightarrow_d \mathcal{N}\left(0, \frac{1}{2\pi^2}\right).$$

The proof of this theorem is given in the last section. For an  $0 < a < 1$ , let  $z_a$  be  $(1 - a)$ 100th percentile of the  $\mathcal{N}(0, 1)$  distribution. By (2.4), the test that rejects  $H_0$  whenever

$$\mathcal{T}_n := \sqrt{2\pi n} \hat{C}_{V,b}^{-1/2} \left| \hat{T}_n - \frac{\hat{C}_{M,b}}{2\pi n} \right| > z_{a/2}$$

has the asymptotic size  $a$ .

Examples of  $g$  that satisfy assumption (A) include gamma distributions with scale  $\gamma$  where  $\kappa = \gamma$ , exponential where  $\kappa = 1$ , and Laplace distribution with location 0 and scale 1 where  $\kappa = 2$ . The class of the regression error densities  $f$  that satisfy assumption (B) includes Laplace where  $r = 2$ , normal and Cauchy for any  $r > 0$ .

2.2. *Super smooth case.* Now we consider the problem of obtaining the limiting distribution of  $\hat{T}_n$  in the super smooth case. Here we need the following assumptions.

- (A'):  $\Phi_g$  satisfies  $\Phi_g(t) \neq 0$ , for any  $t \in \mathbb{R}^p$ . For any  $\beta \in \mathbb{R}^p$ ,  $\beta_k \neq 0$ , for  $k = 1, \dots, p$ ,  $|\Phi_g(\beta't)| \sim C(\beta)|t|^{\lambda_0}e^{-\nu(\beta)|t|^\lambda}$ , as  $|t| \rightarrow \infty$ , for a  $\lambda > 1$ ,  $C(\beta) > 0$ ,  $\nu(\beta) > 0$ , and  $\lambda_0 \in \mathbb{R}$ .
- (B'): The density  $f$  is square-integrable, has bounded first and second derivatives, and  $E\varepsilon^2 < \infty$ .
- (C'):  $\Phi_K$  satisfies (C). Moreover  $\Phi_K(0) = 1$ , and there exist  $A > 0$ ,  $\omega \geq 0$  such that  $\Phi_K(1-t) = At^\omega + o(t^\omega)$ , as  $t \rightarrow 0$ .

From Holzmann and Boysen (2006) we can deduce that under the conditions (A')–(C'), and  $b \rightarrow 0$ ,

$$(2.5) \quad \frac{(2\lambda)^{1+2\omega}\pi C^2(\beta)n}{A^2\nu^{1+2\omega}(\beta)b^{\lambda-1+2\lambda\omega+2\lambda_0}\exp(2\nu(\beta)/b^\lambda)\Gamma(2\omega+1)}T_n(\alpha, \beta) \rightarrow_d \chi_2^2/2,$$

where  $\chi_2^2$  is a r.v. having chi-square distribution with two degrees of freedom, and  $\Gamma(\cdot)$  is the Gamma function. In the current situation we obtain the following theorem.

**THEOREM 2.2.** *Suppose  $H_0$  and the assumptions (A'), (B), (B'), (C'), (D) hold,  $b \rightarrow 0$ , and*

$$(2.6) \quad nb^{-\eta}\exp(-2\nu(\beta)/b^\lambda) \rightarrow \infty, \quad \text{for any } \eta > 0.$$

Then

$$(2.7) \quad \mathcal{T}_{n,s} := \frac{(2\lambda)^{1+2\omega}\pi C(\hat{\beta})^2n}{A^2\nu(\hat{\beta})^{1+2\omega}b^{\lambda-1+2\lambda\omega+2\lambda_0}\exp(2\nu(\hat{\beta})/b^\lambda)\Gamma(2\omega+1)}\hat{T}_n \rightarrow_d \chi_2^2/2.$$

Note that the factor multiplying  $\hat{T}_n$  here is all known. Again, the proof of this theorem appears in the last section. The corresponding test is to reject  $H_0$  with asymptotic size  $a$ , for  $0 < a < 1$ , whenever  $\mathcal{T}_{n,s} > \mathcal{X}_a/2$ , where  $\mathcal{X}_a$  is  $(1-a)$ 100th percentile of the  $\chi_2^2$  distribution.

Examples satisfying assumption (A') include normal densities. If  $g$  is a standard normal density then  $C_g = 1$ ,  $\lambda_0 = 0$ ,  $\lambda = 2$  and  $\nu = 2$ . For kernel functions satisfying assumption (C'), Holzmann and Boysen (2006) used the sinc kernel  $K(x) = \sin(x)/(\pi x)$ , with  $A = 1$  and  $\omega = 0$ , and Fan (1992) used  $\Phi_K(t) = (1-t^2)^3$  with  $A = 8$  and  $\omega = 3$ . Other suitable kernel functions can also be found in Delaigle and Hall (2006).

REMARK 2.1. Here we shall discuss tests for fitting a parametric family of densities. Let  $f_\theta, \theta$  an Euclidean parameter, be family of parametric densities and consider the problem of testing the hypothesis  $H : f = f_{\theta_0}$ , for some  $\theta_0$ . Let  $\tilde{\theta}$  be a  $\sqrt{n}$ -consistent estimate of  $\theta_0$  under  $H$  and define

$$\tilde{T}_n = \int_{\mathbb{R}} (\hat{f}_n(x) - K_b * f_{\tilde{\theta}}(x))^2 dx.$$

One can show that if  $f_\theta, \theta \in \Theta$  is quadratic mean differentiable at  $\theta_0$  with the corresponding vector of derivatives  $\dot{f}_\theta$  satisfying  $\int \|\dot{f}_{\theta_0}(x)\|^2 dx < \infty$ , then the asymptotic distributions of a suitably standardized  $\tilde{T}_n$  are the same as in the simple hypothesis case where  $f_0$  is replaced by the  $f_{\theta_0}$ , in both ordinary and super smooth cases. Obtaining the estimator  $\tilde{\theta}$  is often not difficult. For example, consider the problem of testing  $f = N(0, \sigma_\varepsilon^2)$ ,  $\sigma_\varepsilon > 0$ . Here,  $n^{-1} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}' Z_i)^2 - \hat{\beta}' \Sigma_u \hat{\beta}$  is a  $\sqrt{n}$ -consistent estimator of  $\sigma_\varepsilon^2$ , where  $\hat{\alpha}, \hat{\beta}$  are  $\sqrt{n}$ -consistent estimators of  $\alpha, \beta$ , respectively, and  $\Sigma_u$  is the covariance of  $u$ , which is assumed to be known.

REMARK 2.2. In the case of unknown  $g$  one may consider the following two possibilities. First, if replicated observations of  $Z$  at  $X$  are available, then we can construct a nonparametric estimator  $\tilde{g}$  of  $g$ , and use  $\Phi_{\tilde{g}}$ , instead of  $\Phi_g$ , in the proposed test. The large sample properties of the resulting tests remain unchanged. Second possibility is that the density function of  $X$  might be identifiable in some special cases, such as if  $X$  is non-normal, or  $u$  does not have a normal factor, that is,  $u$  is not a sum of a normal r.v. and of an independent non-normal r.v. In such cases, density  $f$  of  $\varepsilon$  is also identifiable, and one may develop a test based on the sieve estimate of  $f$  proposed in Schennach and Hu (2013). These alternatives deserve an independent study, however, in this paper in order to make our testing procedure more broadly applicable, we refrained from imposing any additional distributional assumptions on  $X$ .

It is perhaps worth emphasizing that correct specification of the measurement error distribution is crucial when applying the proposed tests. We can clearly see that the tests are very different for ordinary and super smooth measurement errors.

REMARK 2.3. The selection of a proper bandwidth for any test based on smoothing techniques is challenging. In the literature, two general approaches are often employed to deal with this problem. One is to use an estimator based optimal bandwidth, such as a cross-validation bandwidth, the other is to consider a set of suitable values for the bandwidth and check



how the test results varies with these values. More discussion on this challenging question can be found in Gao and Gijbels (2008). We may follow Gao and Gijbels (2008)'s idea to develop a procedure to select the bandwidth based on the consideration of size and power functions of the tests. However, such development is not easy and it deserves an independent study.

**3. Consistency and Asymptotic Power.** In this section we shall discuss the consistency and asymptotic power for fixed and local nonparametric alternatives of the above tests for both ordinary and super smooth cases.

*Consistency.* Let  $f_1$  be another fixed density of  $\varepsilon$  such that

$$(3.1) \quad \|f_1 - f_0\| := \left( \int [f_1(x) - f_0(x)]^2 dx \right)^{1/2} > 0.$$

Consider the fixed alternatives,  $H_1 : f(x) = f_1(x)$ , for all  $x \in \mathbb{R}$ .

The following two theorems yield the consistency of the above  $\mathcal{T}_n$  and  $\mathcal{T}_{n,s}$  tests against  $H_1$  for the ordinary and super smooth cases, respectively.

**THEOREM 3.1.** *Suppose assumptions (A) and (C) hold,  $f_0$  and  $f_1$  satisfy (B) with  $r > 3/2$ , and have finite fourth moment, and (2.3) holds under  $H_1$ . Furthermore, suppose (D) holds,  $b \rightarrow 0$ , and  $nb^{\max\{2\kappa+3, 3.5\}} \rightarrow \infty$ . Then*

$$(3.2) \quad \sqrt{2\pi n} \hat{C}_{V,b}^{-1/2} |\hat{T}_n - \frac{\hat{C}_{M,b}}{2\pi n}| \rightarrow_p \infty.$$

**THEOREM 3.2.** *Assume (2.3) holds under  $H_1$ , and that the assumptions of Theorem 2.2 hold. Then*

$$\frac{n}{b^{\lambda-1+2\lambda\omega+2\lambda_0} \exp(2\nu(\hat{\beta})/b^\lambda)} \hat{T}_n \rightarrow_p \infty.$$

*Asymptotic local power.* First we consider the ordinary smooth case. We shall describe the asymptotic distribution of  $\hat{T}_n$  under a sequence of the local nonparametric alternatives

$$f_{1n}(x) = f_0(x) + \delta_{1n}\ell(x), \quad x \in \mathbb{R},$$

with  $\delta_{1n} = (C_{V,b}/2)^{1/4}/(n\pi)^{1/2}$ ,  $f_{1n}$  a nonnegative function,  $\ell \in L_2(\mathbb{R})$  such that  $\int \ell(x)dx = 0$ . We obtain

**THEOREM 3.3.** *Suppose the assumptions of Theorem 2.1 hold and that under  $H_{1n} : f(x) = f_{1n}(x)$ , (2.3) holds. Then, under  $H_{1n}$ ,*

$$\sqrt{2\pi n} \hat{C}_{V,b}^{-1/2} (\hat{T}_n - \hat{C}_{M,b}/(2\pi n)) \rightarrow_d \mathcal{N}(\|\ell\|^2, 1).$$

Similarly for the super smooth case, consider a sequence of the local non-parametric alternatives

$$\begin{aligned} f_{2n}(x) &= f_0(x) + \delta_{2n}\ell(x), \quad x \in \mathbb{R}, \\ \delta_{2n} &= \left( \frac{(2\lambda)^{1+2\omega} \pi C(\beta)^2 n}{A^{2\nu}(\beta)^{1+2\omega} b^{\lambda-1+2\lambda\omega+2\lambda_0} \exp(2\nu(\beta)/b^\lambda) \Gamma(2\omega+1)} \right)^{-1/2}, \end{aligned}$$

with  $f_{2n}$  a nonnegative function,  $\ell \in L_2(\mathbb{R})$ , and  $\int \ell(x)dx = 0$ . We obtain

**THEOREM 3.4.** *Suppose the assumptions of Theorem 2.2 hold and (2.3) holds under  $H_{2n} : f(x) = f_{2n}(x)$ . Then, under  $H_{2n}$ ,*

$$\frac{(2\lambda)^{1+2\omega} \pi C(\hat{\beta})^2 n}{A^{2\nu}(\hat{\beta})^{1+2\omega} b^{\lambda-1+2\lambda\omega+2\lambda_0} \exp(2\nu(\hat{\beta})/b^\lambda) \Gamma(2\omega+1)} \hat{T}_n - \|\ell\|^2 \rightarrow_d \chi_2^2/2.$$

The above two theorems show that the proposed tests can detect alternatives which converge to  $f_0$  at a rate slower than  $n^{-1/2}$ .

*Asymptotic power against a fixed alternative.* Now we describe the asymptotic power for the ordinary smooth case against a fixed alternative  $f_1$  satisfying (3.1). For this purpose we need the following result, which follows from Theorem 2 of Holzmann et al. (2007). Suppose assumptions (A) and (C) hold,  $b \rightarrow 0$ ,  $nb^{\max\{4\kappa+2, 2\kappa+3\}} \rightarrow \infty$ ,  $f_0, f_1$  have bounded second derivatives, satisfy assumption (B) for some  $r > \kappa + 1$  and (3.1) holds. Then

$$(3.3) \quad n^{1/2} (T_n(\alpha, \beta) - \|K_b * (f_1 - f_0)\|^2) \rightarrow_d \mathcal{N}(0, \tau_0^2), \quad (\text{under } H_1),$$

where

$$\tau_0^2 = \frac{1}{2\pi^3} \text{Var} \left( \int e^{-it\varepsilon} \frac{\Phi_{f_1}(t) - \Phi_{f_0}(t)}{\Phi_g(\beta t)} dt \right).$$

We shall use this result to analyze the asymptotic distribution of  $\hat{T}_n$  under the fixed alternative  $H_1$ . To proceed further, let  $\mu_Z := EZ$ , and suppose the first derivatives  $\dot{f}_1$  and  $\dot{f}_0$  exist. Define

$$A_f = 2 \int (f_1 - f_0) \dot{f}_0(x) dx, \quad B_f = 2\mu_Z \int (f_1 - f_0) \dot{f}_1(x) dx.$$

**THEOREM 3.5.** *Assume that (A), (C) and (D) hold,  $f_1$  and  $f_0$  satisfy assumption (B) with  $r > \kappa + 1$ ,  $r > 3/2$ , and  $\kappa$  as in (A) and have bounded second derivatives. Also, assume (3.1) and (2.3) hold under  $H_1$ . Furthermore, if  $b \rightarrow 0$ ,  $nb^{\max\{4\kappa+2, 2\kappa+3\}} \rightarrow \infty$ , then*

$$(3.4) \quad n^{1/2} \left( \hat{T}_n - \|K_b * (f_1 - f_0)\|^2 - (\hat{\alpha} - \alpha)A_f - (\hat{\beta} - \beta)'B_f \right) \rightarrow_d \mathcal{N}(0, \tau_0^2).$$

Note that the effect of estimating  $\alpha$  and  $\beta$  introduces another bias term  $n^{1/2}((\hat{\alpha} - \alpha)A_f + (\hat{\beta} - \beta)'B_f)$  in the asymptotic distribution of the statistic  $\hat{T}_n$ . This bias will vanish if to begin with there is no intercept parameter in the model and  $\mu_Z = 0$ . It also vanishes if, under  $H_1$ , the estimators  $\hat{\alpha}$  and  $\hat{\beta}$  satisfy the linearity conditions

$$(3.5) \quad \hat{\alpha} - \alpha = \frac{1}{n} \sum_{j=1}^n \eta_j + o_p(n^{-1/2}),$$

$$(3.6) \quad \hat{\beta}_k - \beta_k = \frac{1}{n} \sum_{j=1}^n \zeta_{jk} + o_p(n^{-1/2}), \quad k = 1, \dots, p,$$

where  $\eta_j$  are i.i.d. with  $E\eta = 0$ ,  $\text{Var}(\eta) > 0$ ,  $E|\eta|^{2+\vartheta} < \infty$ , for some  $\vartheta > 0$ . Moreover, the same conditions are satisfied by  $\zeta_{jk}$ 's, and also for  $1 \leq i \neq j \neq m \leq n$ ,  $\eta_i$ ,  $\zeta_j := (\zeta_{j1}, \dots, \zeta_{jp})'$  and  $e_m$  are mutually independent.

Examples of the estimators of  $\hat{\alpha}, \hat{\beta}$  that satisfy these two conditions include the naive least square estimators, maximum likelihood estimators (see Huáková and Meintanis (2007)), and the bias-corrected estimators (see Fuller (1987)). Using the above expansion, we obtain the following theorem.

**THEOREM 3.6.** *Assume the conditions of Theorem 3.5 and (3.5)-(3.6) for  $\hat{\alpha}$  and  $\hat{\beta}$  hold. Then, for some  $\tau > 0$ ,*

$$(3.7) \quad n^{1/2}(\hat{T}_n - \|K_b * (f_1 - f_0)\|^2) \rightarrow_d \mathcal{N}(0, \tau^2).$$

The form of  $\tau$  is described in the proof of this theorem in the last section, see (6.26). Although  $\tau$  is complicated to calculate in practice, the bootstrap simulation methods can be used to estimate  $\tau$ .

For the super smooth case, in order to obtain a similar result as above, we need to make the following stronger assumptions on  $f_1$  and  $f_0$ :

(B\*) The characteristic function  $\Phi_f$  of the density  $f$  of  $\varepsilon$  satisfies  $|\Phi_f(t)| = O(|t|^{\xi_0} e^{-|t|^\xi/\zeta})$  for some  $\xi_0 \in \mathbb{R}$ ,  $\zeta > 0$  and  $\xi > \lambda$ .

Assumption (B\*) implies (B) and ensures  $\int |\Phi_f(t)/\Phi_g(\beta t)| dt < \infty$ . An example of  $f$  and  $g$  satisfying (B\*) is where  $f$  is a normal density with variance smaller than 1, and  $g$  is standard normal density.

A result analogous to (3.3) can be obtained in the super smooth case also by following the proof of Theorem 2 in Holzmann et al. (2007) with known  $\alpha$  and  $\beta$ . To be clear, assume  $f_1, f_0$  satisfying (3.1), assumptions (A') and (C') hold, and  $f_1$  and  $f_0$  satisfy assumption (B\*). Assume  $b \rightarrow 0$ , and

$$(3.8) \quad nb^{-\eta} \exp(-4\nu(\beta)/b^\lambda) \rightarrow \infty, \quad \text{for any } \eta > 0.$$

Then (3.3) holds. In the case of unknown  $\alpha$  and  $\beta$ , we obtain the following theorem.

**THEOREM 3.7.** *Suppose assumptions (A'), (C'), and (B\*) hold,  $f_1, f_0$  satisfy (3.1), and have bounded second derivatives. If, in addition,  $b \rightarrow 0$ , and (2.6) holds, then we have (3.4).*

*Furthermore, if  $\hat{\alpha}$  and  $\hat{\beta}$  satisfy (3.5)-(3.6), then (3.7) holds.*

**4. Simulations.** In this section we report the findings of some extensive simulations, which assess some finite sample level and power behavior of a member of the above class of tests. Consider the measurement error model

$$(4.1) \quad Y = 1 + X + \varepsilon, \quad Z = X + u.$$

We wish to test the hypothesis that  $\varepsilon \sim \mathcal{N}(0, \theta_0)$ ,  $\theta_0 = \sigma_\varepsilon^2 > 0$ , i.e.,  $f_{\theta_0}$  in  $H$  of Remark 2.1 is the  $\mathcal{N}(0, \sigma_\varepsilon^2)$  density. As in Koul and Song (2012), we used the bias-corrected estimators  $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{Z}$ ,  $\hat{\beta} = S_{ZY}/(S_{ZZ} - \sigma_u^2)$ , and  $\hat{\sigma}_\varepsilon^2 := n^{-1} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}Z_i)^2 - \hat{\beta}^2 \sigma_u^2$ , where  $\bar{Y}$  and  $\bar{Z}$  denote the sample mean of  $Y$  and  $Z$ , and  $S_{ZY}$  and  $S_{ZZ}$  denote the sample covariance of  $Z$  and  $Y$  and the sample variance of  $Z$ , respectively. In the simulations here, we used  $\sigma_\varepsilon = 1$ .

In the ordinary smooth case, we used  $\Phi_g(t) = 1/(1 + \sigma_u^2 t^2/2)$ , i.e., the measurement error density is double exponential with variance  $\sigma_u^2$ . This  $\Phi_g$  satisfies assumption (A) with  $\kappa = 2$ . In the deconvolution estimator of  $f$ , we used the sinc kernel  $K(x) = \sin x/(\pi x)$ , with the bandwidth  $b = \sigma_u n^{-1/8}$ . The proposed test based on  $\tilde{T}_n$  of Remark 2.1 rejects  $H$  for the large values of  $\hat{\mathcal{T}}_n := n\tilde{C}_{V,b}^{-1/2} |\tilde{T}_n - \hat{C}_{M,b}/(2\pi n)|$ .

In the super smooth case,  $u \sim \mathcal{N}(0, \sigma_u^2)$ . The sinc kernel  $K(x) = \sin x/(\pi x)$  is used in the deconvolution kernel estimator, with the bandwidth  $b = (\sigma_u + 0.05)(\log n)^{-0.1}$ . Thus,  $C_g = 1$ ,  $\nu = 2/\sigma_u^2$ ,  $\lambda_0 = 0$ ,  $\lambda = 2$ ,  $A = 1$ , and  $\omega = 0$  in condition (C'). Then the left side of (2.7) becomes

$$\hat{\mathcal{T}}_{n,s} := \frac{2\pi n \sigma_u^2 \hat{\beta}^2 \tilde{T}_n}{b \exp(|\hat{\beta} \sigma_u|^2/b^2)}.$$

Here we discuss the choice of our bandwidths. Carroll et al. (2006, p282) mention that in the super smooth case of normal measurement error as we have here, an optimal bandwidth in the mean squared error sense of a nonparametric density estimator is proportional to  $(\log n)^{-0.5}$ . We also need  $b \rightarrow 0$  and satisfy (2.6). In the normal measurement error case the parameters  $\lambda$  and  $\nu(\beta)$  of (2.6) are 2 and  $\beta^2 \sigma_u^2/2$ , respectively, and (2.6)

becomes  $d_n := nb^{-\eta} \exp(-\beta^2 \sigma_u^2 b^{-2}) \rightarrow \infty$ , for any  $\eta > 0$ . In our simulation, the largest value of  $\sigma_u^2$  is chosen to be 1,  $\beta = 1$ . For these parameter values the choice of  $b = O((\log n)^{-0.5})$  makes  $d_n = O((\log n)^{-5\eta})$ , a sequence that tends to infinity at a very slow rate compared to the case when  $b = O((\log n)^{-a})$ ,  $0 < a < 0.5$  and  $d_n = O(n(\log n)^{a\eta} \exp(-(\log n)^{2a}))$ . For these reason we used the choice of  $a = 0.1$  in our simulations.

Similar considerations were used for selecting bandwidth in the ordinary smooth case, when the measurement error density is double exponential with  $\kappa = 2$ . By Theorem 2.1, the bandwidth  $b$  should satisfy  $b \rightarrow 0$  and  $nb^7 \rightarrow \infty$ . Both requirements are satisfied by choosing  $b$  to be of the order  $n^{-1/8}$ . The dependence of bandwidths on  $\sigma_u$  is also suggested in the literature. See Stefanski (1990) and Carroll et al. (2006) and the references therein.

To investigate the sensitivity of the tests against the choice of bandwidth, we repeated the simulation using  $b = c\sigma_u n^{-1/8}$  in ordinary smooth case, and  $b = c(\sigma_u + 0.05)(\log n)^{-0.1}$  in super smooth case, with  $c = 0.3, 0.5, 0.8, 0.9, 1.2$ . These simulation results are presented in the supplement to this paper. The simulation results presented in this paper are for the case  $c = 1$ . These simulations show that the results for  $c = 1$  are similar to those for  $c = 0.8, 0.9$  and  $c = 1.2$  but different from those for  $c = 0.3, 0.5$ .

We compared the proposed test with the  $W_n$  test of Koul and Song (2012). To define  $W_n$ , let  $\varphi$  be a density on  $\mathbb{R}$ ,  $c \equiv c_n$  be another bandwidth,  $\varphi_2(u) := \int \varphi(v)\varphi(u+v) dv$ ,  $w$  be a compactly supported density on  $\mathbb{R}$ , and let

$$\begin{aligned} \tilde{h}_n(x) &:= \frac{1}{nc} \sum_{j=1}^n \varphi\left(\frac{x - \hat{e}_j}{c}\right), & \hat{C}_n &:= \frac{1}{n^2 c^2} \sum_{i=1}^n \int \varphi^2\left(\frac{v - \hat{e}_i}{c}\right) w(v) dv, \\ \hat{\Gamma}_n &:= 2 \int \tilde{h}_n^2(x) w^2(x) dx \int [\varphi_2(u)]^2 du. \end{aligned}$$

Then, with  $h_0(x, \hat{\beta}) := \int f_0(x + \hat{\beta}u)g(u)du$ ,

$$W_n := nb^{1/2} \hat{\Gamma}_n^{-1/2} \left| \int (\tilde{h}_n(x) - h_0(x, \hat{\beta}))^2 dw(x) - \hat{C}_n \right|.$$

In this simulation study,  $\varphi$  is chosen to be the standard normal density, the bandwidth  $c = n^{-0.27}$ , and  $w$  is chosen to be the uniform density on the closed interval  $[-6, 6]$ .

To assess the effect of the measurement error on the finite sample level and power of these tests, we conducted simulations for the five values of  $\sigma_u^2 = 0.1, 0.25, 0.5, 0.75, 1$ . The alternatives considered were  $t$ -distributions with  $k$  degrees of freedom, denoted by  $t_k$ , for  $k = 3, 5$ , and double exponential (DE) distribution having zero mean and variance  $1/8$ . The sample sizes chosen were 300 and 500 and the level was chosen to be 0.05.

For each scenario, we repeated the tests 200 times, and the empirical power is computed by counting the proportion of times the given test statistics exceeds the asymptotic critical values based on Theorems 2.1 and 2.2.

In all tables,  $N$  stands for  $\mathcal{N}(0, 1)$  distribution. Table 1 reports the simulation results for both ordinary and super smooth cases. In the ordinary smooth case with the DE measurement error, the empirical levels of both tests are often smaller than 0.05 with the  $\hat{\mathcal{T}}_n$  having a slight advantages over the  $W_n$  test. In terms of the empirical power, at the alternatives  $t_3$  and  $t_5$ , the  $\hat{\mathcal{T}}_n$  test dominates  $W_n$  test for the chosen larger values of  $\sigma_u^2$  while the reverse holds for the chosen smaller values of  $\sigma_u^2$ , for both  $n = 300, 500$ .

In the super smooth case, when  $u \sim N(0, 1)$ , both tests are very conservative, but  $\hat{\mathcal{T}}_n$  test dominates  $W_n$  test in terms of the empirical power at the  $t_3$  and  $t_5$  alternatives for almost all chosen values of  $\sigma_u^2$  and for both values of  $n = 300, 500$ .

n	$\sigma_u^2$	Test	Ordinary Smooth				Super Smooth				
			N	$t_3$	$t_5$	DE	N	$t_3$	$t_5$	DE	
300	0.1	$\hat{\mathcal{T}}_n$	0.030	0.390	0.075	0.300	0.000	0.785	0.155	0.000	
		$W_n$	0.015	0.930	0.375	0.030	0.005	0.885	0.240	0.005	
	0.25	$\hat{\mathcal{T}}_n$	0.020	0.510	0.110	0.080	0.000	0.845	0.220	0.000	
		$W_n$	0.025	0.860	0.255	0.080	0.005	0.785	0.160	0.005	
	0.5	$\hat{\mathcal{T}}_n$	0.035	0.605	0.145	0.040	0.000	0.795	0.150	0.000	
		$W_n$	0.015	0.745	0.180	0.125	0.010	0.645	0.025	0.010	
	0.75	$\hat{\mathcal{T}}_n$	0.045	0.635	0.150	0.025	0.000	0.690	0.080	0.000	
		$W_n$	0.030	0.635	0.145	0.095	0.010	0.475	0.050	0.010	
	1	$\hat{\mathcal{T}}_n$	0.050	0.675	0.155	0.030	0.000	0.555	0.060	0.000	
		$W_n$	0.050	0.535	0.115	0.055	0.010	0.365	0.025	0.010	
	500	0.1	$\hat{\mathcal{T}}_n$	0.055	0.660	0.170	0.505	0.025	0.970	0.405	0.000
			$W_n$	0.020	0.995	0.545	0.005	0.010	0.985	0.515	0.015
0.25		$\hat{\mathcal{T}}_n$	0.045	0.795	0.175	0.165	0.020	0.980	0.500	0.015	
		$W_n$	0.025	0.985	0.440	0.060	0.020	0.955	0.365	0.000	
0.5		$\hat{\mathcal{T}}_n$	0.010	0.815	0.235	0.090	0.000	0.970	0.355	0.000	
		$W_n$	0.020	0.940	0.245	0.125	0.020	0.890	0.205	0.000	
0.75		$\hat{\mathcal{T}}_n$	0.025	0.830	0.240	0.080	0.000	0.925	0.180	0.000	
		$W_n$	0.040	0.880	0.165	0.110	0.015	0.755	0.095	0.000	
1		$\hat{\mathcal{T}}_n$	0.050	0.840	0.270	0.065	0.000	0.855	0.075	0.000	
		$W_n$	0.040	0.800	0.130	0.110	0.020	0.650	0.060	0.000	

TABLE 1

Empirical level  $\mathcal{E}$  power for the case of  $p = 1$ .  $H_0 : \varepsilon \sim N \equiv N(0, \sigma_\varepsilon^2)$ ,  $\sigma_\varepsilon = 1$ .  
 Alternatives:  $t_3, t_5$  and DE distributions

Both tests have little empirical power against the DE alternative. A possible explanation for this is as follows. The DE density is different from the normal density because of the sharp peak at the center. Note that the deconvolution kernel density estimate  $\hat{f}_n$  is a smoothing operator. In finite sample cases, this smoothing nature of  $\hat{f}_n$  may “smooth out” this sharp turning point, “twisting” the density in the direction of normal. This, together with the light tail of the DE distribution may result in  $\hat{f}_n$  to be closer to normal density. For illustration, we generated three random samples, each of size  $n = 300$ , from model (4.1), when  $\varepsilon$  has  $t_3, t_5$ , and the DE distribution with

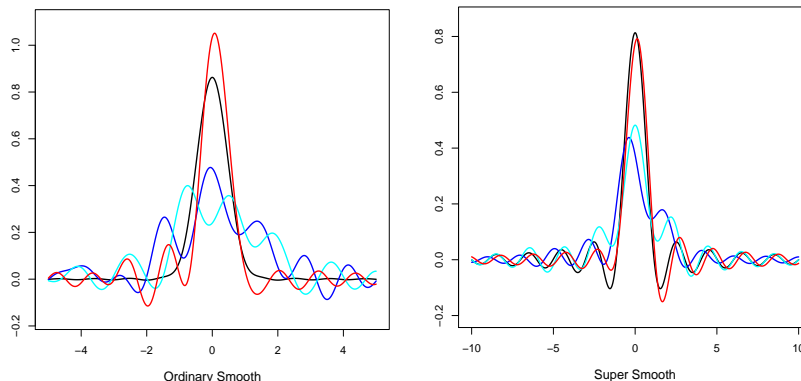


FIG 1. Plots of  $\hat{f}_n(x)$  and  $K * f_0(x)$  when  $\varepsilon \sim t_3, t_5$  and Laplace

mean 0 and variance  $1/8$ , respectively. We used the bandwidth  $0.5\sigma_u n^{-1/8}$  and  $0.5(\sigma_u + 0.05)(\log n)^{-0.1}$  to construct the deconvolution estimates for ordinary smooth and super smooth case, respectively, with  $\sigma_u^2 = 0.5$ . All other simulation setups are the same as before. All the four curves,  $K * f_0$ , and  $\hat{f}_n$  from the three alternatives, are put in the same plots as shown in Figure 1. In both plots, the black curve is  $K * f_0$ , where  $f_0$  is the  $N(0, \hat{\sigma}_\varepsilon^2)$  density; the red curve is  $\hat{f}_n(x)$  when  $\varepsilon \sim DE$  with mean 0 and variance  $1/8$ ; the blue curve is  $\hat{f}_n$  when  $\varepsilon \sim t_3$  and the cyan curve is  $\hat{f}_n$  when  $\varepsilon \sim t_5$ .

Note that in both ordinary and super smooth cases, the red curves, i.e.,  $\hat{f}_n$  when  $\varepsilon \sim DE$ , are much closer to the curve of  $K * f_0$  than the other two curves, implying that the proposed test should have a much lower power in testing the normality of  $\varepsilon$  against the DE alternative. Also note that  $W_n$  test is of a smoothing nature, based on the above observation, its lower power performance is also not surprising.

For the two dimensional  $X$ , the data are generated from the following linear regression model

$$(4.2) \quad Y = 1 + X_1 + 2X_2 + \varepsilon, \quad Z_1 = X_1 + u_1, \quad Z_2 = X_2 + u_2,$$

where  $X_1, X_2$  are independently drawn from  $\mathcal{N}(0, 1)$ . The null hypothesis  $H : \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ , is tested and again the bias-corrected estimators of  $\alpha, \beta_1, \beta_2$  and  $\sigma_\varepsilon^2$  are used. The measurement errors  $u_1$  and  $u_2$  are independent and following the same distributions as in the one dimensional  $X$  case, and other simulation setups are also the same. The simulation results are shown in Table 2. Compared to the one dimensional case, in the ordinary smooth case with DE the measurement error, both tests are liberal as  $\sigma_u^2$  gets larger. In terms of the empirical power, the  $\hat{\mathcal{T}}_n$  test performs better than the  $W_n$  test

for the larger chosen values of  $\sigma_u^2$ , but the latter is much more powerful than the proposed test for the smaller chosen values of  $\sigma_u^2$ .

In the super smooth case, the empirical level performance of the  $\hat{\mathcal{T}}_{n,s}$  test is relatively better than that of the  $W_n$  test for  $n = 300, \sigma_u^2 < 1$ , while for  $n = 500$  both tests are conservative for all choices of  $\sigma_u^2$ . In terms of the empirical power the findings are similar to those in the case of  $p = 1$ .

n	$\sigma_u^2$	Test	Ordinary Smooth				Super Smooth			
			N	$t_3$	$t_5$	DE	N	$t_3$	$t_5$	DE
300	0.1	$\hat{\mathcal{T}}_n$	0.035	0.075	0.030	0.100	0.060	0.095	0.020	0.025
		$W_n$	0.020	0.405	0.080	0.185	0.015	0.980	0.345	0.005
	0.25	$\hat{\mathcal{T}}_n$	0.055	0.100	0.060	0.080	0.050	0.075	0.045	0.030
		$W_n$	0.060	0.150	0.050	0.065	0.010	0.825	0.070	0.015
	0.5	$\hat{\mathcal{T}}_n$	0.080	0.090	0.450	0.075	0.050	0.075	0.045	0.025
		$W_n$	0.075	0.100	0.065	0.065	0.025	0.385	0.030	0.010
	0.75	$\hat{\mathcal{T}}_n$	0.095	0.105	0.065	0.075	0.055	0.190	0.055	0.030
		$W_n$	0.055	0.055	0.075	0.055	0.010	0.130	0.025	0.010
1	$\hat{\mathcal{T}}_n$	0.090	0.095	0.070	0.065	0.035	0.145	0.035	0.020	
	$W_n$	0.075	0.060	0.055	0.070	0.025	0.110	0.020	0.010	
500	0.1	$\hat{\mathcal{T}}_n$	0.015	0.075	0.050	0.055	0.045	0.120	0.085	0.040
		$W_n$	0.040	0.655	0.180	0.400	0.005	0.870	0.215	0.005
	0.25	$\hat{\mathcal{T}}_n$	0.055	0.080	0.080	0.075	0.025	0.120	0.075	0.025
		$W_n$	0.045	0.345	0.080	0.095	0.000	0.520	0.070	0.005
	0.5	$\hat{\mathcal{T}}_n$	0.070	0.100	0.090	0.085	0.025	0.110	0.070	0.045
		$W_n$	0.090	0.185	0.045	0.055	0.005	0.230	0.040	0.015
	0.75	$\hat{\mathcal{T}}_n$	0.075	0.115	0.075	0.085	0.035	0.165	0.060	0.050
		$W_n$	0.100	0.145	0.050	0.060	0.015	0.135	0.040	0.010
	1	$\hat{\mathcal{T}}_n$	0.055	0.145	0.075	0.080	0.040	0.180	0.055	0.040
		$W_n$	0.085	0.100	0.065	0.070	0.010	0.065	0.015	0.010

TABLE 2

*Empirical level & power for the case of  $p = 2$ .  $H_0 : \varepsilon \sim N \equiv N(0, \sigma_\varepsilon^2)$ ,  $\sigma_\varepsilon = 1$ .  
Alternatives:  $t_3, t_5$  and DE distributions*

We also ran an additional simulation comparing the four tests based on  $\hat{\mathcal{T}}_n$ ,  $W_n$ , Kolmogorov-Smirnov and the Cramér-von Mises statistics. Because the asymptotic null distributions of the last two tests are unknown, we used the following parametric bootstrap methodology (see Shao and Tu (1995)) to make this comparison.

Step 1. Use the data set to obtain the bias-corrected estimators  $\hat{\alpha}$  and  $\hat{\beta}$  for  $\alpha$  and  $\beta$ , and  $\hat{\theta}$  for  $\theta$  appearing in the composite null hypothesis  $K_0$  of Remark 2.1.

Step 2. Draw independent sample  $\tilde{e}_j, j = 1, \dots, n$ , with  $\tilde{e}_j \sim h(x; \hat{\alpha}, \hat{\beta}, \hat{\theta}) = \int f_{\hat{\theta}}(x + \hat{\beta}u)g(u)du$ , write  $\tilde{F}_n(x)$  as the empirical distribution function of  $\tilde{e}_j, j = 1, \dots, n$ . Let  $F_{\hat{\theta}, \hat{\alpha}, \hat{\beta}}(x)$  denote the d.f.  $h(x; \hat{\alpha}, \hat{\beta}, \hat{\theta})$ . Calculate the bootstrap values

$$\tilde{\mathcal{T}}_n^* = n\hat{C}_{V,b}^{-1/2}|\tilde{T}_n - \hat{C}_{M,b}/(2\pi n)|, \text{ ordinary smooth}$$



$$\begin{aligned}
\text{or } \tilde{T}_{n,s}^* &:= \frac{2\pi n \sigma_u^2 \hat{\beta}^2 \tilde{T}_n}{b \exp(|\hat{\beta} \sigma_u|^2 / b^2)}, \text{ super smooth} \\
\tilde{T}_{KS} &:= \sup_{x \in \mathbb{R}} n^{1/2} |\tilde{F}_n(x) - F_{\hat{\theta}, \hat{\alpha}, \hat{\beta}}(x)|, \\
\tilde{T}_{CvM} &:= n \int (\tilde{F}_n(x) - F_{\hat{\theta}, \hat{\alpha}, \hat{\beta}}(x))^2 d\tilde{F}_n(x), \\
\tilde{W}_n &:= nb^{1/2} \tilde{\Gamma}_n^{-1} \left| \int (\tilde{h}_n(x) - h(x; \hat{\alpha}, \hat{\beta}, \hat{\theta}))^2 dx - \hat{C}_n \right|.
\end{aligned}$$

where  $\tilde{T}_n$ ,  $\tilde{\Gamma}_n^{-1/2}$  and  $\tilde{h}_n(x)$  denote the analogs of  $\hat{T}_n$ ,  $\hat{\Gamma}_n^{-1}$  and  $\hat{h}_n(x)$ , respectively, where  $\hat{e}_j$  are replaced by  $\tilde{e}_j$ .

Step 3. For each test, repeat step 2 for  $R$  times to obtain  $R$  bootstrap values of the test statistic, define the 5% bootstrap critical value as the 95-th percentile of these  $R$  values.

Step 4. For each test, calculate the given test statistic based on the original data and reject the null hypothesis if it exceeds the corresponding critical value obtained in step 3.

For each test, we repeat steps 1 to 4 for 200 times, then the empirical power is computed by counting the proportion of the times the given test statistic exceeds the bootstrap critical value. The findings of this simulation for the model (4.1) along with some discussion are reported in the supplement to this paper. From this simulation we note that in the ordinary smooth case, for the chosen smaller values of  $\sigma_u^2$ , in terms of the empirical power, the  $W_n$  test dominates  $\hat{T}_n$  test while  $\hat{T}_n$  dominates the KS and CvM tests. For the larger chosen value of  $\sigma_u^2$ , the superiority of the  $\hat{T}_n$  test becomes clear. In the super smooth case, the  $\tilde{T}_{n,s}$  test appears to be inferior to the  $W_n$  test for the chosen small values of  $\sigma_u^2$ , but its empirical power is larger than that of the KS and CvM tests in all scenarios. Compared to  $W_n$  and the proposed tests, the KS and CvM tests are more conservative in both ordinary and super smooth cases.

Following the AE's suggestion, we put the simulation results from both asymptotic-based and bootstrap-based critical values together in Tables 3 and 4, for ordinary and super smooth cases, respectively. From these tables, we see that in the ordinary smooth case the empirical levels of both tests based on bootstrap distributions are much closer to the nominal level of 0.05, compared to those based on the asymptotic null distributions, for the chosen values of  $\sigma_u^2 \geq 0.25$ , with  $\hat{T}_n$  test having a slight advantage over  $W_n$  test. The behavior of the empirical powers of both tests is similar under both asymptotic and bootstrap methodologies.

In the super smooth case neither asymptotic nor bootstrap levels of both

n	$\sigma_u^2$	Test	Asymptotic				Bootstrap				
			N	$t_3$	$t_5$	DE	N	$t_3$	$t_5$	DE	
300	0.1	$\tilde{T}_n$	0.030	0.390	0.075	0.300	0.025	0.495	0.120	0.370	
		$W_n$	0.015	0.930	0.375	0.030	0.085	0.740	0.215	0.095	
	0.25	$\tilde{T}_n$	0.020	0.510	0.110	0.080	0.045	0.615	0.110	0.140	
		$W_n$	0.025	0.860	0.255	0.080	0.055	0.845	0.180	0.060	
	0.5	$\tilde{T}_n$	0.035	0.605	0.145	0.040	0.050	0.655	0.150	0.115	
		$W_n$	0.015	0.745	0.180	0.125	0.065	0.730	0.120	0.060	
	0.75	$\tilde{T}_n$	0.045	0.635	0.150	0.025	0.055	0.650	0.185	0.115	
		$W_n$	0.030	0.635	0.145	0.095	0.055	0.645	0.100	0.095	
	1	$\tilde{T}_n$	0.050	0.675	0.155	0.030	0.055	0.635	0.185	0.105	
		$W_n$	0.050	0.535	0.115	0.055	0.065	0.545	0.095	0.075	
	500	0.1	$\tilde{T}_n$	0.055	0.660	0.170	0.505	0.060	0.705	0.120	0.495
			$W_n$	0.020	0.995	0.545	0.005	0.075	0.910	0.510	0.055
0.25		$\tilde{T}_n$	0.045	0.795	0.175	0.165	0.045	0.815	0.190	0.175	
		$W_n$	0.025	0.985	0.440	0.060	0.090	0.970	0.400	0.025	
0.5		$\tilde{T}_n$	0.010	0.815	0.235	0.090	0.040	0.825	0.310	0.145	
		$W_n$	0.020	0.940	0.245	0.125	0.075	0.935	0.280	0.070	
0.75		$\tilde{T}_n$	0.025	0.830	0.240	0.080	0.050	0.880	0.300	0.120	
		$W_n$	0.040	0.880	0.165	0.110	0.065	0.840	0.200	0.075	
1		$\tilde{T}_n$	0.050	0.840	0.270	0.065	0.075	0.910	0.310	0.155	
		$W_n$	0.040	0.800	0.130	0.110	0.065	0.820	0.165	0.085	

TABLE 3

**Ordinary smooth case.** Empirical level  $\mathcal{E}$  power using asymptotic (left panel) and bootstrap (right panel) critical values, for the case of  $p = 1$ .  $H_0 : \varepsilon \sim N(0, \sigma_\varepsilon^2)$ ,  $\sigma_\varepsilon = 1$ . Alternatives:  $t_3, t_5$  and DE distributions

n	$\sigma_u^2$	Test	Asymptotic				Bootstrap				
			N	$t_3$	$t_5$	DE	N	$t_3$	$t_5$	DE	
300	0.1	$\tilde{T}_n$	0.000	0.785	0.155	0.000	0.015	0.850	0.325	0.080	
		$W_n$	0.005	0.885	0.240	0.005	0.065	0.905	0.375	0.060	
	0.25	$\tilde{T}_n$	0.000	0.845	0.220	0.000	0.015	0.820	0.325	0.000	
		$W_n$	0.005	0.785	0.160	0.005	0.090	0.825	0.225	0.120	
	0.5	$\tilde{T}_n$	0.000	0.795	0.150	0.000	0.005	0.790	0.285	0.000	
		$W_n$	0.010	0.645	0.025	0.010	0.060	0.665	0.150	0.090	
	0.75	$\tilde{T}_n$	0.000	0.690	0.080	0.000	0.000	0.710	0.195	0.000	
		$W_n$	0.010	0.475	0.050	0.010	0.080	0.515	0.090	0.055	
	1	$\tilde{T}_n$	0.000	0.555	0.060	0.000	0.005	0.650	0.110	0.000	
		$W_n$	0.010	0.365	0.025	0.010	0.060	0.435	0.080	0.085	
	500	0.1	$\tilde{T}_n$	0.025	0.970	0.405	0.000	0.030	0.990	0.495	0.215
			$W_n$	0.010	0.985	0.515	0.015	0.075	1.000	0.610	0.155
0.25		$\tilde{T}_n$	0.020	0.980	0.500	0.015	0.035	0.985	0.545	0.010	
		$W_n$	0.020	0.955	0.365	0.000	0.065	0.980	0.440	0.055	
0.5		$\tilde{T}_n$	0.000	0.970	0.355	0.000	0.015	0.970	0.495	0.000	
		$W_n$	0.020	0.890	0.205	0.000	0.060	0.905	0.225	0.085	
0.75		$\tilde{T}_n$	0.000	0.925	0.180	0.000	0.015	0.940	0.380	0.000	
		$W_n$	0.015	0.755	0.095	0.000	0.075	0.760	0.115	0.080	
1		$\tilde{T}_n$	0.000	0.855	0.075	0.000	0.005	0.875	0.245	0.000	
		$W_n$	0.020	0.650	0.060	0.000	0.090	0.620	0.050	0.045	

TABLE 4

**Super smooth case:** Empirical level  $\mathcal{E}$  power using asymptotic (left panel) and bootstrap (right panel) critical values, for the case of  $p = 1$ .  $H_0 : \varepsilon \sim N(0, \sigma_\varepsilon^2)$ ,  $\sigma_\varepsilon = 1$ . Alternatives:  $t_3, t_5$  and DE distributions

tests are anywhere close to 0.05 for any chosen values of  $\sigma_u^2$ , although the  $W_n$  test seems to have slight advantage. Based on the bootstrap distributions, the  $\hat{T}_{n,s}$  test appears to be conservative while the  $W_n$  test to be liberal for all the chosen values of  $\sigma_u^2$  and  $n$ .

Some possible reasons for the above type of empirical findings are as follows. It is well known that the tests based on kernel density estimators require much larger sample sizes for the validity of the asymptotic level because of their slow consistency rate. This rate is much worse in the presence of measurement error. In the super smooth case this rate for the deconvolution kernel estimate (DCE) is notoriously slow and in addition the requirement (2.6) on the bandwidth makes it even slower. The consistency rate of a DCE is determined by its bias and variance. Its bias is the same as of the classical kernel estimate, however, its variance heavily depends on the smoothness of the measurement error. This variance is relatively much larger in the super smooth case than that in ordinary smooth case. Thus, under the null hypothesis, for fixed sample size, the proposed test has much larger variance in the super smooth case than in the ordinary smooth case and the bootstrap critical values in the super smooth case tend to be relatively larger, which leads to conservative tests. In comparison, in the ordinary smooth case, the consistency rate is relatively faster and the bootstrap empirical levels of the proposed and the  $W_n$  tests tend to be closer to the nominal level.

We also compared the relative magnitude of the noncentral parameters from  $\hat{T}_n$  and  $W_n$  tests for some particular hypotheses. These numerical examples indicate that  $\hat{T}_n$  test generally has a larger noncentral parameter than the  $W_n$  test. This indicates that the  $\hat{T}_n$  test has a larger relative efficiency than the  $W_n$  test based on the relative efficiency concept defined in Hong and Lee (2013). However, an independent study on the relative asymptotic efficiency of these two nonparametric tests is needed before formal results can be presented.

**5. A Real Data Example.** The pig data was collected by the Statistical Laboratory of Iowa State University under contract to the Statistical Reporting Service, U.S. Department of Agriculture. The study is originally described in Battese, Fuller and Hickman (1976), and re-investigated by Fuller (1987) under the framework of linear regression model with measurement error. The relationship between the two variables, the number of breeding hogs on hand  $X$  and the number of sows farrowing (giving birth to baby pigs)  $Y$  is modeled by a linear regression model. Since the data are collected by interviewing the farmers, so it is believed both variables are measured with errors. The number of farmers interviewed in this study is

184 and the same group of farmers were interviewed twice with one month gap in between.

In this application, the average of  $Y$ -values from the two interviews from the same farmers is treated as the response variable, and its measurement error is incorporated into the regression error term, the average of the  $X$ -values is treated as the surrogate of “true” number of sows farrowing. We assume the measurement error is normal. From Fuller (1987), the variance of the measurement error is estimated as 69.54 based on the analysis of covariance. We would like to check if the regression error is normal. That is, the null hypothesis we would like to test is  $H_0 : f = N(0, \sigma_\varepsilon^2)$ , for some  $\sigma_\varepsilon^2$ .

The test statistic  $\mathcal{T}_{n,s}$  in Theorem 2.2 gives a value of 0.2073, and it is far less than the critical value 2.9957 designated by the half of the  $\chi^2$ -distribution with degrees of freedom 2. To evaluate the performance of test based on large sample theory, we also apply the proposed goodness of fit test with critical value determined by the bootstrap algorithm. To be specific, the following bootstrap algorithm is used to determine the critical value at the significance level 0.05.

Step 1. Use the date set to obtain the bias-corrected estimators  $\hat{\alpha}$  and  $\hat{\beta}$  for  $\alpha$  and  $\beta$ .

Step 2. Draw independent sample  $\tilde{e}_j$ ,  $j = 1, \dots, n$ , with  $\tilde{e}_j \sim N(0, \hat{\sigma}_\varepsilon^2 + \hat{\beta}^2 \sigma_u^2)$ , and calculate

$$\begin{aligned} \tilde{\mathcal{T}}_{n,s}^* &:= 2\pi n \sigma_u^2 \hat{\beta}^2 \tilde{\mathcal{T}}_n / b \exp(|\hat{\beta} \sigma_u|^2 / b^2), \\ \tilde{\mathcal{T}}_n &= \int_{\mathbb{R}} \left( \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi_K(bt) \frac{n^{-1} \sum_{j=1}^n e^{it\tilde{e}_j}}{\Phi_g(-\beta t)} dt - K_b * f_{\hat{\sigma}_\varepsilon^2}(x) \right)^2 dx. \end{aligned}$$

Step 3. Repeat step 2 for 400 times and denotes the  $\tilde{\mathcal{T}}_{n,s}^*$  as  $\tilde{\mathcal{T}}_1^*, \dots, \tilde{\mathcal{T}}_{400}^*$ . Sort these values in an increasing order, and denote the 95% percentile as  $t_\alpha^*$ .

Step 4. Calculate the  $\tilde{\mathcal{T}}_{n,s}$  with the original data set and reject the null hypothesis if  $\tilde{\mathcal{T}}_{n,s} > t_{0.05}^*$ , otherwise, accept it.

The 0.05 bootstrap critical value is found to be around 0.4590, which is larger than the value of the test statistic 0.2073. Although the result is consistent with the test without using bootstrap, the critical value based on the large sample theory is much larger than the one obtained from bootstrap algorithm. Therefore, the proposed test fails to reject the null hypothesis. We also tried the bootstrap algorithm for the Kolmogorov-Smirnov and Cramér von-Mise tests, and arrived at the same conclusion.

**6. Proofs.** Here we present the proof of Theorems 2.1–3.7. Write  $T_n := T_n(\alpha, \beta)$  and  $f_n(x) := f_n(x, \alpha, \beta)$  and  $\hat{f}_n(x) := f_n(x, \hat{\alpha}, \hat{\beta})$ , for convenience. Let  $\dot{q}$  denote the first derivative of  $q$  for any differentiable function  $q$ .

Since  $C_{V,b} \approx b^{-(4\kappa+1)}$ , we first show

$$(6.1) \quad nb^{2\kappa} \int (\hat{f}_n - f_n)^2(x) dx = o_p(1).$$

Using Parseval's equation, the integral  $\int (\hat{f}_n - f_n)^2(x) dx$  equals to

$$(6.2) \quad \begin{aligned} & \frac{1}{4\pi^2} \int \left( \int e^{-itx} \Phi_K(ht) \left( \frac{\hat{\Psi}_n(t)}{\Phi_g(-\hat{\beta}t)} - \frac{\Psi_n(t)}{\Phi_g(-\beta t)} \right) dt \right)^2 dx \\ &= \frac{1}{2\pi} \int |\Phi_K(ht)|^2 \left| \frac{\hat{\Psi}_n(t)}{\Phi_g(-\hat{\beta}t)} - \frac{\Psi_n(t)}{\Phi_g(-\beta t)} \right|^2 dt \\ &\leq \frac{1}{2\pi} \int |\Phi_K(ht)|^2 \frac{|\hat{\Psi}_n(t) - \Psi_n(t)|^2}{|\Phi_g(-\hat{\beta}t)|^2} dt \\ &\quad + \frac{1}{2\pi} \int |\Phi_K(bt)\Psi_n(t)|^2 \frac{|\Phi_g(-\hat{\beta}t) - \Phi_g(-\beta t)|^2}{|\Phi_g(-\hat{\beta}t)\Phi_g(-\beta t)|^2} dt \\ &= \frac{1}{2\pi} S_1 + \frac{1}{2\pi} S_2, \quad \text{say.} \end{aligned}$$

Since  $\Phi_K$  is supported on  $[-1, 1]$ ,  $\Phi_K(bt) = 0$ , for  $|t| > 1/b$ . Thus in the above two integrals,  $t \in [-1/b, 1/b]$ . Since  $\mu_g := \int |x|g(x)dx < \infty$ ,  $\dot{\Phi}_g$  exists and is uniformly bounded above by  $\mu_g$ . This fact together with (2.3) and assumption (A) imply,

$$(6.3) \quad |\Phi_g(-\hat{\beta}t) - \Phi_g(-\beta t)| \leq \mu_g |t| \|\hat{\beta} - \beta\|,$$

$$(6.4) \quad \begin{aligned} \max_{|t| \leq 1/b} \left| \frac{\Phi_g(-\hat{\beta}t)}{\Phi_g(-\beta t)} - 1 \right| &= \max_{|t| \leq 1/b} \left| \frac{\Phi_g(-\hat{\beta}t) - \Phi_g(-\beta t)}{\Phi_g(-\beta t)} \right| \\ &= O_p(n^{-1/2}b^{-\kappa-1}). \end{aligned}$$

Let  $A_n := \{|\Phi_g(-\hat{\beta}t)| \geq |\Phi_g(-\beta t)|/2, t \in [-1/b, 1/b]\}$ . Since  $nb^{2\kappa+3} \rightarrow \infty$ , (6.4) implies  $P(A_n) \rightarrow 1$ . Thus we need only to restrict our attention to  $A_n$ .

Consider  $S_2$ . Conditions (A) and (B) imply that there exists a  $M$ ,  $c_\beta$ ,  $C_\beta$  and  $C_f$ , such that for all  $|t| > M$ ,  $c_\beta|t|^{-\kappa} \leq |\Phi_g(\beta t)| \leq C_\beta|t|^{-\kappa}$  and  $\Phi_f(t) \leq C_f|t|^{-r}$ . Take  $n$  large enough so that  $M < 1/b$ . Split the integral in  $S_2$  into two ranges, one with  $|t| \leq M$  and the other with  $|t| > M$ . Then by (2.3) and (6.3), on the event  $A_n$ ,  $S_2$  is bounded from the above by

$$4\mu_g^2 \|\hat{\beta} - \beta\|^2 \int_{1/b \geq |t| > M} \frac{|t\Phi_K(bt)\Psi_n(t)|^2}{|\Phi_g(-\beta t)|^4} dt + O_p(n^{-1})$$

$$\leq 8\mu_g^2 \|\hat{\beta} - \beta\|^2 \left\{ \int_{1/b \geq |t| > M} \left[ \frac{|t\Phi_K(bt)|^2 |\Psi_n(t) - \Phi_h(t)|^2}{|\Phi_g(-\beta t)|^4} + \frac{|t\Phi_K(bt)\Phi_h(t)|^2}{|\Phi_g(-\beta t)|^4} \right] dt \right\} + O_p(n^{-1}).$$

By the Parseval's identity

$$(6.5) \quad T_n(\alpha, \beta) = \frac{1}{2\pi} \int \frac{|\Phi_K(bt)|^2 |\Psi_n(t) - \Phi_h(t)|^2}{|\Phi_g(-\beta t)|^2} dt = O_p(n^{-1}b^{-2\kappa-1}),$$

because of (2.1) and (2.2). Because  $|\Phi_g(\beta t)|^{-2} \leq c_\beta^2 |t|^{2\kappa}$ , the first term within the curly brackets in the above bound is bounded above by  $b^{-2\kappa-2}T_n = O_p(n^{-1}b^{-4\kappa-3})$ .

Similarly, assumptions (A) and (B) imply

$$\int_{|t| > M} \frac{|t\Phi_K(bt)\Phi_h(t)|^2}{|\Phi_g(-\beta t)|^4} dt = \int_{|t| > M} \frac{|t\Phi_K(bt)\Phi_f(t)|^2}{|\Phi_g(-\beta t)|^2} dt = O(b^{\min(2r-2\kappa-3, 0)}).$$

Hence, in view of (2.3),

$$(6.6) \quad S_2 = O_p(n^{-2}b^{-4\kappa-3}) + O_p(n^{-1}b^{\min(2r-2\kappa-3, 0)}) = o_p(n^{-1}b^{-2\kappa}).$$

Next, to analyze  $S_1$ . Let

$$\begin{aligned} S_{11} &:= \frac{1}{n^2} \int \frac{|\Phi_K(bt)|^2 \left| \sum_{j=1}^n t(\hat{\beta} - \beta)' Z_j e^{it(Y_j - \alpha - \beta' Z_j)} \right|^2}{|\Phi_g(-\beta t)|^2} dt, \\ S_{12} &:= \frac{1}{n^2} \int \frac{|\Phi_K(bt)|^2 \left| \sum_{j=1}^n t e^{it(Y_j - \alpha - \beta' Z_j)} \right|^2}{|\Phi_g(-\beta t)|^2} dt, \\ S_{13} &:= \frac{1}{n^2} \int \frac{|\Phi_K(bt)|^2 \left| \sum_{j=1}^n t((\hat{\beta} - \beta)' Z_j)^2 e^{it(Y_j - \alpha - \beta' Z_j)} \right|^2}{|\Phi_g(-\beta t)|^2} dt. \end{aligned}$$

Using the fact  $Y_j - \alpha - \beta' Z_j = \varepsilon_j - \beta' u_j$ , we obtain on the event  $A_n$ ,

$$(6.7) \quad \begin{aligned} S_1 &\leq 4 \int |\Phi_K(bt)|^2 \frac{|\hat{\Psi}_n(t) - \Psi_n(t)|^2}{|\Phi_g(-\beta t)|^2} dt \\ &\leq \frac{16}{n^2} \int \frac{|\Phi_K(bt)|^2 \left| \sum_{j=1}^n t(\hat{\beta} - \beta)' Z_j e^{it(\varepsilon_j - \beta' u_j)} \right|^2}{|\Phi_g(-\beta t)|^2} dt \\ &\quad + \frac{16(\hat{\alpha} - \alpha)^2}{n^2} \int \frac{|\Phi_K(bt)|^2 \left| \sum_{j=1}^n t e^{it(\varepsilon_j - \beta' u_j)} \right|^2}{|\Phi_g(-\beta t)|^2} dt \\ &\quad + \frac{16}{n^2 b^2} \int \frac{|\Phi_K(bt)|^2 \left| \sum_{j=1}^n t((\hat{\beta} - \beta)' Z_j)^2 e^{it(\varepsilon_j - \beta' u_j)} \right|^2}{|\Phi_g(-\beta t)|^2} dt \end{aligned}$$

$$\begin{aligned}
& + \frac{16(\hat{\alpha} - \alpha)^4}{n^2 b^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t e^{it(\varepsilon_j - \beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2} dt + O_p(n^{-3} b^{-2\kappa - 7}) \\
& = 16[S_{11} + (\hat{\alpha} - \alpha)^2 S_{12} + b^{-2} S_{13} + (\hat{\alpha} - \alpha)^4 b^{-2} S_{12}] + O_p(n^{-3} b^{-2\kappa - 7}),
\end{aligned}$$

by (2.3), assumption (A), and the fact that  $\sum_{j=1}^n |Z_j|^3 = O_p(n)$ .

Now, consider  $S_{11}$ .

$$S_{11} \leq p \sum_{k=1}^p \frac{(\hat{\beta}_k - \beta_k)^2}{n^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t Z_{kj} e^{it(\varepsilon_j - \beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2} dt.$$

Since  $X$ ,  $u$  and  $\varepsilon$  are mutually independent, for any  $k = 1, \dots, p$ ,

$$\mathbb{E} Z_k e^{it(\varepsilon_j - \beta' u_j)} = \mathbb{E} X_k \Phi_h(t) + \Phi_f(t) \mathbb{E} u_k e^{-it\beta' u}.$$

We use this to obtain

$$\begin{aligned}
(6.8) \quad & \frac{1}{n^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t Z_{kj} e^{it(\varepsilon_j - \beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2} dt \\
& \leq \frac{3}{n^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n [Z_{kj} e^{it(\varepsilon_j - \beta' u_j)} - \mathbb{E} Z_{kj} e^{it(\varepsilon_j - \beta' u_j)}]|^2}{b^2 |\Phi_g(-\beta t)|^2} dt \\
& \quad + \frac{3}{n^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n \mathbb{E} X_k t \Phi_h(t)|^2}{|\Phi_g(-\beta t)|^2} dt \\
& \quad + \frac{3}{n^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t \Phi_f(t) \mathbb{E} u_k e^{-it\beta' u}|}{|\Phi_g(-\beta t)|^2} dt.
\end{aligned}$$

An argument similar to the one used in the proof of Theorem 1 in Holzmann et al. (2007) implies that the first summand in the upper bound of (6.8) is  $O_p(n^{-1} b^{-2\kappa - 3})$ . The second summand is  $O_p(1)$ , by assumption (B), and  $\Phi_h(t)/\Phi_g(-\beta t) = \Phi_f(t)$ . To analyze the third summand, decompose the integral into two ranges,  $|t| > M$  and  $|t| \leq M$ , and use the conditions (A)-(B) to show that the term with integration over  $|t| \leq M$  is  $O_p(1)$ , while the term with  $|t| > M$  is of the order  $O_p(b^{\min(2r - 2\kappa - 3, 0)})$ , thereby showing that the third summand in (6.8) is of the order  $O_p(1) + O_p(b^{2r - 2\kappa - 3})$ . Thus

$$(6.9) \quad S_{11} = O_p(n^{-1} b^{-2\kappa - 3}) + O_p(1) + O_p(b^{2r - 2\kappa - 3}).$$

Similarly one obtains that  $S_{12}$  and  $S_{13}$  are of the same order as  $S_{11}$ . Then (6.7), (6.9),  $nb^{2\kappa + 3} \rightarrow \infty$ ,  $nb^{7/2} \rightarrow \infty$  imply

$$nb^{2\kappa} S_1 = O_p(n^{-1} b^{-3}) + O_p(b^{2\kappa}) + O_p(b^{2r - 3}) + O_p(n^{-2} b^{-7}) = o_p(1).$$

This together with (6.6) completes the proof of (6.1).

From (6.1) and (6.5) we obtain

$$\begin{aligned}\hat{T}_n - T_n &= \int (f_n - \hat{f}_n)^2(x) dx + 2 \int (f_n - \hat{f}_n)(f_n - K_b * f_0)(x) dx \\ &= o_p(n^{-1}b^{-2\kappa-1/2}),\end{aligned}$$

by (2.1) and (2.2). Hence, in view of (2.2),

$$(6.10) \quad n/C_{V,b}^{1/2}(\hat{T}_n - C_{M,b}/(2\pi n)) \rightarrow_d \mathcal{N}(0, 1/2\pi^2).$$

To complete the proof of (2.4), it suffices to show that

$$(6.11) \quad (a) \quad \left|1 - \frac{\hat{C}_{V,b}^{1/2}}{C_{V,b}^{1/2}}\right| = o_p(b^{1/2}), \quad (b) \quad \left|\frac{\hat{C}_{M,b}}{\hat{C}_{V,b}^{1/2}} - \frac{C_{M,b}}{C_{V,b}^{1/2}}\right| = o_p(1).$$

To show (6.11)(a), recall  $\psi(\beta, s, t) := \Phi_g(\beta t + \beta s)\Phi_f(t + s)$ . Then

$$\begin{aligned}& |C_{V,b} - \hat{C}_{V,b}| \\ &= \left| \int \int \frac{|\Phi_K(tb)|^2 |\Phi_K(sb)|^2}{|\Phi_g(\beta t)|^2 |\Phi_g(\beta s)|^2} |\psi(\beta, s, t)|^2 ds dt \right. \\ &\quad \left. - \int \int \frac{|\Phi_K(\hat{b}t)|^2 |\Phi_K(\hat{b}s)|^2}{|\Phi_g(\hat{\beta}t)|^2 |\Phi_g(\hat{\beta}s)|^2} |\psi(\hat{\beta}, s, t)|^2 ds dt \right| \\ &\leq \int \int \frac{|\Phi_K(tb)|^2 |\Phi_K(sb)|^2 \left| |\Phi_g(\beta t)|^2 - |\Phi_g(\hat{\beta}t)|^2 \right|}{|\Phi_g(\beta t)|^2 |\Phi_g(\beta s)|^2 |\Phi_g(\hat{\beta}t)|^2} |\psi(\beta, s, t)|^2 ds dt \\ &\quad + \int \int \frac{|\Phi_K(tb)|^2 |\Phi_K(sb)|^2 \left| |\Phi_g(\beta s)|^2 - |\Phi_g(\hat{\beta}s)|^2 \right|}{|\Phi_g(\hat{\beta}t)|^2 |\Phi_g(\beta s)|^2 |\Phi_g(\hat{\beta}s)|^2} |\psi(\beta, s, t)|^2 ds dt \\ &\quad + \int \int \frac{|\Phi_K(tb)|^2 |\Phi_K(sb)|^2}{|\Phi_g(\hat{\beta}t)|^2 |\Phi_g(\hat{\beta}s)|^2} \left| |\psi(\beta, s, t)|^2 - |\psi(\hat{\beta}, s, t)|^2 \right| ds dt.\end{aligned}$$

In view of (6.4), the first term in the above bound is bounded from the above by

$$\begin{aligned}& \max_{|t| \leq 1/b} \left|1 - \frac{|\Phi_g(\beta t)|^2}{|\Phi_g(\hat{\beta}t)|^2}\right| \int \int \frac{|\Phi_K(tb)|^2 |\Phi_K(sb)|^2}{|\Phi_g(\beta t)|^2 |\Phi_g(\beta s)|^2} |\psi(\beta, s, t)|^2 ds dt \\ &= O_p(n^{-1/2}b^{-1-\kappa}C_{V,b}).\end{aligned}$$

The other two terms in the above bounds are bounded similarly. Together with (2.1) and  $nb^{2\kappa+1} \rightarrow \infty$ , we obtain

$$|1 - \hat{C}_{V,b}/C_{V,b}| = O_p(n^{-1/2}b^{-1-\kappa}) = o_p(b^{-1/2}),$$



which implies (6.11)(a).

Next, consider (6.11)(b). Applying (2.1), (6.11)(a) and  $nb^{2\kappa+1} \rightarrow \infty$ ,

$$\begin{aligned}
& |\hat{C}_{M,b}/\hat{C}_{V,b}^{1/2} - C_{M,b}/C_{V,b}^{1/2}| \\
& \leq |\hat{C}_{M,b} - C_{M,b}| |\hat{C}_{V,b}^{-1/2}| + C_{M,b} C_{V,b}^{-1/2} \left| 1 - \hat{C}_{V,b}^{1/2}/C_{V,b}^{1/2} \right| \\
& \leq \max_{|t| \leq 1/b} \left| 1 - \frac{|\Phi_g(\beta t)|^2}{|\Phi_g(\hat{\beta} t)|^2} \right| C_{M,b} \hat{C}_{V,b}^{1/2} + o_p(1) \\
& = O_p(n^{-1/2} b^{-3/2-\kappa}) = o_p(1).
\end{aligned}$$

This completes the proof of (6.11), which combined with (6.10) also prove (2.4), thereby completing the proof of Theorem 2.1.

**Proof of Theorem 2.2.** Let  $\zeta_\beta(b) := \exp(2\nu(\beta)/b^\lambda)$ ,  $\beta \in \mathbb{R}$ . We shall first show that

$$(6.12) \quad \frac{n}{b^{\lambda-1+2\lambda\omega+2\lambda_0}\zeta_\beta(b)} \int (\hat{f}_n - f_n)^2(x) dx = o_p(1).$$

The proof is similar as in the ordinary smooth case. We only list some main differences.

First, arguing as for (6.4), for the super smooth case, (A') implies

$$\begin{aligned}
(6.13) \quad \max_{|t| \leq 1/b} \left| \frac{\Phi_g(-\hat{\beta}t)}{\Phi_g(-\beta t)} - 1 \right| &= \max_{|t| \leq 1/b} \left| \frac{\Phi_g(-\hat{\beta}t) - \Phi_g(-\beta t)}{\Phi_g(-\beta t)} \right| \\
&= O_p(n^{-1/2} b^{-1+2\lambda_0} \zeta_\beta^{1/2}(b)).
\end{aligned}$$

By (2.6), hence  $P(A_n) \rightarrow 1$ , with  $A_n := \{|\Phi_g(-\hat{\beta}t)| \geq |\Phi_g(-\beta t)|/2, t \in [-1/b, 1/b]\}$ .

Assumptions (A') and (B) imply that there exist constants  $M, c_\beta, C_\beta < \infty$ , such that for  $|t| > M$ ,  $c_\beta |t|^{\lambda_0} e^{-\nu(\beta)|t|^\lambda} \leq |\Phi_g(t)| \leq C_\beta |t|^{\lambda_0} e^{-\nu(\beta)|t|^\lambda}$  and  $|\Phi_f(t)| \leq C_{g1} |t|^{-\lambda_1}$ . Also, on the event  $A_n$ , there exists some  $\tilde{\beta}$  between  $\hat{\beta}$  and  $\beta$ , such that  $S_2$  is bounded from the above by

$$\frac{2\mu_g \|\hat{\beta} - \beta\|^2}{b^2} \int_{|t| \geq M} \frac{|\Phi_K(bt)|^2 (|\Psi_n(t) - \Phi_h(t)|^2 + |\Phi_h(t)|^2)}{|\Phi_g(-\beta t)|^4} dt + O_p(n^{-1}).$$

Based on (2.5),

$$\int_{|t| \geq M} \frac{|\Phi_K(bt)|^2 |\Psi_n(t) - \Phi_g(t)|^2}{|\Phi_g(-\beta t)|^4} dt = O_p(n^{-1} b^{\lambda-1+2\lambda\omega+4\lambda_0} \zeta_\beta^2(b)).$$

From Lemma 5 in van Es and Uh (2005), it follows that

$$\int_{|t| \geq M} \frac{|\Phi_K(bt)\Phi_h(-\beta t)|^2}{|\Phi_g(-\beta t)|^4} dt = O_p(b^{2\lambda_0+2\lambda_1+\lambda(1+2\omega)}\zeta_\beta(b)).$$

So when  $\lambda_1 > 1$  and  $n, b$  satisfy (2.6), we have

$$S_2 = o_p(n^{-1}b^{\lambda-1+2\lambda\omega+2\lambda_0}\zeta_\beta(b)).$$

Next, consider  $S_1$ . Assumptions (A'), (B') and the bound in (6.7) yield

$$(6.14) \quad S_1 \leq 8(\hat{\beta} - \beta)^2/b^2 S_{11} + 8(\hat{\alpha} - \alpha)^2/b^2 S_{12} + O_p\left(\frac{b^{2\lambda_0-1}\zeta_\beta(b)}{n^2 b^4}\right).$$

Consider  $S_{11}$  first. Similar as in (6.8), together with assumptions (A')–(C'), we obtain

$$(6.15) \quad S_{11} = O_p(n^{-1}b^{-1+2\lambda_0}\zeta_\beta(b)) + O_p(1) + O_p(b^{2\lambda_0-1+2\lambda_1+\lambda(1+2\omega)}\zeta_\beta(b)).$$

$S_{12}$  can be considered the same way. Hence, by (2.6), (6.14) and (6.15)

$$\begin{aligned} \frac{nS_1}{b^{\lambda-1+2\lambda\omega+2\lambda_0}\zeta_\beta(b)} &= O_p(n^{-1}b^{-\lambda-2\lambda\omega}) + O_p(n^{-1}b^{-\lambda-2\lambda\omega-1-2\lambda_0}) \\ &\quad + O(b^{2\lambda_1-2}) + O_p(n^{-1}b^{-\lambda-4-2\lambda\omega}) = o_p(1). \end{aligned}$$

This completes the proof of (6.12). Combining this with (2.5), we obtain

$$(6.16) \quad \begin{aligned} \hat{T}_n - T_n &= \int (f_n - \hat{f}_n)^2(x) dx + 2 \int (f_n - \hat{f}_n)(f_n - K_b * f_0)(x) dx \\ &= o_p(n^{-1}b^{\lambda-1+2\lambda\omega+2\lambda_0}\zeta_\beta(b)). \end{aligned}$$

Also, since the first derivatives of  $\nu(\beta)$  and  $C(\beta)$  exist and  $\|\hat{\beta} - \beta\| = O_p(n^{-1/2})$ ,

$$(6.17) \quad |1 - \exp(-2(\nu(\hat{\beta}) - \nu(\beta))/b^\lambda)| = o_p(1).$$

Facts (6.16) and (6.17) yield (2.7). This completes the proof of Theorem 2.2.

**Proof of Theorem 3.1.** Define

$$\tilde{T}_n = \int (\hat{f}_n(x) - K_b * f_1(x))^2 dx.$$

Argue as in the proof of Theorem 2.1 to obtain

$$(6.18) \quad n\tilde{C}_{V,b}^{-1/2}(\tilde{T}_n - \hat{C}_{M,b}/(2\pi n)) \rightarrow_d \mathcal{N}(0, 1/2\pi^2),$$

where  $\tilde{C}_{V,b}$  is same as  $\hat{C}_{V,b}$  with  $f$  replaced by  $f_1$ . Hence,  $\tilde{C}_{V,b} \approx b^{-(4\kappa+1)}$ .

Next, consider

$$\begin{aligned} & nb^{2\kappa+1/2}(\hat{T}_n - \tilde{T}_n) \\ &= nb^{2\kappa+1/2} \int (K_b * f_0(x) - K_b * f_1(x))^2 dx \\ &\quad + 2nb^{2\kappa+1/2} \int (\hat{f}_n(x) - K_b * f_1(x))(K_b * f_1(x) - K_b * f_0(x)) dx. \end{aligned}$$

Because  $\int (K_b * f_0(x) - K_b * f_1(x))^2 dx \rightarrow \|f_1 - f_0\|^2 > 0$  and  $nb^{2\kappa+3} \rightarrow \infty$ , the first term in the right hand side above is of the order  $O(nb^{2\kappa+1/2}) \rightarrow \infty$ , while by (6.18) and the Cauchy-Schwarz inequality, the second term is of the order  $o_p(nb^{2\kappa+1/2})$ . This completes the proof of Theorem 3.1.

The proofs of Theorems 3.2, 3.3 and 3.4 are similar to those of Theorems 3.1 and 2.1, and hence no details are given.

**Proof of Theorem 3.5.** For the sake of completeness of this paper, we first provide a brief proof of (3.3). For  $j = 1, \dots, n$ , let

$$D_j = \frac{1}{\pi} \int |\Phi_K(tb)|^2 \left( \frac{e^{it(\varepsilon_j - \beta' u_j)}}{\Phi_g(-\beta t)} - \Phi_{f_1}(t) \right) \overline{(\Phi_{f_1}(t) - \Phi_{f_0}(t))} dt.$$

Note that since  $K$  is symmetric,  $D_j$  is real. Rewrite

$$\begin{aligned} & T_n - \|K_b * (f_1 - f_0)\|^2 \\ &= \int (f_n - K_b * f_1)^2 dx + 2 \int (f_n - K_b * f_1)(K_b * (f_1 - f_0)) dx. \end{aligned}$$

Recall (2.2) and that  $nb^{4\kappa+2} \rightarrow \infty$ . Hence, the first term on the right hand side above is  $O_p(n^{-1}b^{2\kappa+1}) = o_p(n^{-1/2})$ . Using Parseval's equation, the second term can be written as  $n^{-2} \sum_{j=1}^n D_j$ . Here,  $D_j$ 's are arrays of i.i.d. r.v.'s, with  $ED_1 = 0$ , and  $\text{Var}(D_1)$  converging to

$$\begin{aligned} \tau_0^2 &:= \frac{1}{\pi^2} \int \int \Phi_h(t-s) \frac{(\Phi_{f_1}(s) - \Phi_{f_0}(s)) \overline{(\Phi_{f_1}(t) - \Phi_{f_0}(t))}}{\Phi_g(\beta s) \Phi_g(-\beta t)} ds dt \\ &\quad - \frac{1}{\pi^2} \left( \int \Phi_{f_1}(-t) (\Phi_{f_1}(t) - \Phi_{f_0}(t)) dt \right)^2 \\ &= \frac{1}{2\pi^3} \text{Var} \left( \int e^{-it\varepsilon} \frac{\Phi_{f_1}(t) - \Phi_{f_0}(t)}{\Phi_g(\beta t)} dt \right). \end{aligned}$$

Moreover,

$$\mathbb{E}|D_1|^4 \leq \frac{1}{\pi^4} \left( \int \left( \frac{1}{|\Phi_g(-\beta t)|} + |\Phi_{f_1}(t)| \right) (|\Phi_{f_1}(t)| + |\Phi_{f_0}(t)|) dt \right)^4 = O(1),$$

by the assumption (B) with  $r > \kappa + 1$ . Hence one obtains (3.3), by the Lindeberg-Feller CLT.

To complete the proof of Theorem 3.5, first, consider the case where  $\alpha$  is known, so that  $\hat{f}_n$  is based on the residuals  $Y_i - \alpha - \hat{\beta}'Z_i$ 's only. Without loss of generality, assume  $\alpha = 0$ . Under the alternative  $H_1$ ,

$$\begin{aligned} & n^{1/2}(\hat{T}_n - T_n) \\ &= n^{1/2} \int (\hat{f}_n - f_n)^2(x) dx + 2n^{1/2} \int (\hat{f}_n - f_n)(f_n - K_b * f_1)(x) dx \\ & \quad + 2n^{1/2} \int (\hat{f}_n - f_n)(K_b * f_1 - K_b * f_0)(x) dx. \end{aligned}$$

The same proof as that of (6.1) and  $nb^{4\kappa+2} \rightarrow \infty$  imply

$$(6.19) \quad n^{1/2} \int (\hat{f}_n - f_n)^2(x) dx = o_p(n^{-1/2}b^{-2\kappa}) = o_p(1).$$

This fact together with (2.4) and the Cauchy-Schwarz inequality implies

$$(6.20) \quad 2n^{1/2} \int (\hat{f}_n - f_n)(f_n - K_b * f_1)(x) dx = o_p(n^{-1}b^{-3\kappa-1}) = o_p(1).$$

To deal with the remaining part, let  $\Delta_f(x) := (K_b * f_1 - K_b * f_0)(x)$ . Rewrite  $\hat{f}_n - f_n$  as the sum of the following two terms:

$$\mathcal{D}_1 := \int \int e^{-itx} \Phi_K(bt) \frac{\sum_{j=1}^n (e^{it(\varepsilon_j - \hat{\beta}'u_j)} - e^{it(\varepsilon_j - \beta'u_j)})}{2\pi n \Phi_g(-\hat{\beta}t)} \Delta_f(x) dt dx,$$

$$\mathcal{D}_2 := \int \int e^{-itx} \Phi_K(bt) \frac{\sum_{j=1}^n e^{it(\varepsilon_j - \hat{\beta}'u_j)}}{2\pi n} \left( \frac{1}{\Phi_g(-\hat{\beta}t)} - \frac{1}{\Phi_g(-\beta t)} \right) \Delta_f(x) dt dx.$$

Consider  $\mathcal{D}_1$  first. Since  $nb^{4\kappa+2} \rightarrow \infty$ , and  $\kappa > 1$ , then uniformly in  $|t| \leq 1/b$ ,

$$\frac{1}{n} \sum_{j=1}^n \left( e^{it(\varepsilon_j - \hat{\beta}'u_j)} - e^{it(\varepsilon_j - \beta'u_j)} \right) = \frac{\sum_{j=1}^n it(\beta - \hat{\beta})'Z_j e^{it(\varepsilon_j - \hat{\beta}'u_j)}}{n} + o_p(n^{-1/2}).$$

Let

$$\mathcal{C}_0 := \int \int te^{-itx} \Phi_K(bt) \frac{\sum_{j=1}^n [Z_{jk} e^{it(\varepsilon_j - \beta' u_j)} - \mathbb{E} Z_{jk} e^{it(\varepsilon_j - \beta' u_j)}]}{2\pi n \Phi_g(-\beta t)} dt \Delta_f(x) dx.$$

Then  $\mathbb{E} \mathcal{C}_0 = 0$  and

$$\begin{aligned} (6.21) \quad \mathbb{E} \mathcal{C}_0^2 &\leq \mathbb{E} \left( \sum_{j=1}^n \int \int te^{-itx} \Phi_K(bt) \frac{Z_{jk} e^{it(\varepsilon_j - \beta' u_j)} - \mathbb{E} Z_{jk} e^{it(\varepsilon_j - \beta' u_j)}}{2\pi n \Phi_g(-\beta t)} \right. \\ &\quad \left. \times dt \Delta_f(x) dx \right)^2 \\ &\leq \frac{\mathbb{E} |Z_k|^2}{n} \left( \int \int te^{-itx} \frac{\Phi_K(bt)}{2\pi \Phi_g(-\beta t)} dt \Delta_f(x) dx \right)^2 \\ &\leq \frac{\mathbb{E} |Z_k|^2}{n} \left( \int \int te^{-itx} \Phi_K(bt) \frac{\Phi_K(bt)}{2\pi \Phi_g(-\beta t)} dt \Delta_f(x) dx \right)^2 \\ &\leq \frac{\mathbb{E} |Z_k|^2}{2\pi n} \left( \int \frac{|\Phi_K(bt)|}{|\Phi_g(-\beta t)|} dt \int |\Delta_f(x)| dx \right)^2 = O(n^{-1} b^{-2\kappa-2}) = o(1). \end{aligned}$$

Hence,  $\mathcal{C}_0 = o_p(1)$ . Since

$$\mathbb{E} Z_k e^{it(\varepsilon - \beta' u)} = \mu_Z \Phi_h(t) + \Phi_{f_1}(t) \mathbb{E} u_k e^{-i\beta' u t},$$

assumption (B) with  $r > \kappa + 1$  and the relation  $\Phi_h(t) = \Phi_g(t) \Phi_{f_1}(t)$  imply

$$\int \int te^{-itx} \Phi_K(bt) \frac{\mathbb{E} Z_k e^{it(\varepsilon - \beta' u)}}{\Phi_g(-\beta t)} dt \Delta_f(x) dx = O(1).$$

Together with (6.4), and  $nb^{4\kappa+2} \rightarrow \infty$ , the above analysis yields

$$\begin{aligned} (6.22) \quad \mathcal{D}_1 + \frac{i(\hat{\beta} - \beta)'}{2\pi n} \int \int te^{-itx} \Phi_K(bt) \frac{\mathbb{E} Z e^{it(\varepsilon - \beta' u)}}{\Phi_g(-\beta t)} dt \Delta_f(x) dx \\ = O_p(n^{-1} b^{-\kappa-1}) = o_p(n^{-1/2}). \end{aligned}$$

Next consider  $\mathcal{D}_2$ . Uniformly in  $|t| \leq 1/b$ ,

$$\begin{aligned} &\frac{\Phi_g(-\hat{\beta} t) - \Phi_g(-\beta t)}{\Phi_g^2(-\beta t)} \\ &= \sum_{k=1}^p \left\{ \frac{it(\beta_k - \hat{\beta}_k) \mathbb{E} u_k e^{-i\beta' u t}}{\Phi_g^2(-\beta t)} - \frac{(\beta_k - \hat{\beta}_k)^2 t^2 \mathbb{E} u^2 e^{-i\beta' u t}}{\Phi_g^2(-\beta t)} \right\} \\ &\quad + O_p(n^{-3/2} b^{-3-2\kappa}). \end{aligned}$$

Let

$$\begin{aligned}\mathcal{C}_1 &:= \int \int t e^{-itx} \Phi_K(bt) \frac{\sum_{j=1}^n (e^{it(\varepsilon - \beta' u)} - \Phi_h(t)) \text{E} u_k e^{-i\beta' ut}}{\Phi_g^2(-\beta t)} \Delta_f(x) dt dx, \\ \mathcal{C}_2 &:= \int \int t^2 e^{-itx} \Phi_K(bt) \frac{\sum_{j=1}^n (e^{it(\varepsilon - \beta' u)} - \Phi_h(t)) \text{E} u_k^2 e^{-i\beta' ut}}{\Phi_g^2(-\beta t)} \Delta_f(x) dt dx.\end{aligned}$$

Note that  $\text{EC}_i = 0$ ,  $i = 1, 2$ , and the same arguments as for (6.21) yield

$$\text{EC}_1^2 = O(n^{-1} b^{-4\kappa-2}) = o(1), \quad \text{EC}_2^2 = O(n^{-1} b^{-4\kappa-4}) = o(b^{-2}).$$

Hence,  $\mathcal{C}_1 = o_p(1)$  and  $\mathcal{C}_2 = o_p(b^{-1})$ . Since  $\Phi_h(t) = \Phi_{f_1}(t) \Phi_g(-\beta t)$ ,  $nb^{4\kappa+2} \rightarrow \infty$ , by (6.4) and assumption (B) with  $r > \kappa + 1$ , we obtain

$$\begin{aligned}(6.23) \quad \mathcal{D}_2 &- \frac{i(\hat{\beta} - \beta)'}{2\pi n} \int \int t e^{-itx} \Phi_K(bt) \frac{\Phi_{f_1}(t) \text{E} u e^{-i\beta' ut}}{\Phi_g(-\beta t)} dt \Delta_f(x) dx \\ &= o_p(n^{-1/2}).\end{aligned}$$

Also,

$$-\frac{1}{2\pi} \int i t e^{-itx} \Phi_K(bt) \Phi_{f_1}(t) dt = K_b * \dot{f}_1(x).$$

Combine this with (6.22) and (6.23) to obtain

$$2n^{1/2} \int (\hat{f}_n - f_n)(f_n - K_b * f_0)(x) dx = (\hat{\beta} - \beta)' B_f + o_p(n^{-1/2}).$$

Recall (6.19) and (6.20), immediately

$$(6.24) \quad n^{1/2}(\hat{T}_n - T_n - (\hat{\beta} - \beta)' B_f) = o_p(1).$$

Next, consider the case when the intercept parameter  $\alpha$  is unknown. Let  $a = \alpha - \hat{\alpha}$ . Then

$$\begin{aligned}\hat{T}_n &= \int (f_n(x, \alpha, \hat{\beta}) - K_b * f_0(x+a))^2 dx \\ &= \int (f_n(x, \alpha, \hat{\beta}) - K_b * f_0(x))^2 dx \\ &\quad + \int (K_b * f_0(x+a) - K_b * f_0(x))^2 dx \\ &\quad - 2 \int (f_n(x, \alpha, \hat{\beta}) - K_b * f_0(x))(K_b * f_0(x+a) - K_b * f_0(x)) dx.\end{aligned}$$

The first term on the right side above is  $T_n(\alpha, \hat{\beta})$ , and from (6.24) we obtain

$$n^{1/2}(T_n(\alpha, \hat{\beta}) - T_n - (\hat{\beta} - \beta)'B_f) = o_p(1).$$

Because  $\dot{f}_0$  exists, and is finite, and  $a = O_p(n^{-1/2})$ , the second term is  $O_p(n^{-1})$ . Then to deal with the third term, rewrite the factor multiplying -2 as the sum of the following three terms:

$$\begin{aligned} & \int (f_n(x, \alpha, \hat{\beta}) - f_n(x))(K_b * f_0(x+a) - K_b * f_0(x)) \, dx, \\ & \int (f_n(x) - K_b * f_1(x))(K_b * f_0(x+a) - K_b * f_0(x)) \, dx, \\ & \int (K_b * f_1(x) - K_b * f_0(x))(K_b * f_0(x+a) - K_b * f_0(x)) \, dx. \end{aligned}$$

By using the Cauchy-Schwarz inequality, together with  $a = O_p(n^{-1/2})$ , (6.18) and (6.24), verify that each of the first two terms above are  $o_p(n^{-1/2})$ . The finiteness of  $\dot{f}_0$  implies that the third term is equal to

$$a \int (K_b * f_1(x) - K_b * f_0(x))(K_b * \dot{f}_0(x)) \, dx + o_p(n^{-1/2}).$$

The above analysis and (6.24) imply

$$(6.25) \quad n^{1/2}(\hat{T}_n - T_n - (\hat{\beta} - \beta)'B_f - (\hat{\alpha} - \alpha)A_f) = o_p(1).$$

This fact together with (3.3) completes the proof of Theorem 3.5.

**Proof of Theorem 3.6.** For  $\hat{T}_n$ , recall (3.5), (3.6) and (6.25). Using the details in the proof of Theorem 3.5, we obtain,

$$\hat{T}_n - \|K_b * (f_1 - f_0)\|^2 = \frac{1}{n} \sum_{j=1}^n (D_j + \eta_j A_f + \zeta_j' B_f) + o_p(n^{1/2}).$$

We write

$$(6.26) \quad \tau^2 := \text{Var}(D_1 + \eta_1 A_f + \zeta_1' B_f).$$

Since  $D_j + \eta_j A_f + \zeta_j' B_f$ , for  $j = 1, \dots, n$  are arrays of i.i.d. zero mean r.v.'s and  $E|D_1|^4 = O(1)$ ,  $E|\eta|^2 + \vartheta < \infty$  and  $E\|\zeta\|^2 + \vartheta < \infty$  for some  $\vartheta > 0$ . Thus the claim (3.7) follows by the Lindeberg-Feller CLT, thereby completing the proof.

The proof of Theorem 3.7 is similar as the arguments in the proof of Theorem 3.5 and 3.6. Thus we omit the details of the proof.

**Acknowledgement:** Authors thank the two referees and a diligent Associate Editor for their constructive comments.

## SUPPLEMENTARY MATERIAL

**Some simulation results of GOF tests in measurement error models**

(doi: [COMPLETED BY THE TYPESETTER](#); .pdf). This supplement contains some additional simulation results comparing the test proposed in this paper with some other tests and a bandwidth sensitivity analysis.

**REFERENCES**

- Bachmann, D. and Dette, H. (2005). A note on the Bickel–Rosenblatt test in autoregressive time series. *Statist. Probab. Lett.*, **74(3)**, 221–234.
- Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.*, **1**, 1071–1095.
- Butucea, C. (2004). Asymptotic normality of the integrated square error of a density estimator in the convolution model. *Sort*, **28(1)**, 9–26.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, **83(404)**, 1184–1186.
- Carroll, R. J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C.M. (2006). *Measurement error in nonlinear models: a modern perspective*. CRC Press.
- Cheng, C.L. and Van Ness, J. W. (1999). *Statistical regression with measurement error*. John Wiley & Sons, Ltd.
- Delaigle, A., and Hall, P. (2006). On optimal kernel choice for deconvolution. *Statist. Probab. Lett.*, **76(15)**, 1594–1602.
- Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. CBMS Regional Conference Series in Applied Mathematics **9**. SIAM, Philadelphia.
- Fan, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, **53(1)**, 97–110.
- Fuller, W. A. (1987). *Measurement error models*. John Wiley & Sons, New York.
- Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.*, **103(484)**, 1584–1594.
- Huáková, M., and Meintanis, S. G. (2007). Omnibus tests for the error distribution in the linear regression model. *Statistics*, **41(5)**, 363–376.
- Holzmann, H., Bissantz, N., and Munk, A. (2007). Density testing in a contaminated sample. *J. Multivariate Anal.*, **98(1)**, 57–75.
- Holzmann, H. and Boysen, L. (2006). Integrated square error asymptotics for supersmooth deconvolution. *Scand. J. Statist.*, **33(4)**, 849–860.
- Hong, Y. and Lee, Y. (2013). A loss function approach to model specifica-



- tion testing and its relative efficiency. *Ann. Statist.*, **41(3)**, 1166–1203.
- Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32**, 995–1034.
- Khmaladze, E. V. and Koul, H. L. (2009). Goodness-of-fit problem for errors in nonparametric regression. *Ann. Statist.* **37**, 3165–3185.
- Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models (2nd ed.)*. Lecture Notes in Statistics, **166**, Springer-Verlag, New York.
- Koul, H. L. and Nao, M. (2012). A goodness-of-fit test for GARCH innovation density. *Metrika*, **75(1)**, 127–149.
- Koul, H. L. and Song, W. (2012). A class of goodness-of-fit tests in linear errors-in-variables model. *J. SFdS*, **153(1)**, 52–70.
- Koul, H. L., Song, W. and Zhu, X. (2017). Supplement to ‘Goodness-of-fit Testing of Error Distribution in Linear Measurement Error Models’.
- Laurent, B., Loubes, J. and Marteau, C. (2011). Testing inverse problems: A direct or an indirect problem? *J. Statist. Plann. Inference*, **141**, 1849–1861.
- Lee, S. and Na, S. (2002). On the BickelRosenblatt test for first-order autoregressive models. *Statist. Probab. Lett.*, **56**, 23–35.
- Loynes, R. M. (1980). The empirical d.f. of residuals from generalized regression. *Ann. Statist.*, **8**, 285–298.
- Loubes, J. M. and Marteau, C. (2014). Goodness-of-fit testing strategies from indirect observations. *J. Nonparametr. Stat.*, **26(1)**, 85–99.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer, New York.
- Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.*, **9**, 229–235.
- Stefanski, L. A. and Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics*, **21(2)**, 169–184.
- Van Es, A. and Uh, H. -W. (2004). Asymptotic normality of nonparametric kernel type deconvolution density estimators: crossing the cauchy boundary. *J. Nonparametr. Stat.*, **16(1-2)**, 261–277.

HIRA L. KOUL  
DEPARTMENT OF STATISTICS  
AND PROBABILITY  
MICHIGAN STATE UNIVERSITY  
EAST LANSING, MI 48824-1027  
U. S. A.  
E-MAIL: KOUL@STT.MSU.EDU

WEIXING SONG  
DEPARTMENT OF STATISTICS  
KANSAS STATE UNIVERSITY  
MANHATTAN, KANSAS  
U. S. A.  
E-MAIL:

XIAOQING ZHU  
DEPARTMENT OF STATISTICS  
AND PROBABILITY  
MICHIGAN STATE UNIVERSITY  
EAST LANSING, MI 48824-1027  
U. S. A.  
E-MAIL: ZHUXIAOQ@STT.MSU.EDU