UCL Université catholique de Louvain

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

ISBA

DISCUSSION PAPER

2017/24

A Smooth Nonparametric, Multivariate, Mixed-Data Location-Scale Test

Racine, J. and I. Van Keilegom

A SMOOTH NONPARAMETRIC, MULTIVARIATE, MIXED-DATA LOCATION-SCALE TEST

JEFFREY S. RACINE

Department of Economics and Graduate Program in Statistics, McMaster University, racinej@mcmaster.ca; Department of Economics and Finance, La Trobe University; Info-Metrics Institute, American University; Rimini Center for Economic Analysis; Center for Research in Econometric Analysis of TimE Series (CREATES), Aarhus University.

INGRID VAN KEILEGOM

ORSTAT, KU Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

Abstract. A number of tests have been proposed for assessing the location-scale assumption that is often invoked by practitioners. Existing approaches include Kolmogorov-Smirnov and Cramér-von-Mises statistics that each involve measures of divergence between unknown joint distribution functions and products of marginal distributions. In practice, the unknown distribution functions embedded in these statistics are approximated using non-smooth empirical distribution functions. We demonstrate how replacing the non-smooth distributions with their kernel-smoothed counterparts can lead to substantial power improvements. In so doing we extend existing approaches to the smooth multivariate and mixed continuous and discrete data setting thereby extending the reach of existing approaches. Theoretical underpinnings are provided, Monte Carlo simulations are undertaken to assess finite-sample performance, and illustrative applications are provided.

1. Introduction

A variety of tests have been proposed for assessing the appropriateness of the location-scale assumption that is often invoked in applied settings; see by way of illustration Akritas & Van Keilegom (2001) and Li & Racine (forthcoming), who adopt the location-scale framework, see Einmahl

Date: August 17, 2017.
1991 *Mathematics Subject Classification.* C14 Semiparametric and Nonparametric Methods.
Key words and phrases. Kernel Smoothing; Kolmogorov-Smirnov; Inference.
Racine would like to gratefully acknowledge support from the Natural Sciences and Engineering Research Council of Canada (NSERC:www.nserc.ca), the Social Sciences and Humanities Research Council of Canada (SSHRC:www.sshrc.ca), and the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca). I. Van Keilegom acknowledges financial support from the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409), and from IAP research network P7/06 of the Belgian Government (Belgian Science Policy).

& Van Keilegom (2008), Birke, Neumeyer & Volgushev (2017) and Neumeyer, Noh & Van Keilegom (2016) for various approaches that have been proposed to test the location-scale assumption in a range of settings, and see Neumeyer (2009) for a bootstrap procedure for the error distribution in these models. These approaches employ test statistics that are based on conditional mean models, in particular, the difference between the joint distribution of the predictor and error and the product of the marginal distributions of the predictor and error, and include the Kolmogorov-Smirnov (Kolmogorov 1933, Smirnov 1948), Cramér-von-Mises (Cramér 1928, von Mises 1928) and Anderson-Darling (Anderson & Darling 1952) statistics, among others. In this literature, the unknown joint and marginal distributions are estimated using the respective non-smooth empirical distribution functions (EDFs). However, it turns out that substantial power gains can be realized by replacing the non-smooth EDFs with their kernel-smoothed counterparts. We demonstrate that we retain all of the desirable features of this testing framework yet can realize substantial improvements to existing procedures from the vantage point of finite-sample power without impacting size. Though we consider inference for location-scale models, the results contained herein are of broad applicability and ought to appeal to a wide audience, particularly practitioners concerned with power properties associated with this popular class of test statistics.

The remainder of this paper proceeds as follows: Section 2 outlines the location-scale framework and the proposed smooth testing procedure; Section 3 presents the theoretical underpinnings of the proposed approach; Section 4 presents simulation evidence that demonstrates power gains achievable by a fully data-driven implementation of the proposed approach; Section 5 considers an illustrative application, while Section 6 presents some concluding remarks. The proofs of the main results are given in Appendix A, detailed tables outlining power gains are presented in Appendix B and C, while `R` code (R Core Team 2016) underlying the simulations can be found in Appendix D.

2. Methodology

Consider a smooth location-scale model of the form

$$Y = \mu(X) + \sigma(X)\epsilon,$$

where $\mu(\cdot)$ and $\sigma(\cdot) \geq 0$ are unknown smooth location and scale functions, X is a vector of predictors, and ϵ has zero mean, unit variance, and is otherwise an unknown error process with distribution F_ϵ that is independent of X. We observe n independent copies of (X^{T}, Y), denoted by $(X_1^{\mathrm{T}}, Y_1), \ldots, (X_n^{\mathrm{T}}, Y_n)$.

The location-scale assumption is often invoked as it confers a number of useful properties on the resulting estimator, including i) simpler asymptotic properties than its unstructured counterpart,[1] ii) the ability to nonparametrically estimate the error distribution at a $\sqrt{n}$-rate (Akritas & Van Keilegom 2001, Escanciano & Jacho-Chávez 2012), and iii) more efficient estimation of the conditional distribution of Y given X than its unstructured counterpart.

Even though the independence of the predictors X and error ϵ is a weak and common assumption (see e.g. Akritas & Van Keilegom (2001) and Li & Racine (forthcoming)), particularly in Econometrics, it might be too strong, hence a testing procedure having high power is particularly appealing. For what follows, we define $F_X(x) = P(X \leq x)$, $F_\epsilon(t) = P(\epsilon \leq t)$, and $F_{X,\epsilon}(x,t) = P(X \leq x, \epsilon \leq t)$, and we let

$$H_0 : X \text{ and } \epsilon \text{ are independent.}$$

Consider by way of illustration the test of Einmahl & Van Keilegom (2008) which can be used to assess the adequacy of the location-scale assumption. In essence, Einmahl & Van Keilegom (2008) test for independence between the predictors X and error ϵ in the location-scale model $Y = \mu(X) + \sigma(X)\epsilon$. Given kernel estimates of $\mu(x) = E(Y|X = x)$ and $\sigma^2(x) = V(Y|X = x)$, denoted $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$, one tests for independence of X_i and $\hat{\epsilon}_i = (Y_i - \hat{\mu}(X_i))/\hat{\sigma}(X_i)$ using, for instance, a Kolmogorov-Smirnov test statistic of the form

$$T_{KS} = \sqrt{n} \sup_{x,t} |\hat{F}_{X,\hat{\epsilon}}(x,t) - \hat{F}_X(x)\hat{F}_{\hat{\epsilon}}(t)|, \tag{1}$$

where $\hat{F}_{X,\hat{\epsilon}}(x,t)$, $\hat{F}_X(x)$ and $\hat{F}_{\hat{\epsilon}}(t)$ are the respective EDFs. Due to the inadequacy of using the asymptotic distribution of T_{KS} for inference, a simple bootstrap procedure is used instead to obtain the null distribution from which nonparametric P-values can readily be obtained. This procedure can be easily modified to test for the validity of a homoskedastic model $Y_i = \mu(X_i) + \epsilon_i$ with ϵ_i being independent of X_i, which is also a common assumption in Econometrics, or to test the validity of a

[1] By 'unstructured' we mean a model of the form $Y_i = \mu(X_i) + \epsilon_i$ with $E(\epsilon_i|X_i) = 0$.

transformation model of the form $\Delta(Y_i) = \mu(X_i) + \sigma(X_i)\epsilon_i$ as outlined in Neumeyer et al. (2016), where $\Delta(\cdot)$ is some parametric monotone transformation.

We note in passing that the Cramér-von-Mises statistic, which is also popular in applied settings, is given by

$$T_{CM} = n \int \int \left(\hat{F}_{X,\hat{\epsilon}}(x,t) - \hat{F}_X(x)\hat{F}_{\hat{\epsilon}}(t)\right)^2 d\hat{F}_X(x)\, d\hat{F}_{\hat{\epsilon}}(t).$$

We propose replacing the EDFs in these statistics with their kernel-smoothed counterparts using the approach of Li, Li & Racine (2015) that is briefly described below. Our approach is multivariate in nature and allows for mixed datatypes, features that to the best of our knowledge have not been exploited in the literature. Related work includes Conover (1999, pp. 396–406) who considers a smooth two-sample Kolmogorov-Smirnov test,[2] Bowman, Hall & Prvan (1998) who consider bandwidth selection for univariate kernel smoothed CDFs and Wang et al. (2013) who consider a plug-in bandwidth procedure for smooth univariate kernel smoothed CDFs with a focus on the construction of simultaneous confidence bands.

2.1. **Kernel Estimation of $F_X(x)$, $F_\epsilon(t)$, and $F_{X,\epsilon}(x,t)$ with Mixed Data.** Though Einmahl & Van Keilegom (2008) and others restrict attention to the scalar predictor case, in applied settings one would expect to encounter multivariate predictors that, in addition, might consist of both discrete and continuous datatypes. Though the Kolmogorov-Smirnov and Cramér-von-Mises statistics were developed under the assumption that the random variables possessed continuous distributions $F(\cdot)$ and have been extended to instead admit discrete distributions (Conover 1972, Gleser 1985, Choulakian, Lockhart & Stephens 1994, Lockhart, Spinelli & Stephens 2007), to the best of our knowledge they are unable to handle the multivariate mix of continuous and discrete data often found in regression settings. Our approach tackles this shortcoming by leveraging recent work on nonparametric kernel estimation of distributions involving a mix of discrete and continuous variables.

Li et al. (2015) propose an estimator of a joint distribution function defined over a mix of continuous and discrete random variables, which we will explain by means of the vector of covariates X. We suppose that X_j $(j = 1, \ldots, n)$ is a $(q + r)$-dimensional vector of covariates, consisting of q

[2]This approach compares two kernel smoothed *univariate* distributions; see the function `KS.test` in the `R` package Qiu (2014) which implements this procedure using Wang, Cheng & Yang's (2013) plug-in bandwidth and uses the asymptotic distribution for critical values which is known to be problematic.

continuous covariates denoted by $X_j^c = (X_{j1}^c, \ldots, X_{jq}^c)$ and r (ordered) discrete covariates denoted by $X_j^d = (X_{j1}^d, \ldots, X_{jr}^d)$. Likewise, X consists of a q-dimensional vector of continuous covariates $X_1^c, \ldots, X_q^c$ and an r-dimensional vector of discrete covariates $X_1^d, \ldots, X_r^d$. The support of X is denoted by $R_X = R_{X^c} \times R_{X^d}$, where R_{X^c} is supposed to be a compact subset of $\mathcal{R}^q$. We consider smooth kernel-based estimators of $F_X(x) = P(X \leq x) = P(X^c \leq x^c, X^d \leq x^d)$ (where inequalities should be understood componentwise).

Assume that X_{js}^d takes values in $\{0, 1, \ldots, c_s - 1\}$ $(s = 1, \ldots, r)$, where $c_s \geq 2$ is a positive integer. Let λ_s denote the bandwidth for the s-th discrete variable. We use the kernel function $l(x_s^d, X_{js}^d, \lambda_s) = \eta_s \sum_{z_s^d \leq x_s^d} \lambda_s^{|X_{js}^d - z_s^d|}$, with $\lambda_s^0 = 1$, $0^0 = 1$, $\lambda_s \in [0, 1]$, and η_s a normalizing factor such that $l(c_s - 1, c_s - 1, \lambda_s) = 1$. Write the product (discrete variable) cumulative kernel function as $L_\lambda(x^d, X_j^d) = \prod_{s=1}^r l(x_s^d, X_{js}^d, \lambda_s)$.

Let h_s be the bandwidth associated with X_s^c $(s = 1, \ldots, q)$. The product cumulative kernel function used for the continuous variables is given by $K_h(x^c, X_j^c) = \prod_{s=1}^q \int_{-\infty}^{x_s^c} h_s^{-1} k((z_s^c - X_{js}^c)/h_s)\, dz_s^c$, where $k(\cdot)$ is a univariate density kernel function for a continuous variable such as the standard Epanechnikov or Gaussian kernel function. The cumulative kernel function for the vector of mixed variables is simply the product of $K_h(\cdot)$ and $L_\lambda(\cdot)$ defined above and is given by $G_\gamma(x, X_j) = K_h(x^c, X_j^c) \times L_\lambda(x^d, X_j^d)$, where $\gamma = (h, \lambda)$. Li et al. (2015) consider the mixed-datatype kernel estimator of $F_X(x)$ defined by

$$\hat{F}_X(x) = \frac{1}{n} \sum_{j=1}^n G_\gamma(x, X_j). \tag{2}$$

Next, to estimate $F_\epsilon(t)$, we assume that ϵ is continuous and hence $F_\epsilon(t)$ can be estimated by a (univariate) continuous cumulative kernel estimator. However, ϵ is not observed, so we first need to estimate it. To estimate $\mu(x)$ and $\sigma(x)$, we use local polynomial smoothing for the continuous covariates (see Fan & Gijbels (1996) or Ruppert & Wand (1994), among others), and for the discrete covariates, we use a variation on Aitchison & Aitken (1976)'s kernel function defined by

$$v(x_s^d, X_{js}^d, \nu_s) = \begin{cases} 1 & \text{if } x_s^d = X_{js}^d \\ \nu_s & \text{otherwise.} \end{cases}$$

The range of ν_s is $[0,1]$. Note that when $\nu_s = 0$ the above kernel function becomes an indicator function, and when $\nu_s = 1$, it is a constant function. The product kernel function for the vector x^d of discrete covariates is then given by

$$V_\nu(x^d, X_j^d) = \prod_{s=1}^{r} \nu_s^{1-1(x_s^d = X_{js}^d)}.$$

Combining this with local polynomial smoothing of the continuous covariates (of which the order p will depend on the dimension q and will be determined later – see assumption (A2) in Appendix A), we define $\hat{\mu}(x) = \hat{\beta}_0$, where $\hat{\beta}_0$ is the first component of the vector $\hat{\beta}$, which is the solution of the local minimization problem

$$\min_{\beta} \sum_{j=1}^{n} \Big\{ Y_j - P_j(\beta, x^c, p) \Big\}^2 \prod_{s=1}^{q} \frac{1}{g_s} k\Big(\frac{x_s^c - X_{js}^c}{g_s} \Big) V_\nu(x^d, X_j^d), \tag{3}$$

where $P_j(\beta, x^c, p)$ is a polynomial of order p built up with all $0 \le k \le p$ products of factors of the form $X_{js}^c - x_s^c$ $(s = 1, \ldots, q)$. The vector β is the vector of length $\sum_{k=0}^{p} q^k$, consisting of all coefficients of this polynomial. Here, $g = (g_1, \ldots, g_q)$ is a q-dimensional bandwidth vector. To estimate $\sigma^2(x)$, define $\hat{\sigma}^2(x) = \hat{\gamma}_0$, where $\hat{\gamma}_0$ is defined in the same way as $\hat{\beta}_0$, but with Y_j replaced by $(Y_j - \hat{\mu}(X_j))^2$ in (3) $(j = 1, \ldots, n)$.

Then, let $\hat{\epsilon}_j = (Y_j - \hat{\mu}(X_j))/\hat{\sigma}(X_j)$ be the j-th residual, and define

$$\hat{F}_{\hat{\epsilon}}(t) = \frac{1}{n} \sum_{j=1}^{n} \int_{-\infty}^{t} \frac{1}{b} k\Big(\frac{s - \hat{\epsilon}_j}{b} \Big)\, ds, \tag{4}$$

where $b = b_n$ is the bandwidth for smoothing the residuals. Finally, let

$$\hat{F}_{X,\hat{\epsilon}}(x,t) = \frac{1}{n} \sum_{j=1}^{n} G_\gamma(x, X_j) \int_{-\infty}^{t} \frac{1}{b} k\Big(\frac{s - \hat{\epsilon}_j}{b} \Big)\, ds \tag{5}$$

be an estimator of the joint distribution $F_{X,\epsilon}(x,t)$ of (X, ϵ).

Bandwidth selection proceeds via minimization of a cross-validation function, which we explain for the distribution F_X (for F_ϵ and $F_{X,\epsilon}$ similar ideas apply):

$$CV(\gamma) = \frac{1}{n n_j} \sum_{i=1}^{n} \sum_{j=1}^{n_j} \Big\{ \mathbf{1}(X_i \le x_j^e) - \hat{F}_{X,-i}(x_j^e) \Big\}^2, \tag{6}$$

where x_j^e, $j = 1, \ldots, n_j$, denotes evaluation points, and where $\hat{F}_{X,-i}(x)$ is the estimator defined in (2) except that the i-th data point is removed from the sample. The number of evaluation points can be fixed at, say, $n_j = 100$. This grid of evaluation points plays a role not unlike the number/position of points used for numerical integration. Under quite general conditions and using the cross-validated bandwidths, Li et al. (2015) obtain the result that

$$\sqrt{n}\left(\hat{F}_X(x) - F_X(x) - \frac{\kappa_2}{2}\sum_{s=1}^{q} h_s^2 \frac{\partial^2}{\partial (x_s^c)^2} F_X(x) - \sum_{s=1}^{r} \lambda_s B_s(x)\right) \xrightarrow{d} N(0, V),$$

where $V = F(x)(1 - F(x))$, $\kappa_2 = \int u^2 k(u)\, du$, and

$$B_s(x) = E_{X^d}\Big[\sum_{z^d \le x^d} \mathbf{1}(X_{-s}^d = z_{-s}^d) P(X_s^d = z_s^d) F_{X^c|X^d}(x^c|X^d)\Big], \tag{7}$$

with $F_{X^c|X^d}(x^c|x^d)$ the conditional distribution of X^c given X^d, X_{-s}^d contains all components of X^d except the s-th component, and equalities and inequalities should be understood componentwise.

Li et al. (2015) deliver a smooth nonparametric estimator that, like its non-smooth EDF counterpart, achieves a dimension-free $\sqrt{n}$ rate of convergence. The important point to note is that when the underlying distribution is itself smooth, the kernel estimator is capable of delivering estimators that outperform their non-smooth counterparts in finite-sample settings; see Li et al. (2015) for details. In a typical location-scale model with a continuous response and predictor, smoothness of the joint and marginal distributions of the predictor and error term can be safely assumed in a wide range of applications.

3. Asymptotic Properties

We start with a preliminary result that gives an iid representation for the estimators $\hat{F}_X(x)$, $\hat{F}_{\hat{\epsilon}}(t)$ and $\hat{F}_{X,\hat{\epsilon}}(x,t)$. The regularity conditions mentioned below, as well as the proofs of the results of this section, can be found in Appendix A.

Define $\kappa_2 = \int u^2 k(u)\, du$, and more generally for $p \ge 0$, let κ_{p+1} be the first element of the vector $S^{-1}(s_{p+1}, \ldots, s_{2p+1})^T$, where S is the $(p+1) \times (p+1)$ matrix whose (i,j)-th entry is s_{i+j-2}, with $s_j = \int u^j k(u)\, du$. Also, let

$$C_s(x) = \sum_{z^d} \Big[1(z_s^d \ne x_s^d) \prod_{t \ne s} 1(z_t^d = x_t^d) \mu(x^c, z^d) - \mu(x)\Big] f_X(x^c, z^d),$$

$$D_s(x) = \sum_{z^d} \Big[1(z_s^d \neq x_s^d) \prod_{t \neq s} 1(z_t^d = x_t^d) \sigma^2(x^c, z^d) - \sigma^2(x) \Big] f_X(x^c, z^d),$$

for $s = 1, \ldots, r$, where $f_X(x)$ is the joint probability density function of X.

Theorem 1. *Assume (A1)–(A6). Then, under H_0 and for any x and t,*

$$\begin{aligned}
\hat{F}_X(x) - F_X(x) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \leq x) - F_X(x) + \frac{\kappa_2}{2} \sum_{s=1}^{q} h_s^2 \frac{\partial^2 F_X(x)}{\partial (x_s^c)^2} + \sum_{s=1}^{r} \lambda_s B_s(x) + R_{n,X}(x), \\
\hat{F}_{\hat{\epsilon}}(t) - F_\epsilon(t) &= \frac{1}{n} \sum_{i=1}^{n} \Big[1(\epsilon_i \leq t) - F_\epsilon(t) + f_\epsilon(t) \Big\{ \epsilon_i + \frac{t}{2}(\epsilon_i^2 - 1) \Big\} \Big] \\
&\quad + f_\epsilon(t) \int \frac{1}{\sigma(z)} \Big\{ \frac{\kappa_{p+1}}{(p+1)!} \sum_{s=1}^{q} g_s^{p+1} \frac{\partial^{p+1}}{\partial (z_s^c)^{p+1}} \mu(z) + \sum_{s=1}^{r} \nu_s C_s(z) \Big\} dF_X(z) \\
&\quad + \frac{t}{2} f_\epsilon(t) \int \frac{1}{\sigma^2(z)} \Big\{ \frac{\kappa_{p+1}}{(p+1)!} \sum_{s=1}^{q} g_s^{p+1} \frac{\partial^{p+1}}{\partial (z_s^c)^{p+1}} \sigma^2(z) + \sum_{s=1}^{r} \nu_s D_s(z) \Big\} dF_X(z) \\
&\quad + \frac{\kappa_2}{2} b^2 f_\epsilon'(t) + R_{n,\epsilon}(t), \\
\hat{F}_{X,\hat{\epsilon}}(x,t) - F_{X,\epsilon}(x,t) &= \frac{1}{n} \sum_{i=1}^{n} \Big[\mathbf{1}(X_i \leq x, \epsilon_i \leq t) - F_{X,\epsilon}(x,t) + f_\epsilon(t) \mathbf{1}(X_i \leq x) \Big\{ \epsilon_i + \frac{t}{2}(\epsilon_i^2 - 1) \Big\} \Big] \\
&\quad + f_\epsilon(t) \int_{-\infty}^{x} \frac{1}{\sigma(z)} \Big\{ \frac{\kappa_{p+1}}{(p+1)!} \sum_{s=1}^{q} g_s^{p+1} \frac{\partial^{p+1}}{\partial (z_s^c)^{p+1}} \mu(z) + \sum_{s=1}^{r} \nu_s C_s(z) \Big\} dF_X(z) \\
&\quad + \frac{t}{2} f_\epsilon(t) \int_{-\infty}^{x} \frac{1}{\sigma^2(z)} \Big\{ \frac{\kappa_{p+1}}{(p+1)!} \sum_{s=1}^{q} g_s^{p+1} \frac{\partial^{p+1}}{\partial (z_s^c)^{p+1}} \sigma^2(z) + \sum_{s=1}^{r} \nu_s D_s(z) \Big\} dF_X(z) \\
&\quad + \frac{\kappa_2}{2} \sum_{s=1}^{q} h_s^2 \frac{\partial^2 F_X(x)}{\partial (x_s^c)^2} F_\epsilon(t) + \sum_{s=1}^{r} \lambda_s B_s(x) F_\epsilon(t) + \frac{\kappa_2}{2} b^2 F_X(x) f_\epsilon'(t) \\
&\quad + R_{n,X,\epsilon}(x,t),
\end{aligned}$$

where $\sup_{x \in R_X} |R_{n,X}(x)| = o_P(n^{-1/2})$, $\sup_{t \in \mathcal{R}} |R_{n,\epsilon}(t)| = o_P(n^{-1/2})$ *and* $\sup_{x \in R_X, t \in \mathcal{R}} |R_{n,X,\epsilon}(x,t)| = o_P(n^{-1/2})$.

An immediate consequence of this theorem is the following corollary.

Corollary 2. *Assume (A1)–(A6). Then, under H_0 and for any x and t,*

$$\sqrt{n}\big(\hat{F}_{X,\hat{\epsilon}}(x,t) - \hat{F}_X(x)\hat{F}_{\hat{\epsilon}}(t)\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H(X_i, \epsilon_i, x, t) + b(x,t) + R_n(x,t),$$

where

$$H(X_i,\epsilon_i,x,t) = \mathbf{1}(X_i\le x,\epsilon_i\le t) - F_{X,\epsilon}(x,t) - F_\epsilon(t)\big\{\mathbf{1}(X_i\le x) - F_X(x)\big\} \\ - F_X(x)\big\{1(\epsilon_i\le t) - F_\epsilon(t)\big\} + f_\epsilon(t)\big\{\mathbf{1}(X_i\le x) - F_X(x)\big\}\Big\{\epsilon_i + \frac{t}{2}(\epsilon_i^2-1)\Big\},$$

$$b(x,t) = f_\epsilon(t)\int \big(\mathbf{1}(z\le x) - F_X(x)\big)\frac{1}{\sigma(z)}\Big\{\frac{\kappa_{p+1}}{(p+1)!}\sum_{s=1}^{q} c_{g,s}^{p+1}\frac{\partial^{p+1}}{\partial(z_s^c)^{p+1}}\mu(z) + \sum_{s=1}^{r} c_{\nu,s}C_s(z)\Big\}dF_X(z) \\ + \frac{t}{2}f_\epsilon(t)\int \big(\mathbf{1}(z\le x) - F_X(x)\big)\frac{1}{\sigma^2(z)}\Big\{\frac{\kappa_{p+1}}{(p+1)!}\sum_{s=1}^{q} c_{g,s}^{p+1}\frac{\partial^{p+1}}{\partial(z_s^c)^{p+1}}\sigma^2(z) + \sum_{s=1}^{r} c_{\nu,s}D_s(z)\Big\}dF_X(z),$$

and $\sup_{x\in R_X,t\in\mathcal{R}}|R_n(x,t)| = o_P(1)$.

Remark 3. *Note that some of the bias terms appearing in Theorem 1 cancel out in Corollary 2. Indeed, the biases coming from making the distribution functions $\hat F_X(x)$, $\hat F_{\hat\epsilon}(t)$ and $\hat F_{X,\hat\epsilon}(x,t)$ smooth, disappeared in Corollary 2. Hence, the asymptotic representation of $\sqrt{n}\big(\hat F_{X,\hat\epsilon}(x,t) - \hat F_X(x)\hat F_{\hat\epsilon}(t)\big)$ is the same as in the case where the empirical distributions are not smoothed (see Einmahl & Van Keilegom (2008)). However, inspection of the proof of Theorem 2.2 in the latter paper reveals that their iid expansion of $\sqrt{n}\big(\hat F_{X,\hat\epsilon}(x,t) - \hat F_X(x)\hat F_{\hat\epsilon}(t)\big)$ does not contain the term $f_\epsilon(t)\,n^{-1/2}\sum_{i=1}^n\big\{\mathbf{1}(X_i\le x) - F_X(x)\big\}\big\{\epsilon_i + \frac{t}{2}(\epsilon_i^2-1)\big\}$ that shows up in the formula of $n^{-1/2}\sum_{i=1}^n H(X_i,\epsilon_i,x,t)$ given above. This is because the second statement in their Lemma A.1 is wrong, in the sense that the expression on the left hand side is not centered, and so it is certainly not $o_P(n^{-1/2})$. The correct version of their Theorem 2.2 can be obtained from Corollary 4 below by using undersmoothing of the bandwidth g_1 (and taking $q=1$ and $r=0$).*

We are now ready to state the weak convergence of $\sqrt{n}\big(\hat F_{X,\hat\epsilon} - \hat F_X\hat F_{\hat\epsilon}\big)$ as a process in $\ell^\infty(R_X\times\mathcal{R})$, and the limiting distribution of our test statistics T_{KS} and T_{CM}. Here, $\ell^\infty(R_X\times\mathcal{R})$ is the set of bounded functions from $R_X\times\mathcal{R}$ to $\mathcal{R}$, equipped with the uniform norm.

Corollary 4. *Assume (A1)–(A6).*

(i) *Under H_0, the process $\sqrt{n}\big(\hat F_{X,\hat\epsilon}(x,t) - \hat F_X(x)\hat F_{\hat\epsilon}(t)\big)$ converges weakly in $\ell^\infty(R_X\times\mathcal{R})$ to a Gaussian process $Z(x,t)$ with mean function $b(x,t)$ and covariance function*

$$Cov\big(Z(x_1,t_1),Z(x_2,t_2)\big) = E\big[H(X,\epsilon,x_1,t_1)H(X,\epsilon,x_2,t_2)\big]$$

for $x_1, x_2 \in R_X$ *and* $t_1, t_2 \in \mathcal{R}$.

(ii) *Under* H_0,

$$T_{KS} \xrightarrow{d} \sup_{x \in R_X, t \in \mathcal{R}} |Z(x,t)| \quad \textit{and} \quad T_{CM} \xrightarrow{d} \int \int Z^2(x,t)\, dF_X(x)\, dF_\epsilon(t).$$

4. Finite-Sample Performance

4.1. **The Univariate Continuous Predictor Setting.** In order to assess the finite-sample performance of our proposed approach, we replace the EDFs in (1) with their kernel-smoothed counterparts described in Section 2.1.[3] Bandwidth selection is obtained via cross-validation for h, b and λ (i.e., minimization of Equation (6) above) and via least squares cross-validation for g and ν, and the Epanechnikov kernel function is employed. There are various data-driven permutations one might consider for bandwidth selection, namely a) separate bandwidths for the joint and marginal distributions, b) common bandwidths taken from the joint distribution, and c) common bandwidths taken from the marginal distributions. From the perspective of assessing size, one might expect that the same bandwidths used for estimating the joint distribution be used for the construction of the marginal distributions, otherwise the estimated joint distribution might systematically diverge from the product of the estimated marginals under the null. We investigate this issue empirically and on this basis recommend b) to practitioners.

To assess finite-sample performance, we simulate data for which X is uniform $[0, 1]$ and Y has location $\mu(x) = \sin(2\pi x)$ and scale $\sigma(x)$ that is determined from the error distributions specified below, i.e.,

$$Y_i = \sin(2\pi X_i) + \sigma(X_i)\epsilon_i(X_i).$$

We consider three DGPs in the simulations that follow. For the first the errors are mixtures of two Gaussians, $N(-1, 0.5^2)$ and $N(1, 0.5^2)$ with mixing probabilities $1 - \delta x$ and δx. For the second the errors are drawn from a heavy-tail mixture of two t-distributions, one having a mode at 2 and the other at -2, both having 5 degrees of freedom with the same mixing probabilities as for the Gaussian mixture. For the third the errors are drawn from the Beta distribution with shape

[3]We proceed with the Kolmogorov-Smirnov statistic by way of illustration as the Cramér-von-Mises statistic requires multivariate integration for its computation while the Kolmogorov-Smirnov statistic does not. However, as will be demonstrated for the Kolmogorov-Smirnov approach, replacing the non-smooth distribution functions in the Cramér-von-Mises statistic with their kernel-smoothed counterparts would be expected to reveal similar power gains.

parameters $s_1 = 1 + \delta 10x$ and $s_2 = 11 - \delta 10x$ where $x \in [0,1]$. These errors are then rescaled to have unconditional mean zero and unit variance thereby maintaining a constant signal-to-noise ratio across DGPs and across the range of values for δ considered. In all cases, when $\delta = 0$ the model is a location-scale DGP (i.e. the distribution of ϵ is not a function of x) while when $\delta > 0$ it is a location-shape DGP (i.e. the distribution of ϵ is a function of x). Figure 1 presents the density of $\epsilon(x)$ for various levels of $x \in [0,1]$ when $\delta = 1$.

For what follows we consider $M = 1,000$ Monte Carlo replications drawn from each DGP. For each Monte Carlo replication, we compute each test statistic T_{KS} (non-smooth and smooth, respectively) along with the $B = 399$ (Davidson & MacKinnon 2000) null bootstrap replicates $T^*_{b,KS}$, $b = 1, \ldots, B$ (based on resamples drawn from the (non-smooth) distribution functions), then we compute the empirical P-value as $B^{-1} \sum_{b=1}^{B} \mathbf{1}(T^*_{b,KS} > T_{KS})$, where $\mathbf{1}(A)$ is the usual indicator function taking value one when A is true and zero otherwise. Finally, based on the $M = 1,000$ P-values, we compute empirical rejection probabilities for the non-smooth and smooth test statistics for nominal levels $\alpha = (0.01, 0.05, 0.10)$. To assess size we set $\delta = 0$, and to assess power we let $\delta \in (0,1]$. `R` code to replicate these simulations can be found in Appendix D, and size (i.e. empirical rejection probability when $\delta = 0$) is summarized in Table 1.

Table 1 indicates that both the non-smooth and smooth versions of the test appear to be correctly sized, hence we can proceed to compare their power curves.[4] Next, we vary $\delta \in [0,1]$, and present power curves for the Beta errors in Figure 4 (Figures 2 and 3 present results for the Gaussian and Student-t mixtures).

The percentage gain in power is reported in the tables in Appendix B, and these figures and tables reveal that the improvements in power arising from replacing the EDF with its kernel-smoothed counterpart can be upwards of 100% or more, depending on the nominal size of the test, sample size, and degree of departure from the null. Furthermore, if anything the smooth test appears to be slightly conservative relative to its non-smooth counterpart, particularly for the Beta error case for the smaller sample sizes considered (a positive feature as it has a lower probability of a Type I error than its non-smooth counterpart under the null yet higher power under the alternative).

[4]If anything, the non-smooth version appears to be slightly over-sized for smaller n and the smooth version slightly under-sized for smaller n, but this admits a fair comparison of power curves.

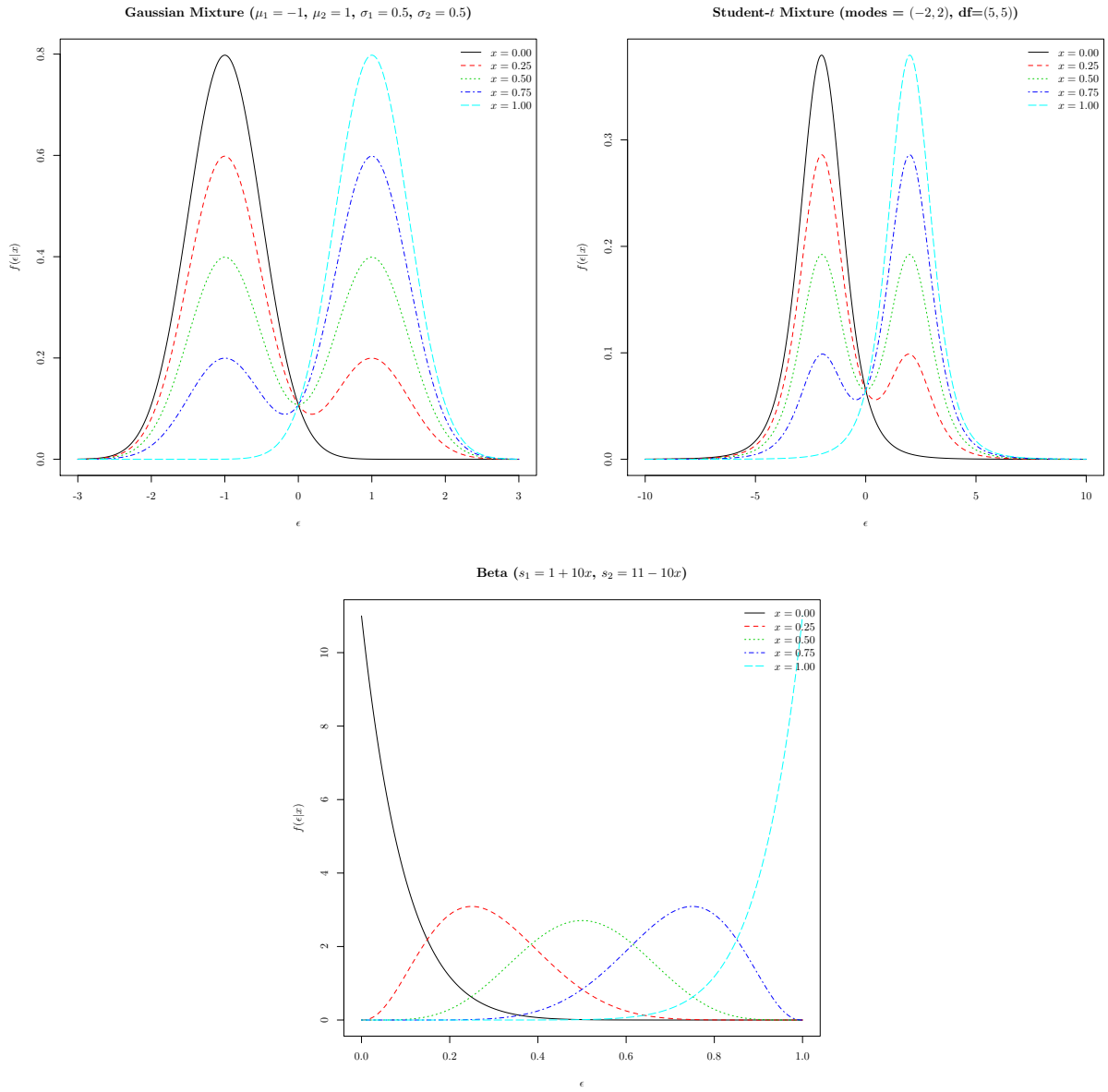

FIGURE 1. Density of $\epsilon(x)$ for various levels of $x \in [0, 1]$, $\delta = 1$, for the Gaussian mixture, Student-t mixture and Beta (the error density for computing size, $\delta = 0$, corresponds to the density when $x = 0$ i.e. the solid black density).

4.2. **The Multivariate Continuous Predictor Setting.** To assess finite-sample performance in the multivariate continuous predictor setting, we simulate data for which X_1 and X_2 are uniform $[0, 1]$ and Y has location $\mu(x) = \sin(\pi(x_1 + x_2))$ and scale $\sigma(x_1 + x_2)/2$ that is determined from

TABLE 1. Size for the Non-smooth and Smooth Tests (Empirical Rejection Probabilities Under the Null, $\delta = 0$, Univariate Continuous Predictor Setting, see Appendix B for Power).

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
Non-smooth, Gaussian Mixture			
100	0.008	0.041	0.101
200	0.008	0.045	0.089
400	0.006	0.048	0.089
800	0.006	0.045	0.102
Smooth, Gaussian Mixture			
100	0.011	0.045	0.104
200	0.006	0.046	0.093
400	0.007	0.057	0.110
800	0.010	0.047	0.092
Non-smooth, Student-t Mixture			
100	0.009	0.051	0.081
200	0.009	0.050	0.104
400	0.005	0.039	0.095
800	0.006	0.038	0.089
Smooth, Student-t Mixture			
100	0.009	0.048	0.101
200	0.010	0.052	0.095
400	0.010	0.045	0.092
800	0.008	0.042	0.095
Non-smooth, Beta			
100	0.012	0.049	0.090
200	0.008	0.050	0.105
400	0.008	0.038	0.099
800	0.011	0.048	0.098
Smooth, Beta			
100	0.011	0.042	0.074
200	0.012	0.051	0.095
400	0.008	0.054	0.110
800	0.013	0.054	0.104

the error distributions specified below, i.e.,

$$Y_i = \sin(\pi(X_{i1} + X_{i2})) + \sigma(X_{i1} + X_{i2})\epsilon_i(X_{i1} + X_{i2}).$$

We consider three DGPs in the simulations that follow. For the first the errors are mixtures of two Gaussians, $N(-1, 0.5^2)$ and $N(1, 0.5^2)$ with mixing probabilities $1 - \delta(x_1 + x_2)/2$ and $\delta(x_1 + x_2)/2$. For the second the errors are drawn from a heavy-tail mixture of two t-distributions, one having a mode at 2 and the other at -2, both having 5 degrees of freedom, with the same mixing probabilities

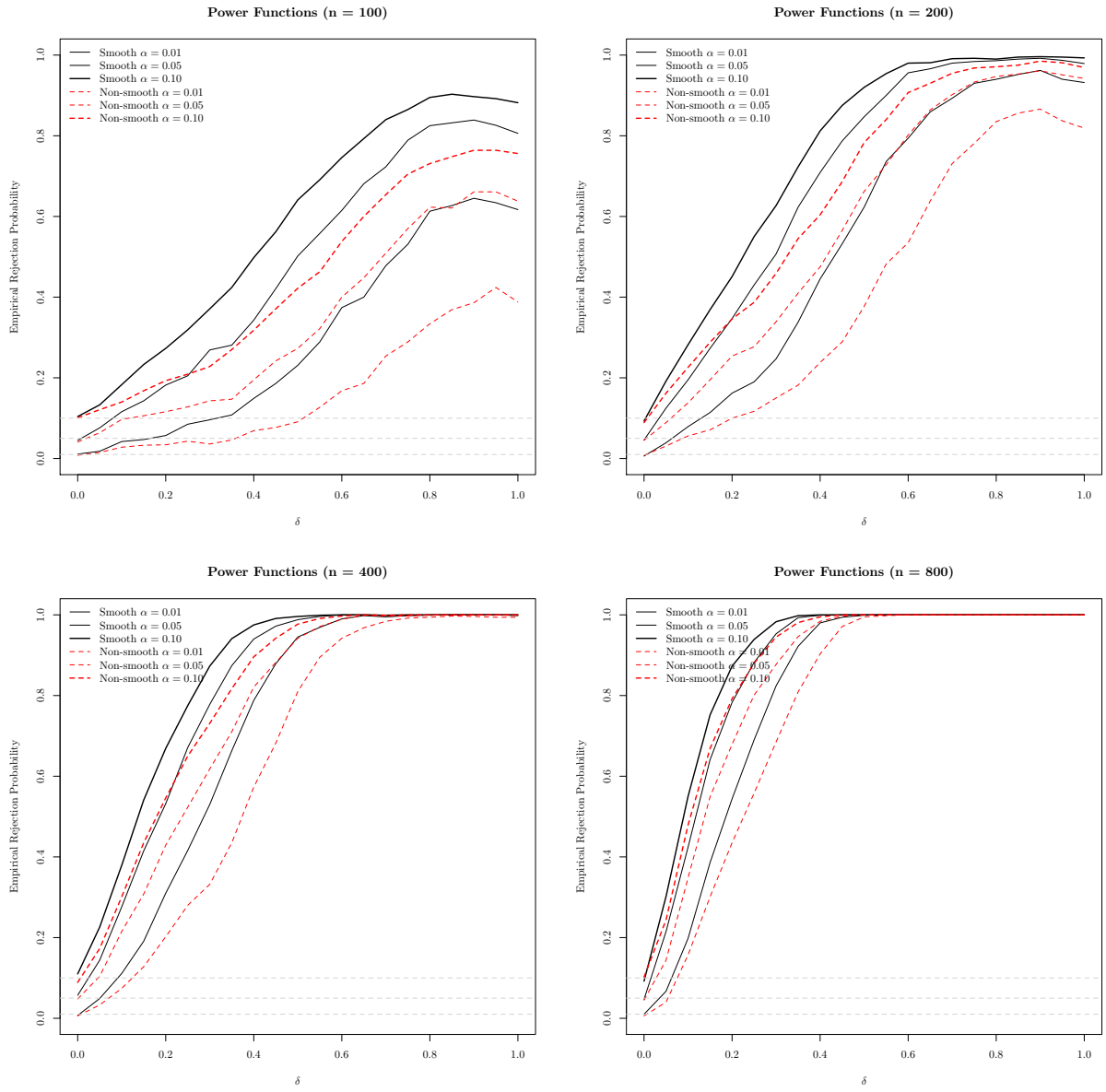

FIGURE 2. Power curves for the non-smooth and smooth versions of the test, Gaussian mixture. Solid lines are for the smooth version, dashed the non-smooth version.

as for the Gaussian mixture. For the third the errors are drawn from the Beta distribution with shape parameters $s_1 = 1 + \delta 5(x_1 + x_2)$ and $s_2 = 11 - \delta 5(x_1 + x_2)$. Per above, these errors are then rescaled to have unconditional mean zero and unit variance.

Table 2 reveals that both the non-smooth and smooth version of the test perform adequately and appear to possess reasonable size when there exist multivariate continuous predictors.

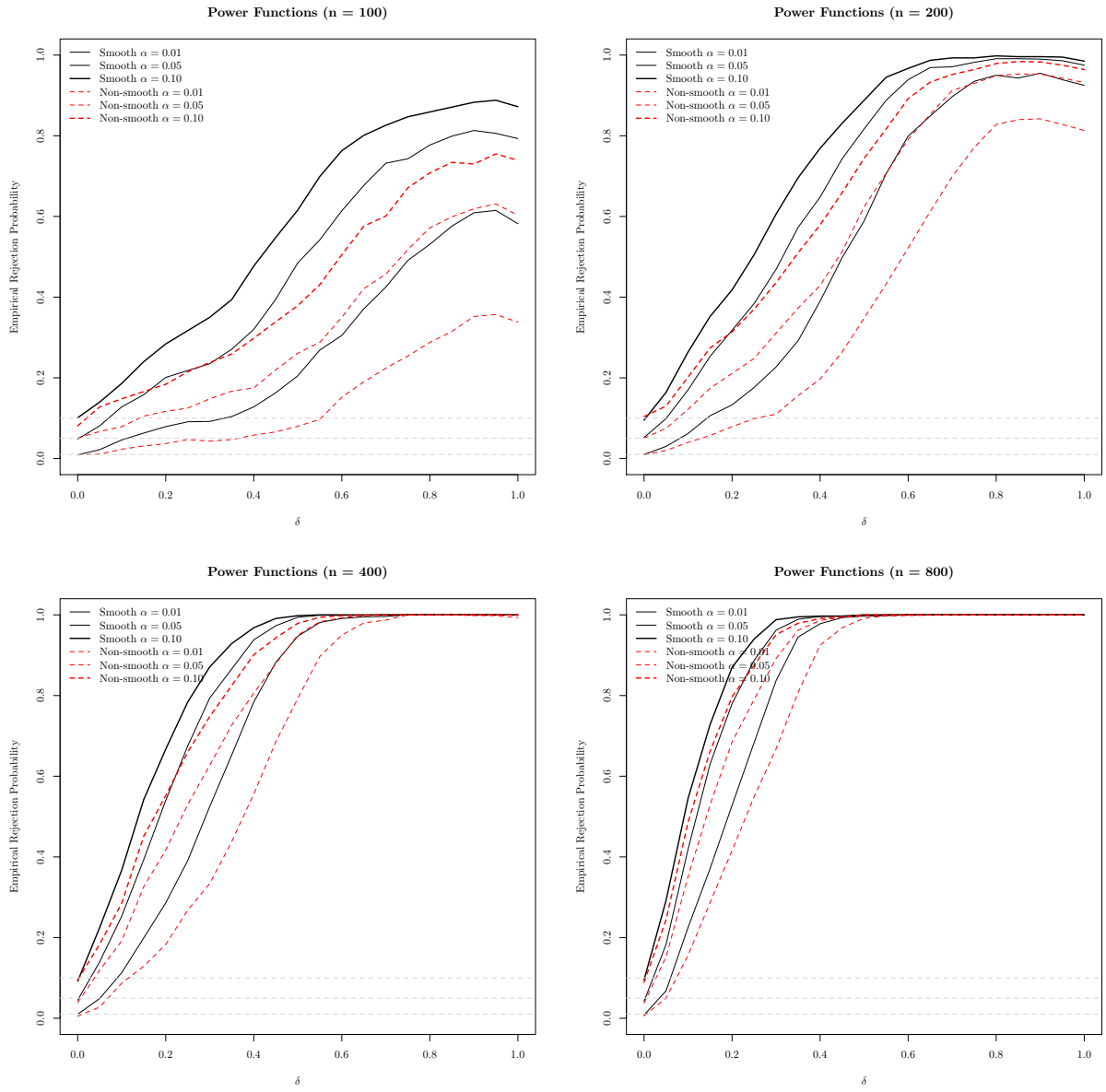

FIGURE 3. Power curves for the non-smooth and smooth versions of the test, Student-t mixture. Solid lines are for the smooth version, dashed the non-smooth version.

4.3. **The Multivariate Mixed Continuous and Discrete Predictor Setting.** To assess finite-sample performance in the multivariate mixed continuous and discrete predictor setting, we simulate data for which X_1 is uniform $[0, 1]$, X_2 is a discrete uniform variable taking value $0, 1/10, 2/10, \ldots, 1$, Y has location $\mu(x) = \sin(2\pi x_1) + x_2$ and scale $\sigma(X_i)$ that is determined from the error distributions

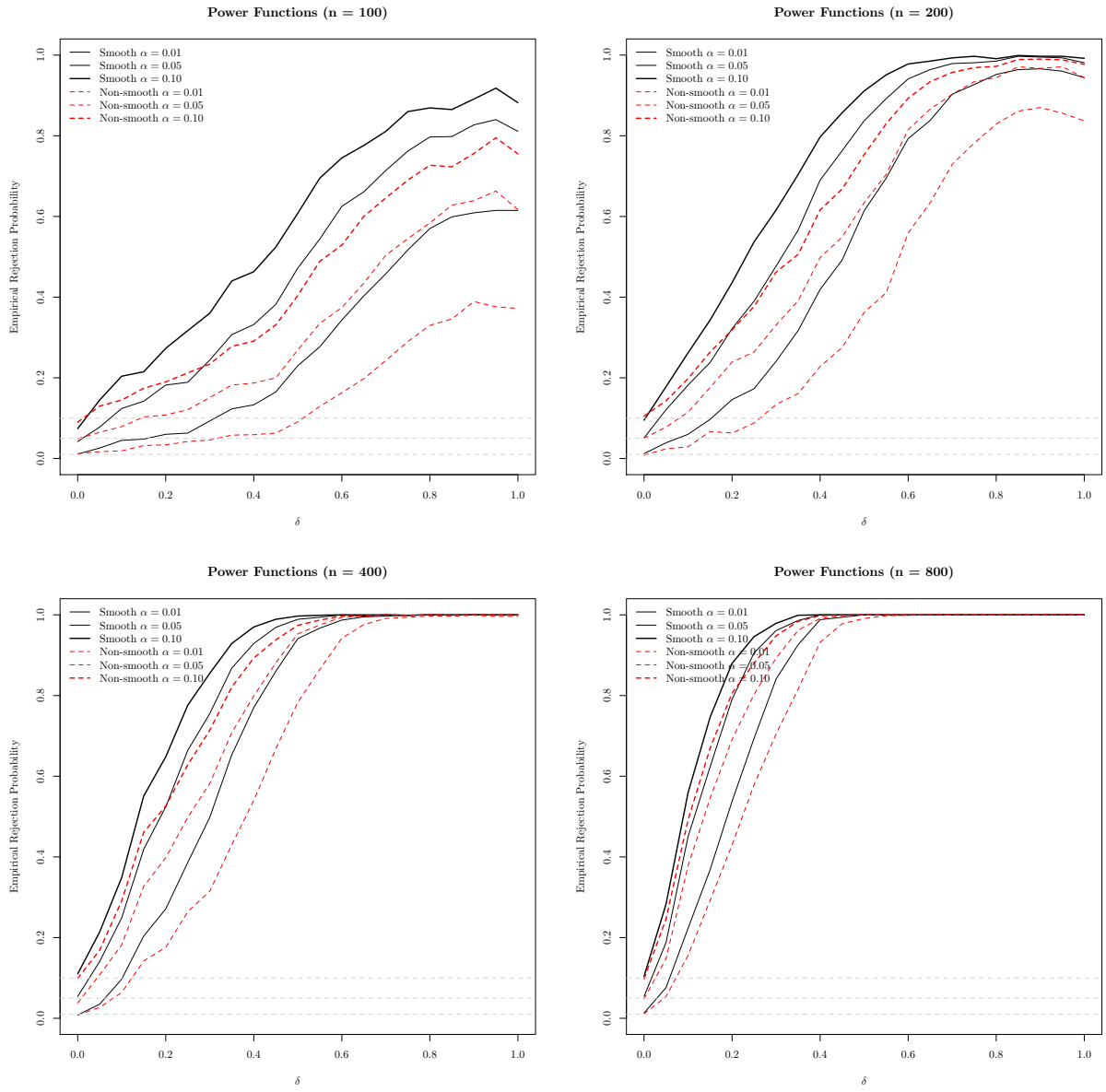

FIGURE 4. Power curves for the non-smooth and smooth versions of the test, Beta errors. Solid lines are for the smooth version, dashed the non-smooth version.

specified below, i.e.,

$$Y_i = \sin(2\pi X_{i1}) + X_{i2} + \sigma(X_{i1} + X_{i2})\epsilon_i(X_{i1} + X_{i2}).$$

We consider three DGPs in the simulations that follow. For the first the errors are mixtures of two Gaussians, $N(-1, 0.5^2)$ and $N(1, 0.5^2)$ with mixing probabilities $1 - \delta(x_1 + x_2)/2$ and $\delta(x_1 + x_2)/2$.

TABLE 2. Size for the Non-smooth and Smooth Tests (Empirical Rejection Probabilities Under the Null, $\delta = 0$, Multivariate Continuous Predictor Setting, see Appendix C for Power).

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
Non-smooth, Gaussian Mixture			
100	0.007	0.042	0.100
200	0.007	0.042	0.094
400	0.012	0.038	0.075
800	0.005	0.037	0.075
Smooth, Gaussian Mixture			
100	0.005	0.033	0.091
200	0.008	0.048	0.087
400	0.006	0.032	0.072
800	0.006	0.037	0.073
Non-smooth, Student-t Mixture			
100	0.018	0.056	0.116
200	0.018	0.080	0.121
400	0.012	0.053	0.107
800	0.008	0.055	0.095
Smooth, Student-t Mixture			
100	0.023	0.078	0.140
200	0.016	0.085	0.140
400	0.017	0.068	0.111
800	0.012	0.057	0.109
Non-smooth, Beta			
100	0.005	0.042	0.100
200	0.021	0.074	0.118
400	0.018	0.062	0.124
800	0.019	0.064	0.108
Smooth, Beta			
100	0.011	0.057	0.104
200	0.015	0.059	0.121
400	0.020	0.070	0.127
800	0.018	0.060	0.112

For the second the errors are drawn from a heavy-tail mixture of two t-distributions, one having a mode at 2 and the other at -2, both having 5 degrees of freedom, with the same mixing probabilities as for the Gaussian mixture. For the third the errors are drawn from the Beta distribution with shape parameters $s_1 = 1 + \delta 5(x_1 + x_2)$ and $s_2 = 11 - \delta 5(x_1 + x_2)$. Per above, these errors are then rescaled to have unconditional mean zero and unit variance.

Table 3 reveals a curious feature of the non-smooth test – in the presence of discrete predictors, the test displays extreme size distortions rendering it completely unsuited for practical application

TABLE 3. Size for the Non-smooth and Smooth Tests (Empirical Rejection Probabilities Under the Null, $\delta = 0$, Multivariate Mixed Continuous and Discrete Predictor Setting).

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
Non-smooth, Gaussian Mixture			
100	0.285	0.545	0.678
200	0.625	0.836	0.918
400	0.950	0.994	0.998
800	1.000	1.000	1.000
Smooth, Gaussian Mixture			
100	0.009	0.046	0.090
200	0.013	0.056	0.100
400	0.015	0.058	0.116
800	0.024	0.081	0.159
Non-smooth, Student-t Mixture			
100	0.361	0.632	0.752
200	0.773	0.933	0.974
400	0.993	1.000	1.000
800	1.000	1.000	1.000
Smooth, Student-t Mixture			
100	0.014	0.054	0.096
200	0.013	0.061	0.108
400	0.014	0.063	0.125
800	0.012	0.074	0.140
Non-smooth, Beta			
100	0.518	0.782	0.877
200	0.951	0.991	0.998
400	1.000	1.000	1.000
800	1.000	1.000	1.000
Smooth, Beta			
100	0.006	0.027	0.074
200	0.004	0.040	0.084
400	0.012	0.051	0.124
800	0.023	0.110	0.200

(its empirical rejection probability under the null approaches 1 as n increases for all conventional levels). However, the smooth version of the test appears to be reasonably sized. The former is perhaps not too surprising given the literature on discrete/discontinuous distributions (Conover 1972, Gleser 1985, Choulakian et al. 1994, Lockhart et al. 2007). Given the extreme size distortions present, we make no attempt at power comparisons.

5. Application

Li & Racine (forthcoming) impose a location-scale quantile model structure on a novel nonparametric quantile estimator that is based on kernel smoothing of a parametric quantile function in a particular manner. A practitioner concerned with their imposition of the location-scale structure might wish to use a pre-test approach, proceeding with the location-scale model if it is deemed appropriate versus an alternative model that does not rely on the location-scale assumption otherwise. They present two illustrative applications, one in which the covariate is continuous and one in which it is discrete, so these illustrations will serve to highlight the potential application of the proposed procedure.

We first consider an Italian gross domestic product (GDP) growth panel for 21 regions covering the period 1951-1998 (millions of Lire, 1990=base). There are $n = 1,008$ observations on 2 variables, 'year' and 'gdp', and 'year' is a discrete predictor and is treated as such in what follows. Next we consider Canadian cross-section wage data consisting of a random sample taken from the 1971 Canadian Census Public Use Tapes for male individuals having common education (grade 13). There are $n = 205$ observations in total on two variables, 'logwage' (logarithm of the wages) and 'age', age being treated as a continuous predictor. We report the test statistics and their bootstrapped P-values in Table 4 based on $B = 999$ bootstrap replications.

Table 4. Application of the Proposed Test to the Italian GDP Dataset and to the Canadian Cross-Section Wage Dataset.

Dataset	T_{KS}	P-value
Canadian Wage	0.4637124	0.3413413
Italian GDP	1.174512	$< 2e-16$

Table 4 reveals that the location-scale presumption is appropriate for the Italian GDP data, but is inappropriate for the Canadian wage data. These results indicate that practitioners can use Li & Racine's (forthcoming) quantile approach for the former but ought to exercise caution when applying it to the latter.

6. Concluding Remarks

Numerous tests have been proposed that rely on the non-smooth empirical distribution function for their implementation; see Einmahl & Van Keilegom (2008), Birke et al. (2017) and Neumeyer

et al. (2016) by way of illustration. We demonstrate how the use of smooth estimators of distribution functions rather than their non-smooth counterparts can deliver tests having superior power profiles, which ought to be particularly appealing for practitioners.

References

Aitchison, J. & Aitken, C. G. G. (1976), 'Multivariate binary discrimination by the kernel method', *Biometrika* **63**, 413–420.

Akritas, M. & Van Keilegom, I. (2001), 'Nonparametric estimation of the residual distribution', *Scandinavian Journal of Statistics* **28**, 549–568.

Anderson, T. W. & Darling, D. A. (1952), 'Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes', *The Annals of Mathematical Statistics* **23**(2), 193–212.

Billingsley, P. (1999), *Convergence of Probability Measures*, Wiley & Sons.

Birke, M., Neumeyer, N. & Volgushev, S. (2017), 'The independence process in conditional quantile location-scale models and an application to testing for monotonicity', *Statistica Sinica* .

Bowman, A. W., Hall, P. & Prvan, T. (1998), 'Bandwidth selection for the smoothing of distribution functions', *Biometrika* **85**, 799–808.

Choulakian, V., Lockhart, R. A. & Stephens, M. A. (1994), 'Cramer-von Mises statistics for discrete distributions', *The Canadian Journal of Statistics* **22**(1), 125–137.

Conover, W. J. (1972), 'A Kolmogorov goodness-of-fit test for discontinuous distributions', *Journal of the American Statistical Association* **67**(339), 591–596.

Conover, W. J. (1999), *Practical Nonparametric Statistics*, third edn, Wiley.

Cramér, H. (1928), 'On the composition of elementary errors', *Scandinavian Actuarial Journal* **1**, 13–74.

Davidson, R. & MacKinnon, J. G. (2000), 'Bootstrap tests: How many bootstraps?', *Econometric Reviews* **19**, 55–68.

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

Escanciano, J. & Jacho-Chávez, D. T. (2012), '$\sqrt{n}$-uniformly consistent density estimation in nonparametric regression models', *Journal of Econometrics* **12**, 305–361.

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

Gleser, L. J. (1985), 'Exact power of goodness-of-fit tests of Kolmogorov type for discontinuous distributions', *Journal of the American Statistical Association* **80**(392), 954–958.

Kolmogorov, A. (1933), 'Sulla determinazione empirica di una legge di distributione', *Giornale dell'Istituto Italiano degli Attuari* **4**, 1–11.

Li, C., Li, H. & Racine, J. (2015), Cross-validated mixed datatype bandwidth selection for nonparametric cumulative distribution/survivor functions, Technical report, McMaster University.

Li, K. & Racine, J. S. (forthcoming), 'Nonparametric conditional quantile estimation: A locally weighted quantile kernel approach', *Journal of Econometrics* .

Li, Q. & Racine, J. S. (2004), 'Cross-validated local linear nonparametric regression', *Statistica Sinica* **14**, 485–512.

Lockhart, R., Spinelli, J. & Stephens, M. (2007), 'Cramér-von Mises statistics for discrete distributions with unknown parameters', *Canadian Journal of Statistics* **35**(1), 125–133.

Neumeyer, N. (2009), 'Smooth residual bootstrap for empirical processes of non-parametric regression residuals', *Scandinavian Journal of Statistics* **36**, 204–228.

Neumeyer, N., Noh, H. & Van Keilegom, I. (2016), 'Heteroscedastic semiparametric transformation models: estimation and testing for validity', *Statistica Sinica* **26**, 925–954.

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

Qiu, D. (2014), *snpar: Supplementary Non-parametric Statistics Methods*. R package version 1.0.
URL: *https://CRAN.R-project.org/package=snpar*

R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Ruppert, D. & Wand, M. P. (1994), 'Multivariate locally weighted least squares regression', *Annals of Statistics* **22**, 1346–1370.

Smirnov, N. (1948), 'Table for estimating the goodness of fit of empirical distributions', *Annals of Mathematical Statistics* **19**, 279–281.

Stute, W. (1982), 'The oscillation behavior of empirical processes', *The Annals of Probability* **10**(1), 86–107.

Van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer.

von Mises, R. E. (1928), *Wahrscheinlichkeit, Statistik und Wahrheit*, Julius Springer.

Wang, J., Cheng, F. & Yang, L. (2013), 'Smooth simultaneous confidence bands for cumulative distribution functions', *Journal of Nonparametric Statistics* **25**, 395–407.

Appendix A. Proofs of the main results

The results of Section 3 are valid under the following regularity conditions.

(A1) k is a symmetric probability density function supported on $[-a, a]$, k is q times continuously differentiable, and $k^{(j)}(\pm a) = 0$ for $j = 0, \ldots, q-1$.

(A2) As $n \to \infty$, $ng_s^{2p+2} \to c_{g,s}^{2p+2}$ for some $0 \leq c_{g,s} < \infty$ $(s = 1, \ldots, q)$, and $n\nu_s^2 \to c_{\nu,s}^2$ for some $0 \leq c_{\nu,s} < \infty$ $(s = 1, \ldots, r)$. Moreover, p is odd and is such that $2p + 2 > 3q$.

(A3) As $n \to \infty$, $nh_s^4 \to c_{h,s}^4$ for some $0 \leq c_{h,s} < \infty$ $(s = 1, \ldots, q)$, $n\lambda_s^2 \to c_{\lambda,s}^2$ for some $0 \leq c_{\lambda,s} < \infty$ $(s = 1, \ldots, r)$, and $nb^4 \to c_b^4$ for some $0 \leq c_b < \infty$.

(A4) All partial derivatives of $F_X(\cdot, x^d)$ up to order $2q+1$ exist on the interior of R_{X^c} (for fixed $x^d \in R_{X^d}$), they are uniformly continuous and $\inf_{x \in R_X} f_X(x) > 0$.

(A5) All partial derivatives of $\mu(\cdot, x^d)$ and $\sigma(\cdot, x^d)$ up to order $p+2$ exist on the interior of R_{X^c} (for fixed $x^d \in R_{X^d}$), they are uniformly continuous and $\inf_{x \in R_X} \sigma(x) > 0$.

(A6) F_ϵ is three times continuously differentiable, $\sup_t |t^2 f_\epsilon'(t)| < \infty$, and $E(\epsilon^6) < \infty$.

Proof of Theorem 1. For $\hat{F}_X(x)$, we refer to Lemma 3.1 in Li et al. (2015). For $\hat{F}_{\hat{\epsilon}}(t)$, note that

$$\hat{F}_{\hat{\epsilon}}(t) = \int_{-\infty}^{\infty} \tilde{F}_{\hat{\epsilon}}(t - vb) k(v)\, dv,$$

where $\tilde{F}_{\hat{\epsilon}}(t) = n^{-1} \sum_{i=1}^n 1(\hat{\epsilon}_i \leq t)$ is the non-smoothed estimator and hence

$$\hat{F}_{\hat{\epsilon}}(t) - F_\epsilon(t) = \int \big[\tilde{F}_{\hat{\epsilon}}(t - vb) - F_\epsilon(t - vb)\big] k(v)\, dv + \int \big[F_\epsilon(t - vb) - F_\epsilon(t)\big] k(v)\, dv.$$

The second term above equals $(1/2)\kappa_2 b^2 f_\epsilon'(t) + o(b^2)$, whereas it follows from the proof of Theorem 2.1 in Neumeyer & Van Keilegom (2010) that the first term equals

$$\begin{aligned}
&n^{-1} \sum_{i=1}^n \int \Big[1(\epsilon_i \leq t - vb) - F_\epsilon(t - vb) + f_\epsilon(t - vb)\Big\{\epsilon_i + \frac{E\hat{\mu}(X_i) - \mu(X_i)}{\sigma(X_i)}\Big\} \\
&\quad + \frac{t - vb}{2} f_\epsilon(t - vb)\Big\{\epsilon_i^2 - 1 + \frac{E\hat{\sigma}^2(X_i) - \sigma^2(X_i)}{\sigma^2(X_i)}\Big\}\Big] k(v)\, dv + o_P(n^{-1/2}) \\
&= n^{-1} \sum_{i=1}^n \Big[1(\epsilon_i \leq t) - F_\epsilon(t) + f_\epsilon(t)\Big\{\epsilon_i + \frac{t}{2}(\epsilon_i^2 - 1)\Big\}\Big] \\
&\quad + f_\epsilon(t) \int \frac{E\hat{\mu}(z) - \mu(z)}{\sigma(z)} dF_X(z) + \frac{t}{2} f_\epsilon(t) \int \frac{E\hat{\sigma}^2(z) - \sigma^2(z)}{\sigma^2(z)} dF_X(z) + o_P(n^{-1/2}) \qquad (8)
\end{aligned}$$

by the smoothness of $f_\epsilon(t)$ and the modulus of continuity of the empirical distribution function given in Theorem 2.14 in Stute (1982). Using the asymptotic expressions of the bias of $\hat{\mu}(z)$ and $\hat{\sigma}(z)$ given in Section 3 in Li & Racine (2004), we find that (8) equals

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\Big[1(\epsilon_i \le t) - F_\epsilon(t) + f_\epsilon(t)\Big\{\epsilon_i + \frac{t}{2}(\epsilon_i^2 - 1)\Big\}\Big] \\
&+ f_\epsilon(t)\int \frac{1}{\sigma(z)}\Big\{\frac{\kappa_{p+1}}{(p+1)!}\sum_{s=1}^{q} g_s^{p+1}\frac{\partial^{p+1}}{\partial (z_s^c)^{p+1}}\mu(z) + \sum_{s=1}^{r}\nu_s C_s(z)\Big\}dF_X(z) \\
&+ \frac{t}{2}f_\epsilon(t)\int \frac{1}{\sigma^2(z)}\Big\{\frac{\kappa_{p+1}}{(p+1)!}\sum_{s=1}^{q} g_s^{p+1}\frac{\partial^{p+1}}{\partial (z_s^c)^{p+1}}\sigma^2(z) + \sum_{s=1}^{r}\nu_s D_s(z)\Big\}dF_X(z) + o_P(n^{-1/2}).
\end{aligned}
$$

Let us finally look at the estimator $\hat{F}_{X,\hat{\epsilon}}(x,t)$ of the joint distribution of (X,ϵ). Write

$$
\begin{aligned}
\hat{F}_{X,\hat{\epsilon}}(x,t) - F_{X,\epsilon}(x,t) &= \big(\hat{F}_{X,\hat{\epsilon}} - \hat{F}_{X,\epsilon} - E(\hat{F}_{X,\hat{\epsilon}}) + E(\hat{F}_{X,\epsilon})\big)(x,t) \\
&\quad + \big(E(\hat{F}_{X,\hat{\epsilon}}) - E(\hat{F}_{X,\epsilon}) - F_{X,\hat{\epsilon}} + F_{X,\epsilon}\big)(x,t) \\
&\quad + \big(\hat{F}_{X,\epsilon} - F_{X,\epsilon}\big)(x,t) + \big(F_{X,\hat{\epsilon}} - F_{X,\epsilon}\big)(x,t) \\
&= T_1 + T_2 + T_3 + T_4,
\end{aligned}
$$

where the expectation $E(\hat{F}_{X,\hat{\epsilon}})(x,t)$ is calculated conditional on $\hat{\epsilon}$ and where $\hat{F}_{X,\epsilon}(x,t)$ is the smoothed empirical distribution based on the true errors $\epsilon_1,\ldots,\epsilon_n$. It follows from Lemma 3.1 in Li et al. (2015) that

$$
\begin{aligned}
T_3 &= \frac{1}{n}\sum_{i=1}^{n}\big[\mathbf{1}(X_i \le x, \epsilon_i \le t) - F_{X,\epsilon}(x,t)\big] \\
&\quad + \frac{\kappa_2}{2}\sum_{s=1}^{q} h_s^2\frac{\partial^2 F_X(x)}{\partial (x_s^c)^2}F_\epsilon(t) + \sum_{s=1}^{r}\lambda_s B_s(x)F_\epsilon(t) + \frac{\kappa_2}{2}b^2 F_X(x)f_\epsilon'(t) + o_P(n^{-1/2}),
\end{aligned}
$$

using the independence between X and ϵ. Next,

$$\begin{aligned}
T_4 &= P(X \le x, \hat{\epsilon} \le t) - P(X \le x, \epsilon \le t) \\
&= \int_{-\infty}^{x} \Big[P\Big(\epsilon \le t\frac{\hat{\sigma}(z)}{\sigma(z)} + \frac{\hat{\mu}(z) - \mu(z)}{\sigma(z)} \Big) - P(\epsilon \le t) \Big] \, dF_X(z) \\
&= f_\epsilon(t) \int_{-\infty}^{x} \Big[t\frac{\hat{\sigma}(z) - \sigma(z)}{\sigma(z)} + \frac{\hat{\mu}(z) - \mu(z)}{\sigma(z)} \Big] \, dF_X(z) + o_P(n^{-1/2}) \\
&= f_\epsilon(t) \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \le x) \Big\{ \epsilon_i + \frac{t}{2}(\epsilon_i^2 - 1) \Big\} \\
&+ f_\epsilon(t) \int_{-\infty}^{x} \frac{1}{\sigma(z)} \Big\{ \frac{\kappa_{p+1}}{(p+1)!} \sum_{s=1}^{q} g_s^{p+1} \frac{\partial^{p+1}}{\partial (z_s^c)^{p+1}} \mu(z) + \sum_{s=1}^{r} \nu_s C_s(z) \Big\} dF_X(z) \\
&+ \frac{t}{2} f_\epsilon(t) \int_{-\infty}^{x} \frac{1}{\sigma^2(z)} \Big\{ \frac{\kappa_{p+1}}{(p+1)!} \sum_{s=1}^{q} g_s^{p+1} \frac{\partial^{p+1}}{\partial (z_s^c)^{p+1}} \sigma^2(z) + \sum_{s=1}^{r} \nu_s D_s(z) \Big\} dF_X(z) + o_P(n^{-1/2}),
\end{aligned}$$

similarly as in the proof of the estimator $\hat{F}_{\hat{\epsilon}}(t)$, and where the integral $\int_{-\infty}^{x}$ should be understood as a $(q+r)$-dimensional integral over the region $\prod_{s=1}^{q}(-\infty, x_s^c] \times \prod_{s=1}^{r}(-\infty, x_s^d]$. Next, we consider the term T_2 :

$$\begin{aligned}
T_2 &= \frac{\kappa_2}{2} \sum_{s=1}^{q} h_s^2 \frac{\partial^2 F_X(x)}{\partial (x_s^c)^2} \big[F_{\hat{\epsilon}}(t) - F_\epsilon(t) \big] + \sum_{s=1}^{r} \lambda_s B_s(x) \big[F_{\hat{\epsilon}}(t) - F_\epsilon(t) \big] + \frac{\kappa_2}{2} b^2 F_X(x) \big[f'_{\hat{\epsilon}}(t) - f'_\epsilon(t) \big] \\
&+ o_P\Big(\sum_{s=1}^{q} h_s^2 + \sum_{s=1}^{r} \lambda_s + b^2 \Big) \\
&= o_P(n^{-1/2}).
\end{aligned}$$

Finally, for the term T_1 first note that we can write

$$\begin{aligned}
\hat{F}_{X,\hat{\epsilon}}(x,t) &= n^{-1} \sum_{i=1}^{n} \int 1(\hat{\epsilon}_i + vb \le t) k(v) \, dv \times \prod_{s=1}^{q} \int_{-\infty}^{\infty} 1(X_{is}^c + v_s h_s \le x_s^c) k(v_s) \, dv_s \\
&\quad \times \prod_{s=1}^{r} \eta_s \sum_{\bar{x}_s^d \le x_s^d} \sum_{z_s^d} 1(X_{is}^d = z_s^d) \lambda_s^{|\bar{x}_s^d - z_s^d|} \\
&\quad = \int \dots \int \eta_1 \dots \eta_r \sum_{\bar{x}_1^d \le x_1^d} \dots \sum_{\bar{x}_r^d \le x_r^d} \sum_{z_1^d} \dots \sum_{z_r^d} \hat{G}_{X,\hat{\epsilon}}(x_1^c - v_1 h_1, \dots, x_q^c - v_q h_q, z_1^d, \dots, z_r^d, t - vb) \\
&\quad \times \lambda_1^{|\bar{x}_1^d - z_1^d|} \dots \lambda_r^{|\bar{x}_r^d - z_r^d|} k(v) k(v_1) \dots k(v_q) \, dv_q \dots dv_1 dv,
\end{aligned}$$

where $\hat{G}_{X,\hat{\epsilon}}(x,t) = n^{-1}\sum_{i=1}^{n}\mathbf{1}(X_i^c \leq x^c, X_i^d = x^d, \hat{\epsilon}_i \leq t)$. We can show that

$$\hat{G}_{X,\hat{\epsilon}}(x,t) - \hat{G}_{X,\epsilon}(x,t) - G_{X,\hat{\epsilon}}(x,t) + G_{X,\epsilon}(x,t) = o_P(n^{-1/2}),$$

uniformly in x and t, in a similar way as in the proof of Lemma A.3 in Neumeyer & Van Keilegom (2010). Hence, $T_1 = o_P(n^{-1/2})$. The iid representation for $\hat{F}_{X,\hat{\epsilon}}(x,t) - F_{X,\epsilon}(x,t)$ now follows. □

Proof of Corollary 4.

(1) For showing the weak convergence of the process $\sqrt{n}\big(\hat{F}_{X,\hat{\epsilon}}(x,t) - \hat{F}_X(x)\hat{F}_{\hat{\epsilon}}(t)\big)$, it suffices to prove that the class $\{(u,v) \to H(u,v,x,t) : x \in R_X, t \in \mathcal{R}\}$ is Donsker (see Van der Vaart & Wellner (1996)). First note that the first three terms of $H(u,v,x,t)$ consist of products and sums of indicators and of distribution functions, and hence it follows from Theorem 2.7.5 and Examples 2.10.7 and 2.10.8 in Van der Vaart & Wellner (1996) that the sum of these three terms forms a Donsker class. Next, for the fourth term note that $f_\epsilon(t)$ (resp. $tf_\epsilon(t)$) is bounded uniformly in t, that v (resp. $v^2 - 1$) does not depend on (x,t), and that $I(u \leq x) - F_X(x)$ consists of bounded and monotone functions. Hence, it is easily seen that the product of these three functions forms a Donsker class.

(2) The convergence of the statistic T_{KS} follows from the continuous mapping theorem, whereas for T_{CM} we use the Helly-Bray Theorem (Billingsley (1999)).

□

Appendix B. Empirical Power Gains, Univariate Continuous Predictor Setting

Table 5. Power and % Increase in Power, $n = 100$, Gaussian Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.028	0.097	0.140	0.042	0.116	0.183	50.0%	19.6%	30.7%
0.2	0.034	0.116	0.193	0.057	0.182	0.273	67.6%	56.9%	41.5%
0.3	0.036	0.143	0.228	0.096	0.269	0.371	166.7%	88.1%	62.7%
0.4	0.069	0.195	0.318	0.149	0.343	0.498	115.9%	75.9%	56.6%
0.5	0.091	0.273	0.422	0.231	0.502	0.641	153.8%	83.9%	51.9%
0.6	0.168	0.400	0.538	0.374	0.615	0.746	122.6%	53.8%	38.7%
0.7	0.254	0.509	0.654	0.478	0.723	0.840	88.2%	42.0%	28.4%
0.8	0.334	0.623	0.731	0.613	0.825	0.895	83.5%	32.4%	22.4%
0.9	0.386	0.661	0.764	0.645	0.839	0.897	67.1%	26.9%	17.4%
1.0	0.388	0.638	0.756	0.617	0.806	0.882	59.0%	26.3%	16.7%

Table 6. Power and % Increase in Power, $n = 200$, Gaussian Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.056	0.138	0.226	0.079	0.195	0.282	41.1%	41.3%	24.8%
0.2	0.100	0.254	0.346	0.162	0.347	0.451	62.0%	36.6%	30.3%
0.3	0.150	0.339	0.459	0.247	0.507	0.627	64.7%	49.6%	36.6%
0.4	0.238	0.475	0.604	0.445	0.709	0.812	87.0%	49.3%	34.4%
0.5	0.377	0.662	0.783	0.623	0.847	0.920	65.3%	27.9%	17.5%
0.6	0.535	0.802	0.907	0.794	0.956	0.980	48.4%	19.2%	8.0%
0.7	0.731	0.902	0.955	0.893	0.980	0.991	22.2%	8.6%	3.8%
0.8	0.835	0.947	0.971	0.940	0.986	0.990	12.6%	4.1%	2.0%
0.9	0.866	0.961	0.985	0.962	0.992	0.996	11.1%	3.2%	1.1%
1.0	0.819	0.942	0.969	0.932	0.979	0.993	13.8%	3.9%	2.5%

Table 7. Power and % Increase in Power, $n = 400$, Gaussian Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.074	0.215	0.301	0.111	0.276	0.379	50.0%	28.4%	25.9%
0.2	0.202	0.429	0.546	0.310	0.532	0.669	53.5%	24.0%	22.5%
0.3	0.332	0.618	0.731	0.530	0.778	0.873	59.6%	25.9%	19.4%
0.4	0.573	0.821	0.896	0.788	0.940	0.975	37.5%	14.5%	8.8%
0.5	0.810	0.942	0.977	0.945	0.988	0.996	16.7%	4.9%	1.9%
0.6	0.942	0.989	0.997	0.990	1.000	1.000	5.1%	1.1%	0.3%
0.7	0.984	0.997	0.999	0.995	0.999	0.999	1.1%	0.2%	0.0%
0.8	0.994	1.000	1.000	0.999	1.000	1.000	0.5%	0.0%	0.0%
0.9	0.996	1.000	1.000	1.000	1.000	1.000	0.4%	0.0%	0.0%
1.0	0.994	1.000	1.000	0.998	1.000	1.000	0.4%	0.0%	0.0%

Table 8. Power and % Increase in Power, $n = 800$, Gaussian Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.157	0.348	0.479	0.198	0.424	0.550	26.1%	21.8%	14.8%
0.2	0.435	0.679	0.792	0.544	0.783	0.873	25.1%	15.3%	10.2%
0.3	0.685	0.877	0.945	0.824	0.953	0.983	20.3%	8.7%	4.0%
0.4	0.903	0.984	0.995	0.980	0.999	1.000	8.5%	1.5%	0.5%
0.5	0.995	1.000	1.000	0.999	1.000	1.000	0.4%	0.0%	0.0%
0.6	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
0.7	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
0.9	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%

TABLE 9. Power and % Increase in Power, $n = 100$, Student-t Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.023	0.079	0.148	0.046	0.128	0.186	100.0%	62.0%	25.7%
0.2	0.037	0.117	0.184	0.079	0.201	0.284	113.5%	71.8%	54.3%
0.3	0.043	0.148	0.238	0.092	0.235	0.350	114.0%	58.8%	47.1%
0.4	0.058	0.175	0.298	0.128	0.320	0.477	120.7%	82.9%	60.1%
0.5	0.080	0.261	0.379	0.205	0.485	0.616	156.2%	85.8%	62.5%
0.6	0.152	0.350	0.505	0.305	0.614	0.763	100.7%	75.4%	51.1%
0.7	0.224	0.457	0.601	0.425	0.732	0.826	89.7%	60.2%	37.4%
0.8	0.288	0.572	0.708	0.531	0.777	0.859	84.4%	35.8%	21.3%
0.9	0.352	0.619	0.730	0.609	0.813	0.883	73.0%	31.3%	21.0%
1.0	0.338	0.603	0.739	0.582	0.793	0.872	72.2%	31.5%	18.0%

TABLE 10. Power and % Increase in Power, $n = 200$, Student-t Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.040	0.122	0.202	0.062	0.170	0.264	55.0%	39.3%	30.7%
0.2	0.079	0.211	0.314	0.133	0.318	0.418	68.4%	50.7%	33.1%
0.3	0.110	0.311	0.436	0.227	0.469	0.606	106.4%	50.8%	39.0%
0.4	0.197	0.429	0.579	0.390	0.648	0.769	98.0%	51.0%	32.8%
0.5	0.348	0.624	0.744	0.589	0.817	0.888	69.3%	30.9%	19.4%
0.6	0.522	0.791	0.892	0.799	0.939	0.967	53.1%	18.7%	8.4%
0.7	0.699	0.912	0.952	0.897	0.971	0.993	28.3%	6.5%	4.3%
0.8	0.828	0.949	0.979	0.950	0.991	0.998	14.7%	4.4%	1.9%
0.9	0.842	0.953	0.983	0.955	0.990	0.996	13.4%	3.9%	1.3%
1.0	0.813	0.932	0.964	0.925	0.975	0.985	13.8%	4.6%	2.2%

TABLE 11. Power and % Increase in Power, $n = 400$, Student-t Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.087	0.192	0.285	0.113	0.253	0.367	29.9%	31.8%	28.8%
0.2	0.183	0.417	0.552	0.286	0.540	0.667	56.3%	29.5%	20.8%
0.3	0.334	0.627	0.748	0.525	0.794	0.871	57.2%	26.6%	16.4%
0.4	0.556	0.806	0.901	0.784	0.938	0.968	41.0%	16.4%	7.4%
0.5	0.793	0.950	0.979	0.947	0.994	0.998	19.4%	4.6%	1.9%
0.6	0.950	0.993	0.998	0.991	0.999	1.000	4.3%	0.6%	0.2%
0.7	0.987	0.999	1.000	0.996	1.000	1.000	0.9%	0.1%	0.0%
0.8	0.999	1.000	1.000	1.000	1.000	1.000	0.1%	0.0%	0.0%
0.9	0.998	1.000	1.000	1.000	1.000	1.000	0.2%	0.0%	0.0%
1.0	0.993	0.999	1.000	0.999	1.000	1.000	0.6%	0.1%	0.0%

TABLE 12. Power and % Increase in Power, $n = 800$, Student-t Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.156	0.351	0.486	0.225	0.419	0.545	44.2%	19.4%	12.1%
0.2	0.416	0.685	0.797	0.528	0.780	0.870	26.9%	13.9%	9.2%
0.3	0.668	0.892	0.951	0.837	0.962	0.988	25.3%	7.8%	3.9%
0.4	0.925	0.986	0.991	0.978	0.996	0.997	5.7%	1.0%	0.6%
0.5	0.992	0.998	1.000	0.997	0.998	1.000	0.5%	0.0%	0.0%
0.6	0.998	1.000	1.000	1.000	1.000	1.000	0.2%	0.0%	0.0%
0.7	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
0.9	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%

TABLE 13. Power and % Increase in Power, $n = 100$, Beta

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.019	0.079	0.145	0.045	0.124	0.204	136.8%	57.0%	40.7%
0.2	0.034	0.108	0.190	0.060	0.182	0.273	76.5%	68.5%	43.7%
0.3	0.046	0.151	0.234	0.093	0.244	0.360	102.2%	61.6%	53.8%
0.4	0.059	0.187	0.291	0.133	0.332	0.463	125.4%	77.5%	59.1%
0.5	0.091	0.269	0.404	0.230	0.472	0.608	152.7%	75.5%	50.5%
0.6	0.163	0.374	0.529	0.343	0.625	0.745	110.4%	67.1%	40.8%
0.7	0.243	0.504	0.646	0.458	0.714	0.811	88.5%	41.7%	25.5%
0.8	0.330	0.584	0.727	0.570	0.797	0.869	72.7%	36.5%	19.5%
0.9	0.389	0.639	0.756	0.609	0.827	0.891	56.6%	29.4%	17.9%
1.0	0.372	0.617	0.755	0.615	0.811	0.882	65.3%	31.4%	16.8%

TABLE 14. Power and % Increase in Power, $n = 200$, Beta

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.029	0.115	0.198	0.060	0.182	0.262	106.9%	58.3%	32.3%
0.2	0.063	0.239	0.319	0.146	0.322	0.436	131.7%	34.7%	36.7%
0.3	0.134	0.331	0.462	0.241	0.477	0.616	79.9%	44.1%	33.3%
0.4	0.228	0.498	0.616	0.419	0.690	0.797	83.8%	38.6%	29.4%
0.5	0.362	0.634	0.753	0.614	0.837	0.911	69.6%	32.0%	21.0%
0.6	0.559	0.815	0.893	0.793	0.941	0.978	41.9%	15.5%	9.5%
0.7	0.729	0.903	0.957	0.903	0.979	0.993	23.9%	8.4%	3.8%
0.8	0.829	0.944	0.972	0.952	0.985	0.991	14.8%	4.3%	2.0%
0.9	0.870	0.967	0.990	0.966	0.995	0.997	11.0%	2.9%	0.7%
1.0	0.837	0.944	0.977	0.944	0.981	0.992	12.8%	3.9%	1.5%

TABLE 15. Power and % Increase in Power, $n = 400$, Beta

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.064	0.182	0.291	0.097	0.249	0.349	51.6%	36.8%	19.9%
0.2	0.176	0.398	0.526	0.271	0.524	0.648	54.0%	31.7%	23.2%
0.3	0.315	0.582	0.714	0.498	0.755	0.855	58.1%	29.7%	19.7%
0.4	0.540	0.799	0.893	0.770	0.929	0.970	42.6%	16.3%	8.6%
0.5	0.783	0.953	0.974	0.941	0.989	0.997	20.2%	3.8%	2.4%
0.6	0.942	0.994	0.997	0.987	1.000	1.000	4.8%	0.6%	0.3%
0.7	0.991	1.000	1.000	0.997	1.000	1.000	0.6%	0.0%	0.0%
0.8	0.997	1.000	1.000	1.000	1.000	1.000	0.3%	0.0%	0.0%
0.9	0.999	1.000	1.000	1.000	1.000	1.000	0.1%	0.0%	0.0%
1.0	0.996	1.000	1.000	0.999	1.000	1.000	0.3%	0.0%	0.0%

TABLE 16. Power and % Increase in Power, $n = 800$, Beta

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.156	0.378	0.492	0.223	0.450	0.559	42.9%	19.0%	13.6%
0.2	0.430	0.691	0.806	0.538	0.791	0.880	25.1%	14.5%	9.2%
0.3	0.705	0.893	0.948	0.841	0.961	0.979	19.3%	7.6%	3.3%
0.4	0.932	0.991	0.998	0.988	1.000	1.000	6.0%	0.9%	0.2%
0.5	0.991	1.000	1.000	1.000	1.000	1.000	0.9%	0.0%	0.0%
0.6	0.999	1.000	1.000	0.999	1.000	1.000	0.0%	0.0%	0.0%
0.7	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
0.9	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%

Appendix C. Empirical Power Gains, Multivariate Continuous Predictor Setting

Table 17. Power and % Increase in Power, $n = 100$, Gaussian Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.018	0.095	0.161	0.031	0.118	0.209	72.2%	24.2%	29.8%
0.2	0.031	0.121	0.226	0.050	0.185	0.292	61.3%	52.9%	29.2%
0.3	0.047	0.146	0.232	0.059	0.185	0.304	25.5%	26.7%	31.0%
0.4	0.046	0.139	0.245	0.053	0.202	0.318	15.2%	45.3%	29.8%
0.5	0.022	0.121	0.221	0.044	0.186	0.325	100.0%	53.7%	47.1%
0.6	0.021	0.126	0.219	0.049	0.190	0.324	133.3%	50.8%	47.9%
0.7	0.032	0.127	0.227	0.042	0.165	0.288	31.2%	29.9%	26.9%
0.8	0.051	0.172	0.272	0.054	0.195	0.298	5.9%	13.4%	9.6%
0.9	0.057	0.188	0.306	0.081	0.226	0.330	42.1%	20.2%	7.8%
1.0	0.094	0.244	0.378	0.113	0.260	0.361	20.2%	6.6%	-4.5%

Table 18. Power and % Increase in Power, $n = 200$, Gaussian Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.036	0.118	0.200	0.050	0.170	0.269	38.9%	44.1%	34.5%
0.2	0.055	0.191	0.301	0.101	0.257	0.382	83.6%	34.6%	26.9%
0.3	0.066	0.216	0.358	0.111	0.305	0.446	68.2%	41.2%	24.6%
0.4	0.093	0.273	0.395	0.146	0.364	0.503	57.0%	33.3%	27.3%
0.5	0.104	0.319	0.457	0.182	0.441	0.597	75.0%	38.2%	30.6%
0.6	0.127	0.359	0.527	0.237	0.527	0.692	86.6%	46.8%	31.3%
0.7	0.180	0.462	0.615	0.298	0.616	0.754	65.6%	33.3%	22.6%
0.8	0.278	0.565	0.708	0.388	0.701	0.814	39.6%	24.1%	15.0%
0.9	0.377	0.653	0.778	0.511	0.733	0.839	35.5%	12.3%	7.8%
1.0	0.454	0.714	0.839	0.559	0.774	0.867	23.1%	8.4%	3.3%

Table 19. Power and % Increase in Power, $n = 400$, Gaussian Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.058	0.198	0.303	0.109	0.263	0.405	87.9%	32.8%	33.7%
0.2	0.124	0.296	0.446	0.179	0.393	0.562	44.4%	32.8%	26.0%
0.3	0.168	0.416	0.597	0.273	0.550	0.684	62.5%	32.2%	14.6%
0.4	0.201	0.505	0.669	0.364	0.661	0.804	81.1%	30.9%	20.2%
0.5	0.291	0.610	0.754	0.494	0.793	0.883	69.8%	30.0%	17.1%
0.6	0.464	0.763	0.875	0.687	0.893	0.950	48.1%	17.0%	8.6%
0.7	0.631	0.876	0.928	0.816	0.947	0.983	29.3%	8.1%	5.9%
0.8	0.820	0.946	0.978	0.928	0.982	0.989	13.2%	3.8%	1.1%
0.9	0.913	0.981	0.991	0.963	0.993	0.996	5.5%	1.2%	0.5%
1.0	0.951	0.992	0.999	0.982	0.998	0.998	3.3%	0.6%	-0.1%

Table 20. Power and % Increase in Power, $n = 800$, Gaussian Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.096	0.282	0.422	0.155	0.373	0.509	61.5%	32.3%	20.6%
0.2	0.210	0.482	0.647	0.301	0.580	0.708	43.3%	20.3%	9.4%
0.3	0.342	0.629	0.773	0.488	0.746	0.857	42.7%	18.6%	10.9%
0.4	0.502	0.799	0.886	0.696	0.896	0.962	38.6%	12.1%	8.6%
0.5	0.706	0.929	0.973	0.890	0.978	0.992	26.1%	5.3%	2.0%
0.6	0.901	0.986	0.997	0.979	0.998	0.999	8.7%	1.2%	0.2%
0.7	0.980	0.997	1.000	0.997	1.000	1.000	1.7%	0.3%	0.0%
0.8	0.999	1.000	1.000	1.000	1.000	1.000	0.1%	0.0%	0.0%
0.9	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%
1.0	1.000	1.000	1.000	1.000	1.000	1.000	0.0%	0.0%	0.0%

TABLE 21. Power and % Increase in Power, $n = 100$, Student-t Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.028	0.100	0.169	0.042	0.122	0.219	50.0%	22.0%	29.6%
0.2	0.035	0.125	0.208	0.048	0.176	0.283	37.1%	40.8%	36.1%
0.3	0.043	0.141	0.228	0.058	0.180	0.309	34.9%	27.7%	35.5%
0.4	0.033	0.135	0.224	0.052	0.165	0.278	57.6%	22.2%	24.1%
0.5	0.027	0.112	0.202	0.047	0.160	0.258	74.1%	42.9%	27.7%
0.6	0.021	0.105	0.182	0.023	0.136	0.255	9.5%	29.5%	40.1%
0.7	0.020	0.087	0.168	0.026	0.117	0.197	30.0%	34.5%	17.3%
0.8	0.029	0.116	0.195	0.039	0.127	0.215	34.5%	9.5%	10.3%
0.9	0.040	0.158	0.244	0.043	0.130	0.221	7.5%	-17.7%	-9.4%
1.0	0.062	0.170	0.275	0.051	0.145	0.228	-17.7%	-14.7%	-17.1%

TABLE 22. Power and % Increase in Power, $n = 200$, Student-t Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.037	0.143	0.221	0.063	0.171	0.272	70.3%	19.6%	23.1%
0.2	0.056	0.207	0.311	0.101	0.252	0.371	80.4%	21.7%	19.3%
0.3	0.079	0.222	0.334	0.116	0.327	0.453	46.8%	47.3%	35.6%
0.4	0.082	0.277	0.397	0.141	0.358	0.499	72.0%	29.2%	25.7%
0.5	0.086	0.263	0.409	0.132	0.359	0.507	53.5%	36.5%	24.0%
0.6	0.111	0.302	0.426	0.153	0.433	0.580	37.8%	43.4%	36.2%
0.7	0.116	0.324	0.488	0.185	0.454	0.596	59.5%	40.1%	22.1%
0.8	0.172	0.416	0.563	0.214	0.484	0.639	24.4%	16.3%	13.5%
0.9	0.233	0.502	0.644	0.298	0.541	0.669	27.9%	7.8%	3.9%
1.0	0.310	0.561	0.691	0.357	0.588	0.701	15.2%	4.8%	1.4%

TABLE 23. Power and % Increase in Power, $n = 400$, Student-t Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.053	0.185	0.266	0.080	0.211	0.317	50.9%	14.1%	19.2%
0.2	0.132	0.293	0.410	0.162	0.366	0.501	22.7%	24.9%	22.2%
0.3	0.156	0.391	0.547	0.238	0.490	0.633	52.6%	25.3%	15.7%
0.4	0.191	0.457	0.629	0.307	0.577	0.719	60.7%	26.3%	14.3%
0.5	0.252	0.536	0.699	0.397	0.695	0.818	57.5%	29.7%	17.0%
0.6	0.357	0.653	0.785	0.518	0.793	0.881	45.1%	21.4%	12.2%
0.7	0.453	0.739	0.864	0.649	0.871	0.942	43.3%	17.9%	9.0%
0.8	0.616	0.856	0.925	0.756	0.923	0.961	22.7%	7.8%	3.9%
0.9	0.752	0.922	0.966	0.839	0.960	0.986	11.6%	4.1%	2.1%
1.0	0.820	0.957	0.981	0.877	0.978	0.987	7.0%	2.2%	0.6%

TABLE 24. Power and % Increase in Power, $n = 800$, Student-t Mixture

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.093	0.252	0.376	0.116	0.301	0.435	24.7%	19.4%	15.7%
0.2	0.187	0.451	0.607	0.256	0.505	0.649	36.9%	12.0%	6.9%
0.3	0.323	0.583	0.735	0.415	0.675	0.793	28.5%	15.8%	7.9%
0.4	0.446	0.724	0.844	0.575	0.823	0.907	28.9%	13.7%	7.5%
0.5	0.598	0.843	0.922	0.757	0.922	0.964	26.6%	9.4%	4.6%
0.6	0.781	0.946	0.980	0.897	0.981	0.993	14.9%	3.7%	1.3%
0.7	0.931	0.989	0.997	0.973	0.999	1.000	4.5%	1.0%	0.3%
0.8	0.981	1.000	1.000	0.997	1.000	1.000	1.6%	0.0%	0.0%
0.9	0.996	1.000	1.000	0.999	1.000	1.000	0.3%	0.0%	0.0%
1.0	0.999	1.000	1.000	1.000	1.000	1.000	0.1%	0.0%	0.0%

TABLE 25. Power and % Increase in Power, $n = 100$, Beta

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.011	0.044	0.085	0.009	0.060	0.113	-18.2%	36.4%	32.9%
0.2	0.005	0.029	0.075	0.004	0.044	0.096	-20.0%	51.7%	28.0%
0.3	0.006	0.041	0.075	0.004	0.033	0.083	-33.3%	-19.5%	10.7%
0.4	0.007	0.039	0.082	0.005	0.036	0.072	-28.6%	-7.7%	-12.2%
0.5	0.004	0.040	0.085	0.008	0.045	0.084	100.0%	12.5%	-1.2%
0.6	0.011	0.058	0.102	0.008	0.051	0.105	-27.3%	-12.1%	2.9%
0.7	0.013	0.056	0.122	0.011	0.063	0.134	-15.4%	12.5%	9.8%
0.8	0.019	0.075	0.143	0.020	0.091	0.175	5.3%	21.3%	22.4%
0.9	0.014	0.091	0.173	0.025	0.121	0.226	78.6%	33.0%	30.6%
1.0	0.026	0.110	0.201	0.044	0.165	0.277	69.2%	50.0%	37.8%

TABLE 26. Power and % Increase in Power, $n = 200$, Beta

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.011	0.047	0.092	0.015	0.057	0.122	36.4%	21.3%	32.6%
0.2	0.014	0.052	0.112	0.016	0.076	0.131	14.3%	46.2%	17.0%
0.3	0.014	0.061	0.109	0.010	0.062	0.126	-28.6%	1.6%	15.6%
0.4	0.013	0.050	0.096	0.019	0.067	0.140	46.2%	34.0%	45.8%
0.5	0.015	0.080	0.155	0.030	0.107	0.186	100.0%	33.8%	20.0%
0.6	0.014	0.092	0.164	0.022	0.128	0.245	57.1%	39.1%	49.4%
0.7	0.034	0.132	0.224	0.048	0.165	0.289	41.2%	25.0%	29.0%
0.8	0.045	0.149	0.243	0.067	0.219	0.349	48.9%	47.0%	43.6%
0.9	0.045	0.170	0.292	0.096	0.289	0.418	113.3%	70.0%	43.2%
1.0	0.057	0.240	0.376	0.142	0.371	0.515	149.1%	54.6%	37.0%

TABLE 27. Power and % Increase in Power, $n = 400$, Beta

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.020	0.087	0.174	0.032	0.133	0.223	60.0%	52.9%	28.2%
0.2	0.014	0.079	0.145	0.025	0.110	0.206	78.6%	39.2%	42.1%
0.3	0.008	0.054	0.123	0.018	0.106	0.203	125.0%	96.3%	65.0%
0.4	0.018	0.094	0.181	0.039	0.163	0.260	116.7%	73.4%	43.6%
0.5	0.036	0.166	0.265	0.049	0.223	0.360	36.1%	34.3%	35.8%
0.6	0.076	0.218	0.336	0.109	0.300	0.457	43.4%	37.6%	36.0%
0.7	0.083	0.246	0.378	0.137	0.370	0.530	65.1%	50.4%	40.2%
0.8	0.101	0.303	0.456	0.187	0.430	0.608	85.1%	41.9%	33.3%
0.9	0.161	0.390	0.549	0.272	0.550	0.713	68.9%	41.0%	29.9%
1.0	0.193	0.490	0.674	0.359	0.687	0.821	86.0%	40.2%	21.8%

TABLE 28. Power and % Increase in Power, $n = 800$, Beta

	Non-Smooth Power			Smooth Power			Percentage Gain		
δ	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
0.1	0.049	0.164	0.256	0.075	0.233	0.373	53.1%	42.1%	45.7%
0.2	0.045	0.144	0.245	0.084	0.237	0.378	86.7%	64.6%	54.3%
0.3	0.044	0.153	0.282	0.085	0.277	0.417	93.2%	81.0%	47.9%
0.4	0.064	0.215	0.350	0.101	0.314	0.467	57.8%	46.0%	33.4%
0.5	0.098	0.305	0.454	0.158	0.430	0.572	61.2%	41.0%	26.0%
0.6	0.158	0.404	0.572	0.267	0.558	0.698	69.0%	38.1%	22.0%
0.7	0.220	0.507	0.669	0.374	0.690	0.819	70.0%	36.1%	22.4%
0.8	0.322	0.612	0.770	0.489	0.759	0.873	51.9%	24.0%	13.4%
0.9	0.381	0.695	0.829	0.596	0.857	0.938	56.4%	23.3%	13.1%
1.0	0.520	0.816	0.902	0.739	0.922	0.970	42.1%	13.0%	7.5%

Appendix D. R Code

Below we present the R code used to generate the simulations contained herein for the univariate continuous predictor setting.

```
## This program implements the test for a location scale model
## outlined in Journal of Econometrics 143 (2008) 88-102
## "Specification tests in nonparametric regression John
## H.J. Einmahla, Ingrid Van Keilegom". The version implemented is the
## KS variant, equation (2.5) page 90, and per their suggestion the
## bootstrap procedure described on page 93. This is simply a test of
## independence of the studentized residuals and the predictors
## X. Note this is for the univariate X case only, though extension to
## multivariate settings is trivial. We also compute the smooth
## version of the test and demonstrate that this results in
## substantial power gains while leaving size unaffected.

rm(list=ls())
library(np)
options(np.tree=TRUE,np.messages=FALSE)
## Use of the Epanechnikov kernel reduces computation time when used in
## conjunction with trees
ckertype <- "epanechnikov"

set.seed(42)

## Function to generate nonparametric P values
P.value <- function(lambda,lambda.boot) { mean(ifelse(lambda.boot > lambda,1,0)) }

## Function to generate the DGP for mu(x) in the model y=mu(x)+sigma(x)e
dgp <- function(x) { sin(2*pi*x) }

## Can switch from size to power: 0 = size, != 0 power
size.power <- scan("size_power.dat")
## Sample size
n <- scan("num_obs.dat")
## Number of bootstrap replications
B <- scan("num_boot.dat")
## Number of Monte Carlo replications
M <- scan("num_monte.dat")

## Some vectors for storage etc.

P.vector <- numeric()
P.smooth.vector <- numeric()

T.ks.star <- numeric(B)
T.smooth.ks.star <- numeric(B)

## Output files with headers...

write("P",file="P_value.out")
write("T.ks",file="T_value.out")

write("P.smooth",file="P_smooth_value.out")
write("T.smooth.ks",file="T_smooth_value.out")

for(m in 1:M) {

    ## x lies in [0,1]

    x <- runif(n)

    ## Beta mixture
    ## epsilon <- rbeta(n,1+size.power*10*x,11-size.power*10*x)
    ## Gaussian mixture
    ## epsilon <- sapply(1:n,function(i){rnormMix(1,mean1=-1,sd1=0.5,mean2=1,sd2=0.5,p.mix=size.power*x[i])})
    ## Student-t mixture
    epsilon <- sapply(1:n,function(i){rtmix(1, prob = size.power*x[i])})

    ## Re-center to have mean zero
    epsilon <- (epsilon - mean(epsilon))/sd(epsilon)
    y <- dgp(x) + epsilon

    ## Local constant estimator, least squares cross-validation
    ## (default)
```

```
model <- npreg(y~x,ckertype=ckertype)
r <- residuals(model)**2

## Fan and Yao's (1998) conditional variance estimator with
## trimming if needed, local constant estimator, least squares
## cross-validation (default)

model.var.fy <- npreg(r~x,ckertype=ckertype)
var.fy <- fitted(model.var.fy)
sd.fy <- sqrt(ifelse(var.fy<=0,.Machine$double.eps,var.fy))

## Studentized residuals

epsilon <- residuals(model)/sd.fy

## Cross-validated bandwidths for smooth distribution functions
## (marginals and joint)

bw.x.eps <- npudistbw(~x+epsilon,ckertype=ckertype)

## Smooth distribution functions, joint bws for marginal x and
## epsilon, joint bws for x and epsilon (common bandwidths used to
## construct joint and marginals)

F.smooth.x <- fitted(npudist(~x,ckertype=ckertype,bws=bw.x.eps$bw[1]))
F.smooth.eps <- fitted(npudist(~epsilon,ckertype=ckertype,bws=bw.x.eps$bw[2]))
F.smooth.x.eps <- fitted(npudist(bws=bw.x.eps))
T.smooth.ks <- sqrt(n)*max(abs(F.smooth.x.eps - F.smooth.x*F.smooth.eps))

## Non-smooth (empirical) distribution functions

F.x <- sapply(1:n,function(i){mean(ifelse(x <= x[i],1,0))})
F.eps <- sapply(1:n,function(i){mean(ifelse(epsilon <= epsilon[i],1,0))})
F.x.eps <- sapply(1:n,function(i){mean(ifelse(x <= x[i] & epsilon <= epsilon[i],1,0))})
T.ks <- sqrt(n)*max(abs(F.x.eps - F.x*F.eps))

## Bootstrap procedure for generating the null distribution of the
## statistics

for(b in 1:B) {

    ## This bootstrap procedure breaks any systematic relationship
    ## between epsilon and X. For bandwidth selection, the
    ## authors write "The bandwidth an, for estimating m, is
    ## selected by means of a least-squares cross-validation
    ## procedure; for computing the bootstrap test statistics the
    ## same bandwidth has been used." (page 95). We therefore take
    ## $bws$bw from the above objects for the model and
    ## conditional variance.

    ## Studentized residual resample

    y.star <- fitted(model) + sd.fy*sample(epsilon,replace=TRUE)
    model.star <- npreg(y.star~x,ckertype=ckertype,bws=model$bws$bw)
    r.star <- residuals(model.star)**2

    ## Recompute conditional standard deviation and bootstrap
    ## studentized residuals

    var.fy.star <- fitted(npreg(r.star~x,ckertype=ckertype,bws=model.var.fy$bws$bw))
    sd.fy.star <- sqrt(ifelse(var.fy.star<=0,.Machine$double.eps,var.fy.star))
    epsilon.star <- residuals(model.star)/sd.fy.star

    ## We don't need to recompute the empirical distribution of F.x here

    ## Smooth distribution functions, joint bws for marginal x and
    ## epsilon, joint bws for x and epsilon

    F.smooth.x.eps.star <- fitted(npudist(~x+epsilon.star,bws=bw.x.eps$bw,ckertype=ckertype))
    F.smooth.eps.star <- fitted(npudist(~epsilon.star,bws=bw.x.eps$bw[2],ckertype=ckertype))
    T.smooth.ks.star[b] <- sqrt(n)*max(abs(F.smooth.x.eps.star - F.smooth.x*F.smooth.eps.star))

    ## Empirical distribution functions

    F.eps.star <- sapply(1:n,function(i){mean(ifelse(epsilon.star <= epsilon.star[i],1,0))})
```

```
        F.x.eps.star <- sapply(1:n,function(i){mean(ifelse(x <= x[i] & epsilon.star <= epsilon.star[i],1,0))})
        T.ks.star[b] <- sqrt(n)*max(abs(F.x.eps.star - F.x*F.eps.star))

    }

    ## Compute the empirical P-values

    P.vector[m] <- P.value(T.ks,T.ks.star)
    P.smooth.vector[m] <- P.value(T.smooth.ks,T.smooth.ks.star)

    write(P.vector[m],file="P_value.out",append=TRUE)
    write(T.ks,file="T_value.out",append=TRUE)

    write(P.smooth.vector[m],file="P_smooth_value.out",append=TRUE)
    write(T.smooth.ks,file="T_smooth_value.out",append=TRUE)

    ## Update empirical rejection frequencies as Monte Carlo proceeds
    ## (1%, 5%, and 10% levels) and write to files

    write(mean(ifelse(P.vector < 0.01, 1, 0)),file="reject_01.out")
    write(mean(ifelse(P.vector < 0.05, 1, 0)),file="reject_05.out")
    write(mean(ifelse(P.vector < 0.10, 1, 0)),file="reject_10.out")

    write(mean(ifelse(P.smooth.vector < 0.01, 1, 0)),file="reject_smooth_01.out")
    write(mean(ifelse(P.smooth.vector < 0.05, 1, 0)),file="reject_smooth_05.out")
    write(mean(ifelse(P.smooth.vector < 0.10, 1, 0)),file="reject_smooth_10.out")

}
```