INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0333

GOODNESS-OF-FIT TESTS IN NONPARAMETRIC REGRESSION

J. EINMAHL and I. VAN KEILEGOM

http://www.stat.ucl.ac.be

Goodness-of-fit tests in nonparametric regression

John H.J. Einmahl — *Tilburg University*

Ingrid Van Keilegom — *Université catholique de Louvain*

December 17, 2003

Abstract

Consider the nonparametric regression model $Y = m(X) + \varepsilon$, where the function m is smooth, but unknown, and ε is independent of X. We construct omnibus goodness-of-fit tests, based on n independent copies of (X, Y), for the independence of ε and X and establish asymptotic results for the proposed tests statistics. We investigate their finite sample properties through a simulation study and present an econometric application to household data. One testing procedure is based on differences of neighboring Y's, whereas the other one makes use of an estimator of m. The proofs are based on delicate weighted empirical process theory.

AMS 2000 subject classifications. Primary: 62G08, 62G10, 62G20, 62G30; secondary: 60F17.

Key words and phrases. Goodness-of-fit, nonparametric regression, test for independence, weak convergence.

1 Introduction

Let (X, Y) be a bivariate random vector where Y is the variable of interest and X is a covariate. We assume that X and Y are related via the nonparametric regression model

$$Y = m(X) + \varepsilon, \tag{1.1}$$

where m is the unknown regression curve and ε is the error. In order to avoid identification problems, we define m as follows. Let T be a given location functional, i.e. for any random variabele Z and any $a > 0$ and b, we have $T(F_{aZ+b}) = aT(F_Z) + b$, where F_{aZ+b} is the distribution function of $aZ + b$. Now we define $m(x) = T(F(\cdot|x))$, with $F(\cdot|x)$ the conditional distribution function of Y, given $X = x$. As a consequence, $T(F_\varepsilon(\cdot|x)) = 0$, with $F_\varepsilon(\cdot|x)$ the conditional distribution function of ε, given $X = x$. In particular we can choose T to be the median (or a quantile), the mode or the mean. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$, n independent replications of (X, Y), be our data.

In this paper we consider the problem of constructing omnibus goodness-of-fit tests for the submodel where

$$\varepsilon \text{ is stochastically independent of } X \tag{1.2}$$

or, in other words, where the conditional distribution of $Y - m(X)$, given $X = x$, does not depend on x. So we will propose procedures for testing the independence between ε and X, that will detect any deviation from the null hypothesis. Although in nonparametric regression the model (1.1) is very standard, testing of (1.2) against the general alternative of dependence seems not to be addressed in the literature. In a number of papers (see, e.g., Liero (2003) and Cao and Gijbels (2003)) tests for homoscedasticity are developed. Instead of looking at the conditional variance only, in this paper we consider the full conditional distribution of ε given X.

Since the errors $\varepsilon_1, \ldots, \varepsilon_n$ are not observed, we cannot use them directly. One approach which deals with this problem considers appropriate differences of Y's corresponding to neighboring X-values. Since m is smooth, m almost cancels out in these differences. Another approach replaces the ε_i by $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$, where $\hat{m}$ is a nonparametric estimator of m, proposed in (3.2) below. The main difficulty is however that for both approaches the thus obtained "errors" are dependent, and hence the classical tests for independence available in the literature cannot be applied here, since most tests assume that the pairs of observations are i.i.d. In this paper we focus on three tests, namely the Kolmogorov-Smirnov, the Cramér-von Mises and the Anderson-Darling test (see, e.g., Shorack and Wellner (1986)). We adapt these tests to the present setup and derive their asymptotic distributions.

Difference-based procedures are widely used in nonparametric regression, especially for the estimation of the error variance (see e.g. Dette et al. (1998), Liero (2003) and Müller et al. (2003)).

The model described in (1.1)-(1.2) has been considered in many papers (see e.g. Akritas and Van Keilegom (2001) and Müller et al. (2004) for some of the more recent contributions) and can be considered as a nonparametric version of the classical parametric linear model. Although the results in this paper will be presented for random design, they can easily be adapted to fixed design. Note that in that case, interest lies in the fact whether or not the error terms $\varepsilon_1, \ldots, \varepsilon_n$ are identically distributed.

Apart from being a goodness-of-fit test for the nonparametric model (1.1)-(1.2), the tests proposed in this paper can also serve for other purposes. Suppose e.g. that one likes to know whether a certain random vector (X, Y) satisfies a parametric model $Y = m_{\boldsymbol{\beta}}(X) + \varepsilon$ (where ε is independent of X and $m_{\boldsymbol{\beta}}$ is a parametric regression curve, of which the form is still to be determined). In such a situation it might be useful to use the nonparametric tests proposed above. If the tests indicate that the independence between ε and X holds, one can then start searching for the particular form of the parametric regression curve.

This paper is organized as follows. In Sections 2 and 3 we propose the two classes of test statistics, respectively, and state and prove the main results. In Section 4 we investigate the small sample performance of the tests of Section 2 in a simulation study and in Section 5 we present an econometric application. Proofs of some lemmas are deferred to the Appendix.

2 The Differences Approach

Consider the model described in (1.1). We will write F_X for the distribution function (df) of X and F_ε for the (unconditional) df of ε. Let $(X_1, \varepsilon_1), ..., (X_n, \varepsilon_n)$ be i.i.d. copies of (X, ε). We want to test

$$H_0 : \varepsilon \text{ is independent of } X$$

against the alternative of dependence, based on $(X_i, Y_i), i = 1, \ldots, n$, with $Y_i = m(X_i)+\varepsilon_i$. In this section we present certain test statistics and derive their asymptotic distribution under H_0. It should be noted that for the approach detailed below the actual choice of the location functional T (see Section 1) has, under H_0, no influence on the distribution of the test statistics below. If H_0 does not hold, the influence of the choice of T on the distribution of the test statistics is typically very minor. So this method is rather robust in this sense.

Assume that the null hypothesis holds true, so $X_1, \ldots, X_n$ and $\varepsilon_1, \ldots, \varepsilon_n$ are two independent i.i.d. samples. Denote with $R_1, ..., R_n$ the ranks of $X_1, ..., X_n$. Observe that $X_1, ..., X_n$ and $\varepsilon_{R_1}, ..., \varepsilon_{R_n}$ are also two independent i.i.d. samples. We consider $(X_1, \varepsilon_{R_1}), ..., (X_n, \varepsilon_{R_n})$. These are i.i.d. random vectors with independent components; clearly ε_{R_i} has df F_ε. Now we redefine our Y_i through $Y_i = m(X_i) + \varepsilon_{R_i}$. Obviously the new data have the same probability distribution as the original ones.

Define $G(y) = P(\varepsilon_2 - \varepsilon_1 \leq y)$ and let g be the corresponding density. So, with f_ε the density of ε, we have

$$g(y) = \int_{-\infty}^{\infty} f_\varepsilon(x) f_\varepsilon(x-y) dx$$

and

$$G(y) = \int_{-\infty}^{\infty} (1 - F_\varepsilon(x-y)) dF_\varepsilon(x).$$

Write

$$F_n(x, y) = \frac{1}{n} \sum_{j=1}^{n} I(X_{j:n} \leq x,\, m(X_{j+1:n}) - m(X_{j:n}) + \varepsilon_{j+1} - \varepsilon_j \leq y).$$

(For convenience we relabeled the original n by $n+1$ in order to define $X_{n+1:n}$ (actually $X_{n+1:n+1}$) and ε_{n+1}; the final sample size is now n.) Note that F_n is the bivariate empirical df of the pairs $(X_i,\, Y_{ri} - Y_i)$, where Y_{ri} is the Y corresponding to the right neighbor of X_i.

Let $\hat{F}_X$ be the empirical df of $X_1, \ldots, X_n$, i.e. $\hat{F}_X(x) = F_n(x, \infty)$; set similarly $\hat{G}(y) =$

$F_n(\infty, y)$. For our testing problem we consider the following test statistics:

$$T_{n,KS} = \sqrt{n} \sup \left| F_n(x,y) - \hat{F}_X(x)\hat{G}(y) \right|, \tag{2.1}$$

$$T_{n,CM} = n \iint (F_n(x,y) - \hat{F}_X(x)\hat{G}(y))^2 d\hat{F}_X(x) d\hat{G}(y), \tag{2.2}$$

$$T_{n,AD} = n \iint \frac{(F_n(x,y) - \hat{F}_X(x)\hat{G}(y))^2}{\hat{F}_X(x)\hat{G}(y)(1-\hat{F}_X(x))(1-\hat{G}(y))} d\hat{F}_X(x) d\hat{G}(y). \tag{2.3}$$

Let V_0 be a centered, bivariate Gaussian process with covariance structure

$$\begin{aligned}
&E\ V_0(x_1, y_1)V_0(x_2, y_2) \\
&= (F_X(x_1 \wedge x_2) - F_X(x_1)F_X(x_2))(G(y_1 \wedge y_2) + H(y_1, y_2) + H(y_2, y_1) - 3G(y_1)G(y_2)), \\
&x_1, x_2, y_1, y_2 \in I\!R,
\end{aligned}$$

where

$$\begin{aligned}
H(y_1, y_2) &= P(\varepsilon_2 - \varepsilon_1 \le y_1, \varepsilon_3 - \varepsilon_2 \le y_2) \\
&= \int_{-\infty}^{\infty} (1 - F_\varepsilon(x - y_1))F_\varepsilon(y_2 + x) dF_\varepsilon(x).
\end{aligned}$$

Observe that V_0 is tied-down at all 4 sides, i.e. $V_0(x,y) = 0$ a.s. if $x = -\infty$ or $x = \infty$ or $y = -\infty$ or $y = \infty$. It will be shown in the proof of Theorem 2.1 below that V_0 is the weak limit of $\sqrt{n}(F_n - \hat{F}_X\hat{G})$.

Denote with D_X the support of X and with f_X its density. We assume that

$$D_X \text{ is a bounded interval and } \inf_{x \in D_X} f_X(x) > 0. \tag{2.4}$$

We also assume that m is differentiable, that

$$\sup_{x \in D_X} |m'(x)| < \infty \tag{2.5}$$

and

$$\sup_{y \in I\!R} f_\varepsilon(y) =: C < \infty. \tag{2.6}$$

Theorem 2.1 *Under H_0 and (2.4), (2.5) and (2.6),*

$$T_{n,KS} \xrightarrow{d} \sup_{\substack{x \in D_X \\ y \in \mathbb{R}}} |V_0(x,y)|, \tag{2.7}$$

$$T_{n,CM} \xrightarrow{d} \iint V_0^2(x,y) dF_X(x) dG(y), \tag{2.8}$$

$$T_{n,AD} \xrightarrow{d} \iint \frac{V_0^2(x,y)}{F_X(x)G(y)(1-F_X(x))(1-G(y))} dF_X(x) dG(y). \tag{2.9}$$

Remark 2.1 The distribution of $\varepsilon_2 - \varepsilon_1$, conditional on $X = x$, is symmetric and hence the third moment is equal to 0, if it exists. Therefore differences in conditional third moment of the ε's are not detected by our test procedures. We can modify our procedures in the following way in order to overcome this problem. Instead of basing the tests on the pairs $(X_i, Y_{ri} - Y_i)$, we will base them on the pairs $(X_i, Y_{ri} - 2Y_i + Y_{li})$, $i = 1, \ldots, n$, where Y_{li} is the Y corresponding to the left neighbor of X_i. Now, mutatis mutandis, Theorem 2.1 remains true. In particular the process V_0 has now covariance structure

$$\begin{aligned}
&E\ V_0(x_1,y_1)V_0(x_2,y_2) \\
&= (F_X(x_1 \wedge x_2) - F_X(x_1)F_X(x_2))(G(y_1 \wedge y_2) + H_1(y_1,y_2) + H_2(y_1,y_2) \\
&\quad + H_1(y_2,y_1) + H_2(y_2,y_1) - 5G(y_1)G(y_2)),
\end{aligned}$$

with

$$\begin{aligned}
G(y) &= P(\varepsilon_3 - 2\varepsilon_2 + \varepsilon_1 \le y), \\
H_1(y_1,y_2) &= P(\varepsilon_3 - 2\varepsilon_2 + \varepsilon_1 \le y_1, \varepsilon_4 - 2\varepsilon_3 + \varepsilon_2 \le y_2), \\
H_2(y_1,y_2) &= P(\varepsilon_3 - 2\varepsilon_2 + \varepsilon_1 \le y_1, \varepsilon_5 - 2\varepsilon_4 + \varepsilon_3 \le y_2).
\end{aligned}$$

The thus obtained tests will be much more sensitive to alternatives where, e.g., the conditional second moments of the ε's are constant in x, whereas the conditional third moments do depend on x. Note that the above linear combination of three Y's (modulo permutations or multiplication with -1 of the coefficients) is the one where the absolute value of the third moment of the corresponding linear combination of ε's is maximal, for a fixed variance.

Remark 2.2 To obtain the critical values for the test statistics $T_{n,KS}, T_{n,CM}$ and $T_{n,AD}$, first observe that $V_0(x,y)$ can be written as

$$V_0(x,y) = Z(F_X(x),y) - F_X(x)Z(1,y), \tag{2.10}$$

where Z is a zero-mean Gaussian process with covariance function

$$\mathrm{Cov}(Z(x_1,y_1),Z(x_2,y_2)) = (x_1 \wedge x_2)(G(y_1 \wedge y_2) + H(y_1,y_2) + H(y_2,y_1) - 3G(y_1)G(y_2)).$$

To simulate an 'estimated' version of Z (for G and H are unknown), first partition the interval $[0,1]$ by means of r_x equidistant points $x_k = k/r_x$ $(k = 1,\ldots,r_x)$ and use a grid of r_y points y_ℓ $(\ell = 1,\ldots,r_y)$ on the real line. Then, simulate r_x i.i.d. r_y-variate normal random vectors $Z_k = (Z_{k1},\ldots,Z_{kr_y})$ $(k = 1,\ldots,r_x)$ with zero mean and covariance matrix

$$\mathrm{Cov}(Z_1) = \Big(r_x^{-1}[\hat{G}(y_i \wedge y_j) + \hat{H}(y_i,y_j) + \hat{H}(y_j,y_i) - 3\hat{G}(y_i)\hat{G}(y_j)]\Big)_{i,j=1}^{r_y},$$

where

$$\hat{G}(y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_{i+1:n} - Y_{i:n} \le y),$$

$$\hat{H}(y_1,y_2) = \frac{1}{n-1}\sum_{i=1}^{n-1} I(Y_{i+1:n} - Y_{i:n} \le y_1, Y_{i+2:n} - Y_{i+1:n} \le y_2).$$

Note that $Z_1,\ldots,Z_{r_x}$ can be simulated by noting that $Z_k = \sqrt{\mathrm{Cov}(Z_1)}(W_1^{(k)},\ldots,W_{r_y}^{(k)})'$ $(k = 1,\ldots,r_x)$, where $W_1^{(k)},\ldots,W_{r_y}^{(k)}$ are independent standard normal random variables. The process Z is now approximated by the $(r_x \times r_y)-$variate random vector $\tilde{Z}(x_k,y_\ell) = \sum_{j=1}^{k} Z_{j\ell}$. Now V_0 can be approximated by using the approximation of Z and by replacing F_X with $\hat{F}_X$ in (2.10). After repeating this procedure a large number of times, the critical values of the three tests can be approximated very well.

Proof of Theorem 2.1 First we show that $F_n(x,y)$ can be approximated by

$$\begin{aligned}\widetilde{F}_n(x,y) &= \frac{1}{n}\sum_{j=1}^{n} I(X_{j:n} \le x,\, \varepsilon_{j+1} - \varepsilon_j \le y)\\ &= \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x,\, \varepsilon_{R_i+1} - \varepsilon_{R_i} \le y).\end{aligned}$$

Using (2.4) we obtain

$$\max_j (X_{j+1:n} - X_{j:n}) = O_P\left(\frac{\log n}{n}\right).$$

This in combination with (2.5) yields

$$\begin{aligned} & \max_{1\le j\le n} |m(X_{j+1:n}) - m(X_{j:n})| \\ = & \sup_{x\in D_X} |m'(x)| O_P\left(\frac{\log n}{n}\right) = O_P\left(\frac{\log n}{n}\right) = o_P\left(\frac{\log^2 n}{n}\right). \end{aligned}$$

So with arbitrarily high probability for large n

$$F_n(x,y) \le \widetilde{F}_n\left(x, y + \frac{\log^2 n}{n}\right).$$

Set $F(x,y) = F_X(x)G(y)$. Then

$$\begin{aligned} \alpha_n(x,y) &:= \sqrt{n}(F_n(x,y) - F(x,y)) \\ &\le \sqrt{n}\left(\widetilde{F}_n\left(x, y + \frac{\log^2 n}{n}\right) - F\left(x, y + \frac{\log^2 n}{n}\right)\right) \\ &\quad + \sqrt{n}\left(F\left(x, y + \frac{\log^2 n}{n}\right) - F(x,y)\right) \\ &=: \widetilde{\alpha}_n\left(x, y + \frac{\log^2 n}{n}\right) + \sqrt{n}F_X(x)\left(G\left(y + \frac{\log^2 n}{n}\right) - G(y)\right). \end{aligned}$$

From (2.6) we see

$$\sup_{\substack{x\in D_X \\ y\in \mathbb{R}}} \sqrt{n}F_X(x)\left(G\left(y + \frac{\log^2 n}{n}\right) - G(y)\right) \le C\frac{\log^2 n}{\sqrt{n}}.$$

So we have with arbitrarily high probability for large n and uniformly in x and y

$$\left.\begin{aligned} \alpha_n(x,y) \le \widetilde{\alpha}_n\left(x, y + \tfrac{\log^2 n}{n}\right) + C\tfrac{\log^2 n}{\sqrt{n}} \\ \alpha_n(x,y) \ge \widetilde{\alpha}_n\left(x, y - \tfrac{\log^2 n}{n}\right) - C\tfrac{\log^2 n}{\sqrt{n}} \end{aligned}\right\} \qquad (2.11)$$

where the latter inequality follows similarly.

We now consider the weak convergence of

$$\widetilde{\alpha}_n(x,y) = \sqrt{n}\left(\widetilde{F}_n(x,y) - F(x,y)\right),$$

where

$$\widetilde{F}_n(x,y) = \frac{1}{n}\sum_{j=1}^{n} I(X_{j:n} \le x, V_j \le y),$$

with $V_j = \varepsilon_{j+1} - \varepsilon_j$. Clearly the V_j are 1-dependent. Now

$$\widetilde{F}_n(x,y) = \frac{1}{n}\sum_{j=1}^{n\hat{F}_X(x)} I(V_j \le y).$$

Write

$$\hat{G}_x(y) = \frac{1}{\lfloor nx \rfloor}\sum_{j=1}^{\lfloor nx \rfloor} I(V_j \le y),$$

and observe that $\widetilde{F}_n(x,y) = \hat{F}_X(x)\hat{G}_{\hat{F}_X(x)}(y)$ (and $\hat{G}_1 = \hat{G}$). Define also for $0 < z \le n$

$$\widetilde{\alpha}_{2z}(y) = \frac{1}{\sqrt{\lfloor z \rfloor}}\sum_{j=1}^{\lfloor z \rfloor} (I(V_j \le y) - G(y)) \quad (0/0 = 0)$$

and

$$Z_n(x,y) = \sqrt{\frac{\lfloor nx \rfloor}{n}}\widetilde{\alpha}_{2nx}(y), 0 \le x \le 1,\ y \in I\!R.$$

Observe that $Z_n(1,y) = \widetilde{\alpha}_{2n}(y)$. Since the V_j are 1-dependent, $Z_n(x,y)$ can be written as the sum of two dependent sequential empirical processes based on i.i.d. rv's, namely the odd and even indexed V_j, respectively. So Z_n is tight. It remains to prove the weak convergence of the finite dimensional distributions. Consider $(x_1, y_1), \ldots, (x_k, y_k)$, with $x_1 \le x_2 \le \cdots \le x_k$. By the Cramér-Wold device it suffices to consider linear combinations, i.e. $\sum_{r=1}^{k} a_r Z_n(x_r, y_r)$. Now using the central limit theorem for triangular arrays of m-dependent rv's and the fact that

$$\begin{aligned}
& \sum_{r=1}^{k} a_r Z_n(x_r, y_r) \\
= \; & \sum_{r=1}^{k} a_r Z_n(x_1, y_r) + \sum_{r=2}^{k} a_r (Z_n(x_2, y_r) - Z_n(x_1, y_r)) \\
& + \cdots + \sum_{r=k}^{k} a_r (Z_n(x_k, y_r) - Z_n(x_{k-1}, y_r)),
\end{aligned}$$

where these k terms are almost independent, we see that $\sum_{r=1}^{k} a_r Z_n(x_r, y_r)$ converges weakly. In summary, since g is bounded (use (2.6)), Z_n converges weakly on $D([0,1] \times$

$I\!R$) to a centered, uniformly continuous, bounded Gaussian process Z with covariance structure

$$
\begin{aligned}
& E\ Z(x_1,y_1)Z(x_{2,}y_2) \\
= \ & (x_1 \wedge x_2)(G(y_1 \wedge y_2) + H(y_1,y_2) + H(y_2,y_1) - 3G(y_1)G(y_2)).
\end{aligned}
$$

So $Var\ Z(x,y) = x(G(y) + 2H(y,y) - 3G^2(y))$. Obviously $H(y,y) \leq P(\varepsilon_2 - \varepsilon_1 \leq y) = G(y)$.

We have

$$
\begin{aligned}
\widetilde{\alpha}_n(x,y) \ = \ & \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n\hat{F}_X(x)}(I(V_j \leq y) - G(y))\right) \\
& + G(y)\sqrt{n}(\hat{F}_X(x) - F_X(x)).
\end{aligned} \tag{2.12}
$$

It is well known that $\sqrt{n}(\hat{F}_X - F_X)$ converges weakly to $B \circ F_X$, with B a Brownian bridge. We also have that $\sqrt{n}(\hat{F}_X - F_X)$ and Z_n are independent, and hence so are B and Z. Using the Skorohod construction (keeping the same notation for the new processes) we see that the right hand side of (2.12) is, almost surely, equal to

$$
\begin{aligned}
& Z(\hat{F}_X(x),y) + G(y)B(F_X(x)) + o(1) \\
= \ & Z(F_X(x),y) + G(y)B(F_X(x)) + o(1), \text{ uniformly in } x \text{ and } y.
\end{aligned}
$$

So $\{\tilde{\alpha}_n(x,y),\, x \in D_X,\, y \in I\!R\}$, converges weakly to

$$
\{Z(F_X(x),y) + G(y)B(F_X(x)),\, x \in D_X,\, y \in I\!R\}.
$$

Write $V(x,y) = Z(F_X(x),y) + G(y)B(F_X(x))$. Using this and the fact that V is uniformly continuous with respect to $d\left((x_1,y_1),(x_2,y_2)\right) = |F_X(x_1) - F_X(x_2)| + |y_1 - y_2|$, we see from (2.11), that α_n converges to the same limit, i.e. we have

$$
\alpha_n \xrightarrow{d} V. \tag{2.13}
$$

In particular we have that

$$
\sqrt{n}(\hat{F}_X - F_X) \xrightarrow{d} V(\cdot,\infty) \stackrel{d}{=} B \circ F_X, \tag{2.14}
$$

and, similarly, with $\alpha_{2n}(y) = \alpha_n(\infty, y)$,

$$\alpha_{2n} \xrightarrow{d} V(\infty, \cdot) \stackrel{d}{=} Z(1, \cdot). \tag{2.15}$$

Since

$$\begin{aligned}
& \sqrt{n}(F_n(x,y) - \hat{F}_X(x)\hat{G}(y)) \\
= \; & \sqrt{n}(F_n(x,y) - F(x,y)) - G(y)\sqrt{n}(\hat{F}_X(x) - F_X(x)) \\
& - \hat{F}_X(x)\sqrt{n}(\hat{G}(y) - G(y)),
\end{aligned}$$

we obtain from (2.13), (2.14), (2.15),

$$\begin{aligned}
& \sqrt{n}(F_n - \hat{F}_X\hat{G}) \xrightarrow{d} V - G(y)V(\cdot, \infty) - F_X(x)V(\infty, \cdot) \\
= \; & Z(F_X, \cdot) - F_X Z(1, \cdot) =: V_0.
\end{aligned} \tag{2.16}$$

Now we are prepared to prove (2.7)-(2.9). Statement (2.7) is immediate from (2.16) and statement (2.8) follows easily from (2.16) and the Helly-Bray theorem.

We are now going to prove (2.9). From (2.11) we have

$$\left.\begin{array}{l}
\sqrt{n}(F_n(x,y) - \hat{F}_X(x)\hat{G}(y)) \\
\leq \widetilde{\alpha}_n\left(x, y + \frac{\log^2 n}{n}\right) - G(y)\alpha_n(x, \infty) \\
\quad - \hat{F}_X(x)\widetilde{\alpha}_n\left(\infty, y - \frac{\log^2 n}{n}\right) + 2C\frac{\log^2 n}{\sqrt{n}}, \\
 \\
\sqrt{n}(F_n(x,y) - \hat{F}_X(x)\hat{G}(y)) \\
\geq \widetilde{\alpha}_n\left(x, y - \frac{\log^2 n}{n}\right) - G(y)\alpha_n(x, \infty) \\
\quad - \hat{F}_X(x)\widetilde{\alpha}_n\left(\infty, y + \frac{\log^2 n}{n}\right) - 2C\frac{\log^2 n}{\sqrt{n}}.
\end{array}\right\} \tag{2.17}$$

Set $V_{n,0} = \sqrt{n}(F_n - \hat{F}_X\hat{G})$. From (2.13) and (2.16), we have, using the Skorohod construction for (2.13) (but keeping the same notation),

$$\sup_{\substack{x \in D_X \\ y \in I\!R}} |\alpha_n(x,y) - V(x,y)| \to 0 \text{ a.s.} \tag{2.18}$$

and

$$\sup_{\substack{x \in D_X \\ y \in I\!R}} |V_{n,0}(x,y) - V_0(x,y)| \to 0 \text{ a.s.} \tag{2.19}$$

Set

$$M(x,y) = F_X(x)G(y)(1-F_X(x))(1-G(y))$$

and

$$\hat{M}(x,y) = \hat{F}_X(x)\hat{G}(y)(1-\hat{F}_X(x))(1-\hat{G}(y)).$$

Let $0 < \varepsilon < \frac{1}{4}$ be arbitrary and let $\delta(\varepsilon) > 0$ be a function of ε to be chosen later on, such that $\lim_{\varepsilon \downarrow 0} \delta(\varepsilon) = 0$. Denote with $q_{1\varepsilon}$ and $\widetilde{q}_{1\varepsilon}$ the $\delta(\varepsilon)$-th and $(1-\delta(\varepsilon))$-th quantiles of F_X, respectively, and with $q_{2\varepsilon}$, $\widetilde{q}_{2\varepsilon}$ the same quantiles of G. Write $S_\varepsilon = (q_{1\varepsilon}, \widetilde{q}_{1\varepsilon}) \times (q_{2\varepsilon}, \widetilde{q}_{2\varepsilon})$. We have

$$\begin{aligned}
&\left| \iint_{S_\varepsilon} \frac{V_{n,0}^2(x,y)}{\hat{M}(x,y)} d\hat{F}_X(x) d\hat{G}(y) - \iint_{S_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} dF_X(x) dG(y) \right| \\
\leq\ & \iint_{S_\varepsilon} \frac{|V_{n,0}^2(x,y) - V_0^2(x,y)|}{\hat{M}(x,y)} d\hat{F}_X(x) d\hat{G}(y) \\
& + \iint_{S_\varepsilon} \frac{|M(x,y) - \hat{M}(x,y)|}{\hat{M}(x,y) M(x,y)} V_0^2(x,y) d\hat{F}_X(x) d\hat{G}(y) \\
& + \left| \iint_{S_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} (d\hat{F}_X(x) d\hat{G}(y) - dF_X(x) dG(y)) \right|.
\end{aligned}$$

From (2.19) and (2.18) we now see that the first and second term on the right converge to 0 a.s. The a.s. convergence to 0 of the third term follows from the Helly-Bray theorem.

Set $A_\varepsilon = I\!R^2 \backslash S_\varepsilon$. In view of what we just proved, it is now sufficient for the proof of (2.9) to show that for large n

$$P\left(\iint_{A_\varepsilon} \frac{V_{n,0}^2(x,y)}{\hat{M}(x,y)} d\hat{F}_X(x) d\hat{G}(y) \geq 17\varepsilon \right) \leq 4\varepsilon \tag{2.20}$$

and

$$P\left(\iint_{A_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} dF_X(x) dG(y) \geq \varepsilon \right) \leq \varepsilon. \tag{2.21}$$

From (2.16) it follows by a straightforward computation that $EV_0^2(x,y) = F_X(x)(1 - F_X(x))(G(y) + 2H(y,y) - 3G^2(y))$. Hence

$$\begin{aligned}
\frac{EV_0^2(x,y)}{M(x,y)} &= \frac{G(y)(1-G(y)) + 2(H(y,y) - G^2(y))}{G(y)(1-G(y))} \\
&= 1 + 2\frac{H(y,y) - G^2(y)}{G(y)(1-G(y))} \leq 3.
\end{aligned}$$

So by the Markov inequality we obtain

$$
\begin{aligned}
& P\left(\iint_{A_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} dF_X(x)dG(y) \geq \varepsilon\right) \\
\leq\ & \frac{1}{\varepsilon} E \iint_{A_\varepsilon} \frac{V_0^2(x,y)}{M(x,y)} dF_X(x)dG(y) = \frac{1}{\varepsilon} \iint_{A_\varepsilon} \frac{EV_0^2(x,y)}{M(x,y)} dF_X(x)dG(y) \\
\leq\ & \frac{3}{\varepsilon} \iint_{A_\varepsilon} dF_X(x)dG(y) \leq \frac{12\delta(\varepsilon)}{\varepsilon} \leq \varepsilon,
\end{aligned}
$$

for $\delta(\varepsilon) \leq \varepsilon^2/12$. This is (2.21).

Set $B_\varepsilon = A_\varepsilon \cap ((-\infty, m_1) \times (-\infty, m_2))$, with m_1 and m_2 the medians of F_X and G, respectively. Because of symmetry and because of (2.14) and (2.15), (2.20) follows from

$$
P\left(\iint_{B_\varepsilon} \frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x)d\hat{G}(y) \geq \varepsilon\right) \leq \varepsilon, \tag{2.22}
$$

for large n. Let Q_{1n} and Q_{2n} be the empirical quantile functions corresponding to $\hat{F}_X$ and $\hat{G}$ respectively, and set $a_n = \frac{1}{n^{3/4}}$. Trivially

$$
\begin{aligned}
& n \int_{-\infty}^{Q_{2n}(a_n)} \int_{-\infty}^{Q_{1n}(a_n)} \frac{(F_n(x,y) - \hat{F}_X(x)\hat{G}(y))^2}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x)d\hat{G}(y) \\
\leq\ & 2n \int_{-\infty}^{Q_{2n}(a_n)} \int_{-\infty}^{Q_{1n}(a_n)} \frac{F_n^2(x,y) + \hat{F}_X^2(x)\hat{G}^2(y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x)d\hat{G}(y) \\
\leq\ & 4n\hat{F}_X(Q_{1n}(a_n))\hat{G}(Q_{2n}(a_n)) \overset{a.s.}{\leq} 4n\left(a_n + \frac{1}{n}\right)^2 \to 0 \quad (n \to \infty).
\end{aligned}
$$

Now assume $x \geq Q_{1n}(a_n)$, $y \leq Q_{2n}(a_n)$ and $(x,y) \in B_\varepsilon$. We have

$$
\frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} \leq 2n\left(\frac{\hat{G}(y)}{\hat{F}_X(x)} + \hat{F}_X(x)\hat{G}(y)\right).
$$

Hence

$$
\begin{aligned}
&\int\limits_{Q_{1n}(a_n)}^{m_1} \int\limits_{-\infty}^{Q_{2n}(a_n)} \frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{G}(y) d\hat{F}_X(x) \\
\leq\ & 2n \int\limits_{Q_{1n}(a_n)}^{m_1} \left(\frac{1}{\hat{F}_X(x)} + \hat{F}_X(x) \right) d\hat{F}_X(x) \int\limits_{-\infty}^{Q_{2n}(a_n)} \hat{G}(y) d\hat{G}(y) \\
\leq\ & n \left(1 + \log \frac{1}{\hat{F}_X(Q_{1n}(a_n))} \right) \hat{G}^2(Q_{2n}(a_n)) \\
\leq\ & n \left(1 + \log \frac{1}{a_n} \right) \left(a_n + \frac{1}{n} \right)^2 \to 0 \ \text{ a.s.}
\end{aligned}
$$

Similarly we can deal with the region where $x \leq Q_{1n}(a_n)$, $y \geq Q_{2n}(a_n)$ and $(x,y) \in B_\varepsilon$. Set $T_\varepsilon = \{(x,y) \in B_\varepsilon : x \leq Q_{1n}(a_n) \text{ or } y \leq Q_{2n}(a_n)\}$. We have now shown that

$$
P\left(\iint_{T_\varepsilon} \frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y) \geq \frac{1}{2}\varepsilon \right) \leq \frac{1}{2}\varepsilon.
$$

So it remains to consider that part of the integral in (2.22) where $Q_{1n}(a_n) \leq x \leq q_{1\varepsilon}$ and $Q_{2n}(a_n) \leq y \leq m_2$, or $Q_{2n}(a_n) \leq y \leq q_{2\varepsilon}$ and $Q_{1n}(a_n) \leq x \leq m_{1.}$ This part of the proof is the crucial part. The proof for both regions, however, is rather similar, so we confine ourselves to studying the integral over the latter region, i.e. we will show

$$
P\left(\int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \int\limits_{Q_{1n}(a_n)}^{m_1} \frac{V_{n,0}^2(x,y)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y) \geq \frac{1}{2}\varepsilon \right) \leq \frac{1}{2}\varepsilon. \tag{2.23}
$$

Because of (2.17) and the fact that $(a+b+c+d)^2 \leq 4(a^2+b^2+c^2+d^2)$, it suffices to prove (2.23) with $V_{n,0}$ replaced by $\widetilde{\alpha}_n\left(x, y \pm \frac{\log^2 n}{n}\right)$, $G(y)\alpha_n(x,\infty)$, $\hat{F}_X(x)\widetilde{\alpha}_n\left(\infty, y \pm \frac{\log^2 n}{\sqrt{n}}\right)$, $2C\frac{\log^2 n}{\sqrt{n}}$, respectively. Denote the resulting statements as (2.23a$\pm$), (2.23b), (2.23c$\pm$), (2.23d). For (2.23d) we have

$$
\begin{aligned}
&\int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \int\limits_{Q_{1n}(a_n)}^{m_1} \frac{4C^2 \log^4 n}{n\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y) \\
\leq\ & \frac{4C^2 \log^4 n}{n} \log \frac{1}{\hat{F}_X(Q_{1n}(a_n))} \log \frac{1}{\hat{G}(Q_{2n}(a_n))} \overset{a.s.}{\leq} \frac{4C^2 \log^6 n}{n} \to 0.
\end{aligned}
$$

For (2.23b) we see, writing $\alpha_{1n}(x) = \alpha_n(x, \infty)$,

$$\int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \int\limits_{Q_{1n}(a_n)}^{m_1} \frac{(G(y)\alpha_{1n}(x))^2}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y)$$

$$= \int\limits_{Q_{1n}(a_n)}^{m_1} \frac{\alpha_{1n}^2(x)}{\hat{F}_X(x)} d\hat{F}_X(x) \int\limits_{Q_{2n}(a_n)}^{q_{1\varepsilon}} \frac{G^2(y)}{\hat{G}(y)} d\hat{G}(y)$$

$$\leq \left(\sup_{Q_{1n}(a_n) \leq x \leq m_1} \frac{|\alpha_{1n}(x)|}{\hat{F}_X^{\frac{1}{4}}(x)} \right)^2 \int\limits_{Q_{1n}(a_n)}^{m_1} \frac{1}{\hat{F}_X^{\frac{1}{2}}(x)} d\hat{F}_X(x) \int\limits_{Q_{2n}(a_n)}^{q_{1\varepsilon}} \frac{G^2(y)}{\hat{G}(y)} d\hat{G}(y)$$

$$= O_P(1) \int\limits_{Q_{2n}(a_n)}^{q_{1\varepsilon}} \hat{G}(y) d\hat{G}(y) = O_P(1) \cdot \delta^2(\varepsilon).$$

For (2.23c+) we obtain ((2.23c−) can be dealt with similarly)

$$\int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \int\limits_{Q_{1n}(a_n)}^{m_1} \frac{\hat{F}_X^2(x) \widetilde{\alpha}_{2n}^2 \left(y + \frac{\log^2 n}{n} \right)}{\hat{F}_X(x)\hat{G}(y)} d\hat{F}_X(x) d\hat{G}(y)$$

$$= \int\limits_{Q_{1n}(a_n)}^{m_1} \hat{F}_X(x) d\hat{F}_X(x) \int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \frac{\widetilde{\alpha}_{2n}^2 \left(y + \frac{\log^2 n}{n} \right)}{\hat{G}(y)} d\hat{G}(y)$$

$$\leq \int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \frac{\widetilde{\alpha}_{2n}^2 \left(y + \frac{\log^2 n}{n} \right)}{G\left(y + \frac{\log^2 n}{n} \right)} \frac{G\left(y + \frac{\log^2 n}{n} \right)}{\hat{G}(y)} d\hat{G}(y)$$

$$= O_P(1) \left(\sup_{Q_{2n}(a_n) \leq y \leq q_{2\varepsilon}} \frac{\left| \widetilde{\alpha}_{2n} \left(y + \frac{\log^2 n}{n} \right) \right|}{G^{\frac{1}{4}} \left(y + \frac{\log^2 n}{n} \right)} \right)^2 \int\limits_{Q_{2n}(a_n)}^{q_{2\varepsilon}} \frac{1}{G^{\frac{1}{2}} \left(y + \frac{\log^2 n}{n} \right)} d\hat{G}(y)$$

$$= O_P(1) \hat{G}^{\frac{1}{2}}(q_{2\varepsilon}) = O_P(1) \delta^{\frac{1}{2}}(\varepsilon).$$

Here and in the sequel it is used that (for even n) $\hat{G}$ can be written as the average of two dependent empirical df's of the $\frac{n}{2}$ odd an even indexed V_j, respectively. Since the V_j are 1-dependent, these two empirical df's are based on i.i.d. random variables with df G. Hence we can use the well-known properties of classical empirical df's and processes, in particular weak convergence of weighted empirical processes and results on the ratio of the empirical and the true df.

Remains (2.23a+) ((2.23a−) can be dealt with similarly). In accordance with the above we let

$$\begin{aligned}\widetilde{F}_{n(1)}(x,y) &= \frac{2}{n}\sum_{\substack{j=1\\ j\ odd}}^{n} I(X_{j:n}\le x, V_j\le y)\\ &= \frac{2}{n}\sum_{j=1}^{\frac{n}{2}} I(X_{2j-1:n}\le x, V_{2j-1}\le y)\end{aligned}$$

and

$$\widetilde{\alpha}_{n(1)}(x,y)=\sqrt{\frac{n}{2}}(\widetilde{F}_{n(1)}(x,y)-F_X(x)G(y));$$

$\widetilde{F}_{n(2)}$ and $\widetilde{\alpha}_{n(2)}$ are defined similarly for the even indices. Clearly for (2.23a+) , it is sufficient (use $(a+b)^2\le 2(a^2+b^2)$) to consider $\widetilde{\alpha}_{n(1)}\left(x,y+\frac{\log^2 n}{n}\right)$ instead of $\widetilde{\alpha}_n\left(x,y+\frac{\log^2 n}{n}\right)$; $\widetilde{\alpha}_{n(2)}$ can be dealt with similarly. So we need to show

$$P\left(\int_{Q_{2n}(a_n)}^{q_{2\varepsilon}}\int_{Q_{1n}(a_n)}^{m_1}\frac{\widetilde{\alpha}^2_{n(1)}\left(x,y+\frac{\log^2 n}{n}\right)}{\hat{F}_X(x)\hat{G}(y)}d\hat{F}_X(x)d\hat{G}(y)\ge\frac{1}{4}\varepsilon\right)\le\frac{1}{4}\varepsilon. \tag{2.24}$$

Set $\hat{F}_{X(1)}(x)=\widetilde{F}_{n(1)}(x,\infty)=\frac{2}{n}\sum_{j=1}^{\frac{n}{2}}I(X_{2j-1:n}\le x)$ and

$$\hat{G}_{u(1)}(y)=\frac{1}{\lfloor\frac{n}{2}u\rfloor}\sum_{j=1}^{\lfloor\frac{n}{2}u\rfloor}I(V_{2j-1}\le y)\quad(0/0=0)\,.$$

Observe that

$$\widetilde{F}_{n(1)}(x,y)=\hat{F}_{X(1)}(x)\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y).$$

So, with $y_n=y+\frac{\log^2 n}{n}$,

$$\begin{aligned}\widetilde{\alpha}_{n(1)}(x,y_n) &= \sqrt{\frac{n}{2}}(\hat{F}_{X(1)}(x)\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y_n)-F_X(x)G(y_n))\\ &= \sqrt{\frac{n}{2}}\hat{F}_{X(1)}(x)(\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y_n)-G(y_n))\\ &\quad+\sqrt{\frac{n}{2}}G(y_n)(\hat{F}_{X(1)}(x)-F_X(x)).\end{aligned} \tag{2.25}$$

Since $\hat{F}_X(x)-\frac{1}{n}\le\hat{F}_{X(1)}(x)\le\hat{F}_X(x)$, for all x, we see similar as for (2.23b), that the latter term contributes $O_P(1)\delta^2(\varepsilon)$, to the integral in (2.24), where we again used

$(a+b)^2 \le 2(a^2+b^2)$. So it remains to consider the contribution of the first term. $\left(\left\lfloor \frac{n}{2}u \right\rfloor / \sqrt{\frac{n}{2}}\right)(\hat{G}_{u(1)}(y) - G(y))$ is the sequential empirical process based on $\frac{n}{2}$ i.i.d random variables with df G. Using the Komlós, Major and Tusnády (1975) Kiefer process approximation of the sequential empirical process, we have that there exists a Kiefer process K such that

$$\sup_{\substack{0\le u\le 1 \\ y\in \mathbb{R}}} \left| \left(\left\lfloor \frac{n}{2}u \right\rfloor / \sqrt{\frac{n}{2}} \right) (\hat{G}_{u(1)}(y) - y) - \frac{K\left(G(y), \left\lfloor \frac{n}{2}u \right\rfloor\right)}{\sqrt{\frac{n}{2}}} \right| \stackrel{a.s.}{=} O\left(\frac{\log^2 n}{\sqrt{n}}\right).$$

Hence

$$\sup_{\substack{x\in D_X \\ y\in \mathbb{R}}} \left| \sqrt{\frac{n}{2}}\ \hat{F}_{X(1)}(x)(\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y) - G(y)) - \frac{K(G(y), \frac{n}{2}\hat{F}_{X(1)}(x))}{\sqrt{\frac{n}{2}}} \right|$$
$$\stackrel{a.s.}{=} O\left(\frac{\log^2 n}{\sqrt{n}}\right). \tag{2.26}$$

Since $|\hat{F}_{X(1)}(x) - \hat{F}_X(x)| \le \frac{1}{n}$, for all x, we have by the law of the iterated logarithm for the ordinary empirical process that

$$\sup_{x\in D_X} |\hat{F}_{X(1)}(x) - \hat{F}_X(x)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{a.s.}$$

Using this, the relation between a Kiefer process and a bivariate Wiener process and Lemma 1.11.1 in Csőrgő and Révész (1981) on the oscillations of a bivariate Wiener process it follows immediately that

$$\sqrt{\frac{2}{n}} \sup_{\substack{x\in D_X \\ y\in \mathbb{R}}} \left| K(G(y), \frac{n}{2}\hat{F}_{X(1)}(x)) - K(G(y), \frac{n}{2}F_X(x)) \right| = O\left(\frac{1}{n^{1/5}}\right) \quad \text{a.s.} \tag{2.27}$$

Since $(K(x, \frac{n}{2}y))_{x,y} \stackrel{d}{=} (\sqrt{\frac{n}{2}}K(x,y))_{x,y}$, we obtain from F and G that for a Kiefer process $\widetilde{K}$

$$\sup_{\substack{x\in D_X \\ y\in \mathbb{R}}} \left| \sqrt{\frac{n}{2}}\hat{F}_{X(1)}(x)(\hat{G}_{\hat{F}_{X(1)}(x)(1)}(y) - G(y)) - \widetilde{K}(G(y), F_X(x)) \right| \stackrel{a.s.}{=} O\left(\frac{1}{n^{1/5}}\right). \tag{2.28}$$

Now we turn back to (2.24), and in particular to the first term on the right in (2.25). From (2.28) we see that this term can be replaced by $\widetilde{K}(G(y_n), F_X(x))$; the remaining

$O\left(\frac{1}{n^{1/5}}\right)$ can be treated as (2.23d) before. So it remains to consider

$$\int\limits_{Q_2(a_n)}^{q_{2\varepsilon}} \int\limits_{Q_{1n}(a_n)}^{m_1} \frac{\widetilde{K}^2(G(y_n), F_X(x))}{F_X^{\frac{1}{2}}(x) G^{\frac{1}{2}}(y_n)} \frac{1}{\hat{F}_X^{\frac{1}{2}}(x)\hat{G}^{\frac{1}{2}}(y)} d\hat{F}_X(x) d\hat{G}(y)$$

$$\leq \left(\sup_{\substack{x \geq Q_{1n}(a_n) \\ y \geq Q_{2n}(a_n)}} \frac{\widetilde{K}(G(y_n), F_X(x))}{(F_X(x)G(y_n))^{\frac{1}{4}}} \right)^2 \hat{F}_X^{\frac{1}{2}}(m_1) \hat{G}^{\frac{1}{2}}(q_{2\varepsilon}). \tag{2.29}$$

It follows from Corollary 1.12.2 in Csőrgő and Révész (1981) that the first term on the right of (2.29) is $O_P(1)$, so the right hand side of (2.29) is $O_P(1) \cdot \delta^{\frac{1}{2}}(\varepsilon)$.

Collecting everything we proved for (2.24) and (2.23), we now see that for a proper, small enough choice of $\delta(\varepsilon)$, we indeed proved (2.24) and (2.23). □

3 The Estimation Approach

Recall $F_X(x) = P(X \leq x)$, $F_\varepsilon(y) = P(\varepsilon \leq y)$, $F(y|x) = P(Y \leq y | X = x)$ and set $F_{X,\varepsilon}(x, y) = P(X \leq x, \varepsilon \leq y)$. The probability density functions of these distributions will be denoted with lower case letters. In this section we assume throughout that

$$m(x) = \int\limits_0^1 F^{-1}(s|x) J(s)\, ds, \tag{3.1}$$

where $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ is the quantile function of Y given x and $J(s)$ is a given score function satisfying $\int_0^1 J(s)\, ds = 1$. (E.g., the choice $J \equiv 1$ leads to $m(x) = E(Y|x)$.)

The proposed tests will be based on the difference $\hat{F}_{X,\hat{\varepsilon}}(x, y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)$ for appropriate estimators $\hat{F}_{\hat{\varepsilon}}$ and $\hat{F}_{X,\hat{\varepsilon}}$ of F_ε and $F_{X,\varepsilon}$ respectively. (The distribution function F_X of X is estimated, as in Section 2, by its empirical df $\hat{F}_X$.) To estimate the distribution of ε, first estimate $m(x)$ by

$$\hat{m}(x) = \int\limits_0^1 \hat{F}^{-1}(s|x) J(s)\, ds, \tag{3.2}$$

where

$$\hat{F}(y|x) = \sum_{i=1}^n W_i(x, a_n) I(Y_i \leq y) \tag{3.3}$$

is the Stone (1977) estimator and the $W_i(x,a_n)$ $(i=1,\ldots,n)$ are the Nadaraya-Watson weights

$$W_i(x,a_n)=\frac{K\left(\frac{x-X_i}{a_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{a_n}\right)}$$

(with K a given kernel function and $(a_n)_{n\in I\!N}$ a bandwidth sequence). Now define $\hat{\varepsilon}_i = Y_i-\hat{m}(X_i)$ for the resulting residuals, and let

$$\hat{F}_{\hat{\varepsilon}}(y)=n^{-1}\sum_{i=1}^n I(\hat{\varepsilon}_i\le y). \tag{3.4}$$

Finally, $F_{X,\varepsilon}(x,y)$ is estimated by

$$\hat{F}_{X,\hat{\varepsilon}}(x,y)=\frac{1}{n}\sum_{i=1}^n I(X_i\le x,\hat{\varepsilon}_i\le y).$$

To test the null hypothesis, we define the following test statistics, which are the natural analogues of the difference based statistics $T_{n,KS}, T_{n,CM}$ and $T_{n,AD}$ of Theorem 2.1:

$$\tilde{T}_{n,KS}=\sqrt{n}\ \sup|\hat{F}_{X,\hat{\varepsilon}}(x,y)-\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)|, \tag{3.5}$$

$$\tilde{T}_{n,CM}=n\iint(\hat{F}_{X,\hat{\varepsilon}}(x,y)-\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y))^2\,d\hat{F}_X(x)\,d\hat{F}_{\hat{\varepsilon}}(y), \tag{3.6}$$

$$\tilde{T}_{n,AD}=n\iint\frac{(\hat{F}_{X,\hat{\varepsilon}}(x,y)-\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y))^2}{\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)(1-\hat{F}_X(x))(1-\hat{F}_{\hat{\varepsilon}}(y))}\,d\hat{F}_X(x)\,d\hat{F}_{\hat{\varepsilon}}(y). \tag{3.7}$$

In the first theorem we obtain an i.i.d. representation for the difference $\hat{F}_{X,\hat{\varepsilon}}(x,y)-\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)$ $(x\in D_X, y\in I\!R)$ (weighted in an appropriate way), on which all three test statistics are based. Based on this result, the weak convergence will then be established. The assumptions mentioned below are given in the Appendix.

Theorem 3.1 *Assume (A),(K),(J) and (F). Then, under H_0, for $0\le\beta<\frac{1}{2}$,*

$$\sup_{\substack{x\in D_X\\ y\in I\!R}}\left|\frac{\hat{F}_{X,\hat{\varepsilon}}(x,y)-\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)}{N(x,y)^\beta}\right.$$

$$\left.-\quad\frac{(\hat{F}_{X,\varepsilon}(x,y)-F_{X,\varepsilon}(x,y))-F_X(x)(\hat{F}_\varepsilon(y)-F_\varepsilon(y))-F_\varepsilon(y)(\hat{F}_X(x)-F_X(x))}{N(x,y)^\beta}\right|=o_P(n^{-1/2}),$$

with

$$\hat{F}_{X,\varepsilon}(x,y) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x, \varepsilon_i \le y),$$

$$\hat{F}_{\varepsilon}(y) = \frac{1}{n}\sum_{i=1}^{n} I(\varepsilon_i \le y),$$

and

$$N(x,y) = F_X(x)F_\varepsilon(y)(1-F_X(x))(1-F_\varepsilon(y)).$$

Proof Write

$$\begin{aligned}
&N(x,y)^{-\beta}\{[\hat{F}_{X,\hat{\varepsilon}}(x,y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)] - [F_{X,\varepsilon}(x,y) - F_X(x)F_\varepsilon(y)]\} \\
&= N(x,y)^{-\beta}\{[\hat{F}_{X,\hat{\varepsilon}}(x,y) - F_{X,\varepsilon}(x,y)] - F_X(x)[\hat{F}_{\hat{\varepsilon}}(y) - F_\varepsilon(y)] - \hat{F}_{\hat{\varepsilon}}(y)[\hat{F}_X(x) - F_X(x)]\} \\
&= N(x,y)^{-\beta}\,(\hat{F}_{X,\varepsilon}(x,y) - F_{X,\varepsilon}(x,y)) - F_X(x)(\hat{F}_\varepsilon(y) - F_\varepsilon(y)) - F_\varepsilon(y)(\hat{F}_X(x) - F_X(x)) \\
&\quad + N(x,y)^{-\beta}\,n^{-1}\sum_{i=1}^{n}[I(X_i \le x) - F_X(x)][I(\hat{\varepsilon}_i \le y) - I(\varepsilon_i \le y)] \\
&\quad - N(x,y)^{-\beta}[\hat{F}_X(x) - F_X(x)][\hat{F}_{\hat{\varepsilon}}(y) - F_\varepsilon(y)]. \qquad (3.8)
\end{aligned}$$

From Lemma A.1 it follows that the second term on the right hand side of (3.8) is equal to (using the notation of that lemma)

$$N(x,y)^{-\beta}(\hat{F}_X(x) - F_X(x))(F_{\hat{\varepsilon}}(y) - F_\varepsilon(y)) + o_P(n^{-1/2}),$$

uniformly in x and y. This term is $o_P(n^{-1/2})$, since by the Chibisov-O'Reilly theorem,

$$\sup_x \left|\frac{\hat{F}_X(x) - F_X(x)}{[F_X(x)(1-F_X(x))]^\beta}\right| = O_P(n^{-1/2}),$$

and since

$$\sup_y \left|\frac{F_{\hat{\varepsilon}}(y) - F_\varepsilon(y)}{[F_\varepsilon(y)(1-F_\varepsilon(y))]^\beta}\right| = o_P(1), \qquad (3.9)$$

which can be shown in a similar way as in the beginning of the proof of Lemma A.1. Using again Lemma A.1, the third term of (3.8) can be written as

$$\begin{aligned}
&N(x,y)^{-\beta}[\hat{F}_X(x) - F_X(x)][\{\hat{F}_\varepsilon(y) - F_\varepsilon(y)\} + \{F_{\hat{\varepsilon}}(y) - F_\varepsilon(y)\}] + o_P(n^{-1/2}) \\
&= o_P(n^{-1/2}),
\end{aligned}$$

uniformly in x and y. Hence, the result follows. $\square$

The next result follows readily from Theorem 3.1, by using standard empirical process theory.

Theorem 3.2 *Assume (A),(K),(J) and (F). Let W_0 be a 4-sided tied-down Wiener process on $[0,1]^2$, defined by $W_0(u,v) = W(u,v) - uW(1,v) - vW(u,1) + uvW(1,1)$, $u,v \in [0,1]$, where W is a standard bivariate Wiener process. Under H_0, for $0 \leq \beta < \frac{1}{2}$, the process*

$$\sqrt{n}\,\frac{\hat{F}_{X,\hat{\varepsilon}}(x,y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)}{N(x,y)^\beta}, \quad x \in D_X, y \in I\!R,$$

converges weakly to $W_0(F_X(x), F_\varepsilon(y))/N(x,y)^\beta$.

As a consequence, we find the limiting distribution of the three test statistics. Note that these limits are distribution-free and identical to the ones in the classical case, i.e. when m is not estimated, but known.

Theorem 3.3 *Assume (A),(K),(J) and (F). Then, under H_0,*

$$\tilde{T}_{n,KS} \xrightarrow{d} \sup_{0<u,v<1} |W_0(u,v)|,$$

$$\tilde{T}_{n,CM} \xrightarrow{d} \iint W_0^2(u,v)\,dudv,$$

$$\tilde{T}_{n,AD} \xrightarrow{d} \iint \frac{W_0^2(u,v)}{uv(1-u)(1-v)}\,dudv.$$

Proof The result for $\tilde{T}_{n,KS}$ follows readily from Theorem 3.2 and the continuous mapping theorem. The result for $\tilde{T}_{n,CM}$ follows from Theorem 3.2, Lemma A.1, (3.9) and the Helly-Bray theorem.

Now we present the proof for $\tilde{T}_{n,AD}$. From the Skorohod construction and Theorem 3.2 it follows that (keeping the same notation for the new processes)

$$\sup_{x,y} \left| \frac{\sqrt{n}\hat{T}(x,y) - W_0(F_X(x), F_\varepsilon(y))}{N(x,y)^\beta} \right| \xrightarrow{P} 0, \tag{3.10}$$

where $\hat{T}(x,y) = \hat{F}_{X,\hat{\varepsilon}}(x,y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)$ and $0 \le \beta < 1/2$. In what follows we will show that

$$\iint \frac{n\hat{T}^2(x,y)}{\hat{N}(x,y)}\, d\hat{F}_X(x)\, d\hat{F}_{\hat{\varepsilon}}(y) - \iint \frac{W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)}\, dF_X(x)\, dF_\varepsilon(y) \xrightarrow{P} 0, \qquad (3.11)$$

where $\hat{N}(x,y) = \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)(1-\hat{F}_X(x))(1-\hat{F}_{\hat{\varepsilon}}(y))$. Define $A_n = (\hat{F}_X^{-1}(n^{-3/4}), \hat{F}_X^{-1}(1-n^{-3/4})) \times (\hat{F}_{\hat{\varepsilon}}^{-1}(n^{-3/4}), \hat{F}_{\hat{\varepsilon}}^{-1}(1-n^{-3/4}))$. The left hand side of (3.11) can be written as

$$\begin{aligned}
&\iint_{A_n} \frac{n\hat{T}^2(x,y) - W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)^{1/3}} \frac{N(x,y)^{1/3}}{\hat{N}(x,y)^{1/4}} \frac{d\hat{F}_X(x)\, d\hat{F}_{\hat{\varepsilon}}(y)}{\hat{N}(x,y)^{3/4}} \\
&+ \iint_{A_n} \frac{N(x,y) - \hat{N}(x,y)}{N(x,y)^{2/5}} \frac{W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)^{4/5}} \frac{N(x,y)^{1/5}}{\hat{N}(x,y)^{1/6}} \frac{d\hat{F}_X(x)\, d\hat{F}_{\hat{\varepsilon}}(y)}{\hat{N}(x,y)^{5/6}} \\
&+ \iint_{A_n} \frac{W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)} d\hat{F}_X(x)\, d\hat{F}_{\hat{\varepsilon}}(y) - \iint_{D_X \times I\!R} \frac{W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)} dF_X(x)\, dF_\varepsilon(y) \\
&+ \iint_{A_n^c} \frac{n\hat{T}^2(x,y)}{\hat{N}(x,y)}\, d\hat{F}_X(x)\, d\hat{F}_{\hat{\varepsilon}}(y) \\
&= \sum_{i=1}^{5} T_i.
\end{aligned}$$

The term T_1 is $o_P(1)$ by (3.10) and Lemma A.2. For showing that $T_2 = o_P(1)$ use is made of Lemmas A.1 and A.2 and the Chibisov-O'Reilly theorem. The convergence in probability to 0 of $T_3 + T_4$ follows from the Helly-Bray theorem. Remains to consider T_5. We will only show that

$$\iint_{B_n} \frac{n\hat{T}^2(x,y)}{\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)}\, d\hat{F}_X(x)\, d\hat{F}_{\hat{\varepsilon}}(y) \xrightarrow{P} 0,$$

where B_n is the intersection of A_n^c and $(-\infty, m_1) \times (-\infty, m_\varepsilon)$, with m_1 and m_ε the medians of F_X and F_ε, respectively; the other parts can be dealt with similarly. First consider (with $c_n = \hat{F}_X^{-1}(n^{-3/4}), d_n = \hat{F}_{\hat{\varepsilon}}^{-1}(n^{-3/4})$)

$$\begin{aligned}
&\int_{-\infty}^{c_n} \int_{-\infty}^{d_n} \frac{n\hat{T}^2(x,y)}{\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)}\, d\hat{F}_{\hat{\varepsilon}}(y)\, d\hat{F}_X(x) \\
&\le 2n \int_{-\infty}^{c_n} \int_{-\infty}^{d_n} \frac{\hat{F}_{X,\hat{\varepsilon}}^2(x,y) + \hat{F}_X^2(x)\hat{F}_{\hat{\varepsilon}}^2(y)}{\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)}\, d\hat{F}_{\hat{\varepsilon}}(y)\, d\hat{F}_X(x) \\
&\le 4n\hat{F}_X(c_n)\hat{F}_{\hat{\varepsilon}}(d_n) \\
&\le 4n(n^{-3/4} + n^{-1})^2 \to 0.
\end{aligned}$$

Next, consider

$$\begin{aligned}
&\int_{c_n}^{m_1}\int_{-\infty}^{d_n}\frac{n\hat{T}^2(x,y)}{\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)}\,d\hat{F}_{\hat{\varepsilon}}(y)\,d\hat{F}_X(x)\\
&\leq 2n\int_{c_n}^{m_1}\int_{-\infty}^{d_n}\Big(\frac{\hat{F}_{\hat{\varepsilon}}(y)}{\hat{F}_X(x)}+\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)\Big)\,d\hat{F}_{\hat{\varepsilon}}(y)\,d\hat{F}_X(x)\\
&\leq n(\log n^{3/4}+1)(n^{-3/4}+n^{-1})^2\to 0.
\end{aligned}$$

Finally, the integral over $(-\infty, c_n)\times(d_n, m_\varepsilon)$ can be dealt with in a similar way. □

4 Simulations

Suppose that X has a uniform distribution on the interval $[0,1]$, that $m(x)=E(Y|X=x)=x-0.5x^2$ and suppose that, under the null hypothesis, ε is a zero-mean normal random variable with a standard deviation of 0.1. The simulations are carried out for samples of sizes $n=50,100,200$ and 400 and the significance level $\alpha=0.05$. Each simulation consists of 2000 replications. (Note that if m is not the conditional mean of Y given X, but another location functional (like trimmed mean or median), then the simulation results under H_0 remain valid, and when H_0 is violated, only minor changes are to be expected.)

Consider the following alternative hypotheses:

$$H_{1,A}:\varepsilon|X=x\sim N(0,\frac{1+ax}{100}),$$

with $a>0$. Also, let

$$H_{1,B}:\varepsilon|X=x\overset{d}{=}\frac{W_x-r_x}{10\sqrt{2r_x}},$$

where $W_x\sim\chi^2_{r_x}$, $r_x=1/(ax)$ and $a>0$ controls the skewness of the distribution. Note that the first and second moment of the variable ε created in the latter way do not depend on x and coincide with the respective moments under H_0. When a tends to 0, the distribution of $\varepsilon|X=x$ converges to its null-distribution, since it is well known that a standardized χ^2_r-distribution converges to a normal distribution when $r\to\infty$. Finally, let

$$H_{1,C}:\varepsilon|X=x\sim\frac{1}{10}\sqrt{1-(ax)^{1/4}}\,t_{2/(ax)^{1/4}},$$

where $0 < a \leq 1$ is a parameter controlling the kurtosis of the distribution. By construction, the conditional moments up to order three of ε given X are constant and coincide with the respective moments under the null hypothesis, while the fourth conditional moment does depend on X. The distribution of ε under $H_{1,C}$ converges to the null-distribution of ε when a tends to 0. It should be emphasized that tests for homoscedasticity fail to detect the dependence of ε and X for the latter two alternative hypotheses.

Table 1 shows the results of the simulations under $H_{1,A}$ where we used the difference based testing procedures of Section 2. We observe that the empirical α-levels (see $a = 0$) are reasonably close to their nominal value of 0.05 and that the method succeeds well in detecting dependence of the second moment of ε on X. For $H_{1,B}$ we considered the method described in Remark 2.1, i.e. we used triplets of the form $Y_3 - 2Y_2 + Y_1$ instead of differences $Y_2 - Y_1$, as it is explained in Remark 2.1 that the latter procedure is insensitive to dependence of the third moment of ε on X. The result shown in Table 2 leads to similar conclusions as for Table 1. Finally, the first, difference based, testing procedure works also well for detecting dependence on the fourth moment of ε on X as defined in $H_{1,C}$; see Table 3. In most cases, the Anderson-Darling statistic is the most powerful of the three test procedures, closely followed by the Cramér-von Mises test, while the Kolmogorov-Smirnov test is usually the least powerful procedure.

a	$n = 50$			$n = 100$			$n = 200$			$n = 400$		
	KS	CM	AD	KS	CM	AD	KS	CM	AD	KS	CM	AD
0	.051	.086	.058	.061	.067	.070	.060	.049	.077	.055	.043	.074
1	.067	.117	.096	.086	.147	.184	.187	.255	.346	.310	.431	.603
2.5	.097	.205	.183	.203	.330	.407	.428	.608	.724	.740	.891	.957
5	.144	.314	.285	.347	.545	.632	.684	.857	.933	.954	.991	.998
10	.194	.427	.393	.487	.719	.785	.858	.961	.989	.994	1.00	1.00
100	.324	.606	.572	.715	.906	.946	.977	1.00	1.00	1.00	1.00	1.00

Table 1: *Power of* $T_{n,KS}$, $T_{n,CM}$ *and* $T_{n,AD}$ *under* $H_{1,A}$.

Other simulations have shown that the second method (based on estimating m) is quite sensitive to the choice of the bandwidth, in the sense that the empirical α-levels depend heavily on the bandwidth and can be far from the nominal level for a wide range of

a	$n=50$			$n=100$			$n=200$			$n=400$		
	KS	CM	AD	KS	CM	AD	KS	CM	AD	KS	CM	AD
0.0	.022	.055	.051	.029	.044	.059	.039	.050	.086	.046	.042	.086
1	.070	.125	.107	.129	.158	.172	.242	.265	.306	.394	.452	.506
2.5	.172	.248	.199	.335	.403	.388	.575	.632	.636	.876	.912	.922
5	.340	.459	.371	.602	.694	.676	.869	.919	.918	.994	.997	.998
10	.474	.634	.563	.831	.892	.889	.981	.990	.990	1.00	1.00	1.00
100	.653	.857	.848	.967	.993	.996	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: *Power of* $T_{n,KS}$, $T_{n,CM}$ *and* $T_{n,AD}$ *under* $H_{1,B}$.

a	$n=50$			$n=100$			$n=200$			$n=400$		
	KS	CM	AD	KS	CM	AD	KS	CM	AD	KS	CM	AD
0.0	.051	.086	.058	.061	.067	.070	.060	.049	.077	.055	.043	.074
0.2	.090	.139	.101	.094	.116	.122	.133	.149	.195	.177	.208	.282
0.4	.115	.183	.138	.141	.181	.202	.225	.283	.344	.375	.461	.558
0.6	.161	.254	.200	.230	.314	.341	.402	.504	.559	.667	.785	.834
0.8	.246	.391	.337	.410	.563	.576	.674	.801	.847	.936	.980	.991
1.0	.465	.705	.631	.786	.927	.936	.973	.997	.998	1.00	1.00	1.00

Table 3: *Power of* $T_{n,KS}$, $T_{n,CM}$ *and* $T_{n,AD}$ *under* $H_{1,C}$.

bandwidths. We therefore only present results for the first method (based on differences) as we believe that this method should be preferred above the other one in practice.

5 Data analysis

The data we consider consist of monthly expenditures in Dfl. of Dutch households on several commodity categories, as well as on a number of background variables (Dfl. = Dutch guilders, 1 Dfl. is about 0.4 US \$). We use expenditures on food and total expenditures from October 1986 to September 1987. We selected the households consisting of two persons; the sample size is equal to 159. The data are extracted from the Data Archive of the Journal of Applied Econometrics and have been analyzed in Adang and Melenberg (1995).

Our interest lies in testing model (1.1)-(1.2) by using the differences approach developed in Section 2 with two different variables of interest:

Y = share of food expenditure in household budget

Y = log(expenditure on food per household)

For both cases X = log(total expenditures).

Method	Test	Foodshare	log Food exp.
Pairs	KS	0.061	0.436
	CM	≤ 0.001	0.572
	AD	≤ 0.001	0.427
Triplets	KS	0.027	0.980
	CM	0.002	0.770
	AD	0.002	0.561

Table 4: *P-values for the household data.*

The P-values of the tests of Section 2 are presented in Table 4. Our table shows that model (1.1)-(1.2) is violated by the foodshare data, whereas log food expenditure seems to have 'errors' independent of the covariate.

Appendix

Assumptions

$\boldsymbol{(A)}$ The sequence a_n satisfies $na_n^4 \to 0$ and $na_n^{3+2\delta}(\log a_n^{-1})^{-1} \to \infty$ for some $\delta > 0$.
$\boldsymbol{(K)}$ The probability density function K has compact support, $\int uK(u)\,du = 0$ and K is twice continuously differentiable.

$\boldsymbol{(J)(i)}$ $J(s) = I(0 \leq s \leq 1)$ or
$\boldsymbol{(ii)}$ there exist $0 \leq s_0 \leq s_1 \leq 1$ such that $s_0 \leq \inf\{s \in [0,1]; J(s) \neq 0\}$, $s_1 \geq \sup\{s \in [0,1]; J(s) \neq 0\}$ and $\inf_{x \in D_X} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) > 0$ and J is twice continuously differentiable on $(0,1)$, $\int_0^1 J(s)\,ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.

$\boldsymbol{(F)(i)}$ The support D_X of X is a bounded interval, F_X is twice continuously differentiable and $\inf_{x \in D_X} f_X(x) > 0$.

(ii) $F(y|x)$ is differentiable in y and twice differentiable in x and the derivatives are continuous in (x, y).

(iii) For every $\gamma \in (0,1)$ there exists an $\alpha \in (0,1)$, such that

$$\sup_{y,|z|\le\alpha} f_\varepsilon(y+z)/\min(F_\varepsilon(y), 1-F_\varepsilon(y))^{1-\gamma} < \infty,$$

and

$$\sup_{y,|z|\le\alpha} \frac{f_\varepsilon(y+z)}{f_\varepsilon(y)} \min(F_\varepsilon(y), 1-F_\varepsilon(y))^{\gamma} < \infty.$$

In addition to F_ε, $\hat{F}_\varepsilon$ and $\hat{F}_{\hat{\varepsilon}}$, we will need $F_{\hat{\varepsilon}}(y) = P(Y - \hat{m}(X) \le y|\hat{m})$, where (X, Y) is independent of $(X_1, Y_1), \ldots, (X_n, Y_n)$. The proofs of Section 3 are based on the two following crucial results.

Lemma A.1 *Assume (A),(K),(J) and (F). Then, for* $0 \le \beta < \frac{1}{2}$,

$$\sup_{y \in I\!\!R} \Big| [F_\varepsilon(y)(1-F_\varepsilon(y))]^{-\beta} [\hat{F}_{\hat{\varepsilon}}(y) - \hat{F}_\varepsilon(y) - F_{\hat{\varepsilon}}(y) + F_\varepsilon(y)] \Big| = o_P(n^{-1/2})$$

and

$$\sup_{\substack{x \in D_X \\ y \in I\!\!R}} \Big| N(x,y)^{-\beta} n^{-1} \sum_{i=1}^n [I(X_i \le x) - F_X(x)]$$
$$\times [I(\hat{\varepsilon}_i \le y) - I(\varepsilon_i \le y) - F_{\hat{\varepsilon}}(y) + F_\varepsilon(y)] \Big| = o_P(n^{-1/2}).$$

Proof We will show the first statement. The second one can be proved in a similar way. For reasons of symmetry we restrict attention to the case where $y < F_\varepsilon^{-1}(1/2)$. Since $1 - F_\varepsilon(y)$ is bounded away from 0 in this case, we only need to consider $F_\varepsilon(y)$ in the denominator. Choose $0 < \delta_1 < (\frac{1}{2} - \beta)/(\frac{1}{2} + \beta)$. Write

$$\frac{F_{\hat{\varepsilon}}(y)}{F_\varepsilon(y)^{1-\delta_1}} - F_\varepsilon(y)^{\delta_1} = \int \frac{P(\varepsilon \le y + \hat{m}(x) - m(x)|\hat{m}) - P(\varepsilon \le y)}{F_\varepsilon(y)^{1-\delta_1}} dF_X(x)$$
$$= \int \frac{f_\varepsilon(\xi_y(x))}{F_\varepsilon(y)^{1-\delta_1}} (\hat{m}(x) - m(x))\, dF_X(x),$$

for some $\xi_y(x)$ between y and $y+\hat{m}(x)-m(x)$. Since $\sup_{y,|z|\le\alpha} f_\varepsilon(y+z)/F_\varepsilon(y)^{1-\delta_1}<\infty$ (for some $\alpha>0$) and $\sup_x|\hat{m}(x)-m(x)|=o_P(1)$ (see Proposition 4.3 in Akritas and Van Keilegom (2001)), the above is $o_P(1)$, uniformly in y. Hence, it follows that

$$\sup_y \frac{F_{\hat{\varepsilon}}(y)^{\beta+\delta_2}}{F_\varepsilon(y)^\beta}=O_P(1)$$

with $\delta_2=\beta\delta_1/(1-\delta_1)$ and so it suffices to consider $F_{\hat{\varepsilon}}(y)^{-\beta-\delta_2}\,n^{-1}\sum_{i=1}^n[I(\hat{\varepsilon}_i\le y)-I(\varepsilon_i\le y)-F_{\hat{\varepsilon}}(y)+F_\varepsilon(y)]$. Next, note that in a similar way (but with replacing δ_1 by $\delta_1/(1+\delta_1)$), we can show that

$$\sup_y\left|\frac{F_\varepsilon(y)^{\beta+2\delta_2}}{F_{\hat{\varepsilon}}(y)^{\beta+\delta_2}}-F_\varepsilon(y)^{\delta_2}\right|\xrightarrow{P}0.$$

So since $\beta+2\delta_2<\frac{1}{2}$, it follows from the Chibisov-O'Reilly theorem that

$$\sup_y\left|F_\varepsilon(y)^{-\beta-2\delta_2}n^{-1}\sum_{i=1}^n[I(\varepsilon_i\le y)-F_\varepsilon(y)]\right|=O_P(n^{-1/2}),$$

and hence it suffices to show that

$$\sup_y\left|n^{-1}\sum_{i=1}^n\left[\frac{I(\hat{\varepsilon}_i\le y)}{F_{\hat{\varepsilon}}(y)^a}-\frac{I(\varepsilon_i\le y)}{F_\varepsilon(y)^a}-F_{\hat{\varepsilon}}(y)^{1-a}+F_\varepsilon(y)^{1-a}\right]\right|=o_P(n^{-1/2}),\qquad\text{(A.1)}$$

where $a=\beta+\delta_2$ throughout the proof. Note that $0\le a<1/2$. Let $d_n(x)=\hat{m}(x)-m(x)$, and consider the class

$$\mathcal{F}=\Big\{(x,e)\mapsto\frac{I(e\le y+d(x))}{P(\varepsilon\le y+d(X))^a}-\frac{I(e\le y)}{P(\varepsilon\le y)^a}-P(\varepsilon\le y+d(X))^{1-a}+P(\varepsilon\le y)^{1-a};$$
$$y<F_\varepsilon^{-1}(1/2),\ d\in C_1^{1+\delta}(D_X)\Big\},$$

where $C_1^{1+\delta}(D_X)$ (with $\delta>0$ as in assumption (A)) is the class of all differentiable functions d defined on D_X such that $\|d\|_{1+\delta}\le\alpha/2$ (with $\alpha>0$ as in assumption $(F)(iii)$), where

$$\|d\|_{1+\delta}=\max\{\sup_x|d(x)|,\sup_x|d'(x)|\}+\sup_{x,x'}\frac{|d'(x)-d'(x')|}{|x-x'|^\delta}.$$

Note that by Propositions 4.3, 4.4 and 4.5 in Akritas and Van Keilegom (2001), we have that $P(d_n\in C_1^{1+\delta}(D_X))\to1$ as $n\to\infty$. In the next part of this proof we will show that the class $\mathcal{F}$ is Donsker, i.e. we will establish the weak convergence of $n^{-1/2}\sum_{i=1}^n f(X_i,\varepsilon_i)$,

$f \in \mathcal{F}$. This is done by verifying the conditions of Theorem 2.11.9 in van der Vaart and Wellner (1996) :

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\bar{\varepsilon}, \mathcal{F}, L_2^n)} d\bar{\varepsilon} \to 0 \qquad \text{for every } \delta_n \downarrow 0 \tag{A.2}$$

$$n^{1/2} E\left\{\sup_{f\in\mathcal{F}} |f(X,\varepsilon)| I(\sup_{f\in\mathcal{F}} |f(X,\varepsilon)| > n^{1/2}\eta)\right\} \to 0 \qquad \text{for every } \eta > 0, \tag{A.3}$$

where $N_{[]}(\bar{\varepsilon}, \mathcal{F}, L_2^n)$ is the bracketing number, defined as the minimal number of sets $N_{\bar{\varepsilon}}$ in a partition $\mathcal{F} = \cup_{j=1}^{N_{\bar{\varepsilon}}} \mathcal{F}_{\bar{\varepsilon}j}$, such that for every $j = 1, \ldots, N_{\bar{\varepsilon}}$:

$$E\left\{\sup_{f,g\in\mathcal{F}_{\bar{\varepsilon}j}} |f(X,\varepsilon) - g(X,\varepsilon)|^2\right\} \leq \bar{\varepsilon}^2. \tag{A.4}$$

According to Theorem 2.10.6 in van der Vaart and Wellner (1996), we can deal with the four terms in the definition of $\mathcal{F}$ separately. We will restrict ourselves to showing (A.2) and (A.3) for

$$\mathcal{F}_1 = \left\{(x,e) \mapsto \frac{I(e \leq y + d(x))}{P(\varepsilon \leq y + d(X))^a}; y < F_\varepsilon^{-1}(1/2), d \in C_1^{1+\delta}(D_X)\right\},$$

since the other terms are similar, but much easier. We will assume $0 < \bar{\varepsilon} \leq 1$. In Corollary 2.7.2 of the aforementioned book it is stated that $m = N_{[]}((K_1\bar{\varepsilon})^2, C_1^{1+\delta}(D_X), L_2(P))$ is bounded by $\exp(K(K_1\bar{\varepsilon})^{-\frac{2}{1+\delta}})$, with $K_1 > 0$ to be determined later. Let $d_1^L \leq d_1^U, \ldots, d_m^L \leq d_m^U$ be the functions defining the m brackets for $C_1^{1+\delta}(D_X)$. Thus, for each d and each fixed y :

$$I(\varepsilon \leq y + d_i^L(X)) \leq I(\varepsilon \leq y + d(X)) \leq I(\varepsilon \leq y + d_i^U(X)).$$

Let $b = \min(2a, 1-2a)$. Define $F_i^L(y) = P(\varepsilon \leq y + d_i^L(X))$ and let $-\infty = y_{i1}^L < y_{i2}^L < \ldots < y_{i,m_L}^L = +\infty$ ($m_L = O(\bar{\varepsilon}^{-2/b})$) partition the line in segments having F_i^L-probability less than or equal to $K_2\bar{\varepsilon}^{2/b}$ where $K_2 > 0$ will be chosen later. Similarly, define $F_i^U(y) = P(\varepsilon \leq y + d_i^U(X))$ and let $-\infty = y_{i1}^U < y_{i2}^U < \ldots < y_{i,m_U}^U = +\infty$ ($m_U = O(\bar{\varepsilon}^{-2/b})$) partition the line in segments having F_i^U-probability less than or equal to $K_2\bar{\varepsilon}^{2/b}$.

Let $\mathcal{F}_{\bar{\varepsilon}ik}$ ($i = 1, \ldots, m$, $k = 1, \ldots, m_L - 1$) be the subset of $\mathcal{F}_1$ defined by the functions $d_i^L \leq d \leq d_i^U$ and $\tilde{y}_{ik}^L \leq y \leq \tilde{y}_{ik}^U$, where $\tilde{y}_{ik}^L = y_{ik}^L$ and $\tilde{y}_{ik}^U$ is the smallest of the y_{ik}^U which is larger than (or equal to) $y_{i,k+1}^L$. Fix i, k and fix X and ε. We consider three cases :

Case 1 : For all $f \in \mathcal{F}_{\bar{\varepsilon}ik}, f(X,\varepsilon) = 0$. The supremum in (A.4) equals zero in that case.

Case 2 : For certain $f \in \mathcal{F}_{\bar{\varepsilon}ik}, f(X,\varepsilon) = 0$ and for certain $f \in \mathcal{F}_{\bar{\varepsilon}ik}, f(X,\varepsilon) \neq 0$. This happens only if $\tilde{y}^L_{ik} + d^L_i(X) \le \varepsilon \le \tilde{y}^U_{ik} + d^U_i(X)$. Also, the supremum in (A.4) is bounded by $F_\varepsilon(\varepsilon)^{-2a}$ in that case. Hence, the expected value in (A.4), restricted to those (X,ε) that belong to case 2, is bounded by

$$
\begin{aligned}
&\int \int_{\tilde{y}^L_{ik}+d^L_i(x)}^{\tilde{y}^U_{ik}+d^U_i(x)} F_\varepsilon(y)^{-2a}\, dF_\varepsilon(y)\, dF_X(x) \\
&= \frac{1}{1-2a} \int [F_\varepsilon(\tilde{y}^U_{ik} + d^U_i(x))^{1-2a} - F_\varepsilon(\tilde{y}^L_{ik} + d^L_i(x))^{1-2a}]\, dF_X(x) \\
&= \frac{1}{1-2a} \int [F_\varepsilon(\tilde{y}^U_{ik} + d^U_i(x))^{1-2a} - F_\varepsilon(\tilde{y}^L_{i,k+1} + d^U_i(x))^{1-2a}]\, dF_X(x) \\
&\quad + \frac{1}{1-2a} \int [F_\varepsilon(\tilde{y}^L_{i,k+1} + d^U_i(x))^{1-2a} - F_\varepsilon(\tilde{y}^L_{i,k+1} + d^L_i(x))^{1-2a}]\, dF_X(x) \\
&\quad + \frac{1}{1-2a} \int [F_\varepsilon(\tilde{y}^L_{i,k+1} + d^L_i(x))^{1-2a} - F_\varepsilon(\tilde{y}^L_{ik} + d^L_i(x))^{1-2a}]\, dF_X(x) \\
&\le \frac{1}{1-2a}[F^U_i(\tilde{y}^U_{ik}) - F^U_i(\tilde{y}^L_{i,k+1})]^{1-2a} + \int F_\varepsilon(\xi_{ik}(x))^{-2a} f_\varepsilon(\xi_{ik}(x))(d^U_i(x) - d^L_i(x))\, dF_X(x) \\
&\quad + \frac{1}{1-2a}[F^L_i(\tilde{y}^L_{i,k+1}) - F^L_i(\tilde{y}^L_{ik})]^{1-2a} \\
&\le \frac{2K_2^{1-2a}}{1-2a} \bar{\varepsilon}^{2(1-2a)/b} + K' \|d^U_i - d^L_i\|_{L_1(P)},
\end{aligned}
$$

and this is bounded by $\bar{\varepsilon}^2$ for proper choice of K_1 and K_2, where $K' > 0$ and where $\xi_{ik}(x)$ is between $\tilde{y}^L_{i,k+1} + d^L_i(x)$ and $\tilde{y}^L_{i,k+1} + d^U_i(x)$.

Case 3 : For all $f \in \mathcal{F}_{\bar{\varepsilon}ik}, f(X,\varepsilon) \neq 0$. This implies that $k > 1$ and hence $F^L_i(\tilde{y}^L_{ik}) \ge K\bar{\varepsilon}^{2/b}$. Hence, the expected value at the left hand side of (A.4), restricted to those (X,ε) that satisfy the condition of case 3, is bounded by

$$
\begin{aligned}
&[F^L_i(\tilde{y}^L_{ik})^{-a} - F^U_i(\tilde{y}^U_{ik})^{-a}]^2 F^L_i(\tilde{y}^L_{ik}) \\
&= \{[F^L_i(\tilde{y}^L_{ik})^{-a} - F^L_i(\tilde{y}^L_{i,k+1})^{-a}] + [F^L_i(\tilde{y}^L_{i,k+1})^{-a} - F^U_i(\tilde{y}^L_{i,k+1})^{-a}] \\
&\quad + [F^U_i(\tilde{y}^L_{i,k+1})^{-a} - F^U_i(\tilde{y}^U_{ik})^{-a}]\}^2 F^L_i(\tilde{y}^L_{ik}) \\
&=: \{T_1 + T_2 + T_3\}^2 F^L_i(\tilde{y}^L_{ik}) \le 3\{T_1^2 + T_2^2 + T_3^2\} F^L_i(\tilde{y}^L_{ik}).
\end{aligned}
$$

It is easy to see that

$$
T_1^2 F^L_i(\tilde{y}^L_{ik}) \le [F^L_i(\tilde{y}^L_{i,k+1}) - F^L_i(\tilde{y}^L_{ik})]^{2a} F^L_i(\tilde{y}^L_{ik})^{1-4a}
$$

and this is bounded by $\bar{\varepsilon}^2$ for proper choice of $K_2 > 0$ (consider separately $a \leq 1/4$ and $a > 1/4$). It can be shown in a similar way that $T_\ell^2 F_i^L(\tilde{y}_{ik}^L) \leq \bar{\varepsilon}^2$ for $\ell = 2, 3$ and for $K_1, K_2 > 0$ small enough. This shows that (A.4) is satisfied and hence

$$N_{[\,]}(\bar{\varepsilon}, \mathcal{F}_1, L_2^n) = O(\exp(2K(K_1\bar{\varepsilon})^{-2/(1+\delta)})).$$

It now follows that (A.2) holds, since

$$\int_0^{\delta_n} \sqrt{\log N_{[\,]}(\bar{\varepsilon}, \mathcal{F}_1, L_2^n)}\, d\bar{\varepsilon} \leq 2K \int_0^{\delta_n} (K_1\bar{\varepsilon})^{-1/(1+\delta)}\, d\bar{\varepsilon} = 2K\frac{1+\delta}{\delta}(K_1\delta_n)^{\delta/(1+\delta)} \to 0.$$

Next, by writing

$$\begin{aligned}
\sup_{f\in\mathcal{F}_1} |f(X,\varepsilon)| &\leq \sup_{d\in C_1^{1+\delta}(D_X), -\infty<y<\infty} \frac{I(\varepsilon \leq y + d(X))}{P(\varepsilon \leq y + d(X))^a} \\
&= \sup_{d\in C_1^{1+\delta}(D_X)} \left[\int F_\varepsilon(\varepsilon - d(X) + d(x))^a\, dF_X(x)\right]^{-1} \\
&\leq F_\varepsilon(\varepsilon - \alpha)^{-a},
\end{aligned}$$

it follows that the left hand side of (A.3) is bounded by

$$n^{1/2} E\{F_\varepsilon(\varepsilon-\alpha)^{-a} I(F_\varepsilon(\varepsilon-\alpha)^{-a} > n^{1/2}\eta)\} = n^{1/2}\int_{-\infty}^{\kappa_n} F_\varepsilon(y-\alpha)^{-a} f_\varepsilon(y)\, dy, \qquad \text{(A.5)}$$

where $\kappa_n = F_\varepsilon^{-1}[(n^{1/2}\eta)^{-1/a}] + \alpha$. It now follows from condition $(F)(iii)$ that $(A.5)$ is bounded, for n large enough, by (where $\gamma > 0$ is chosen such that $a + \gamma < 1/2$ and K is some positive constant)

$$\begin{aligned}
&K n^{1/2} \int_{-\infty}^{\kappa_n} F_\varepsilon(y-\alpha)^{-(a+\gamma)}\, dF_\varepsilon(y-\alpha) \\
&= K n^{1/2} \int_0^{(n^{1/2}\eta)^{-1/a}} u^{-(a+\gamma)}\, du \\
&= \frac{K n^{1/2}}{1-a-\gamma}(n^{1/2}\eta)^{-(1-a-\gamma)/a} = O(n^{(2a+\gamma-1)/(2a)}) = o(1).
\end{aligned}$$

This shows that the class $\mathcal{F}_1$ (and hence $\mathcal{F}$) is Donsker. Next, let us calculate

$$\begin{aligned}
&Var\left[\frac{I(\varepsilon \leq y + d_n(X))}{P(\varepsilon \leq y + d_n(X))^a} - \frac{I(\varepsilon \leq y)}{P(\varepsilon \leq y)^a} - P(\varepsilon \leq y + d_n(X))^{1-a} + P(\varepsilon \leq y)^{1-a} \,\middle|\, d_n\right] \\
&\leq E\left[E\left(\left\{\frac{I(\varepsilon \leq y + d_n(X))}{P(\varepsilon \leq y + d_n(X))^a} - \frac{I(\varepsilon \leq y)}{P(\varepsilon \leq y)^a}\right\}^2 \,\middle|\, X, d_n\right) \,\middle|\, d_n\right]. \qquad \text{(A.6)}
\end{aligned}$$

The conditional expectation is equal to (suppose that $d_n(X) \geq 0$ for simplicity)

$$
\begin{aligned}
& E\left[\frac{I(\varepsilon \leq y + d_n(X))}{F_\varepsilon(y + d_n(X))^{2a}} + \frac{I(\varepsilon \leq y)}{F_\varepsilon(y)^{2a}} - 2\frac{I(\varepsilon \leq y)}{F_\varepsilon(y + d_n(X))^a F_\varepsilon(y)^a} \middle| X, d_n\right] \\
&= F_\varepsilon(y + d_n(X))^{1-2a} - F_\varepsilon(y)^{1-2a} + 2\frac{F_\varepsilon(y)^{1-2a}}{F_\varepsilon(y + d_n(X))^a}[F_\varepsilon(y + d_n(X))^a - F_\varepsilon(y)^a] \\
&= (1-2a)F_\varepsilon(\xi_y(X))^{-2a} f_\varepsilon(\xi_y(X)) d_n(X) \\
&\quad + 2a\frac{F_\varepsilon(y)^{1-2a}}{F_\varepsilon(y + d_n(X))^a} F_\varepsilon(\tilde{\xi}_y(X))^{a-1} f_\varepsilon(\tilde{\xi}_y(X)) d_n(X) \\
&\leq F_\varepsilon(\xi_y(X))^{-2a} f_\varepsilon(\xi_y(X)) d_n(X)
\end{aligned}
$$

and, by condition $(F)(iii)$, this is bounded by $Kd_n(X)$ for some $K > 0$, where $\xi_y(X)$ and $\tilde{\xi}_y(X)$ are between y and $y + d_n(X)$. A similar derivation can be given when $d_n(X) \leq 0$. It follows that the right hand side of (A.6) is bounded by $K \sup_x |d_n(x)| = o_P(1)$, by Proposition 4.3 in Akritas and Van Keilegom (2001).

Since the class $\mathcal{F}$ is Donsker, it follows from Corollary 2.3.12 in van der Vaart and Wellner (1996) that

$$
\lim_{\eta \downarrow 0} \limsup_{n \to \infty} P\left(\sup_{f \in \mathcal{F}, Var(f) < \eta} n^{-1/2}\left|\sum_{i=1}^n f(X_i, \varepsilon_i)\right| > \bar{\varepsilon}\right) = 0,
$$

for each $\bar{\varepsilon} > 0$. By restricting the supremum inside this probability to the elements in $\mathcal{F}$ corresponding to $d(X) = d_n(X)$ as defined above, (A.1) follows. □

Lemma A.2 *Assume (A),(K),(J) and (F), let* $\eta > 0, 0 < \zeta < 1$ *and* $b_n = n^{-(1-\zeta)}$. *Then,*

$$
\sup_{\hat{F}_{\hat{\varepsilon}}(y) \geq b_n} \left\{F_\varepsilon(y)^\eta \left|\frac{F_\varepsilon(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1\right|\right\} = o_P(1)
$$

and

$$
\sup_{1-\hat{F}_{\hat{\varepsilon}}(y) \geq b_n} \left\{(1 - F_\varepsilon(y))^\eta \left|\frac{1 - F_\varepsilon(y)}{1 - \hat{F}_{\hat{\varepsilon}}(y)} - 1\right|\right\} = o_P(1).
$$

Proof We only prove the first statement. The second one follows in a similar way. Choose

$\nu > 0$ such that $\nu < 1/2$ and $(1-\zeta)(1-\nu) < 1/2$. Then,

$$\begin{aligned}
\sup_{F_{\hat\varepsilon}(y)\geq b_n}\Big|\frac{\hat F_{\hat\varepsilon}(y)}{F_{\hat\varepsilon}(y)}-1\Big| &\leq \sup_{F_{\hat\varepsilon}(y)\geq b_n}\frac{|\hat F_{\hat\varepsilon}(y)-F_{\hat\varepsilon}(y)|}{F_{\hat\varepsilon}(y)^\nu b_n^{1-\nu}}\\
&= o(1)\sup_{F_{\hat\varepsilon}(y)\geq b_n}\frac{|\sqrt{n}(\hat F_{\hat\varepsilon}(y)-F_{\hat\varepsilon}(y))|}{F_{\hat\varepsilon}(y)^\nu}\\
&= o(1)\Big(\sup\frac{|\sqrt{n}(\hat F_{\varepsilon}(y)-F_{\varepsilon}(y))|}{F_{\varepsilon}(y)^\nu}+o_P(1)\Big)=o_P(1),
\end{aligned}$$

where the last equality follows from the Chibisov-O'Reilly theorem and the one but last equality from the proof of Lemma A.1. Hence, it follows that

$$\sup_{F_{\hat\varepsilon}(y)\geq b_n}\Big|\frac{F_{\hat\varepsilon}(y)}{\hat F_{\hat\varepsilon}(y)}-1\Big| = o_P(1). \tag{A.7}$$

We next show that the supremum in (A.7) can be replaced by the supremum over $\{y : \hat F_{\hat\varepsilon}(y)\geq b_n\}$. Indeed, it follows from (A.7) that there exists a sequence $\delta_n\downarrow 0$ such that

$$P\left(\sup_{F_{\hat\varepsilon}(y)\geq b_n}\Big|\frac{F_{\hat\varepsilon}(y)}{\hat F_{\hat\varepsilon}(y)}-1\Big|\geq\delta_n\right)\to 0$$

as $n\to\infty$. Hence,

$$\begin{aligned}
&P\left(\sup_{\hat F_{\hat\varepsilon}(y)\geq b_n}\Big|\frac{F_{\hat\varepsilon}(y)}{\hat F_{\hat\varepsilon}(y)}-1\Big|\geq\varepsilon\right)\\
&\leq P\left(\sup_{F_{\hat\varepsilon}(y)\geq b_n(1-\delta_n)}\Big|\frac{F_{\hat\varepsilon}(y)}{\hat F_{\hat\varepsilon}(y)}-1\Big|\geq\varepsilon\right)+P\left(\sup_{F_{\hat\varepsilon}(y)\geq b_n}\Big|\frac{F_{\hat\varepsilon}(y)}{\hat F_{\hat\varepsilon}(y)}-1\Big|\geq\delta_n\right)\to 0.
\end{aligned}$$

Finally, consider

$$\sup_{\hat F_{\hat\varepsilon}(y)\geq b_n}\Big\{F_\varepsilon(y)^\eta\Big|\frac{F_\varepsilon(y)}{\hat F_{\hat\varepsilon}(y)}-1\Big|\Big\}=\sup_{\hat F_{\hat\varepsilon}(y)\geq b_n}\Big\{F_{\hat\varepsilon}(y)^\eta\Big|\frac{F_{\hat\varepsilon}(y)}{\hat F_{\hat\varepsilon}(y)}-1\Big|\Big\}+o_P(1)=o_P(1),$$

since $\sup_y|F_{\hat\varepsilon}(y)^\eta-F_\varepsilon(y)^\eta| = o_P(1)$ and since it follows from the proof of Lemma A.1 that

$$\sup_y\Big|\frac{F_\varepsilon(y)^{1+\eta}}{F_{\hat\varepsilon}(y)}-F_\varepsilon(y)^\eta\Big|=o_P(1).$$

$\square$

Acknowledgment We like to thank Arthur van Soest for providing the data on food expenditures and for very helpful discussions concerning the corresponding data analysis. The research of Ingrid Van Keilegom was supported by the contract 'Projet d'Actions de Recherche Concertées' nr. 98/03–217, and by Belgian IAP research network grant nr. P5/24.

References

Adang, P.J.M. and Melenberg, B. (1995) Nonnegativity constraints and intratemporal uncertainty in multi-good life-cycle models. *J. of Appl. Econometrics*, **10**, 1 - 15.

Akritas, M.G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.*, **28**, 549-568.

Cao, R. and Gijbels, I. (2003). Testing homoscedasticity via the integrated conditional variance in nonparametric regression. In preparation.

Csörgő, M. and Révész, P. (1981). *Strong approximations in probability and statistics.* Academic Press, New York.

Dette, H., Munk, A. and Wagner, T. (1998). Estimating the variance in nonparametric regression - what is a reasonable choice? *J. R. Statist. Soc. - Ser. B*, **60**, 751–764.

Komlós, J. Major, P. and Tusnády G. (1975). An approximation of partial sums of independent rv's and the sample df, I. *Z. Warsch. Verw. Gebiete*, **32**, 111-131.

Liero, H. (2003). Testing homoscedasticity in nonparametric regression. *J. Nonpar. Statist.*, **15**, 31–51.

Müller, U., Schick, A. and Wefelmeyer, W. (2003). Estimating the error variance in nonparametric regression by a covariate-matched U-statistic. *Statistics*, **37**, 179–188.

Müller, U., Schick, A. and Wefelmeyer, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *J. Statist. Planning Infer.*, **119**, 75–93.

Shorack, G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595-645.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer-Verlag, New York.

Dept. of Econometrics & OR
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands
Email: j.h.j.einmahl@uvt.nl

Institut de Statistique
Université catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve
Belgium
Email: vankeilegom@stat.ucl.ac.be