Estimating the error distribution function in semiparametric regression *

Ursula U. Müller, Anton Schick, Wolfgang Wefelmeyer

Abstract

We prove a stochastic expansion for a residual-based estimator of the error distribution function in a partly linear regression model. It implies a functional central limit theorem. As special cases we cover nonparametric, nonlinear and linear regression models.

1 Introduction

We consider the partly linear regression model

$$Y = \vartheta^\top U + \varrho(X) + \varepsilon,$$

where the error ε is independent of the covariate pair (U, X) and the parameter ϑ is k-dimensional. We make the following assumptions.

(F) *The error ε has mean zero, a finite moment of order $\beta > 8/3$, and a density f which is Hölder with exponent $\xi > 1/3$.*

(G) *The distribution G of X is* quasi-uniform *on $[0, 1]$ in the sense that $G([0, 1]) = 1$ and G has a density g that is bounded and bounded away from zero on $[0, 1]$.*

(H) *The covariate vector U satisfies $E[|U|^2] < \infty$, the matrix $E[(U - \mu(X))(U - \mu(X)^\top]$ is positive definite, μ is continuous and τg is bounded, where $\mu(X) = E(U|X)$ and $\tau(X) = E(|U|^2|X)$.*

(R) *The function ϱ is twice continuously differentiable.*

---

*AMS 2000 subject classification: Primary 62G05, 62G08, 62G20, 62J02
Key words and Phrases: Local linear smoother, i.i.d. representation, Donsker class, efficiency.

Our goal is to estimate the distribution function F of ε based on n independent copies (U_j, X_j, Y_j) of (U, X, Y). Our estimator of F will be the empirical distribution function based on residuals. To obtain residuals we need estimators of ϑ and ϱ. Under the above assumptions, there exist $n^{1/2}$-consistent estimators of ϑ; see e.g. Schick (1996). Given such an estimator $\hat{\vartheta}$ of ϑ, we estimate ϱ by a *local linear smoother* $\hat{\varrho}$ as follows. For a fixed x in $[0,1]$, the estimator $\hat{\varrho}(x)$ is the first component of the minimizer $(\hat{\beta}_0, \hat{\beta}_1)$ of

$$\sum_{j=1}^{n} \Big(Y_j - \hat{\vartheta}^\top U_j - \beta_0 - \beta_1 \frac{X_j - x}{c_n} \Big)^2 w\Big(\frac{X_j - x}{c_n} \Big), \tag{1.1}$$

where c_n is a bandwidth and w is a kernel with the following properties.

(W) *The kernel w is a three times continuously differentiable symmetric density with compact support* $[-1, 1]$.

We then form the residuals

$$\hat{\varepsilon}_j = Y_j - \hat{\vartheta}^\top U_j - \hat{\varrho}(X_j), \quad j = 1, \dots, n,$$

and use as estimator of F the empirical distribution function of these residuals,

$$\hat{\mathbb{F}}(t) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}[\hat{\varepsilon}_j \le t], \quad t \in \mathbb{R}.$$

We denote the empirical distribution function based on the errors by

$$\mathbb{F}(t) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}[\varepsilon_j \le t], \quad t \in \mathbb{R}.$$

We can now state our main result.

Theorem 1.1 *Assume that (F), (G), (H), (R) and (W) hold and $c_n \sim (n \log n)^{-1/4}$. Let $\hat{\vartheta}$ be a $n^{1/2}$-consistent estimator of ϑ. Then*

$$\sup_{t \in \mathbb{R}} \Big| \hat{\mathbb{F}}(t) - \mathbb{F}(t) - f(t) \, \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j \Big| = o_p(n^{-1/2}). \tag{1.2}$$

Under our assumptions (G) and (R), the optimal choice of bandwidth for estimating ϱ is of order $n^{-1/5}$. Our proof requires an undersmoothed estimator of ϱ with a bias that is of order $o(n^{-1/2})$. This is guaranteed by the choice of bandwidth in the theorem.

The nonparametric regression model $Y = \varrho(X) + \varepsilon$ is a special case of the partly linear regression corresponding to $\vartheta = 0$. Taking $\hat{\vartheta} = 0$, the above theorem carries over to this model, giving (1.2) without condition (H).

Our approach is motivated by Akritas and Van Keilegom (2001) who consider the heteroscedastic nonparametric regression model $Y = \varrho(X) + s(X)\varepsilon$. In our model, $s(X) = 1$ and $\varrho(X) = E(Y|X)$, which corresponds to their function J being 1. Their assumption (A2) would then imply that ε is quasi-uniform on some finite interval. We get by with considerably weaker conditions.

Kiwitt, Nagel and Neumeyer (2005) treat the nonparametric regression model $Y = \varrho(X) + \varepsilon$ with additional linear constraints on the error distribution F. They rely on the results and assumptions of Akritas and Van Keilegom and use a kernel estimator for ϱ. The kernel estimator requires stronger assumptions on the design density g than our linear smoother.

In the nonparametric regression model our estimator $\hat{\mathbb{F}}(t)$ has influence function

$$\mathbf{1}[\varepsilon \le t] - F(t) + f(t)\varepsilon$$

and is therefore efficient by Müller, Schick and Wefelmeyer (2004). Since this is also the influence function of $\hat{\mathbb{F}}(t)$ in the larger partly linear model, $\hat{\mathbb{F}}(t)$ is also efficient there.

The linear regression model $Y = \vartheta^\top U + \varepsilon$ corresponds to the case $\varrho = 0$. In this model one can take $\hat{\varepsilon}_j = Y_j - \hat{\vartheta}^\top U_j$ with $\hat{\vartheta}$ the least squares estimator of ϑ and obtains the following result.

Theorem 1.2 *Assume that F has mean zero, finite variance and a uniformly continuous density f, and the matrix $E[UU^\top]$ is positive definite. Then*

$$\sup_{t\in\mathbb{R}} \left| \hat{\mathbb{F}}(t) - \mathbb{F}(t) - f(t)\ (\hat{\vartheta} - \vartheta)^\top E[U] \right| = o_p(n^{-1/2}).$$

This was first proved by Koul (1969) for fixed design. See also Koul (2002) and, for increasing dimension, Mammen (1996). Theorem 1.2 follows from Theorem 2.3 about nonlinear regression.

Our paper is organized as follows. In Section 2 we adapt a result of Akritas and Van Keilegom (2001) on uniform stochastic expansions of residual-based empirical distribution functions to our setting. In Section 3 we prove Theorem 1.1. Technical details about kernel type estimators are in Section 4.

2 General results

Let ε be a random variable with distribution function F, and let Z be a random vector with distribution Q independent of ε. Let D be a non-negative function in $L_2(Q)$, and let $\mathcal{D}$ be a set of measurable functions a such that $|a| \le D$ and $0 \in \mathcal{D}$. We now give conditions on the class $\mathcal{D}$ that imply that the class $\mathcal{H} = \{h_{a,t} : a \in \mathcal{D}, t \in \mathbb{R}\}$ is $F \otimes Q$-Donsker, where

$$h_{a,t}(\varepsilon, Z) = \mathbf{1}[\varepsilon - a(Z) \le t], \quad a \in \mathcal{D}, t \in \mathbb{R}.$$

For this we endow $\mathcal{D}$ with the $L_1(Q)$-pseudo-norm. By an η-bracket for $(\mathcal{D}, L_1(Q))$ we mean a set $[\underline{a}, \overline{a}] = [a \in \mathcal{D} : \underline{a} \le a \le \overline{a}\}$ where $\underline{a}$ and $\overline{a}$ belong to $L_1(Q)$ and satisfy $\int |\underline{a} - \overline{a}|\, dQ \le \eta$. Recall that the *bracketing number* $N_{[\,]}(\eta, \mathcal{D}, L_1(Q))$ is the smallest integer m for which there are m η-brackets $[\underline{a}_1, \overline{a}_1], \dots, [\underline{a}_m, \overline{a}_m]$ which cover $\mathcal{D}$ in the sense that the union of the brackets contains $\mathcal{D}$.

Theorem 2.1 *Assume that F has a finite second moment and a bounded density and that the bracketing numbers satisfy*

$$\int_0^1 \sqrt{\log N_{[\,]}(\eta^2, \mathcal{D}, L_1(Q))}\, d\eta < \infty. \tag{2.1}$$

Then $\mathcal{H}$ is $F \otimes Q$-Donsker.

Proof: Let F_+ denote the distribution function of $\varepsilon + D(Z)$ and F_- the distribution function of $\varepsilon - D(Z)$. Since these random variables have finite second moments by the assumptions on F and D, we see that $F_-(t) \le C^2/t^2$ for negative t and $(1 - F_+(t)) \le C^2/t^2$ for positive t, where C is some positive constant C. Note that $\mathcal{H}$ has envelope 1. We shall show that

$$\int_0^\infty \sqrt{\log N_{[\,]}(\eta, \mathcal{H}, L_2(F \otimes Q))}\, d\eta < \infty. \tag{2.2}$$

The desired result then follows from Ossiander (1987), see also van der Vaart and Wellner (1996, Theorem 2.5.6). Let L denote the Lipschitz constant of F. Let $[\underline{a}, \overline{a}]$ be an $\eta^2/(2L)$-bracket for $(\mathcal{D}, L_1(Q))$ and $a \in [\underline{a}, \overline{a}]$. We may assume that $|\underline{a}| \leq D$ and $|\overline{a}| \leq D$. Let $u < v$ be real numbers such that $v - u \leq \eta^2/(2L)$. Then, for $t \in [u, v]$, we have

$$\mathbf{1}[\varepsilon - \underline{a}(Z) \leq u] \leq \mathbf{1}[\varepsilon - a(Z) \leq t] \leq \mathbf{1}[\varepsilon - \overline{a}(Z) \leq v]$$

and

$$\begin{aligned} E[(\mathbf{1}[\varepsilon - \underline{a}(Z) \leq u] - \mathbf{1}[\varepsilon - \overline{a}(Z) \leq v])^2] & \\ &= E[F(v + \overline{a}(Z)) - F(u + \underline{a}(Z))] \\ &\leq L(v - u + E[\overline{a}(Z) - \underline{a}(Z)]) \\ &\leq \eta^2; \end{aligned}$$

for $t \leq -C/\eta$, we have

$$0 \leq \mathbf{1}[\varepsilon - a(Z) \leq t] \leq \mathbf{1}[\varepsilon - D(Z) \leq -C/\eta]$$

and

$$E[\mathbf{1}[\varepsilon - D(Z) \leq -C/\eta]^2] \leq F_{-}(-C/\eta) \leq \eta^2;$$

and, for $t \geq C/\eta$, we have

$$\mathbf{1}[\varepsilon + D(Z) \leq C/\eta] \leq \mathbf{1}[\varepsilon - a(Z) \leq t] \leq 1$$

and

$$E[(1 - \mathbf{1}[\varepsilon + D(Z) \leq C/\eta])^2] \leq 1 - F_{+}(C/\eta) \leq \eta^2.$$

This shows that the bracketing numbers $N_{[\,]}(\eta, \mathcal{H}, L_2(F \otimes Q))$ are bounded by

$$K\eta^{-3} N_{[\,]}(\eta^2/(2L), \mathcal{D}, L_1(Q))$$

for all $0 < \eta \leq 1$ and some constant K and are bounded by 1 for $\eta \geq 1$ (take the bracket $[0, 1]$). Since $\sqrt{x + y} \leq \sqrt{x} + \sqrt{y}$ for non-negative x and y and since $\int_0^1 \sqrt{\log(\eta^{-3})}\, d\eta$ is finite, we see that (2.1) implies the desired (2.2). □

Now consider a regression model

$$Y = r(Z) + \varepsilon$$

and independent copies (Y_j, Z_j) of (Y, Z). For an estimator $\hat{r}$ of r define the residuals $\hat{\varepsilon}_j = Y_j - \hat{r}(Z_j)$. As before we set

$$\hat{\mathbb{F}}(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}[\hat{\varepsilon}_j \le t] \quad \text{and} \quad \mathbb{F}(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}[\varepsilon_j \le t], \quad t \in \mathbb{R}.$$

Theorem 2.2 *Let $\mathcal{D}$ be as in the previous theorem. Let F have a finite second moment and a density f that is Hölder with exponent $\xi \in (0,1]$. Assume that there is an $\hat{a}$ such that*

$$P(\hat{a} \in \mathcal{D}) \to 1, \tag{2.3}$$

$$\int |\hat{a}|^{1+\xi}\, dQ = o_p(n^{-1/2}), \tag{2.4}$$

$$\sup_z |\hat{r}(z) - r(z) - \hat{a}(z)| = o_p(n^{-1/2}). \tag{2.5}$$

Then

$$\sup_{t\in\mathbb{R}} \left| \hat{\mathbb{F}}(t) - \mathbb{F}(t) - f(t)\int \hat{a}\, dQ \right| = o_p(n^{-1/2}).$$

Proof: Without loss of generality we may assume $\hat{a}$ is $\mathcal{D}$-valued; otherwise replace $\hat{a}$ by $\hat{a}\mathbf{1}[\hat{a} \in \mathcal{D}]$. Let

$$\tilde{\mathbb{F}}(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}[\varepsilon_j - \hat{a}(Z_j) \le t] \quad \text{and} \quad F_a(t) = \int F(t + a(z))\, dQ(z).$$

Then we can write

$$\hat{\mathbb{F}}(t) - \mathbb{F}(t) - f(t)\int \hat{a}\, dQ = T_1(t) + T_2(t) + T_3(t),$$

where

$$\begin{aligned} T_1(t) &= \hat{\mathbb{F}}(t) - \tilde{\mathbb{F}}(t), \\ T_2(t) &= \tilde{\mathbb{F}}(t) - F_{\hat{a}}(t) - \mathbb{F}(t) + F(t), \\ T_3(t) &= F_{\hat{a}}(t) - F(t) - f(t)\int \hat{a}\, dQ. \end{aligned}$$

Since f is Hölder, say with constant Λ, we obtain that

$$\begin{aligned}|T_3(t)| &\le \int |F(t+\hat{a}(z)) - F(t) - f(t)\hat{a}(z)|\, dQ(z) \\ &\le \Lambda \int |\hat{a}|^{1+\xi}\, dQ = o_p(n^{-1/2}).\end{aligned}$$

To deal with T_1 and T_2, we introduce the empirical process

$$\begin{aligned}\nu_n(a,t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\mathbf{1}[\varepsilon_j - a(Z_j) \le t] - F_a(t)) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (h_{a,t}(\varepsilon_j, Z_j) - E[h_{a,t}(\varepsilon, Z)]), \quad a \in \mathcal{D}, t \in \mathbb{R},\end{aligned}$$

associated with the Donsker class $\mathcal{H}$. Then we have the identity

$$n^{1/2} T_2(t) = \nu_n(\hat{a}, t) - \nu_n(0, t)$$

and the bound

$$\begin{aligned}|n^{1/2} T_1(t)| &\le n^{1/2}(\tilde{F}(t+R_n) - \tilde{F}(t-R_n)) \\ &\le |\nu_n(\hat{a}, t+R_n) - \nu_n(\hat{a}, t-R_n)| \\ &\qquad + n^{1/2}(F_{\hat{a}}(t+R_n) - F_{\hat{a}}(t-R_n)),\end{aligned}$$

where R_n denotes the left-hand side of (2.5). Since f is Hölder, f is also bounded and the functions F_a are Lipschitz with Lipschitz constant $\|f\|_\infty$. Thus we have

$$n^{1/2}(F_{\hat{a}}(t+R_n) - F_{\hat{a}}(t-R_n)) \le 2\|f\|_\infty n^{1/2} R_n = o_p(1). \qquad (2.6)$$

Moreover, for $s, t \in \mathbb{R}$ and $a, b \in \mathcal{D}$,

$$\begin{aligned}E[(h_{a,s}(\varepsilon, Z) - h_{b,t}(\varepsilon, Z))^2] &\le E[|F(s + a(Z)) - F(t + b(Z))|] \\ &\le \|f\|_\infty \big(|s-t| + E[|a(Z) - b(Z)|]\big).\end{aligned}$$

In view of this and the stochastic equi-continuity of the empirical process, for every $\eta > 0$ there is a $\delta > 0$ such that, with P^* denoting outer measure,

$$\sup_n P^*\Big(\sup_{t\in\mathbb{R}, a\in\mathcal{D}, \int |a|\, dQ < \delta} |\nu_n(a,t) - \nu_n(0,t)| > \eta \Big) < \eta,$$

$$\sup_n P^*\Big(\sup_{a\in\mathcal{D}, s,t\in\mathbb{R}, |s-t|<\delta} |\nu_n(a,s) - \nu_n(a,t)| > \eta \Big) < \eta.$$

The first of these statements and (2.4) imply $\sup_t |T_2(t)| = o_p(n^{-1/2})$, while the second, (2.5) and (2.6) imply $\sup_t |T_1(t)| = o_p(n^{-1/2})$. $\square$

Theorem 2.2 was formulated with semiparametric regression in mind. In parametric regression models one typically has

$$\int |\hat{a}| \, dQ = O_p(n^{-1/2}) \tag{2.7}$$

in which case the Hölder condition on f can be relaxed to uniform continuity, as is easily seen by inspecting the proof. To state this result, we look at the parametric regression model $Y = r_\vartheta(Z) + \varepsilon$ with regression function r_ϑ indexed by a k-dimensional parameter ϑ and differentiable in the parameter in the following sense.

(D) *There is function $\dot{r}_\vartheta$ into $\mathbb{R}^k$ such that $|\dot{r}_\vartheta| \in L_2(Q)$ and*

$$\sup_z |r_{\vartheta+t}(z) - r_\vartheta(z) - t^\top \dot{r}_\vartheta(z)| = o(|t|).$$

Given a $n^{1/2}$-consistent estimator $\hat{\vartheta}$ of ϑ, assumptions (2.3), (2.7) and (2.5) are met with $\hat{r} = r_{\hat{\vartheta}}$ and $\hat{a} = (\hat{\vartheta} - \vartheta)^\top \dot{r}_\vartheta$, and with $\mathcal{D} = \{t^\top \dot{r}_\vartheta : |t| \le 1\}$. Since

$$N_{[\,]}(\eta, \mathcal{D}, L_1(Q)) \le M\eta^{-k}$$

for some constant M, the entropy condition (2.1) holds. Thus we have the following result for parametric regression.

Theorem 2.3 *Assume that (D) holds and $\hat{r} = r_{\hat{\vartheta}}$ with $\hat{\vartheta}$ a $n^{1/2}$-consistent estimator of ϑ. Let F have a finite second moment and a uniformly continuous density f. Then*

$$\sup_{t \in \mathbb{R}} \left| \hat{\mathbb{F}}(t) - \mathbb{F}(t) - f(t)(\hat{\vartheta} - \vartheta)^\top \int \dot{r}_\vartheta \, dQ \right| = o_p(n^{-1/2}).$$

A special case is the linear regression model, for which $r_\vartheta(z) = \vartheta^\top z$. In this model, assumption (D) is trivially satisfied with $\dot{r}_\vartheta(z) = z$, and a $n^{1/2}$-consistent estimator of ϑ is given by the least squares estimator, provided $E[\varepsilon] = 0$, $E[\varepsilon^2]$ is finite and the matrix $E[ZZ^\top]$ is positive definite.

3 Partly linear regression

To prove our main result, Theorem 1.1, we apply Theorem 2.2 with the choices $Z = (U, X)$ and $r(Z) = \vartheta^\top U + \varrho(X)$. We take $\mathcal{D}$ to be the class of functions

$$a(u, x) = b^\top(u - \mu(x)) + c(x)$$

with $b \in [-1, 1]^k$ and c belonging to a class $\mathcal{C}$ of functions to be introduced next. Let $1 < \alpha \leq 2$. For a function h defined on $[0, 1]$ set

$$\|h\| = \sup_{0 \leq x \leq 1} |h(x)|.$$

If h is also differentiable on $[0, 1]$, let

$$\|h\|_\alpha = \|h\| + \|h'\| + \sup_{0 \leq x < y \leq 1} \frac{|h'(x) - h'(y)|}{|x - y|^{\alpha - 1}}.$$

We take $\mathcal{C} = C_1^\alpha([0, 1])$, the set of all such functions h with $\|h\|_\alpha \leq 1$. By Corollary 2.7.2 in van der Vaart and Wellner (1996), there is a constant K such that

$$\log N_{[\,]}(\eta^2, C_1^\alpha([0, 1]), L_1(G)) \leq K\eta^{-2/\alpha}, \quad \eta \leq 1. \tag{3.1}$$

Note that

$$N_{[\,]}(\eta, \mathcal{D}, L_1(Q)) \leq N_{[\,]}(\eta, \mathcal{D}_0, L_1(Q)) N_{[\,]}(\eta, C_1^\alpha([0, 1]), L_1(G)),$$

where

$$\mathcal{D}_0 = \{(u, x) \mapsto b^\top(u - \mu(x)) : b \in [-1, 1]^k\}.$$

Since $N_{[\,]}(\eta, \mathcal{D}_0, L_1(Q)) \leq M\eta^{-k}$ for some constant M, the desired entropy condition (2.1) follows from (3.1) and $\alpha > 1$.

Lemma 3.2 below yields (2.5) with

$$\hat{a}(u, x) = (\hat{\vartheta} - \vartheta)^\top(u - \mu(x)) + \hat{c}(x)$$

and $\hat{c}$ defined in (3.11). By the $n^{1/2}$-consistency of $\hat{\vartheta}$, we have (2.3) if $P(\hat{c} \in C_1^\alpha([0, 1]))$ tends to one. Sufficient conditions for the latter are given in Lemma 3.1 for arbitrary $\hat{c}$ and verified in Lemma 3.3 for the choice $\hat{c}$ in (3.11). By the definition of μ, we have

$$\int \hat{a}^2\, dQ = (\hat{\vartheta} - \vartheta)^\top E[(U - \mu(X))(U - \mu(X))^\top](\hat{\vartheta} - \vartheta) + \int \hat{c}^2\, dG. \tag{3.2}$$

In view of the inequality

$$\int |\hat{a}|^{1+\xi}\, dQ \le \Big(\int \hat{a}^2\, dQ\Big)^{(1+\xi)/2},$$

relation (2.4) follows from the $n^{1/2}$-consistency of $\hat{\vartheta}$, Lemma 3.4 and the choice of bandwidth. Thus the assumptions of Theorem 2.2 hold and we obtain

$$\sup_{t\in\mathbb{R}} \Big|\hat{\mathbb{F}}(t) - F(t) - f(t)\int \hat{a}\, dQ\Big| = o_p(n^{-1/2}).$$

By the definition of μ and Lemma 3.5,

$$\int \hat{a}\, dQ = \int \hat{c}\, dG = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j + o_p(n^{-1/2}).$$

Since f is Hölder and hence bounded, the desired expansion (1.2) follows.

We now give sufficient conditions for an estimator $\hat{c}$ to belong to the class $C_1^{\alpha}([0,1])$.

Lemma 3.1 *Let G be such that $G([0,1]) = 1$. Let $\hat{c}$ be twice continuously differentiable on $[0,1]$ and satisfy $\|\hat{c}\| = o_p(1)$, $\|\hat{c}'\| = O_p(n^{-\beta_1})$ and $\|\hat{c}''\| = O_p(n^{\beta_2})$ for constants β_1 and β_2 in $(0,1)$. Let $1 < \alpha < 1 + \beta_1 \wedge (1-\beta_2)$. Then $P(\hat{c} \in C_1^{\alpha}([0,1]) \to 1$.*

Proof: We need to show that $P(\|\hat{c}\|_\alpha > 1) \to 0$. In view of $\|\hat{c}\| = o_p(1)$, $\|\hat{c}'\| = o_p(1)$ and the definition of $\|h\|_\alpha$, it is enough to show that

$$\sup_{0\le x<y\le 1} \frac{|\hat{c}'(x) - \hat{c}'(y)|}{|x-y|^{\alpha-1}} = o_p(1).$$

Since $\|\hat{c}'\| = O_p(n^{-\beta_1})$, we have

$$\sup_{y-x>1/n} \frac{|\hat{c}'(x) - \hat{c}'(y)|}{|x-y|^{\alpha-1}} \le 2n^{\alpha-1}\|\hat{c}'\| = O_p(n^{\alpha-1-\beta_1});$$

since $\|\hat{c}''\| = O_p(n^{\beta_2})$, we have

$$\sup_{0<y-x\le 1/n} \frac{|\hat{c}'(x) - \hat{c}'(y)|}{|x-y|^{\alpha-1}} \le \|\hat{c}''\| \sup_{0<y-x\le 1/n} |x-y|^{2-\alpha} = O_p(n^{\beta_2+\alpha-2}).$$

The desired result follows as the exponents $\alpha - 1 - \beta_1$ and $\alpha + \beta_2 - 2$ are negative. $\square$

Recall that $\hat{\varrho}(x)$ is the first component of the minimizer $(\hat{\beta}_0, \hat{\beta}_1)$ of (1.1). This minimizer obeys the normal equations

$$\begin{bmatrix} \hat{p}_0(x) & \hat{p}_1(x) \\ \hat{p}_1(x) & \hat{p}_2(x) \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \hat{q}_0(x) \\ \hat{q}_1(x) \end{bmatrix},$$

where, for $i = 0, 1, \dots$ and with $w_i(y) = y^i w(y)$,

$$\hat{p}_i(x) = \frac{1}{nc_n} \sum_{j=1}^n w_i\Big(\frac{X_j - x}{c_n}\Big),$$
$$\hat{q}_i(x) = \frac{1}{nc_n} \sum_{j=1}^n (Y_j - \hat{\vartheta}^\top U_j) w_i\Big(\frac{X_j - x}{c_n}\Big).$$

To describe $\hat{\varrho}(x)$ we set

$$\bar{p}_i(x) = E[\hat{p}_i(x)] = \int \frac{1}{c_n} w_i\Big(\frac{u - x}{c_n}\Big) g(u)\, du = \int w_i(y) g(x + c_n y)\, dy.$$

For the rest of this section we assume without further mention that the bandwidth c_n satisfies $c_n \to 0$ and $c_n^{-1} n^{-1} \log n \to 0$ and make additional assumptions as needed. This allows one to obtain versions of Theorem 1.1 for more general choices of bandwidth. We also assume (without loss of generality) that $c_n < 1/2$. Under (G) and (W) we have

$$\|\hat{p}_i - \bar{p}_i\| = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big), \tag{3.3}$$

$$\|\bar{p}_i\| \le \|\bar{p}_0\| \le \|g\| < \infty. \tag{3.4}$$

The former follows from Corollary 4.2 applied with $v = w_i$, $T_j = T = 1$, $\beta = \infty$ and $\delta = 0$. By (G) there is an $\eta > 0$ such that $g(x) \ge \eta$ for $0 \le x \le 1$. Thus, for such x,

$$g(x + c_n y) \ge \eta\, \mathbf{1}\Big[\frac{-x}{c_n} \le y \le \frac{1 - x}{c_n}\Big]$$

and therefore

$$\bar{p}_0(x)\bar{p}_2(x) - \bar{p}_1^2(x) = \bar{p}_0(x) \int \Big(y - \frac{\bar{p}_1(x)}{\bar{p}_0(x)}\Big)^2 g(x + c_n y) w(y)\, dy$$

is bounded below by

$$\frac{\eta^2}{4} \int_{-x/c_n}^{(1-x)/c_n} 2w(y)\, dy \int_{-x/c_n}^{(1-x)/c_n} \Big(y - \frac{\bar{p}_1(x)}{\bar{p}_0(x)}\Big)^2 2w(y)\, dy.$$

Since we assumed $c_n < 1/2$, the range of integration always includes one of the intervals $[-1, 0]$ and $[0, 1]$, and $2w$ restricted to either interval is a density. This and the symmetry of w imply

$$\bar{p}_0(x)\bar{p}_2(x) - \bar{p}_1^2(x) \geq \frac{\eta^2\sigma^2}{4} \tag{3.5}$$

with σ^2 the variance of $2w$ restricted to $[0, 1]$. It follows from (3.3) – (3.5) that

$$P\Big(\inf_{0\leq x\leq 1}(\hat{p}_0(x)\hat{p}_2(x) - \hat{p}_1^2(x)) \geq \frac{\eta^2\sigma^2}{8}\Big) \to 1.$$

Hence, with probability tending to one, we can write

$$\hat{\varrho} = \frac{\hat{q}_0\hat{p}_2 - \hat{q}_1\hat{p}_1}{\hat{p}_0\hat{p}_2 - \hat{p}_1^2}.$$

For $i = 1, 2$, let

$$\hat{s}_i = \frac{\hat{p}_i}{\hat{p}_2\hat{p}_0 - \hat{p}_1^2} \quad \text{and} \quad \bar{s}_i = \frac{\bar{p}_i}{\bar{p}_2\bar{p}_0 - \bar{p}_1^2}.$$

It follows from (3.4) and (3.5) that

$$\|\bar{s}_i\| \leq \frac{4\|g\|}{\eta^2\sigma^2} < \infty, \tag{3.6}$$

and from (3.3) – (3.5) that

$$\|\hat{s}_i - \bar{s}_i\| = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big). \tag{3.7}$$

Let

$$A_i(x) = \frac{1}{nc_n} \sum_{j=1}^{n} \varepsilon_j w_i\Big(\frac{X_j - x}{c_n}\Big),$$

$$B_i(x) = \frac{1}{nc_n}\sum_{j=1}^n U_j w_i\Big(\frac{X_j - x}{c_n}\Big),$$

$$C_i(x) = \frac{1}{nc_n}\sum_{j=1}^n \big(\varrho(X_j) - \varrho(x) - \varrho'(x)(X_j - x)\big) w_i\Big(\frac{X_j - x}{c_n}\Big).$$

Using the identities $\hat{q}_i = A_i - (\hat{\vartheta} - \vartheta)^\top B_i + C_i + \varrho\hat{p}_i + c_n\varrho'\hat{p}_{i+1}$, $\hat{p}_0\hat{s}_2 - \hat{p}_1\hat{s}_1 = 1$ and $\hat{p}_1\hat{s}_2 - \hat{p}_2\hat{s}_1 = 0$, we can write, with probability tending to one,

$$\begin{aligned}\hat{\varrho} - \varrho &= (A_0 - (\hat{\vartheta} - \vartheta)^\top B_0 + C_0 + \varrho\hat{p}_0 + c_n\varrho'\hat{p}_1)\hat{s}_2 \\ &\qquad - (A_1 - (\hat{\vartheta} - \vartheta)^\top B_1 + C_1 + \varrho\hat{p}_1 + c_n\varrho'\hat{p}_2)\hat{s}_1 - \varrho \\ &= A_0\hat{s}_2 - A_1\hat{s}_1 - (\hat{\vartheta} - \vartheta)^\top (B_0\hat{s}_2 - B_1\hat{s}_1) + C_0\hat{s}_2 - C_1\hat{s}_1.\end{aligned}$$

Note that under (R) and in view of (3.3) and (3.4) we have

$$\|C_i\| \le c_n^2\|\varrho''\|\|\hat{p}_0\| = O_p(c_n^2), \qquad i = 0, 1. \tag{3.8}$$

Applying Corollary 4.3 with $v = w_i$ and $T_j = e^\top U_j$ with $e \in \mathbb{R}^k$, and utilizing (H), we obtain

$$\|e^\top(B_i - \bar{p}_i\mu)\| = o_p(1), \qquad i = 0, 1,$$

if $c_n^{-1}n^{-\delta}\log n$ is bounded for some $0 < \delta < 1/2$. Hence, using (3.4), (3.6), (3.7), boundedness of μ on $[0, 1]$, and the identity $\bar{p}_0\bar{s}_2 - \bar{p}_1\bar{s}_1 = 1$, we find that

$$\|e^\top(B_0\hat{s}_2 - B_1\hat{s}_1 - \mu)\| = o_p(1). \tag{3.9}$$

Applying Corollary 4.2 with $v = w_i$, $T_j = \varepsilon_j$ and $\delta = 0$, we have

$$\|A_i\| = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big), \qquad i = 0, 1, \tag{3.10}$$

if F has a finite moment of order $\beta > 2$ and $c_n^{-1}n^{-1+2/\beta}\log n$ is bounded. Applying (3.6)–(3.10) to the above representation of $\hat{\varrho} - \varrho$, and setting

$$\hat{c} = A_0\bar{s}_2 - A_1\bar{s}_1, \tag{3.11}$$

we obtain the following result.

Lemma 3.2 *Assume that (H), (G), (R) and (W) hold and that F has a finite moment of order $\beta > 8/3$. Let $nc_n^4 \to 0$ and $c_n^{-1}n^{-\delta}\log n \to 0$ be bounded for some δ in $(1/4, 1/2) \cap (1/4, 1-2/\beta]$. Let $\hat\vartheta$ be $n^{1/2}$-consistent. Then (2.5) holds with*

$$\hat a(u,x) = (\hat\vartheta - \vartheta)^\top (u - \mu(x)) + \hat c(x).$$

The condition $\beta > 8/3$ implies $1 - 2/\beta > 1/4$ and ensures the existence of a sequence of bandwidths with the required properties. For example, we can pick $c_n \sim n^{-\gamma}$ with $1/4 < \gamma < (1-2/\beta) \wedge (1/2)$.

Condition (W) implies that w_i, w_i' and w_i'' are integrable and Lipschitz. Hence, by Corollary 4.2, applied with $v = w_i^{(j)}$ and $T_j = \varepsilon_j$, we obtain, for $j = 0, 1, 2$,

$$\|A_0^{(j)}\| + \|A_1^{(j)}\| = c_n^{-j} O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big)$$

provided F has a finite moment of order $\beta > 2$ and $c_n^{-1}n^{-1+2/\beta}\log n$ is bounded. Furthermore, since g is bounded, we have

$$\|\bar p_0^{(j)}\| + \|\bar p_1^{(j)}\| + \|\bar p_2^{(j)}\| = O(c_n^{-j}).$$

Thus, in view of (3.4) and (3.5),

$$\|\bar s_1^{(j)}\| + \|\bar s_2^{(j)}\| = O(c_n^{-j}).$$

The above rates show that

$$\|\hat c^{(j)}\| = c_n^{-j} O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big), \quad j = 0, 1, 2.$$

Thus, if $c_n^{-1} = O(n^\gamma)$ with $\gamma < 1/3$, then the assumptions of Lemma 3.1 hold for all β_1, β_2 with $0 < 2\beta_1 < 1 - 3\gamma$ and $\beta_2 \le 1/3$. Thus we have the following result for $\hat c$ as defined in (3.11).

Lemma 3.3 *Assume that (G) and (W) hold, F has a finite moment of order $\beta > 2$, and $c_n^{-1} = O(n^\gamma)$ for some $\gamma < 1/3$. Then $P(\hat c \in C_1^\alpha([0,1])) \to 1$ for some $\alpha > 1$.*

Direct calculation show that $E[A_i^2(x)] \le E[\varepsilon^2]\|g\|\|w\|_\infty/(nc_n)$. This and the boundedness of $\bar s_1$ and $\bar s_2$ yield the following result for $\hat c$ as defined in (3.11).

Lemma 3.4 *Assume that (G) and (W) hold and F has a finite second moment. Then*

$$\int \hat{c}^2\, dG = O_p\Big(\frac{1}{nc_n}\Big).$$

Finally, we have the following result.

Lemma 3.5 *Assume that (G) and (W) hold. Then*

$$\int \hat{c}\, dG = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j + o_p(n^{-1/2}). \tag{3.12}$$

Proof: Using $\bar{p}_0\bar{s}_2 - \bar{p}_1\bar{s}_1 = 1$, we find that the integral in (3.12) equals

$$\begin{aligned}
J &= \frac{1}{nc_n}\sum_{j=1}^{n} \varepsilon_j \int \Big(w_0\Big(\frac{X_j - x}{c_n}\Big)\bar{s}_2(x) - w_1\Big(\frac{X_j - x}{c_n}\Big)\bar{s}_1(x)\Big)\, dG(x) \\
&= \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \big(\bar{p}_0(X_j)\bar{s}_2(X_j) - \bar{p}_1(X_j)\bar{s}_1(X_j) - \Delta_0(X_j) + \Delta_1(X_j)\big) \\
&= \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \big(1 - \Delta_0(X_j) + \Delta_1(X_j)\big)
\end{aligned}$$

with

$$\begin{aligned}
\Delta_0(X_j) &= \int \frac{1}{c_n} w_0\Big(\frac{X_j - x}{c_n}\Big)(\bar{s}_2(X_j) - \bar{s}_2(x))\, dG(x) \\
&= \int w_0(y)(\bar{s}_2(X_j) - \bar{s}_2(X_j - c_n y))g(X_j - c_n y)\, dy; \\
\Delta_1(X_j) &= \int \frac{1}{c_n} w_1\Big(\frac{X_j - x}{c_n}\Big)(\bar{s}_1(X_j) - \bar{s}_1(x))\, dG(x) \\
&= \int w_1(y)(\bar{s}_1(X_j) - \bar{s}_1(X_j - c_n y))g(X_j - c_n y)\, dy.
\end{aligned}$$

Thus the assertion (3.12) follows if $\int (\Delta_0 - \Delta_1)^2\, dG = o_p(1)$. Since w and g are bounded, this is implied by

$$\iint_{I_n} (\bar{s}_i(x) - \bar{s}_i(x - c_n y))^2\, dx\, dy \to 0, \quad i = 1, 2,$$

with $I_n = \{(x,y) : 0 \le x \le 1, -1 \le y \le 1, 0 \le x - c_n y \le 1\}$. In view of (3.4) and (3.5), the latter is implied by

$$\iint_{I_n} (\bar{p}_i(x) - \bar{p}_i(x - c_n y))^2 \, dx \, dy \to 0, \quad i = 0, 1, 2.$$

These three integrals can be bounded by a multiple of

$$\sup_{|t| \le c_n} \int (g(x+t) - g(x))^2 \, dx,$$

which converges to zero by continuity of shifts in L_2; see Theorem 9.5 in Rudin (1974). □

4 Auxiliary results

Throughout this section let $Z, Z_1, Z_2, \dots$ be independent and identically distributed m-dimensional random vectors, and, for each x in $\mathbb{R}$, let h_{nx} be a bounded measurable function from $\mathbb{R}^m$ into $\mathbb{R}$.

Proposition 4.1 *Let B_n be a sequence of positive numbers such that $B_n = O(n^\alpha)$ for some $\alpha > 0$. Assume that*

$$\sup_{|x| \le B_n} \|h_{nx}\|_\infty = O\Big(\frac{n}{\log n}\Big), \tag{4.1}$$

$$\sup_{|x| \le B_n} E[h_{nx}^2(Z)] = O\Big(\frac{n}{\log n}\Big), \tag{4.2}$$

and, for positive numbers κ_1, κ_2 and A,

$$\|h_{ny} - h_{nx}\|_\infty \le A n^{\kappa_2} |y - x|^{\kappa_1}, \quad |x|, |y| \le B_n, \quad |y - x| \le 1. \tag{4.3}$$

Then

$$\sup_{|x| \le B_n} \Big| \frac{1}{n} \sum_{j=1}^n h_{nx}(Z_j) - E[h_{nx}(Z)] \Big| = O_p(1). \tag{4.4}$$

Proof: Let $H_n(x)$ denote the expression inside the absolute value in (4.4). We use an inequality of Hoeffding (1963): If $\xi_1, \dots, \xi_n$ are independent random variables that have mean zero and variance σ^2 and are bounded by M, then for $\eta > 0$,

$$P\Big(\Big|\frac{1}{n}\sum_{j=1}^{n}\xi_j\Big| \ge \eta\Big) \le 2\exp\Big(-\frac{n\eta^2}{2\sigma^2 + (2/3)M\eta}\Big).$$

Applying this inequality with $\xi_j = h_{nx}(Z_j) - E[h_{nx}(Z)]$, we obtain for $\eta > 0$:

$$P(|H_n(x)| \ge \eta) \le 2\exp\Big(-\frac{n\eta^2}{2E[h_{nx}^2(Z)] + 2\eta\|h_{nx}\|_\infty}\Big).$$

Thus there is a positive number a such that for all $\eta > 0$,

$$\sup_{|x|\le B_n} P(|H_n(x)| \ge \eta) \le 2\exp\Big(-\frac{\eta^2}{1\vee\eta}a\log n\Big).$$

Now let $x_{nk} = -B_n + 2kB_n n^{-m}$ for $k = 0, 1, \ldots, n^m$, with m an integer greater than $\alpha + \kappa_2/\kappa_1$. The above yields for large enough $\eta > 0$,

$$P\Big(\max_{k=0,\ldots,n^m} |H_n(x_{nk})| > \eta\Big) \le \sum_{k=0}^{n^m} P(|H_n(x_{nk})| > \eta) = o(1).$$

This shows that

$$H_{n,1} = \max_{k=0,\ldots,n^m} |H_n(x_{nk})| = O_p(1).$$

It follows from (4.3) that

$$\begin{aligned} H_{n,2} &= \max_{k=0,\ldots,n^m} \sup_{|x-x_{nk}|\le B_n n^{-m}} |H_n(x) - H_n(x_{nk})| \\ &= O(B_n^{\kappa_1} n^{-m\kappa_1} n^{\kappa_2}) = O_p(1). \end{aligned}$$

In view of the inequality

$$\sup_{|x|\le B_n} |H_n(x)| \le H_{n,1} + H_{n,2}$$

we have the desired result (4.4). □

In the following corollary we interpret $1/\beta$ as zero if β is infinity.

Corollary 4.2 *Assume that the function v is integrable and Hölder with positive exponent κ, the random variable X has a bounded density g, the random variable T is in L_β for some $2 \le \beta \le \infty$, and τg is bounded, where $\tau(X) = E(T^2|X)$. Let $c_n \to 0$ and $c_n^{-1} n^{-1-\delta+2/\beta} \log n$ be bounded for some $\delta \ge 0$. Then, for i.i.d. copies (T_j, X_j) of (T, X), we have*

$$\sup_{0\le x\le 1} \Big| \frac{1}{nc_n} \sum_{j=1}^n \Big(T_j v\Big(\frac{X_j - x}{c_n}\Big) - E\Big[T v\Big(\frac{X-x}{c_n}\Big)\Big]\Big)\Big| = O_p(\zeta_n^{-1/2})$$

with $\zeta_n = n^{1-\delta} c_n / \log n$.

Proof: Set $K = 2\|T\|_{L_\beta}$. Define

$$\begin{aligned}
R_{nj}(x) &= \zeta_n^{1/2} T_j \mathbf{1}[|T_j| \le K n^{1/\beta}] \frac{1}{c_n} v\Big(\frac{X_j - x}{c_n}\Big),\\
S_{nj}(x) &= \zeta_n^{1/2} T_j \mathbf{1}[|T_j| > K n^{1/\beta}] \frac{1}{c_n} v\Big(\frac{X_j - x}{c_n}\Big).
\end{aligned}$$

It suffices to show that

$$\sup_{0\le x\le 1} \Big| \frac{1}{n} \sum_{j=1}^n (R_{nj}(x) - E[R_{nj}(x)])\Big| = O_p(1) \tag{4.5}$$

and

$$\sup_{0\le x\le 1} \Big| \frac{1}{n} \sum_{j=1}^n (S_{nj}(x) - E[S_{nj}(x)])\Big| = o_p(1). \tag{4.6}$$

Statement (4.6) is true for $\beta = \infty$ as then $S_{nj}(x) = 0$. For $\beta < \infty$ we have

$$\begin{aligned}
P\Big(\max_{1\le j\le n} |T_j| > K n^{1/\beta}\Big) &\le \sum_{j=1}^n P(|T_j| > K n^{1/\beta})\\
&\le K^{-\beta} E[|T|^\beta \mathbf{1}[|T| > K n^{1/\beta}]] \to 0
\end{aligned}$$

and thus

$$P\Big(\sup_{0\le x\le 1} \Big| \frac{1}{n} \sum_{j=1}^n S_{nj}(x)\Big| > 0\Big) \le P\Big(\max_{1\le j\le n} |T_j| > K n^{1/\beta}\Big) \to 0.$$

The assumptions on v imply that v is bounded. Hence we also have

$$\begin{aligned}
&\sup_{0\le x\le 1}\Big|\frac{1}{n}\sum_{j=1}^{n} E[S_{nj}(x)]\Big| \\
&\qquad \le n^{(1-\delta)/2}c_n^{-1/2}(\log n)^{-1/2}\|v\|_\infty E[|T|\mathbf{1}[|T|>Kn^{1/\beta}]] \\
&\qquad \le \|v\|_\infty E[|T|^\beta] n^{(1-\delta)/2}c_n^{-1/2}(\log n)^{-1/2}(Kn^{1/\beta})^{1-\beta} \\
&\qquad = O(n^{-(1+\delta)/2+1/\beta}c_n^{-1/2}(\log n)^{-1/2}) = o(1).
\end{aligned}$$

This shows that (4.6) holds for $\beta<\infty$ as well.

To show (4.5) we apply the previous proposition with $B_n = 1$ and $h_{nx}(T_j, X_j) = R_{nj}(x)$. We have

$$\sup_{0\le x\le 1}\|h_{nx}\|_\infty \le K\|v\|_\infty n^{(1-\delta)/2+1/\beta}c_n^{-1/2}(\log n)^{-1/2} = O\Big(\frac{n}{\log n}\Big).$$

Furthermore,

$$\begin{aligned}
\sup_{0\le x\le 1} E[h_{nx}^2(T,X)] &\le \frac{n^{1-\delta}}{c_n\log n}E\Big[\tau(X)v^2\Big(\frac{X-x}{c_n}\Big)\Big] \\
&= \frac{n^{1-\delta}}{\log n}\int v^2(y)\tau(x+c_ny)g(x+c_ny)\,dy \\
&\le \frac{n^{1-\delta}}{\log n}\|\tau g\|_\infty \int v^2(y)\,dy.
\end{aligned}$$

Since v is Hölder with exponent κ, we obtain, with Λ denoting the Hölder constant,

$$\|h_{ny}-h_{nx}\|_\infty \le \Big(\frac{n^{1-\delta}c_n}{\log n}\Big)^{1/2} n^{1/\beta}c_n^{-1-\kappa}\Lambda|y-x|^\kappa \le Cn^{1+\kappa+1/\beta}|y-x|^\kappa.$$

Thus the assumptions of the proposition hold, and we obtain (4.5). □

Corollary 4.3 *Assume that the function v is Hölder with positive exponent κ and has compact support, the random variable X has a bounded density g, the random variable T has a finite second moment, μ is uniformly continuous and τg is bounded, where $\mu(X) = E(T|X)$ and $\tau(X) = E(T^2|X)$. Let $c_n\to 0$ and $c_n^{-1}n^{-\delta}\log n$ be bounded for some $0<\delta<1/2$. Then, for i.i.d. copies (T_j, X_j) of (T, X), we have*

$$\sup_{0\le x\le 1}\Big|\frac{1}{nc_n}\sum_{j=1}^{n} T_j v\Big(\frac{X_j-x}{c_n}\Big) - \mu(x)\int v(y)g(x+c_ny)\,dy\Big| = o_p(1).$$

Proof: Write

$$\frac{1}{c_n}E\Big[Tv\Big(\frac{X-x}{c_n}\Big)\Big] = \int \mu(x+c_n y)v(y)g(x+c_n y)\,dy$$

and note that $\int |v(y)|g(x+c_n y)\,dy \le \|g\| \int |v(y)|\,dy$. The desired result now follows from the uniform continuity of μ and Corollary 4.2 applied with $\beta = 2$. □

Acknowledgments
Anton Schick was supported by NSF Grants DMS 0072174 and DMS 0405791. An earlier version of the paper was finished while two of the authors stayed at the Oberwolfach Institute under the program "Research in Pairs". We thank a referee whose suggestions led us to rewrite the paper completely.

References

[1] M. G. Akritas and I. Van Keilegom, Non-parametric estimation of the residual distribution, *Scand. J. Statist.* **28**, (2001), 549–567.

[2] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58**, (1963), 13–30.

[3] S. Kiwitt, E.-R. Nagel and N. Neumeyer, Empirical likelihood estimators for the error distribution in nonparametric regression models. Technical Report, Ruhr-Universität Bochum, 2005.

[4] H. L. Koul, Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression, *Ann. Math. Statist.* **40**, (1969), 1950–1979.

[5] H. L. Koul, *Weighted Empirical Processes in Dynamic Nonlinear Models, 2nd ed.*, Lecture Notes in Statistics 166, Springer, New York, 2002.

[6] E. Mammen, Empirical process of residuals for high-dimensional linear models, *Ann. Statist.* **24**, (1996), 307–335.

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

[8] M. Ossiander, A central limit theorem under metric entropy with L_2 bracketing, *Ann. Probab.* **15**, (1987), 897–919.

[9] W. Rudin, *Real and Complex Analysis, 2nd ed.*, McGraw-Hill, New York, 1974.

[10] A. Schick, Root-n consistent estimation in partly linear regression models. *Statist. Probab. Lett.*, **28**, (1996), 353–358.

[11] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes, With Applications to Statistics*, Springer, New York, 1996.

Ursula U. Müller
Department of Statistics
Texas A&M University
College Station, TX 77843-3143, USA

Anton Schick
Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902-6000, USA

Wolfgang Wefelmeyer
Mathematisches Institut
Universität zu Köln
Weyertal 86–90
50931 Köln, Germany