



**Estimating the error distribution in non-  
parametric multiple regression  
with applications to model testing**

NATALIE NEUMEYER & INGRID VAN KEILEGOM

**Preprint–No. 2008-01**

**July 2008**

DEPARTMENT MATHEMATIK  
ARBEITSBEREICH MATHEMATISCHE STATISTIK  
UND STOCHASTISCHE PROZESSE

# Estimating the error distribution in nonparametric multiple regression with applications to model testing

Natalie NEUMEYER\*      Ingrid VAN KEILEGOM†

July 7, 2008

## Abstract

In this paper we consider the estimation of the error distribution in a heteroscedastic nonparametric regression model with multivariate covariates. As estimator we consider the empirical distribution function of residuals, which are obtained from multivariate local polynomial fits of the regression and variance functions, respectively. Weak convergence of the empirical residual process to a Gaussian process is proved. We also consider various applications for testing model assumptions in nonparametric multiple regression. The obtained model tests are able to detect local alternatives that converge to zero at  $n^{-1/2}$ -rate, independent of the covariate dimension. We consider in detail a test for additivity of the regression function.

**Key words:** Additive model, Goodness-of-fit, Hypothesis testing, Nonparametric regression, Residual distribution, Semiparametric regression.

---

\*Department of Mathematics, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany, E-mail: neumeyer@math.uni-hamburg.de

†Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, E-mail: ingrid.vankeilegom@uclouvain.be

# 1 Introduction

In mathematical statistics nonparametric regression models constitute very important methods of analyzing relations between observed random variables. In this paper we regard the often neglected case of multivariate covariates, which is of special importance in applications. To this end consider the random vector  $(X, Y)$ , where  $X$  is  $d$ -dimensional and  $Y$  is one dimensional, and suppose the relation between  $X$  and  $Y$  is given by

$$Y = m(X) + \sigma(X)\varepsilon, \quad (1.1)$$

where  $m(\cdot) = E(Y|X = \cdot)$ ,  $\sigma^2(\cdot) = \text{Var}(Y|X = \cdot)$  and where it is assumed that  $\varepsilon$  and  $X$  are independent. We are interested in estimating the distribution of the error  $\varepsilon$ , and in applying this estimated error distribution to develop tests for model assumptions. As an estimator for the error distribution function we consider the empirical distribution of residuals, that are obtained from multivariate local polynomial fits of the regression and variance functions, respectively. We show weak convergence of the corresponding empirical residual process to a Gaussian process. Comparable results in a model with univariate covariates ( $d = 1$ ) were developed by Akritas and Van Keilegom (2001). In the case of multivariate covariates we are only aware of the work by Müller, Schick and Wefelmeyer (2007) for a partially linear model. So far estimating the error distribution in the nonparametric model has not been considered in the literature in the case of multivariate covariates. Moreover the proofs as given by Akritas and Van Keilegom (2001) are not straightforwardly generalized to the case of multivariate covariates.

Akritas and Van Keilegom's (2001) results for the nonparametric regression model with univariate covariates have been successfully applied to develop tests for model assumptions. In this context Van Keilegom, González Manteiga and Sánchez Sellero (2008) consider goodness-of-fit tests for the regression function and Dette, Neumeyer and Van Keilegom (2007) propose a goodness-of-fit test for the variance function. For comparison of several independent regression models, Pardo Fernández, Van Keilegom and González Manteiga (2007) and Pardo Fernández (2007) investigated tests for equality of regression functions and tests for equality of error distributions, respectively. Neumeyer, Dette and Nagel (2005) suggested a goodness-of-fit test for the error distribution.

Thanks to the results on estimation of the error distribution developed in this paper, all of the above tests are also valid in the important case of multivariate covariates. An important advantage of the proposed test statistics is that they are able to detect local alternatives that converge to zero at  $n^{-1/2}$ -rate, independent of the dimension of the

covariate, whereas for the classical tests based on smoothing (with bandwidth parameter  $h$ ) this rate is  $n^{-1/2}h^{-d/4}$ , and this will be substantially slower when  $d$  is large. Moreover the new theory opens various possibilities to test for parametric or semiparametric models in the context of multiple regression. We explain the general idea for these tests and consider testing for additivity of the regression function as detailed example. Here we prove weak convergence of the residual empirical process on which the test statistics are based.

The paper is organized as follows. In Section 2 we define the estimator of the error distribution and give the asymptotic results under regularity conditions. In Section 3 we explain in general how the results can be applied for model testing. The case of testing for additivity of the regression function is considered in detail in Section 4. All proofs are given in an Appendix.

## 2 Estimation of the error distribution

As mentioned in the Introduction, the aim of this section is to propose and study an estimator of the distribution of the error  $\varepsilon$  under model (1.1).

Let  $F_X(x) = P(X \leq x)$  and  $F_\varepsilon(y) = P(\varepsilon \leq y)$  and let  $f_X(x)$  and  $f_\varepsilon(y)$  denote the probability density functions of  $X$  and  $\varepsilon$ . Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be an i.i.d. sample taken from model (1.1), where we denote the components of  $X_i$  by  $(X_{i1}, \dots, X_{id})$  ( $i = 1, \dots, n$ ). We start by estimating the regression function  $m(x)$  and the variance function  $\sigma^2(x)$  for an arbitrary point  $x = (x_1, \dots, x_d)$  in the support  $R_X$  of  $X$  in  $\mathbb{R}^d$ , which we suppose to be compact. We estimate  $m(x)$  by a local polynomial estimator of degree  $p$  [see Fan and Gijbels (1996) or Ruppert and Wand (1994), among others], i.e.  $\widehat{m}(x) = \widehat{\beta}_0$ , where  $\widehat{\beta}_0$  is the first component of the vector  $\widehat{\beta}$ , which is the solution of the local minimization problem

$$\min_{\beta} \sum_{i=1}^n \left\{ Y_i - P_i(\beta, x, p) \right\}^2 K_h(X_i - x), \quad (2.1)$$

where  $P_i(\beta, x, p)$  is a polynomial of order  $p$  built up with all  $0 \leq k \leq p$  products of factors of the form  $X_{ij} - x_j$  ( $j = 1, \dots, d$ ). The vector  $\beta$  is the vector of length  $\sum_{k=0}^p d^k$ , consisting of all coefficients of this polynomial. Here, for  $u = (u_1, \dots, u_d) \in \mathbb{R}^d$ ,  $K(u) = \prod_{j=1}^d k(u_j)$  is a  $d$ -dimensional product kernel,  $k$  is a univariate kernel function,  $h = (h_1, \dots, h_d)$  is a  $d$ -dimensional bandwidth vector converging to zero when  $n$  tends to infinity, and

$K_h(u) = \prod_{j=1}^d k(u_j/h_j)/h_j$ . To estimate  $\sigma^2(x)$ , define

$$\widehat{\sigma}^2(x) = \widehat{\gamma}_0 - \widehat{\beta}_0^2, \quad (2.2)$$

where  $\widehat{\gamma}_0$  is defined in the same way as  $\widehat{\beta}_0$ , but with  $Y_i$  replaced by  $Y_i^2$  in (2.1) ( $i = 1, \dots, n$ ). See also Härdle and Tsybakov (1997), where this estimator is considered for a one-dimensional covariate.

Next, let for  $i = 1, \dots, n$ ,

$$\widehat{\varepsilon}_i = \frac{Y_i - \widehat{m}(X_i)}{\widehat{\sigma}(X_i)},$$

and define the estimator of the error distribution  $F_\varepsilon(y)$  by

$$\widehat{F}_\varepsilon(y) = n^{-1} \sum_{i=1}^n I(\widehat{\varepsilon}_i \leq y). \quad (2.3)$$

We will need the following conditions:

- (C1)  $k$  is a symmetric probability density function supported on  $[-1, 1]$ ,  $k$  is  $d$  times continuously differentiable, and  $k^{(j)}(\pm 1) = 0$  for  $j = 0, \dots, d-1$ .
- (C2)  $h_j$  ( $j = 1, \dots, d$ ) satisfies  $h_j/h \rightarrow c_j$  for some  $0 < c_j < \infty$  and some baseline bandwidth  $h$  satisfying  $nh^{2p+2} \rightarrow 0$  and  $nh^{3d+\delta} \rightarrow \infty$  for some small  $\delta > 0$ .
- (C3) All partial derivatives of  $F_X$  up to order  $2d+1$  exist on the interior of  $R_X$ , they are uniformly continuous and  $\inf_{x \in R_X} f_X(x) > 0$ .
- (C4) All partial derivatives of  $m$  and  $\sigma$  up to order  $p+2$  exist on the interior of  $R_X$ , they are uniformly continuous and  $\inf_{x \in R_X} \sigma(x) > 0$ .
- (C5)  $F_\varepsilon$  is twice continuously differentiable,  $\sup_y |y^2 f'_\varepsilon(y)| < \infty$ , and  $E|Y|^6 < \infty$ .

Note that (C2) implies that the order  $p$  of the local polynomial fit should satisfy  $p+1 > (3d)/2$ , e.g. when  $d = 1$  we can take  $p = 1$ , when  $d = 2$  a local cubic fit suffices, etc.

Also, note that the condition  $nh^{2p+2} \rightarrow 0$  in (C2) comes from the fact that the asymptotic bias, which is of order  $O(h^{p+1})$ , should be asymptotically negligible with respect to terms of order  $O(n^{1/2})$ . However, the order of this bias can be refined, in a similar way as was done in e.g. Fan and Gijbels (1996) (p. 62) when  $d = 1$ , which leads to the following refined condition :  $nh^{2p+4} \rightarrow 0$  when  $p$  is even, and  $nh^{2p+2} \rightarrow 0$  when  $p$  is odd. Taking this refinement into account, we get that  $p = 0$  suffices when  $X$  is one-dimensional, and this coincides with what has been done in Akritas and Van Keilegom (2001).

We are now ready to state the two main results of this Section.

**Theorem 2.1** *Assume (C1)-(C5). Then,*

$$\widehat{F}_\varepsilon(y) - F_\varepsilon(y) = n^{-1} \sum_{i=1}^n \left\{ I(\varepsilon_i \leq y) - F_\varepsilon(y) + \varphi(\varepsilon_i, y) \right\} + R_n(y),$$

where  $\sup_{-\infty < y < \infty} |R_n(y)| = o_P(n^{-1/2})$  and

$$\varphi(z, y) = f_\varepsilon(y) \left\{ z + \frac{y}{2} [z^2 - 1] \right\}.$$

**Corollary 2.2** *Assume (C1)-(C5). Then, the process  $n^{1/2}(\widehat{F}_\varepsilon(y) - F_\varepsilon(y))$  ( $-\infty < y < \infty$ ) converges weakly to a zero-mean Gaussian process  $Z(y)$  with covariance function*

$$\text{Cov}(Z(y_1), Z(y_2)) = E \left[ \{I(\varepsilon \leq y_1) - F_\varepsilon(y_1) + \varphi(\varepsilon, y_1)\} \{I(\varepsilon \leq y_2) - F_\varepsilon(y_2) + \varphi(\varepsilon, y_2)\} \right].$$

Note that the above results are considerably more difficult to obtain than the corresponding results for  $d = 1$  obtained by Akritas and Van Keilegom (2001) by using local constant estimators. A direct extension of the results in the latter paper to dimensions  $d$  larger than one is in fact not possible because that would lead to two contradictory conditions on the bandwidth, namely on the one hand  $nh^{3d+\delta} \rightarrow \infty$ , and on the other hand  $nh^4 \rightarrow 0$  (where the latter condition comes from the bias term). It was therefore necessary to consider other estimators of  $m$  and  $\sigma$ , that improve the rate of convergence of the bias term. We chose to use local polynomial estimators, because of their nice bias properties and also because of their excellent behavior in practice, which has been widely demonstrated in the literature (see e.g. Fan and Gijbels (1996)). Also note that without any exception, all papers in the literature related to estimation or testing problems involving the nonparametric estimation of the error distribution, are developed for  $d = 1$ .

### 3 Model tests

Tests for various hypotheses can be based on the suggested estimated error distribution. Due to the curse of dimensionality in nonparametric multiple regression, investigators often prefer parametric models

$$m \in \{m_\vartheta \mid \vartheta \in \Theta\}, \tag{3.1}$$

or semiparametric models such as partially linear models,

$$m(x_1, \dots, x_d) = \beta_1 x_1 + \dots + \beta_{d-1} x_{d-1} + g(x_d) \quad \text{for some } \beta_1, \dots, \beta_{d-1} \in \mathbb{R}, \tag{3.2}$$

single index models

$$m(x_1, \dots, x_d) = g(\beta_1 x_1 + \dots + \beta_d x_d) \quad \text{for some } \beta_1, \dots, \beta_d \in \mathbb{R}, \quad (3.3)$$

or additive models

$$m(x_1, \dots, x_d) = m_1(x_1) + \dots + m_d(x_d) \quad (3.4)$$

with univariate nonparametric functions  $g$  and  $m_1, \dots, m_d$ , respectively. Hence there is a great interest in goodness-of-fit tests for the regression function. The results displayed in Section 2 can be applied to test for each of the hypotheses (3.1)–(3.4), and in the next section we consider in detail the testing for an additive regression model (3.4). General testing procedures for semiparametric regression models have also been considered by Rodríguez-Póo, Sperlich and Vieu (2005) and Chen and Van Keilegom (2006). Their tests are based on smoothing techniques (with a bandwidth  $h$ ), and they are able to detect local alternatives of the order  $n^{-1/2}h^{-d/4}$ . The tests proposed here can detect however local alternatives that converge to zero at  $n^{-1/2}$ -rate, which will be substantially faster when  $d$  is large.

The general idea to apply the new results for hypotheses testing is to compare the estimated error distribution  $\widehat{F}_\varepsilon$  under the full nonparametric model (as in Section 2) with the empirical distribution function  $\widetilde{F}_\varepsilon$  of residuals estimated under the null model, and to apply Kolmogorov-Smirnov or Cramér-von Mises tests based on the process

$$\sqrt{n}(\widehat{F}_\varepsilon(\cdot) - \widetilde{F}_\varepsilon(\cdot)),$$

which converges to a Gaussian process. Analogous tests for hypothesis (3.1) were proposed by Van Keilegom, González Manteiga and Sánchez Sellero (2008) when the covariate is one-dimensional. Similar tests for the hypothesis

$$\sigma^2 \in \{\sigma_\vartheta^2 \mid \vartheta \in \Theta\}$$

for the variance function (which includes tests for homoscedasticity as special case) were considered by Dette, Neumeyer and Van Keilegom (2007), whereas Neumeyer, Dette and Nagel (2005) test goodness-of-fit of the error distribution, i. e.

$$F_\varepsilon \in \{F_\vartheta \mid \vartheta \in \Theta\}.$$

Thanks to the results of Section 2, the three latter tests are now also valid when the covariate is multi-dimensional.

Moreover the tests by Pardo Fernández, Van Keilegom and González Manteiga (2007) and Pardo Fernández (2007) for equality of regression functions, i. e.

$$m_1 = \cdots = m_k,$$

and equality of error distributions, i. e.

$$F_{\varepsilon_1} = \cdots = F_{\varepsilon_k},$$

respectively, are now carried over to the case of multivariate covariates. Those tests are in the context of  $k$  independent regression models

$$Y_{ij} = m_i(X_{ij}) + \sigma_i(X_{ij})\varepsilon_{ij}, \quad j = 1, \dots, n_i, i = 1, \dots, k.$$

All the considered tests provide the possibility to detect local alternatives of rate  $n^{-1/2}$ , independent of the covariate dimension  $d$ . As explained already above, this is in big contrast with smoothing based tests, that are based e.g. on the  $L^2$ -distance between the nonparametrically and parametrically estimated regression function in the case of hypothesis (3.1). These tests usually can only detect local alternatives of the rate  $n^{-1/2}h^{-d/4}$ .

## 4 Testing for additivity

In this section we consider in detail the application of the residual-based empirical process to testing additivity of a multivariate regression function. Different tests for additivity of regression models were proposed by Gozalo and Linton (2001), Dette and von Lieres und Wilkau (2001), Yang, Park, Xue and Härdle (2006), among others. Our aim is to test validity of the hypothesis  $H_0$  of an additive regression model, i. e.

$$H_0 : m(x_1, \dots, x_d) = m_1(x_1) + \cdots + m_d(x_d) + c \quad \text{for all } (x_1, \dots, x_d) \in R_X, \quad (4.1)$$

against the general nonparametric alternative as considered in Section 2. Here in model (4.1) we assume  $E[m_\ell(X_{i\ell})] = 0$  for all  $\ell = 1, \dots, d$  to identify the univariate regression functions. Let  $\hat{m}$ ,  $\hat{\sigma}^2$  and  $\hat{F}_\varepsilon$  denote the estimators defined in (2.1), (2.2) and (2.3), respectively. Further, let  $X_{i,-\ell} = (X_{i1}, \dots, X_{i,\ell-1}, X_{i,\ell+1}, \dots, X_{id})$  and denote its density by  $f_{X_{-\ell}}$ , whereas the density of  $X_{i\ell}$  is denoted by  $f_{X_\ell}$ . To estimate the additive regression components we apply the marginal integration estimator [see Newey (1994), Tjøstheim and Auestad (1994), Linton and Nielsen (1995)] and define

$$\hat{m}_\ell(x_\ell) = \frac{1}{n} \sum_{j=1}^n \hat{m}(X_{j1}, \dots, X_{j,\ell-1}, x_\ell, X_{j,\ell+1}, \dots, X_{jd}) - \bar{Y}_n, \quad (4.2)$$

where  $\bar{Y}_n = n^{-1} \sum_{j=1}^n Y_j$ . Let

$$\tilde{m}(x_1, \dots, x_d) = \hat{m}_1(x_1) + \dots + \hat{m}_d(x_d) + \bar{Y}_n \quad (4.3)$$

denote the additive regression estimator and let  $\tilde{F}_\varepsilon$  be the empirical distribution function of residuals

$$\tilde{\varepsilon}_i = \frac{Y_i - \tilde{m}(X_i)}{\hat{\sigma}(X_i)}, \quad i = 1, \dots, n,$$

estimated under the null model. Tests for additivity can be based on Kolmogorov-Smirnov or Cramér-von Mises type functionals of the process  $n^{1/2}(\hat{F}_\varepsilon - \tilde{F}_\varepsilon)$ , given by

$$T_{KS} = n^{1/2} \sup_{-\infty < y < \infty} |\hat{F}_\varepsilon(y) - \tilde{F}_\varepsilon(y)|$$

$$T_{CM} = n \int (\hat{F}_\varepsilon(y) - \tilde{F}_\varepsilon(y))^2 d\hat{F}_\varepsilon(y).$$

Please note that  $\hat{F}_\varepsilon$  consistently estimates  $F_\varepsilon$ , whereas  $\tilde{F}_\varepsilon$  consistently estimates the distribution  $F_{\tilde{\varepsilon}}$  of  $\tilde{\varepsilon}_i = (Y_i - m_1(X_{i1}) - \dots - m_d(X_{id}) - c)\sigma^{-1}(X_i)$  for  $c = E[Y_i]$  and  $m_\ell(x_\ell) = E[m(X_{i1}, \dots, X_{i,\ell-1}, x_\ell, X_{i,\ell+1}, \dots, X_{id})]$ , where  $m(x) = E[Y_i | X_i = x]$ . Exactly as in Theorem 2.1 by Van Keilegom, González Manteiga and Sánchez Sellero (2008) it follows that  $F_\varepsilon = F_{\tilde{\varepsilon}}$  is equivalent to the null hypothesis (4.1).

To derive the following asymptotic results we will need an additional assumption.

(C6) All partial derivatives of  $f_X$  and  $f_{X_{-\ell}}$  ( $\ell = 1, \dots, d$ ) up to order  $p + 1$  exist and are uniformly continuous.

**Theorem 4.1** *Assume (C1)-(C6) and the null hypothesis (4.1). Then,*

$$\hat{F}_\varepsilon(y) - \tilde{F}_\varepsilon(y) = \frac{f_\varepsilon(y)}{n} \sum_{i=1}^n \varepsilon_i H(X_i) + \tilde{R}_n(y),$$

where  $\sup_{-\infty < y < \infty} |\tilde{R}_n(y)| = o_P(n^{-1/2})$ . Here,  $H$  is defined by

$$H(X_i) = 1 - \sum_{\ell=1}^d \sigma(X_i) \frac{g_\ell(X_{i\ell}) f_{X_{-\ell}}(X_{i,-\ell})}{f_X(X_i)} + (d-1) \sigma(X_i) \int \frac{f_X(x)}{\sigma(x)} dx,$$

where for  $\ell = 1, \dots, d$ ,

$$g_\ell(X_{i\ell}) = \int \frac{f_X(x_1, \dots, x_{\ell-1}, X_{i\ell}, x_{\ell+1}, \dots, x_d)}{\sigma(x_1, \dots, x_{\ell-1}, X_{i\ell}, x_{\ell+1}, \dots, x_d)} d(x_1, \dots, x_{\ell-1}, x_{\ell+1}, \dots, x_d).$$

**Corollary 4.2** *Assume (C1)-(C6) and the null hypothesis (4.1). Then, the process  $n^{1/2}(\widehat{F}_\varepsilon(y) - \widetilde{F}_\varepsilon(y))$  ( $-\infty < y < \infty$ ) converges weakly to  $f_\varepsilon(y)Z$ , where  $Z$  is a zero-mean normal random variable with variance  $\text{Var}(Z) = E[H^2(X)]$ .*

Please note that for a univariate model ( $d = 1$ ) the limiting process is degenerate as could be expected because then each regression model is ‘additive’. Further in models with homoscedastic variance the dominating part of the expansion in Theorem 4.1 simplifies to

$$\frac{f_\varepsilon(y)}{n} \sum_{i=1}^n \varepsilon_i \sum_{\ell=1}^d \left\{ 1 - \frac{f_{X_\ell}(X_{i\ell})f_{X_{-\ell}}(X_{i,-\ell})}{f_X(X_i)} \right\}$$

which vanishes in the case where all covariate components are independent.

**Corollary 4.3** *Assume (C1)-(C6) and the null hypothesis (4.1). Then,*

$$T_{KS} \xrightarrow{d} \sup_{-\infty < y < \infty} f_\varepsilon(y)|Z|$$

$$T_{CM} \xrightarrow{d} \int f_\varepsilon^2(y)dF_\varepsilon(y) Z^2,$$

where  $Z$  is defined in Corollary 4.2.

The proof of Corollary 4.3 is very similar to the proof of Corollary 3.3 in Van Keilegom, González Manteiga and Sánchez Sellero (2008) and is therefore omitted.

To apply the test we recommend the application of smooth residual bootstrap. A description of the method and asymptotic theory for the univariate case can be found in Neumeyer (2006).

**Remark 4.4** Consider the local alternative

$$H_{1n} : m(x_1, \dots, x_d) = m_1(x_1) + \dots + m_d(x_d) + c + n^{-1/2}r(x_1, \dots, x_d),$$

for all  $(x_1, \dots, x_d) \in R_X$ , where  $E[m_\ell(X_{i\ell})] = 0$  ( $\ell = 1, \dots, d$ ) and the function  $r$  satisfies  $E(r^2(X)) < \infty$ . Then, it can be shown that under  $H_{1n}$ ,

$$T_{KS} \xrightarrow{d} \sup_{-\infty < y < \infty} f_\varepsilon(y)|Z + b|$$

$$T_{CM} \xrightarrow{d} \int f_\varepsilon^2(y)dF_\varepsilon(y) (Z + b)^2,$$

for some  $b \in \mathbb{R}$ . The proof is similar to the proof of Theorem 3.4 in Van Keilegom, González Manteiga and Sánchez Sellero (2008), and we therefore refer to that paper for more details.

**Remark 4.5** Assume we want to test for a separable model with regression function

$$m(x_1, \dots, x_d) = G(m_1(x_1), \dots, m_d(x_d))$$

with known link function  $G$ , where functions  $q_\ell$  ( $\ell = 1, \dots, d$ ) are known (or consistently estimable) such that  $\int G(m_1(x_1), \dots, m_d(x_d))q_\ell(x_{-\ell}) dx_{-\ell} = m_\ell(x_\ell)$ , where  $x_{-\ell} = (x_1, \dots, x_{\ell-1}, x_{\ell+1}, \dots, x_d)$ . With

$$\widehat{m}_\ell(x_\ell) = \int \widehat{m}(x_1, \dots, x_d) \widehat{q}_\ell(x_{-\ell}) dx_{-\ell}$$

and  $\widetilde{m}(x_1, \dots, x_d) = G(\widehat{m}_1(x_1), \dots, \widehat{m}_d(x_d))$  the analogous testing procedure as explained for testing additivity can be applied.

**Remark 4.6** Combining the methods developed in Section 2 with those considered by Pardo Fernández, Van Keilegom and González Manteiga (2007) and Neumeyer and Sperlich (2006) one can also test for equality of additive components, when  $k$  regression models

$$Y_{ij} = m_i(X_{ij}) + \sigma_i(X_{ij})\varepsilon_{ij}, \quad j = 1, \dots, n_i, i = 1, \dots, k,$$

with additive structure

$$m_i(X_{ij}) = m_i(Z_{ij}, W_{ij}) = r_i(Z_{ij}) + g_i(W_{ij}), \quad E[r_i(Z_{ij})] = 0, \quad i = 1, \dots, k,$$

are given and one is interested in the hypothesis

$$H_0 : r_1 = r_2 = \dots = r_k.$$

For  $i = 1, \dots, k$  denote by  $\widehat{F}_{\varepsilon_i}$  the empirical distribution function of the residuals  $\widehat{\varepsilon}_{ij} = (Y_{ij} - \widehat{m}(X_{ij}))/\widehat{\sigma}(X_{ij})$  ( $j = 1, \dots, n_i$ ), and by  $\widetilde{F}_{\varepsilon_i}$  the empirical distribution function of the residuals  $\widetilde{\varepsilon}_{ij} = (Y_{ij} - \widehat{r}(Z_{ij}) - \widehat{g}_i(W_{ij}))/\widehat{\sigma}(X_{ij})$ , where  $\widehat{g}_i$  denotes the marginal integration estimator for  $g_i$  (within the  $i$ th sample), and  $\widehat{r}$  denotes a marginal integration estimator for  $r = r_1 = \dots = r_k$  under  $H_0$  obtained from the pooled sample [see Neumeyer and Sperlich (2006)]. A test can be obtained from comparing  $\widehat{F}_{\varepsilon_i}$  with  $\widetilde{F}_{\varepsilon_i}$  ( $i = 1, \dots, k$ ) in the same manner as shown by Pardo Fernández, Van Keilegom and González Manteiga (2007) in the context of testing for equality of regression functions.

## Appendix: Proofs

In this Appendix the proofs will be given of the main theorems and of several lemmas.

**Lemma A.1** Assume (C1)-(C5). Then,

$$\|\widehat{m} - m\|_{d+\alpha} = o_P(1), \quad \|\widehat{\sigma} - \sigma\|_{d+\alpha} = o_P(1),$$

where  $0 < \alpha < \delta/2$ ,  $\delta$  is defined as in condition (C2), and where for any function  $f$  defined on  $R_X$ ,

$$\|f\|_{d+\alpha} = \max_{k \leq d} \sup_{x \in R_X} |D^k f(x)| + \max_{k=d} \sup_{x, x' \in R_X} \frac{|D^k f(x) - D^k f(x')|}{\|x - x'\|^\alpha},$$

$$k = (k_1, \dots, k_d),$$

$$D^k = \frac{\partial^k}{\partial x_1^{k_1} \dots \partial x_d^{k_d}},$$

$$k. = \sum_{j=1}^d k_j, \text{ and } \|\cdot\| \text{ is the Euclidean norm on } \mathbb{R}^d.$$

**Proof.** First note that it follows from Theorem 6 in Masry (1996) that

$$\sup_{x \in R_X} |\widehat{m}(x) - m(x)| = O_P((nh^d)^{-1/2}(\log n)^{1/2}) + O(h^{p+1}) = o_P(1).$$

Next, note that based on this result it can now be shown that for  $k. \leq d$ ,

$$\sup_{x \in R_X} |D^k \widehat{m}(x) - D^k m(x)| = O_P((nh^{d+2k.})^{-1/2}(\log n)^{1/2}) + O(h^{p+1-k.}) = o_P(1),$$

and for  $k. = d$ ,

$$\begin{aligned} & \sup_{x, x' \in R_X} \frac{|D^k \widehat{m}(x) - D^k m(x) - D^k \widehat{m}(x') + D^k m(x')|}{\|x - x'\|^\alpha} \\ &= O_P((nh^{3d+2\alpha})^{-1/2}(\log n)^{1/2}) + O(h^{p+1-d-\alpha}) = o_P(1). \end{aligned}$$

A detailed proof of the latter two results can be found in Proposition 3.2 and Theorem 3.2 in Ojeda (2008) respectively for the case where  $d = 1$ . In the multiple regression case, the proof is similar but more technical, and is therefore omitted.

For  $\widehat{\sigma} - \sigma$ , note that we can again apply Theorem 6 in Masry (1996), but with  $Y_i$  replaced by  $Y_i^2$  ( $i = 1, \dots, n$ ). Hence, the same reasoning as for  $\widehat{m} - m$  applies.  $\square$

**Lemma A.2** Assume (C1)-(C5). Then,

$$\int \frac{\widehat{m}(x) - m(x)}{\sigma(x)} f_X(x) dx = n^{-1} \sum_{i=1}^n \varepsilon_i + o_P(n^{-1/2}),$$

and

$$\int \frac{\widehat{\sigma}(x) - \sigma(x)}{\sigma(x)} f_X(x) dx = \frac{1}{2} n^{-1} \sum_{i=1}^n \{\varepsilon_i^2 - 1\} + o_P(n^{-1/2}).$$

**Proof.** We prove the first statement. The second one can be shown in a very similar way. The proof is based on the notion of ‘equivalent kernels’, introduced by Fan and Gijbels (1996), p. 63-64. Although their development is valid in the case where  $d = 1$ , it can be seen (but the proof is technical) that equivalent kernels can be extended to the context of multiple regression. In fact, for the extension to  $d > 1$  one needs to stack the  $\sum_{k=0}^p d^k$  coefficients of the local polynomial of order  $p$  into one big vector, and apply the same kind of development as in the case  $d = 1$ . Details are omitted. This development yields that the estimator  $\widehat{m}(x)$  can be written in the form of a Nadaraya-Watson type estimator:

$$\widehat{m}(x) = \sum_{i=1}^n W_0^n\left(\frac{x - X_i}{h}\right) Y_i,$$

where the weights  $W_0^n(\cdot)$  depend on  $x$  and satisfy

$$\sum_{i=1}^n W_0^n\left(\frac{x - X_i}{h}\right) = 1 \quad \text{and} \quad W_0^n(u) = \frac{1}{nh} \frac{1}{f_X(x)} K^*(u) \{1 + o_P(1)\}$$

uniformly in  $u \in [-1, 1]^d$  and  $x \in R_X$ . Here,  $K^*(\cdot)$  is the so-called equivalent kernel and is a product kernel  $K^*(u_1, \dots, u_d) = \prod_{j=1}^d k^*(u_j)$ , where  $k^*$  is a univariate kernel satisfying

$$\int u^q k^*(u) du = \delta_{0q} \quad (0 \leq q \leq p). \quad (\text{A.1})$$

It now follows that we can write

$$\widehat{m}(x) - m(x) = n^{-1} f_X^{-1}(x) \sum_{i=1}^n K_h^*(x - X_i) (Y_i - m(x)) \{1 + o_P(1)\},$$

and hence (we take  $\sigma \equiv 1$  for simplicity)

$$\begin{aligned} & \int (\widehat{m}(x) - m(x)) f_X(x) dx \\ &= n^{-1} \sum_{i=1}^n \int K_h^*(x - X_i) (Y_i - m(x)) dx \{1 + o_P(1)\} \\ &= \left[ n^{-1} \sum_{i=1}^n \varepsilon_i + O(h^{p+1}) \right] \{1 + o_P(1)\} = n^{-1} \sum_{i=1}^n \varepsilon_i + o_P(n^{-1/2}), \end{aligned}$$

where the second equality follows from a Taylor expansion of  $m(X_i) - m(x)$  of order  $p + 1$  and from equation (A.1), and where the last equality follows from condition (C2).  $\square$

**Lemma A.3** *Assume (C1)-(C5). Then,*

$$\sup_{-\infty < y < \infty} \left| n^{-1} \sum_{i=1}^n \left\{ I(\widehat{\varepsilon}_i \leq y) - I(\varepsilon_i \leq y) - F_{\widehat{\varepsilon}}(y) + F_{\varepsilon}(y) \right\} \right| = o_P(n^{-1/2}),$$

where  $F_{\widehat{\varepsilon}}$  is the distribution of  $\widehat{\varepsilon} = (Y - \widehat{m}(X))/\widehat{\sigma}(X)$  conditionally on the data  $(X_j, Y_j)$ ,  $j = 1, \dots, n$  (i.e. considering  $\widehat{m}$  and  $\widehat{\sigma}$  as fixed functions).

**Proof.** The proof is very similar to that of Lemma 1 in Akritas and Van Keilegom (2001). Therefore, we will restrict attention to explaining the main changes with respect to that proof. First note that Lemma A.1 implies that, with probability tending to one,  $(\widehat{m} - m)/\sigma$  and  $\widehat{\sigma}/\sigma$  belong to  $C_1^{d+\alpha}(R_X)$  and  $\widetilde{C}_2^{d+\alpha}(R_X)$  respectively, where  $0 < \alpha < \delta/2$  is as in Lemma A.1. Here,  $C_1^{d+\alpha}(R_X)$  is the class of  $d$  times differentiable functions  $f$  defined on  $R_X$  such that  $\|f\|_{d+\alpha} \leq 1$  (with  $\|f\|_{d+\alpha}$  defined in Lemma A.1), and  $\widetilde{C}_2^{d+\alpha}(R_X)$  is the class of  $d$  times differentiable functions  $f$  defined on  $R_X$  such that  $\|f\|_{d+\alpha} \leq 2$  and  $\inf_{x \in R_X} f(x) \geq 1/2$ .

Next, note that for any  $\bar{\varepsilon} > 0$ , the  $\bar{\varepsilon}^2$ -bracketing numbers of these two classes are bounded by

$$\begin{aligned} N_{[]}(\bar{\varepsilon}^2, C_1^{d+\alpha}(R_X), L_2(P)) &\leq \exp(K\bar{\varepsilon}^{-2d/(d+\alpha)}) \\ N_{[]}(\bar{\varepsilon}^2, \widetilde{C}_2^{d+\alpha}(R_X), L_2(P)) &\leq \exp(K\bar{\varepsilon}^{-2d/(d+\alpha)}), \end{aligned}$$

where  $P$  is the joint probability measure of  $(X, \varepsilon)$  and  $K > 0$  [see Theorem 2.7.1 in van der Vaart and Wellner (1996)]. It now follows using the same arguments as in Akritas and Van Keilegom (2001) that the  $\bar{\varepsilon}$ -bracketing number of the class

$$\mathcal{F}_1 = \left\{ (x, e) \rightarrow I(e \leq y f_2(x) + f_1(x)) : y \in \mathbb{R}, f_1 \in C_1^{d+\alpha}(R_X) \text{ and } f_2 \in \widetilde{C}_2^{d+\alpha}(R_X) \right\}$$

is at most  $O(\bar{\varepsilon}^{-2} \exp(K\bar{\varepsilon}^{-2d/(d+\alpha)}))$  and hence

$$\int_0^\infty \sqrt{\log N_{[]}(\bar{\varepsilon}, \mathcal{F}_1, L_2(P))} d\bar{\varepsilon} < \infty.$$

The rest of the proof is now exactly the same as in Akritas and Van Keilegom (2001) and is therefore omitted.  $\square$

**Proof of Theorem 2.1.** From Lemma A.3 it follows that

$$\begin{aligned}
& \widehat{F}_\varepsilon(y) - F_\varepsilon(y) \\
&= n^{-1} \sum_{i=1}^n I(\varepsilon_i \leq y) - F_\varepsilon(y) \\
&\quad + \int \left\{ F_\varepsilon \left( \frac{y\widehat{\sigma}(x) + \widehat{m}(x) - m(x)}{\sigma(x)} \right) - F_\varepsilon(y) \right\} dF_X(x) + o_P(n^{-1/2}) \\
&= n^{-1} \sum_{i=1}^n I(\varepsilon_i \leq y) - F_\varepsilon(y) \\
&\quad + f_\varepsilon(y) \int \sigma^{-1}(x) \{y(\widehat{\sigma}(x) - \sigma(x)) + \widehat{m}(x) - m(x)\} dF_X(x) \\
&\quad + \frac{1}{2} \int f'_\varepsilon(\xi_x) \sigma^{-2}(x) \{y(\widehat{\sigma}(x) - \sigma(x)) + \widehat{m}(x) - m(x)\}^2 dF_X(x) + o_P(n^{-1/2}),
\end{aligned}$$

for some  $\xi_x$  between  $y$  and  $\sigma^{-1}(x)\{y\widehat{\sigma}(x) + \widehat{m}(x) - m(x)\}$ . The third term above is  $O_P((nh^d)^{-1} \log n) = o_P(n^{-1/2})$ , which follows from the proof of Lemma A.1 and from conditions (C2) and (C5), while the second term equals  $n^{-1} \sum_{i=1}^n \varphi(\varepsilon_i, y) + o_P(n^{-1/2})$ , which follows from Lemma A.2.  $\square$

**Proof of Corollary 2.2.** The proof is very similar to that of Theorem 2 in Akritas and Van Keilegom (2001) and is therefore omitted.  $\square$

**Lemma A.4** *Assume (C1)-(C5) and (4.1). Then, for  $\widetilde{m}$  defined in (4.3),*

$$\|\widetilde{m} - m\|_{d+\alpha} = o_P(1),$$

where  $0 < \alpha < \delta/2$ , with  $\delta$  defined in condition (C2), and  $\|\cdot\|_{d+\alpha}$  defined in Lemma A.1.

**Proof.** From (4.3) and (4.1) we have for  $x = (x_1, \dots, x_d)$

$$\widetilde{m}(x) - m(x) = \sum_{\ell=1}^d (\widehat{m}_\ell(x_\ell) - m_\ell(x_\ell)) + \bar{Y}_n - c,$$

where  $\bar{Y}_n - c = n^{-1} \sum_{i=1}^n (Y_i - E[Y_i]) = O_P(n^{-1/2})$ , and it does not depend on  $x$ . For the ease of notation we consider  $\widehat{m}_\ell(x_\ell) - m_\ell(x_\ell)$  in detail for  $\ell = 1$ . With the assumption  $E[m_\ell(X_{i\ell})] = 0$  ( $\ell = 1, \dots, d$ ) we obtain

$$\begin{aligned}
\widehat{m}_1(x_1) - m_1(x_1) &= \frac{1}{n} \sum_{j=1}^n \left( \widehat{m}(x_1, X_{j2}, \dots, X_{jd}) - m(x_1, X_{j2}, \dots, X_{jd}) \right) \\
&\quad + \sum_{\ell=2}^d \frac{1}{n} \sum_{j=1}^n m_\ell(X_{j\ell}) + c - \bar{Y}_n,
\end{aligned}$$

where the terms in the last line are of order  $O_P(n^{-1/2})$  and independent of  $x_1$ . Hence, we obtain

$$\sup_{x_1} |\widehat{m}_1(x_1) - m_1(x_1)| \leq \sup_x |\widehat{m}(x) - m(x)| + O_P(n^{-1/2}) = o_P(1)$$

by Lemma A.1. For the derivatives note that for  $k \leq d$ , and for  $j = 1, \dots, n$ ,

$$m_1^{(k)}(x_1) = \frac{\partial^k m(x_1, X_{j2}, \dots, X_{jd})}{\partial x_1^k}, \quad \widehat{m}_1^{(k)}(x_1) = \frac{1}{n} \sum_{j=1}^n \frac{\partial^k \widehat{m}(x_1, X_{j2}, \dots, X_{jd})}{\partial x_1^k},$$

and hence

$$\sup_{x_1} |\widehat{m}_1^{(k)}(x_1) - m_1^{(k)}(x_1)| \leq \sup_x \left| \frac{\partial^k \widehat{m}(x)}{\partial x_1^k} - \frac{\partial^k m(x)}{\partial x_1^k} \right| = o_P(1)$$

by Lemma A.1. Further, we have

$$\begin{aligned} & \sup_{x_1, x'_1} \frac{|\widehat{m}_1^{(k)}(x_1) - m_1^{(k)}(x_1) - \widehat{m}_1^{(k)}(x'_1) + m_1^{(k)}(x'_1)|}{|x_1 - x'_1|^\alpha} \\ & \leq \sup_{x_1, x'_1} \frac{1}{n} \sum_{j=1}^n \frac{\left| \frac{\partial^k \widehat{m}(x_1, X_{j,-1})}{\partial x_1^k} - \frac{\partial^k m(x_1, X_{j,-1})}{\partial x_1^k} - \frac{\partial^k \widehat{m}(x'_1, X_{j,-1})}{\partial x_1^k} + \frac{\partial^k m(x'_1, X_{j,-1})}{\partial x_1^k} \right|}{\|(x_1, X_{j,-1})^t - (x'_1, X_{j,-1})^t\|^\alpha} \\ & \leq \sup_{x, x'} \frac{|D^{(k,0,\dots,0)}(\widehat{m} - m)(x) - D^{(k,0,\dots,0)}(\widehat{m} - m)(x')|}{\|x - x'\|^\alpha} = o_P(1) \end{aligned}$$

by Lemma A.1. □

**Lemma A.5** *Assume (C1)-(C6) and (4.1). Then, for  $\ell = 1, \dots, d$ ,*

$$\begin{aligned} \int \frac{\widehat{m}_\ell(x_\ell) - m_\ell(x_\ell)}{\sigma(x)} dF_X(x) &= \frac{1}{n} \sum_{i=1}^n \sigma(X_i) \varepsilon_i \frac{g_\ell(X_{i\ell}) f_{X_{-\ell}}(X_{i,-\ell})}{f_X(X_i)} \\ &\quad - \int \frac{f_X(x)}{\sigma(x)} dx \frac{1}{n} \sum_{i=1}^n \left[ \sigma(X_i) \varepsilon_i + m_\ell(X_{i\ell}) \right] + o_P(n^{-1/2}) \end{aligned}$$

where  $g_\ell$  is defined in Theorem 4.1.

**Proof.** For ease of notation we only consider the case  $\ell = 1$ . Then from (A.2) and the proof of Lemma A.2 we have by standard arguments (using condition (C6))

$$\begin{aligned}
& \int \frac{\widehat{m}_1(x_1) - m_1(x_1)}{\sigma(x)} dF_X(x) \\
&= \int \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sigma(X_i) \varepsilon_i \frac{K_h^*(x_1 - X_{i1}, X_{j2} - X_{i2}, \dots, X_{jd} - X_{id})}{f_X(x_1, X_{j2}, \dots, X_{jd})} \frac{dF_X(x)}{\sigma(x)} \\
&\quad + \int \frac{f_X(x)}{\sigma(x)} dx \left[ (c - \bar{Y}_n) + \sum_{k=2}^d \frac{1}{n} \sum_{j=1}^n m_k(X_{jk}) \right] + o_P(n^{-1/2}) \\
&= \int \frac{1}{nh_1} \sum_{i=1}^n \sigma(X_i) \varepsilon_i k^* \left( \frac{x_1 - X_{i1}}{h_1} \right) \frac{f_{X_{-1}}(X_{i2}, \dots, X_{id})}{f_X(x_1, X_{i2}, \dots, X_{id})} \frac{dF_X(x)}{\sigma(x)} \\
&\quad + \int \frac{f_X(x)}{\sigma(x)} dx \frac{1}{n} \sum_{j=1}^n \left[ c - Y_j + m_2(X_{j2}) + \dots + m_d(X_{jd}) \right] + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n \sigma(X_i) \varepsilon_i \frac{f_{X_{-1}}(X_{i2}, \dots, X_{id})}{f_X(X_i)} \int \frac{f_X(X_{i1}, x_2, \dots, x_d)}{\sigma(X_{i1}, x_2, \dots, x_d)} d(x_2, \dots, x_d) \\
&\quad - \int \frac{f_X(x)}{\sigma(x)} dx \frac{1}{n} \sum_{j=1}^n \left[ \sigma(X_j) \varepsilon_j + m_1(X_{j1}) \right] + o_P(n^{-1/2}),
\end{aligned}$$

which is the desired expansion.  $\square$

**Proof of Theorem 4.1.** Exactly as in the proof of Theorem 2.1 we obtain the expansion

$$\begin{aligned}
\widetilde{F}_\varepsilon(y) - F_\varepsilon(y) &= n^{-1} \sum_{i=1}^n I(\varepsilon_i \leq y) - F_\varepsilon(y) \\
&\quad + f_\varepsilon(y) \int \sigma^{-1}(x) \{y(\widehat{\sigma}(x) - \sigma(x)) + \widetilde{m}(x) - m(x)\} dF_X(x) + o_P(n^{-1/2})
\end{aligned}$$

uniformly with respect to  $y$ , and hence,

$$\begin{aligned}
\widehat{F}_\varepsilon(y) - \widetilde{F}_\varepsilon(y) &= f_\varepsilon(y) \left[ \int \sigma^{-1}(x) \{\widehat{m}(x) - m(x)\} dF_X(x) - (\bar{Y}_n - c) \int \sigma^{-1}(x) dF_X(x) \right. \\
&\quad \left. - \sum_{\ell=1}^d \int \sigma^{-1}(x) \{\widehat{m}_\ell(x_\ell) - m_\ell(x_\ell)\} dF_X(x) \right] + o_P(n^{-1/2}).
\end{aligned}$$

Now Lemma A.2 and Lemma A.5 yield

$$\begin{aligned}
& \widehat{F}_\varepsilon(y) - \widetilde{F}_\varepsilon(y) \\
= & f_\varepsilon(y) \left( \frac{1}{n} \sum_{i=1}^n \varepsilon_i - (\bar{Y}_n - c) \int \sigma^{-1}(x) dF_X(x) \right. \\
& \left. - \sum_{\ell=1}^d \left\{ \frac{1}{n} \sum_{i=1}^n \sigma(X_i) \varepsilon_i \frac{g_\ell(X_{i\ell}) f_{X_{-\ell}}(X_{i,-\ell})}{f_X(X_i)} - \int \frac{f_X(x)}{\sigma(x)} dx \frac{1}{n} \sum_{i=1}^n [\sigma(X_i) \varepsilon_i + m_\ell(X_{i\ell})] \right\} \right) \\
& + o_P(n^{-1/2}) \\
= & f_\varepsilon(y) \left( \frac{1}{n} \sum_{i=1}^n \varepsilon_i \left[ 1 - \sum_{\ell=1}^d \sigma(X_i) \frac{g_\ell(X_{i\ell}) f_{X_{-\ell}}(X_{i,-\ell})}{f_X(X_i)} \right] + (d-1) \int \frac{f_X(x)}{\sigma(x)} dx \frac{1}{n} \sum_{i=1}^n \sigma(X_i) \varepsilon_i \right) \\
& + o_P(n^{-1/2})
\end{aligned}$$

and the assertion follows.  $\square$

**Acknowledgments.** The second author acknowledges support from IAP research network nr. P6/03 of the Belgian government (Belgian Science Policy).

## References

- M. G. Akritas and I. Van Keilegom** (2001). *Nonparametric estimation of the residual distribution*. Scand. J. Statist. 28, 549–567.
- S. X. Chen and I. Van Keilegom** (2006). *A goodness-of-fit test for parametric and semiparametric models in multiresponse regression* (submitted) (paper available at <http://www.stat.ucl.ac.be/ISpub/dp/2006/dp0616.pdf>) (under revision for Bernoulli).
- H. Dette, N. Neumeyer and I. Van Keilegom** (2007). *A new test for the parametric form of the variance function in non-parametric regression*. J. Roy. Statist. Soc. Ser. B 69, 903–917.
- H. Dette and C. von Lieres und Wilkau** (2001). *Testing additivity by kernel based methods – what is a reasonable test?* Bernoulli 7, 669–697.
- J. Fan and I. Gijbels** (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

- P. L. Gozalo and O. B. Linton** (2001). *A nonparametric test of additivity in generalized nonparametric regression with estimated parameters*, J. Econometrics 104, 1–48.
- W. Härdle and A. Tsybakov** (1997). *Local polynomial estimators of the volatility function in nonparametric autoregression*. J. Econometrics 81, 223–242.
- O. B. Linton and J. P. Nielsen** (1995). *A kernel method of estimating structured nonparametric regression based on marginal integration*. Biometrika 82, 93–101.
- E. Masry** (1996). *Multivariate local polynomial regression for time series: uniform strong consistency and rates*. J. Time Series Analysis 17, 571–599.
- U. U. Müller, A. Schick and W. Wefelmeyer** (2007). *Estimating the error distribution function in semiparametric regression*. Statist. Decisions 25, 1–18.
- N. Neumeyer** (2006). Smooth residual bootstrap for empirical processes of nonparametric regression residuals (submitted) (paper available at <http://www.math.uni-hamburg.de/research/ims.html>) (under revision for Scand. J. Statist.).
- N. Neumeyer, H. Dette and E.–R. Nagel** (2005). *Bootstrap tests for the error distribution in linear and nonparametric regression models*. Austr. New Zealand J. Statist. 48, 129–156.
- N. Neumeyer, S. Sperlich** (2006). *Comparison of separable components in different samples*. Scand. J. Statist. 33, 477–501.
- W. K. Newey** (1994). *Kernel estimation of partial means*. Econometric Theory 10, 233–253.
- J. L. Ojeda** (2008). *Hölder continuity properties of the local polynomial estimator* (submitted) (paper available at <http://www.unizar.es/galdeano/preprints/lista.html>).
- J. C. Pardo–Fernández** (2007). *Comparison of error distributions in nonparametric regression*. Statist. Probab. Lett. 77, 350–356.
- J. C. Pardo–Fernández, I. Van Keilegom and W. González–Manteiga** (2007). *Comparison of regression curves based on the estimation of the error distribution*. Statistica Sinica 17, 1115–1137.

- J. M. Rodríguez-Póo, S. Sperlich and P. Vieu** (2005). *An adaptive specification test for semiparametric models* (submitted)  
(paper available at <http://papers.ssrn.com/sol3/>).
- D. Ruppert and M. P. Wand** (1994). *Multivariate locally weighted least squares regression*. *Ann. Statist.* 22, 1346–1370.
- D. Tjøstheim and B. H. Auestad** (1994). *Nonparametric identification of nonlinear time series: projections*. *J. Amer. Statist. Assoc.* 89, 1398–1409.
- A. W. van der Vaart and J. A. Wellner** (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- I. Van Keilegom, W. González–Manteiga and C. Sánchez Sellero** (2008). *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution*. *TEST*, to appear.
- L. Yang, B. U. Park, L. Xue and W. Härdle** (2006). *Estimation and testing for varying coefficients in additive models with marginal integration*. *J. Amer. Statist. Assoc.* 101, 1212–1227.