Estimating the error distribution function in semiparametric additive regression models

Ursula U. Müller, Anton Schick and Wolfgang Wefelmeyer

Abstract. We consider semiparametric additive regression models with a linear parametric part and a nonparametric part, both involving multivariate covariates. For the nonparametric part we assume two models. In the first, the regression function is unspecified and smooth; in the second, the regression function is additive with smooth components. Depending on the model, the regression curve is estimated by suitable least squares methods. The resulting residual-based empirical distribution function is shown to differ from the error-based empirical distribution function by an additive expression, up to a uniformly negligible remainder term. This result implies a functional central limit theorem for the residual-based empirical distribution function. It is used to test for normal errors.

Key words: Partly linear regression model, nonparametric additive regression, uniform Bahadur representation, local polynomial smoother, orthogonal series estimator, Hölder space, test for normal errors, martingale transform test.

1. Introduction and Main Results

This article considers the partly linear regression model

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

where the error ε has mean zero, finite variance and a density f, and is independent of the covariate pair (U, X), with U a p-dimensional random vector and X a q-dimensional random vector. We assume two different models for the nonlinear part ρ of the regression function.

Model 1: The function ρ is smooth.

Model 2: The function ρ is additive, $\rho(x) = \rho_1(x_1) + \cdots + \rho_q(x_q)$, with smooth components $\rho_1, \ldots, \rho_q$.

We are interested in estimating the error distribution function $F(t)$ by a residual-based empirical distribution function $\hat{\mathbb{F}}(t) = (1/n) \sum_{j=1}^n \mathbf{1}[\hat{\varepsilon}_j \le t]$ based on n independent copies $(U_1, X_1, Y_1), \ldots, (U_n, X_n, Y_n)$ of (U, X, Y). The residuals are of the form

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

where $\hat{\vartheta}$ is some $\sqrt{n}$-consistent estimator of ϑ and $\hat{\rho}$ is an appropriate estimator based on the covariates X_j and the "observations" $Y_j - \hat{\vartheta}^\top U_j$. In Model 1 we estimate ρ by local polynomial smoothers, and in Model 2 we estimate $\rho_1, \dots, \rho_q$ by orthogonal series estimators. We show that, in both models, the residual-based empirical distribution function $\hat{\mathbb{F}}(t)$ differs from the error-based empirical distribution function $\mathbb{F}(t) = (1/n)\sum_{j=1}^n \mathbf{1}[\varepsilon_j \le t]$ by an additive expression, up to a uniformly negligible remainder term,

$$\sup_{t\in\mathbb{R}} \left| \hat{\mathbb{F}}(t) - \mathbb{F}(t) - f(t)\frac{1}{n}\sum_{j=1}^n \varepsilon_j \right| = o_p(n^{-1/2}). \tag{1.1}$$

This immediately implies a functional central limit theorem for the residual-based empirical distribution function.

Most of the literature on this problem is concerned with cases in which the regression function is *parametric*, in particular with linear regression, and in which ρ is not there. We refer to Koul (1969, 1970, 2002), Durbin (1973), Loynes (1980), Shorack (1984), and, for increasing dimension of ϑ, to Portnoy (1986) and Mammen (1996). Parametric regression functions are easier to handle since the finite-dimensional regression parameter ϑ can be estimated at the root-n rate. If a nonparametric regression function ρ is involved, different arguments are needed to obtain a stochastic expansion, and hence the root-n rate, and asymptotic normality for the residual-based empirical distribution function.

For heteroscedastic *nonparametric* regression and one-dimensional covariate, Akritas and Van Keilegom (2001) give a functional central limit theorem for a residual-based empirical distribution function. Neumeyer and Van Keilegom (2010) treat multivariate covariates. A related result for the homoscedastic case is given by Cheng (2005) who uses separate parts of the sample for estimating the regression function and the error distribution function, and also estimates the error density. Kiwitt et al. (2008) assume linear constraints on the error distribution and obtain an improved residual-based empirical distribution function. Müller et al. (2009b) consider multivariate covariates and estimate the regression function by local polynomial smoothers. A related result for nonparametric *autoregression* and one-dimensional covariate is in Müller et al. (2009a).

Müller et al. (2007) consider the *partly linear regression model* above, but only for one-dimensional X. They use a local *linear* smoother for the regression function. Here we follow the approach from that paper, but, in order to handle the more complex case when X is a random vector, we use local *polynomial* smoothers as in Müller et al. (2009b). In the one-dimensional case both methods can be used. Which one to choose will depend on the smoothness assumptions about the regression function and the error distribution one is willing to make. For more explanations we refer to Müller et al. (2009b), in particular to their Remark 5. In the following we will refer to the above two papers as MSW1 and MSW2.

We make the following standard assumptions on U and X.

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

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

In Theorem 2.1 we prove expansion (1.1) for Model 1, using a local polynomial smoother whose degree depends on the dimension of the covariate. This generalizes Theorem 1 of MSW2 from the nonparametric regression model to the partly linear regression model; the proof follows in part that of MSW2. Also note that our Theorem 2.1 yields Theorem 1.1 in MSW1 where we study the partly linear regression model with $\rho = \rho_1$. An alternative approach could be based on orthogonal series estimators. We demonstrate this approach in Theorem 4.1 for the additive Model 2, where we use orthogonal series estimators for the components $\rho_1, \ldots, \rho_q$ of ρ; this requires a different proof.

In the case of a one-dimensional covariate X the two models coincide and we have an alternative estimator for Model 1. Theorem 2.1 allows weaker smoothness assumptions on ρ at the expense of stronger moment assumptions on the error variable. In turn, for a twice continuously differentiable ρ, Theorem 4.1 improves on Theorem 2.1 by relaxing the moment assumptions on the error variable and removing the extra conditions on the conditional moments of U given X.

There are two reasons why we chose to estimate ρ in Model 2 by a series estimator. First of all, we wanted to demonstrate an approach that is not only new in this context but also rather convenient: we can work with an orthonormal basis for the space of (square-integrable) *additive* functions to approximate the *additive* function ρ (which is straightforward using least squares estimators for the coefficients). Secondly, we are interested in an accurate estimator of the error distribution function, that is, in a (residual-based) estimator $\hat{\mathbb{F}}$ that has the asymptotic expansion (1.1). The series estimator accomplishes this under weak assumptions.

The components of the regression function in additive regression models can be estimated in several other ways. Stone (1985) uses an additive spline estimator. The backfitting method of Breiman and Friedman (1985), and Buja et al. (1987), estimates the additive components one by one and iterates this procedure. The marginal integration method of Newey (1994), Tjøstheim and Auestad (1994), and Linton and Nielsen (1995) starts with an estimator for a multivariate nonparametric regression function and obtains estimators for the additive components by integrating out all but one of the variables, usually with empirical estimators based on the remaining components of the covariates. Linton (1997) uses marginal integration to provide an initial estimator, and then a single backfitting step. See also Fan et al. (1998), and Mammen et al. (1999). The estimators are compared by Sperlich et al. (1999), Delecroix and Protopopescu (2000), and Dette et al. (2005).

Residual-based empirical distribution functions can be used for a number of different purposes. Using a result of Gill (1989) on compact differentiability of quantile functions, we obtain an approximation for the residual-based empirical quantile function, uniformly over bounded intervals: For $0 < \alpha < \beta < 1$,

$$\sup_{\alpha \le u \le \beta} \Big| \hat{\mathbb{F}}^{-1}(u) - F^{-1}(u) + \frac{1}{n} \sum_{j=1}^{n} \Big(\frac{\mathbf{1}[\varepsilon_j \le F^{-1}(u)] - u}{f(F^{-1}(u))} + \varepsilon_j \Big) \Big| = o_p(n^{-1/2}).$$

Residual-based empirical distribution functions can also be used to test various hypotheses about regression models. Tests for parametric hypotheses about the regression function are considered in nonparametric regression by Stute (1997) and Khmaladze and Koul (2004, 2009). Tests for a parametric regression function and for additivity in heteroscedastic nonparametric regression are studied in Van Keilegom et al. (2008), and in Neumeyer and Van Keilegom (2010), respectively. In Section 3 we use Theorem 2.1 to test for normality of the errors.

It is straightforward to adapt our results to the heteroscedastic model $Y = \vartheta^\top U + \rho(X) + \sigma(U,X)\eta$, with a standardized error variable η and a conditional variance function $\sigma^2(U,X)$, and work with residuals that involve an estimator of the conditional variance, $\hat\varepsilon_j = [Y_j - \hat\vartheta^\top U_j - \hat\rho(X_j)]/\hat\sigma(U,X)$. For the proofs it would be convenient to make the usual assumption that the conditional variance function is bounded and bounded away from zero.

Our paper is organized as follows. In Section 2 we treat Model 1. The key result is Theorem 2.1, which gives the uniform stochastic expansion (1.1) of the residual-based empirical distribution function $\hat{\mathbb{F}}$ when the nonlinear part of the regression function is estimated by a local polynomial smoother. A test for normality of errors is discussed in Section 3. In Section 4 we treat Model 2. As explained there, this covers the purely nonparametric additive regression model. In contrast to Section 2 we now use orthogonal series estimators for the additive components of the regression function, based on the trigonometric basis. The main result is Theorem 4.1. The assertion is the same as in Theorem 2.1, i.e. expansion (1.1) for the corresponding residual-based empirical distribution function. We also address testing for normality of errors, which can be done analogously to Section 3. The performance of the test is investigated with a small simulation study. Theorem 4.1 is based on a technical result of independent interest, Proposition 4.1. The proofs of Theorems 2.1 and 4.1 and of Proposition 4.1 are in Sections 5–7.

2. Estimating the error distribution in Model 1

In this section we treat the model

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

where the error ε has mean zero and finite variance σ^2, and is independent of the covariate pair (U,X), with U a p-dimensional random vector, $\vartheta \in \mathbb{R}^p$ an unknown parameter vector, ρ an unknown smooth function, and X a q-dimensional random vector.

For a non-negative integer m and a $\gamma \in (0,1]$ we introduce the Hölder space $\mathcal{H}(m,\gamma)$ as follows. A function h from $\mathscr{C}$ to $\mathbb{R}$ belongs to $\mathcal{H}(m,\gamma)$ if it has continuous partial derivatives up to order m and the partial derivatives of order m are Hölder with exponent γ.

Let $\hat\vartheta$ denote a $\sqrt{n}$-consistent estimator of ϑ. Such estimators exist, as shown in Schick (1996). We assume that the function ρ belongs to $\mathcal{H}(m,\gamma)$, and estimate it by a local polynomial smoother of degree m; see Stone (1980, 1982), and Ruppert and Wand (1994) for general results on multivariate local polynomial smoothers. Such estimators were used in MSW2 in the case $\vartheta = 0$, i.e., for the nonparametric regression model. Since ϑ is not zero here, we need to work with the difference $Y_j - \hat\vartheta^\top U_j$ instead of the response variable Y_j.

In order to define the local polynomial smoother we introduce some notation. By a *multi-index* we mean a q-dimensional vector $i = (i_1, \dots, i_q)$ whose components are non-negative integers. For a multi-index i let ψ_i denote the function on $\mathbb{R}^q$ defined by

$$\psi_i(x) = \frac{x_1^{i_1}}{i_1!} \cdots \frac{x_q^{i_q}}{i_q!}, \quad x = (x_1, \dots, x_q) \in \mathbb{R}^q.$$

Set $i_\bullet = i_1 + \cdots + i_q$. Let $I(m)$ denote the set of multi-indices i with $i_\bullet \le m$, and $J(m)$ the set of multi-indices i with $i_\bullet = m$. Now fix densities $w_1, \dots, w_q$ and set

$$w(x) = w_1(x_1) \cdots w_q(x_q), \quad x = (x_1, \dots, x_q) \in \mathbb{R}^q.$$

Let c_n be a bandwidth. Then the *local polynomial smoother $\hat\rho$ (of degree m)* is defined as follows. For a fixed x in $\mathscr{C}$, the estimator $\hat\rho(x)$ is the component $\hat\beta_0(x)$ corresponding to the multi-index $0 = (0, \dots, 0)$ of a minimizer

$$\hat\beta(x) = \operatorname*{arg\,min}_{\beta=(\beta_i)_{i\in I(m)}} \sum_{j=1}^n w\Big(\frac{X_j - x}{c_n}\Big)\bigg(Y_j - \hat\vartheta^\top U_j - \sum_{i\in I(m)} \beta_i \psi_i\Big(\frac{X_j - x}{c_n}\Big)\bigg)^2.$$

For $j = 1, \dots, n$ we estimate the error ε_j by the residual

$$\hat\varepsilon_j = Y_j - \hat\vartheta^\top U_j - \hat\rho(X_j).$$

We prove Theorem 2.1 under the assumption that the function ρ is smooth in the sense that it belongs to the Hölder space $\mathcal{H}(m, \gamma)$ with smoothness parameter $s = m + \gamma$ greater than $3q/2$. This means that if the dimension q of the covariate vector X is high then we need more partial derivatives for ρ than in the low-dimensional case. The reason for the smoothness assumption is a technical one. Our proof relies on arguments from empirical process theory: the class $\mathbf{1}(\varepsilon \le t)$, $t \in \mathbb{R}$, is Donsker, but typically not anymore if ε is replaced by a residual $\hat\varepsilon$. Smoothness of ρ is necessary to cope with the random shift $\hat\varepsilon - \varepsilon$ (which is done by verifying equation (5.1) in Section 5). This is more difficult if the dimension of the covariate vector is high. Here we handle the shift by working with polynomial smoothers of degree m. This requires that the aforementioned smoothness condition on ρ is satisfied. Further details are given in Remark 2.2 below. The proof of Theorem 2.1 is in Section 5.

THEOREM 2.1. *Suppose (G) and (H) hold, $\|U\|$ has a moment greater than 2, μ is continuous and τg is bounded, where $\mu(X) = E(U|X)$ and $\tau(X) = E(|U|^2|X)$. Suppose that ρ belongs to $\mathcal{H}(m, \gamma)$ with $s = m + \gamma > 3q/2$. Let the error density f have mean zero, a finite moment of order greater than $4s/(2s - q)$, and be Hölder with exponent greater than $q/(2s - q)$. Let $\hat\vartheta$ be a $\sqrt{n}$-consistent estimator of ϑ. Let the densities $w_1, \dots, w_q$ be $(q + 2)$ times continuously differentiable with compact support $[-1, 1]$. Choose a bandwidth $c_n \sim (n \log n)^{-1/(2s)}$. Then the uniform stochastic expansion*

$$\sup_{t\in\mathbb{R}} \Big|\frac{1}{n}\sum_{j=1}^n \mathbf{1}[\hat\varepsilon_j \le t] - \frac{1}{n}\sum_{j=1}^n \mathbf{1}[\varepsilon_j \le t] - f(t)\frac{1}{n}\sum_{j=1}^n \varepsilon_j\Big| = o_p(n^{-1/2})$$

holds.

REMARK 2.1. We should point out that $4s/(2s-q) < 3$ and $q/(2s-q) < 1/2$ if $s > 3q/2$. Thus the assumptions on the error density f are satisfied if f has mean zero, a finite third moment, and is Hölder with exponent $1/2$. The Hölder condition is met by all densities with finite Fisher information for location.

REMARK 2.2. The use of local polynomial smoothers allows us to construct estimators of ρ that possess a sufficiently fast rate of uniform convergence, are sufficiently smooth, and have a bias that is uniformly of order $o(n^{-1/2})$. To obtain these properties we need sufficient smoothness of the regression function to control the bias, and a sufficiently large bandwidth for good uniform accuracy. The rate of this accuracy is of order $a_n = (\log n/(nc^q))^{-1/2}$ and needs to satisfy $a_n^{1+\xi} = o(n^{-1/2})$ with ξ the Hölder exponent of f. To control the bias we need $c^s = o(n^{-1/2})$. We were able to meet these goals with the smoothness condition $s > 3q/2$ on the regression function ρ and the bandwidth choice in the theorem under a mild moment assumption on f and a fairly small ξ.

Our smoothness requirements are less stringent than those of Neumeyer and Van Keilegom (2010) in the nonparametric heteroscedastic regression model. They require the regression function to have uniformly continuous partial derivatives of order ν with $\nu - 1 > 3q/2$, see their (C4) and the comment after their (C5). In addition, they require the density of X to have continuous partial derivatives of order $2q$, see their (C3), and impose stronger smoothness assumptions on f in their (C5).

In Section 3 we introduce a test for normality of the errors. The test requires a uniform expansion of the empirical distribution function based on *normalized* residuals, which now involve estimators of σ. The expansion is a result of separate interest and derived in the following Remark 2.3, using Theorem 2.1.

REMARK 2.3. Let us now assume that $\varepsilon = \sigma Z$, where Z has mean zero, variance one, and a finite fourth moment. Let f_* and F_* denote the density and distribution function of Z. Then $f(x) = f_*(x/\sigma)/\sigma$ and $F(x) = F_*(x/\sigma)$. We want to estimate F_*. The obvious estimator is the empirical distribution function based on the normalized residuals $\hat{Z}_j = \hat{\varepsilon}_j/\hat{\sigma}$, where $\hat{\sigma}$ is the square root of $(1/n)\sum_{j=1}^n \hat{\varepsilon}_j^2$. The resulting estimator of F_* is

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

with $\hat{\mathbb{F}}(t) = (1/n)\sum_{j=1}^n \mathbf{1}[\hat{\varepsilon}_j \le t]$ as in Theorem 2.1. Under the above moment conditions, and the conditions of Theorem 2.1 (other than those on f), we have

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^n \varepsilon_j^2 + o_p(n^{-1/2}) = \sigma^2\Big(1 + \frac{1}{n}\sum_{j=1}^n (Z_j^2 - 1)\Big) + o_p(n^{-1/2}).$$

From this we conclude

$$S = \frac{\hat{\sigma} - \sigma}{\sigma} = \frac{\hat{\sigma}^2 - \sigma^2}{(\hat{\sigma} + \sigma)\sigma} = \frac{1}{n}\sum_{j=1}^n \frac{Z_j^2 - 1}{2} + o_p(n^{-1/2}).$$

Set $\bar{\varepsilon} = (1/n)\sum_{j=1}^n \varepsilon_j$. We can write

$$\begin{aligned}\hat{\mathbb{F}}_*(t) &= \mathbb{F}(\sigma t) + f_*(t)\frac{1}{n}\sum_{j=1}^n \Big(Z_j + t\frac{Z_j^2-1}{2}\Big) + \big(\hat{\mathbb{F}}(\hat{\sigma}t) - \mathbb{F}(\hat{\sigma}t) - f(\hat{\sigma}t)\bar{\varepsilon}\big) \\ &\quad + \big(\mathbb{F}(\hat{\sigma}t) - F(\hat{\sigma}t) - \mathbb{F}(\sigma t) + F(\sigma t)\big) + \big(F_*((1+S)t) - F_*(t) - St f_*(t)\big) \\ &\quad + \big(f_*((1+S)t) - f_*(t)\big)\frac{1}{n}\sum_{j=1}^n Z_j + t f_*(t)\Big(S - \frac{1}{n}\sum_{j=1}^n \frac{Z_j^2-1}{2}\Big).\end{aligned}$$

Thus we obtain the uniform expansion

$$\sup_{t\in\mathbb{R}}\Big|\frac{1}{n}\sum_{j=1}^n \mathbf{1}[\hat{Z}_j \le t] - \frac{1}{n}\sum_{j=1}^n \mathbf{1}[Z_j \le t] - f_*(t)\frac{1}{n}\sum_{j=1}^n \Big(Z_j + t\frac{Z_j^2-1}{2}\Big)\Big| = o_p(n^{-1/2})$$

under the assumptions of Theorem 2.1, with the assumptions on f replaced by the assumptions that the density f_* has mean zero, variance one, a finite fourth moment, is Hölder with exponent greater than $q/(2s-q)$, and the function $t \mapsto t f_*(t)$ is uniformly continuous. The expansion agrees with that of Theorem 2.1 in Neumeyer and Van Keilegom (2010) obtained for the model $Y = \rho(X) + \sigma(X)Z$ in the case of a constant σ.

3. Testing for normality of the errors

Let us keep the notation from Remark 2.3 and write ϕ and Φ for the standard normal density and distribution function, respectively. The result in Remark 2.3 can be used to test whether the errors have a normal distribution: the expansion for $\hat{\mathbb{F}}_*(t)$ implies a functional central limit theorem, and therefore yields the null limit distribution for statistics based on it. In terms of the density f_*, the null hypothesis is

$$H_0 : f_*(x) = \phi(x) = \exp(-x^2/2)/\sqrt{2\pi}, \quad x \in \mathbb{R}.$$

Possible test statistics are the Kolmogorov–Smirnov statistic

$$T_{KS} = \sup_{t\in\mathbb{R}} n^{1/2}|\hat{\mathbb{F}}_*(t) - \Phi(t)|$$

and the Cramér–von Mises statistic

$$T_{CM} = n\int \big(\hat{\mathbb{F}}_*(t) - \Phi(t)\big)^2\phi(t)\,dt.$$

We prefer to work with a martingale transform test statistic. This has the advantage that the limiting distribution does not depend on features of the unknown distribution F, which otherwise would have to be estimated. Our proposed statistic is

$$T_{MT} = \sup_{t\in\mathbb{R}} n^{1/2}\Big|\hat{\mathbb{F}}_*(t) - \int_{-\infty}^t h^\top(x)\Gamma^{-1}(x)\int_x^\infty h(y)\,d\hat{\mathbb{F}}_*(y)\;\phi(x)\,dx\Big|,$$

where

$$h(x) = (1, x, x^2-1)^\top = \big(1, -\phi'(x)/\phi(x), -(x\phi(x))'/\phi(x)\big)^\top$$

and

$$\Gamma(x) = \int_x^\infty h(z)h^\top(z)\phi(z)\,dz.$$

This is a version of the test statistic studied in Khmaladze and Koul (2009) for nonparametric regression, adapted to testing for normality. One calculates that

$$\Gamma(x) = \begin{bmatrix} 1-\Phi(x) & \phi(x) & x\phi(x) \\ \phi(x) & 1-\Phi(x)+x\phi(x) & (x^2+1)\phi(x) \\ x\phi(x) & (x^2+1)\phi(x) & 2(1-\Phi(x))+(x^3+x)\phi(x) \end{bmatrix}.$$

Since the last two coordinates of h are the score function for location and scale at the parameter value $(0,1)$ in the normal location-scale model, respectively, it follows from Khmaladze and Koul (2009, Remark 4.2) that, under the null hypothesis, T_{MT} converges in distribution to $B_* = \sup_{0\le t\le 1}|B(t)|$, where B is a standard Brownian motion. Our test for normality of errors rejects the null hypothesis if T_{MT} exceeds the $(1-\alpha)$-quantile of the distribution of B_*. The asymptotic size of the test is therefore α.

If we set

$$H(t) = \int_{-\infty}^t h^\top(x)\Gamma^{-1}(x)\phi(x)\,dx,$$

then we can write

$$T_{MT} = \sup_{t\in\mathbb{R}} n^{1/2}\Big|\hat{\mathbb{F}}_*(t) - \frac{1}{n}\sum_{j=1}^n H(t\wedge\hat{Z}_j)h(\hat{Z}_j)\Big|.$$

Remark 3.1. The above test can easily be extended to testing the null hypothesis $H_0 : F_* = F_0$ for some fixed distribution F_0 with finite Fisher information for location and scale. In this case one has to work with $h(x) = \big(1, -f_0'(x)/f_0(x), -(xf_0(x))'/f_0(x)\big)^\top$.

Martingale transform tests for a parametric form of the error distribution in heteroscedastic regression are considered by Mora and Pérez-Alonso (2009). Bootstrap tests for parametric models of the error distribution in nonparametric regression are studied by Neumeyer et al. (2006) and by Heuchenne and Van Keilegom (2010).

4. Estimating the error distribution in Model 2

Now consider the model

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

where ρ is an additive regression function,

$$\rho(x) = \rho_1(x_1) + \cdots + \rho_q(x_q), \quad x = (x_1,\dots,x_q) \in \mathscr{C},$$

with twice continuously differentiable components $\rho_1,\dots,\rho_q$. Again we assume that the error variable ε has mean zero and finite variance, and that it is independent of the covariate pair (U,X), with U a p-dimensional random vector and X a q-dimensional random vector. We estimate $\rho_1,\dots,\rho_q$ by orthogonal series estimators. Properties of such series estimators for semiparametric regression models are studied by Eubank et al. (1990), Andrews (1991), Donald and Newey (1994), Eubank (1999), Li (2000), and Delecroix and Protopopescu (2001).

Write r for the regression function,

$$r(u,x) = \vartheta^\top u + \rho(x), \quad u \in \mathbb{R}^p, x \in \mathscr{C}.$$

For $i = 1, \ldots, q$ and $j = 1, 2, \ldots$, let ψ_{ij} denote the function defined by

$$\psi_{ij}(x) = \sqrt{2}\cos(j\pi x_i), \quad x = (x_1, \ldots, x_q)^\top \in \mathscr{C}.$$

Note that the functions 1 and ψ_{ij} for $i = 1, \ldots, q$ and $j = 1, 2, \ldots$ form an orthonormal basis of the space of additive and square-integrable functions on $\mathscr{C}$. It follows from assumption (G) that

$$g_* \int_{\mathscr{C}} h^2(x)\,dx \le \int h^2(x)\,dG(x) \le g^* \int_{\mathscr{C}} h^2(x)\,dx \tag{4.1}$$

where $g_* > 0$ and $g^* < \infty$ are the infimum and supremum of g on $\mathscr{C}$. This shows that the functions 1 and ψ_{ij} for $i = 1, \ldots, q$ and $j = 1, 2, \ldots$ form a Hilbert space basis of the additive G-square-integrable functions. Let Q denote the joint distribution of U and X, and set

$$\psi_m = (1, \psi_{11}, \ldots, \psi_{1m}, \ldots, \psi_{q1}, \ldots, \psi_{qm})^\top.$$

For large m, the regression function r is well approximated by r_m, its projection in $L_2(Q)$ onto the linear space L_m that consists of the functions

$$(u,x) \to c^\top u + a^\top \psi_m(x)$$

with $c \in \mathbb{R}^p$ and $a \in \mathbb{R}^{1+qm}$. Indeed, we have

$$E[(r(U,X) - r_m(U,X))^2] = O(m^{-3}). \tag{4.2}$$

This follows from the calculations

$$\begin{aligned} E[(r(U,X) - r_m(U,X))^2] &= \inf_{c,a} E[(r(U,X) - c^\top U - a^\top \psi_m(X))^2] \\ &\le E[(\rho(X) - \alpha^\top \psi_m(X))^2] \\ &\le g^* \int_{\mathscr{C}} \big(\rho(x) - \alpha^\top \psi_m(x)\big)^2\,dx \\ &= g^* \sum_{i=1}^{q} \int_0^1 \Big(\rho_i(t) - \bar\rho_i - \sum_{j=1}^{m} \alpha_{ij}\psi_{ij}(t)\Big)^2\,dt, \end{aligned}$$

where $\alpha = (\bar\rho_1 + \cdots + \bar\rho_q, \alpha_{11}, \ldots, \alpha_{1m}, \ldots, \alpha_{q1}, \ldots, \alpha_{qm})^\top$ is the vector of Fourier coefficients

$$\bar\rho_i = \int_0^1 \rho_i(t)\,dt \quad \text{and} \quad \alpha_{ij} = \int_{\mathscr{C}} \rho(x)\psi_{ij}(x)\,dx = \int \rho_i(t)\sqrt{2}\cos(j\pi t)\,dt.$$

Twice integrating by parts yields the formula

$$(j\pi)^2 \int_0^1 h(t)\cos(j\pi t)\,dt = (-1)^j h'(1) - h'(0) - \int_0^1 h''(t)\cos(j\pi t)\,dt, \quad j = 1, 2, \ldots,$$

for every twice continuously differentiable function h defined on $[0,1]$. The desired result is now immediately apparent.

We can write $L_m = \{b^\top h_m : b \in \mathbb{R}^{p+1+qm}\}$, where

$$h_m(u,x) = (u^\top, \psi_m^\top(x))^\top.$$

The projection r_m is of the form $\beta_m^\top h_m$. We estimate β_m by the least squares estimator

$$\hat\beta_m = \arg\min_\beta \sum_{j=1}^n \big(Y_j - \beta^\top h_m(U_j, X_j)\big)^2.$$

The resulting estimator of r is then $\hat r = \hat\beta_m^\top h_m$, yielding the residuals

$$\hat\varepsilon_j = Y_j - \hat\beta_m^\top h_m(U_j, X_j), \quad j = 1, \dots, n.$$

In our asymptotics we assume that $m = m_n$ increases with the sample size. The proof of the next theorem is in Section 6.

THEOREM 4.1. *Suppose that (G) and (H) hold, that ε has mean zero and finite variance, and that its density f is Hölder with exponent ξ greater than $1/3$. Assume ρ is additive with twice continuously differentiable components $\rho_1, \dots, \rho_q$. Let m be of the form $m = m_n \sim n^{1/4}$. Then the uniform stochastic expansion*

$$\sup_{t\in\mathbb{R}} \Big| \frac1n \sum_{j=1}^n \mathbf{1}[\hat\varepsilon_j \le t] - \frac1n \sum_{j=1}^n \mathbf{1}[\varepsilon_j \le t] - f(t) \frac1n \sum_{j=1}^n \varepsilon_j \Big| = o_p(n^{-1/2})$$

holds.

REMARK 4.1. The above approach is easily modified to cover the nonparametric additive regression model, which corresponds to Model 2 with $\vartheta = 0$. One now works with the residuals $\hat\varepsilon_j = Y_j - \hat\beta_m^\top \psi_m(X_j)$ where $\hat\beta_m$ minimizes

$$\sum_{j=1}^n (Y_j - \beta^\top \psi_m(X_j))^2.$$

Since the covariate U is not present, we no longer need to impose condition (H). We obtain the conclusion of Theorem 4.1 under the other assumptions of this theorem.

REMARK 4.2. As in Remark 2.3 let us now assume that $\varepsilon = \sigma Z$, where Z has mean zero, variance one, and a finite fourth moment. Let f_* and F_* denote the density and distribution function of Z. The obvious estimator of F_* is the empirical distribution function

$$\hat{\mathbb{F}}_*(t) = \frac1n \sum_{j=1}^n \mathbf{1}[\hat Z_j \le t], \quad t \in \mathbb{R},$$

based on the normalized residuals $\hat Z_j = \hat\varepsilon_j/\hat\sigma$, where $\hat\sigma$ is the square root of $(1/n)\sum_{j=1}^n \hat\varepsilon_j^2$. Proceeding as in Remark 2.3 we derive the uniform expansion

$$\sup_{t\in\mathbb{R}} \Big| \frac1n \sum_{j=1}^n \mathbf{1}[\hat Z_j \le t] - \frac1n \sum_{j=1}^n \mathbf{1}[Z_j \le t] - f_*(t) \frac1n \sum_{j=1}^n \Big(Z_j + t\frac{Z_j^2 - 1}{2} \Big) \Big| = o_p(n^{-1/2})$$

under the assumptions of Theorem 4.1, with the assumptions on f replaced by the assumptions that the density f_* has mean zero, variance one, a finite fourth moment, and is Hölder with exponent greater than 1/3, and that the function $t \mapsto tf_*(t)$ is uniformly continuous.

REMARK 4.3. The result from the previous remark can be used as in Section 3 to derive a test of normality. In terms of the density f_* the null hypothesis is $H_0 : f_* = \phi$, where ϕ is the standard normal density. The martingale transform test statistic for normality again has the form

$$T_{MT} = \sup_{t\in\mathbb{R}} n^{1/2}\Big|\frac{1}{n}\sum_{j=1}^{n}\mathbf{1}[\hat{Z}_j \le t] - \frac{1}{n}\sum_{j=1}^{n} H(t\wedge\hat{Z}_j)h(\hat{Z}_j)\Big|.$$

where h and H are defined in Section 3. Its limiting distribution under the null hypothesis is that of $\sup_{0\le t\le 1}|B(t)|$, where B is a standard Brownian motion.

We performed a small simulation study for the model $Y = \rho(X_1, X_2) + \varepsilon$, where ρ has two additive terms, $\rho(x_1, x_2) = \rho_1(x_1) + \rho_2(x_2)$, which we estimate with our proposed series estimator. For the simulations we chose $\rho(X_1, X_2) = X_1 + X_2$ (Table 1) and $\rho(X_1, X_2) = e^{-X_1} + \sqrt{1 + X_2^2}$ (Table 2), with independent uniformly distributed covariates X_1 and X_2. In order to test for normality of the errors we consider, firstly, the case when the errors indeed come from a normal distribution. Secondly, we studied four scenarios under the alternative hypothesis, a t-distribution with 5 degress of freedom; a centered chi-square distribution with five degrees of freedom; a double-exponential distribution with scale parameter 1; and a mixture of two normal distributions, where the first has mean -5 and standard deviation 1 and is given weight 1/3 and the second has mean 2.5 and standard deviation 2 and is given weight 2/3. We considered samples of size $n = 50$ and 100, for which the choice of $m = 2, 3$ or 4 basis functions seems to be reasonable, and a 5% significance level α. The quantiles are evaluated numerically, using the formula provided in Shorack and Wellner (2009; page 34, equation (7)). The (upper) 5% quantile is 2.2414, i.e. the null hypothesis is rejected if $T_{MT} > 2.2414$. (The quantiles for $\alpha = 0.01$ and $\alpha = 0.1$ are 2.807 and 1.960, respectively.)

In Tables 1 and 2 we give the proportion of tests (based on 1000 replications) that reject the null hypothesis. The test appears to be slightly conservative: the rejection rate under the null hypothesis, which is given in row (a) in both tables, is always below $\alpha =$

TABLE 1. Tests for normal errors: Simulation results for $\rho(X_1, X_2) = X_1 + X_2$

	$n = 50$			$n = 100$		
F	$m = 2$	$m = 3$	$m = 4$	$m = 2$	$m = 3$	$m = 4$
(a) standard normal	0.027	0.032	0.030	0.030	0.040	0.026
(b) standard $t(5)$	0.214	0.208	0.173	0.359	0.355	0.320
(c) centered chi-square(5)	0.493	0.464	0.426	0.784	0.771	0.739
(d) double-exponential	0.290	0.244	0.198	0.505	0.467	0.425
(e) normal mixture	0.052	0.043	0.020	0.707	0.596	0.525

The figures are the proportions of tests that reject the null hypothesis in 1000 trials.

TABLE 2. Tests for normal errors: Simulation results for $\rho(X_1, X_2) = e^{-X_1} + \sqrt{1 + X_2^2}$.

	$n = 50$			$n = 100$		
F	$m = 2$	$m = 3$	$m = 4$	$m = 2$	$m = 3$	$m = 4$
(a) standard normal	0.034	0.037	0.034	0.044	0.041	0.035
(b) standard $t(5)$	0.205	0.193	0.165	0.365	0.356	0.335
(c) centered chi-square(5)	0.488	0.448	0.408	0.779	0.774	0.730
(d) double-exponential	0.282	0.256	0.213	0.495	0.502	0.484
(e) normal mixture	0.063	0.029	0.028	0.748	0.639	0.538

The figures are the proportions of tests that reject the null hypothesis in 1000 trials.

0.05. Nevertheless, the results indicate that the proposed test is quite powerful. The only exception is the mixture of normals, where the test has no power if the sample size is relatively small ($n = 50$). However, already for $n = 100$ it performs quite well. Also it turns out that in all but one of our examples (Table 2 (d), n=100) the choice $m = 2$ works best.

The proof of Theorem 4.1 relies on the following technical result which is of independent interest. Its proof is given in the Section 7.

PROPOSITION 4.1. *For $n \geq 1$, let $\varepsilon_{n1}, \ldots, \varepsilon_{nn}$ be independent random variables with common distribution function F with mean zero and finite variance, let $Z_{n1}, \ldots, Z_{nn}$ be independent k_n-dimensional random vectors independent of $(\varepsilon_{n1}, \ldots, \varepsilon_{nn})$, and let $\hat{\Delta}_n$ be a k_n-dimensional random vector. Suppose that F has a density f that is Hölder with exponent ξ for some $0 < \xi \leq 1$, that*

$$k_n \to \infty \quad \text{and} \quad n^{-1/2} k_n^{3/2} \log n \to 0, \tag{4.3}$$

$$\|\hat{\Delta}_n\| = O_p(1), \tag{4.4}$$

$$\frac{1}{n} \sum_{j=1}^{n} |\hat{\Delta}_n^\top Z_{nj}|^{1+\xi} = o_p(n^{-1/2}), \tag{4.5}$$

$$\sum_{j=1}^{n} \|Z_{nj}\| = O_p(\sqrt{n} k_n), \tag{4.6}$$

and there is a constant B such that for all finite C,

$$\sup_{|\delta| \leq C} \sum_{j=1}^{n} E[|\hat{\Delta}_n^\top Z_{nj}|] \leq BC\sqrt{nk_n}. \tag{4.7}$$

Then the following uniform stochastic expansion holds,

$$\sup_{t \in \mathbb{R}} \Big| \frac{1}{n} \sum_{j=1}^{n} \big(\mathbf{1}[\varepsilon_{nj} \leq t + \hat{\Delta}_n^\top Z_{nj}] - \mathbf{1}[\varepsilon_{nj} \leq t] - f(t) \hat{\Delta}_n^\top Z_{nj} \big) \Big| = o_p(n^{-1/2}).$$

5. Proof of Theorem 2.1

Order the multi-indices $i \in I(m)$ lexicographically. Let ψ be the vector with components ψ_i, $i \in I(m)$. Note that $\hat{\beta}$ is a solution of the normal equation

$$R(x) = W(x)\hat{\beta}(x),$$

where

$$R(x) = \frac{1}{nc_n^q}\sum_{j=1}^n w\Big(\frac{X_j - x}{c_n}\Big)(Y_j - U_j^\top \hat{\vartheta})\psi\Big(\frac{X_j - x}{c_n}\Big)$$

and

$$W(x) = \frac{1}{nc_n^q}\sum_{j=1}^n w\Big(\frac{X_j - x}{c_n}\Big)\psi\Big(\frac{X_j - x}{c_n}\Big)\psi^\top\Big(\frac{X_j - x}{c_n}\Big).$$

We can write $R(x) = A(x) - \big(B(x) + C(x) + D(x)\mu^\top(x)\big)(\hat{\vartheta} - \vartheta)$, where

$$A(x) = \frac{1}{nc_n^q}\sum_{j=1}^n w\Big(\frac{X_j - x}{c_n}\Big)(Y_j - U_j^\top \vartheta)\psi\Big(\frac{X_j - x}{c_n}\Big),$$

$$B(x) = \frac{1}{nc_n^q}\sum_{j=1}^n w\Big(\frac{X_j - x}{c_n}\Big)\psi\Big(\frac{X_j - x}{c_n}\Big)(U_j - \mu(X_j))^\top,$$

$$C(x) = \frac{1}{nc_n^q}\sum_{j=1}^n w\Big(\frac{X_j - x}{c_n}\Big)\psi\Big(\frac{X_j - x}{c_n}\Big)(\mu(X_j) - \mu(x))^\top,$$

$$D(x) = \frac{1}{nc_n^q}\sum_{j=1}^n w\Big(\frac{X_j - x}{c_n}\Big)\psi\Big(\frac{X_j - x}{c_n}\Big).$$

Note that $D(x)$ is the first column of $W(x)$. We can write this as $D(x) = W(x)e$, where e is the vector $(e_i)_{i\in I(m)}$ with $e_0 = 1$ and $e_i = 0$ for $i \neq 0$. Applying Corollary 1 of MSW2 to the entries of the matrices $B(x)$ and $C(x)$ we obtain

$$\sup_{x\in\mathscr{C}} \|B(x)\| = o_p(1) \quad \text{and} \quad \sup_{x\in\mathscr{C}} \|C(x)\| = o_p(1).$$

It follows from the proof of Lemma 1 in MSW2 that $W(x)$ is invertible for all $x \in \mathscr{C}$ on an event whose probability tends to one. On this event we have $\hat{\rho}(x) = e^\top\hat{\beta}(x) = e^\top W^{-1}(x)R(x)$ and therefore

$$\begin{aligned}&\hat{\vartheta}^\top u + \hat{\rho}(x) - \vartheta^\top u - \rho(x)\\ &= (\hat{\vartheta} - \vartheta)^\top(u - \mu(x)) + e^\top W^{-1}(x)A(x) - \rho(x) - e^\top W^{-1}(x)(B(x) + C(x))(\hat{\vartheta} - \vartheta).\end{aligned}$$

The norm of a function h in the Hölder space $\mathcal{H}(m,\gamma)$ is defined by

$$\|h\|_{m,\gamma} = \max_{i\in I(m)}\sup_{x\in\mathscr{C}}|D^i h(x)| + \max_{i\in J(m)}\sup_{x,y\in\mathscr{C},x\neq y}\frac{|D^i h(y) - D^i h(x)|}{\|x - y\|^\gamma}$$

with

$$D^i h(x) = \frac{\partial^{i_1+\cdots+i_q}}{\partial x_1^{i_1}\cdots\partial x_q^{i_q}} h(x), \quad x = (x_1,\dots,x_q) \in \mathscr{C}.$$

Let $\mathcal{H}_1(m,\gamma)$ denote the unit ball of $\mathcal{H}(m,\gamma)$ for this norm.

Since $Y - \vartheta^\top U = \rho(X) + \varepsilon$, we obtain from the results in MSW2 that there exists a random function $\hat{c}$ based on $Y_1 - \vartheta^\top U_1, X_1, \dots, Y_n - \vartheta^\top U_n, X_n$ such that the following four properties hold:

$$P\big(\hat{c} \in \mathcal{H}_1(q,\alpha)\big) \to 1 \tag{5.1}$$

for some $\alpha > 0$;

$$\int |\hat{c}(x)|^{1+\xi} g(x)\,dx = o_p(n^{-1/2}) \tag{5.2}$$

for $\xi > q/(2s-q)$;

$$\int \hat{c}(x) g(x)\,dx = \frac{1}{n}\sum_{j=1}^n \varepsilon_j + o_p(n^{-1/2});$$

and

$$\sup_{x\in\mathscr{C}} \big|e^\top W^{-1}(x) A(x) - \rho(x) - \hat{c}(x)\big| = o_p(n^{-1/2}).$$

Actually, we can take

$$\hat{c}(x) = e^\top (E[W(x)])^{-1} \frac{1}{n c_n^q} \sum_{j=1}^n w\Big(\frac{X_j - x}{c_n}\Big)\varepsilon_j \psi\Big(\frac{X_j - x}{c_n}\Big).$$

Set $r(u,x) = \vartheta^\top u + \rho(x)$ and $\hat{r}(u,x) = \hat{\vartheta}^\top u + \hat{\rho}(x)$. Let $\mathcal{D}$ be the class of functions

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

with $b \in [-1,1]^p$ and $c \in \mathcal{H}_1(q,\alpha)$. Set

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

Write Q for the joint distribution of (U,X). Then

$$P\big(\hat{a} \in \mathcal{D}\big) \to 1;$$

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

$$\int \hat{a}\,dQ = \frac{1}{n}\sum_{j=1}^n \varepsilon_j + o_p(n^{-1/2}); \tag{5.3}$$

$$\sup_{u\in\mathbb{R}^p, x\in\mathscr{C}} \big|\hat{r}(u,x) - r(u,x) - \hat{a}(u,x)\big| = o_p(n^{-1/2}). \tag{5.4}$$

The assertion follows from Theorem 2.2 in MSW1, the fact that f is bounded, and (5.3). Requirement (2.1) of MSW1 on the bracketing numbers of our class $\mathcal{D}$ is verified as in that paper, but now using (1.5) of MSW2, replacing (3.1) of MSW1.

6. Proof of Theorem 4.1

We use $\|A\|_o$ and $\|A\|_2$ to denote the spectral and Euclidean norms of a matrix A. Recall that these are defined by

$$\|A\|_o = \sup_{\|v\|=1} \|Av\| = \sup_{\|v\|=1,\|w\|=1} w^\top Av \quad \text{and} \quad \|A\|_2^2 = \sum A_{ij}^2 = \operatorname{trace}(A^\top A).$$

It is easy to see that $\|A\|_o^2$ equals the maximal eigenvalue of $A^\top A$ and that $\|A\|_o \le \|A\|_2$. For a nonnegative definite matrix, $\|A\|_o$ equals the largest eigenvalue of A.

Since $\|U\|$ has a finite second moment, $a_n = E[\|U\|^2 \mathbf{1}[\|U\| > n^{1/4}]]$ converges to zero. Now set $\bar{a}_n = \max(n^{-1/4}, a_n^{1/3})$. Then we have

$$P\Big(\max_{1\le j\le n} \|U_j\| > \bar{a}_n\sqrt{n}\Big) \le nP(\|U_1\| > \bar{a}_n\sqrt{n}) \le \frac{E[\|U\|^2\mathbf{1}[\|U\| > \bar{a}_n\sqrt{n}]]}{\bar{a}_n^2} \le a_n^{1/3}.$$

Let us now set $U_{nj} = U_j\mathbf{1}[\|U_j\| \le \bar{a}_n\sqrt{n}]$ and $Y_{nj} = \vartheta^\top U_{nj} + \rho(X_j) + \varepsilon_j$. Since the original data $(Y_1, U_1, X_1), \dots, (Y_n, U_n, X_n)$ and the modified data $(Y_{n1}, U_{n1}, X_1), \dots, (Y_{nn}, U_{nn}, X_n)$ differ only on the event $\{\max_{1\le j\le n}\|U_j\| > \bar{a}_n n^{1/2}\}$, whose probability tends to zero, it suffices to prove the result with the modified data in place of the original data. We still use $\hat{\beta}_m$ to denote the modified least squares estimator which is defined as

$$\hat{\beta}_m = \arg\min_b \sum_{j=1}^n \big(Y_{nj} - b^\top h_m(U_{nj}, X_j)\big)^2$$

and also write $r_m(U_{nj}, X_j) = \beta_m^\top h_m(U_{nj}, X_j)$ for the projection of $r(U_{nj}, X_j)$ onto the linear space $\{b^\top h_m(U_{nj}, X_j) : b \in \mathbb{R}^\nu\}$ with $\nu = p + 1 + qm$.

The modified residuals are $\hat{\varepsilon}_{nj} = Y_{nj} - \hat{\beta}_m^\top h_m(U_{nj}, X_j)$ and can be expressed as

$$\hat{\varepsilon}_{nj} = \varepsilon_j + r(U_{nj}, X_j) - r_m(U_{nj}, X_j) - (\hat{\beta}_m - \beta_m)^\top h_m(U_{nj}, X_j) = \varepsilon_j - \hat{\Delta}_n^\top Z_{n,j},$$

where

$$\hat{\Delta}_n = \begin{pmatrix} 1 \\ \sqrt{n/m}\,(\hat{\beta} - \beta) \end{pmatrix} \quad \text{and} \quad Z_{n,j} = \begin{pmatrix} r_m(U_{nj}, X_j) - r(U_{nj}, X_j) \\ \sqrt{m/n}\, h_m(U_{nj}, X_j) \end{pmatrix}.$$

Because we work with a column space that contains 1, the sum of the residuals is zero. Since $\hat{\Delta}_n^\top Z_{n,j}$ equals $\varepsilon_j - \hat{\varepsilon}_{nj}$, we have

$$\frac{1}{n}\sum_{j=1}^n f(t)\hat{\Delta}_n^\top Z_{n,j} = f(t)\frac{1}{n}\sum_{j=1}^n \varepsilon_j.$$

Thus Theorem 4.1 follows from Proposition 4.1 if (4.3)–(4.6) hold for the present choices of $\hat{\Delta}_n$ and $Z_{n,j}$, and with $\xi > 1/3$. Since $k_n = 1 + \nu = 2 + p + qm$ and $m \sim n^{1/4}$, we obtain (4.3). Note also that (4.4) is equivalent to

$$\hat{\beta}_m - \beta_m = O_p(\sqrt{m/n}). \tag{6.1}$$

The argument used to derive (4.2) also yields

$$E[(r_m(U_{n1}, X_1) - r(U_{n1}, X_1))^2] = O(m^{-3}). \tag{6.2}$$

This implies

$$\frac{1}{n}\sum_{j=1}^{n}\big(r_m(U_{nj},X_j)-r(U_{nj},X_j)\big)^2 = O_p(m^{-3}) = O_p(n^{-3/4}) \tag{6.3}$$

and also (4.6) in view of $E[\|h_m(U_{n1},X_1)\|^2] \le E[\|U_1\|^2]+1+2qm$. It follows from the moment inequality that

$$\frac{1}{n}\sum_{j=1}^{n}|\hat\Delta_n^\top Z_{n,j}|^{1+\xi} \le \Big(\frac{1}{n}\sum_{j=1}^{n}(\hat\Delta_n^\top Z_{n,j})^2\Big)^{(1+\xi)/2}.$$

As ξ is assumed to be greater than $1/3$, condition (4.5) follows from (6.3) and

$$\frac{1}{n}\sum_{j=1}^{n}\big((\hat\beta_m-\beta_m)^\top h_m(U_{nj},X_j)\big)^2 = O_p(m/n) = O_p(n^{-3/4}). \tag{6.4}$$

To prove (4.7), we note that

$$\big(E[|v^\top h_m(U_{n1},X_1)|]\big)^2 \le E[(v^\top h_m(U_{n1},X_1))^2] = v^\top M_n v$$

with

$$M_n = E[h_m(U_{n1},X_1)h_m^\top(U_{n1},X_1)].$$

We shall show that

$$\lambda = \liminf_{m\to\infty}\inf_{\|v\|=1} v^\top M_m v > 0 \quad\text{and}\quad \Lambda = \limsup_{m\to\infty}\sup_{\|v\|=1} v^\top M_m v < \infty. \tag{6.5}$$

Thus (4.7) follows from (6.2) and the second part of (6.5).

To prove (6.4), we write its left-hand side as $(\hat\beta_m-\beta_m)^\top \hat M_m(\hat\beta_m-\beta_m)$, where

$$\hat M_m = \frac{1}{n}\sum_{j=1}^{n} h_m(U_{nj},X_j)h_m(U_{nj},X_j)^\top.$$

We shall show that

$$\|\hat M_m - M_m\|_o = o_p(1). \tag{6.6}$$

Let $\hat\lambda_m$ and $\hat\Lambda_m$ denote the smallest and largest eigenvalue of $\hat M_m$, so that

$$\hat\lambda_m = \inf_{\|v\|=1} v^\top \hat M_m v \quad\text{and}\quad \hat\Lambda = \sup_{\|v\|=1} v^\top \hat M_m v.$$

It follows from (6.5) and (6.6) that

$$P(\lambda/2 < \hat\lambda_m \le \hat\Lambda_m < 2\Lambda) \to 1. \tag{6.7}$$

Since the left-hand side of (6.4) is bounded by $\hat\Lambda\|\hat\beta_m-\beta_m\|^2$, (6.4) is a consequence of (6.7) and (6.1).

To prove (6.1), we note that the least squares estimator $\hat\beta_m$ satisfies the normal equations $\hat M_m\hat\beta_m = \hat V_m$, where

$$\hat V_m = \frac{1}{n}\sum_{j=1}^n Y_j h_m(U_{nj},X_j) = \frac{1}{n}\sum_{j=1}^n \big(\varepsilon_j + r(U_{nj},X_j) - r_m(U_{nj},X_j)\big)h_m(U_{nj},X_j) + \hat M_m\beta_m.$$

In view of (6.7), the desired (6.1) follows if we show that

$$T_1 = \frac{1}{n}\sum_{j=1}^n \varepsilon_j h_m(U_{nj},X_j) = O_p\Big(\sqrt{\frac{m}{n}}\Big)$$

and

$$T_2 = \frac{1}{n}\sum_{j=1}^n \big(r(U_{nj},X_j) - r_m(U_{nj},X_j)\big)h_m(U_{nj},X_j) = O_p(m^{-3/2}) = O_p\Big(\sqrt{\frac{m}{n}}\Big).$$

Since the summands in T_1 and T_2 are independent and centered, we have

$$nE[\|T_1\|^2] = E[\varepsilon_1^2\|h_m(U_{n1},X_1)\|^2] = E[\varepsilon^2]\big(E[\|U\|^2] + 1 + 2qm\big) = O(m)$$

and, with the help of (6.2),

$$\begin{aligned} nE[\|T_2\|^2] &= E[|r(U_{n1},X_1) - r_m(U_{n1},X_1)|^2\|h_m(U_{n1},X_1)\|^2] \\ &\le (\bar a_n^2 n + 1 + 2qm)E[|r(U_{n1},X_1) - r_m(U_{n1},X_1)|^2] = o(n)m^{-3}. \end{aligned}$$

We are left to verify (6.5) and (6.6). Equation (6.6) follows from the fact that

$$\begin{aligned} nE[\|\hat M_m - M_m\|_2^2] &= \sum_{i=1}^\nu\sum_{j=1}^\nu n\mathrm{Var}\Big(\frac{1}{n}\sum_{l=1}^n h_{mi}(U_{nl},X_l)h_{mj}(U_{nl},X_l)\Big) \\ &\le \sum_{i=1}^\nu\sum_{j=1}^\nu E[h_{mi}^2(U_{n1},X_1)h_{mj}^2(U_{n1},X_1)] = n^{-1}E[\|h_m(U,X)\|^4] \\ &\le 2E[\|U\|^4\mathbf{1}[\|U\| \le \bar a_n\sqrt{n}]] + 2(1+2qm)^2 = o(n). \end{aligned}$$

Next we show (6.5). Note that the matrix $\Psi_m = E[\psi_m(X)\psi_m^\top(X)]$ is invertible. Indeed, its eigenvalues are between g_* and g^*, see (4.1). Set $U_{n1}^* = U_{n1} - C_n^\top\psi_m(X_1)$ where $C_n = \Psi_m^{-1}E[\psi_m(X_1)U_{n1}^\top]$. Then $E[U_{n1}^*\psi_m^\top(X_1)] = 0$. For a unit vector $v = (v_1^\top, v_2^\top)^\top$, with $v_1 \in \mathbb{R}^p$ and $v_2 \in \mathbb{R}^{1+qm}$, we thus find

$$\begin{aligned} v^\top M_m v &= E[(v^\top h_m(U_{n1},X_1))^2] \\ &= E[(v_1^\top U_{n1}^* + (v_2 + C_n v_1)^\top\psi_m(X_1))^2] \\ &= E[(v_1^\top U_{n1}^*)^2] + E[((v_2 + C_n v_1)^\top\psi_m(X_1))^2]. \end{aligned}$$

It is now easy to see that

$$E[v_1(U_{n1} - E(U_{n1}|X_1))^2] + g_*\|v_2 + C_n v_1\|^2 \le v^\top M_m v \le E[(v_1^\top U_{n1})^2] + g^*\|v_2 + C_n v_1\|^2.$$

With the help of the Cauchy–Schwarz inequality we derive the bound

$$\|C_n\|_o \le B = (E[\|U\|^2]g^*)^{1/2}/g_*.$$

This shows that $\Lambda \le E[\|U\|^2] + g^*(1+B)^2$. Since $E[\|U_{n1} - U_1\|^2] \to 0$, it follows from (H) that, for large m, $v^\top M_m v \ge \eta\|v_1\|^2 + g_*(\|v_2\| - \|C_n v_1\|)^2$, where 2η is the smallest eigenvalue of $E[(U-\mu(X))(U-\mu(X))^\top]$. Now if $\|v_1\| > 1/(2+B)$, we have $v^\top M_m v \ge \eta/(2+B)^2$, while for $\|v_1\| \le 1/(2+B)$, we have $v^\top M v \ge g_*(\sqrt{1-1/(2+B)^2} - B/(2+B))^2 \ge g_*/(2+B)^2$. This shows that $\lambda > \min\{\eta, g_*\}/(2+B)^2$.

7. Proof of Proposition 4.1

To simplify notation we abbreviate ε_{nj} by ε_j, Z_{nj} by Z_j, and set $S_n = \sum_{j=1}^n \|Z_j\|$. Let

$$H(s,t,\Delta) = \frac{1}{n}\sum_{j=1}^n \big(\mathbf{1}[\varepsilon_j \le t + \Delta^\top Z_j + s\|Z_j\|] - F(t + \Delta^\top Z_j + s\|Z_j\|)\big)$$

for $s, t \in \mathbb{R}$ and $\Delta \in \mathbb{R}^{k_n}$. It follows from the properties of f and (4.5) that

$$\sup_{t\in\mathbb{R}} \frac{1}{n}\sum_{j=1}^n \big|F(t + \hat\Delta_n^\top Z_j) - F(t) - f(t)\hat\Delta_n^\top Z_j\big| = O_p\Big(\frac{1}{n}\sum_{j=1}^n |\hat\Delta_n^\top Z_j|^{1+\xi}\Big) = o_p(n^{-1/2}).$$

In view of this and (4.4), it suffices to show that

$$\sup_{t\in\mathbb{R},\ \|\Delta\|\le C} \big|H(0,t,\Delta) - H(0,t,0)\big| = o_p(n^{-1/2})$$

for each finite positive constant C. Fix such a C. It follows from (4.6) and $k_n = o(n^{1/2})$, that the probability of the event $A_n = \{\max_{1\le j\le n}\|Z_j\| \le n\}$ tends to one. Since F has a finite second moment, the probability of the event $B_n = \{\max_{1\le j\le n}|\varepsilon_j| \le \sqrt{n}\}$ tends to one and $t^2(F(-t) + (1-F(t)) \to 0$ as $t \to \infty$. On the intersection $A_n \cap B_n$ of these events we have

$$\sup_{|t|>Cn+\sqrt{n},\ \|\Delta\|\le C} \big|H(0,t,\Delta) - H(0,t,0)\big| \le F(-n^{1/2}) + (1 - F(n^{1/2})) = o(n^{-1}).$$

Thus we are left to show that $R_n = o_p(n^{-1/2})$, where

$$R_n = \sup_{|t|\le Cn+\sqrt{n},\ \|\Delta\|\le C} \big|H(0,t,\Delta) - H(0,t,0)\big|.$$

Now let δ and η be small positive numbers. Let $t_1, \dots, t_M$ be real numbers in $[-Cn - \sqrt{n}, Cn + \sqrt{n}]$ such that the intervals $[t_i - \delta, t_i + \delta]$ cover $[-Cn - \sqrt{n}, Cn + \sqrt{n}]$, and let $\Delta_1, \dots, \Delta_N$ be vectors in $\{x \in \mathbb{R}^{k_n} : \|x\| \le C\}$ such that the balls $\{\Delta : \|\Delta - \Delta_i\| \le \eta\}$ cover the ball $\{x \in \mathbb{R}^{k_n} : \|x\| \le C\}$. We can choose these points such that $M \le 1 + (Cn + \sqrt{n})/\delta$ and $N \le (1 + C\sqrt{k_n}/\eta)^{k_n}$.

Then we have the bound

$$R_n \le \max_{i,l}\Big(\big|H(0,t_i,\Delta_l) - H(0,t_i,0)\big| + \sup_{|t-t_i|\le\delta,\ \|\Delta-\Delta_l\|\le\eta} \big|H(0,t,\Delta) - H(0,t_i,\Delta_l)\big|\Big).$$

Using monotonicity of F and of the maps $x \mapsto \mathbf{1}[\varepsilon_j \le x]$, the supremum term can be further bounded by

$$H(\eta, t_i + \delta, \Delta_l) - H(-\eta, t_i - \delta, \Delta_l) + \frac{2}{n}\sum_{j=1}^{n} \big|F(t_i + \Delta_l^\top Z_j + T_j) - F(t_i + \Delta_l^\top Z_j - T_j)\big|,$$

where $T_j = \delta + \eta\|Z_j\|$. Since F is Lipschitz with constant $L = \sup_{t\in\mathbb{R}} f(t)$, we see that

$$R_n \le R_{n,1} + R_{n,2} + 4L(\delta + \eta S_n/n)$$

with

$$R_{n,1} = \max_{i,l} \big|H(0, t_i, \Delta_l) - H(0, t_i, 0)\big|$$

and

$$R_{n,2} = \max_{i,l} \big|H(\eta, t_i + \delta, \Delta_l) - H(-\eta, t_i - \delta, \Delta_l)\big|.$$

Thus for positive K and y we have the inequality

$$\begin{aligned} P\big(\sqrt{n}R_n > 3y\big) &\le P\big(S_n > K\sqrt{n}k_n\big) + P\big(\sqrt{n}R_{n,1} > y\big) \\ &\quad + P\big(\sqrt{n}R_{n,2} > y, S_n \le K\sqrt{n}k_n\big) + P\big(4L(\sqrt{n}\delta + \eta K k_n) > y\big). \end{aligned}$$

We use the Bernstein inequality of Hoeffding (1963): If $\xi_1, \dots, \xi_n$ are independent random variables that have zero means and are bounded by a constant M, then

$$P\Big(\Big|\sum_{j=1}^{n} \xi_j\Big| \ge x\Big) \le 2\exp\Big(-\frac{x^2}{2\sum_{j=1}^{n} E[\xi_j^2] + (2/3)Mx}\Big), \quad x > 0.$$

The summands $\xi_j = \mathbf{1}[\varepsilon_j \le t_i + \Delta_l^\top Z_j] - \mathbf{1}[\varepsilon_j \le t_i] - F(t_i + \Delta_l^\top Z_j) + F(t_i)$ in the average $H(0, t_i, \Delta_l) - H(0, t_i, 0)$ are independent, centered, bounded by 2, and satisfy

$$\sum_{j=1}^{n} E[\xi_j^2] \le \sum_{j=1}^{n} E[|F(t_i + \Delta_l^\top Z_j) - F(t_i)|] \le \sum_{j=1}^{n} LE[|\Delta_l^\top Z_j|] \le LBC\sqrt{nk_n},$$

so that

$$P\big(\sqrt{n}|H(0, t_i, \Delta_l) - H(0, t_i, 0)| > y\big) \le 2\exp\Big(-\frac{ny^2}{2LBC\sqrt{nk_n} + 2y\sqrt{n}}\Big), \quad y > 0.$$

Now write P_Z and E_Z for the conditional probability measure and expectation given $Z = (Z_1, \dots, Z_n)$. Conditionally given Z, the summands

$$\begin{aligned} \xi_j = \mathbf{1}[\varepsilon_j \le t_i + \delta + \Delta_l^\top Z_j + \eta\|Z_j\|] - \mathbf{1}[\varepsilon_j \le t_i - \delta + \Delta_l^\top Z_j - \eta\|Z_j\|] \\ - F(t_i + \delta + \Delta_l^\top Z_j + \eta\|Z_j\|) + F(t_i - \delta + \Delta_l^\top Z_j - \eta\|Z_j\|) \end{aligned}$$

in the average $H(\eta, t_i + \delta, \Delta_l) - H(-\eta, t_i + \delta, \Delta_l)$ are independent, centered, bounded by 2 and satisfy

$$\begin{aligned} \sum_{j=1}^{n} E_Z(\xi_j^2) &\le \sum_{j=1}^{n} \big(F(t_i + \delta + \Delta_l^\top Z_j + \eta\|Z_j\|) - F(t_i - \delta + \Delta_l^\top Z_j - \eta\|Z_j\|)\big) \\ &\le 2L(n\delta + \eta S_n). \end{aligned}$$

This yields

$$P_Z\big(\sqrt{n}|H(\eta,t_i+\delta,\Delta_l)-H(-\eta,t_i-\delta,\Delta_l)|>y\big)\le 2\exp\Big(-\frac{ny^2}{4L(n\delta+\eta S_n)+2y\sqrt{n}}\Big)$$

for $y>0$. Thus we have

$$\begin{aligned}P\big(\sqrt{n}R_{n,1}>y\big)&\le\sum_{i,l}P\big(\sqrt{n}|H(0,t_i,\Delta_l)-H(0,t_i,0)|>y\big)\\&\le 2MN\exp\Big(-\frac{ny^2}{2LBC\sqrt{nk_n}+2y\sqrt{n}}\Big),\quad y>0,\end{aligned}$$

and

$$P\big(\sqrt{n}R_{n,2}>y,S_n\le K\sqrt{n}k_n\big)\le 2MN\exp\Big(-\frac{ny^2}{4L(n\delta+\eta K\sqrt{n}k_n)+2y\sqrt{n}}\Big),\quad y>0.$$

Taking $\delta=\eta=1/n$, we can choose $M=O(n^2)$ and $\log N=O(k_n\log n)=o(\sqrt{n}/k_n)$ and obtain

$$\limsup_{n\to\infty}P\big(\sqrt{n}R_n>3y\big)\le\limsup_{n\to\infty}P\big(S_n>K\sqrt{n}k_n\big)\le\frac{1}{K}$$

for all finite K. This is the desired result in view of (4.6).

Acknowledgements

Ursula U. Müller was supported by NSF Grant DMS 0907014. Anton Schick was supported by NSF Grant DMS 0906551. The authors thank two referees for suggestions which improved the manuscript. For example, they led us to add tests for normality and simulations.

References

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

[2] Andrews, D.W.K. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* **59**, 307–345.

[3] Breiman, L. and Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation. With discussion and with a reply by the authors. *J. Amer. Statist. Assoc.* **80**, 580–619.

[4] Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17**, 453–555.

[5] Cheng, F. (2005). Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *J. Statist. Plann. Inference* **128**, 327–349.

[6] Delecroix, M. and Protopopescu, C. (2000). Are regression series estimators efficient in practice? A computational comparison study. *Comput. Statist.* **15**, 511–529.

[7] Delecroix, M. and Protopopescu, C. (2001). Regression series estimators: the MISE approach. *J. Nonparametr. Stat.* **13**, 453–483.

[8] Dette, H., Von Lieres und Wilkau, C. and Sperlich, S. (2005). A comparison of different nonparametric methods for inference on additive models. *J. Nonparametr. Stat.* **17**, 57–81.

[9] Donald, S.G. and Newey, W.K. (1994) Series estimation of semilinear models. *J. Multivariate Anal.* **50**, 30–40.

[10] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279–290.
[11] Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing.* Second edition. Statistics: Textbooks and Monographs 157, Dekker, New York.
[12] Eubank, R. L., Hart, J.D. and Speckman, P. (1990). Trigonometric series regression estimators with an application to partially linear models. *J. Multivariate Anal.* **32**, 70–83.
[13] Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *Ann. Statist.* **26**, 943–971.
[14] Gill, R.D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. With a discussion by J.A. Wellner and J. Præstgaard and a reply by the author. *Scand. J. Statist.* **16**, 97–128.
[15] Heuchenne, C. and Van Keilegom, I. (2010). Goodness-of-fit tests for the error distribution in nonparametric regression. *Comput. Statist. Data Anal.* **54**, 1942–1951.
[16] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30.
[17] Khmaladze, E.V. and Koul, H.L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32**, 995–1034.
[18] Khmaladze, E.V. and Koul, H.L. (2009). Goodness-of-fit problem for errors in non-parametric regression: distribution free approach. *Ann. Statist.* **37**, 3165–3185.
[19] Kiwitt, S., Nagel, E.-R. and Neumeyer, N. (2008). Empirical likelihood estimators for the error distribution in nonparametric regression models. *Math. Methods Statist.* **17**, 241–260.
[20] Koul, H.L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40**, 1950–1979.
[21] Koul, H.L. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.* **41**, 1768–1773.
[22] Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* Lecture Notes in Statistics 166. Springer, New York.
[23] Li, Q. (2000). Efficient estimation of additive partially linear models. *Internat. Econom. Rev.* **41**, 1073–1092.
[24] Linton, O.B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika* **84**, 469–473.
[25] Linton, O. and Nielsen, J.P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93–100.
[26] Loynes, R.M. (1980). The empirical distribution function of residuals from generalised regression. *Ann. Statist.* **8**, 285–299.
[27] Mammen, E. (1996). Empirical process of residuals for high-dimensional linear models. *Ann. Statist.* **24**, 307–335.
[28] Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27**, 1443–1490.
[29] Mora, J. and Pérez-Alonso, A. (2009). Specification tests for the distribution of errors in nonparametric regression: a martingale approach. *J. Nonparametr. Stat.* **21**, 441–452.
[30] Müller, U.U., Schick, A. and Wefelmeyer, W. (2007). Estimating the error distribution in semiparametric regression. *Statist. Decisions* **25**, 1–18.
[31] Müller, U.U., Schick, A. and Wefelmeyer, W. (2009a). Estimating the innovation distribution in nonparametric autoregression. *Probab. Theory Related Fields* **14**, 53–77.
[32] Müller, U.U., Schick, A. and Wefelmeyer, W. (2009b). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statist. Probab. Lett.* **79**, 957–964.
[33] Neumeyer, N., Dette, H. and Nagel, E.-R. (2006). Bootstrap tests for the error distribution in linear and nonparametric regression models. *Aust. N. Z. J. Stat.* **48**, 129–156.
[34] Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.* **101**, 1067–1078.

[35] Newey, W.K. (1994). Series estimation of regression functionals. *Econometric Theory* **10**, 1–28.
[36] Portnoy, S. (1986). Asymptotic behavior of the empiric distribution of M-estimated residuals from a regression model with many parameters. *Ann. Statist.* **14**, 1152–1170.
[37] Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22**, 1346–1370.
[38] Schick, A. (1996). Root-n consistent estimation in partly linear regression models. *Statist. Probab. Lett.* **28**, 353–358.
[39] Shorack, G.R. and Wellner J. (2009). *Empirical Processes with Applications to Statistics.* Second edition. SIAM.
[40] Shorack, G.R. (1984). Empirical and rank processes of observations and residuals. *Canad. J. Statist.* **12**, 319–332.
[41] Sperlich, S., Linton, O.B. and Härdle, W. (1999). Integration and backfitting methods in additive models — finite sample properties and comparison. *TEST* **8**, 419–458.
[42] Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348–1360.
[43] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric estimators. *Ann. Statist.* **10**, 1040–1053.
[44] Stone, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 689-705.
[45] Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613–641.
[46] Tjøstheim, D. and Auestad, B. H. (1994). Nonparametric identification of nonlinear time series: projections. *J. Amer. Statist. Assoc.* **89**, 1398–1409.
[47] Van Keilegom, I, González Manteiga, W. and Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST* **17**, 401–415.

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, Mathematical Institute, University of Cologne, Weyertal 86–90, 50931 Cologne, Germany