Estimating the error distribution in a single-index model

Hira L. Koul
Michigan State University

Ursula U. Müller
Texas A&M University

Anton Schick
Binghamton University

Abstract. This paper addresses the problem of estimating the error distribution in single-index regression models. We estimate the error distribution function with a weighted nonparametric residual empirical distribution function. Our main result is a first order uniform stochastic expansion of the estimator. This expansion makes it possible to derive asymptotically distribution free goodness-of-fit tests about the error distribution. Our approach is to regard the single-index model as a nonparametric regression model, but with *estimated* covariates (the estimated indices). However, the usual assumption in classical nonparametric regression, that the covariate distribution is quasi-uniform (bounded and bounded away from zero on its compact support), is not reasonable here. We handle this by introducing weights which restrict the estimation of the link function to intervals.

1. Introduction

We consider the single-index regression model in which the response variable Y is linked to a p-dimensional covariate vector X via the formula

$$Y = \varrho(\theta_0^\top X) + \varepsilon, \tag{1.1}$$

where ϱ is a smooth function, θ_0 is a p-dimensional unit vector, and the error variable ε is independent of the covariate X, has mean zero and a finite variance. In order to guarantee identifiability, we require that the matrix $E[XX^\top]$ is positive definite and that θ_0 belongs to Θ, the set of all p-dimensional unit vectors whose first coordinate is positive, see e.g., Cui, Härdle and Zhu (2011). Furthermore, we assume that ε has a density f and that $\theta^\top X$ has a density g_θ for each θ in Θ.

The single-index regression model was introduced to overcome the curse of dimensionality. Numerous applications and theoretical results can be found in Stoker (1986), Li (1991), Ichimura (1993), Xia and Li (1999), Xia, Tong and Li (2002), Xia, Tong, Li and Zhu (2002), Xia and Härdle (2006), Xia (2008), and references therein. The primary focus of these and related papers has been the estimation of the parameter θ_0 and of the link function ϱ. Stute and Zhu (2005)

Key words and phrases. Weighted residual empirical distribution function; local quadratic smoother.

provide asymptotically distribution free maximin tests for fitting a single-index model to the regression function against a large class of local alternatives.

Here we are interested in the estimation of the error distribution function F based on independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of (X, Y). Our goal is to derive a first order uniform stochastic expansion for a suitably weighted residual empirical distribution function. Such a uniform expansion has been obtained in linear, partially linear and nonparametric regression models by Koul (1969, 1970, 2002), Akritas and Van Keilegom (2001), Müller, Schick and Wefelmeyer (2007, 2009a), and Neumeyer and Van Keilegom (2010). In the context of time series, expansions of this type have been obtained in Boldin (1982, 1990, 1998), Koul (1991, 2002), Müller, Schick and Wefelmeyer (2009b), and Neumeyer and Selk (2013). The existing literature does not cover the case of interest here.

If we denote the single-index $\theta_0^\top X$ by S, then we can write the regression model as a nonparametric regression model

$$Y = \varrho(S) + \varepsilon.$$

A common assumption for estimating ϱ in such nonparametric regression models is that the single covariate S has a density that is bounded and bounded away from zero on its compact support. We call distributions of this type *quasi-uniform*. Thus, if θ_0 were known and if S were quasi-uniform, we could estimate ϱ by classical nonparametric curve estimators, and the results of Müller, Schick and Wefelmeyer ("MSW" 2007) would yield the desired expansion of the corresponding residual empirical process. However, the assumption that S is quasi-uniform is not reasonable in our case as the following two examples demonstrate.

EXAMPLE 1. Suppose that X is uniformly distributed on the unit disk $D = \{x \in \mathbb{R}^2 : \|x\| \le 1\}$. In this case $\theta^\top X$ has density

$$g_\theta(s) = g(s) = \frac{2}{\pi}\sqrt{1-s^2}\,\mathbf{1}[|s|<1]$$

for all θ in Θ, and this density is not bounded away from zero on $[-1, 1]$.

EXAMPLE 2. Suppose that X is uniformly distributed on the unit square $[0, 1] \times [0, 1]$. Let $\theta = (a, b)^\top$ with $0 < a \le 1/\sqrt{2}$ and $b = \sqrt{1-a^2}$. Then the density of $\theta^\top X$ is given by

$$g_\theta(s) = \frac{1}{ab}\Big[\mathbf{1}[0 \le s \le b]\min(s, a) + \mathbf{1}[b < s < a+b](a+b-s)\Big].$$

This density is piecewise linear, and its support depends on a.

Let $\hat\theta$ be an estimator of θ_0 and set

$$\hat S_j = \hat\theta^\top X_j \quad \text{and} \quad \hat\delta_j = \mathbf{1}[\hat S_j \in \hat I], \quad j = 1, \dots, n,$$

where $\hat I$ is the random interval $[\hat l, \hat u]$ whose endpoints are functions of the estimated indices $\hat S_1, \dots, \hat S_n$ and the estimator $\hat\theta$, say

$$\hat l = \phi_{n,l}(\hat S_1, \dots, \hat S_n, \hat\theta) \quad \text{and} \quad \hat u = \phi_{n,u}(\hat S_1, \dots, \hat S_n, \hat\theta). \tag{1.2}$$

Choices of such random intervals are discussed in Remark 1.1 below.

We estimate the link function ϱ by a local quadratic smoother $\hat\varrho$ treating $\hat S_j$ as the regressor. Our estimator $\hat{\mathbb{F}}_n$ of the distribution function F is based on the residuals $Y_j - \hat{\varrho}(\hat{S}_j)$ for which $\hat{\delta}_j = 1$, i.e.,

$$\hat{\mathbb{F}}_n(t) = \frac{1}{N_n} \sum_{j=1}^{n} \hat{\delta}_j \mathbf{1}[Y_j - \hat{\varrho}(\hat{S}_j) \le t], \quad t \in \mathbb{R}, \tag{1.3}$$

with $N_n = \sum_{j=1}^{n} \hat{\delta}_j$.

Remark 1.1. Let us briefly comment on choices of $\hat{I}$. The goal is to have g_{θ_0} bounded away from zero on $\hat{I}$ with high probability. If g_θ were known, we could choose intervals $I(\theta)$ on which g_θ is bounded away from zero, and then take $\hat{I} = I(\hat{\theta})$. If g_θ is known up to parameters, say g_θ is a normal density with mean $\theta^\top \mu$ and variance $\theta^\top \Sigma \theta$ for some unknown vector μ and some unknown dispersion matrix Σ, then we could take $\hat{I} = [\hat{\nu} - c\hat{\sigma}, \hat{\nu} + c\hat{\sigma}]$, where $\hat{\nu}$ is the sample mean and $\hat{\sigma}$ the sample standard deviation of the estimated indices $\hat{S}_1, \dots, \hat{S}_n$. Another choice for $\hat{I}$ is $[\hat{q}_{\alpha_1}, \hat{q}_{\alpha_2}]$ for $0 < \alpha_1 < \alpha_2 < 1$, where $\hat{q}_\alpha$ denotes the α-th sample quantile of the estimated indices. Using only values $\hat{S}_j$ that are densely distributed in an interval $\hat{I}$ around the mean or median, e.g., between the upper and lower five percent quantiles, seems to be a natural choice: it ensures that the local smoother $\hat{\varrho}$ has enough 'observations' available to estimate the link function reasonably well.

The remainder of the paper is organized as follows. In Section 2 we describe our main result, a first order uniform stochastic expansion of $\hat{\mathbb{F}}_n$, and discuss in detail the assumptions used. An application of the main result is described in Section 3 by constructing asymptotically distribution free tests for fitting an error distribution in model (1.1). Some properties of local quadratic smoothers are given in Section 4. In Section 5 we generalize results from MSW (2007) for nonparametric regression with quasi-uniform covariates to the case when quasi-uniformity cannot be assumed. Sections 4 and 5 play a major role in the proof of our main result given in Section 6. Our approach is to regard the single-index model as a nonparametric regression model with estimated covariates $\hat{S}_j$. The randomness caused by the estimated parameters $\hat{\theta}$ is handled using discretization and contiguity arguments, which are standard techniques in the construction of efficient estimators in semiparametric models.

2. Main Result

We begin by describing the local quadratic smoother. The value $\hat{\varrho}(s)$ of this estimator at $s \in \mathbb{R}$ equals the first component $\hat{\beta}_0$ of the minimizer $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ of

$$\frac{1}{nb_n} \sum_{j=1}^{n} \Big(Y_j - \beta_0 - \beta_1 \frac{\hat{S}_j - s}{b_n} - \beta_2 \Big(\frac{\hat{S}_j - s}{b_n} \Big)^2 \Big)^2 K\Big(\frac{\hat{S}_j - s}{b_n} \Big),$$

where K is a symmetric density with compact support $[-1, 1]$ and b_n is a bandwidth, i.e., b_n is a sequence of positive numbers that converges to zero.

We prove our main result, a uniform stochastic expansion of $\hat{\mathbb{F}}_n$, under the following conditions. Let $I = [a, b]$ be a compact interval of $\mathbb{R}$.

(R1) The regression function ϱ is twice continuously differentiable and satisfies

$$E[|\varrho(\theta^\top X) - \varrho(\theta_0^\top X) - (\theta - \theta_0)^\top X \varrho'(\theta_0{}^\top X)|^2] = o(\|\theta - \theta_0\|^2),$$

as $\|\theta - \theta_0\| \to 0$.

(R2) The $p \times p$ matrix
$$M = E[(\varrho'(S))^2 (X - E[X|S])(X - E[X|S])^\top]$$
has rank $p-1$.

(T) The estimator $\hat\theta$ satisfies $n^{1/2}(\hat\theta - \theta_0) = O_p(1)$ and is discretized.

(I) There are interior points $l_0 < u_0$ of I and functions $\bar\phi_{n,l}$ and $\bar\phi_{n,u}$ such that for all θ_n in Θ with $n^{1/2}(\theta_n - \theta_0)$ bounded we have
$$\phi_{n,l}(S_1, \ldots, S_n, \theta_n) = \bar\phi_{n,l}(\theta_n) + o_p(n^{-1/4}) = l_0 + o_p(1),$$
$$\phi_{n,u}(S_1, \ldots, S_n, \theta_n) = \bar\phi_{n,u}(\theta_n) + o_p(n^{-1/4}) = u_0 + o_p(1),$$
with $\phi_{n,l}$ and $\phi_{n,u}$ as in (1.2) and $S_j = \theta_0^\top X_j$.

(G1) The density g_{θ_0} is bounded and also bounded away from zero on I.

(G2) The map $\theta \mapsto \sqrt{g_\theta}$ is differentiable at θ_0 in L_2, i.e., there is a measurable function $\dot g_{\theta_0}$ from $\mathbb{R}$ into $\theta_0^\perp = \{v \in \mathbb{R}^p : v^\top \theta_0 = 0\}$ such that $\|\dot g_{\theta_0}\|$ is square-integrable and
$$\int \Big(\sqrt{g_\theta(s)} - \sqrt{g_{\theta_0}(s)} - (\theta - \theta_0)^\top \dot g_{\theta_0}(s)\Big)^2 \, ds = o(\|\theta - \theta_0\|^2)$$
holds as $\|\theta - \theta_0\| \to 0$.

(F1) The error variable has a finite third moment.

(F2) The error density f has finite Fisher information for location.

We shall now discuss these assumption. The first part of (R1) is used to derive appropriate properties of the local quadratic smoother of ϱ. The bias of this estimator is of order $o(b_n^2)$ which needs to be of order $o(n^{-1/2})$. The choice $b_n \sim n^{-1/4}$ used in Theorem 2.1 is the largest bandwidth satisfying this requirement. Larger bandwidth are allowed under additional smoothness assumptions on ϱ. For example, if the second derivative of ϱ is Hölder with exponent α, then we can take larger b_n subject to the constraint $b_n^{2+\alpha} = o(n^{-1/2})$. In particular, for $\alpha > 1/2$, the familiar choice $b_n \sim n^{-2/5}$ works. Instead of a local quadratic smoother, we could have worked with a local linear smoother. The bias of this estimator is of order $O(b_n^2)$. This would require a smaller bandwidth such as $b_n \sim (n \log n)^{-1/4}$ to guarantee that the bias is of order $o(n^{-1/2})$. A local linear smoother with this choice of bandwidth was used in MSW (2007).

The matrix M in condition (R2) cannot have full rank p as
$$\theta_0^\top M \theta_0 = E[(\varrho'(S))(\theta_0^\top (X - E[X|S]))^2] = E[(\varrho'(S))(S - E[S|S])^2] = 0.$$
Condition (R2) guarantees that $v^\top M v > 0$ for every unit vector v orthogonal to θ_0. This is needed to guarantee the existence of a root-n consistent estimator of θ_0, as required in condition (T).

The set $\theta_0^\perp$ appearing in (G2) is the tangent space of Θ at θ_0. The requirement in (G2) that $\dot g_{\theta_0}$ takes values in $\theta_0^\perp$ ensures that the derivative $\dot g_{\theta_0}$ is uniquely determined (up to almost everywhere equivalence). Without this assumption, the differentiability requirement would also hold with $\dot g_{\theta_0}$ replaced by $\dot g_{\theta_0} + h\theta_0$ for each square-integrable h. This follows from the fact that $(\theta - \theta_0)^\top \theta_0$ equals $-\|\theta - \theta_0\|^2/2$ for all θ in Θ. On the other hand, it suffices to verify the differentiability condition for some $\dot g_{\theta_0}$ that is not $\theta_0^\perp$-valued, because it then holds with $\dot g_{\theta_0}$ replaced by $(I_p - \theta_0\theta_0^\top)\dot g_{\theta_0}$, where I_p is the $p \times p$ identity matrix, and this replacement is $\theta_0^\perp$-valued in view of $\theta_0^\top(I_p - \theta_0\theta_0^\top) = 0$.

By the same token, we can replace in (R1) the derivative $X\varrho'(\theta_0^\top X)$ by the $\theta_0^\perp$-valued derivative $(I_p - \theta_0\theta_0^\top)X\varrho'(\theta_0^\top X)$. This and (F2) show that the score function for θ_0 is given by

$$\ell(\varepsilon)\varrho'(S)(I_p - \theta_0\theta_0^\top)X,$$

with $\ell = -f'/f$, the score function for location. The tangent space $\mathscr{T}$ for the nuisance parameter (ϱ, F, G), with G the distribution of X, consists of the function

$$\ell(\varepsilon)a(S) + c(\varepsilon) + b(X)$$

with $E[a^2(S)]$ finite, $E[b(X)] = 0$ and $E[b^2(X)]$ finite, $E[c(\varepsilon)] = E[\varepsilon c(\varepsilon)] = 0$ and $E[c^2(\varepsilon)]$ finite. The projection of the score function onto $\mathscr{T}^p$ is given by

$$\ell(\varepsilon)\varrho'(S)(I_p - \theta_0\theta_0^\top)E[X|S].$$

Thus the efficient score for estimating θ_0 is

$$\ell(\varepsilon)\varrho'(S)(I_p - \theta_0\theta_0^\top)(X - E[X|S])$$

and the efficient information matrix is

$$J_* = E[\ell^2(\varepsilon)](I_p - \theta_0\theta_0^\top)M(I_p - \theta_0\theta_0^\top) = E[\ell^2(\varepsilon)]M.$$

While the information matrix is not invertible, the map ϕ from ${\theta_0}^\perp$ to ${\theta_0}^\perp$ defined by it, i.e. $\phi(v) = J_*v$, $v \in {\theta_0}^\perp$, is invertible. Finally, the efficient influence function for estimation $F(t)$ is given by

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

This can be deduced from the results in Müller and Schick (2016) and the form of the present tangent space.

For the construction of root-n consistent estimators of θ_0 we refer to Carroll, Fan, Gijbels and Wand (1997), Wang, Xue, Zhu, and Chong (2010), and Xia and Härdle (2006), who develop $n^{1/2}$-consistent estimators of the underlying Euclidean parameters in a class of partially linear single-index models. Cui, Härdle and Zhu (2011) use a method of estimating functions to develop estimators of θ_0 that satisfy condition (T) for a large class of single-index model. Their estimator of θ_0 is found to have smaller or equal limiting variance than that of Carroll et al. (1997). See also the correction note by Li et al. (2011) pertaining to the reference Wang et al. (2010). The method of Hall and Yao (2005) provides yet another approach to obtain a root-n consistent estimator.

Condition (T) also requires that the root-n consistent estimator of θ_0 is discretized. Such an estimator can be obtained by discretizing any preliminary root-n consistent estimator $\hat\theta$ on grids with mesh width $n^{-1/2}$, e.g., by replacing it by the closest point on the grid, so the change is at most $n^{-1/2}$ and consistency is preserved. This trick simplifies the proofs since we can replace $\hat\theta_n$ by a *nonrandom* sequence $\theta_n = \theta_0 + O(n^{-1/2})$, see, e.g., Le Cam (1985) or van der Vaart (1998).

We use condition (G2) to establish that the distributions of $(\theta_n^\top X_1, \dots, \theta_n^\top X_n)$ and $(\theta_0^\top X_1, \dots, \theta_0^\top X_n)$ are mutually contiguous whenever $\theta_n = \theta_0 + O(n^{-1/2})$. This implies that (I) holds with each $S_j = \theta_0^\top X_j$ replaced by $\theta_n^\top X_j$. This and (T) then allow us to conclude that $\hat l$ is a consistent estimator of l_0, more precisely, we have

$$\hat l = \bar\phi_{n,l}(\hat\theta) + o_p(n^{-1/4}) = l_0 + o_p(1).$$

An analogous statement holds for $\hat{u}$. Similar arguments yield

$$\frac{N_n}{n} = \frac{1}{n}\sum_{j=1}^{n}\hat{\delta}_j = P(l_0 \le \theta_0^\top X \le u_0) + o_p(1). \tag{2.1}$$

Let $\hat{q}_\alpha$ denote the α-th sample quantile constructed from the estimated indices $\hat{S}_1,\dots,\hat{S}_n$. Recall that the sample quantile based on independent observations from a density is a root-n consistent estimator of the quantile whenever the density is positive and continuous at this quantile. Thus condition (I) is met by

$$\hat{I} = [\hat{l},\hat{u}] = [\hat{q}_{\alpha_1},\hat{q}_{\alpha_2}],$$

with $0 < \alpha_1 < \alpha_2 < 1$, if g_{θ_0} is continuous and positive on an open interval containing the α_1 and α_2-quantiles of g_{θ_0}. In particular, condition (I) holds for any such α_1 and α_2 if g_{θ_0} is continuous and the set $\{g_{\theta_0} > 0\}$ is an interval.

The moment assumption (F1) is used to derive properties of the local quadratic smoothers. The assumption (F2) together with (R1) is used to obtain contiguity. It also guarantees that the density f is Hölder with exponent $1/2$, which meets one of the requirements in MSW (2007), namely, the density f to be Hölder with exponent greater than $1/3$.

We now state our main result, the uniform stochastic expansion of the estimator $\hat{\mathbb{F}}_n$ introduced in (1.3). This expansion is similar to the expansions obtained in MSW (2007, 2009a) in semiparametric and nonparametric regression models. The difference is the presence of weights $w(\theta_0^\top X_j)$, where

$$w(s) = \frac{\mathbf{1}[l_0 \le s \le u_0]}{P(l_0 \le \theta_0^\top X \le u_0)}, \quad s \in \mathbb{R}.$$

Theorem 2.1. *Suppose the model* (1.1) *and the conditions (R1), (R2), (T), (I), (G1), (G2), (F1) and (F2) hold. In addition, assume that the kernel K has a Hölder continuous second derivative, and the bandwidth b_n satisfies $b_n \sim n^{-1/4}$. Then we have the uniform stochastic expansion*

$$\sup_{t\in\mathbb{R}} \left|\hat{\mathbb{F}}_n(t) - F(t) - \mathbb{W}_n(t)\right| = o_p(n^{-1/2}) \tag{2.2}$$

with

$$\mathbb{W}_n(t) = \frac{1}{n}\sum_{j=1}^{n} w(\theta_0^\top X_j)\big[\mathbf{1}[\varepsilon_j \le t] - F(t) + f(t)\varepsilon_j\big], \quad t \in \mathbb{R}.$$

Remark 2.1. The above result shows that the influence function of the estimator $\hat{\mathbb{F}}_n(t)$ is

$$\phi_t(Y,X) = w(\theta_0^\top X)\big[\mathbf{1}[\varepsilon \le t] - F(t) + f(t)\varepsilon\big],$$

which is the efficient influence function for estimating $F(t)$ multiplied by $w(\theta_0^\top X)$. The asymptotic variance of our estimator thus equals the efficient variance multiplied by $E[w^2(\theta_0^\top X)]$. This factor equals $1/p_0$ with

$$p_0 = P(l_0 \le \theta_0^\top X \le u_0).$$

Thus our estimator is nearly efficient if p_0 is close to one.

3. An application

We shall now discuss an application of (2.2) for deriving an asymptotically distribution free (ADF) test for fitting a known error distribution in the model (1.1). For this we introduce the process

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

Note that $n\mathrm{Cov}(\mathbb{Z}_n(s), \mathbb{Z}_n(t)) = C(s,t)$ and $n\mathrm{Cov}(\mathbb{W}_n(s), \mathbb{W}_n(t)) = (1/p_0)C(s,t)$ with

$$C(s,t) = \mathrm{Cov}(\mathbf{1}[\varepsilon \le s] + f(s)\varepsilon, \mathbf{1}[\varepsilon \le t] + f(t)\varepsilon), \quad s,t \in \mathbb{R}.$$

Recall, say from Koul (2002), that $n^{1/2}\mathbb{Z}_n$ converges weakly to a continuous Gaussian process $\mathbb{Z}$ with mean zero and covariance function C. Thus Theorem 2.1 implies

$$n^{1/2}(\hat{\mathbb{F}}_n - F) \to_D p_0^{-1/2}\mathbb{Z},$$

where $\to_D$ denotes weak convergence in the Skorokhod space $D[-\infty,\infty]$ and uniform metric. By (2.1), $\hat{p}_n = N_n/n$ is a consistent estimator of p_0 and we conclude

$$N_n^{1/2}(\hat{\mathbb{F}}_n - F) \to_D \mathbb{Z}. \tag{3.1}$$

An analog of Theorem 2.1 is obtained in MSW (2009a) for the ordinary nonparametric residual empirical process $\hat{F}_n$ in a class of nonparametric regression models. They established, under some conditions on the regression function and F, the expansion

$$n^{1/2}\sup_{t\in\mathbb{R}}|\hat{F}_n(t) - F(t) - \mathbb{Z}_n(t)| = o_p(1). \tag{3.2}$$

Let F_0 be a known distribution function having zero mean, a finite third moment and finite Fisher information for location. Consider the problem of testing $H_0 : F = F_0$ versus the alternative that H_0 is not true. In the context of nonparametric regression models, Khmaladze and Koul (2009) (KK) used the expansion (3.2) to show that under H_0 a certain transform of $\hat{F}_n$ converges weakly in $D[-\infty,\infty]$ and uniform metric to $B \circ F_0$, where B is standard Brownian motion on $[0,\infty)$. The results (2.2) and (3.1) used with $F = F_0$ enable one to conclude that the analog of this transform will also converge weakly, under H_0, to $B \circ F_0$. For the sake of completeness we describe this transform here.

Let f_0 be density of F_0 and f_0' be its a.e. derivative. Define

$$h(x) = \big(1, -f_0'(x)/f_0(x)\big)^\top, \quad \sigma^2(x) = \int_x^\infty \Big(\frac{f_0'(y)}{f_0(y)}\Big)^2 dF_0(y),$$
$$\Gamma_{F_0(x)} = \int_x^\infty h(x)h^\top(x)dF_0(x) = \begin{pmatrix} 1 - F_0(x) & f_0(x) \\ f_0(x) & \sigma^2(x)\end{pmatrix}, \quad x \in \mathbb{R}.$$

Let

$$K_n(t) = \int_{-\infty}^t h^\top(s)\Gamma^{-1}_{F_0(s)}\int_s^\infty h(z)\,d\hat{\mathbb{F}}_n(z)\,dF_0(s), \quad t \in \mathbb{R}.$$

The transformed process is

$$U_n(t) = n^{1/2}\big(\hat{\mathbb{F}}_n(t) - K_n(t)\big), \quad t \in \mathbb{R}.$$

See the discussion in KK for the existence of this transform. Arguing as in KK, one can show with the help of (2.2) and (3.1) that under appropriate conditions $\hat{p}_n^{1/2} U_n \to_D B \circ F_0$. As a consequence, under H_0,

$$D_n = \sup_{t \in \mathbb{R}} |\hat{p}_n^{1/2} U_n(t)| = \sup_{t \in \mathbb{R}} |N_n^{1/2}(\hat{\mathbb{F}}_n(t) - K_n(t))| \to_D \sup_{0 \le s \le 1} |B(s)|,$$

and the test based on D_n is ADF for testing H_0 in the single-index model (1.1). Perhaps it is worth pointing out that this test differs from its analog used for fitting an error distribution in the one sample location model only in that the scale factor $n^{1/2}$ is replaced by $N_n^{1/2}$ and the ordinary residual empirical distribution function of the one sample location model is replaced by $\hat{\mathbb{F}}_n$.

4. Properties of Local Quadratic Smoothers

In this section we describe some large sample properties of a local quadratic smoother of the regression function r on a closed interval $I = [a, b]$ with $a < b$ for the nonparametric regression model

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

where ε and Z are independent random variables, ε has mean zero and a finite third moment, Z has a *bounded* density g, and r is twice continuously differentiable. Let $(Y_1, Z_1), \dots, (Y_n, Z_n)$ denote independent copies of the pair (Y, Z) from the above regression model.

The local quadratic smoother $\hat{r}$ associated with a kernel K and a bandwidth c_n is defined as follows. The value $\hat{r}(z)$ of this estimator at z is given by the first component of the minimizer $\hat{\beta}(z) = (\hat{\beta}_0(z), \hat{\beta}_1(z), \hat{\beta}_2(z))^\top$ of

$$L(\beta) = \frac{1}{nc_n} \sum_{j=1}^n \Big(Y_j - \beta_0 - \beta_1 \frac{Z_j - z}{c_n} - \beta_2 \Big(\frac{Z_j - z}{c_n} \Big)^2 \Big)^2 K\Big(\frac{Z_j - z}{c_n} \Big).$$

We assume throughout that K is a symmetric density with support $[-1, 1]$ and make additional assumption as needed.

In what follows we use the following notation. For a $k \times m$ matrix A, we let $\|A\|$ denote its Euclidean norm

$$\|A\| = \left(\sum_{i=1}^k \sum_{j=1}^m A_{ij}^2 \right)^{1/2}.$$

For a function M from the interval I to the set of $k \times m$ matrices we set

$$\|M\|_* = \sup_{x \in I} \|M(x)\|.$$

If this function is differentiable with derivative M', then we set

$$\|M\|_{1,\gamma} = \|M\|_* + \|M'\|_* + \sup_{x,y \in I, x < y} \frac{\|M'(x) - M'(y)\|}{|x - y|^\gamma}, \qquad 0 < \gamma < 1.$$

These norms apply to vectors ($m = 1$) and scalars ($k = m = 1$).

Set $\psi(x) = (1, x, x^2)^\top$ for $x \in \mathbb{R}$. Then the above criterion function becomes

$$L(\beta) = \frac{1}{nc_n} \sum_{j=1}^n \Big(Y_j - \beta^\top \psi\Big(\frac{Z_j - z}{c_n} \Big) \Big)^2 K\Big(\frac{Z_j - z}{c_n} \Big).$$

Routine calculations show that the minimizer $\hat{\beta}(z)$ of $L(\beta)$ solves the normal equations

$$\hat{W}(z)\hat{\beta}(z) = \hat{A}(z) + \hat{B}(z),$$

where

$$\hat{W}(z) = \frac{1}{nc_n}\sum_{j=1}^{n}\psi\Big(\frac{Z_j - z}{c_n}\Big)\psi^\top\Big(\frac{Z_j - z}{c_n}\Big)K\Big(\frac{Z_j - z}{c_n}\Big),$$

$$\hat{A}(z) = \frac{1}{nc_n}\sum_{j=1}^{n}\varepsilon_j\psi\Big(\frac{Z_j - z}{c_n}\Big)K\Big(\frac{Z_j - z}{c_n}\Big),$$

$$\hat{B}(z) = \frac{1}{nc_n}\sum_{j=1}^{n}r(Z_j)\psi\Big(\frac{Z_j - z}{c_n}\Big)K\Big(\frac{Z_j - z}{c_n}\Big).$$

Since K has support $[-1,1]$ and r'' is uniformly continuous on compact sets, a Taylor expansion yields

$$\|\hat{B} - \hat{W}\dot{r}_{c_n}\|_* = \sup_{z\in I}|\hat{B}(z) - \hat{W}(z)\dot{r}_{c_n}(z)| = o(c_n^2)$$

with $\dot{r}_{c_n}(z) = (r'(z), c_n r'(z), c_n^2 r''(z)/2)^\top$. Direct calculations show that

$$\bar{W}(z) = E[\hat{W}(z)] = \int \psi(u)\psi^\top(u)K(u)g(z + c_n u)\,du.$$

In order to prove Theorem 4.1 below, which lists some important properties of the smoother, we need the following two lemmas. The first lemma is an immediate consequence of the definition of $\bar{W}$ and the fact that the matrix

$$\int_A \psi(u)\psi^\top(u)K(u)\,du$$

is positive definite for any subinterval A of $[-1,1]$ of positive length. Recall that g is bounded.

LEMMA 4.1. *Suppose g is also bounded away from zero on I. Then there is an α, $0<\alpha<1$, such that the eigenvalues of $\bar{W}(z)$ fall into the interval $[\alpha, 1/\alpha]$ for all z in I and all c_n satisfying $c_n \le l/2$, where l is the length of the interval I.*

The next lemma is a consequence of Corollary 4.2 in MSW (2007) with their δ equal to zero. Note that Z has a bounded density as required there. We also use the fact that ε has finite third moment.

LEMMA 4.2. *Suppose w is an integrable and Hölder continuous function and $\log n/(nc_n)$ is bounded. Then the rate*

$$\sup_{z\in I}\Big|\frac{1}{nc_n}\sum_{j=1}^{n}w\Big(\frac{Z_j - z}{c_n}\Big) - E[w\Big(\frac{Z_j - z}{c_n}\Big)]\Big| = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big)$$

holds. Moreover, if $E|\varepsilon|^3 < \infty$ and $\log n/(c_n n^{1/3})$ is bounded, then the rate

$$\sup_{z\in I}\Big|\frac{1}{nc_n}\sum_{j=1}^{n}\varepsilon_j w\Big(\frac{Z_j - z}{c_n}\Big)\Big| = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big)$$

holds.

In view of Lemma 4.2, from now on we assume that $\log n/(c_n n^{1/3})$ is bounded. It then follows from Lemma 4.1 that

$$\|\bar{W}\|_* = O(1) \quad \text{and} \quad \|\bar{W}^{-1}\|_* = O(1).$$

Furthermore, Lemma 4.2 applied to the entries of the matrices implies that

$$\|\hat{W} - \bar{W}\|_* = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big)$$

and

$$\|\hat{A}\|_* = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big),$$

provided K is Hölder. It follows that the matrices $\hat{W}(z)$, $z \in I$, are invertible on the event $\{\|\hat{W} - \bar{W}\|_* < \alpha\}$, whose probability converges to one. On this event we have

$$\|\hat{W}\|_* = O(1), \qquad \|\hat{W}^{-1}\|_* = O(1),$$

and

$$\|\hat{W}^{-1} - \bar{W}^{-1}\|_* = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big).$$

Moreover, on this event we have the identity

$$\begin{aligned} &\hat{\beta}(z) - \dot{r}_{c_n}(z) - \bar{W}(z)^{-1}\hat{A}(z) \\ &\quad = \ (\hat{W}(z)^{-1} - \bar{W}(z)^{-1})\hat{A}(z) + \hat{W}(z)^{-1}(\hat{B}(z) - \hat{W}(z)\dot{r}_{c_n}(z)), \quad z \in I. \end{aligned}$$

Using the above properties we obtain the following result. Recall that we assumed that g is bounded, that ε has a finite third moment and that K is a symmetric density with support $[-1, 1]$.

PROPOSITION 4.1. *Suppose g is also bounded away from zero on I and the kernel K is also Hölder. Then the uniform stochastic expansion*

$$\|\hat{\beta} - \dot{r}_{c_n} - \bar{W}^{-1}\hat{A}\|_* = O_p\Big(\frac{\log n}{nc}\Big) + o_p(c_n^2)$$

holds.

Thus, under the assumptions of the proposition and $c_n \sim n^{-1/4}$, we have the expansion

$$\sup_{z\in I} |\hat{r}(z) - r(z) - [1, 0, 0]\bar{W}(z)^{-1}\hat{A}(z)| = o_p(n^{-1/2}).$$

Next, we investigate the magnitude of the process

$$\hat{C}(z) = \bar{W}^{-1}(z)\hat{A}(z), \quad z \in I.$$

PROPOSITION 4.2. *Suppose g is bounded away from zero on I and K has a Hölder continuous second derivative. Then the map $z \mapsto \hat{C}(z)$ is twice differentiable and the following rates hold.*

$$\|\hat{C}\|_* = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big),$$

$$\|c_n\hat{C}'\|_* = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big),$$

$$\|c_n^2\hat{C}''\|_* = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big).$$

PROOF. Note that

$$\|\hat{C}\|_* \le \|\bar{W}^{-1}\|_* \|\hat{A}\|_* = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big).$$

By the properties of the kernel K, the function $z \mapsto \hat{C}(z)$ is twice continuously differentiable with the first derivative given by

$$\hat{C}'(z) = \bar{W}^{-1}(z)\hat{A}'(z) - \bar{W}^{-1}(z)\bar{W}'(z)\bar{W}^{-1}(z)\hat{A}(z)$$

and the second derivative given by

$$\begin{aligned}\hat{C}''(z) &= \bar{W}^{-1}(z)\hat{A}''(z) - 2\bar{W}^{-1}(z)\bar{W}'(z)\bar{W}^{-1}(z)\hat{A}'(z)\\ &\quad + 2\bar{W}^{-1}(z)\bar{W}'(z)\bar{W}^{-1}(z)\bar{W}'(z)\bar{W}^{-1}(z)\hat{A}(z) - \bar{W}^{-1}(z)\bar{W}''(z)\bar{W}^{-1}(z)\hat{A}(z).\end{aligned}$$

We write the matrix $[c_n\hat{A}'(z), c_n^2\hat{A}''(z)]$ as

$$\frac{1}{nc_n}\sum_{j=1}^n \varepsilon_j \Psi\Big(\frac{Z_j - z}{c_n}\Big)$$

with $\Psi = [(K\psi)', (K\psi)'']$. By the assumption on K, the entries of Ψ are integrable and Hölder. Thus we obtain from Lemma 4.2 that

$$\|c_n\hat{A}'\|_* + \|c_n^2\hat{A}''\|_* = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big).$$

Rewrite the matrix $[c_n\bar{W}'(z), c_n^2\bar{W}''(z)]$ as

$$\int g(u + c_n x)[V'(x), V''(x)]\,dx$$

with $V = K\psi\psi^\top$. From this we conclude that $\|c_n\bar{W}'\|_* + \|c_n^2\bar{W}''\|_* = O(1)$. Combining the above we obtain

$$\|c_n\hat{C}'\|_* \le \|\bar{W}^{-1}\|_*\|c_n\hat{A}'\|_* + \|\bar{W}^{-1}\|_*^2\|c_n\bar{W}'\|_*\|\hat{A}\|_*$$

and

$$\begin{aligned}\|c_n^2\hat{C}''\|_* &\le \|\bar{W}^{-1}\|_*\|c_n^2\hat{A}''\|_* + 2\|\bar{W}^{-1}\|_*^2\|c_n\bar{W}'\|_*\|c_n\hat{A}'\|_*\\ &\quad + 2\|\bar{W}^{-1}\|_*^3\|c_n\bar{W}'\|_*^2\|\hat{A}\|_* + \|\bar{W}^{-1}\|_*^2\|c_n^2\bar{W}''\|_*\|\hat{A}\|_*.\end{aligned}$$

This immediately yields the desired rates. □

We use Proposition 4.2 to obtain rates on the Hölder norms $\|\hat{C}\|_{1,\gamma}$, $0 < \gamma < 1$. Since we can bound $\|\hat{C}'(s) - \hat{C}'(t)\||s-t|^{-\gamma}$ by $\|\hat{C}''\|_* c_n^{1-\gamma}$ for $0 < |s-t| \le c_n$ and by $2\|\hat{C}'\|_* c_n^{-\gamma}$ for $|t-s| > c_n$, we have the following result.

PROPOSITION 4.3. *Suppose the assumptions of Proposition 4.2 hold and* $0 < \gamma < 1$. *Then the rate*

$$\|\hat{C}\|_{1,\gamma} = O_p\Big(\Big(\frac{\log n}{nc_n^{3+2\gamma}}\Big)^{1/2}\Big)$$

holds. In particular, for $c_n \sim n^{-1/4}$ *and* $\gamma < 1/2$, *one has*

$$\|\hat{C}\|_{1,\gamma} = o_p(1).$$

The next result summarizes properties of the local quadratic smoother $\hat{r}$ if the bandwidth is proportional to $n^{-1/4}$.

THEOREM 4.1. *Suppose g is bounded away from zero on I, K has a Hölder continuous second derivative and the bandwidth satisfies $c_n \sim n^{-1/4}$. Then the following hold with $\hat{c} = [1,0,0]\hat{C}$ the first coordinate of $\hat{C}$.*

$$\sup_{z\in I} |\hat{r}(z) - r(z) - \hat{c}(z)| = o_p(n^{-1/2}), \tag{4.1}$$

$$\int_I \hat{c}^2(z) g(z)\, dz = O_p(n^{-3/4}), \tag{4.2}$$

and, for $0 < \gamma < 1/2$,

$$\sup_{z\in I} |\hat{c}(z)| + \sup_{z\in I} |\hat{c}'(z)| + \sup_{s,t\in I, s<t} \frac{|\hat{c}'(t) - \hat{c}'(s)|}{|t-s|^\gamma} = o_p(1). \tag{4.3}$$

Moreover, for any square-integrable functions $v, v_1, v_2, \dots$ satisfying

$$\int_I (v_n(z) - v(z))^2\, dz = o(1),$$

we have the expansion

$$\int_I \hat{c}(z) v_n(z) g(z)\, dz = \frac{1}{n}\sum_{j=1}^n \varepsilon_j \mathbf{1}[Z_j \in I] v(Z_j) + o_p(n^{-1/2}). \tag{4.4}$$

PROOF. Claim (4.1) is a consequence of Proposition 4.1 and (4.3) of Proposition 4.3. Statement (4.2) follows from the bounds $\|\bar{W}^{-1}\|_* = O(1)$ and

$$nc_n E[\|\hat{A}(z)\|^2] \le E[\varepsilon^2] E\Big[\frac{3}{c_n} K^2\Big(\frac{Z-z}{c_n}\Big)\Big] \le 3E[\varepsilon^2] \int g(z + c_n u) K^2(u)\, du$$

and the boundedness of g. Here we used $\|\psi K\|^2 \le 3K^2$.

In order to prove (4.4) we set

$$\tilde{v}_n(z) = \mathbf{1}[z \in I] v_n(z) g(z) [1,0,0] \bar{W}^{-1}(z),$$

and

$$V_n(z) = \int \tilde{v}_n(z - c_n u)\psi(u) K(u)\, du \quad \text{and} \quad V(z) = \mathbf{1}[z\in I] v(z).$$

Then rewrite the left-hand side of (4.4) as

$$\frac{1}{nc_n}\sum_{j=1}^n \varepsilon_j \int \tilde{v}_n(z)\psi\Big(\frac{Z_j - z}{c_n}\Big) K\Big(\frac{Z_j - z}{c_n}\Big)\, dz = \frac{1}{n}\sum_{j=1}^n \varepsilon_j V_n(Z_j).$$

Let $\Psi = \int \psi(u)\psi^\top(u) K(u)\, du$. Since the density g is bounded, it is square-integrable. This and the translation continuity in L_2 yield the convergence

$$\int \Big\| \int g(z + c_n u)\psi(u)\psi^\top(u) K(u)\, du - g(z)\Psi \Big\|^2 dz \to 0.$$

Since g is bounded away from zero on I, we conclude from this that the map $z \mapsto \mathbf{1}[z \in I]\bar{W}^{-1}(z)$ converges to the map $z \mapsto \mathbf{1}[z \in I](g(z)\Psi)^{-1}$ in Lebesgue measure. An application of Lebesgue's dominated convergence theorem now yields that $\tilde{v}_n$ converges in L_2 to $\tilde{v}$, where

$$\tilde{v}(z) = \mathbf{1}[z \in I] v(z) [1,0,0] \Psi^{-1}.$$

Using this, the identity $\tilde{v}(z)\int \psi(u)K(u)\,du = \tilde{v}(z)\Psi[1,0,0]^\top = \mathbf{1}[z \in I]v(z) = V(z)$, and the translation continuity in L_2, we derive

$$\begin{aligned}\Delta_n &= \int (V_n(z) - V(z))^2\,dz \\ &= \int \Big| \int \tilde{v}_n(z - c_n u)\psi(u)K(u)\,du - \mathbf{1}[z \in I]v(z)\Big|^2\,dz \\ &\le 2\int \|\tilde{v}_n(z) - \tilde{v}(z)\|^2\,dz \int \|\psi(u)K(u)\|^2\,du \\ &\quad + 2\int \Big|\int (\tilde{v}(z - c_n u) - \tilde{v}(z))\psi(u)K(u)\,du\Big|^2\,dz = o(1).\end{aligned}$$

From the above we conclude that n times the second moment of the difference

$$\int_I \hat{c}(z)v_n(z)g(z)\,dz - \frac{1}{n}\sum_{j=1}^n \varepsilon_j v(Z_j) = \frac{1}{n}\sum_{j=1}^n \varepsilon_j(V_n(Z_j) - V(Z_j))$$

equals $E[\varepsilon^2]E[(V_n(Z) - V(Z))^2]$, which is bounded by a constant times Δ_n. This implies the desired (4.4). $\square$

Remark 4.1. Let $\hat{v}$ be an estimator of some square-integrable function v_0. Suppose there is a sequence of square-integrable functions v_n such that

$$\int_I (\hat{v}(z) - v_n(z))^2\,dz = o(n^{-1/4}) \quad \text{and} \quad \int_I (v_n(z) - v_0(z))^2\,dz = o(1).$$

Then under the assumption of the previous theorem the expansion

$$\int_I \hat{c}(z)\hat{v}(z)g(z)\,dz = \frac{1}{n}\sum_{j=1}^n \varepsilon_j \mathbf{1}[Z_j \in I]v_0(Z_j) + o_p(n^{-1/2})$$

holds. This follows from (4.2), (4.4), the inequality

$$\Big|\int_I \hat{c}(z)(\hat{v}(z) - v_n(z))g(z)\,dz\Big|^2 \le \int_I \hat{c}^2(z)g(z)\,dz \int (\hat{v}(z) - v_n(z))^2 g(z)\,dz$$

and the fact that g is bounded.

5. Estimating the error distribution in nonparametric regression

In this section we modify results from MSW (2007) to the case when the regressor is not quasi-uniform. We begin by extending their Theorems 2.1 and 2.2.

Let ε be a random variable with distribution function F, and let Z be a k-dimensional random vector with distribution Q, independent of ε. Let D be a non-negative function in $L_2(Q)$, and $\mathscr{D}$ be a set of measurable functions a such that $|a| \le D$ and $0 \in \mathscr{D}$. Let $\mathscr{V}$ be a class of measurable functions from $\mathbb{R}^k$ into $[0,1]$. We now give conditions on the classes $\mathscr{D}$ and $\mathscr{V}$ that imply that the class

$$\mathscr{H} = \{h_{a,v,t} : a \in \mathscr{D}, v \in \mathscr{V}, t \in \mathbb{R}\}$$

is $F \otimes Q$-Donsker, where

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

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

PROPOSITION 5.1. *Suppose that $\mathscr{V}$ is Q-Donsker. Assume that F has a finite second moment and a bounded density and that the bracketing numbers satisfy*

$$\int_0^1 \sqrt{\log N_{[\,]}(\eta^2, \mathscr{D}, L_1(Q))}\, d\eta < \infty. \tag{5.1}$$

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

PROOF. Let ϕ be the projection map from $\mathbb{R} \times \mathbb{R}^k$ into $\mathbb{R}^k$ so that $\phi(\varepsilon, Z) = Z$. Since $\mathscr{V}$ is Q-Donsker, the class $\tilde{\mathscr{V}} = \{v \circ \phi : v \in \mathscr{V}\}$ is $F \otimes Q$-Donsker. Let $\mathscr{H}_1 = \{h_{a,1,t} : a \in \mathscr{D}, t \in \mathbb{R}\}$ with $h_{a,1,t}(\varepsilon, Z) = \mathbf{1}[\varepsilon - a(Z) \le t]$. It follows from Theorem 2.1 of MSW (2007) that the class $\mathscr{H}_1$ is $F \otimes Q$ is Donsker. Since $\tilde{\mathscr{V}}$ and $\mathscr{H}_1$ are uniformly bounded (by 1) $F \otimes Q$-Donsker classes, their pairwise product $\tilde{\mathscr{V}} \cdot \mathscr{H}_1 = \{\tilde{v}h : \tilde{v} \in \tilde{\mathscr{V}}, h \in \mathscr{H}_1\}$ forms a $F \otimes Q$-Donsker class by Example 2.10.8 in van der Vaart and Wellner (1996). This is the desired result as $\mathscr{H}$ equals $\tilde{\mathscr{V}} \cdot \mathscr{H}_1$. $\square$

Now consider a regression model

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

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

$$\hat{W}(t,v) = \frac{1}{n}\sum_{j=1}^n v(Z_j)\mathbf{1}[\hat{\varepsilon}_j \le t], \quad W(t,v) = \frac{1}{n}\sum_{j=1}^n v(Z_j)\mathbf{1}[\varepsilon_j \le t], \quad t \in \mathbb{R},\ v \in \mathscr{V}.$$

PROPOSITION 5.2. *Let $\mathscr{D}$ and $\mathscr{V}$ be as in Proposition 5.1. Let $\mathscr{V}$ have envelope $\mathbf{1}_I$ for some compact convex set I with nonempty interior. Let F have a finite second moment and a density f that is Hölder with exponent $\xi \in (0,1]$. Additionally, assume that there is an $\hat{a}$ such that*

$$P(\hat{a} \in \mathscr{D}) \to 1, \tag{5.2}$$

$$\int \mathbf{1}_I |\hat{a}|^{1+\xi}\, dQ = o_p(n^{-1/2}), \tag{5.3}$$

$$\sup_{z \in I} |\hat{r}(z) - r(z) - \hat{a}(z)| = o_p(n^{-1/2}). \tag{5.4}$$

Then the uniform expansion

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

holds.

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

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

Then we can write

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

where

$$T_1(t,v) = \hat{W}(t,v) - \tilde{W}(t,v),$$
$$T_2(t,v) = \tilde{W}(t,v) - W_{\hat{a}}(t,v) - W(t,v) + W_0(t,v),$$
$$T_3(t,v) = W_{\hat{a}}(t,v) - W_0(t,v) - f(t)\int \hat{a}\, v\, dQ.$$

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

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

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

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

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

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

and the bound

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

where R_n denotes the left-hand side of (5.4). Since f is Hölder, f is bounded and F is Lipschitz with Lipschitz constant $\|f\|_\infty$. Thus we obtain

$$n^{1/2}(W_{\hat{a}}(t+R_n,v) - W_{\hat{a}}(t-R_n,v)) \le 2\|f\|_\infty n^{1/2}R_n = o_p(1). \tag{5.5}$$

Moreover, for $s,t \in \mathbb{R}$ and $a,b \in \mathscr{D}$, we have the bound

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

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

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

The first of these statements and (5.3) imply

$$\sup_{t\in\mathbb{R}, v\in\mathscr{V}} |T_2(t,v)| = o_p(n^{-1/2}),$$

while the second, (5.4) and (5.5) imply

$$\sup_{t\in\mathbb{R}, v\in\mathscr{V}} |T_1(t,v)| = o_p(n^{-1/2}).$$

This completes the proof. □

Now fix a v_0 in $\mathscr{V}$ and let $\hat{v}$ denote an estimator of v_0. Suppose this estimator satisfies

$$P(\hat{v} \in \mathscr{V}) \to 1 \tag{5.6}$$

and

$$\int (\hat{v}(z)) - v_0(z))^2 \, dQ(z) = o_p(1). \tag{5.7}$$

It follows that

$$\hat{v}_* = \frac{1}{n} \sum_{j=1}^{n} \hat{v}(Z_j) = \int v_0 \, dQ + o_p(1). \tag{5.8}$$

Moreover, under the assumptions of Proposition 5.2, the uniform expansion

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

holds. We write $W(t, \hat{v}) = \hat{v}_* F(t) + U(t, \hat{v})$, where

$$U(t, v) = \frac{1}{n} \sum_{j=1}^{n} v(Z_j)(\mathbf{1}[\varepsilon_j \le t] - F(t)).$$

Note that the functions $(y, z) \mapsto v(z)(\mathbf{1}[\varepsilon \le t] - F(t))$ with $v \in \mathscr{V}$ and $t \in \mathbb{R}$ form an $F \otimes Q$-Donsker class. Thus we find

$$\sup_{t \in \mathbb{R}} \Big| W(t, \hat{v}) - \hat{v}_* F(t) - U(t, v_0) \Big| = o_p(n^{-1/2}).$$

Combining the above yields the uniform expansion

$$\sup_{t \in \mathbb{R}} \Big| \hat{W}(t, \hat{v}) - \hat{v}_* F(t) - W_0(t, v_0) - f(t) \int \hat{a} \, \hat{v} \, dQ \Big| = o_p(n^{-1/2}).$$

This finding is summarized in the following theorem.

Proposition 5.3. *Suppose the assumptions of Proposition 5.2 are met and $\hat{v}$ is an estimator which satisfies (5.6) and (5.7) for some $v_0 \in \mathscr{V}$ with $\bar{v}_0 = \int v_0 \, dQ$ positive. Then the uniform expansion*

$$\sup_{t \in \mathbb{R}} \Big| \hat{W}(t, \hat{v})/\hat{v}_* - F(t) - U(t, v_0)/\bar{v}_0 - f(t) \int \hat{a} \hat{v} \, dQ / \hat{v}_* \Big| = o_p(n^{-1/2})$$

holds with $\hat{v}_*$ as in (5.8).

Now assume that Z has dimension 1 with a density g that is bounded and bounded away on the interval $I = [a, b]$ with $-\infty < a < b < \infty$. We take $\mathscr{D}$ to be the set of all functions h that vanish off I and satisfy

$$\|h\|_{1,1/4} = \sup_{z \in I} |h(z)| + \sup_{z \in I} |h'(z)| + \sup_{a \le s < t \le b} \frac{|h'(s) - h(t)|}{|t - s|^{1/4}} \le 1.$$

Here we have to understand h' as the derivative of the restriction of h to I so that $h'(a)$ is the right-hand derivative of h at a and $h'(b)$ is the left-hand derivative at b. It follows from Theorem 2.7.1 in van der Vaart and Wellner (1996) that the entropy condition (5.1) holds as $\log N_{[\,]}(\eta^2, \mathscr{D}, L_1(Q))$ is bounded by $C(b - a)(1/\eta)^{8/5}$, for some positive constant C. It follows from the results in the previous section that a local quadratic smoother with bandwidth $c_n \sim n^{-1/4}$ and appropriate kernel K satisfies the conditions (5.2) to (5.4) with $\hat{a} = \hat{c}$ and $\xi > 1/3$. Let $\mathscr{V}$ be the set of indicator functions of intervals $[l, u]$ with $a \leq l < u \leq b$. This is clearly a Donsker class. Now take

$$\hat{v} = \mathbf{1}_{[\hat{l}, \hat{u}]}, \quad v_n = \mathbf{1}_{[l_n, u_n]} \quad \text{and} \quad v_0 = \mathbf{1}_{[l_0, u_0]}$$

with $l_0 < u_0$ interior points of I. Note that

$$\int (\mathbf{1}_{[s,t]}(z) - \mathbf{1}_{[l,u]}(z))^2 \, dz \leq |s - l| + |t - u|.$$

We have the following result for the regression problem with a one-dimensional Z.

THEOREM 5.1. *Suppose Z has a bounded density that is bounded away from zero on the interval $I = [a, b]$, ε has mean zero, a finite third moment and a density f that is Hölder with exponent greater than $1/3$, the kernel K has a Hölder continuous second derivative, the bandwidth satisfies $c_n \sim n^{-1/4}$. the lower endpoints of the above intervals satisfy $\hat{l} = l_n + o_p(n^{-1/4})$ and $l_n \to l_0$ and the upper endpoints satisfy $\hat{u} = u_n + o_p(n^{-1/4})$ and $u_n \to u_0$. Then the estimator*

$$\hat{F}(t) = \frac{1}{\hat{N}} \sum_{j=1}^{n} \mathbf{1}[\hat{l} \leq Z_j \leq \hat{u}] \mathbf{1}[\hat{\varepsilon}_j \leq t], \quad t \in \mathbb{R},$$

with $\hat{N} = \sum_{j=1}^{n} \mathbf{1}[\hat{l} \leq Z_j \leq \hat{u}]$, satisfies the uniform expansion

$$\sup_{t \in \mathbb{R}} \left| \hat{F}(t) - F(t) - \frac{1}{n} \sum_{j=1}^{n} \frac{\mathbf{1}[l_0 \leq Z_j \leq u_0]}{P(l_0 \leq Z \leq u_0)} \Big[\mathbf{1}[\varepsilon_j \leq t] - F(t)) + f(t)\varepsilon_j \Big] \right| = o_p(n^{-1/2}).$$

6. Proof of Theorem 2.1

A key technical tool for proving Theorem 2.1 will be the use of two contiguity results. For the sake of self-containment, we shall briefly review the notion of contiguity of Le Cam (1960) and give the needed contiguity results that will be used in the proof; see also Le Cam (1986) and Hájek and Šidák (1967).

Let $(\Omega_n, \mathscr{A}_n, \{P_n, Q_n\})$ be a sequence of binary experiments. Then Q_n is *contiguous* to P_n if for every sequence A_n, $A_n \in \mathscr{A}_n$, $P_n(A_n) \to 0$ implies $Q_n(A_n) \to 0$. We say P_n and Q_n are *mutually contiguous* if Q_n is contiguous to P_n and P_n is contiguous to Q_n.

We now state a sufficient condition for contiguity of product measures. For this we assume that $(\Omega, \mathscr{A}, \mu)$ is a measure space and $\{\Gamma_\theta : \theta \in \Theta\}$ is a family of probability measures dominated by μ. Denote by γ_θ a density of Γ_θ with respect to μ. Suppose there is a measurable function $\dot{\gamma}_{\theta_0}$ from Ω into ${\theta_0}^\perp$ such that $\|\dot{\gamma}_{\theta_0}\|$ belongs to $L_2(\mu)$ and

$$\int (\gamma_\theta^{1/2} - \gamma_{\theta_0}^{1/2} - (\theta - \theta_0)^\top \dot{\gamma}_{\theta_0})^2 \, d\mu = o(\|\theta - \theta_0\|^2) \tag{6.1}$$

holds. Then the product measures $\Gamma_{\theta_n}^n$ and $\Gamma_{\theta_0}^n$ are mutually contiguous whenever $n^{1/2}(\theta_n - \theta_0)$ is bounded. See, e.g., van der Vaart (1998).

We shall use this result first with

$$\gamma_\theta(x.y) = \gamma_{1,\theta}(x, y) = f(y - \rho(\theta^\top x)), \quad x \in \mathbb{R}^p, y \in \mathbb{R},$$

and $\mu = G \otimes \lambda$ where G is the distribution of X and λ is the Lebesgue measure. It follows from (R1), (R2) and (F2) that (6.1) holds with

$$\dot\gamma_{\theta_0}(x,y) = \frac{-f'(y-\rho({\theta_0}^\top x)}{2f^{1/2}(y-\rho({\theta_0}^\top x)}\rho'({\theta_0}^\top x)(I_p - \theta_0\theta_0^\top)x.$$

Then we shall apply the result with

$$\gamma_\theta(x,y) = \gamma_{2,\theta}(x,y) = f(y)g_\theta(x), \quad x\in\mathbb{R}, y\in\mathbb{R},$$

and $\mu = \lambda\otimes\lambda$. It follows from (G2) that (6.1) holds with $\dot\gamma_{\theta_0}(x,y) = f^{1/2}(y)\dot g_{\theta_0}(x)$.

By the properties of $\hat\theta$ specified in (T), it suffices to prove the result with $\hat\theta$ replaced by non-stochastic sequences θ_n such that $n^{1/2}(\theta_n - \theta_0)$ is bounded. This is a standard argument used in the construction of efficient estimators in semiparametric models, see, e.g., Schick (1986) and references therein.

Now fix such a sequence θ_n and set

$$S_{n,j} = \theta_n^\top X_j, \quad \delta_{n,j} = \mathbf{1}[S_{n,j}\in I(\theta_n)] \quad \text{and} \quad \varepsilon_{n,j} = Y_j - \varrho(S_{n,j})$$

for $j = 1,\dots,n$. Let $\tilde\varrho$ denote the local linear smoother associated with minimizing

$$\frac{1}{nb_n}\sum_{j=1}^n \Big(Y_j - \beta_0 - \beta_1\frac{S_{n,j}-s}{b_n} - \beta_2\Big(\frac{S_{n,j}-s}{b_n}\Big)^2\Big)^2 K\Big(\frac{S_{n,j}-s}{b_n}\Big).$$

Moreover, we introduce

$$\tilde F(t) = \frac{\sum_{j=1}^n \mathbf{1}[\tilde l_n \le S_{nj} \le \tilde u_n]\mathbf{1}[Y_j - \tilde\varrho(S_{nj})\le t]}{\sum_{j=1}^n \mathbf{1}[\tilde l_n \le S_{nj}\le u_n]},$$

with $\tilde l_n = \phi_{n,l}(S_{n,1},\dots,S_{n,n},\theta_n)$ and $\tilde u_n = \phi_{n,u}(S_{n,1},\dots,S_{n,n},\theta_n)$ and set

$$\tilde{\mathbb{W}}_n(t) = \frac{1}{n}\sum_{j=1}^n w(S_{n,j})\big[\mathbf{1}[\varepsilon_{n,j}\le t] - F(t)) - f(t)\varepsilon_{n,j}\big], \quad t\in\mathbb{R}.$$

We achieve our goal by verifying the uniform stochastic expansions

$$\sup_{t\in\mathbb{R}}\Big|\tilde F(t) - F(t) - \tilde{\mathbb{W}}_n(t)\Big| = o_p(n^{-1/2}) \tag{6.2}$$

and

$$\sup_{t\in\mathbb{R}}|\tilde{\mathbb{W}}_n(t) - \mathbb{W}_n(t)| = o_p(n^{-1/2}). \tag{6.3}$$

To stress dependence on the parameter θ_0 we now write P_θ for the underlying probability measure when $\theta_0 = \theta$ and write $P_{n,\theta}$ for the joint distribution of the data $X_1, Y_1,\dots,X_n,Y_n$ under P_θ, for each $\theta\in\Theta$. It follows from the above that the sequences of distributions P_{n,θ_n} and P_{n,θ_0} are mutually contiguous. Thus it suffices to prove (6.2) under the measure P_{θ_n}. Under the measure P_{θ_n}, we have

$$Y_j = \varrho(S_{n,j}) + \varepsilon_{n,j}, \quad j = 1,\dots,n,$$

and derive that that the left-hand side of (6.2) is a function of the random vectors $(\varepsilon_{n,1}, S_{n,1})^\top,\dots,(\varepsilon_{n,n},S_{n,n})^\top$. Under P_{θ_n} these variables are independent with common density γ_{2,θ_n}. By another contiguity argument is thus suffices to prove (6.2) under the assumption that the random vectors $(\varepsilon_{n,1},S_{n,1})^\top,\dots,(\varepsilon_{n,n},S_{n,n})^\top$ are independent with density γ_{2,θ_0}. The desired (6.2) then follows from Theorem 5.1.

Note that the distribution of the process defined by the first average in (6.3) under the measure P_{θ_n} equals the distribution of the process defined by the second average in (6.3) under P_{θ_0}. Thus, by contiguity, the difference of these two processes is tight under P_{θ_0}. It suffices to prove (6.3) without the supremum, but for all t in $\mathbb{R}$. Fix such a t. We are left to verify

$$\frac{1}{n}\sum_{j=1}^{n} h_{\theta_n}(X_j,Y_j) - \frac{1}{n}\sum_{j=1}^{n} h_{\theta_0}(X_j,Y_j) = o_p(n^{-1/2}) \tag{6.4}$$

with

$$h_\theta(X,Y) = w(\theta^\top X)\Big\{\mathbf{1}[Y-\rho(\theta^\top X)\le t] - F(t) + f(t)(Y-\rho(\theta^\top X)).$$

Using the translation continuity in L_2, we verify

$$\iint \Big|h_\theta(x,y)\sqrt{\gamma_{1,\theta}(x,y)} - h_{\theta_0}(x,y)\sqrt{\gamma_{1,\theta_0}(x,y)}\Big|^2\, dG(x)dy \to 0$$

as $\theta\to\theta_0$. With $\ell = -f'/f$ the score function for location, we verify

$$\begin{aligned} D_{\theta_0} &= -\iint h_{\theta_0}(x,y)\ell(y-\varrho({\theta_0}^\top x))\varrho'({\theta_0}^\top x)(I_p-\theta_0\theta_0^\top)x\gamma_{1,\theta_0}(x,y)\, dG(x)dy \\ &= -E\big[(\mathbf{1}[\varepsilon\le t]-F(t)+f(t)\varepsilon)\ell(\varepsilon)\big]\; E\big[w(\theta_0^\top X)\varrho'(\theta_0^\top X)(I_p-\theta_0\theta_0^\top)X\big] = 0, \end{aligned}$$

because the first expectation in the product equals $-f(t)-0+f(t)=0$. Since the densities $\gamma_{1,\theta}$ are Hellinger differentiable at θ_0 with Hellinger derivative

$$\kappa_{\theta_0}(x,y) = \ell(y-\varrho({\theta_0}^\top x))\varrho'({\theta_0}^\top x)(I_p-\theta_0\theta_0^\top)x,$$

as shown above, the claim (6.4) follows from Theorem 2.3 in Schick (2001), which extends to the present parameter set Θ. His result is stated for open subsets of $\mathbb{R}^p$.

Dedication

The authors congratulate Winfried Stute on his 70th birthday and wish him uncountable years of good health and research productivity.

H.L.K.: Working with Winfried on several research projects has been one of my truly rewarding experiences. He is a great scholar and a great friend.

U.U.M.: In appreciation of encouragement and support on several occasions, generously given and gratefully received.

References

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

[2] Boldin, M.V. (1982). An estimate of the distribution of the noise in an autoregressive scheme. *Teor. Veroyatnost. i Primenen.*, **27(4)**, 805–810.

[3] Boldin, M.V. (1990). On testing hypotheses in the sliding average scheme by the Kolmogorov Smirnov and ω^2 tests. *Theory of Probability and Its Applications*, **34(4)**, 699–704.

[4] Boldin, M.V. (1998). On residual empirical distribution functions in ARCH models with applications to testing and estimation. *Mitt. Math. Sem. Giessen*, No. **235**, 40–66.

[5] Carroll, R.J., Fan, J., Gijbels, I., and Wand, M.P. (1997). Generalized Partially Linear Single-Index Models. *J. Amer. Statist. Assoc.*, **92**, 477–489.

[6] Cui, X., Härdle, W. and Zhu, L. (2011). Generalized single-index models: The EFM approach. *Ann. Statist.*, **39**, 1658–1688.

[7] Hájek, J. and Sidák, Z. (1967). *Theory of rank tests.* Academic Press, New York-London; Academia Publishing House of the Czechoslovak Academy of Sciences, Prague.

[8] Hall, P.J. and Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *Ann. Statist.*, **33**, 1404–1421.

[9] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics*, **58**, 71–120.

[10] Khmaladze E.V. and Koul H.L. (2009). Goodness-of-fit problem for errors in nonparametric regression: Distribution free approach. *Ann. Statist.*, **37**, 3165–3185.

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

[12] Koul, H.L. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.*, **41**, 1768–1773.

[13] Koul, H.L. (1991). A weak convergence result useful in robust autoregression. *J. Statist. Plann. Inference*, **29**, 291–308.

[14] Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* Lecture Notes in Statistics, **166**. Springer-Verlag, New York.

[15] Le Cam, L. (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, **3**, 37–98.

[16] Le Cam, L. (1986). *Asymptotic methods in statistical decision theory.* Springer Series in Statistics. Springer-Verlag, New York.

[17] Li, K.C. (1991). Sliced Inverse Regression for Dimension Reduction. *J. Amer. Statist. Assoc.*, **86**, 316–342.

[18] Li, T.T., Yang, H., Wang, J.L., Xue, L.G. and Zhu, L. (2011). Correction on Estimation for a partial linear single-index model. *Ann. Statist.*, **39**, 3441–3443.

[19] Müller, U.U., Schick, A. (2016). Efficiency transfer for regression models with responses missing at random. To appear in: *Bernoulli.*

[20] Müller, U.U., Schick, A. and Wefelmeyer, W. (2007). Estimating the error distribution function in semiparametric regression. *Statist. Decisions*, **25**, 1–18.

[21] Müller, U.U., Schick, A. and Wefelmeyer, W. (2009a). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statist. Probab. Lett.*, **79**, 957–964.

[22] Müller, U.U., Schick, A. and Wefelmeyer, W. (2009b). Estimating the innovation distribution in nonparametric autoregression. *Probab. Theory Related Fields*, **144**, 53–77.

[23] Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. of Mult. Analysis,* **101(5)**, 1067–1078.

[24] Neumeyer, N. and Selk, L. (2013). A note on non-parametric testing for gaussian innovations in ararch models. *J. of Time Series Analysis*, **34**, 362–367.

[25] Schick, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.*, **14**, 1139–1151.

[26] Schick, A. (2001). On asymptotic differentiability of averages. *Statist. Probab. Letters* **51**, 15–23.

[27] Stoker, T.M. (1986). Consistent estimation of scaled coefficients. *Econometrics*, **54**, 1461–1481.

[28] Stute, W. and Zhu, L. (2005). Nonparametric checks for single-index models. *Ann. Statist.*, **33**, 1048–1083.

[29] van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press.

[30] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes*, Springer.

[31] Xia, Y. (2008). A multiple-index model and dimension reduction. *J. Amer. Statist. Assoc.*, **103**, 1631–1640.

[32] Xia, Y. and Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *J. Multivariate Anal.*, **97**, 1162–1184.

[33] Xia, Y. and Li, W.K. (1999). On the estimation and testing of functional-coefficient linear models. *Statist. Sinica*, **9**, 735–758.

[34] Xia, Y., Tong, H. and Li, W.K. (2002). Single-index volatility models and estimation. *Statist. Sinica*, **12(3)**, 785–799.

[35] Xia, Y., Tong, H., Li, W.K. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. *J.R. Stat. Soc. B*, **64**, 363–410.

[36] Wang, J.L., Xue, L., Zhu, L. and Chong, Y.S. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.*, **38**, 246–274

Hira L. Koul, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA

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