The Annals of Statistics
1997, Vol. 25, No. 2, 613–641

NONPARAMETRIC MODEL CHECKS FOR REGRESSION[1]

BY WINFRIED STUTE

Justus-Liebig-University of Giessen

In this paper we study a marked empirical process based on residuals. Results on its large-sample behavior may be used to provide nonparametric full-model checks for regression. Their decomposition into principal components gives new insight into the question: which kind of departure from a hypothetical model may be well detected by residual-based goodness-of-fit methods? The work also contains a small simulation study on straight-line regression.

1. Introduction and main results. The purpose of the present paper is to study a general method for testing the goodness of fit of a parametric regression model. More precisely, let (X, Y) denote a random vector in $d+1$-dimensional Euclidean space. Assume that Y is integrable so that the associated regression function

$$m(x) = \mathbb{E}[Y|X = x], \qquad x \in \mathbb{R}^d,$$

is well defined (up to a null set). Much of the existing literature is concerned with parametric modeling in that m is assumed to belong to a given family

$$\mathscr{M} = \{m(\cdot, \theta)\colon \theta \in \Theta\}$$

of functions, where $\Theta \subset \mathbb{R}^p$ is a proper parameter set. In other words, it is assumed that $m(x) = m(x, \theta_0)$ for some "true parameter" θ_0. Many efforts have been devoted to the problem of how to estimate or to test for hypotheses about θ_0. The best known case is the linear model in which $m(x, \theta) = g^T(x)\theta$, g is a known vector-valued function and T denotes the transpose. For a thorough discussion of nonlinear regression, see Seber and Wild (1989).

Model diagnostics have been addressed, in the context of a given model, in most textbooks on regression. The proposed procedures are based on the observed residuals and are used mainly to detect and discuss the influence of selected data points. See Atkinson (1985), Cook and Weisberg (1982) and Davison and Tsai (1992) for details.

Starting with Nadaraya (1964) and Watson (1964), there has been much activity on how to estimate m in a completely nonparametric framework. This approach requires smoothing of the data rather than estimation of the finite-dimensional parameters and leads to less precise fits. Furthermore, parametric models are still attractive among practitioners because the parameter θ together with the functional form of $m(\cdot, \theta)$ describes, in a concise way, the

Received July 1994; revised June 1995.
[1]Work partially supported by the Deutsche Forschungsgemeinschaft.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62G10, 62G30, 62J02.
Key words and phrases. Marked empirical process, residuals, model check for regression, principal components, Cramér–von Mises, smooth and directional tests.

impact of X on the output Y. Since there may be several competing models, in order to prevent wrong conclusions, every statistical inference that is based on a model $\mathscr{M}$ should be accompanied by a proper model check, that is, by a test for

$$H_0\colon m \in \mathscr{M} \quad \text{versus} \quad H_1\colon m \notin \mathscr{M}. \tag{1.1}$$

Interestingly enough, a systematic study of (1.1) only started in the late 1980s. See, for example, Barry and Hartigan (1990), Buckley (1991), Cox, Koh, Wahba and Yandell (1988), Eubank and Hart (1992, 1993), Eubank and LaRiccia (1992), Eubank and Spiegelman (1990), Firth, Glosup and Hinkley (1991), Härdle and Mammen (1993), Joglekar, Schuenemeyer and LaRiccia (1989), Müller (1992), Staniswalis and Severini (1991), Stute and González Manteiga (1995) and Su and Wei (1991). See also Randles (1984), Shillington (1979), Neill and Johnson (1985) and Utts (1982) for some earlier references.

In most of these papers the X-variable was fixed rather than random, and sometimes further very restrictive distributional assumptions on the error distribution (like normality) and the design (equally distributed in a given compact set) had to be added. Most of these assumptions cannot be justified in a random design setup.

Eubank and Hart (1993) nicely pointed out that a common feature in most of the above work is as follows: assuming (for simplicity) that $\mathscr{M}$ is a finite-dimensional subspace of some L^2 (the space of square-integrable functions), we may find an orthonormal expansion of the true m into some $m(\cdot, \theta_0)$ from $\mathscr{M}$ and a Fourier series in terms of orthonormal functions $\{u_i\}$ forming a base of the orthogonal complement of $\mathscr{M}$, say

$$m = m(\cdot, \theta_0) + f \equiv m(\cdot, \theta_0) + \sum_i a_i u_i.$$

Testing for H_0 then becomes tantamount to testing for $f \equiv 0$. Denoting by $\hat{f}$ a suitable estimate of f and letting Γ be a quadratic functional such that $\Gamma = 0$ at $f \equiv 0$, then H_0 is rejected if $\Gamma(\hat{f})$ exceeds a critical value. Of course, the performance of this test heavily depends on Γ and the choice of $\hat{f}$. As long as $\hat{f}$ includes some smoothing parameters, the local asymptotic power of the test is bigger than its size only if the alternatives approach the hypothesis at a rate $\gg n^{-1/2}$. There are two notable exceptions, however. Buckley (1991) and Eubank and Hart (1992) discussed two (related) tests (for constant regression) which are based on an integrated version of $\hat{f}$ and which allow for alternatives tending to the hypothesis at the rate $n^{-1/2}$. Though the main results of Eubank and Hart (1992) are formulated for more general test problems, the regularity assumptions required for the orthonormal system $\{u_i\}$ seem to eliminate many other examples of interest.

In the present paper we discuss in detail an approach which avoids smoothing of the data and leads to tests which are consistent on the whole of H_1. They are based on certain marked empirical processes based on residuals which are introduced below and which are shown to converge in distribution to a certain Gaussian process. The eigenvalues and eigenvectors of its covariance kernel

are analyzed in detail in Section 2. Decompositions of the underlying processes into their principal components are obtained which give new insight into the question of what kind of departure from a hypothetical model may be detected by residual-based goodness-of-fit procedures. In Section 3, we shall derive, among other things, optimal tests for H_0 versus alternatives approaching $\mathscr{M}$ from a specified direction (at the rate $n^{-1/2}$). Section 4 contains some simulation results and numerical computations for straight-line regression. Proofs of the major results are deferred to Section 5. To simplify the notation, we shall only consider real-valued X's. Most of our results have immediate extensions to the multivariate case. See the remarks in Section 6.

Now, let F denote the unknown distribution function (d.f.) of X. Put

$$I(x) = \int_{-\infty}^{x} m\, dF, \qquad x \in \mathbb{R},$$

the integrated regression function. By a measure-theoretic argument, for a given F, the function I uniquely determines m. Hence it is tempting to draw any conclusion about m from an analysis of I. Note that, by the definition of m, we have

$$I(x) = \mathbb{E}\{1_{\{X \le x\}} Y\}. \tag{1.2}$$

Hence, if (X_i, Y_i), $1 \le i \le n$, is an independent sample with the same distribution as (X, Y), the empirical analog of I becomes

$$I_n(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \le x\}} Y_i.$$

Actually, $I_n(x)$ is an unbiased estimator of $I(x)$ for each x and every $n \ge 1$:

$$\mathbb{E}[I_n(x)] = I(x). \tag{1.3}$$

Moreover, I_n is a random step function which is constant between consecutive X-order statistics $X_{1:\,n} \le \cdots \le X_{n:\,n}$ and has jumps $Y_{[1:\,n]}, \ldots, Y_{[n:\,n]}$ there. Here $Y_{[i:\,n]}$ is the Y-value associated with $X_{i:\,n}$ (the concomitant). In other words, I_n is a random element in the Skorokhod space $D(-\infty, \infty)$. Furthermore, for a fixed x, the SLLN applies to yield

$$\lim_{n\to\infty} I_n(x) = I(x) \quad \text{with probability 1.}$$

A typical uniformity argument even gives Glivenko–Cantelli convergence:

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |I_n(x) - I(x)| = 0 \quad \text{with probability 1.}$$

To obtain nondegenerate limit results, we need to consider the standardized process

$$R_n(x) = n^{-1/2} \sum_{i=1}^{n} 1_{\{X_i \le x\}} [Y_i - m(X_i)], \qquad x \in \mathbb{R},$$

where $R_n(x)$ is a sum of independent identically distributed (i.i.d.) random variables conditionally centered at X_i, $1 \le i \le n$, with variance

$$T(x) = \int_{-\infty}^{x} \mathrm{Var}(Y|X = u)F(du). \tag{1.4}$$

By the CLT,

$$R_n(x) \to \mathcal{N}(0, T(x)) \quad \text{in distribution.} \tag{1.5}$$

Theorem 1.1 provides the extension of (1.5) to a functional invariance principle, namely that R_n converges in distribution to a suitable Gaussian process as $n \to \infty$. Such a result is useful if one wants to test the simple hypothesis $m = m_0$. Just replace m by m_0 in the definition of R_n and reject the hypothesis if a proper discrepancy in R_n exceeds a critical value. For full-model checks the process R_n requires some modification. Assume, for the moment, that H_0 holds, and let θ_n be any reasonable estimator of θ_0, such as the LSE. Put

$$R_n^1(x) = n^{-1/2} \sum_{i=1}^{n} 1_{\{X_i \le x\}}[Y_i - m(X_i, \theta_n)],$$

a marked empirical process, where the marks in brackets are given by the classical residuals.

Here R_n^1 is uniquely determined by the X's and the residuals and vice versa. Though residuals have been proposed for model diagnostics for a long time, the process R_n^1 has not been investigated much before. Nor are we aware of any nonparametric principal component analysis in regression.

As already mentioned, the process R_n takes its value in the Skorokhod space $D(-\infty, \infty)$. Refer to Pollard (1984) for a thorough discussion of $D[0, \infty)$. Informally speaking, convergence in $D(-\infty, \infty)$ is equivalent to distributional convergence on compacta. This excludes the possibility of handling goodness-of-fit statistics such as $\sup_x |R_n(x)|$, where the sup extends over the whole real line. To also deal with such statistics, we continuously extend R_n to $\pm\infty$ by putting

$$R_n(-\infty) = 0 \quad \text{and} \quad R_n(\infty) = n^{-1/2} \sum_{i=1}^{n} [Y_i - m(X_i)].$$

Then R_n becomes a process in $D[-\infty, \infty]$ which, modulo a continuous transformation, is the same as the more familiar $D[0, 1]$. Actually, by incorporating a classical quantile transformation, we may write

$$R_n(x) = \bar{R}_n(F(x)), \qquad x \in [-\infty, \infty], \tag{1.6}$$

where $\bar{R}_n$ is the empirical residual-based process for an X-sample from the uniform distribution on $[0, 1]$. For our first result no condition other than the square integrability of Y will be required.

THEOREM 1.1. *Assume* $\mathbb{E}Y^2 < \infty$. *Then*

$$R_n \to R_\infty \quad \textit{in distribution in the space } D[-\infty, \infty].$$

The limit process is a centered Gaussian process with covariance function [*see* (1.4)]

$$K(s, t) = T(s \wedge t). \tag{1.7}$$

REMARK. Since T is nondecreasing and nonnegative, R_∞ admits a representation

$$R_\infty(x) = B(T(x)), \tag{1.8}$$

where B is a standard Brownian motion on the positive real line. Hence R_∞ is continuous whenever T is. This in turn holds true if F is continuous.

The representation (1.8) in connection with Theorem 1.1 and the continuous mapping theorem also yields (for continuous T)

$$\sup_{x \in \mathbb{R}} |R_n(x)| \to \sup_{0 \le t \le T(\infty)} |B(t)| \quad \text{in distribution.}$$

From this (asymptotic) critical values for the Kolmogorov–Smirnov test based on R_n are readily available. Just estimate (under $m = m_0$) $T(\infty) = \mathrm{Var}[Y - m(X)]$ by the sample variance T_n of $Y_i - m_0(X_i)$, $1 \le i \le n$, and compute the boundary crossing probabilities for a Brownian motion on $[0, T_n]$ upon observing that, due to the scaling property of B,

$$\sup_{0 \le t \le T} |B(t)| = \sqrt{T} \sup_{0 \le t \le 1} |B(t)| \quad \text{in law.}$$

The distribution of the right-hand side is, however, tabulated.

A process related to our R_n has also been studied (in a two-sample setup) by Delgado (1993). Apart from various continuity assumptions on F and the involved regression function(s), he also assumed independence of X and $Y - m(X)$. Naturally, Theorem 1.1, also covers the special model studied by Diebolt (1995), who was able to derive a Hungarian-type strong approximation result for R_n. As was shown by Theorem 1.1, the invariance principle for R_n holds true without any distributional restrictions.

For full-model checks things unfortunately become more complicated. This is due to the fact that the limit covariance even under H_0 may depend on $\mathscr{M}$ and the true but unknown θ_0. In most cases no simple transformation to a Brownian motion comparable to (1.8) is available.

For the large-sample behavior of R_n^1, some regularity assumptions on θ_n will be needed. They are summarized under Assumption 1. Assumption 2 requires some smoothness of the model $\mathscr{M}$. Note that these assumptions are not only of a technical nature but the involved quantities become an intrinsic part of the covariance functions.

ASSUMPTION 1. Under H_0, that is, $m = m(\cdot, \theta_0)$ for some unknown $\theta_0 \in \Theta$, θ_n admits an expansion

$$n^{1/2}(\theta_n - \theta_0) = n^{-1/2} \sum_{i=1}^{n} l(X_i, Y_i, \theta_0) + o_{\mathbb{P}}(1)$$

for some vector-valued function l such that:

(i) $\mathbb{E}\{l(X, Y, \theta_0)\} = 0$;

(ii) $L(\theta_0) := \mathbb{E}\{l(X, Y, \theta_0) l^T(X, Y, \theta_0)\}$ exists.

ASSUMPTION 2. (i) $m(x, \theta)$ is continuously differentiable at each θ in the interior set Θ^0 of Θ. Set

$$g(x, \theta) = \frac{\partial m(x, \theta)}{\partial \theta} \equiv (g_1(x, \theta), \ldots, g_p(x, \theta))^T.$$

(ii) There exists an F-integrable function $M(x)$ such that

$$|g_i(x, \theta)| \le M(x) \quad \text{for all } \theta \in \Theta \text{ and } 1 \le i \le p.$$

REMARK. Assumption 2 implies that the function

$$G(x, \theta) = (G_1(x, \theta), \ldots, G_p(x, \theta))^T,$$

with

$$G_i(x, \theta) = \int_{-\infty}^{x} g_i(u, \theta) F(du), \qquad 1 \le i \le p,$$

is well defined and continuous at θ for each $\theta \in \Theta^0$. Assumption 2(ii) will also be needed for the uniform convergence of empirical integrals; cf. Theorem 2 in Jennrich (1969). Also continuity of G guarantees continuity of the limit process of R_n^1.

THEOREM 1.2. *Assume* $\mathbb{E}Y^2 < \infty$ *and let Assumptions* 1 *and* 2 *be satisfied. Then, under* $m = m(\cdot, \theta_0)$, *we have, uniformly in* x,

$$R_n^1(x) = R_n(x) - n^{-1/2} \sum_{i=1}^{n} G^T(x, \theta_0) l(X_i, Y_i, \theta_0) + o_{\mathbb{P}}(1).$$

The proofs of Theorems 1.1 and 1.2 will be deferred to Section 5.

For a generalized linear model, the process R_n^1 has also been studied by Su and Wei (1991), even for multivariate X's. Their proof relies on a decomposition of X into a subvector, which is purely discrete, and another one, which is continuous. Clearly, such a decomposition need not exist, and even if it exists the independence argument for the V's introduced on page 426 of their paper is simply not true. Moreover, their simulation results seem to be based on a fixed design while the theoretical results are formulated for random X's.

COROLLARY 1.3. *Under the assumptions of Theorem* 1.2,

$$R_n^1 \to R_\infty^1 \quad \textit{in distribution in the space } D[-\infty, \infty],$$

where R_∞^1 *is a zero-mean Gaussian process with covariance function*

$$\begin{aligned} K^1(s, t) = {} & K(s, t) + G^T(s, \theta_0) L(\theta_0) G(t, \theta_0) \\ & - G^T(s, \theta_0) \mathbb{E}\left\{1_{\{X \le t\}} [Y - m(X, \theta_0)] l(X, Y, \theta_0)\right\} \\ & - G^T(t, \theta_0) \mathbb{E}\left\{1_{\{X \le s\}} [Y - m(X, \theta_0)] l(X, Y, \theta_0)\right\}. \end{aligned}$$

We already mentioned that the complicated structure of K^1 does not allow for a simple representation of R^1_∞ in terms of a process for which boundary crossing probabilities are readily available. The situation is similar to full-model checks for the underlying distribution of the X's, rather than regression. A general nonparametric approach to such problems has been first proposed in a landmark paper by Durbin (1973). Durbin, Knott and Taylor (1975) derived the principal components of the involved empirical processes. Their decompositions highlight the types of departure from a hypothetical model, which are satisfactorily detected, for example, by a Cramér–von Mises test.

It is the goal of Section 2 to derive the corresponding components, in a regression setup, for our processes R^1_n and R^1_∞, respectively, when $\mathscr{M}$ is a linear model. Here θ_n will be the least-squares estimator (LSE). Section 3 contains some power considerations, while in Section 4 several numerical computations will be provided. As a result we shall obtain a fairly good impression of what residual-based goodness-of-fit procedures may achieve in polynomial regression.

We close this section by adding some further notation. For a given parametric family $\mathscr{M} = \{m(\cdot, \theta)\colon \theta \in \Theta\}$ of (potential) regression functions, we put

$$D^2(\theta) := \mathbb{E}\left\{[Y - m(X, \theta)]^2\right\}, \qquad \theta \in \Theta.$$

By the definition of m, irrespective of whether $H_0\colon m \in \mathscr{M}$ holds true or not, we have

$$D^2(\theta) = \mathbb{E}\{[Y - m(X)]^2\} + \int [m(x) - m(x, \theta)]^2 F(dx).$$

The function D^2 is minimized at $\tilde{\theta}_0$ if and only if $\tilde{\theta}_0$ is a minimizer of

$$\tilde{D}^2(\theta)\colon = \int [m(x) - m(x, \theta)]^2 F(dx).$$

The quantity $\tilde{D}^2(\tilde{\theta}_0)$ is a means to measure the distance between the true m and the hypothetical model $\mathscr{M}$. The parameter $\tilde{\theta}_0$ is called the minimum-distance parameter for m. Clearly, if $m = m(\cdot, \theta_0) \in \mathscr{M}$, then $\theta_0 = \tilde{\theta}_0$. For identifiability reasons uniqueness of the minimizer will be postulated without further discussion. Given a sample (X_i, Y_i), $1 \le i \le n$, $\tilde{\theta}_0$ will be estimated by the minimizer (LSE) θ_n of

$$D^2_n(\theta) := n^{-1} \sum_{i=1}^{n} [Y_i - m(X_i, \theta)]^2.$$

Strong consistency and asymptotic normality of θ_n were first obtained, under H_0, by Jennrich (1969). In the next section we analyze the eigenvalues and eigenfunctions of K^1 in the case when θ_n is the LSE. From that we shall be able to derive decompositions of R^1_n and R^1_∞ into their principal components. For simplicity, we shall restrict ourselves to linear models, that is,

$$m(x, \theta) = g^T(x)\theta = g_1(x)\theta_1 + \cdots + g_p(x)\theta_p,$$

where $g = (g_1, \ldots, g_p)^T$ is a known vector-valued function.

2. Nonparametric principal component analysis of linear models. In this section we derive decompositions of R_n, R_n^1 and their limits into their principal components. In Section 3 this will enable us to study the power functions of various goodness-of-fit tests based on components. To simplify the analysis, we will restrict ourselves to the linear model. Hence the elements of $\mathscr{M}$ may be specified in the following way:

$$m(x, \theta) = g_1(x)\theta_1 + \cdots + g_p(x)\theta_p, \tag{2.1}$$

where $g_1, \ldots, g_p$ are known real-valued functions. When $g_i(x) = x^{i-1}$ for $1 \le i \le p$, $\mathscr{M}$ consists of all polynomials of degree less than or equal to $p-1$. Under (2.1), Assumption 2 is clearly satisfied with

$$g(x, \theta) \equiv g(x) = (g_1(x), \ldots, g_p(x))^T$$

and

$$G_i(x, \theta) = \int_{-\infty}^{x} g_i(u)F(du)$$

not depending on θ, provided that each of the g_i's is (square-) F-integrable. As to Assumption 1, the LSE θ_n equals

$$\theta_n = \left(X^T(n)X(n)\right)^{-1} X^T(n)Y(n).$$

Here

$$X(n) = \begin{pmatrix} g_1(X_1) & \cdots & g_p(X_1) \\ \vdots & & \vdots \\ g_1(X_n) & \cdots & g_p(X_n) \end{pmatrix}$$

is the associated (random) design matrix and

$$Y(n) = (Y_1, \ldots, Y_n)^T$$

is the output vector. Recall $\tilde{\theta}_0$, the minimum-distance parameter for m. We may conclude that

$$n^{1/2}(\theta_n - \tilde{\theta}_0) = A_n^{-1} n^{-1/2} \sum_{i=1}^{n} g(X_i)[\Delta(X_i) + \varepsilon_i]. \tag{2.2}$$

Here

$$A_n = n^{-1}X^T(n)X(n) = \left\{ n^{-1} \sum_{i=1}^{n} g_l(X_i)g_m(X_i) \right\}_{1 \le l, m \le p},$$

$$\varepsilon_i = Y_i - m(X_i), \qquad 1 \le i \le n,$$

denote the true errors and

$$\Delta(X_i) := m(X_i) - m(X_i, \tilde{\theta}_0), \qquad 1 \le i \le n,$$

measure the observed deviations from the model. Under H_0, $\tilde{\theta}_0 = \theta_0$ and $\Delta(X_i) = 0$. Setting

$$A := \left\{ \int g_l(x) g_m(x) F(dx) \right\}_{1 \le l,\, m \le p},$$

the SLLN yields $A_n \to A$ with probability 1 as $n \to \infty$. Assuming that A is regular, we obtain from (2.2), under H_0,

$$n^{1/2}(\theta_n - \theta_0) = n^{-1/2} \sum_{i=1}^{n} A^{-1} g(X_i) \varepsilon_i + o_{\mathbb{P}}(1).$$

Hence Assumption 1 holds with

$$l(x, y, \theta_0) = A^{-1} g(x)[y - m(x, \theta_0)]$$

and

$$L(\theta_0) = A^{-1} \int g(x) \sigma^2(x) g^T(x) F(dx) A^{-1}. \tag{2.3}$$

The covariance kernel K^1 becomes

$$\begin{aligned} K^1(s,t) = {} & \int_{-\infty}^{s \wedge t} \sigma^2(x) F(dx) + G^T(s) A^{-1} \int g(x) \sigma^2(x) g^T(x) F(dx) A^{-1} G(t) \\ & - G^T(s) A^{-1} \int_{-\infty}^{t} \sigma^2(x) g(x) F(dx) \\ & - G^T(t) A^{-1} \int_{-\infty}^{s} \sigma^2(x) g(x) F(dx). \end{aligned} \tag{2.4}$$

Further simplification occurs if the noise variable ε is independent of X. In this case $\sigma^2(x) = \operatorname{Var} \varepsilon \equiv \sigma^2$ does not depend on x so that (2.4) reduces to

$$K^1(s,t) = \sigma^2 \left\{ F(s \wedge t) - G^T(t) A^{-1} G(s) \right\}. \tag{2.5}$$

Also K^1 does not depend on the unknown parameter θ_0. When testing for "no effect," $p = 1$ and $g_1 \equiv 1$. In this case K^1 is the covariance of a scaled Brownian bridge with respect to time F. When $p > 1$ and $\{g_1, \ldots, g_p\}$ contains the function $\equiv 1$, then $\operatorname{Var} R^1_\infty(\infty) = K^1(\infty, \infty) = 0$, that is, $R^1_\infty(\infty) \equiv 0$. In other words, the paths of R^1_∞ are also of bridge type in this case. This follows also from the fact that $R^1_n(\infty) = 0$ for each n, which means nothing else but the well-known fact that the residual sum equals 0.

For theoretical investigations, since R^1_n also admits a representation $R^1_n = \bar{R}^1_n \circ F$ in terms of a uniform X-sample, it will suffice to consider this case. Also $\sigma^2 = 1$ w.l.o.g. $\bar{R}_\infty$ then constitutes a standard Brownian motion on $[0, 1]$. It is well known that the eigenfunctions and eigenvalues of its kernel K: $(s,t) \to s \wedge t$ are given by

$$l_j(t) = \sqrt{2} \sin\left[\left(j - \tfrac{1}{2} \right) \pi t \right]$$

and

$$\lambda_j = 1 / \left(j - \tfrac{1}{2} \right)^2 \pi^2,$$

respectively. Note that $\{l_j\}_{j\geq 1}$ is a complete orthonormal base for $L^2[0,1]$. Denote by

$$\xi_{nj} := \lambda_j^{-1/2} \int_0^1 \bar{R}_n(t) l_j(t)\, dt,$$

$$\xi_j := \lambda_j^{-1/2} \int_0^1 \bar{R}_\infty(t) l_j(t)\, dt,$$

$$\eta_{nj} := \lambda_j^{-1/2} \int_0^1 \bar{R}_n^1(t) l_j(t)\, dt$$

and

$$\eta_j := \lambda_j^{-1/2} \int_0^1 \bar{R}_\infty^1(t) l_j(t)\, dt$$

the associated (standardized) Fourier coefficients. Put

$$h_j(t) = \sqrt{2}\cos\big[\big(j - \tfrac{1}{2}\big)\pi t\big], \qquad j \geq 1.$$

Integrate ξ_{nj} and η_{nj} by parts to obtain

$$\xi_{nj} = n^{-1/2} \sum_{i=1}^n \varepsilon_i h_j(X_i)$$

and

$$\eta_{nj} = n^{-1/2} \sum_{i=1}^n [Y_i - \theta_n^T g(X_i)] h_j(X_i).$$

From Theorems 1.1 and 1.2 and the continuous mapping theorem, finitely many of the ξ_{nj}'s converge in distribution to the corresponding ξ_j's; similarly for the η's. Furthermore, we have in $L^2[0,1]$ and hence in distribution

$$\bar{R}_n = \sum_{j=1}^\infty \lambda_j^{1/2} \xi_{nj} l_j$$

and

$$\bar{R}_\infty = \sum_{j=1}^\infty \lambda_j^{1/2} \xi_j l_j.$$

Since the l_j's and λ_j's are the eigenfunctions and eigenvalues of K, that is,

$$\int_0^1 K(s,t) l_j(t)\, dt = \lambda_j l_j(s),$$

$\xi_1, \xi_2, \ldots$ are independent $\mathcal{N}(0,1)$, while, for a finite $n \geq 1$, the variables $\xi_{n1}, \xi_{n2}, \ldots$ have variance 1 and are at least uncorrelated. Of course, we also have, for example,

$$\bar{R}_\infty^1 = \sum_{j=1}^\infty \lambda_j^{1/2} \eta_j l_j. \tag{2.6}$$

Note, however, that, in general, K^1 has eigenfunctions and eigenvalues different from l_j and λ_j, so that the η's are no longer independent. In other words, (2.6) does not constitute the desired expansion of $\bar{R}^1_\infty$ into its principal components. Since the determination of the eigenfunctions of K^1 is not obvious at all, our first goal will be to find a representation of $\bar{R}^1_\infty$ of the following form:

$$\bar{R}^1_\infty = \sum_{j=1}^{\infty} \xi_j D_j \tag{2.7}$$

for a suitable set of functions $\{D_j\}_{j\geq 1}$. Hence the right-hand side of (2.7) is a series with independent summands, but the (deterministic) D_j's need not be orthonormal in $L^2[0, 1]$. To define the D_j's, recall

$$h_j(t) = \sqrt{2}\cos\left[\left(j - \tfrac{1}{2}\right)\pi t\right]$$

and put

$$\langle g_k, h_j\rangle := \int_0^1 g_k(x)h_j(x)\,dx, \qquad j \geq 1.$$

Write

$$\delta_j^T := (\langle g_1, h_j\rangle, \ldots, \langle g_p, h_j\rangle)$$

and put, for $0 \leq t \leq 1$,

$$D_j(t) := \lambda_j^{1/2} l_j(t) - G^T(t)A^{-1}\delta_j. \tag{2.8}$$

Theorem 2.1 gives the desired expansion of $\bar{R}^1_\infty$. It may be used to generate sample paths of $\bar{R}^1_\infty$ via Monte Carlo and thus to approximate, for example, some selected boundary crossing probabilities for $\bar{R}^1_\infty$. It will also serve to characterize the Fourier coefficients $\eta_1, \eta_2, \ldots$ (Theorem 2.2).

THEOREM 2.1. *Let D_j be defined as in* (2.8). *Then, in distribution, we have*

$$\bar{R}^1_\infty = \sum_{j=1}^{\infty} \xi_j D_j = \bar{R}_\infty - \left[A^{-1}\sum_{j=1}^{\infty} \xi_j \delta_j\right]^T G.$$

Next note that along with the l_j's the h_j's also form a complete orthonormal base of $L^2[0,\ 1]$. Hence

$$g_k = \sum_{j=1}^{\infty} \langle g_k, h_j\rangle h_j. \tag{2.9}$$

Put

$$\Delta = \begin{pmatrix} \delta_1^T \\ \delta_2^T \\ \vdots \end{pmatrix} = \begin{pmatrix} \langle g_1, h_1\rangle & \cdots & \langle g_p, h_1\rangle \\ \langle g_1, h_2\rangle & \cdots & \langle g_p, h_2\rangle \\ \vdots & & \vdots \end{pmatrix}.$$

From Parseval's identity (cf. Hewitt and Stromberg (1965), page 246) we immediately obtain

$$A = \Delta^T \Delta = \sum_{j=1}^{\infty} \delta_j \delta_j^T. \tag{2.10}$$

The functions l_j and h_j are related via the trivial identities

$$l_j' = \lambda_j^{-1/2} h_j, \qquad h_j' = -\lambda_j^{-1/2} l_j. \tag{2.11}$$

Integrate $\int g_k h_j$ by parts and apply (2.11) to obtain

$$\langle g_k, h_j \rangle = \lambda_j^{-1/2} \int_0^1 G_k(x) l_j(x)\, dx. \tag{2.12}$$

Introduce the infinite-dimensional vectors

$$\xi_n = (\xi_{nj})_{j \geq 1}, \qquad \xi = (\xi_j)_{j \geq 1},$$
$$\eta_n = (\eta_{nj})_{j \geq 1}, \qquad \eta = (\eta_j)_{j \geq 1},$$

and put

$$l(t) = (l_j(t))_{j \geq 1}, \qquad 0 \leq t \leq 1.$$

The distributions of ξ_n and so on are uniquely determined by their finite-dimensional distributions. Hence, to verify that two vectors have the same distribution, it would suffice to compare finitely many of their components. Moreover, all vectors have components which are absolutely summable. Denote by I the infinite-dimensional identity matrix and, for each real $r > 0$, let

$$\lambda^r := \begin{pmatrix} \lambda_1^r & & 0 \\ & \lambda_2^r & \\ 0 & & \ddots \end{pmatrix}.$$

A crucial role will be played by the matrix

$$M := I - \Delta(\Delta^T \Delta)^{-1} \Delta^T = I - \Delta A^{-1} \Delta^T.$$

In Theorem 2.2 we show that, in distribution, $\eta = M\xi$. Conclude that $\Delta^T \eta = 0$. The situation is similar to (4.14) in Durbin, Knott and Taylor (1975), who observed that, in a different setup, estimation of the nuisance parameters results in a projection of the Fourier coefficients on the orthogonal complement of the columns of Δ.

THEOREM 2.2. *The matrix M is symmetric and idempotent. Furthermore,*

$$\eta = M\xi \sim \mathcal{N}_\infty(0, M) \quad \textit{in distribution}.$$

In Theorem 2.3 we give a useful characterization of the eigenvalues and eigenfunctions of K^1 in terms of M and $\lambda^{1/2}$. This result parallels the discussion (in a different context) on page 223 of Durbin, Knott and Taylor (1975).

THEOREM 2.3. *There exists an orthonormal matrix N and a diagonal matrix*

$$\mu = \begin{pmatrix} \mu_1 & & 0 \\ & \mu_2 & \\ 0 & & \ddots \end{pmatrix}$$

such that:

(i) $\lambda^{1/2} M \lambda^{1/2} = N \mu N^T$;
(ii) $\mu_1 > \mu_2 > \cdots > 0$ *are the eigenvalues of* K^1;
(iii) $N^T l(t)$ *is the vector of eigenfunctions of* K^1.

Conclude from Theorem 2.3 that the eigenfunctions of K^1 are certain linear combinations of the l_j's. Similarly, the corresponding Fourier coefficients are linear combinations of the η's. Actually, write $N = (n_{ij})_{i,j}$ for a moment. From (iii) the jth eigenfunction of K^1 is given as

$$m_j(t) = \sum_{i=1}^{\infty} n_{ij} l_i(t),$$

while

$$\int_0^1 \bar{R}^1_\infty(t) m_j(t)\, dt = \sum_{i=1}^{\infty} n_{ij} \lambda_i^{1/2} \eta_i. \tag{2.13}$$

Finally,

$$\int_0^1 \left[\bar{R}^1_\infty(t)\right]^2 dt = \sum_{j=1}^{\infty} \left[\sum_{i=1}^{\infty} n_{ij} \lambda_i^{1/2} \eta_i \right]^2$$

constitutes the decomposition of the corresponding Cramér–von Mises functional into its principal components. Similarly, for $\bar{R}^1_n$,

$$\begin{aligned} \int \bar{R}^1_n(t) m_j(t)\, dt &= \sum_{i=1}^{\infty} n_{ij} \lambda_i^{1/2} \eta_{ni} \\ &= n^{-1/2} \sum_{k=1}^{n} \left[Y_k - \theta_n^T g(X_k) \right] \left\{ \sum_{i=1}^{\infty} n_{ij} \lambda_i^{1/2} h_i(X_k) \right\}. \end{aligned}$$

So far we have assumed that the X's come from the uniform distribution on $[0, 1]$. For the theoretical investigations this may be justified by the feasibility of representing a general R^1_n in terms of $\bar{R}^1_n$ properly transformed in time. Now, in order to apply the Cramér–von Mises test to a given set of data (X_i, Y_i), $1 \le i \le n$, for example, we have to proceed as follows:

1. Estimate the noise variance by the residual sum of squares

$$\sigma_n^2 = n^{-1} \sum_{i=1}^{n} [Y_i - \theta_n^T g(X_i)]^2$$

or any other consistent (model independent) estimate of σ^2; see, for example, Rice (1984).

2. Let

$$F_n(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \le x\}}$$

denote the empirical distribution function of $X_1, \ldots, X_n$, and put

$$W_n^2 = \sigma_n^{-2} \int_{-\infty}^{\infty} [R_n^1(x)]^2 F_n(dx).$$

For a continuous F, Corollary 1.3 and the continuous mapping theorem immediately yield, under H_0,

$$W_n^2 \to W_\infty^2 \equiv \int_0^1 [\bar{R}_\infty^1(t)]^2 \, dt \quad \text{in distribution,}$$

where, as before, $\bar{R}_\infty^1$ denotes a centered Gaussian process with covariance kernel (2.5), $\sigma^2 = 1$ and F is the uniform distribution on $0 \le t \le 1$.

3. For computational purposes, let $X_{1:n} \le \cdots \le X_{n:n}$ denote the ordered X-sample, and let $Y_{[i:n]}$ be the Y-value associated with $X_{i:n}$. Finally, write

$$e_{[i:n]} = Y_{[i:n]} - \theta_n^T g(X_{i:n})$$

for the ith residual concomitant of $X_{i:n}$. Obviously,

$$W_n^2 = \sigma_n^{-2} n^{-2} \sum_{k=1}^{n} \left\{ \sum_{i=1}^{k} e_{[i:n]} \right\}^2. \tag{2.14}$$

Its jth component may be approximated by

$$\begin{aligned} \rho_{nj} &\equiv \sigma_n^{-1} \int_{-\infty}^{\infty} R_n^1(x) m_j(F_n(x)) F_n(dx) \\ &= \sigma_n^{-1} n^{-1} \sum_{k=1}^{n} R_n^1(X_k) m_j(\text{Rank } X_k/n) \qquad (2.15) \\ &= \sigma_n^{-1} n^{-1} \sum_{k=1}^{n} m_j(k/n) \left\{ \sum_{i=1}^{k} e_{[i:n]} \right\}. \end{aligned}$$

The terms in brackets constitute the cusums of the residual concomitants. Equation (2.14) also appears on page 258 of Buckley (1991) as a statistic designed for testing for constant regression (i.e., $p = 1$ and $g_1 \equiv 1$), assuming that the errors are normal. That paper also contains an interesting historical review of the earlier literature on residual cusums and their applications in model diagnostics.

As we will see in the next section, the components ρ_{nj} and their asymptotic counterparts

$$\rho_j = \int_0^1 \bar{R}_\infty^1(u) m_j(u) \, du$$

not only play an important role in the representation and analysis of the Cramér–von Mises statistic, but also constitute an intrinsic tool for constructing new smooth and directional tests for H_0 versus specific alternatives.

Finally, we briefly discuss the problem of how to actually compute the components of $\bar{R}^1_\infty$. From Theorem 2.3 we need to diagonalize the infinite matrix $\lambda^{1/2}M\lambda^{1/2}$. In practice, this will be possible only in exceptional cases. Rather, following a suggestion of Durbin, Knott and Taylor (1975), we propose to approximate the first q components by first computing the eigenvectors and eigenvalues of $\varepsilon_q^{1/2}M_q\varepsilon_q^{1/2}$, where M_q is the $q \times q$ matrix

$$M_q = I_q - \Delta_q(\Delta_q^T\Delta_q)^{-1}\Delta_q^T$$

and

$$\Delta_q = \begin{pmatrix} \delta_1^T \\ \vdots \\ \delta_q^T \end{pmatrix}, \qquad \varepsilon_q^{1/2} = \begin{pmatrix} \lambda_1^{1/2} & & 0 \\ & \ddots & \\ 0 & & \lambda_q^{1/2} \end{pmatrix}.$$

Put

$$\tilde{M} = \begin{pmatrix} M_q & 0 \\ 0 & I \end{pmatrix}$$

and observe that

$$\lambda^{1/2}\tilde{M}\lambda^{1/2} = \begin{pmatrix} \varepsilon_q^{1/2}M_q\varepsilon_q^{1/2} & 0 \\ 0 & \lambda_{\infty-q} \end{pmatrix},$$

with

$$\lambda_{\infty-q}^{1/2} = \begin{pmatrix} \lambda_{q+1}^{1/2} & & 0 \\ & \lambda_{q+2}^{1/2} & \\ 0 & & \ddots \end{pmatrix}.$$

Suppose that $\varepsilon_q^{1/2}M_q\varepsilon_q^{1/2}$ has eigenvalues $\tilde{\mu}_1, \ldots, \tilde{\mu}_q$ and eigenvectors $\tilde{n}_i = (n_{1i}, \ldots, n_{qi})^T$, $1 \le i \le q$. Then $\lambda^{1/2}\tilde{M}\lambda^{1/2}$ has eigenvalues $\tilde{\mu}_1, \ldots, \tilde{\mu}_q$, λ_{q+1}, $\lambda_{q+2}, \ldots$ and eigenvectors

$$n_{i0} = (\tilde{n}_i, 0, 0, \ldots)^T \quad \text{for } 1 \le i \le q$$

and

$$n_{i0} = e_i\text{—the } i\text{th unit vector—for } q < i.$$

In Section 4 we will use these values to approximate the true eigenvalues and eigenvectors of $\lambda^{1/2}M\lambda^{1/2}$. The distribution of

$$W_\infty^2 = \sum_{j=1}^{\infty} \mu_j Z_j^2 = \sum_{j=1}^{\infty} \rho_j^2 \tag{2.16}$$

will be approximated by that of

$$W^2_{\infty q} := \sum_{i=1}^{q} \tilde{\mu}_j Z_j^2 + \lambda \chi_k^2,$$

where the Z_j's are independent standard normal, χ_k^2 is a suitable χ^2-variable with k degrees of freedom and λ is such that $W^2_{\infty q}$ has the same mean and variance as W^2_∞. The distribution of $W^2_{\infty q}$ will then be computed by Imhof's (1961) method.

3. Cramér–von Mises, smooth and directional tests. In this section we shall employ principal component analysis in order to investigate the power of various tests based on R_n^1. To this end, suppose that, for some function r and some $\theta_0 \in \Theta$,

$$m(x) \equiv m_n(x) = m(x, \theta_0) + n^{-1/2} r(x). \tag{3.1}$$

Up to square integrability no further assumptions on r will be needed. Recall $\tilde{\theta}_0$ from Section 1, the minimum-distance parameter for m. Under the sequence of alternatives (3.1), $\tilde{\theta}_0 = \tilde{\theta}_{0n}$, of course, depends on n. Check that θ_0 and $\tilde{\theta}_{0n}$ satisfy the equation

$$n^{1/2}(\theta_0 - \tilde{\theta}_{0n}) = -A^{-1} \begin{pmatrix} \int r g_1 \, dF \\ \vdots \\ \int r g_p \, dF \end{pmatrix}. \tag{3.2}$$

It is easy to see from (3.2) that

$$\begin{aligned} R_n^1(x) = R_n(x) &- n^{-1/2} \sum_{i=1}^{n} G^T(x) l(X_i, Y_i, \theta_0) \\ &+ \int_{-\infty}^{x} r(u) F(du) - G^T(x) A^{-1} \int_{-\infty}^{\infty} g(u) r(u) F(du) + o_{\mathbb{P}}(1), \end{aligned} \tag{3.3}$$

where the right-hand side of (3.3) is computed under θ_0. In other words, a change from H_0 to (3.1) results in a nonrandom shift function. In particular, R_n^1 has the same limit covariance function as before. Therefore the eigenfunctions and eigenvalues remain the same, while the components of R_∞^1 become noncentered normal random functions. First, it will be worthwhile studying the underlying structure in greater detail from a geometric point of view. For this, introduce the operator

$$\Pi\colon s \to s - G^T A^{-1} \begin{pmatrix} \int g_1 \dfrac{ds}{dF} \, dF \\ \vdots \\ \int g_p \dfrac{ds}{dF} \, dF \end{pmatrix}$$

acting on all functions s admitting a square-integrable F-density. Needless to say that when $s = \int^{\bullet} r \, dF$, $\Pi(s)$ results in the shift function appearing in (3.3).

LEMMA 3.1. *The operator* Π *constitutes the projection onto the orthogonal complement of the functions* $\int^{\bullet} g_i \, dF$, $1 \le i \le p$.

Actually, check that $\Pi^2 = \Pi$ and that each of the above p functions is mapped into 0. Now, write

$$s_0 = \int^{\bullet} r \, dF \quad \text{and put} \quad \pi_0 = \Pi s_0.$$

If r is orthogonal to $g_1, \ldots, g_p$, then $\Pi s_0 = s_0$. In the general case, π_0 is closer to the subspace spanned by the components of G^T than s_0. This is nicely reflected by the Fourier representation of π_0. For the sake of simplicity, assume that F is the uniform distribution on the unit interval. In terms of the set of eigenfunctions $\{m_j\}_{j \ge 1}$, one has

$$\begin{aligned} \langle \pi_0, m_j \rangle &= \int_0^1 \pi_0(t) m_j(t) \, dt = \sum_{i=1}^{\infty} n_{ij} \int_0^1 \pi_0(t) l_i(t) \, dt \\ &= \sum_{i=1}^{\infty} n_{ij} \lambda_i^{1/2} \int_0^1 h_i(t) \pi_0'(t) \, dt, \end{aligned}$$

where the last equality followed from integration by parts. Hence

$$\begin{aligned} \langle \pi_0, m_j \rangle &= \sum_{i=1}^{\infty} n_{ij} \lambda_i^{1/2} \int_0^1 h_i(t) \left[r(t) - g^T(t) A^{-1} \begin{pmatrix} \langle g_1, r \rangle \\ \vdots \\ \langle g_p, r \rangle \end{pmatrix} \right] dt \\ &= \sum_{i=1}^{\infty} n_{ij} \lambda_i^{1/2} \beta_i. \end{aligned} \tag{3.4}$$

In summary, under the sequence of alternatives (3.1), the limit process $\bar{R}_{\infty}^1$ has the representation

$$\bar{R}_{\infty}^1 = \sum_{j=1}^{\infty} m_j \sum_{i=1}^{\infty} n_{ij} \lambda_i^{1/2} (\eta_i + \beta_i), \tag{3.5}$$

where the η's are the same as in (2.13). Parallel to Lemma 3.1, the function $[\ldots]$ in the integral (3.4) may also be viewed as a projection, namely as that of r onto the orthogonal complement of the subspace spanned by $g_1, \ldots, g_p$. Under orthogonality, we again obtain r. Note, however, that in the general case $\langle , \rangle$ is computed w.r.t. the unknown F, so that, for given $g_1, \ldots, g_p$ and r, orthogonality will be the exception rather than the rule.

The decomposition (3.5) implies that the local asymptotic power of the Cramér–von Mises test, for example, exceeds the size of the test if at least one of the β's does not vanish. It will be bounded away from the size uniformly in r, if the r's under consideration are such that their projections stay away from $\operatorname{span}\{g_1, \ldots, g_p\}$ uniformly in r. Finally, for $p = 1$, $[\ldots]$ in (3.4)

equals the function which comes up in the first step of the Gram–Schmidt orthogonalization algorithm.

We now want to discuss the local asymptotic power of the Cramér–von Mises test in a more quantitative fashion. For this recall from (3.5) that $\bar{R}^1_\infty$ has, under the sequence (3.1), the representation

$$\sum_{j=1}^{\infty} m_j \langle \bar{R}^1_\infty + \pi_0, m_j \rangle, \tag{3.6}$$

where the $\bar{R}^1_\infty$ in (3.6) is computed under H_0. In other words, the distribution of $\bar{R}^1_\infty$ under (3.1) equals that of $\bar{R}^1_\infty$ under H_0 shifted by the function

$$\sum_{j=1}^{\infty} m_j \langle \pi_0, m_j \rangle = \pi_0.$$

Denote by ν_0 and ν_1 the distribution of $\bar{R}^1_\infty$ under H_0 and (3.1), respectively. A fundamental result by Grenander (1952) yields that (in our notation), under the mild technical assumption

$$\sum_{j=1}^{\infty} \mu_j^{-1} \langle \pi_0, m_j \rangle^2 < \infty,$$

ν_1 is absolutely continuous w.r.t. ν_0. From a statistical point of view, this is the more interesting case since, informally speaking, the convergence entails that H_0 and H_1 are very close. Actually, if the series diverges, ν_0 and ν_1 are orthogonal so that the asymptotic power is 1.

Now, the Radon-Nikodym derivative equals

$$g \to \exp\left\{ - \int_0^1 A(t) \left[g(t) - \frac{\pi_0(t)}{2} \right] dt \right\}, \tag{3.7}$$

where the function A is defined as

$$A(t) = - \sum_{j=1}^{\infty} \langle \pi_0, m_j \rangle \mu_j^{-1} m_j(t).$$

Moreover, replacing g by $\bar{R}^1_\infty$, the exponent $\{\ldots\}$ has a normal distribution with

$$\mathbb{E}_0\{\ldots\} = \tfrac{1}{2} \int_0^1 A(t) \pi_0(t)\, dt = \tfrac{-1}{2} \sum_{j=1}^{\infty} \mu_j^{-1} \langle \pi_0, m_j \rangle^2 \tag{3.8}$$

and

$$\mathrm{Var}_0\{\ldots\} = \sum_{j=1}^{\infty} \mu_j^{-1} \langle \pi_0, m_j \rangle^2. \tag{3.9}$$

The situation here is similar to the one considered by Neuhaus (1976) in his Lemma 2.5, where he investigated the local asymptotic power of the classical Cramér–von Mises test. More precisely, fix the direction r, as in (3.1), but let

the distance to $\mathscr{M}$ tend to 0, that is, consider ar rather than r, with $a \to 0$. Assume $\sigma^2 = 1$ first. Write $\|\cdot\|$ for the L_2-norm and introduce

$$\beta(c, r, a) := \mathbb{P}_{ar}(\|\bar{R}^1_\infty\|^2 > c),$$

where the probability on the right-hand side is computed under the shift ar. Here $\beta(c, r, a)$ is the local asymptotic power of the Cramér–von Mises test with critical value c, as we approach $\mathscr{M}$ from the direction $n^{-1/2}ar$. Let $\pi_0 = \pi_0^r$ be as before. Trivially,

$$\pi_0^{ar} = a\pi_0^r$$

and, similarly, in an obvious notation,

$$A^{ar} = aA^r = aA.$$

Conclude from (3.7) that

$$\beta(c, r, a) = \int_{\{\|\bar{R}^1_\infty\|^2 > c\}} \exp\left\{-aN + \frac{a^2}{2}\langle A, \pi_0\rangle\right\} d\mathbb{P}_0,$$

where $N = \langle A, \bar{R}^1_\infty\rangle$ is a zero-mean normal variable with variance (3.9). Differentiate the right-hand side w.r.t. a and use a symmetry argument to show that the first derivative vanishes at zero to finally get

$$\begin{aligned}\beta(c, r, a) &= \beta(c, r, 0) + \frac{a^2}{2}\left[\langle A, \pi_0\rangle\beta(c, r, 0) + \int_{\{\|\bar{R}^1_\infty\|^2 > c\}} N^2\, d\mathbb{P}_0\right] + o(a^2)\\ &= \beta(c, r, 0) + \frac{a^2}{2}\left[-\langle A, \pi_0\rangle(1 - \beta(c, r, 0)) - \int_{\{\|\bar{R}^1_\infty\|^2 \le c\}} N^2\, d\mathbb{P}_0\right]\\ &\quad + o(a^2).\end{aligned}$$

The term in brackets constitutes the curvature of the asymptotic power function. Let $c = c_\alpha$ be chosen such that

$$\beta(c, r, 0) = \alpha,$$

the size of the test. Expand N into its Fourier series,

$$N = \langle A, \bar{R}^1_\infty\rangle = -\sum_{j=1}^{\infty}\langle \pi_0, m_j\rangle\mu_j^{-1/2}Z_j, \tag{3.10}$$

where, as in (2.16), $Z_1, Z_2, \ldots$ are i.i.d. standard normal. Use a symmetry argument to show that if (3.10) is plugged into $\int_{\{\ \}} N^2$, the integrals $\int_{\{\ \}} Z_i Z_j$ vanish for $i \neq j$. Consequently, from (3.8) and (3.10), the power function admits an expansion

$$\beta(c_\alpha, r, a) = \alpha + \frac{a^2}{2}\sum_{j=1}^{\infty}\mu_j^{-1}\langle \pi_0, m_j\rangle^2\left[1 - \alpha - \int_{\{\|\bar{R}^1_\infty\|^2 \le c_\alpha\}} Z_j^2\, d\mathbb{P}_0\right] + o(a^2).$$

Since in the case of an arbitrary (unknown) σ^2 the test is based on R^1_n/σ_n rather than R^1_n, the shift ar needs to be replaced by ar/σ in the general case.

In the above expansion of β, a^2 thus becomes a^2/σ^2 which nicely features the loss of power if the noise variance increases. The Z_j-integral may be calculated numerically upon utilizing techniques elaborated, for example, in Johnson and Kotz (1970), Chapter 6, Sections 5 and 6.

To continue, similar to the ordinary Cramér–von Mises test, our W_n^2 also leads to a test which is consistent against all alternatives, but, on the other hand, is not able to detect specific alternatives one might have in mind. In particular, tests for higher-frequency alternatives to H_0 will only have a moderate if not low power. This is immediately seen from the Fourier expansions of W_∞^2 and W_n^2, since possible high-frequency deviations from H_0 are heavily downweighted by μ_j already for $j \geq 3$. In such a situation a reweighting of the components ρ_{nj} in (2.15) leads to tests which, in spirit, are similar to Neyman's (1937) smooth test for densities. In our case we may fix some $m \geq 1$ and put

$$W_{nm}^2 := \sum_{j=1}^{m} \mu_j^{-1} \rho_{nj}^2.$$

Then, in distribution,

$$W_{nm}^2 \to \chi_m^2(0) \quad \text{under } H_0$$

and

$$W_{nm}^2 \to \chi_m^2\left(\sum_{j=1}^{m} \frac{\langle \pi_0, m_j \rangle^2}{\sigma^2} \right) \quad \text{under (3.1).}$$

The local asymptotic power of these tests may be studied along the same lines as in Milbrodt and Strasser (1990), page 7. One further possibility would be to let $m \to \infty$ as $n \to \infty$, as was done, in a different setup, by Eubank and LaRiccia (1992). Hence, as the sample size increases, deviations with high frequencies will be discovered eventually.

As Eubank and LaRiccia (1992), page 2072, nicely pointed out, smooth tests represent a compromise between directional and omnibus tests. To derive such a directional test in the context of regression, with a composite H_0 and F, σ^2 unspecified, of course, recall (3.1) again. To motivate the procedure, again Grenander's (1952) formula (3.7) will be of great importance. Now, the representation (3.3) suggests that each test of H_0 versus (3.1), which is based on R_n^1/σ_n, asymptotically becomes one of testing the simple hypothesis $\tilde{H}_0$: No shift versus $\tilde{H}_1$: Shift = π_0 in the exponential model defined by (3.7). By the Neyman–Pearson lemma, the optimal test rejects $\tilde{H}_0$ in favor of $\tilde{H}_1$ if and only if

$$-\int_0^1 A(t) \left[\bar{R}_\infty^1(t) - \frac{\pi_0(t)}{2} \right] dt \geq c_1$$

or, equivalently,

$$\sum_{j=1}^{\infty} \frac{\langle \pi_0, m_j \rangle \langle \bar{R}_\infty^1, m_j \rangle}{\mu_j} \geq c_2. \tag{3.11}$$

Under H_0, the last series constitutes a zero-mean normal random variable with variance

$$\gamma^2 \equiv \sum_{j=1}^{\infty} \frac{\langle \pi_0, m_j \rangle^2}{\mu_j}.$$

For finite sample size the approximate Neyman–Pearson test for H_0 versus (3.1) is thus given by the critical region

$$\sum_{j=1}^{m} \frac{\langle \pi_0, m_j \rangle^* \rho_{nj}}{\mu_j} \geq c, \tag{3.12}$$

where ρ_{nj} is defined in (2.15), m is a proper truncation point and

$$\langle \pi_0, m_j \rangle^* = \int \pi_0^*(x) m_j(F_n(x)) F_n(dx)$$

in which π_0^* results from π_0 by replacing the unknown F by F_n. Again some asymptotic theory may be elaborated for situations, in which $m \to \infty$ as $n \to \infty$. In the ordinary setup, that is, when one is interested in the distribution of the X's rather than regression, tests similar to (3.11) and (3.12) have been investigated by Schoenfeld (1977, 1980). The notable fact about our approach is that the optimality as stated, for example, in Theorem 4 of Schoenfeld (1977), may be derived in a very elegant way from the "functional" Neyman–Pearson lemma, upon utilizing (3.7).

4. A numerical example. In this section we let $\mathscr{M} = \{m(x, \theta)\colon \theta \in \Theta\}$ with $m(x, \theta) = \theta x$ denoting the one-parameter linear model on the straight line. Here F will be the uniform distribution on the unit interval. So, $A = 1/3$ and $G(x) = x^2/2$. The function r in (3.1) will be $r(x) = x^2$ (or a multiple of it). The shift function in (3.3) then becomes

$$\pi_0(x) := \frac{x^3}{3} - \frac{3}{8}x^2, \qquad 0 \leq x \leq 1.$$

Note that π_0 is always less than or equal to 0. Under the sequence of alternatives (3.1), the limit Cramér–von Mises statistic has the orthonormal decomposition

$$W_\infty^2 = \int_0^1 \left[\bar{R}_\infty^1(x)\right]^2 dx = \sum_{j=1}^{\infty} \mu_j Z_j^2$$

(in distribution), where the Z_j are independent $\mathscr{N}(\alpha_j, 1)$ random variables with

$$\alpha_j = \int_0^1 \pi_0(x) m_j(x)\, dx.$$

To compute the null distribution of W_∞^2, observe that for δ_j we obtain

$$\delta_j = \langle g, h_j \rangle = \sqrt{2} \int_0^1 x \cos\left[\left(j - \frac{1}{2}\right)\pi x\right] dx$$
$$= \frac{\sqrt{2}}{(j-1/2)\pi}\left[(-1)^{j+1} - \frac{1}{(j-1/2)\pi}\right].$$

The matrix M_q is given as

$$m_{ij} = \begin{cases} -3\delta_i\delta_j, & \text{for } 1 \le i \ne j \le q, \\ 1 - 3\delta_i^2, & \text{for } 1 \le i = j \le q. \end{cases}$$

For $W_{\infty q}^2$, we have to solve the equations

$$\mathbb{E}W_{\infty q}^2 = \mathbb{E}W_\infty^2 \tag{4.1}$$

and

$$\operatorname{Var} W_{\infty q}^2 = \operatorname{Var} W_\infty^2. \tag{4.2}$$

Since

$$K^1(s,t) = \min(s,t) - \tfrac{3}{4}s^2t^2$$

and

$$\mathbb{E}W_\infty^2 = \int_0^1 K^1(t,t)dt = \tfrac{7}{20},$$

$$\operatorname{Var} W_\infty^2 = \int_0^1 \int_0^1 [K^1(s,t)]^2 \, ds\, dt = \tfrac{5}{42} + \tfrac{9}{200},$$

(4.1) and (4.2) become

$$\tfrac{7}{20} = \sum_{j=1}^q \tilde{\mu}_j + \lambda k$$

and

$$\tfrac{1}{2}\left[\tfrac{5}{42} + \tfrac{9}{200}\right] = \sum_{j=1}^q \tilde{\mu}_j^2 + \lambda^2 k.$$

From this

$$\lambda = \frac{(1/2)[5/42 + 9/200] - \sum_{j=1}^q \tilde{\mu}_j^2}{7/20 - \sum_{j=1}^q \tilde{\mu}_j}$$

and

$$k = \frac{[7/20 - \sum_{j=1}^q \tilde{\mu}_j]^2}{(1/2)[5/42 + 9/200] - \sum_{j=1}^q \tilde{\mu}_j^2}.$$

TABLE 1

x	$\mathbb{P}(W^2_{\infty 50} > x)$	$\mathbb{P}(W^2_{\infty 100} > x)$
0.1	0.7318	0.7350
0.2	0.4975	0.4992
0.4	0.2791	0.2798
0.6	0.1708	0.1712
0.8	0.1083	0.1085
1.0	0.0701	0.0702
1.2	0.0460	0.0461
1.4	0.0304	0.0305
1.6	0.0203	0.0203
1.8	0.0136	0.0136
2.0	0.0092	0.0092
2.2	0.0062	0.0062
2.4	0.0042	0.0042

Table 1 presents, for $q = 50$ and $q = 100$, some selected values of $\mathbb{P}(W^2_{\infty q} > x)$.

Recall that the null distributions of W^2_∞ and $W^2_{\infty q}$ do not depend on F for a continuous F.

Table 2 contains, for $\alpha = 0.05$ and $\alpha = 0.01$, the fractiles z_α of $\mathbb{P}(W^2_{\infty q} > x)$.

In the following we report on a small simulation study, in which independent data (X_i, Y_i) were generated as follows:

$$Y_i = 5X_i + aX_i^2 + \varepsilon_i, \qquad 1 \le i \le n. \tag{4.3}$$

The variable X_i is uniformly distributed on the unit interval and ε_i is $\mathcal{N}(0, \sigma^2)$ and independent of X_i. For $a = 0$ the regression function belongs to the set $\mathcal{M}$ of all linear functions on $[0, 1]$ (through the origin). The statistic W^2_n has been computed, for sample size $n = 400$, to test the hypothesis H_0: $m \in \mathcal{M}$. The significance levels were $1 - \alpha = 0.95$ and 0.99, and H_0 was rejected whenever $W^2_n \ge z_\alpha$ with z_α from Table 2, with $q = 100$. This was repeated 1000 times, for each of the sample situations mentioned in Table 3. Then the actual percentage of times H_0 was rejected was computed and compared with the nominal level.

We see that, under H_0 (i.e., $a = 0$), the actual percentage of times H_0 has been rejected is pretty close to the nominal level. Alternatively, there is a rapid increase in power as a increases, that is, as the true m deviates from the hypothetical $\mathcal{M}$. This effect becomes less obvious for large σ^2. From

TABLE 2

$\mathbb{P}(W^2_{\infty q} > z_\alpha) = \alpha$	$z_\alpha (q = 50)$	$z_\alpha (q = 100)$
$\alpha = 0.05$	1.159915	1.160904
$\alpha = 0.01$	1.954048	1.955040

TABLE 3
Percentage of times H_0 was rejected

σ^2	a	$\alpha = 0.05$	$\alpha = 0.01$
1	0	5.1%	1.2%
1	1	40.8%	17.4%
1	5	100.0%	100.0%
2	0	4.5%	1.0%
2	1	22.8%	4.4%
2	5	100.0%	97.8%
3	0	5.5%	0.9%
3	1	16.7%	2.2%
3	5	89.3%	67.1%

a statistical point of view this is not unexpected since in this case it is more difficult to decide whether irregularities in the data are caused either by noise or by a deviation from $\mathscr{M}$. See also the discussion of the power function β in the last section.

Similar effects occur if, rather than (4.3), the true m is not linear but, for example, convex and hence "supported" by a line. Also note that in our simulation study the X's were uniformly distributed on the unit interval. Here the deviation between θx and $\theta x + ax^2$ is not that obvious for small or moderately large a's. If, on the other hand, the distribution of X is supported by the whole real line and has at least medium tails, the difference between m and $\mathscr{M}$ becomes more obvious resulting in a larger power of the test.

A more extensive simulation study together with a detailed analysis of a bootstrap approximation will appear elsewhere.

5. Proofs. Our first lemma will be useful for proving the tightness of $\{R_n\}_{n\geq 1}$.

LEMMA 5.1. *Let (α_i, β_i), $1 \leq i \leq n$, be i.i.d. square-integrable bivariate random vectors such that*

$$\mathbb{E}\alpha_i = 0 = \mathbb{E}\beta_i.$$

Then

$$\begin{aligned}\mathbb{E}\left\{\left[\sum_{i=1}^n \alpha_i\right]^2\left[\sum_{j=1}^n \beta_j\right]^2\right\}&\\ = n\mathbb{E}[\alpha_1^2\beta_1^2] + n(n-1)\mathbb{E}[\alpha_1^2]\mathbb{E}[\beta_1^2] + 2n(n-1)\mathbb{E}^2[\alpha_1\beta_1]&\\ \leq n\mathbb{E}[\alpha_1^2\beta_1^2] + 3n(n-1)\mathbb{E}[\alpha_1^2]\mathbb{E}[\beta_1^2].&\end{aligned}$$

PROOF. The equality is hidden in the proof of Theorem 13.1 in Billingsley (1968), but stated there only for centered zero–one variables. Its extension

to the general case is straightforward. The inequality follows from Cauchy–Schwarz. □

REMARK. To make the statement of Lemma 5.1 meaningful, $\mathbb{E}[\alpha_1^2\beta_1^2]$ needs to be finite. By Cauchy–Schwarz this is guaranteed if both α_1 and β_1 have finite fourth moments. In tightness proofs, however, it happens quite often that the above expectation exists under only the assumption that α_1 and β_1 are square-integrable.

PROOF OF THEOREM 1.1. Recalling

$$R_n(x) = n^{-1/2} \sum_{i=1}^{n} 1_{\{X_i \le x\}}[Y_i - m(X_i)],$$

we immediately see that $R_n(x)$ has independent zero-mean summands such that

$$\operatorname{Cov}[R_n(x_1), R_n(x_2)] = \int_{-\infty}^{x_1 \wedge x_2} \sigma^2(v) F(dv) = K(x_1, x_2),$$

where, as in Section 1,

$$\sigma^2(v) = \operatorname{Var}[Y | X = v]$$

denotes the conditional variance of Y given X.

Apply the multivariate CLT to show that the finite dimensional distributions of R_n converge to those of R_∞. For tightness, and also for the representation (1.6), introduce the standard quantile representation

$$X_i = F^{-1}(U_i), \qquad 1 \le i \le n,$$

of X_i in terms of a uniform random variable U_i. Then we have

$$R_n(x) = \bar{R}_n(F(x)),$$

where

$$\bar{R}_n(u) = n^{-1/2} \sum_{i=1}^{n} 1_{\{U_i \le u\}}[Y_i - m \circ F^{-1}(U_i)].$$

Observe that

$$\mathbb{E}[Y_i | U = u] = m \circ F^{-1}(u).$$

In other words, we may and do assume in the following that F is the uniform distribution on [0, 1]. For arbitrary $0 \le u_1 \le u \le u_2 \le 1$, set

$$\alpha_i = 1_{\{u_1 < U_i \le u\}}[Y_i - m(U_i)]$$

and

$$\beta_i = 1_{\{u < U_i \le u_2\}}[Y_i - m(U_i)].$$

As to our previous remark, note that $\alpha_i \beta_i \equiv 0$. From Lemma 5.1 we therefore obtain

$$\mathbb{E}\left\{[\bar{R}_n(u_2) - \bar{R}_n(u)]^2 [\bar{R}_n(u) - \bar{R}_n(u_1)]^2\right\} \le 3 \int_{u_1}^{u} \sigma^2(v)\, dv \int_{u}^{u_2} \sigma^2(v)\, dv$$
$$\le 3[H(u_2) - H(u_1)]^2,$$

where

$$H(u) := \int_0^u \sigma^2(v)\, dv$$

is a continuous nondecreasing function. According to Theorem 15.7 in Billingsley (1968), the proof of Theorem 1.1 is complete. □

Next we will give the proof for the representation and convergence of R_n^1.

PROOFS OF THEOREM 1.2 AND COROLLARY 1.3. Write

$$R_n^1(x) = \bar{R}_n(F(x)) - n^{-1/2} \sum_{i=1}^{n} 1_{\{X_i \le x\}} [m(X_i, \theta_n) - m(X_i, \theta_0)]$$
$$= \mathrm{I} - \mathrm{II}.$$

The second sum admits the representation

$$II = n^{1/2}(\theta_n - \theta_0)^T n^{-1} \sum_{i=1}^{n} 1_{\{X_i \le x\}} [g(X_i, \theta_{ni}) - g(X_i, \theta_0)]$$
$$+ n^{1/2}(\theta_n - \theta_0)^T n^{-1} \sum_{i=1}^{n} \left[1_{\{X_i \le x\}} g(X_i, \theta_0) - G(x, \theta_0)\right]$$
$$+ n^{1/2}(\theta_n - \theta_0)^T G(x, \theta_0),$$

where θ_{ni} denote proper points between θ_n and θ_0. Apply Assumption 2 and the SLLN to show that the first two terms are $o_{\mathbb{P}}(1)$. From Assumption 1, the last process is asymptotically equivalent to

$$n^{-1/2} \sum_{i=1}^{n} l(X_i, Y_i, \theta_0)^T G(x, \theta_0). \tag{5.1}$$

Convergence of the fidis and computation of the limit covariance are now straightforward. For a continuous F, since (5.1) is asymptotically C-tight, as is I by Theorem 1.1, R_n^1 is also asymptotically C-tight. Apply a quantile transformation to deal with the general case. □

PROOF OF THEOREM 2.1. The second equality in Theorem 2.1 is trivial in view of (2.8) and $R_\infty = \sum_{j=1}^{\infty} \lambda_j^{1/2} \xi_j l_j$. For the first, note that both sides are zero-mean Gaussian processes. Hence it suffices to show that the covariance function of the series equals (2.5) (with $\sigma^2 = 1$ and F = uniform distribution on

[0, 1]). Since $\xi_1, \xi_2, \ldots$ are independent and standard normal, the covariance of $\sum_{j=1}^{\infty} \xi_j D_j$ becomes

$$\sum_{j=1}^{\infty} D_j(s)D_j(t) = \sum_{j=1}^{\infty} \lambda_j l_j(s)l_j(t) - \sum_{j=1}^{\infty} \lambda_j^{1/2} l_j(t)G^T(s)A^{-1}\delta_j$$
$$- \sum_{j=1}^{\infty} \lambda_j^{1/2} l_j(s)G^T(t)A^{-1}\delta_j + \sum_{j=1}^{\infty} G^T(t)A^{-1}\delta_j G^T(s)A^{-1}\delta_j.$$

By Mercer's theorem, the first series becomes $K(s, t)$. The second equals $G^T(t)A^{-1}G(s)$, by (2.12). Similarly for the third. In view of (2.10), the fourth also equals $G^T(t)A^{-1}G(s)$. The proof is complete. □

PROOF OF THEOREM 2.2. It is readily checked that M is symmetric and idempotent. Since ξ consists of independent standard normal coordinates, $M\xi$ is a zero-mean normal vector with covariance function $MM^T = M^2 = M$. Finally, the equation $\eta = M\xi$ (in distribution) follows from Theorem 2.1, (2.12) and the very definition of ξ, η and M. □

PROOF OF THEOREM 2.3. Assertion (i) follows from Theorem 4.2 on page 355 and Problem 1 on page 307 of Taylor and Lay (1980). Verification of the assumptions for that problem is simple but tedious and will therefore be omitted. For (ii) and (iii) recall from (2.12) that

$$G = \Delta^T \lambda^{1/2} l.$$

From this, (2.12) and (i) we obtain (in matrix form)

$$\int_0^1 N^T l(s)K^1(s, t)\,ds$$
$$= N^T \int_0^1 l(s)(s \wedge t)\,ds - N^T \int_0^1 l(s)G^T(s)\,ds(\Delta^T\Delta)^{-1}\Delta^T\lambda^{1/2}l(t)$$
$$= N^T\lambda l(t) - N^T\lambda^{1/2}\Delta(\Delta^T\Delta)^{-1}\Delta^T\lambda^{1/2}l(t)$$
$$= N^T\lambda^{1/2}M\lambda^{1/2}l(t) = N^T N\mu N^T l(t) = \mu N^T l(t).$$

This proves that $N^T l$ is a set of eigenfunctions with eigenvalues μ_i, $i = 1, 2, \ldots$. It is easily seen that there are no other eigenvalues. □

6. Concluding remarks. We would like to end this paper by adding some comments on possible extensions to a multivariate design. As to the measure-theoretic motivation, the class of intervals in the definition of I needs to be replaced by any measure-determining family of sets in the multivariate case. Theorem 1.1 can be immediately extended to quadrants by verifying the distributional convergence criteria for multiparameter processes due to Neuhaus (1971) and Straf (1971). Under the assumption that (in our terms) the design

variable is independent of the errors, Gaenssler (1994) proved distributional convergence of R_n when parametrized by a Vapnik–Chervonenkis class. The primary purpose of the present paper, however, has been not only to provide invariance principles for R_n and R_n^1, but also to find their principal components. As we have seen, these components play a crucial role for both creating new tests and discussing their power properties. For multivariate X's we could find corresponding decompositions if we parametrize R_n^1 by a linearly ordered class of sets. For example, take all balls with a fixed center. Of course, such a class would not be measure determining. On the other hand, if we parametrize by all quadrants respectively points in the sample space of X, with possible dependences among its components, no explicit decomposition is available so far. In this context, we are also not aware of any explicit (i.e., not only theoretical) decomposition of the process studied by Durbin, Knott and Taylor (1975) in the multivariate case.

REFERENCES

ATKINSON, A. C. (1985). *Plots, Transformations and Regression*. Oxford Univ. Press.

BARRY, D. and HARTIGAN, J. A. (1990). An omnibus test for departures from constant mean. *Ann. Statist.* **18** 1340–1357.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BUCKLEY, M. J. (1991). Detecting a smooth signal: optimality of cusum based procedures. *Biometrika* **78** 253–262.

COOK, R. D. and WEISBERG, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, New York.

COX, D., KOH, E., WAHBA, G. and YANDELL, B. S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16** 113–119.

DAVISON, A. C. and TSAI, C. L. (1992). Regression model diagnostics. *Internat. Statist. Rev.* **60** 337–353.

DELGADO, M. A. (1993). Testing the equality of nonparametric curves. *Statist. Probab. Lett.* **17** 199–204.

DIEBOLT, J. (1995). A nonparametric test for the regression function: asymptotic theory. *J. Statist. Plann. Inference* **44** 1–17.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279–290.

DURBIN, J., KNOTT, M. and TAYLOR, C. C. (1975). Components of Cramér–von Mises statistics. II. *J. Roy. Statist. Soc. Ser. B* **37** 216–237.

EUBANK, R. L. and HART, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20** 1412–1425.

EUBANK, R. L. and HART, J. D. (1993). Commonality of cusum, von Neumann and smoothing-based goodness-of-fit tests. *Biometrika* **80** 89–98.

EUBANK, R. L. and LARICCIA, V. N. (1992). Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Statist.* **20** 2071–2086.

EUBANK, R. L. and SPIEGELMAN, C. H. (1990). Testing the goodness-of-fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85** 387–392.

FIRTH, D., GLOSUP, J. and HINKLEY, D. V. (1991). Model checking with nonparametric curves. *Biometrika* **78** 245–252.

GAENSSLER, P. (1994). On recent developments in the theory of set-indexed processes. In *Asymptotic Statistics* (P. Mandl and M. Husková, eds.) 87–109. Physica, Heidelberg.

GRENANDER, U. (1952). Stochastic processes and statistical inference. *Ark. Mat.* **1** 195–277.

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947.

HEWITT, E. and STROMBERG, K. (1965). *Real and Abstract Analysis*. Springer, Berlin.

IMHOF, P. J. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48** 419–426.

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least-squares estimators. *Ann. Math. Statist.* **40** 633–643.

JOGLEKAR, G., SCHUENEMEYER, J. H. and LARICCIA, V. (1989). Lack of fit testing when replicates are not available. *Amer. Statist.* **43** 135–143.

JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Univariate Distributions* **2**. Houghton-Mifflin, Boston.

MILBRODT, H. and STRASSER, H. (1990). On the asymptotic power of the two-sided Kolmogorov–Smirnov-test. *J. Statist. Plann. Inference* **26** 1–23.

MÜLLER, H. G. (1992). Goodness-of-fit diagnostics for regression models. *Scand. J. Statist.* **19** 157–172.

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

NEILL, J. W. and JOHNSON, D. E. (1985). Testing linear regression function adequacy without replication. *Ann. Statist.* **13** 1482–1489.

NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.

NEUHAUS, G. (1976). Asymptotic power properties of the Cramér–von Mises test under contiguous alternatives. *J. Multivariate Anal.* **6** 95–110.

NEYMAN, J. (1937). Smooth test for goodness-of-fit. *Skand. Aktuarietidskr.* **20** 149–199.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

RANDLES, R. H. (1984). On tests applied to residuals. *J. Amer. Statist. Assoc.* **79** 349–354.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

SCHOENFELD, D. A. (1977). Asymptotic properties of tests based on linear combinations of the orthogonal components of the Cramér–von Mises statistic. *Ann. Statist.* **5** 1017–1026.

SCHOENFELD, D. A. (1980). Tests based on linear combinations of the orthogonal components of the Cramér–von Mises statistic when parameters are estimated. *Ann. Statist.* **8** 1017–1022.

SEBER, G. A. F. and WILD, C. J. (1989). *Nonlinear Regression*. Wiley, New York.

SHILLINGTON, E. R. (1979). Testing lack of fit in regression without replication. *Canad. J. Statist.* **7** 137–146.

STANISWALIS, J. G. and SEVERINI, T. A. (1991). Diagnostics for assessing regression models. *J. Amer. Statist. Assoc.* **86** 684–692.

STRAF, M. L. (1971). Weak convergence of stochastic processes with several parameters. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 187–221. Univ. California Press, Berkeley.

STUTE, W. and GONZÁLEZ-MANTEIGA, W. (1995). NN goodness-of-fit tests for linear models. *J. Statist. Plann. Inference* **53** 75–92.

SU, J. Q. and WEI, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *J. Amer. Statist. Assoc.* **86** 420–426.

TAYLOR, A. E. and LAY, D. C. (1980). *Introduction to Functional Analysis*. Wiley, New York.

UTTS, J. M. (1982). The rainbow test for lack of fit in regression. *Comm. Statist. Theory Methods* **11** 2801–2815.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

MATHEMATICAL INSTITUTE
UNIVERSITY OF GIESSEN
ARNDTSTRASSE 2
D-35392 GIESSEN
GERMANY
E-MAIL: winfried.stute@math.uni-giessen.de