Distribution-free specification tests of conditional models

Miguel A. Delgado[a,*], Winfried Stute[b]

[a]Departamento de Economía, Universidad Carlos III de Madrid, C./ Madrid 126-128, Getafe 28903, Madrid, Spain
[b]Mathematical Institute, University of Giessen, Arndtstr. 2, Germany

Available online 19 August 2007

Abstract

This article proposes a class of asymptotically distribution-free specification tests for parametric conditional distributions. These tests are based on a martingale transform of a proper sequential empirical process of conditionally transformed data. Standard continuous functionals of this martingale provide omnibus tests while linear combinations of the orthogonal components in its spectral representation form a basis for directional tests. Finally, Neyman-type smooth tests, a compromise between directional and omnibus tests, are discussed. As a special example we study in detail the construction of directional tests for the null hypothesis of conditional normality versus heteroskedastic contiguous alternatives. A small Monte Carlo study shows that our tests attain the nominal level already for small sample sizes.

Keywords: Conditional models; Martingale transformation; Sequential empirical process; Specification tests

1. Introduction

The correct specification of a statistical model is important for several reasons. First, it provides a convenient framework to describe and understand, for example, the dynamics of a time series or a causal relation between independent and dependent variables in regression. In each case it turns out that conditional quantities like autoregressive functions or conditional distributions are of major interest, while marginal distributions of explanatory variables may be considered as parametric or nonparametric nuisance parameter functions. The choice of the model has some consequences on the estimation of unknown parameters and hence on the interpretation of data or the prediction of future values of a dependent variable. The validity of statistical inferences based on conditional maximum likelihood principle, e.g., relies on the correct specification of the conditional distribution model. In particular, the popular Lagrange multiplier and likelihood ratio tests on parameter restrictions are invalid under misspecification, though robust but inefficient inferences are possible. However, classical procedures are optimal under a correct specification. Applications using conditional maximum likelihood are available in abundant supply in economics, as well as in any other disciplines where statistical inference is indispensable. The correct specification of conditional distributions is especially crucial in microeconometrics and biostatistics, where parameter identification is sustained by a correct specification. In these cases, parameter estimates are inconsistent under misspecification. See the

*Corresponding author.
E-mail addresses: miguelangel.delgado@uc3m.es (M.A. Delgado), winfried.stute@math.uni-giessen.de (W. Stute).

classical monograph by Maddala (1983) on limited-dependent and qualitative variables models, Cameron and Trivedi (1998) for count data models, or Lancaster (1990) for duration models.

In the simple case of independent identically distributed observations the history of goodness-of-fit tests started with the classical χ^2-test for cell probabilities. For continuous variables most of the procedures, like Kolmogorov–Smirnov and Cramér–von Mises tests, are based on proper functionals of the empirical process. When the model to be tested is composite, the need to estimate unknown parameters has some impact on the distributional character under the null model so that available tables of critical values are no longer valid. See the work of Gikhman (1953) and Kac et al. (1955) for some early fundamental contributions in this context. A formal derivation of the limit process is due to Durbin (1973) and Neuhaus (1973, 1976), among others. For practical purposes, critical values of the tests can be obtained either through resampling or through the orthogonal components in the spectral representation of the underlying empirical process, as suggested by Durbin et al. (1975).

A different approach was initiated by Khmaladze (1981), who proposed to transform the empirical process to an appropriate martingale, which in distribution may then be approximated by a time-transformed Brownian Motion. As a consequence, classical functionals of these processes like the Kolmogorov–Smirnov or Cramér–von Mises test statistics become asymptotically distribution-free so that existing tables can be used.

In this paper we are interested, for a multivariate observation (X, Y), in the conditional distribution of Y given $X = x$. For the related question of testing just the conditional mean and not the whole conditional distributional structure, the literature is much more elaborate. Härdle and Mammen (1993) were among the first to compare parametric and nonparametric fits. These tests require some smoothing to the effect that the power of these tests may depend on the choice of the smoothing parameter. Stute (1997) investigated so-called integrated regression function (or cusum) processes which avoid smoothing and at the same time allow for a principal component analysis. If we replace (in our notation) Y by indicators $1_{\{Y \leqslant y\}}$, these approaches lead to tests of conditional probability models and may be found in Andrews (1997). In particular he investigated the Kolmogorov–Smirnov test. Due to the complicated distributional character of the test statistic, a bootstrap approximation was proposed and studied. The martingale transformation of the cusum process for fixed design and linear regression is due to Brown et al. (1975). The random design case with a possibly nonlinear regression function has been dealt with in Stute et al. (1998), while applications to time series and generalized linear models may be found in Koul and Stute (1999) and Stute and Zhu (2002). See also Nikabadze and Stute (1997) and Khmaladze and Koul (2004). Zheng (2000) has extended the smoothing approach to specification tests of conditional distributions, while Bai (2003) has applied Khmaladze's martingale approach to tests of the marginal distribution of time series innovations.

To motivate the approach of the present paper we recall a fundamental result due to Rosenblatt (1952). Namely, let (X, Y) be a bivariate random vector with an unknown continuous distribution function F. Denote with F_X the marginal distribution function of X and let $F_{Y|X}(y|x)$ be the conditional distribution function of Y given $X = x$ evaluated at y. Given F_X, F is uniquely determined through $F_{Y|X}$ and vice versa.

In nonparametric testing for F, it is known that tests based on the empirical distribution function are no longer distribution-free. In this context, Rosenblatt (1952) used F_X and $F_{Y|X}$ to introduce a transformation $T = T(X, Y) = (U, V)$ of (X, Y), which maps (X, Y) into a vector (U, V) such that U and V are independent and uniformly distributed on [0,1]. Just put $U = F_X(X)$ and $V = F_{Y|X}(Y|X)$. It is easy to recover (X, Y) from (U, V). Actually, we have with probability one $(X, Y) = (F_X^{-1}(U), F_{Y|X}^{-1}(V|F_X^{-1}(U))$, where G^{-1} denotes the quantile function of a distribution function G. The transformation T can be extended to higher dimensions, but in this paper, for most of the time, we shall stick to the bivariate case. We rather study the important situation when $X = Z^{\mathrm{T}}\delta_0$, for a $p \times 1$ random vector Z and an unknown parameter vector δ_0, so that the multidimensionality of the model enters through a proper projection of a random vector Z. The extension to the case where $X = m(Z, \delta_0)$ for a suitably smooth m is routine. These so-called dimension reducing models are popular in applied fields and naturally lead to an input–output analysis in which, at an intermediate step, the independent variable is univariate. This is relevant in many econometric applications, where one assumes a regression model with innovations independent of the explanatory variables, e.g., limited-dependent variable models.

The Rosenblatt transform T constitutes the extension of the transformation $U = F_X(X)$, which is basic in the analysis of univariate data and leads to many distribution-free procedures based on ranks or

Kolmogorov–Smirnov and Cramér–von Mises discrepancies. Since ordering is unavailable in the multivariate case we propose to order the inputs through the U's treating the V's as the associated concomitants. This leads to a sequential version of an empirical process based on concomitants. Its statistical analysis will be the focus of this paper.

To be more precise, assume that we observe a sample of independent identically distributed data with the same distribution as (X, Y), say $(X_1, Y_1), \ldots, (X_n, Y_n)$. Set

$$(U_i, V_i) = T(X_i, Y_i), \quad 1 \leqslant i \leqslant n$$

and consider the associated uniform empirical distribution function

$$G_n(u, v) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{U_i \leqslant u\}} 1_{\{V_i \leqslant v\}} \quad \text{for } 0 \leqslant u, v \leqslant 1.$$

Here 1_A is the indicator function of the event A. The empirical process

$$\alpha_n(u, v) := \sqrt{n}[G_n(u, v) - uv] \quad \text{for } 0 \leqslant u, v \leqslant 1$$

is a random element in the Skorokhod space $D[0,1]^2$, endowed with a proper topology. See, for example, Straf (1971), Neuhaus (1971) and Bickel and Wichura (1971). Note that the distribution of α_n is free of F. Throughout this paper we shall denote with "$\longrightarrow_{\mathrm{d}}$" weak convergence or convergence in distribution. It is then well known that in $D[0,1]^2$ we have

$$\alpha_n \longrightarrow_{\mathrm{d}} B^1, \tag{1.1}$$

where B^1 is a tied-down Brownian sheet. That is, a centered Gaussian process on the unit square with covariance kernel

$$\mathbb{E}[B^1(u_1, v_1) B^1(u_2, v_2)] = (u_1 \wedge u_2) \cdot (v_1 \wedge v_2) - u_1 u_2 v_1 v_2.$$

Functionals of the empirical process α_n are distribution-free and form a basis for goodness-of-fit tests of simple hypotheses on F. They are, however, unsuitable for testing the specification of $F_{Y|X}$ when F_X is unknown. In order to circumvent this problem we propose to substitute U_i by the normalized ranks of the X_i's:

$$U_{ni} = F_{Xn}(X_i), \quad 1 \leqslant i \leqslant n,$$

with F_{Xn} denoting the empirical distribution function of $X_1, \ldots, X_n$. This leads to

$$\begin{aligned}\bar{G}_n(u, v) &= \frac{1}{n} \sum_{i=1}^{n} 1_{\{U_{ni} \leqslant u\}} 1_{\{V_i \leqslant v\}} \\ &= \frac{1}{n} \sum_{i=1}^{n} 1_{\{i/n \leqslant u\}} 1_{\{V_{[i:n]} \leqslant v\}} \\ &= \frac{1}{n} \sum_{i=1}^{\lfloor nu \rfloor} 1_{\{V_{[i:n]} \leqslant v\}}.\end{aligned}$$

Here, $V_{[i:n]}$ is the V-concomitant associated with $X_{i:n}$, that is, $V_{[i:n]} = V_j$ if $X_{i:n} = X_j$ with $X_{1:n} \leqslant X_{2:n} \leqslant \cdots \leqslant X_{n:n}$ denoting the set of X-order statistics. The empirical process associated with $\bar{G}_n$ becomes

$$\begin{aligned}\bar{\alpha}_n(u, v) &:= n^{1/2}[\bar{G}_n(u, v) - u \cdot v] \\ &= n^{1/2}[\bar{G}_n(u, v) - v \cdot \bar{G}_n(u, 1)] + v \cdot \frac{\lfloor nu \rfloor - nu}{n^{1/2}}.\end{aligned}$$

Since the second term is negligible, it is natural to consider

$$\beta_n(u,v) := n^{1/2}[\bar{G}_n(u,v) - v \cdot \bar{G}_n(u,1)] = \frac{1}{n^{1/2}} \sum_{i=1}^{\lfloor nu \rfloor} [1_{\{V_{[i:n]} \leqslant v\}} - v],$$

which is the standard sequential empirical process of the concomitants. Notice that, since $\{V_1, \ldots, V_n\}$ and $\{X_1, \ldots, X_n\}$ are independent, $\{V_{[1:n]}, \ldots, V_{[n:n]}\}$ is a random permutation of $\{V_1, \ldots, V_n\}$. That is, $\{V_{[1:n]}, \ldots, V_{[n:n]}\}$ are independent identically distributed copies of V. It follows from classical empirical process theory, see Shorack and Wellner (1986), that

$$\beta_n \longrightarrow_{\mathrm{d}} K \text{ in the space } D[0,1]^2,$$

where K is the standard Kiefer process, a centered biparameter Gaussian process on the unit square with covariance function

$$\mathbb{E}[K(u_1,v_1) \cdot K(u_2,v_2)] = (u_1 \wedge u_2)(v_1 \wedge v_2 - v_1 \cdot v_2).$$

The Kiefer process can be represented in terms of the standard Brownian sheet B, a zero mean Gaussian process with covariance function

$$\mathbb{E}[B(u_1,v_1) \cdot B(u_2,v_2)] = (u_1 \wedge u_2)(v_1 \wedge v_2),$$

namely as

$$K(u,v) = (1-v) \int_0^v \int_0^u \frac{1}{1-\bar{v}} B(\mathrm{d}\bar{u}, \mathrm{d}\bar{v}).$$

In practical situations, the conditional distribution functions $F_{Y|X}$ are parametrically modeled, and the hypothesis to be tested becomes

$$\mathrm{H}_0 : F_{Y|X} \in \mathscr{F}.$$

Here, $\mathscr{F}$ is a given family of parametric conditional distribution functions

$$\mathscr{F} = \{F_{Y|X,\theta} : \theta \in \Theta\},$$

and $\Theta \subset \mathbb{R}^p$ is a proper parameter space. The alternative hypothesis may be specified or not. Under H_0, there exists a $\theta_0 \in \Theta$ such that $F_{Y|X} = F_{Y|X,\theta_0}$, and given a $\sqrt{n}$*-consistent* estimator of θ_0, say θ_n, $\bar{G}_n(u,v)$ can be replaced by

$$\hat{G}_n(u,v) := \frac{1}{n} \sum_{i=1}^{\lfloor nu \rfloor} 1_{\{\hat{V}_{n[i:n]} \leqslant v\}},$$

with $\hat{V}_{ni} = F_{Y|X,\theta_n}(Y_i|X_i)$ and $\hat{V}_{n[i:n]}$ denoting the $\hat{V}$-concomitant of $X_{i:n}$. The final version of β_n then becomes

$$\hat{\beta}_n(u,v) := n^{1/2}[\hat{G}_n(u,v) - v \cdot \hat{G}_n(u,1)] = \frac{1}{n^{1/2}} \sum_{i=1}^{\lfloor nu \rfloor} [1_{\{\hat{V}_{n[i:n]} \leqslant v\}} - v].$$

The asymptotic distribution of $\hat{\beta}_n(1,\cdot)$ may be derived along the lines of Durbin (1973), who as already mentioned established the weak limit of the univariate empirical process with estimated parameters. The empirical process $\hat{\beta}_n(1,\cdot)$ has also been considered by Bai (2003) for testing $\dot{\mathrm{H}}_0$: $\mathbb{E}(F_{Y|X,\theta_0}(y|X)) = F_Y(y)$ for some $\theta_0 \in \Theta$, with F_Y denoting the marginal distribution function of Y. The resulting test has trivial power for testing H_0 in all directions where $\dot{\mathrm{H}}_0$ is satisfied. Neuhaus (1971, 1976) extended Durbin's (1973) results to the multiparameter case and considered general contiguous nonparametric alternatives. See also Bai (1994, 1996).

We derive the asymptotic distribution of $\hat{\beta}_n$ under the type of regularity conditions on $\mathscr{F}$ corresponding to Neuhaus (1976) and Durbin (1973):

A1. Assume that $\partial F_{Y|X,\theta}(y|x)/\partial\theta$ exists for all $(x, y) \in \mathbb{R}^2$ and each component of the vector of functions

$$q_\theta(u, v) := \int_0^u \frac{\partial}{\partial\theta} F_{Y|X,\theta}(F^{-1}_{Y|X,\theta}(v|F_X^{-1}(\bar{u}))|F_X^{-1}(\bar{u}))\,\mathrm{d}\bar{u}$$

is continuous on $[0, 1]^2 \times \Theta$.

Our first result is crucial for proving the weak convergence of $\hat{\beta}_n$. It provides a convenient representation of $\hat{\beta}_n$ in terms of β_n and $\theta_n - \theta_0$.

Theorem 1. *Under* H_0 *and for* $\mathscr{F}$ *satisfying* A1, *suppose that* $\theta_n = \theta_0 + \mathrm{O}_{\mathbb{P}}(n^{-1/2})$. *Then we have*

$$\sup_{(u,v)\in[0,1]^2} |\hat{\beta}_n(u, v) - \beta_n(u, v) + q_{\theta_0}(u, v)^{\mathrm{T}} n^{1/2}(\theta_n - \theta_0)| = \mathrm{o}_{\mathbb{P}}(1).$$

In many situations θ_n admits a linear representation in terms of independent identically distributed random variables, in which case we can identify the limit of $\hat{\beta}_n$.

A2. Assume that

$$\theta_n = \theta_0 + \frac{1}{n}\sum_{i=1}^n \ell_{\theta_0}(X_i, Y_i) + \mathrm{o}_{\mathbb{P}}(n^{-1/2}),$$

where, for each $x \in \mathbb{R}$ and every $\theta \in \Theta$,

$$\int_{\mathbb{R}} \ell_\theta(x, y) F_{Y|X,\theta}(\mathrm{d}y|x) = 0$$

and

$$\sup_{x\in\mathbb{R}} \left\| \int_{\mathbb{R}} \ell_\theta(x, y)\ell_\theta(x, y)^{\mathrm{T}} F_{Y|X,\theta}(\mathrm{d}y|x) \right\| < \infty.$$

When $\mathscr{F}$ is given through its conditional densities $f_{Y|X,\theta}$, say, a natural estimator of θ_0 is the conditional maximum likelihood estimator:

$$\theta_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^n \ln f_{Y|X,\theta}(Y_i|X_i).$$

In this case,

$$\ell_\theta(x, y) = \mathscr{I}_\theta^{-1} \frac{\partial}{\partial\theta} \ln f_{Y|X,\theta}(y|x),$$

where

$$\mathscr{I}_\theta = \mathbb{E}\left[\frac{\partial}{\partial\theta} \ln f_{Y|X,\theta}(Y|X) \frac{\partial}{\partial\theta^{\mathrm{T}}} \ln f_{Y|X,\theta}(Y|X)\right]$$

is the "conditional" information matrix:

$$q_\theta(u, v) = \int_0^v \int_0^u \frac{\partial}{\partial\theta} \ln f_{Y|X,\theta}(F^{-1}_{Y|X,\theta}(\bar{v}|F_X^{-1}(\bar{u}))|F_X^{-1}(\bar{u}))\,\mathrm{d}\bar{u}\,\mathrm{d}\bar{v}$$

$$\equiv \int_0^v \int_0^u \varphi_\theta(\bar{u}, \bar{v})\,\mathrm{d}\bar{u}\,\mathrm{d}\bar{v}. \tag{1.2}$$

The next result is a consequence of Theorem 1 and A2.

Corollary 1. *Under the conditions in Theorem* 1 *and* A2,

$$\hat{\beta}_n \longrightarrow_{\mathrm{d}} \hat{\beta}_\infty \text{ in the space } D[0, 1]^2,$$

with

$$\hat{\beta}_\infty(u,v) = K(u,v) - q_{\theta_0}(u,v)^{\mathrm{T}} \cdot \int_0^1 \int_0^1 \ell_{\theta_0}(F_X^{-1}(\bar{u}), F_{Y|X,\theta_0}^{-1}(\bar{v}|F_X^{-1}(\bar{u})))B(\mathrm{d}\bar{u}, \mathrm{d}\bar{v}).$$

If an observation $(X, Y, Z, \ldots)$ is multivariate with more than two components, the Rosenblatt transformation also works but requires, besides F and $F_{Y|X}$, also the specifications of $F_{Z|X,Y,\ldots}$. Rather than this, we now discuss the case when $X = Z^{\mathrm{T}}\delta_0$. Along with θ_n, let δ_n be a $\sqrt{n}$-consistent estimator of δ_0. For example we could take

$$(\theta_n, \delta_n) = \underset{(\theta,\delta)}{\operatorname{argmax}} \sum_{i=1}^n \ln f_{Y|X,\theta}(Y_i|Z_i^{\mathrm{T}}\delta).$$

Consider the following modification of $\hat{\beta}_n$:

$$\tilde{\beta}_n(u,v) := \frac{1}{n^{1/2}} \sum_{i=1}^{\lfloor nu \rfloor} [1_{\{\tilde{V}_{n[i:n]} \leqslant v\}} - v] \equiv n^{1/2}[\tilde{G}_n(u,v) - v\tilde{G}_n(u,1)],$$

where now $\tilde{V}_{n[i:n]}$ is the ith $\hat{V}$-concomitant with respect to the ordered $\tilde{X}_{n1}, \ldots, \tilde{X}_{nn}$, where $\tilde{X}_{ni} = Z_i^{\mathrm{T}}\delta_n$ is in place of $X_i = Z_i^{\mathrm{T}}\delta_0$. In this case the need to estimate δ_0 requires an additional correction in the expansion of the associated $\tilde{G}_n$.

For the sake of simplicity we only consider the case when θ and δ have no coordinates in common. Otherwise the derivative needs to be taken only with respect to the components of δ which do not appear in θ.

Theorem 2. *Under the conditions of Theorem* 1, *assume that $F_{Y|X,\theta}(y|x)$ is also differentiable with respect to x and let δ_n and θ_n be $\sqrt{n}$-consistent estimators of δ_0 and θ_0, respectively. Assume also that Z has finite second moments. Then*

$$\sup_{0 \leqslant u,v \leqslant 1} |\tilde{\beta}_n(u,v) - \beta_n(u,v) + q_{\theta_0,\delta_0}(u,v)^{\mathrm{T}} n^{1/2}(\theta_n - \theta_0) + q^1_{\theta_0,\delta_0}(u,v)^{\mathrm{T}} n^{1/2}(\delta_n - \delta_0)| = \mathrm{o}_{\mathbb{P}}(1).$$

Here q_{θ_0,δ_0} is the q-function from before, but with $F_X(x) = \mathbb{P}(Z^{\mathrm{T}}\delta_0 \leqslant x)$ now depending on the unknown δ_0 and

$$\begin{aligned} q^1_{\theta_0,\delta_0}(u,v) &:= \mathbb{E}\left[1_{\{F_X(Z^{\mathrm{T}}\delta_0) \leqslant u\}} Z \frac{\partial}{\partial x} F_{Y|X,\theta_0}(F_{Y|X,\theta_0}^{-1}(v|Z^{\mathrm{T}}\delta_0)|Z^{\mathrm{T}}\delta_0)\right] \\ &= \int_0^u r(F_X^{-1}(\bar{u})) \frac{\partial}{\partial x} F_{Y|X,\theta_0}(F_{Y|X,\theta_0}^{-1}(v|F_X^{-1}(\bar{u}))|F_X^{-1}(\bar{u}))\,\mathrm{d}\bar{u} \end{aligned} \tag{1.3}$$

with $r(x) = \mathbb{E}[Z|X = x]$ denoting the vector-valued regression function of Z given $X = Z^{\mathrm{T}}\delta_0 = x$.

Typically, δ_n also admits a representation in terms of independent identically distributed random variables. We also obtain an analogue of Corollary 1. Since, however, the limit process depends on unknown parameters, the unknown F_X and the model $\mathscr{F}$, tests based on $\hat{\beta}_n$ and $\tilde{\beta}_n$ are still not (asymptotically) distribution-free.

The rest of the paper is organized as follows. The next section presents a transformation of the sequential empirical process of estimated concomitants, which converges in distribution to the standard biparameter Brownian sheet. Hence, continuous functionals of this transformed process are suitable for testing composite hypotheses. Power considerations are studied in Section 3, where we provide the limiting distribution of the transformed process under contiguous alternatives converging to the null at the parametric rate $n^{-1/2}$. In this section, we also provide the spectral decomposition of the transformed process and propose test statistics based on linear combinations of the principal components. Furthermore we derive test statistics consisting of the optimal combination of principal components, thus maximizing the power in the direction of a particular contiguous alternative. The results of a Monte Carlo experiment are reported in Section 4. Proofs are postponed to the appendix.

2. Distribution-free transformation of the sequential empirical process with estimated concomitants

The martingale transformation of $\hat{\beta}_n$ to be discussed now will turn out to be a composition of two operators. In the first step we transform $\hat{\beta}_n$ so that in the limit the Kiefer process will be replaced by the Brownian sheet. In the next step we shall apply a model-dependent transformation which is designed to give us distribution-free processes.

Now, as mentioned earlier, the Kiefer process can be represented in terms of independent Gaussian increments, namely as a stochastic integral with respect to a Brownian sheet:

$$K(u, v) = (1 - v) \int_0^v \int_0^u \frac{1}{1 - \bar{v}} B(\mathrm{d}\bar{u}, \mathrm{d}\bar{v}).$$

Inverting this last expression, we obtain

$$B = \mathscr{L}_0 K,$$

where $\mathscr{L}_0$ is the linear operator defined as

$$\mathscr{L}_0 m(u, v) = m(u, v) - \int_0^v \frac{1}{1 - \bar{v}} \int_{\bar{v}}^1 \int_0^u m(\mathrm{d}\tilde{u}, \mathrm{d}\tilde{v})\, \mathrm{d}\bar{v},$$

for a generic function $m : [0, 1]^2 \to \mathbb{R}$.

Hence, tests on simple hypotheses on $F_{Y|X}$ can alternatively be based on the transformed process

$$\mathscr{L}_0 \beta_n(u, v) = n^{1/2} \mathscr{L}_0 \bar{G}_n(u, v) = \frac{1}{n^{1/2}} \sum_{i=1}^{\lfloor nu \rfloor} [1_{\{V_{[i:n]} \leqslant v\}} + \log[1 - (v \wedge V_{[i:n]})]].$$

Note that this is the time-sequential version of the martingale part in the Doob–Meyer decomposition of the uniform empirical process. Applying the continuous mapping theorem and the weak convergence of β_n, we have, under H_0,

$$\mathscr{L}_0 \beta_n \longrightarrow_{\mathrm{d}} B \text{ in the space } D[0, 1]^2.$$

Similarly

$$\mathscr{L}_0 \hat{\beta}_n(u, v) = n^{1/2} \mathscr{L}_0 \hat{G}_n(u, v) = \frac{1}{n^{1/2}} \sum_{i=1}^{\lfloor nu \rfloor} [1_{\{\hat{V}_{n[i:n]} \leqslant v\}} + \log[1 - (v \wedge \hat{V}_{n[i:n]})]],$$

while for $\mathscr{L}_0 \tilde{\beta}_n$ the $\hat{V}_{n[i:n]}$ need to be replaced with $\tilde{V}_{n[i:n]}$.

Assuming that the conditions in Corollary 1 are satisfied, then

$$\mathscr{L}_0 \hat{\beta}_n \longrightarrow_{\mathrm{d}} \mathscr{L}_0 \hat{\beta}_\infty,$$

with

$$\mathscr{L}_0 \hat{\beta}_\infty(u, v) = B(u, v) - \int_0^v \int_0^u h_{\theta_0}(\bar{u}, \bar{v})^{\mathrm{T}}\, \mathrm{d}\bar{u}\, \mathrm{d}\bar{v} \cdot \int_0^1 \int_0^1 \bar{\ell}_{\theta_0}(\tilde{u}, \tilde{v}) B(\mathrm{d}\tilde{u}, \mathrm{d}\tilde{v}),$$

where

$$\mathscr{L}_0 q_{\theta_0}(u, v) = \int_0^v \int_0^u h_{\theta_0}(\bar{u}, \bar{v})\, \mathrm{d}\bar{u}\, \mathrm{d}\bar{v}$$

and

$$\bar{\ell}_\theta(u, v) = \ell_\theta(F_X^{-1}(u), F_{Y|X,\theta}^{-1}(v|F_X^{-1}(u))).$$

If, as in the case of the maximum likelihood estimator, see (1.2), q_θ has a Lebesgue density φ_θ, we have

$$h_\theta(u, v) = \varphi_\theta(u, v) - \frac{1}{1 - v} \int_v^1 \varphi_\theta(u, \bar{v})\, \mathrm{d}\bar{v}. \tag{2.1}$$

From the above representation of $\mathscr{L}_0\hat{\beta}_n$ we see that K has been replaced by B. Actually, unlike $\hat{\beta}_\infty$, $\mathscr{L}_0\hat{\beta}_\infty$ admits the same type of representation as the limiting distribution of the standard biparameter empirical process with estimated parameters. This fact suggests to apply the scanning innovation approach proposed by Khmaladze (1988, 1993) in order to obtain an empirical process converging in distribution to the biparameter Brownian sheet under the null. For this, let us consider a family of measurable subsets,

$$\mathscr{S} = \{S_{(u,v)} : (u,v) \in [0,1]^2\},$$

satisfying the following properties:

1. For every $(u_1,v_1),(u_2,v_2) \in [0,1]^2$, $S_{(u_1,v_1)} \subset S_{(u_2,v_2)}$ or $S_{(u_2,v_2)} \subset S_{(u_1,v_1)}$, that is, S is linearly ordered.
2. $\bigcup_{(u,v)} S(u,v) = [0,1]^2$ and $\bigcap_{(u,v)} S(u,v) = \emptyset$.
3. If $S_{(u_i,v_i)} \in \mathscr{S}$, $i = 1,2,\ldots$ then $\liminf_n S_{(u_n,v_n)} \in \mathscr{S}$.
4. $S_{(u_1,v_1)} \backslash S_{(u_2,v_2)} \to S_0$ as $(u_1,v_1) \to (u_2,v_2)$,

where S_0 is a set with Lebesgue measure equal to zero.

Examples of sets satisfying these conditions are

$$\mathscr{S} = \{[0,1] \times [0,v], v \in [0,1]\}, \tag{2.2}$$

$$\mathscr{S} = \{[0,v] \times [0,v], v \in [0,1]\}. \tag{2.3}$$

For any particular family of sets $\mathscr{S}$, let us define the matrix

$$A_\theta(u,v) = \int\int_{\bar{S}(u,v)} h_\theta(\bar{u},\bar{v}) h_\theta(\bar{u},\bar{v})^{\mathrm{T}} \,\mathrm{d}\bar{u}\,\mathrm{d}\bar{v},$$

where $\bar{S}(u,v)$ denotes the complement of $S(u,v)$. The scanning innovation of $\mathscr{L}_0\hat{\beta}_\infty$ is given by $(\mathscr{L}_{\theta_0} \circ \mathscr{L}_0)\hat{\beta}_\infty$, where $\mathscr{L}_\theta$ is the linear operator defined as

$$\mathscr{L}_\theta m(u,v) = m(u,v) - \int_0^v \int_0^u h_\theta(\bar{u},\bar{v})^{\mathrm{T}} A_\theta^{-1}(\bar{u},\bar{v}) \int\int_{\bar{S}_{(\bar{u},\bar{v})}} h_\theta(\tilde{u},\tilde{v}) m(\mathrm{d}\tilde{u},\mathrm{d}\tilde{v}) \,\mathrm{d}\bar{u}\,\mathrm{d}\bar{v},$$

for a generic function $m : [0,1]^2 \to \mathbb{R}$.

Usually, as it will be the case in this paper, it is assumed that the matrix $A_{\theta_0}(u,v)$ is nonsingular for $(u,v) \in [0,1)^2$, that is, that the components of h_{θ_0} are linearly independent in every interval $[0,u] \times [0,v]$. However, there are families of distributions where this condition is not fulfilled. In such a situation $A_\theta^{-1}(\cdot,\cdot)$ is the generalized inverse of $A_\theta(\cdot,\cdot)$ satisfying

$$A_\theta^{-1}(\cdot,\cdot)[A_\theta(\cdot,\cdot)\xi] = \begin{cases} \xi & \text{if } \xi \in \text{ Image } (A_\theta(\cdot,\cdot)), \\ 0 & \text{otherwise.} \end{cases}$$

Interestingly, the transformation provided by the operator $\mathscr{L}_\theta$ is unique irrespective of the generalized inverse used, as proved by Nikabadze (1997).

The choice of the sets in (2.2) is very convenient from the computational view point. In this case,

$$(\mathscr{L}_\theta \circ \mathscr{L}_0)\hat{\beta}_n(u,v) = \mathscr{L}_0\hat{\beta}_n(u,v) - \int_0^v \int_0^u h_\theta(\bar{u},\bar{v})^{\mathrm{T}} A_\theta^{-1}(\bar{v}) \int_0^1 \int_{\bar{v}}^1 h_\theta(\tilde{u},\tilde{v}) \mathscr{L}_0\hat{\beta}_n(\mathrm{d}\tilde{u},\mathrm{d}\tilde{v}) \,\mathrm{d}\bar{u}\,\mathrm{d}\bar{v},$$

where

$$A_\theta(v) = \int_0^1 \int_v^1 h_\theta(\bar{u},\bar{v}) h_\theta(\bar{u},\bar{v})^{\mathrm{T}} \,\mathrm{d}\bar{v}\,\mathrm{d}\bar{u}$$

only depends on v.

The following theorem provides the weak convergence of the transformed sequential empirical process. Since in most examples A_θ is the null matrix when u or v equals 1, we shall, in the following, restrict our processes to $[0,1)^2$. The associated space $D[0,1)^2$ is endowed with the topology of Skorokhod convergence on compact subsets of $[0,1)^2$. For a related discussion of $D[0,\infty)$, see Pollard (1984).

Theorem 3. *Under* H_0 *and the conditions in Theorem* 1,

$$(\mathscr{L}_{\theta_0} \circ \mathscr{L}_0)\hat{\beta}_n \longrightarrow_{\mathrm{d}} B \text{ in the space } D[0,1)^2.$$

Since F_X and θ_0 are unknown, the transformation $\mathscr{L}_{\theta_0}$ is unavailable in practice and needs to be replaced by its data-dependent analogue. For this, put

$$\widehat{\mathscr{L}}_{\theta_n} m(u,v) = m(u,v) - \int_0^v \int_0^u \hat{h}_{\theta_n}(\bar{u},\bar{v})^{\mathrm{T}} \hat{A}_{\theta_n}^{-1}(\bar{u},\bar{v}) \int\int_{\bar{S}(\bar{u},\bar{v})} \hat{h}_{\theta_n}(\tilde{u},\tilde{v}) m(\mathrm{d}\tilde{u},\mathrm{d}\tilde{v})\,\mathrm{d}\bar{u}\,\mathrm{d}\bar{v},$$

with

$$\hat{A}_\theta(u,v) = \int\int_{\bar{S}(u,v)} \hat{h}_\theta(\bar{u},\bar{v})\hat{h}_\theta(\bar{u},\bar{v})^{\mathrm{T}}\,\mathrm{d}\bar{u}\,\mathrm{d}\bar{v}.$$

Here $\hat{h}_\theta$ is defined through

$$\mathscr{L}_0 \hat{q}_\theta(u,v) = \int_0^v \int_0^u \hat{h}_\theta(\bar{u},\bar{v})\,\mathrm{d}\bar{u}\,\mathrm{d}\bar{v}$$

and $\hat{q}_\theta$ is defined as q_θ, but with F_X replaced with F_{Xn}.

Theorem 4. *Under* H_0 *and the conditions in Theorem* 1,

$$(\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n \longrightarrow_{\mathrm{d}} B \text{ in the space } D[0,1)^2.$$

Theorem 4 reveals that for the operators replacement of θ_0 by θ_n has no effect on the limit, in contrast to the processes β_n and $\hat{\beta}_n$. See also Stute et al. (1998).

Test statistics are based on continuous functionals of $(\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n$. The following corollary is a straightforward consequence of Theorem 4 and the continuous mapping theorem,

Corollary 2. *Under* H_0 *and the conditions in Theorem* 1,

$$\Gamma((\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n) \longrightarrow_{\mathrm{d}} \Gamma(B),$$

for any functional Γ *on* $D[0,1)^2$ *being continuous at the sample paths of* B.

Remark 1. The results of this section continue to hold in the situation of Theorem 2. For this, replace the function q_θ by the function $(q_{\theta,\delta}^{\mathrm{T}}, q_{\theta,\delta}^{1\mathrm{T}})^{\mathrm{T}}$. Since $q_{\theta,\delta}^1$ is an integral, it may be estimated at parametric rates though it contains the unknown regression function r. In fact, in view of (1.3), $q_{\theta_0,\delta_0}^1(u,v)$ can be estimated by

$$\hat{q}_{\theta_n,\delta_n}^1(u,v) = \frac{1}{n}\sum_{i=1}^n 1_{\{\tilde{F}_{Xn}(Z_i^{\mathrm{T}}\delta_n) \leqslant u\}} Z_i \frac{\partial}{\partial x} F_{Y|X,\theta_n}(F_{Y|X,\theta_n}^{-1}(v|Z_i^{\mathrm{T}}\delta_n)|Z_i^{\mathrm{T}}\delta_n),$$

whose increments are free of nonparametric components. Here, $\tilde{F}_{Xn}$ is the sample distribution of $Z_i^{\mathrm{T}}\delta_n$, $i \geqslant 1$.

The *Kolmogorov–Smirnov* and *Cramér–von Mises* statistics pertain to the functionals

$$\Gamma(f) = \sup_{0 \leqslant u,v < 1} |f(u,v)| \quad \text{and} \quad \Gamma(f) = \int_0^1 \int_0^1 f(u,v)^2\,\mathrm{d}u\,\mathrm{d}v,$$

respectively, resulting in the test statistics

$$K_n = \sup_{0 \leqslant u,v < 1} |(\hat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n(u,v)|$$

and

$$C_n = \int_0^1 \int_0^1 |(\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n(u,v)|^2\,\mathrm{d}u\,\mathrm{d}v,$$

Table 1
Critical values of C_∞ and K_∞

	C_∞	K_∞
$\alpha = 0.10$	0.53	2.21
$\alpha = 0.05$	0.72	2.46
$\alpha = 0.01$	1.18	3.03

respectively. Under H_0 and the conditions in Corollary 1,

$$K_n \longrightarrow_{\mathrm{d}} K_\infty = \sup_{0 \leqslant u,v < 1} |B(u,v)|,$$

$$C_n \longrightarrow_{\mathrm{d}} C_\infty = \int_0^1 \int_0^1 B(u,v)^2 \,\mathrm{d}u\,\mathrm{d}v$$

in distribution. Table 1 provides some quantiles of K_∞ and C_∞. The distribution of suprema for the two parameter Brownian Motion (K_∞) has been tabulated by Brownrigg (2005). We have obtained the critical values of C_∞ by simulation, using the spectral representation in (3.2).

From the computational viewpoint, it is more convenient to use the asymptotically equivalent versions

$$\hat{K}_n = \sup_{1 \leqslant i,\ j \leqslant n} \left| (\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n\left(\frac{i}{n}, \hat{V}_{nj}\right) \right|,$$

$$\hat{C}_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| (\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n\left(\frac{i}{n}, \hat{V}_{nj}\right) \right|^2.$$

The resulting tests are omnibus, but power in particular directions can be improved by using linear combinations of the principal components of $(\hat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n$, as will be discussed in the next section.

For the sets in (2.2), the transformation of $\hat{\beta}_n$ can be written as

$$\begin{aligned}(\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n(u,v) &= n^{1/2}(\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{G}_n(u,v) \\ &= \frac{1}{n^{1/2}} \sum_{i=1}^{\lfloor nu \rfloor} [1_{\{\hat{V}_{n[i:n]} \leqslant v\}} + \log[1 - (v \wedge \hat{V}_{n[i:n]})]] \\ &\quad - n^{1/2} \int_0^v \left(\frac{1}{n} \sum_{i=1}^{\lfloor nu \rfloor} \hat{h}_{\theta_n}\left(\frac{i}{n}, \bar{v}\right)^{\mathrm{T}} \right) \hat{A}_{\theta_n}^{-1}(\bar{v}) \int_0^1 \int_{\bar{v}}^1 \hat{h}_{\theta_n}(\tilde{u}, \tilde{v}) \mathscr{L}_0 \hat{G}_n(\mathrm{d}\tilde{u}, \mathrm{d}\tilde{v})\,\mathrm{d}\bar{v}.\end{aligned}$$

It may happen that the function φ_θ in (1.2) and hence h_θ does not depend on u:

$$\varphi_\theta(u,v) = \varphi_\theta(v), \quad h_\theta(u,v) = h_\theta(v).$$

This may be the case, for example, when φ pertains to the maximum likelihood estimator and $\mathscr{F}$ is the normal location-scale family. See Section 4 for details. In such a situation, $\hat{h} = h$ and the transformation of $\hat{\beta}_n$ becomes In such a situation, $\hat{h} = h$ and the transformation of $\hat{\beta}_n$ becomes

$$\begin{aligned}(\widehat{\mathscr{L}}_{\theta_n} \circ \mathscr{L}_0)\hat{\beta}_n(u,v) &= \frac{1}{n^{1/2}} \sum_{i=1}^{\lfloor nu \rfloor} [1_{\{\hat{V}_{n[i:n]} \leqslant v\}} + \log[1 - (v \wedge \hat{V}_{n[i:n]})]] \\ &\quad - n^{1/2} u \int_0^v h_{\theta_n}(\bar{v}) \hat{A}_{\theta_n}^{-1}(\bar{v}) \int_{\bar{v}}^1 h_{\theta_n}(\tilde{v}) \mathscr{L}_0 \hat{G}_n(1, \mathrm{d}\tilde{v})\,\mathrm{d}\bar{v}.\end{aligned}$$

Here

$$\hat{A}_\theta(v) = \int_v^1 h_\theta(\bar{v})h_\theta(\bar{v})^{\mathrm{T}}\,\mathrm{d}\bar{v},$$

while the last double integral may be seen to be equal to

$$\int_0^v \begin{pmatrix} 0 \\ h_{\theta_n}(\bar{v}) \end{pmatrix}^{\mathrm{T}} \begin{bmatrix} 1-\bar{v} & \int_{\bar{v}}^1 \varphi_{\theta_n}(\tilde{v})^{\mathrm{T}}\,\mathrm{d}\tilde{v} \\ \int_{\bar{v}}^1 \varphi_{\theta_n}(\tilde{v})\,\mathrm{d}\tilde{v} & \int_{\bar{v}}^1 \varphi_{\theta_n}(\tilde{v})\varphi_{\theta_n}(\tilde{v})^{\mathrm{T}}\,\mathrm{d}\tilde{v} \end{bmatrix}^{-1} \begin{pmatrix} \int_{\bar{v}}^1 \hat{G}_n(1,\mathrm{d}\tilde{v}) \\ \int_{\bar{v}}^1 \varphi_{\theta_n}(\tilde{v})\hat{G}_n(1,\mathrm{d}\tilde{v}) \end{pmatrix}\mathrm{d}\bar{v}$$

$$= \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} 1 \\ \varphi_{\theta_n}(\hat{V}_{n[i:n]}) \end{pmatrix}^{\mathrm{T}} \int_0^{v\wedge \hat{V}_{n[i:n]}} \begin{bmatrix} 1-\bar{v} & \int_{\bar{v}}^1 \varphi_{\theta_n}(\tilde{v})^{\mathrm{T}}\,\mathrm{d}\tilde{v} \\ \int_{\bar{v}}^1 \varphi_{\theta_n}(\tilde{v})\,\mathrm{d}\tilde{v} & \int_{\bar{v}}^1 \varphi_{\theta_n}(\tilde{v})\varphi_{\theta_n}(\tilde{v})^{\mathrm{T}}\,\mathrm{d}\tilde{v} \end{bmatrix}^{-1} \begin{pmatrix} 0 \\ h_{\theta_n}(\bar{v}) \end{pmatrix}\mathrm{d}\bar{v}.$$

In our simulations the integrals were computed using numerical methods. See Section 4.

3. Contiguous alternatives and directional tests

Consider the contiguous alternatives

A3. $$\mathrm{H}_{1n}: \frac{F_{Y|X}(\mathrm{d}y|x)}{F_{Y|X,\theta_0}(\mathrm{d}y|x)} = 1 + \frac{t_{n\theta_0}(y,x)}{n^{1/2}} \quad \text{some } \theta_0 \in \Theta,$$

where $t_{n\theta}: \mathbb{R}^2 \to \mathbb{R}$ is such that

$$\int_{\mathbb{R}} t_{n\theta}(y,x)F_{Y|X,\theta}(\mathrm{d}y|x) = 0 \quad \text{and} \quad t_{n\theta} \to t_\theta \text{ as } n\to\infty \text{ in } L_2$$

for each $x \in \mathbb{R}$ and all $\theta \in \Theta$.

The restriction on $t_{n\theta}$ allows modeling particular departures from the null hypothesis, which are properly defined conditional distribution functions. In order to illustrate these local alternatives, consider the example of testing conditional normality under homoskedasticity, i.e.

$$\mathrm{H}_0: F_{Y|X,\theta_0}(y|x) = \Phi\left(\frac{y-x}{\sigma}\right),$$

where $\Phi(\varepsilon) = \int_{-\infty}^{\varepsilon}\phi(\bar{\varepsilon})\,\mathrm{d}\bar{\varepsilon}$ and $\phi(\varepsilon) = \exp(-\varepsilon^2/2)/\sqrt{2\pi}$ is the standard normal probability density function. Here, σ^2 is the conditional variance under H_0, that is, the model is homoskedastic. An interesting local alternative is

$$\mathrm{H}_{1n}: F_{Y|X}(y|x) = \Phi\left(\frac{y-x}{\sigma_n(x)}\right) \text{ with } \sigma_n^2(x) = \sigma^2\left[1+\frac{\gamma(x)}{n^{1/2}}\right] \quad \text{for some } \sigma > 0,$$

for a particular positive function γ. This contiguous alternative can be alternatively written as

$$\mathrm{H}_{1n}: \frac{F_{Y|X}(\mathrm{d}y|x)}{F_{Y|X,\theta_0}(\mathrm{d}y|x)} = \frac{\sigma}{\sigma_n(x)}\exp\left\{-\frac{(y-x)^2}{2}\left[\frac{1}{\sigma_n^2(x)} - \frac{1}{\sigma^2}\right]\right\} = 1 + \frac{t_{n\theta_0}(y,x)}{n^{1/2}},$$

with

$$t_{n\theta_0}(y,x) = -n^{1/2}\left[1 - \frac{\sigma}{\sigma_n(x)}\exp\left\{-\frac{(y-x)^2}{2}\left[\frac{1}{\sigma_n^2(x)} - \frac{1}{\sigma^2}\right]\right\}\right].$$

Therefore,

$$t_{n\theta_0}(y,x) \to t_{\theta_0}(y,x) = \gamma(x)\left[\frac{(y-x)^2}{2\sigma^2} - 1\right] \quad \text{as } n\to\infty.$$

To study $\hat{\beta}_n$ under H_{1n} in A3, we may again proceed in steps. To compensate for the deviation from the null model, the expansion of $\hat{G}_n$ under H_{1n} now becomes

$$\sup_{0\leqslant u,v\leqslant 1} |\hat{G}_n(u,v) - \bar{G}_n(u,v) + q_{\theta_0}(u,v)^{\mathrm{T}}(\theta_n - \theta_0) + n^{-1/2}T^1_{\theta_0}(u,v)| = \mathrm{o}_{\mathbb{P}}(n^{-1/2}), \tag{3.1}$$

where

$$T^1_\theta(u,v) = \int_0^u \int_0^v t_\theta(F^{-1}_{Y|X,\theta}(\bar{v}|F^{-1}_X(\bar{u})), F^{-1}_X(\bar{u}))\,\mathrm{d}\bar{v}\,\mathrm{d}\bar{u}.$$

Under contiguous alternatives the expansion A2 of θ_n still continues to hold, but the ℓ_{θ_0}-terms typically are not centered anymore. See Behnen and Neuhaus (1975). This results in the additional shift

$$T^2_\theta(u,v) = q^{\mathrm{T}}_\theta(u,v) \int_0^1 \int_0^1 \bar{\ell}_\theta(\bar{u},\bar{v}) t_\theta(F^{-1}_{Y|X,\theta}(\bar{v}|F^{-1}_X(\bar{u})), F^{-1}_X(\bar{u}))\,\mathrm{d}\bar{v}\,\mathrm{d}\bar{u}.$$

Put

$$T_\theta(u,v) = T^1_\theta(u,v) - T^2_\theta(u,v).$$

Then, under H_{1n}, $\hat{\beta}_n - T_{\theta_0}$ has the same limit as $\hat{\beta}_n$ under H_0. This yields the following result.

Theorem 5. *Under* H_{1n} *and the conditions in Theorem* 1,

$$(\widehat{\mathscr{L}_{\theta_n} \circ \mathscr{L}_0})(\hat{\beta}_n - T_{\theta_0}) \longrightarrow_{\mathrm{d}} B \text{ in the space } D[0,1)^2.$$

The associated shift function T_{θ_0} will be in charge of the local power of the test. Through the additional term T^2_θ it is possible that, though parameters may be known, their estimation increases the power of the test.

It is well known, see Kuelbs (1968), that B has the Kac–Siegert representation:

$$B(u,v) = \sum_{i=1}^\infty \sum_{j=1}^\infty z_{ij} \lambda_{ij}^{1/2} \Phi_{ij}(u,v),$$

where

$$\lambda_{ij} = \frac{16}{[(2i-1)(2j-1)\pi^2]^2}, \quad \Phi_{ij}(u,v) = 2\sin\left[\frac{(2i-1)\pi u}{2}\right] \sin\left[\frac{(2j-1)\pi v}{2}\right]$$

and

$$z_{ij} = \int_0^1 \int_0^1 \frac{B(u,v)\Phi_{ij}(u,v)}{\lambda_{ij}^{1/2}}\,\mathrm{d}u\,\mathrm{d}v, \quad i,j = 1,2,3,\ldots$$

are the principal components of B.

The principal components of $(\widehat{\mathscr{L}_{\theta_n} \circ \mathscr{L}_0})\hat{\beta}_n$ are

$$\hat{z}_{ij} = \int_0^1 \int_0^1 \frac{(\widehat{\mathscr{L}_{\theta_n} \circ \mathscr{L}_0})\hat{\beta}_n(u,v)\Phi_{ij}(u,v)}{\lambda_{ij}^{1/2}}\,\mathrm{d}u\,\mathrm{d}v.$$

Hence, applying the continuous mapping theorem, $\hat{z}_{ij} \rightarrow_{\mathrm{d}} \mathrm{N}(\tau_{ij}, 1)$ under H_{1n} with

$$\tau_{ij} = \int_0^1 \int_0^1 \frac{\left(\mathscr{L}_{\theta_0} \circ \mathscr{L}_0\right) T_{\theta_0}(u,v)\Phi_{ij}(u,v)}{\lambda_{ij}^{1/2}}\,\mathrm{d}u\,\mathrm{d}v.$$

Tests can be based on linear combinations of some $\hat{z}_{ij}$, as has been suggested, in the context of goodness-of-fit testing of marginal distributions, by Durbin et al. (1975). Notice that, under H_{1n}, upon applying

Parseval's Theorem,

$$\hat{C}_n = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[(\widehat{\mathscr{L}}_{\theta_n}\circ\mathscr{L}_0)\hat{\beta}_n\left(\frac{i}{n},\hat{V}_{nj}\right)\right]^2 \rightarrow_{\mathrm{d}} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(z_{ij}+\tau_{ij})^2\lambda_{ij}. \tag{3.2}$$

Conclude that the resulting tests will hardly detect high frequency alternatives, since λ_{ij} will take very small values when i and j become large. See Eubank and La Riccia (1992) for a discussion. This suggests to use Neyman-type test statistics. See Neyman (1937). For this fix m_1 and m_2. Then

$$S_{n,m_1,m_2} = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\hat{z}_{ij}^2 \longrightarrow_{\mathrm{d}} \chi^2_{m_1+m_2}\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\tau_{ij}^2\right) \quad \text{under } \mathrm{H}_{1n},$$

with $\chi^2_m(\Lambda)$ denoting a noncentral chi-square variate with noncentrality parameter Λ. These smooth tests are expected to perform better than those based on the Cramér–von Mises or Kolmogorov–Smirnov criteria in the direction of high frequency alternatives. It is also relevant to find the optimal linear combination of principal components such that the resulting test maximizes the power in the direction of particular contiguous alternatives, along the lines suggested by Schoenfeld (1977, 1980) and Stute (1997). In fact, as it happens with Neyman-type statistics, S_{n,m_1,m_2} can be interpreted as a Lagrange multiplier test for testing that V and U are independent and uniformly distributed in $[0,1]$ in the direction of an exponential density, along the lines of Kallenberg and Ledwina (1999) for a related problem.

Now, under H_{1n},

$$(\widehat{\mathscr{L}}_{\theta_n}\circ\mathscr{L}_0)\hat{\beta}_n \longrightarrow_{\mathrm{d}} M = B + (\mathscr{L}_{\theta_0}\circ\mathscr{L}_0)T_{\theta_0}.$$

M has the spectral representation,

$$M(u,v) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} r_{ij}\lambda_{ij}^{1/2}\Phi_{ij}(u,v),$$

where r_{ij} is distributed as $\mathrm{N}(\tau_{ij},1)$. Conclude that we may consider a test of the hypothesis

$$\bar{\mathrm{H}}_0 : \mathbb{E}[r_{ij}] = 0 \quad \text{all } i,j = 1,2,\ldots,$$

versus

$$\bar{\mathrm{H}}_1 : \mathbb{E}[r_{ij}] = \tau_{ij} \quad \text{some } i,j = 1,2,\ldots.$$

The asymptotic likelihood-ratio test statistic based on $r_{ij}, i = 1,\ldots,m_1,\ j = 1,\ldots,m_2$ is given by

$$\begin{aligned}\Lambda_{m_1m_2} &= \exp\left\{\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\tau_{ij}\left(r_{ij}-\frac{\tau_{ij}}{2}\right)\right\}\\ &= \exp\left\{\int_0^1\int_0^1 \Delta_{m_1m_2}(u,v)\left[M(u,v)-\frac{(\mathscr{L}_{\theta_0}\circ\mathscr{L}_0)T_{\theta_0}(u,v)}{2}\right]\mathrm{d}u\,\mathrm{d}v\right\},\end{aligned}$$

with

$$\Delta_{m_1m_2}(u,v) = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\frac{\tau_{ij}\Phi_{ij}(u,v)}{\lambda_{ij}^{1/2}}.$$

Grenander (1950) showed that if $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\tau_{ij}^2 < \infty$, the most powerful test, at the significance level α, consists of rejecting $\bar{\mathrm{H}}_0$ when

$$\Lambda_\infty > k \quad \text{with } \mathbb{P}(\Lambda_\infty > k) = \alpha.$$

Here

$$\Lambda_\infty = \exp\left\{\int_0^1\int_0^1 \Delta_\infty(u,v)\left[M(u,v)-\frac{(\mathscr{L}_{\theta_0}\circ\mathscr{L}_0)T_{\theta_0}(u,v)}{2}\right]\mathrm{d}u\,\mathrm{d}v\right\}$$

with

$$\Delta_\infty(u,v)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{\tau_{ij}\Phi_{ij}(u,v)}{\lambda_{ij}^{1/2}}.$$

We can use, as a test statistic,

$$\varphi=\frac{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} r_{ij}\cdot\tau_{ij}}{(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\tau_{ij}^2)^{1/2}}
=\frac{\int_0^1\int_0^1 \Delta_\infty(u,v)M(u,v)\,\mathrm{d}u\,\mathrm{d}v}{(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\tau_{ij}^2)^{1/2}}.$$

Then $\varphi\sim\mathrm{N}(0,1)$ under $\bar{\mathrm{H}}_0$. $\bar{\mathrm{H}}_0$ is rejected when

$$\varphi\geqslant c_{1-\alpha},$$

with $c_{1-\alpha}$ denoting the $(1-\alpha)$th quantile of $\mathrm{N}(0,1)$.

In practice, we must estimate τ_{ij}, truncate and rescale the series to come up with an upper one-sided test based on

$$\hat{\varphi}_{n,m_1m_2}=\frac{\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\hat{\tau}_{ij}\cdot\hat{z}_{ij}}{(\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\hat{\tau}_{ij}^2)^{1/2}}\longrightarrow_{\mathrm{d}}\mathrm{N}(0,1)\quad\text{under } \mathrm{H}_0,$$

with m_1 and m_2 fixed integers,

$$\hat{\tau}_{ij}=\int_0^1\int_0^1\frac{(\widehat{\mathscr{L}_{\theta_n}\circ\mathscr{L}_0})\hat{T}_{n\theta_n}(u,v)\Phi_{ij}(u,v)}{\lambda_{ij}^{1/2}}\,\mathrm{d}u\,\mathrm{d}v,$$

$$\hat{T}_{n\theta}(u,v)=\frac{1}{n}\sum_{i=1}^{\lfloor nu\rfloor} t_{n\theta}(Y_{[i:n]},X_{i:n})1_{\{\hat{V}_{n[i:n]}\leqslant v\}}-\hat{q}_\theta(u,v)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\ell_\theta(X_i,Y_i)t_{n\theta}(Y_i,X_i).$$

Again, replacement of the operator $\mathscr{L}_{\theta_0}$ by an estimated operator $\mathscr{L}_{\theta_n}$ does not change the limit. For a given parametric conditional model and a specified alternative an analysis of the components which guarantee high power depends on the model. In the following section we discuss how our method applies for testing conditional normality.

4. Monte Carlo

In this section we apply the Cramér–von Mises test based on $\hat{C}_n$ to test for conditional normality with homoscedastic disturbances, that is,

$$F_{Y|X,\theta}(y|x)=\Phi\Big(\frac{y-x}{\sigma}\Big),$$

with $x=\delta_{00}+\delta_{01}z$, $\theta_0=(\delta_0^{\mathrm{T}},\sigma^2)^{\mathrm{T}}\in\mathbb{R}^2\times\mathbb{R}^+$ and $\delta_0=(\delta_{00},\delta_{01})^{\mathrm{T}}$, where Φ is the standard normal distribution. Conclude that

$$f_{Y|X,\theta}(y|x)=\frac{1}{\sigma}\phi\Big(\frac{y-x}{\sigma}\Big)$$

with ϕ the standard normal probability density function. Therefore,

$$\frac{\partial}{\partial\theta}\ln f_{Y|X,\theta}(y|x)=\frac{1}{\sigma^2}\begin{pmatrix}\frac{1}{2}\Big(\frac{(y-x)^2}{\sigma^2}-1\Big)\\ y-x\\ z(y-x)\end{pmatrix}.$$

Notice that, for all $\theta \in \mathbb{R}^2 \times \mathbb{R}^+$,

$$F^{-1}_{Y|X}(v|x) = x + \sigma \cdot \Phi^{-1}(v).$$

Hence,

$$\frac{\partial}{\partial \theta} \ln f_{Y|X}(F^{-1}_{Y|X}(v|x)|x) = \frac{1}{\sigma^2} \begin{pmatrix} \frac{1}{2}(\Phi^{-1}(v)^2 - 1) \\ \sigma \cdot \Phi^{-1}(v) \\ z \cdot \sigma \cdot \Phi^{-1}(v) \end{pmatrix},$$

which is used for computing q_θ in (1.2). It is immediate that the function φ_θ in (1.2) and hence h_θ in (2.1) does not depend on u. The random variable X is always distributed as $U(0,1)$ with $\sigma = \delta_{00} = \delta_{01} = 1$. Programs were written in double precision FORTRAN 90 and run using a Intel Pentium 4 processor with the

Table 2
Proportion of rejection under $H_0 : Y|X \sim N(Z, \sigma^2)$. $Z = \delta_{00} + \delta_{01} X$

	No estimated parameters	σ^2 Estimated
	$n = 15$	$n = 15$
$\alpha = 0.10$	0.1236	0.1188
$\alpha = 0.05$	0.0646	0.0680
$\alpha = 0.01$	0.0206	0.0240
	$n = 25$	$n = 25$
$\alpha = 0.10$	0.1080	0.1052
$\alpha = 0.05$	0.0578	0.0582
$\alpha = 0.01$	0.0146	0.0142
	$n = 50$	$n = 50$
$\alpha = 0.10$	0.1030	0.1038
$\alpha = 0.05$	0.0522	0.0548
$\alpha = 0.01$	0.0126	0.0132
	$n = 100$	$n = 100$
$\alpha = 0.10$	0.0976	0.1010
$\alpha = 0.05$	0.0506	0.0508
$\alpha = 0.01$	0.0094	0.0100

Table 3
Proportion of rejection under fixed alternative H_1: $Y|X \sim N(Z, 12 \cdot (X - 0.5)^2)$

	No estimated parameters	σ^2 Estimated
	$n = 50$	$n = 50$
$\alpha = 0.10$	0.0950	0.1650
$\alpha = 0.05$	0.0370	0.0814
$\alpha = 0.01$	0.0064	0.0208
	$n = 100$	$n = 100$
$\alpha = 0.10$	0.2038	0.3282
$\alpha = 0.05$	0.0724	0.1834
$\alpha = 0.01$	0.0080	0.0496
	$n = 200$	$n = 200$
$\alpha = 0.10$	0.6426	0.6962
$\alpha = 0.05$	0.2982	0.4722
$\alpha = 0.01$	0.0246	0.1620

Microsoft Developer Studio Compiler, and the IMSL library was used for generating the random numbers (routines DRNUN and DRNNOR), for computing the inverse of the standard normal distribution (routine DNORDF), for numerical integration taking into account possible singularities at the end points (routine DQDAGS). Monte Carlo experiments are based on 5000 simulations.

We have considered sample sizes of $n = 15$, 25, 50 and 100. We report on the percentages of rejection for the cases where (a) θ_0 is completely known and (b) δ_0 is known but σ^2 unknown (and estimated).

The proportion of rejections under H_0 is reported on in Table 2. The attained level is very good, even for small sample sizes like $n = 25$.

Table 3 reports on the proportion of rejections under the alternative hypothesis

$$H_1 : F_{Y|X,\theta}(y|x) = \Phi\left(\frac{y-x}{\sigma(x)}\right) \quad \text{with } \sigma^2(x) = 12\,(z-0.5)^2.$$

Note that $\sigma^2 = \mathbb{E}(Var(Y|X)) = \mathbb{E}(\sigma^2(X)) = 1$, as under H_0.

Acknowledgment

Research funded by the Spanish Ministerio de Educación y Ciencia reference number SEJ2004-04583/ECON.

Appendix A

In the following lemma we analyze the local behavior of the sequential empirical process associated with the concomitants of the V_i's. For this, define for $0 \leqslant u, v \leqslant 1$ and real $\kappa_1, \kappa_2, \ldots, \kappa_n$,

$$\beta_n^0(u, v, \kappa_1, \ldots, \kappa_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nu \rfloor} [1_{\{V_{[i:n]} \leqslant v + \kappa_i n^{-1/2}\}} - 1_{\{V_{[i:n]} \leqslant v\}} - \kappa_i n^{-1/2}].$$

We shall see that β_n^0 converges to zero uniformly in (u, v) and $\kappa_1, \ldots, \kappa_n$, as long as the κ_i range in a compact interval.

Lemma 1. *For each finite K, as* $n \to \infty$,

$$\sup_{\substack{0 \leqslant u,v \leqslant 1 \\ \{|\kappa_i| \leqslant K, i=1,\ldots,n\}}} |\beta_n^0(u, v, \kappa_1, \ldots, \kappa_n)| = o_{\mathbb{P}}(1).$$

Proof. The proof follows standard arguments when dealing with residual empirical processes, see e.g., Koul (2002) monograph. For fixed u, v and $\kappa_i, i = 1, \ldots, n$, the assertion is trivial. Just observe that the concomitants are independent and identically distributed as a U(0, 1) random variable. Obviously β_n^0 converges to zero in squared mean and hence in probability. For a given sequence $\kappa_1, \kappa_2, \ldots$, β_n^0 is also tight in (u, v), since it is only a variation of a time-sequential empirical process, which is well known to be tight. In order to get uniformity in κ, use monotonicity of the indicators, decompose the interval $[-K, K]$ into small subintervals and reduce the analysis, up to a small error, to a finite grid. Since this is standard, details are omitted. □

Proof of Theorem 1. Since

$$\hat{V}_{ni} = F_{Y|X,\theta_n}(F_{Y|X}^{-1}(V_i|X_i)|X_i),$$

we have, by continuity,

$$1_{\{\hat{V}_{ni} \leqslant v\}} = 1_{\{V_i \leqslant F_{Y|X}(F_{Y|X,\theta_n}^{-1}(v|X_i)|X_i)\}}.$$

Applying a mean value theorem argument, for $1 \le i \le n$,

$$F_{Y|X}(F_{Y|X,\theta_n}^{-1}(v|X_i)|X_i)$$
$$= v + (\theta_0 - \theta_n)^{\mathrm{T}} \frac{\partial}{\partial\theta} F_{Y|X,\theta}(F_{Y|X,\theta_n}^{-1}(v|X_i)|X_i)\Big|_{\theta=\theta_{ni}^*}, \tag{A.1}$$

$i=1,2,\ldots$, where $\|\theta^*_{ni}-\theta_0\|\leqslant\|\theta_n-\theta_0\|$. Since $\partial F_{Y|X,\theta}(F^{-1}_{Y|X,\theta_n}(v|X_i)|X_i)/\partial\theta$ is bounded in a neighborhood of θ_0, and since $\theta_n=\theta_0+\mathrm{O}_{\mathbb{P}}(n^{-1/2})$, (A.1) implies that

$$F_{Y|X}(F^{-1}_{Y|X,\theta_n}(v|X_i)|X_i)=v+\kappa_i n^{-1/2},\quad i=1,2,\ldots,$$

where with large probability κ_i ranges in a possibly large but compact set. Hence, from Lemma 1, we obtain uniformly in $(u,v)\in[0,1]^2$ that, up to a remainder $\mathrm{o}_{\mathbb{P}}(1)$,

$$\hat{\beta}_n(u,v)=\beta_n(u,v)-n^{1/2}(\theta_n-\theta_0)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{\lfloor nu\rfloor}\frac{\partial}{\partial\theta}F_{Y|X,\theta}(F^{-1}_{Y|X,\theta_n}(v|X_{i:n})|X_{i:n})\bigg|_{\theta=\theta^*_{ni}}.$$

The result now follows from the assumed continuity of

$$\partial F_{Y|X,\theta}(F^{-1}_{Y|X,\theta_n}(v|X_i)|X_i)/\partial\theta,$$

the consistency of θ_n, and the uniform convergence of the involved empirical integrals. □

Proof of Theorem 2. Compared with the previous proof, we now have

$$\hat{V}_{ni}=F_{Y|X,\theta_n}(Y_i|\tilde{X}_{ni}),$$

with $\tilde{X}_{ni}=Z_i^{\mathrm{T}}\delta_n$ and $Y_i=F^{-1}_{Y|X,\theta_0}(V_i|Z_i^{\mathrm{T}}\delta_0)$. Hence

$$1_{\{\hat{V}_{ni}\leqslant v\}}=1_{\{V_i\leqslant F_{Y|X,\theta_0}(F^{-1}_{Y|X,\theta_n}(v|\tilde{X}_{ni})|X_i)\}}.$$

But

$$\begin{aligned}&F_{Y|X,\theta_0}(F^{-1}_{Y|X,\theta_n}(v|\tilde{X}_{ni})|X_i)\\&\quad=v+(\theta_0-\theta_n)^{\mathrm{T}}\frac{\partial}{\partial\theta}F_{Y|X,\theta}(F^{-1}_{Y|X,\theta_n}(v|\tilde{X}_{ni})|Z_i^{\mathrm{T}}\delta_0)_{\theta=\theta^*_{ni}}\\&\qquad+(\delta_0-\delta_n)^{\mathrm{T}}Z_i\frac{\partial}{\partial x}F_{Y|X,\theta_n}(F_{Y|X,\theta_n}(v|\tilde{X}_{ni})|x)_{x=x^*_{ni}},\end{aligned}$$

where x^*_{ni} is between $Z_i^{\mathrm{T}}\delta_n$ and $Z_i^{\mathrm{T}}\delta_0$. If we sum these terms up for the first $\lfloor nu\rfloor$ ordered $X_i=Z_i^{\mathrm{T}}\delta_0$, note that in probability and uniformly in $0\leqslant u,v\leqslant 1$:

$$\frac{1}{n}\sum_{i=1}^{\lfloor nu\rfloor}\frac{\partial}{\partial\theta}F_{Y|X,\theta}(F^{-1}_{Y|X,\theta_n}(v|\tilde{X}_{[i:n]})|X_{i:n})_{\theta=\theta_n}\to q_{\theta_0,\delta_0}(u,v),$$

$$\frac{1}{n}\sum_{i=1}^{n}1_{\{\tilde{F}_{Xn}(\tilde{X}_{ni})\leqslant u\}}Z_i\frac{\partial}{\partial x}F_{Y|X,\theta_n}(F^{-1}_{Y|X,\theta_n}(v|\tilde{X}_{ni})|x^*_{ni})\to q^1_{\theta_0,\delta_0}(u,v),$$

where $\tilde{F}_{Xn}$ is the sample distribution of $\tilde{X}_{ni}$, $i\geqslant 1$. Actually, this follows from the continuity of the involved functions, upon noticing that because of the $n^{1/2}$-consistency of δ_n and the fact that Z has finite second moments we have

$$\max_{1\leqslant i\leqslant n}Z_i^{\mathrm{T}}(\delta_n-\delta_0)=\mathrm{o}_{\mathbb{P}}(1).\qquad\square$$

Proof of Theorem 3. It follows from Corollary 1 that $\hat{\beta}_n$ is tight. It is then not difficult to show that also $(\mathscr{L}_{\theta_0}\circ\mathscr{L}_0)\hat{\beta}_n$ is tight. Since also the finite dimensional distributions converge, it suffices to show that in distribution $(\mathscr{L}_{\theta_0}\circ\mathscr{L}_0)\hat{\beta}_\infty$ equals a Brownian sheet. First, the operator $\mathscr{L}_{\theta_0}\circ\mathscr{L}_0$ is linear so that the limit is a centered Gaussian process. Check the covariance structure to get the assertion of the theorem. See also Khmaladze (1988, 1993) or Lemma 3.1 in Stute et al. (1998) for related arguments. □

Proof of Theorem 4. To prove Theorem 4 it suffices to show that

$$(\hat{\mathscr{L}}_{\theta_n}\circ\mathscr{L}_0)\hat{\beta}_n-(\mathscr{L}_{\theta_0}\circ\mathscr{L}_0)\hat{\beta}_n\to 0 \text{ in probability.}$$

This may be proved along the lines of Stute et al. (1998), where similar things have been done in the context of model checks in regression. □

Proof of Theorem 5. We already pointed out that Theorem 5 is a consequence of the expansion (3.1) and the central limit theorem under contiguous alternatives due to Behnen and Neuhaus (1975). To show (3.1), recall

$$F_{Y|X}(\mathrm{d}y|x) = (1 + n^{-1/2} t_{n\theta_0}(y, x)) F_{Y|X,\theta_0}(\mathrm{d}y|x).$$

Hence, compared to the proof of Theorem 1, we have to add another term, namely

$$n^{-1/2} \int_{-\infty}^{F^{-1}_{Y|X,\theta_n}(v|X_i)} t_{n\theta_0}(y, X_i) F_{Y|X,\theta_0}(\mathrm{d}y|X_i),$$

to the right-hand side of (A.1). Summation over the first $\lfloor nu \rfloor$ X-order statistics and using a continuity argument as well as assumption A3 yield the representation (3.1) and hence the assertion of Theorem 5. $\square$

References

Andrews, D.W.K., 1997. A conditional Kolmogorov test. Econometrica 65, 1097–1128.
Bai, J., 1994. Weak convergence of sequential empirical processes of residuals in ARMA models. Annals of Statistics 22, 2051–2061.
Bai, J., 1996. Testing for parameter constancy in linear regressions: an empirical distribution function approach. Econometrica 64, 597–622.
Bai, J., 2003. Testing parametric conditional distributions of dynamic models. The Review of Economics and Statistics 85, 531–549.
Behnen, K., Neuhaus, G., 1975. A central limit theorem under contiguous alternatives. Annals of Statistics 3, 1349–1353.
Bickel, P., Wichura, M., 1971. Convergence criteria for multiparameter stochastic processes. Annals of Mathematical Statistics 42, 1656–1670.
Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society Series B 37, 149–192.
Brownrigg, R.D., 2005. Tables of distribution functions of suprema of Brownian Motion on a line or in 2-space. At ⟨http://www.mcs.vuw.ac.nz/ray/Brownian/⟩.
Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis on Count Data. Cambridge University Press, Cambridge.
Durbin, J., 1973. Weak convergence of the sample distribution function when parameters are estimated. Annals of Statistics 1, 279–290.
Durbin, J., Knott, M., Taylor, C.C., 1975. Components of Cramér–von Mises statistics II. Journal of the Royal Statistical Society Series B 37, 216–237.
Eubank, R.L., La Riccia, V.N., 1992. Asymptotic comparison of Cramér–von Mises and nonparametric function techniques for testing goodness-of-fit. Annals of Statistics 20, 2071–2086.
Gikhman, I.I., 1953. Some remarks on A. Kolmogorov's goodness of fit test. Dokladi Akademii Nauk 91, 715–718 (in Russian).
Grenander, U., 1950. Stochastic processes and statistical inference. Arkiv för Matematik 1, 195–277.
Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.
Kac, M., Kiefer, J., Wolfowitz, J., 1955. On tests of normality and other goodness of fit based on distance methods. Annals of Mathematical Statistics 26, 189–211.
Kallenberg, C.M., Ledwina, T., 1999. Data-driven rank tests for independence. Journal of the American Statistical Association 94, 285–301.
Khmaladze, E.V., 1981. Martingale approach to the goodness of fit tests. Theory of Probabilities and Applications 26, 246–265.
Khmaladze, E.V., 1988. An innovation approach in goodness-of-fit tests in $\mathbb{R}^m$. Annals of Statistics 16, 1503–1516.
Khmaladze, E.V., 1993. Goodness of fit problem and scanning innovation martingales. Annals of Statistics 21, 798–829.
Khmaladze, E.V., Koul, H.L., 2004. Martingale transforms goodness-of-fit tests in regression models. Annals of Statistics 32, 995–1034.
Koul, H., 2002. Weighted Empirical Processes in Dynamic Nonlinear Models. Springer, New York.
Koul, H., Stute, W., 1999. Nonparametric model checks for time series. Annals of Statistics 27, 204–236.
Kuelbs, J., 1968. The invariance principle for a lattice of random variables. Annals of Mathematical Statistics 39, 382–389.
Lancaster, T., 1990. The Econometric Analysis of Transition Data. Cambridge University Press, Cambridge.
Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge.
Neuhaus, G., 1971. On weak convergence of stochastic processes with multi-dimensional time parameter. Annals of Mathematical Statistics 42, 1285–1295.
Neuhaus, G., 1973. Asymptotic properties of the Cramér–von Mises statistic when parameters are estimated. In: Proceedings of the Prague Symposium on Asymptotic Statistics, vol. 2, pp. 257–297.
Neuhaus, G., 1976. Weak convergence under contiguous alternatives of the empirical process when parameters are estimated: the D_k approach. Lecture Notes in Mathematics, vol. 566. Springer, Berlin, pp. 68–82.
Neyman, J., 1937. Smooth tests for goodness of fit. Skandinavan Aktuarietidskrijt 20, 149–199.
Nikabadze, A., 1997. Scanning innovations and goodness of fit tests for vector random variables against the general alternative. A. Razmadze Mathematical Institute, Tbilisi, Preprint.
Nikabadze, A., Stute, W., 1997. Model checks under random censorship. Statistics and Probability Letters 32, 249–259.
Pollard, D., 1984. Convergence of Stochastic Processes. Springer, New York, Berlin.
Rosenblatt, M., 1952. Remarks on a multivariate transformation. Annals of Mathematical Statistics 23, 470–472.

Schoenfeld, D.A., 1977. Asymptotic properties of tests based on linear combinations of the orthogonal components of the Cramér–von Mises statistic. Annals of Statistics 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. Annals of Statistics 8, 1017–1022.
Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.
Straf, M.L., 1971. Weak convergence of stochastic processes with several parameters. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2. University of California Press, Berkeley, CA, pp. 187–221.
Stute, W., 1997. Nonparametric model checks for regression. Annals of Statistics 25, 613–641.
Stute, W., Zhu, L.X., 2002. Model checks for generalized linear models. Scandinavan Journal of Statistics 29, 535–545.
Stute, W., Thies, S., Zhu, L.X., 1998. Model checks for regression: an innovation process approach. Annals of Statistics 26, 1916–1934.
Zheng, J.X., 2000. A conditional test of conditional parametric distributions. Econometric Theory 16, 667–691.