Test (2013) 22:361–411
DOI 10.1007/s11749-013-0327-5

INVITED PAPER

An updated review of Goodness-of-Fit tests for regression models

Wenceslao González-Manteiga · Rosa M. Crujeiras

Published online: 25 July 2013

Abstract This survey intends to collect the developments on Goodness-of-Fit for regression models during the last 20 years, from the very first origins with the proposals based on the idea of the tests for density and distribution, until the most recent advances for complex data and models. Far from being exhaustive, the contents in this paper are focused on two main classes of tests statistics: smoothing-based tests (kernel-based) and tests based on empirical regression processes, although other tests based on Maximum Likelihood ideas will be also considered. Starting from the simplest case of testing a parametric family for the regression curves, the contributions in this field provide also testing procedures in semiparametric, nonparametric, and functional models, dealing also with more complex settings, as those ones involving dependent or incomplete data.

Keywords Bootstrap calibration · Empirical distribution of the residuals · Empirical regression process · Likelihood ratio tests · Smoothing tests

Mathematics Subject Classification 62G08 · 62G09 · 62G10

This invited paper is discussed in the comments available at doi:10.1007/s11749-013-0328-4, doi:10.1007/s11749-013-0329-3, doi:10.1007/s11749-013-0330-x, doi:10.1007/s11749-013-0331-9, doi:10.1007/s11749-013-0332-8, doi:10.1007/s11749-013-0333-7.

W. González-Manteiga (✉) · R.M. Crujeiras
Dpt. of Statistics and Operations Research, Faculty of Mathematics, University of Santiago de Compostela, Santiago de Compostela, Spain
e-mail: wenceslao.gonzalez@usc.es

R.M. Crujeiras
e-mail: rosa.crujeiras@usc.es

1 Introduction. From density to regression

With the aim of testing if a data distribution belongs to a certain parametric family, Pearson introduced at the beginning of the twentieth century the term Goodness-of-Fit (GoF). Since then, there has been an enormous amount of papers on this topic. As an example, MathScinet database reported in October 2012 approximately 3000 references from the search by *Goodness-of-Fit*, referring the oldest entries to GoF for distribution models. On the other hand, the most recent papers on this topic consider much more complicated settings, such as diffusion models or multidimensional covariance structures. This simple notice states the impossibility of condensing in a single paper an exhaustive review of all the contributions in this field.

More recently, since the early 1990s, there has been also a large amount of contributions concerning GoF for regression models, which will be the focus of this work. Another search in MathScinet provides approximately 250 references on this topic, showing the interest that has been aroused within the statistical community in the last 20 years.

An up-to-date review of the most important recent contributions on GoF for regression models will be given in this paper. The final goal is to show the clear connection between the developments of GoF for regression models with the previous ideas on GoF for density or distribution, jointly with the possible extensions and applications to different contexts of increasing complexity, both regarding the data characteristics and/or the model assumptions. As will be soon noticed by the reader, and all along this paper, smoothing is crucial in most of the methodological developments, and the authors are concerned about kernel smoothing methods although other smoothing techniques could also be applied. Other testing procedures such as the ones considering empirical regression processes and related tests based on likelihood functions and residuals distributions will also be commented on.

In this introductory section, some basic ideas of classical tests for density and distribution functions will be presented, jointly with the first works on GoF for regression models, starting from the 1990s. Yet in these early works, two main approaches to the GoF testing problem can be considered: a first approach is based on smoothing methods for regression, whereas the second one considers the construction of empirical regression processes. Section 2 explores some more recent perspectives for GoF testing in regression models, introduced in the beginning of this century, collecting likelihood-based tests, tests based on the empirical distribution of the residuals, and also tests designed for avoiding the curse of dimensionality. Test calibration in practice and power analysis is briefly commented on in Sect. 3, and GoF tests in more complex semiparametric and nonparametric models are introduced in Sect. 4. Temporal and/or spatial correlation structures must be taking into account in order to adapt GoF tests to dependent data, and this issue is considered in Sect. 5. GoF tests with censored, truncated, and biased data and data with measurement errors are revised in Sect. 6, whereas the methodological related topic of comparison of regression curves is commented on in Sect. 7. Finally, some very recent advances on GoF testing are presented in Sect. 8, concerning models with random effects, quantile, and functional regression.

1.1 Distribution and density tests

The basic and fundamental ideas, in the roots of the recent developments on GoF theory for regression models, are limited to the comparison of a nonparametric pilot estimator of the distribution F or the density f of a certain random variable (r.v.) X, with a consistent (under the null hypothesis) estimator of the target function. The pilot estimator is usually given by the empirical cumulative distribution function, for the distribution F, and by a kernel density estimator for f. In 1973, two essential contributions for further methodological developments are published: the paper by Durbin (1973) on the distribution and the one by Bickel and Rosenblatt (1973) on the density.

The general statement of these tests is the following. Given a random sample $\{X_1, \ldots, X_n\}$ of a r.v. X, the goal is to test the following hypothesis in an omnibus way:

$$H_0 : F \in \mathcal{F}_{\text{dist}} = \{F_\theta\}_{\theta \in \Theta \subset \mathbb{R}^q}, \quad \text{vs.} \quad H_a : F \notin \mathcal{F}_{\text{dist}},$$

for the distribution function or, for the density:

$$H_0 : F \in \mathcal{F}_{\text{dens}} = \{f_\theta\}_{\theta \in \Theta \subset \mathbb{R}^q}, \quad \text{vs.} \quad H_a : F \notin \mathcal{F}_{\text{dens}},$$

assuming obviously that these functions exist. Formally, the test statistics are usually based on a discrepancy between the pilot estimator, which is universally consistent, and the corresponding consistent estimator under the null hypothesis H_0. Hence, for the distribution case, the test statistic can be written in generic form as

$$T_n = T(F_n, F_{\widehat{\theta}}) \equiv T(\alpha_n), \tag{1}$$

where $F_n(x) = n^{-1}\text{card}\{j; X_j \leq x\}$ is the empirical cumulative distribution function (where card denotes the cardinality of a set) and $F_{\widehat{\theta}}$ is a parametric estimator under H_0, $\widehat{\theta}$ being a $\sqrt{n}$-consistent estimator of $\theta_0 \in \Theta$, the true parameter under H_0. Expression (1) can also be written in terms of α_n, which denotes an empirical process with estimated parameter $\widehat{\theta}$. Specifically, $\alpha_n(\cdot) = \sqrt{n}(F_n(\cdot) - F_{\widehat{\theta}}(\cdot))$. Weak convergence of α_n, studied in detail by Durbin (1973), is the key for deriving the asymptotic behavior of any continuous functional of this process. For instance, $T_n = \sup_x |\alpha_n(x)| = \|\alpha_n(\cdot)\|_\infty$ (where $\|\cdot\|_\infty$ denotes the supremum norm), corresponding to the Kolmogorov–Smirnov test, or $T_n = \int \alpha_n^2(x)\, dF_n(x)$, the well-known Cramér–von Mises test. In general, the asymptotic behavior of the tests $T(\alpha_n)$ is determined by the continuous functional operating on a Gaussian limit process, denoted by α (see Durbin 1973 for further details).

The test statistic for a density can be written as $T_n = T(f_{nh}, f_{\widehat{\theta}})$, or more exactly as

$$T_n = T\big(f_{nh}, \mathbb{E}_{\widehat{\theta}}(f_{nh})\big) \equiv T(\widetilde{\alpha_n}) \tag{2}$$

where $f_{nh}(x) = n^{-1}\sum_{i=1}^n K_h(x - X_i)$ is the kernel density estimator (see Rosenblatt 1956 and Parzen 1962), with K a kernel density function and h the smoothing parameter or bandwidth. K_h denotes the rescaled kernel, which in the one-dimensional case is $K_h(x) = h^{-1}K(x/h)$. Considering $\mathbb{E}_{\widehat{\theta}}(f_{nh})$ in (2), with $\mathbb{E}_\theta(f_{nh}(x)) = \int K_h(x - u)\,dF_\theta(u)$ instead of $f_{\widehat{\theta}}$, avoids the bias inherent to nonparametric density estimation.

Similar to the distribution case (1), the test statistic in (2) can be expressed as $T_n \equiv T(\widetilde{\alpha_n})$ where $\widetilde{\alpha_n}(\cdot) = \sqrt{nh}(f_{nh}(\cdot) - \mathbb{E}_{\widehat{\theta}}(f_{nh}(\cdot)))$ is the empirical process associated to the density whose limit process $\widetilde{\alpha}$ is also Gaussian (see Rosenblatt 1991). Unlike the limit process α for the distribution case, the covariance structure of $\widetilde{\alpha}$ is independent of the distribution of $\widehat{\theta}$. This alternative route for GoF is related to the seminal paper by Bickel and Rosenblatt (1973), whose ideas were extended to the p-dimensional setting in the 1990s. Specifically, an L^2-test in dimension p in given by $T_n \equiv T(\widetilde{\alpha_n}) = \int \widetilde{\alpha_n}^2(x)\omega(x)\,dx$, with ω a weight function to mitigate edge-effects and, in this case, $f_{nh}(x)$ a p-dimensional kernel density estimator (with $K_h(x) = h^{-p}K(x/h)$ a p-dimensional rescaled kernel) and a rate $\sqrt{nh^p}$ in the definition of the empirical process. Then, it can be seen that

$$h^{-p/2}\left(T_n - \int K^2(x)\,dx \int f(x)\omega(x)\,dx\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, 2\int (K*K)^2(x)\,dx \int f^2(x)\omega^2(x)\,dx\right)$$

($\xrightarrow{d}$ denoting convergence in distribution) where $K*K$ denotes the self-convolution of the kernel, $h \equiv h_n \to 0$ and $nh^p \to \infty$ (see Fan 1994, 1998). This convergence result can be derived using arguments from continuous functionals on Gaussian processes, or Central Limit Theorems (CLT) for U-statistics with kernels varying with n (see, for instance, de Jong 1987). Analogous CLT structures to the one above will appear along this paper for GoF tests based on smooth estimators.

Bickel and Rosenblatt test has been the subject for a large collection of statistical papers, adjusting the methodology for different data contexts or exploring other functionals beyond the L^2 distance. The adaptation of the test for multivariate probability density functions was studied by Ahmad and Cerrito (1993) while Gouriéroux and Tenreiro (2001) and Chebana (2004) derived its asymptotic power properties, under local alternatives. Location-scale invariant GoF tests are considered by Tenreiro (2007), for multidimensional random vectors, whereas test for assessing normality were studied by Tenreiro (2009). The Bickel and Rosenblatt test has also been adapted for diffusion processes by Lee (2006), showing also asymptotic normality. The law of the iterated logarithm for T_n was discussed by Liang and Jing (2007) under fixed alternatives, motivated by the previous work by Giné and Mason (2004). For weakly dependent observations, Neumann and Paparoditis (2000) modified T_n including a parametric estimate of the stationary density, using bootstrap for calibration. The asymptotic distribution of such a test statistic is derived by Lee and Na (2002) for autoregressive models. The initial distributional results on T_n were extended by Bachmann and Dette (2005) for fixed alternatives, explaining also the asymptotic behavior of the tests statistic proposed by Lee and Na (2002) in the autoregressive setting. From a different perspective, Chebana (2006) established the functional asymptotic normality of T_n as a process indexed on a family of weight functions, providing a finite-dimensional limit law and a result guaranteeing the stochastic equicontinuity of the process. The discrepancy between the nonparametric estimator and the corresponding density under the null hypothesis may be measured with an L^1 distance, as proposed by Cao and Lugosi (2005) and Albers and Schaafsma (2008). Liero et al. (1998) derived the behavior of a test based on the supremum norm, and compare its power behavior under Pitman alternatives with the classical test. Finally, inspired by the previous ideas, Fermanian (2005) proposed distribution free GoF for copulas, investigating their asymptotic distribution, and Bücher and Dette (2010) derived the asymptotic properties under fixed alternatives.

1.2 Basic ideas on GoF tests for regression

The ideas of GoF for density and distribution have been naturally extended in the 1990s to regression models. Considering as a reference a regression model with random design $Y = m(X) + \varepsilon$, $\{(X_i, Y_i)\}_{i=1}^n$ being a random sample of $(X, Y) \in \mathbb{R}^{p+1}$ (that is, (X_i, Y_i) independent and identically distributed (i.i.d.) as (X, Y)), the goal is to test:

$$H_0 : m \in \mathcal{M}_\theta = \{m_\theta\}_{\theta \in \Theta \subset \mathbb{R}^q}, \quad \text{vs.} \quad H_a : m \notin \mathcal{M}_\theta,$$

where $m(x) = \mathbb{E}(Y|X = x)$ is the regression function of Y over X, with $\mathbb{E}(\varepsilon|X) = 0$. In the regression context, apart from the target function m, there are usually some nuisance functions such as the conditional variance $\sigma^2(x) = \mathrm{Var}(Y|X = x)$ or the density of the explanatory variable X, namely f, playing a role in the test statistics distribution.

Smoothing-based tests Although there exist a large variety of smoothing methods for regression models, this review will mainly consider kernel type estimators such as the Nadaraya–Watson estimator (Nadaraya 1964; Watson 1964), given by

$$m_{nh}(x) = \sum_{i=1}^n W_{ni}(x) Y_i, \quad \text{with } W_{ni}(x) = \frac{K_h(x - X_i)}{\sum_{j=1}^n K_h(x - X_j)}, \ i = 1, \ldots, n.$$

With the same spirit as for the previous tests for the p-dimensional density function, the empirical process for the p-dimensional regression problem (in the sense that the dimension of the explanatory variable is p) is given by

$$\begin{aligned} \overline{\alpha_n}(x) &= \sqrt{nh^p}\big(m_{nh}(x) - \mathbb{E}_{\widehat{\theta}}\big(m_{nh}(x)\big)\big) \\ &= \sqrt{nh^p} \sum_{i=1}^n W_{ni}(x)\big(Y_i - m_{\widehat{\theta}}(X_i)\big) \\ &= \sqrt{nh^p} \sum_{i=1}^n W_{ni}(x) \widehat{\varepsilon}_{i0}, \end{aligned}$$

which can be interpreted as a smoothing over the residuals $\{\widehat{\varepsilon}_{i0}\}_{i=1}^n$, with $\widehat{\varepsilon}_{i0} = Y_i - m_{\widehat{\theta}}(X_i)$, and providing $\mathbb{E}_{\widehat{\theta}}$ an estimate of $\mathbb{E}_{\theta_0}$, θ_0 being the true parameter under H_0 and $\widehat{\theta}$ a $\sqrt{n}$-consistent estimator of θ_0, such as the one obtained by least squares or maximum likelihood for Gaussian errors.

Again, a general test based on $\overline{\alpha_n}$ can be devised by applying a continuous functional on the empirical process, such as

$$T_n = \int \overline{\alpha_n}^2(x)\omega(x)\,dx. \tag{3}$$

The limit distribution of T_n, under some regularity conditions, can be obtained from empirical process theory or using the aforementioned results by de Jong (1987):

$$\begin{aligned} &h^{-p/2}\left(T_n - \int K^2(x)\,dx \int \frac{\sigma^2(x)\omega(x)}{f(x)}\,dx\right) \\ &\quad \xrightarrow{d} \mathcal{N}\left(0, 2\int (K*K)^2(x)\,dx \int \frac{\sigma^4(x)\omega^2(x)}{f^2(x)}\,dx\right). \end{aligned} \tag{4}$$

Denoting by T_{1n} the test statistic introduced by Härdle and Mammen (1993)

$$T_{1n} = \int \left(m_{nh}(x) - m_{nh}(x,\widehat{\theta})\right)^2\omega(x)\,dx, \tag{5}$$

the test statistic T_n in (3) can be written as follows:

$$T_n = nh^p T_{1n} = nh^p \int \left(m_{nh}(x) - m_{nh}(x,\widehat{\theta})\right)^2\omega(x)\,dx, \quad \text{with}$$

$$m_{nh}(x,\widehat{\theta}) = \sum_{i=1}^n W_{ni}(x) m_{\widehat{\theta}}(X_i)$$

the data smoother under H_0. A discretized version of this test statistic can be found in González-Manteiga and Cao (1993). In terms of T_{1n}, the limit distribution result in (4) can be rewritten as

$$\begin{aligned} &nh^{p/2}\left(T_{1n} - \left(nh^p\right)^{-1} \int K^2(x)\,dx \int \frac{\sigma^2(x)\omega(x)}{f(x)}\,dx\right) \\ &\quad \xrightarrow{d} \mathcal{N}\left(0, 2\int (K*K)^2(x)\,dx \int \frac{\sigma^4(x)\omega^2(x)}{f^2(x)}\,dx\right). \end{aligned} \tag{6}$$

Moreover, the discretized version of T_{1n}, that is, $T_{1n}^D = n^{-1}\sum_{i=1}^n (m_{nh}(X_i) - m_{nh}(X_i,\widehat{\theta}))^2\omega(X_i)$, estimates consistently $\mathbb{E}(\mathbb{E}^2(\varepsilon_0|X)\omega(X))$ which is null under H_0, with $\varepsilon_0 = Y - m_{\theta_0}(X)$.

Alternatively, as can be seen in Zhang and Dette (2004), it is possible to define other test statistics based on consistent estimators of different characteristics of the null hypothesis to test. For instance, by estimating $\mathbb{E}(\varepsilon_0\mathbb{E}(\varepsilon_0|X)f(X)\omega(X))$ which is null nuder H_0. Except for a negligible bias term, a natural estimate for this quantity is given by

$$T_{2n} = \frac{1}{n(n-1)} \sum_{i\neq j} K_h(X_i - X_j)\left(Y_i - m_{\widehat{\theta}}(X_i)\right)\left(Y_j - m_{\widehat{\theta}}(X_j)\right)\omega(X_i), \tag{7}$$

the test statistic introduced by Zheng (1996), which presents no asymptotic bias. A recent extension to the case where covariates may have a discrete component can be seen in Hsiao et al. (2007).

Also $\mathbb{E}([\varepsilon_0^2 - (\varepsilon_0 - \mathbb{E}(\varepsilon_0|X))^2]\omega(X))$ is null under H_0 and an estimate for this quantity is

$$T_{3n} = \frac{1}{n}\sum_{i=1}^{n}\big(Y_i - m_{\widehat{\theta}}(X_i)\big)^2\omega(X_i) - \frac{1}{n}\sum_{i=1}^{n}\big(Y_i - m_{nh}(X_i)\big)^2\omega(X_i), \tag{8}$$

based on the differences of the error variance estimates in the regression model, introduced by Dette (1999) and closely related to the generalized likelihood ratio test (see Fan et al. 2001 and Fan and Jiang 2007) to be revised later.

Asymptotic distributions of both T_{2n} and T_{3n} follow a similar architecture to T_{1n} in (6). Precisely, for T_{2n}:

$$nh^{p/2}T_{2n} \xrightarrow{d} \mathcal{N}\left(0, 2\int K^2(x)\,dx \int \sigma^4(x) f^2(x)\omega^2(x)\,dx\right) \tag{9}$$

and for the variance difference statistic T_{3n}, with $K^{2*} = 2K - K*K$, the asymptotic distribution is

$$\begin{aligned} &nh^{p/2}\left(T_{3n} - \big(nh^p\big)^{-1}K^{2*}(0)\int \sigma^2(x)\omega(x)\,dx\right) \\ &\quad \xrightarrow{d} \mathcal{N}\left(0, 2\int K^{2*}(x)\,dx \int \sigma^4(x)\omega^2(x)\,dx\right). \end{aligned} \tag{10}$$

Although other tests could be chosen in this preliminary section, it should be noticed that T_{1n}, as well as T_{2n} and T_{3n} are naturally motivated from conditions characterizing the null hypothesis.

The test proposed by Härdle and Mammen (1993) owns a large collection of variants, with different smooth estimators for the regression function, with alternative estimators under the null hypothesis or considering other discrepancy measures. For instance, Kozek (1991) presented a supremum-norm-based test comparing the nonparametric kernel estimate of a regression function with the corresponding least-squares estimator, under the null parametric hypothesis, proving the consistency of such a test and Koul and Ni (2004) studied a class of minimum distance tests for multidimensional covariates and heteroscedatiscity. Alcalá et al. (1999) checked a parametric null hypothesis using local polynomial regression (see Fan and Gijbels 1996 for a survey on local polynomial modeling). Also based on local polynomial fitting, Liu et al. (2000) introduced a consistent model especification test and compared its performance with simple kernel regression estimators. Stute and González-Manteiga (1996) proposed a test for linearity based on the comparison of a nearest neighbor estimator and a parametric estimator of m. Li (2005) assessed the lack of fit of a nonlinear regression model, comparing the local linear kernel and parametric fits. Resembling the classical theory for the F test, Raz (1990) proposed an approximation to a permutation test where a kernel estimator is used. Müller (1992) compared a nonparametric fit with the least squares estimator by a kernel approach, for fixed design regression models. Diagnostic tests have been proposed by Staniswalis and Severini (1991), comparing the fitted models in a collection of points, and by Samarov (1993), by means integral type functionals. A test for parametric nonlinearities has been also studied by Wooldridge (1992).

The use of the asymptotic distributions in (6), (9) or (10) for testing in practice, entails selecting the smoothing parameter h, a broadly studied problem in regression estimation but with serious gaps for testing problems. Hart (1997), Fan et al. (2001), and Eubank et al. (2005) gave some strategies on bandwidth selection. Although the literature is relatively scarce on this topic, it is worth citing the works by Kulasekera and Wang (1997), Zhang (2003, 2004) or more recently, Gao and Gijbels (2008). Added to this difficulty, there is also the need to estimate nonparametrically the nuisance functions, and also the slow rate of convergence to the Gaussian limit distribution.

Although this paper is focused on nonparametric kernel regression estimators, different pilot estimator such as splines, wavelets or orthogonal expansions can also be used for GoF tests. Eubank and Hart (1992) studied the large-sample properties of GoF for linearity, considering cubic smoothing splines over regression residuals whereas Hart and Wehrly (1992) introduced smoothing splines to deal with the edge-effect phenomenon in nonparametric regression and proposed a test for the adequacy of a polynomial regression function. Eubank and Hart (1993) also considered smoothing splines for comparing the performance of different type of tests. Eubank and LaRiccia (1993) proposed different tests based on weighted sums of Fourier coefficients and investigated their asymptotic properties.

It should also be mentioned the book by Hart (1997), which collects a survey on the use of nonparametric smoothing methods for testing the fit of a parametric model.

Tests based on empirical regression processes An alternative methodology for avoiding the selection of a smoothness parameter, inspired by the GoF methods for distributions, is based on the empirical estimator of the integrated regression function $\mathcal{I}(x) = \int_{-\infty}^{x} m(t)\,dF(t) = \mathbb{E}(Y\mathbb{I}(X \leq x))$, where $\mathbb{I}$ denotes the indicator function. The integrated regression function $\mathcal{I}$ can be estimated by $\mathcal{I}_n(x) = n^{-1}\sum_{i=1}^{n} Y_i\mathbb{I}(X_i \leq x)$ and a new empirical process can be constructed:

$$\overline{\overline{\alpha_n}}(x) = \sqrt{n}\big(\mathcal{I}_n(x) - \mathbb{E}_{\widehat{\theta}}\big(\mathcal{I}_n(x)\big)\big) = \sqrt{n}\sum_{i=1}^{n}\widehat{\varepsilon}_{i0}\mathbb{I}(X_i \leq x).$$

This empirical process can be taken again as the basis for generating test statistics, such as a Cramér–von Mises or Kolmogorov–Smirnov type tests, introduced for the distribution function case. The study of the asymptotic distribution of this type of tests is based on the weak convergence of $\overline{\overline{\alpha_n}}$ to a Gaussian limit process, being Stute (1997) an obliqued reference in this context, with a preliminary approximation given by Bierens (1982) and an earlier work by Su and Wei (1991). GoF tests based on empirical process for regression models with non-random design have been studied by Koul and Stute (1998) and Diebolt (1995), for a nonlinear parametric regression function. The extension to nonlinear and heteroscedastic regression is the goal of Diebolt and Zuber (1999, 2001).

1.3 Some notes for the reader

For the sake of simplicity, the notation that will be repeatedly used along this paper is introduced now, some already established in this first section. In a parametric setting, $\mathcal{M}_\theta$ denotes a parametric family of regression functions, with $\theta \in \Theta \subset \mathbb{R}^q$. Note that q is the dimension of the parameter space, and p will be used for the dimension of the covariate in the regression model. Dimension of a vector or a matrix is denoted by dim. A $\sqrt{n}$-consistent estimator for θ will be denoted by $\widehat{\theta}$.

In the general formulation of smoothing-based tests, a nonnegative weight function is needed, and it will be denoted by ω. The indicator function is written as $\mathbb{I}$. The superscript t denotes the transpose of a vector or a matrix, $\|\cdot\|$ is the Euclidean norm, and $*$ is the convolution operator.

Sup stands for the support of a random variable (with capital S) and sup denotes the supremum. The symbol $\sim$ will be used to denote equality in distribution, whereas convergence in distribution is denoted as $\xrightarrow{d}$. As usual, iff is a shortcut for the if and only if condition. RSS_0 and RSS_1 denote the average residual sums of squares under the null and the alternative hypotheses.

A collection of test statistics for GoF in different contexts will be presented along this paper. A generic test statistic will be denoted by T_n, and its expression may be different according to the regression setting considered. For smoothing-based tests statistics, notation T_{1n}, T_{2n} and T_{3n} is reserved for the tests presented above, specifically, in (5), (7) and (8), or suitable adaptations. Kolmogorov–Smirnov or Cramér–von Mises type tests, both computed from empirical processes or from the empirical distribution of the residuals, will be denoted by T_{nKS} and T_{nCM}, respectively. Empirical regression processes are denoted by $\overline{\overline{\alpha_n}}$, as an abuse of notation, since they also depend on h, but the bandwidth dependence will be made explicit only if necessary. The Gaussian limit of such a process will be denoted by α along the paper, except in Sect. 3, devoted to test calibration. In this section, following the standard notation, α denotes the significance level of a test.

Finally, likelihood ratio test statistics will be denoted by Λ_n (generalized likelihood ratio test), Λ_n^E (generalized empirical likelihood ratio test) and Λ_n^L (local empirical likelihood ratio test).

2 Some recent tests from the last ten years

In the last decade, a variety of alternative procedures to the previous testing methods, based on smoothing techniques and empirical processes, have been introduced for solving the GoF problem. Some of them are inspired by the classical likelihood ratio test or in the novel notion of empirical likelihood; other ones make use of the empirical distribution of the residuals under the null hypothesis and some other tests are designed to avoid the curse of dimensionality. In what follows, these previous ideas will be revised and some references will be given for each of these methodological settings.

2.1 The generalized likelihood ratio test

Assuming that the error process ε in a regression models follows a Gaussian distribution $\mathcal{N}(0,\sigma^2)$, it is possible to build the generalized likelihood ratio test (GLRT) as

$$\Lambda_n = l(m_{nh}, \widehat{\sigma}) - l(m_{\widehat{\theta}}, \widehat{\sigma}_0)$$

where $l(m,\sigma) = -n\log(\sqrt{2\pi\sigma^2}) - (2\sigma^2)^{-1}\sum_{i=1}^n (Y_i - m(X_i))^2$ is the Gaussian log-likelihood with $\widehat{\sigma}_0^2 = n^{-1}\sum_{i=1}^n (Y_i - m_{\widehat{\theta}}(X_i))^2 = \mathrm{RSS}_0$, the maximum likelihood estimators for the error variance under H_0 and $\mathrm{RSS}_1 = \widehat{\sigma}^2 = n^{-1}\sum_{i=1}^n (Y_i - m_{nh}(X_i))^2$, the corresponding generalized maximum likelihood estimator in the nonparametric context. This testing procedure has been studied by Fan et al. (2001) and more recently by Fan and Jiang (2007), in a survey for *Test* journal. The proposed method is a natural extension of the classical likelihood ratio test where, under the alternative, the likelihood function is evaluated in a nonparametric estimator of the regression function. The GLRT test, as noted in Fan et al. (2001), is given by

$$\Lambda_n = \frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1} \approx \frac{n}{2}\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1} \tag{11}$$

which resembles the F-test construction for regression models (see, for instance, Seber 1977 or Seber and Wild 1989). The numerator in (11) is essentially the test statistic in (8). See also Gijbels and Rousson (2001) for an F-test in local linear regression.

The asymptotic properties of the GLRT are similar to the previous tests, but a significant property of the GLRT is that the asymptotic distribution does not depend on nuisance functions, exhibiting what is known as Wilks phenomenon: the limit distribution of Λ_n in (11) is asymptotically χ^2. Specifically, $\tau\Lambda_n \sim \chi^2_{\mu_n}$, with $\mu_n \to \infty$ and τ a constant such that

$$(2\mu_n)^{-1/2}(\tau\Lambda_n - \mu_n) \xrightarrow{d} \mathcal{N}(0,1), \tag{12}$$

and neither μ_n nor τ in (12) depend on the nuisance parameters or functions (see Fan et al. 2001 and Fan and Jiang 2007 for details).

As an example, assume that $\sigma^2(x) = \mathrm{Var}(Y|X=x) = \sigma^2$, the univariate explanatory variable X has density f with $\mathrm{Sup}(X) = [0,1]$ and the error $\varepsilon \sim \mathcal{N}(0,\sigma^2)$. When testing the null hypothesis of linearity, that is, $H_0 : m \in \mathcal{M}_\theta$ with $\mathcal{M}_\theta = \{m_\theta : m_\theta(x) = \theta_0 + \theta_1 x\}$, then

$$\tau = \frac{K(0) - 2^{-1}\int K^2(x)\,dx}{\int (K(x) - 2^{-1}(K*K)(x))^2\,dx} \quad \text{and}$$

$$\mu_n = \frac{1}{h}\frac{(K(0) - 2^{-1}\int K^2(x)\,dx)^2}{\int (K(x) - 2^{-1}(K*K)(x))^2\,dx}.$$

Regarding the structure of Λ_n in (11) as an approximated generalized F-test, another class of GoF tests for regression models can be introduced. Some early contributions to the F-test type procedures, with spline smoothing, are Ramil-Novo and

González-Manteiga (1998) and Ramil-Novo and González-Manteiga (2000). More recently, Huang and Chen (2008) used kernel smoothing for the same purpose. See also Huang and Davidson (2010) for an extension to partially linear models.

2.2 The empirical process based on the empirical likelihood ratio test

The ideas of empirical likelihood (see Owen 2001) can be used to provide likelihood ratio tests, representing also an option to the test based on the integrated regression function (Sect. 1) or simply using the local empirical likelihood as an alternative to smoothing methods, and that will be seen in next section.

Taking into account that establishing a null parametric hypothesis $H_0 : m \in \mathcal{M}_\theta$ is equivalent to set $\mathbb{E}(\mathbb{I}(X \le x)(Y - m_{\theta_0}(X))) = 0$ for some $\theta_0 \in \Theta$ and $x \in \mathrm{Sup}(X)$, the empirical likelihood based on an i.i.d. sample of $(X, Y) \sim \overline{F}$ can be written as

$$\max_{\overline{F}} \prod_{i=1}^{n} \mathbb{P}_{\overline{F}}\big((X_i, Y_i) = (X, Y)\big) = L_{\overline{F}}, \quad \text{subject to} \quad \mathbb{E}\big(\mathbb{I}(X \le x)\varepsilon_0\big) = 0.$$

For instance, for a unidimensional X, the empirical likelihood is given by

$$L_{\overline{F}} = \prod_{i=1}^{n} \big(\overline{F}(X_i, Y_i) - \overline{F}\big(X_i^-, Y_i\big) - \overline{F}\big(X_i, Y_i^-\big) + \overline{F}\big(X_i^-, Y_i^-\big)\big)$$

and the generalized empirical likelihood ratio test statistic is

$$\begin{aligned} \Lambda_n^E(x) &= \frac{\sup\{L_{\bar{F}}; \mathbb{E}_{\bar{F}}(\mathbb{I}(X \le x)(Y - m_{\widehat{\theta}}(X))) = 0\}}{\sup L_{\overline{F}}} \\ &= \sup\left\{ n^n \prod_{i=1}^{n} p_i, \, p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \mathbb{I}(X_i \le x)\big(Y_i - m_{\widehat{\theta}}(X_i)\big) = 0 \right\}. \end{aligned}$$

A test statistic can be given by any continuous functional on Λ_n^E, such as $T_n = -2 \int \log \Lambda_n^E(x) \omega(x)\, dx$. See the works by Hjort et al. (2009) and Van Keilegom et al. (2008b) for more details on this methodology.

2.3 The local empirical likelihood ratio test

A local version of the empirical likelihood for the test given in Sect. 2.2 is

$$\Lambda_n^L = -2 \int \log\big(L_n\big(\tilde{m}(x, \widehat{\theta}) n^n\big)\big)\omega(x)\, dx,$$

where $L_n(\tilde{m}(x, \widehat{\theta})) = \max \prod_{i=1}^{n} p_i(x)$, subject to

$$\sum_{i=1}^{n} p_i(x) = 1, \quad \sum_{i=1}^{n} p_i(x) K_h(x - X_i)\big(Y_i - \tilde{m}(x, \widehat{\theta})\big) = 0$$

with $\tilde{m}(x, \widehat{\theta}) = \mathbb{E}_{\widehat{\theta}}(m_{nh}(x))$ the empirical local likelihood under H_0 (see Chen and Cui 2003 and Chen and Van Keilegom 2009b for more details). In the exhaustive review (with discussion) of Chen and Van Keilegom (2009b) in *Test*, the neat asymptotic result can be checked. Under $H_0 : m \in \mathcal{M}_\theta$, it can be established that

$$h^{-p/2}\big(\Lambda_n^L - 1\big) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\int (K * K)^2(x)\,dx \int \omega(x)\,dx}{\int K^2(x)\,dx}\right) \tag{13}$$

under some regularity conditions, shows a distribution free asymptotic behavior similar to the Wilks phenomenon exhibited by (12). Some previous references on this methodology can be consulted in Tripathi and Kitamura (2003) and Kitamura et al. (2004), and an extension to multidimensional response models can be found in Chen and Van Keilegom (2009a).

2.4 Tests based on the empirical distribution of the residuals

Assume that the regression model can be written in a location-scale form as

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

with ε independent of X and with error distribution $F_\varepsilon(y) = \mathbb{P}(\varepsilon \le y) = \mathbb{P}(\frac{Y-m(X)}{\sigma(X)} \le y)$. If $\tilde{\theta}_0$ denotes the argument that minimizes $\mathbb{E}((m(X) - m_\theta(X))^2)$ over the parameter set $\Theta \subset \mathbb{R}^q$, then $m_{\tilde{\theta}_0}$ is the parametric model with minimum distance to m, and the error distribution under this model is built as

$$F_{\varepsilon_0}(y) = \mathbb{P}(\varepsilon_0 \le y) = \mathbb{P}\left(\frac{Y - m_{\tilde{\theta}_0}(X)}{\sigma(X)} \le y\right).$$

Hence, the null hypothesis $H_0 : m \in \mathcal{M}_\theta$ is true if and only if the error distributions F_ε and F_{ε_0} are the same. This result opens a way for doing GoF by considering continuous functionals of the process $\{\hat{F}_\varepsilon(\cdot) - \hat{F}_{\varepsilon_0}(\cdot)\}$, where the estimators of the error distribution can be given by

$$\hat{F}_\varepsilon(y) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\left(\frac{Y_i - m_{nh}(X_i)}{\widehat{\sigma}(X_i)} \le y\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}(\widehat{\varepsilon}_i \le y)$$

and

$$\hat{F}_{\varepsilon_0}(y) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\left(\frac{Y_i - m_{\widehat{\theta}}(X_i)}{\widehat{\sigma}(X_i)} \le y\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}(\widehat{\varepsilon}_{i0} \le y)$$

respectively, where the variance estimator is

$$\widehat{\sigma}^2(x) = \sum_{i=1}^n W_{ni}(x) Y_i^2 - m_{nh}^2(x)$$

$\{W_{ni}\}_{i=1}^n$ being a sequence of Nadaraya–Watson weights and $\widehat{\theta}$ a least squares estimator. See Van Keilegom et al. (2008a) and Khmadladze and Koul (2009) for $p = 1$

(one-dimensional covariate) and Neumeyer (2009) and Neumeyer and Van Keilegom (2010) for $p \geq 1$.

Based on the empirical distribution of the residuals, the Kolmogorov–Smirnov and Cramér–von Mises tests are given by

$$T_{n\mathrm{KS}} = n^{1/2} \sup_{y\in\mathbb{R}} \left|\hat{F}_\varepsilon(y) - \hat{F}_{\varepsilon_0}(y)\right|, \quad \text{and} \quad T_{n\mathrm{CM}} = n \int \left(\hat{F}_\varepsilon(y) - \hat{F}_{\varepsilon_0}(y)\right)^2 d\hat{F}_{\varepsilon_0}(y). \tag{14}$$

From this methodology, a test for the error distribution can also be constructed, without further assumptions on m and σ, just comparing the empirical distribution of the residuals $\{\widehat{\varepsilon}_i\}_{i=1}^n$ with the one estimated under $H_0 : F_\varepsilon \in \mathcal{F}_\theta$.

A pioneer work for heteroscedastic regression models is the one by Akritas and Van Keilegom (2001). Jiménez-Gamero et al. (2005) studied the GoF testing problem in a multivariate linear model. Also for multivariate covariates, Müller et al. (2009) provided a result for the empirical distribution of the residuals when these are obtained from an undersmoothed local polynomial approximation. Mora and Pérez-Alonso (2009) tested a parametric family for the regression errors, considering a martingale transform of the empirical process. GoF tests for the error distribution were studied by Heuchenne and Van Keilegom (2010) without imposing a parametric for the regression or the variance.

The equality of the error distribution can also be interpreted in terms of the characteristic functions: if F_ε and F_{ε_0} coincide, the same happens with their characteristic functions. Hence, test statistics can be designed in terms of the corresponding empirical counterparts, $\hat{\phi}_\varepsilon(t) = n^{-1}\sum_{j=1}^n e^{it\widehat{\varepsilon}_j}$ and $\hat{\phi}_{\varepsilon_0}(t) = n^{-1}\sum_{j=1}^n e^{it\widehat{\varepsilon}_{j0}}$, denoting i the imaginary number. This also extends to functionals on the empirical characteristic functions, such as

$$T_n = n \int \left|\hat{\phi}_\varepsilon(t) - \hat{\phi}_{\varepsilon_0}(t)\right|^2 \omega(t)\, dt, \tag{15}$$

among others (see Huskova and Meintanis 2009 for further details). Some other related papers are Huskova and Meintanis (2007, 2010).

Finally, it should be mentioned that the asymptotic limit distribution of (14) and (15) can be derived based on the weak convergence of the processes $\{\hat{F}_\varepsilon(\cdot) - \hat{F}_{\varepsilon_0}(\cdot)\}$ and $\{\hat{\phi}_\varepsilon(\cdot) - \hat{\phi}_{\varepsilon_0}(\cdot)\}$, respectively, given that both tests are obtained from continuous functionals operating on these processes.

2.5 Tests design for avoiding the curse of dimensionality

A great deal of the theory developed during the 1990s, already introduced in Sect. 1, considers tests statistics constructed from the comparison of a nonparametric estimator of the regression model and an estimator under the null hypothesis (that is, based on the $\overline{\alpha_n}$ process), or in the comparison of the corresponding integrated regression function estimators (based on the $\overline{\overline{\alpha_n}}$ process). In both cases, the curse of dimensionality as p increases, p being the dimension of the explanatory variable, can be appreciated. For those tests based on $\overline{\alpha_n}$, the effect of the increasing dimension is clear when regarding the asymptotic power, although it is not that obvious for the other class of tests. Nevertheless, recent simulation studies have shown that the curse of dimensionality for tests based on $\overline{\overline{\alpha_n}}$ also appears for small samples and some further comments on the asymptotic power of GoF tests will be given in next section.

The difficulties aforementioned lead to different modifications of the previous methods in order to avoid the curse of dimensionality. For the tests based on smoothing methods, corresponding to process $\overline{\alpha_n}$, the works by Lavergne and Patilea (2008) and Xia (2009) should be noticed. Inspired on the projection pursuit ideas, the null hypothesis $H_0 : m \in \mathcal{M}_\theta$ is true if and only if $m = m_{\theta_0} \in \mathcal{M}_\theta$, and this is also equivalent to $\mathbb{E}(\varepsilon|X) = \mathbb{E}(\varepsilon_0|X) = \mathbb{E}(Y - m_{\theta_0}(X)|X) = 0$. In addition, this is also equivalent to:

$$\sup_{\beta, \|\beta\|=1} \sup_{v} \left|\mathbb{E}\left(\varepsilon|\beta^t X = v\right)\right| = 0 \Leftrightarrow \sup_{\beta, \|\beta\|=1} \mathbb{E}\left(\varepsilon \mathbb{E}\left(\varepsilon|\beta^t X\right)\right) = 0$$

under some regularity conditions, and this allows for the construction of some tests, similar to (7) (see Lavergne and Patilea 2008) which adapted to this context is given by

$$T_n = \sup_{\beta, \|\beta\|=1} \sum_{i<j} K_h\left(\beta^t (X_i - X_j)\right)\left(Y_i - m_{\widehat{\theta}}(X_i)\right)\left(Y_j - m_{\widehat{\theta}}(X_j)\right).$$

Another interesting idea consists on projecting the covariate X in the direction of $\beta = \beta_0$ such that β_0 (with $\|\beta_0\| = 1$) minimizes $\mathbb{E}^2(\varepsilon - \mathbb{E}(\varepsilon|\beta^t X)) = \mathbb{E}^2(\varepsilon - m_\beta(X))$, the single-indexing procedure obtained through the corresponding empirical counterparts (see Xia 2009). This enables to construct test statistics such as

$$T_n = \frac{1}{n} \sum_{i=1}^{n} \omega(X_i)\left(\widehat{\varepsilon}_{i0} - \widehat{m}_{\widehat{\beta}_i}\left(\widehat{\beta}_i^t X_i\right)\right)^2,$$

where

$$\hat{\beta}_i = \arg \min_{\beta, \|\beta\|=1} \sum_{i \neq j} \left(\widehat{\varepsilon}_{j0} - \widehat{m}^i_\beta(X_j)\right)^2, \quad i = 1, \ldots, n$$

and

$$\widehat{m}^i_\beta(x) = \frac{1}{n \hat{f}^i_\beta(X_i)} \sum_{i \neq j} K_h\left(\beta^t (x - X_j)\right)\widehat{\varepsilon}_{j0}, \quad \text{and} \quad \hat{f}^i_\beta(x) = \frac{1}{n} \sum_{i \neq j} K_h\left(\beta^t (x - X_j)\right).$$

In Xia (2009), the author proposed a single-indexing cross-validation, as a measure for the fit of the residuals, concluding that if the residuals cannot be predicted from the covariates, then the model is adequate.

Regarding the tests based on empirical regression processes, in Stute et al. (2008), the authors replaced the empirical process $\overline{\overline{\alpha_n}}$ by

$$\overline{\overline{\alpha_n}}^g(t) = n^{-1/2} \sum_{i=1}^{n} \left(g(X_i) - \overline{g}\right)\mathbb{I}(\widehat{\varepsilon}_{i0} \leq t), \quad t \in \mathbb{R} \tag{16}$$

indexed unidimensionally in t, with $\overline{g} = n^{-1} \sum_{i=1}^n g(X_i)$. The key for the adequate behavior of the tests based on (16) lies in the fourth term of the asymptotic representation (see Stute et al. 2008). Under the assumption that ε is independent of X in the regression model, this term is given by the empirical counterpart of

$$\mathcal{A} = \mathbb{E}\big[\big(g(X) - \mathbb{E}\big(g(X)\big)\big)H(t, X, \tilde{\theta}_0)\big]$$

with $H(t, x, \theta) = \mathbb{P}(\varepsilon \leq t + m_\theta(X) - m(X)|X = x)$ and $\tilde{\theta}_0$ defined at the beginning of Sect. 2.4. If the null hypothesis $H_0 : m \in \mathcal{M}_\theta$ does not hold, then $\mathcal{A} \neq 0$, guaranteeing the power of the test for fixed alternatives. The selection of the function g is also discussed in Stute et al. (2008) with the goal of maximizing power.

It is also worth it mentioning the contribution made by Escanciano (2006b), based on the almost sure characterization of the null hypothesis as $\mathbb{E}(\varepsilon_0 \mathbb{I}(\beta^t X \leq u)) = 0$, for some $\theta_0 \in \Theta$, $\forall u \in \mathbb{R}$ and $\forall \beta$ such that $\|\beta\| = 1$. This leads to a process $\overline{\overline{\alpha_n}}(\beta, u) = n^{-1/2} \sum_{i=1}^n \widehat{\varepsilon}_{i0} \mathbb{I}(\beta^t X_i \leq u)$, indexed in β and u. A detailed study on the distribution of functionals of $\overline{\overline{\alpha_n}}$ can be found in Escanciano (2004) and Escanciano and Velasco (2006b). Specifically, the Kolmogorov–Smirnov and the Cramér–von Mises tests can be extended to this setting, with the following statistics:

$$T_{n\mathrm{KS}} = \sup_u \sup_{\beta,\, \|\beta\|=1} \left| n^{-1/2} \sum_{i=1}^n \widehat{\varepsilon}_{i0} \mathbb{I}\big(\beta^t X_i \leq u\big) \right| = \sup_u \sup_{\beta,\, \|\beta\|=1} \left| \overline{\overline{\alpha_n}}(\beta, u) \right|,$$

$$T_{n\mathrm{CM}} = \int_{\mathbb{S}^p \times \mathbb{R}} \big(\overline{\overline{\alpha_n}}(\beta, u)\big)^2 \, dF_{n\beta}(u)\, d\omega(\beta)$$

$F_{n\beta}$ being the empirical distribution of $\{\beta^t X_i\}_{i=1}^n$ and ω a weight function over the projection direction. To some extent, $T_{n\mathrm{CM}}$ can be interpreted as the average limit of the Cramér–von Mises statistic on the projection directions β with associated probability ω.

3 Approximating the test distribution

Consider a generic testing problem in statistical inference,[1]

$$H_0 : g \in \mathcal{G} = \{g_\theta\}_{\theta \in \Theta}, \quad \text{vs.} \quad H_a : g \notin \mathcal{G}$$

to be solved by the construction of a test statistic T_n, where g may be the distribution F, the density f, the regression function m or the integrated regression function $\mathcal{I}$. Once a suitable test statistic is available, a crucial task is the calibration of critical points for a given level α, namely c_α. Usually, the estimation of these critical points c_α such that $\mathbb{P}_{H_0}(T_n \geq c_\alpha) = \alpha$ can be done by means of the asymptotic distribution, which is Gaussian for the tests designed for $g = f$ and $g = m$, density and regression, (4), (6), (9) and (10) or taking advantage of the Wilks phenomenon for (11). This can also be done using the limit distribution of continuous functionals associated to the empirical process, for instance, for the distribution function in (1). For the regression setting, the same applies for functions over Λ_n^E in Sect. 2.2, or for functionals on the empirical process (14) or (15).

[1] Note that α is the significance level of the test. Empirical processes are denoted by $\overline{\alpha_n}$ or $\overline{\alpha_{nh}}$, to make the dependence on h explicit. Along this section, g denotes a target function to test (density, distribution or regression function).

3.1 Distribution approximation under H_0

The use of asymptotic theory for calibration poses some problems such as the need to estimate some nuisance functions, a slow convergence rate to the limit distribution, or the difficulty of determining the limit distribution associated to continuous functionals on Gaussian processes. Under these circumstances, calibration can be done by means of resampling procedures, such as bootstrap (see Efron 1979).

One of the earlier references in this approach for calibrating the distribution of $T_n = T(F_n, F_{\widehat{\theta}})$ can be found in Stute et al. (1993). Specifically, if $T_n^* = T(F_n^*, F_{\widehat{\theta}}^*)$ is the bootstrap test statistic, then c_α can be estimated by $\widehat{c}_\alpha$ such that $\mathbb{P}_{H_0}^*(T_n^* \geq \widehat{c}_\alpha) = \alpha$, where F_n^* and $F_{\widehat{\theta}}^*$ are obtained from a bootstrap sample $\{X_1^*, \dots, X_n^*\}$ with $\mathbb{P}^*$ denoting the probability in the resampling. The bootstrap sample is a collection of i.i.d. generations from $X^* \sim F_{\widehat{\theta}}$ (parametric bootstrap). This procedure is extended for the calibration of $T_n = \int \overline{\alpha_n}^2(x)\omega(x)\,dx$ by Härdle and Mammen (1993), using Nadaraya–Watson kernel estimation, or with local linear methods by Alcalá et al. (1999). Stute et al. (1998a) considered the same idea for calibrating tests based on empirical regression processes. Actually, most of the papers on GoF methods after the cited ones include details about bootstrap algorithms for calibrating the test statistic distribution. The idea in this setting is based on obtaining bootstrap samples $\{(X_i^*, Y_i^*)\}_{i=1}^n$, from the bootstrap regression model $Y_i^* = \widehat{m}_{H_0}(X_i^*) + \varepsilon_i^*$. The errors in the bootstrap model ε_i^* may follow the empirical distribution of the residuals $\{\widehat{\varepsilon}_{i0}\}_{i=1}^n$ (naive bootstrap), may be generated by wild bootstrap (see Wu 1986 or Liu 1988), or can be obtained by smooth bootstrap, considering the convolution of the empirical distribution function of the residuals with a kernel (see Van Keilegom et al. 2008a and Cao and González-Manteiga 1993 for other smoothing methods), among many other choices. The choice of the bootstrap resampling method (naive, wild or smooth) is driven by the different characteristics of the regression model. Hence, naive bootstrap works for homocedastic model, although it is proved to be inconsistent when heterocedasticity is present. In this case, wild bootstrap is the alternative. In addition, if there is information available on the regularity of the error distribution, then smooth bootstrap can be used.

It should also be mentioned that, in the bootstrap regression model, $\widehat{m}_{H_0}$ represents the estimation of the regression function under the null hypothesis, which can be obtained by parametric methods, nonparametrically or in a semiparametric way, as will be seen later.

An alternative to bootstrap methods for calibration can be found in the martingale transform of the empirical processes (see Stute et al. 1998b and Khmadladze and Koul 2004), or also in Monte Carlo methods as proposed by Zhu (2005), introducing a randomization of the addends in the i.i.d. representation of the test statistics by some noise with zero mean and unit variance. See, for instance, Cao and González-Manteiga (2008) as an example of the application of this technique in some new tests based on smoothing methods. A comparison between the bootstrap and the martingale transform was provided by Koul and Sakhanenko (2005), and a recent contribution on the martingale approach was given by Song (2010). A justification for the accuracy of bootstrap can be always found in the high order expansions for the test statistic distributions (see Fan and Linton 2003).

3.2 Power comparison

Once a test has been properly calibrated for a certain level α, the one with maximum power should be chosen. In the regression setting, this can be done considering Pitman alternatives, that is:

$$H_0 : m \in \mathcal{M}_\theta, \quad \text{vs.} \quad H_a : m \equiv m_n = m_\theta + c_n d$$

where $c_n \to 0$ and d is a deterministic function collecting the deviation direction for the alternative hypothesis. Among all those tests with level α, the most powerful is that one with asymptotic power tending to one with a smallest c_n. As an example, the classical F-test for parametric regression models and parametric alternative with higher dimension verifies that $c_n \sim n^{-1/2}$, a parametric rate.

A test for a linear null hypothesis $H_0 : m_\theta(x) = \theta^t x$ based on smoothing the residuals such as $T_n = \int \overline{\alpha}_n^2(x)\omega(x)\,dx$, and all the previous tests based on smoothing, verify that $c_n \sim n^{-1/2}h^{-p/4}$. This is the price to pay for setting a nonparametric estimation under the alternative hypothesis, namely a contiguous or Pitman alternative. A parametric rate for c_n can be achieved for $T_n = \int \overline{\overline{\alpha_n}}^2(x)\,dF_n(x)$ (see Stute 1997). These remarks on the convergence rates also hold for the two testing approaches, based on smoothing or based on empirical regression processes, for more complex hypothesis, discussed in next sections.

From the previous considerations, could it be concluded that tests based on empirical regression processes are more powerful than tests based on smoothed residuals? Although this assertion holds asymptotically, it is not the same for small samples as shown by different simulation studies, as the one carried out by Miles and Mora (2002). Actually, the convergence rate for smoothed residuals test can be improved by using the following modified test statistic:

$$T_n = \max_{h \in H_n} \frac{\int \overline{\alpha_{nh}}^2(x)\omega(x)\,dx - \widehat{\mathbb{E}}_{H_0}\left(\int \overline{\alpha_{nh}}^2(x)\omega(x)\,dx\right)}{\widehat{\mathrm{Var}}^{1/2}\left(\int \overline{\alpha_{nh}}^2(x)\omega(x)\,dx\right)}, \tag{17}$$

which is just a studentized version of T_n for different $h \in H_n$, a suitably chosen grid for the bandwidth values. Although (17) considers an L_2 norm, this philosophy is directly applied to other tests. In Horowitz and Spokoiny (2001), it can be seen that the former test has a rate $c_n \sim n^{-1/2}(\log\log n)^{1/2}$ (see also Spokoiny 2001 for mathematical details). In addition, the data-driven modification in Guerre and Lavergne (2005) leads to rates close to $c_n \sim n^{-1/2}$. The rate $n^{-1/2}h^{-p/4}$, for the smoothed tests, clearly highlights the curse of dimensionality, can be ameliorated following Lavergne and Patilea (2008). In general, as noticed by Hall and Yatchew (2005), those tests based on $\sqrt{n}$-rate estimates of characteristics under the null hypothesis, detect alternatives at this same rate. As an example, tests based on the empirical distribution of the residuals also achieve this rate.

Another way of analyzing the asymptotic power of a test is considering Ingster's minimax approximation (see Ingster 1982, 1993a, 1993b, 1993c). From this perspective, m is assumed to belong to a certain space of differentiable functions on $\mathbb{R}^p$, denoted by $\mathcal{B}$, departing from the null hypothesis at a distance $c_n \to 0$. The goal of the minimax approach is to find the c_n rate with the fastest convergence to zero, guaranteeing that the test is uniformly consistent in $\mathcal{B}$, leading to the optimal rate of testing.

If $\lim_{n\to\infty} \inf_{m\in\mathcal{B}} \mathbb{P}(\text{Reject } H_0|m) = 1$, then a test is uniformly consistent in a space $\mathcal{B}$. Thus, for Hölder, Sobolev, and Besov spaces with derivatives of order $s \geq p/4$ and known s, the rate is $n^{-2s/(4s+p)}$ (see Ingster works and Guerre and Lavergne 2002). For unknown s, Spokoiny (1996) showed that the rate is $(n^{-1}\sqrt{\log\log n})^{2s/(4s+p)}$ and for $s < p/4$, it is shown in Guerre and Lavergne (2002) that the rate is $n^{-1/4}$. The modified test statistic in (17) achieves the optimal minimax rate.

4 Goodness-of-Fit in semiparametric and nonparametric models

This section revises different procedures for testing more complex hypotheses for the regression model, beyond the purely parametric case, as well as hypothesis about the nuisance functions appearing in the model, such as the conditional variance.

4.1 Tests on the regression function

After the extensive statistical literature on parametric regression models from the first half of the 1990s, the spotlight turned later to more complex nonparametric and semiparametric hypotheses. The primary goal is the statement of GoF tests for simplified models, with higher interpretability or tests for mitigating the curse of dimensionality.

Along this section, GoF tests for the following hypotheses will be commented:

– Partially linear model, with a linear function on a set of covariates X_1:

$$H_{0\mathrm{PL}} : \mathbb{E}(Y|X) = \theta_0^t X_1 + m_2(X_2) \tag{18}$$

with $X = (X_1, X_2)$ and $\dim(X_1) = p_1$, $\dim(X_2) = p_2$ (with $p_1 + p_2 = p$).

– Simplified model, with regression function depending only on a part of the covariates:

$$H_{0\mathrm{SM}} : \mathbb{E}(Y|X) = m_1(X_1) \tag{19}$$

with $\dim(X_1) = p_1$.

– Additive model:

$$H_{0\mathrm{AM}} : \mathbb{E}(Y|X) = c + m_1(X_1) + \cdots + m_p(X_p) \tag{20}$$

with $X = (X_1, \ldots, X_p)$.

– Single index model, with unknown link function $\mathcal{H}$:

$$H_{0\mathrm{SIM}} : \mathbb{E}(Y|X) = \mathcal{H}\big(\theta_0^t X\big). \tag{21}$$

– Generalized additive model, with link function $\mathcal{H}$, including possible interactions of different orders:

$$H_{0\mathrm{GAM}} : \mathbb{E}(Y|X) = \mathcal{H}\left(c + \sum_{d=1}^{p} m_d(X_d) + \sum_{i<j} m_{ij}(X_i, X_j) + \cdots\right), \tag{22}$$

m_{ij} representing the function associated with coordinates i, j of the predictor vector X.

The testing problem in (18) has been solved both using smoothing methods and empirical regression processes. For instance, the test statistic T_{2n} in (7) can be adapted to this setting. Denote the errors in the model under $H_{0\text{PL}}$ by $\varepsilon_{i0} = Y_i - \theta_0^t X_{1i} - m_2(X_{2i})$, for $i = 1, \ldots, n$. The null hypothesis $H_{0\text{PL}}$ holds if and only if $\mathbb{E}(\varepsilon_{i0}\mathbb{E}(\varepsilon_{i0}|X_i)) = 0$ and a natural estimator of this characteristic is given by

$$T_{2n}^{\text{PL}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} K_h(X_i - X_j)\widehat{\varepsilon}_{i0}\widehat{\varepsilon}_{j0}\hat{f}_2(X_{2i})\hat{f}_2(X_{2j}),$$

with $\hat{f}_2$ is a p_2-dimensional kernel density estimator (to avoid a random denominator), and $\widehat{\varepsilon}_{i0} = Y_i - \widehat{\theta}X_{i1} - \widehat{m}_{2\tilde{h}}(X_{2i})$, with bandwidth $\tilde{h}$ associated to a Nadaraya–Watson p_2-dimensional weight, estimated in two stages (see Robinson 1988 or Speckman 1988). Asymptotic theory results for $nh^{p/2}T_{2n}^{\text{PL}}$ can be found in Fan and Li (1996), obtaining an asymptotic variance equal to

$$2\int K^2(x)\,dx \int \sigma^4(x) f^2(x) f_2^4(x)\,dx,$$

f_2 being the density of the X_2 component.

A similar argument also holds for $H_{0\text{SM}}$. Now, the model errors are $\varepsilon_{i0} = Y_i - m_1(X_{1i})$ and the test in (7) can be adapted as

$$T_{2n}^{\text{SM}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} K_h(X_i - X_j)\widehat{\varepsilon}_{i0}\widehat{\varepsilon}_{j0}\hat{f}_1(X_{1i})\hat{f}_1(X_{1j}),$$

with $\hat{f}_1$ a p_1-dimensional kernel density estimator, $\widehat{\varepsilon}_{i0} = Y_i - \widehat{m}_{1\tilde{h}}(X_{1i})$, for $i = 1, \ldots, n$, and $\widehat{m}_{1\tilde{h}}$ being a p_1-dimensional Nadaraya–Watson estimator with smoothing parameter $\tilde{h}$.

The asymptotic behavior of $nh^{p/2}T_{2n}^{\text{SM}}$ was derived by Fan and Li (1996) where the asymptotic variance is equal to $2\int K^2(x)\,dx \int \sigma^4(x) f^2(x) f_1^4(x)\,dx$. The previous asymptotic expressions are useful for test calibration based on the limit distribution. In Li and Wang (1998) and Gu et al. (2007), for instance, some other alternatives based on bootstrap methods for calibration were explored.

Regarding $H_{0\text{SIM}}$ in (21) for the single index model, and denoting by $\varepsilon_{i0} = Y_i - \mathcal{H}(\theta_0^t X_i)$, for $i = 1, \ldots, n$, a test statistic could be based on an adequate estimate of

$$\mathbb{E}\big(\varepsilon_{i0} f_{\theta_0}\big(\theta_0^t X_i\big)\mathbb{E}\big(\varepsilon_{i0} f_{\theta_0}\big(\theta_0^t X_i\big)|X_i\big) f(X_i)\big)$$

where f_{θ_0} denotes the density of $\theta_0^t X_i$, also avoiding random denominators in kernel smoothing. The analogous test statistic is given by

$$T_{2n}^{\text{SIM}} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} \widehat{\varepsilon}_{i0}\widehat{\varepsilon}_{j0} \hat{f}_{\widehat{\theta}}\big(\widehat{\theta}^t X_i\big) \hat{f}_{\widehat{\theta}}\big(\widehat{\theta}^t X_j\big) K_h(X_i - X_j).$$

The asymptotic variance of the rescaled previous test $nh^{p/2}T_{2n}^{\mathrm{SIM}}$, for the single index model, is equal to $2\int K^2(x)\,dx\int \sigma^4(x)f^2(x)f_{\theta_0}^4(\theta_0^t x)\,dx$. One could proceed similarly adapting the tests in (5) and (8), based on smoothing methods.

The alternative testing route based on empirical regression processes was studied in Zhu and Ng (2003) and Delgado and González-Manteiga (2001). For instance, for the testing problem associated with $H_{0\mathrm{SM}}$, it is enough to consider the empirical process

$$\widetilde{\alpha}_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{f}_1(X_{1i})\widehat{\varepsilon}_{i0}\mathbb{I}(X_i\le x) \tag{23}$$

which results from the almost sure characterization of $H_{0\mathrm{SM}}$ as

$$f_1(X_1)\mathbb{E}\big(Y-m(X_1)|X\big)=0 \Leftrightarrow T(x)=\mathbb{E}\big(f_1(X_1)\big(Y-m_1(X_1)\big)\mathbb{I}(X\le x)\big)=0,$$

for all $x\in\mathrm{Supp}(X)$, which has a Gaussian limit (see Delgado and González-Manteiga 2001 for details). Similarly, for the partial linear model, $H_{0\mathrm{PL}}$ holds almost surely iff $\mathbb{E}[\{(Y-\mathbb{E}(Y|X_2))-\theta_0^t(X_1-\mathbb{E}(X_1|X_2))\}\mathbb{I}(X\le x)f_2(X_2)]=0$ for all $x\in\mathrm{Supp}(X)$ and the associated empirical process is (see Delgado and González-Manteiga 2001)

$$\widetilde{\widetilde{\alpha}}_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{\varepsilon}_{i0}\hat{f}_2(X_{2i})\mathbb{I}(X_i\le x). \tag{24}$$

With respect to the single index model, $H_{0\mathrm{SIM}}$ is almost surely equivalent to $\mathbb{E}((Y-\mathcal{H}(\theta_0^t X))\mathbb{I}(X\le x))=0$, for all $x\in\mathrm{Supp}(X)$, and the empirical process is now given by

$$\widetilde{\widetilde{\widetilde{\alpha}}}_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{\varepsilon}_{i0}\hat{f}_{\widehat{\theta}}\big(\widehat{\theta}^t X_i\big)\mathbb{I}(X_i\le x), \tag{25}$$

whose asymptotic behavior was studied by Xia et al. (2004), jointly with different estimators for θ_0.

For the calibration of the previous tests distribution, based on (23), (24) and (25), the bootstrap ideas introduced in Stute et al. (1998a) can be adapted. It should also be mentioned that, regarding the power and the asymptotic properties of the tests based on smoothing methods or in empirical regression processes, the remarks in Sect. 4 for parametric regression models also apply in these settings. Besides, a simple modification of Λ_n^E (Λ_n^L) in the empirical process based on the global (or local) empirical likelihood, replacing the error estimates for the corresponding ones according to the hypothesis to test, allows the extension of the methodology discussed in Sects. 2.2 and 2.3 to these general contexts (see Van Keilegom et al. 2008b).

Apart from the smoothing and empirical processes-based tests, there are also other alternatives in the literature. For instance, Stute and Zhu (2005a) propose a score-type test for $H_{0\mathrm{SIM}}$, which detects Pitman alternatives at a $\sqrt{n}$-rate, and also peak alternatives, being undetectable for the previous tests, as discussed in Horowitz and Spokoiny (2001).

From the former work by Eubank et al. (1995), there has been an extensive literature on tests for $H_{0\mathrm{AM}}$ in (20), such as the contributions by Dette and von Lieres und Wilkau (2001), Gozalo and Linton (2001), Härdle et al. (2001), Derbort et al. (2002), and Fan and Jiang (2005). Denoting by $\widehat{m}$ the general nonparametric estimator and $\widehat{m}_0$ the estimator under the null hypothesis, usually obtained by backfitting or marginal integration (see Sperlich et al. 1999 for a comparison of both methods), the following test statistic can be constructed:

$$T_{1n}^{\mathrm{AM}} = \frac{1}{n}\sum_{i=1}^{n}\big(\widehat{m}(X_i) - \widehat{m}_0(X_i)\big)^2, \tag{26}$$

which is just a natural generalization of the tests in T_{1n} in (5) or its discretized version T_{1n}^{D}, with asymptotic bias and variance as in (4). Motivated by the incorrelation between ε_0 and $(m(X) - m_0(X))$ (see Gozalo and Linton 2001), the following test statistic:

$$T_{1n}^{\mathrm{AM}*} = \frac{1}{n}\sum_{i=1}^{n}\big(\widehat{m}(X_i) - \widehat{m}_0(X_i)\big)\widehat{\varepsilon}_{i0} \tag{27}$$

has asymptotic bias $K(0)\int \sigma^2(x)\,dx$ and asymptotic variance

$$2\int K^2(x)\,dx \int \sigma^4(x)\,dx.$$

Following the test statistic in (8),

$$T_{3n}^{\mathrm{AM}} = \frac{1}{n}\sum_{i=1}^{n}\big(\widehat{\varepsilon}_{i0}^2 - \widehat{\varepsilon}_i^2\big) \tag{28}$$

has asymptotic bias and variance similar to (10), and following the construction in (7)

$$T_{2n}^{\mathrm{AM}} = \frac{1}{n(n-1)}\sum_{i\neq j} K_h(X_i - X_j)\widehat{\varepsilon}_{i0}\widehat{\varepsilon}_{j0}, \tag{29}$$

presents asymptotic behavior mimicking (9). Finally, the GLRT test in (11), can be adapted to this context as

$$\Lambda_n^{\mathrm{AM}} = \frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1} \approx \frac{n}{2}\,\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}, \tag{30}$$

with $\mathrm{RSS}_0 = \sum_{i=1}^{n}(Y_i - \widehat{c} - \sum_{j=1}^{p}\widehat{m}_j(X_{ij}))^2$ and $\mathrm{RSS}_1 = \sum_{i=1}^{n}(Y_i - \widehat{m}(X_i))^2$.

The test statistics in (26)–(30) follow the same asymptotic distributions as the corresponding ones in the fully parametric context, also for Pitman alternatives. This is a consequence of the CLT in de Jong (1987), related to the asymptotic behavior of U-statistics, recalling that although the additive model is nonparametric, its dimension is lower than under the general hypothesis. Similar comments can be made regarding fixed alternatives (see Dette et al. 2005).

Finally, for the generalized additive model (22), or its simpler version (21) with known link $\mathcal{H}$, Stute and Zhu (2005b) extended the empirical regression process for generalized linear models (GLMs). Rodríguez-Campos et al. (1998) (and also Azzalini and Bowman 1993 and Azzalini et al. 1989 as previous references) adapt the test from González-Manteiga and Cao (1993) for GLMs and Härdle et al. (1998) and Müller (2001) provided an extension for the generalized partial linear model (GPLM)

$$H_{0\mathrm{GPL}} : \mathbb{E}(Y|X) = \mathcal{H}\big(\theta_0^t X_1 + m_2(X_2)\big), \tag{31}$$

in an approach inspired by the likelihood ratio test. Sperlich et al. (2002) and Roca-Pardiñas et al. (2005) studied interaction tests under a generalized additive model (22), and Liang et al. (2010), using generalized versions of the GLRT, introduced test statistics for the more general partial linear single index model

$$H_{0\mathrm{PLSIM}} : \mathbb{E}(Y|X) = \theta_0^t X_1 + \mathcal{H}\big(\bar{\theta}_0^t X_2\big). \tag{32}$$

An exhaustive comparative study for different testing methods for (21), (22) and (31) is given by Roca-Pardiñas and Sperlich (2007).

The inmense flow of contributions on GoF tests for complex regression models, such as (32), in the last decade, makes it extremely difficult to condense a detailed list of reference. Some alternatives to the aforementioned test can be found in Escanciano and Song (2009), Song (2010), and Maity et al. (2009).

4.2 Tests for the variance function

In testing methods for the regression function, from parametric, nonparametric or semiparametric settings, in order to use the asymptotic distribution or to obtain a simpler functional formulation, the nuisance functions behavior must be determined. Among the possible nuisance functions, the conditional variance $\sigma^2(x) = \mathrm{Var}(Y|X = x)$, is the most frequent one, existing many contributions devoted to test $H_0 : \sigma^2 \in \mathcal{S}$ vs. $H_a : \sigma^2 \notin \mathcal{S}$. The homocedasticity test, (σ^2 constant), is a particular case. Tests for constant variance have been studied by Diblasi and Bowman (1997), smoothing the residuals on a suitably transformed scale, and Dette and Munk (1998). A residual-based tests for heteroscedasticity was proposed by Dette (2002) and Liero (2003) derived a tests for checking the hypothesis of constant conditional variance against its dependence on the design of the covariate. The empirical process approach has been considered by Zhu et al. (2001) in this context. See also the book by Carroll and Ruppert (1988) for previous references on the topic.

For a location-scale model, $Y = m(X) + \sigma(X)\varepsilon$, with ε independent of X, a more general hypothesis than homocedasticity may be of interest:

$$H_0 : \sigma^2 \in \mathcal{S}_\gamma = \big\{\sigma_\gamma^2,\ \gamma \in \Gamma\big\}, \quad \text{vs.} \quad H_a : \sigma^2 \notin \mathcal{S}_\gamma.$$

Considering $\varepsilon = (Y - m(X))/\sigma(X)$ and $\varepsilon_0 = (Y - m(X))/\sigma_{\tilde{\gamma}_0}(X)$, with

$$\tilde{\gamma}_0 = \arg\min_{\gamma\in\Gamma} \mathbb{E}^2\big(\big(Y - m(X)\big)^2 - \sigma_\gamma^2(X)\big) = \arg\min_{\gamma\in\Gamma}\big(\sigma^2(X) - \sigma_\gamma^2(X)\big)^2,$$

the null hypothesis H_0 holds iff the distributions of the errors ε and ε_0 are the same. Hence, the methodology for GoF tests for regression based on the empirical distribution of the residuals can be generalized for testing about the variance (see Dette et al. 2007 for further details) m being the nuisance function in this scenario.

Apart from these previous ideas, there have been also other proposals for tests on the variance inspired by the estimation of the regression and the integrated regression functions, as detailed in Chap. 7 in Zhu (2005) or, more recently, by Samarakoon and Song (2010). It is also possible to test other hypothesis about σ^2, with nonparametric or semiparametric assumptions on m, similar to Sect. 5.1. For some examples, see You and Chen (2005), Dette and Marchlewski (2008), and Wong et al. (2009).

5 Testing when dependence is present

As in many other areas of statistical inference, GoF methods for regression models have been also adapted to account for dependent data. In this section, testing on the trend and on the correlation function for fixed and random design models will be revised, both in time series, spatial and spatio-temporal models and for continuos time processes.

5.1 Testing in time series

Consider the following simple fixed design regression model:

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n \tag{33}$$

where m is the regression function (or trend, in the time series glossary), x_i is an empirical known predictor. For instance, $x_i = i/n$, for $i = 1, \ldots, n$ a sequence of time moments which is typically assumed to belong to the unit interval, $x_i \in [0, 1]$. This predictors are linked to some particular design, such as $i/n = \int_0^{x_i} f(t)\,dt$ f being the design density, allowing this formulation also for non-equally spaced moments. In addition, the zero-mean error sequence $\{\varepsilon_i\}_{i=1}^n$ in (33) is assumed to show a certain second-order stationary dependence structure, with covariance function $C(k) = \mathbb{E}(\varepsilon_1 \varepsilon_{k+1})$. This stationarity condition means that the covariance is just a function of the lag between two time moments.

Most of the test based on smoothing methods and described in Sect. 1 can be adapted to this context for assessing $H_0 : m \in \mathcal{M}_\theta$. Specifically, the tests statistic T_{1n} in (5) or its discretized version, T_{1n}^D, can be directly applied for this testing problem, as well as the test T_{3n} in (7). The key is using a smoothing method suitably adapted for fixed design (see, for instance, Priestley and Chao 1972 or Gasser and Müller 1979).

The asymptotic behavior of the smoothed test statistics with time dependence is quite similar to the independent case, just replacing the conditional variance by the covariance sum, $\sum_{k=-\infty}^{\infty} C(k)$.

GoF tests in linear regression models with correlated errors have been studied by González-Manteiga and Vilar-Fernández (1995), considering a MA(∞) structure. Bootstrap calibration of tests with ARMA structure was performed by Vilar-

Fernández and González-Manteiga (1996, 2000). Biederman and Dette (2000) extended the results in González-Manteiga and Vilar-Fernández (1995) under fixed alternatives. GoF testing with dependent data have been also extended to partial linear regression models (see González-Manteiga and Aneiros-Pérez 2003), also under long-memory dependence (see Aneiros-Pérez et al. 2004). Biederman and Dette (2001) proposed optimal designs for testing the functional form of a regression model, under fixed design with heterocedastic errors and their ideas could be extended to the dependent case.

In a random design problem, the random sample is given by $\{(X_t, Y_t)\}_{t=1}^n$, a collection of $(p+1)$-dimensional vectors from a series $\{(X_t, Y_t)\}_{t\in\mathbb{Z}}$, following the model $Y_t = m(X_t) + \varepsilon_t$. Considering a strictly stationary distribution $\overline{F}(x, y)$ associated with a prototype variable (X_0, Y_0), a test for the trend function can be done based on the smoothed tests with convenient modifications, usually assuming that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a strictly stationary process with $\mathbb{E}(\varepsilon_t|\mathcal{F}_t) = 0$, $\mathcal{F}_t$ being the σ-algebra generated by the past observation of the process up to a time t, $\{(X_k, Y_{k-1})\}_{k\le t}$. A suitable adaptation of (5) is

$$T_{1n}^{\mathrm{TS}} = \int \left(\frac{1}{n} \sum_{t=1}^{n} K_h(x - X_t)\bigl(Y_t - \widehat{m}_{H_0}(X_t)\bigr) \right)^2 \omega(x)\, dx \tag{34}$$

with $\widehat{m}_{H_0}$ a consistent estimator under the null hypothesis to test, which may be parametric, nonparametric or semiparametric. It turns out that the asymptotic bias of $nh^{p/2}T_{1n}^{\mathrm{TS}}$ is given by

$$h^{-p/2} \int K^2(u)\omega(x + hu)\Pi(x)\sigma^2(x)\, dx\, du$$

and the asymptotic variance

$$2\int \sigma^4(x)\Pi^2(x)\omega(x)\, dx \int K(u)K(v)K(u - z)K(v - z)\, du\, dv\, dz,$$

where $\sigma^2(x) = \mathrm{Var}(Y_t|X_t = x)$ denotes the conditional variance and Π is the stationary density of $\{X_t\}_{t\in\mathbb{Z}}$. Hence, the asymptotic bias and variance of the previous test is the same as for the original T_{1n}, as long as the dimension of the null hypothesis is smaller than the nonparametric dimension of the alternative. Similar results can also be seen in González-Manteiga et al. (2002).

Kreiss et al. (2008) provided an extensive review on this type of tests, justifying the consistency of a variety of resampling procedures for test calibration (see also Franke et al. 2002). In Gao (2007), the adaptation of the test (7) can be found. For other tests based on empirical likelihood ideas, see Chen et al. (2003) and Chen and Gao (2007).

There are also some previous works on simpler hypothesis, such as the linearity of the trend, which implies that $X_t = (Y_{t-1}, \ldots, Y_{t-k})$, defining an AR($k$) model (see Hjellvik and Tjøstheim 1995, 1996, and Hjellvik et al. 1998). Also for this AR(k) model, Li and Tkacz (2006) propose a testing procedure for a parametric model for the conditional density: $H_0 : \Pi(y|x) = \Pi(y|x, \theta)$, $\Pi(\cdot|x)$ being the stationary density of $Y_t|(Y_{t-1}, \ldots, Y_{t-k}) = x$.

The adaptation of empirical processes based GoF tests for AR(k) processes is also possible. Consider $X_t = (Y_{t-1}, \ldots, Y_{t-k}, Z_t)$ with $k \leq t$ and Z_t a p-dimensional variable. For testing $H_0 : m \in \mathcal{M}_\theta$, the following empirical process can be constructed:

$$\alpha_n^*(x, \omega) = \sqrt{n} \sum_{t=k}^{n} \big(Y_t - m_{\widehat{\theta}}(X_t)\big)\omega(X_t, x).$$

The convergence of α_n^*, as well as the properties of the associated tests and resampling methods for calibration, have been studied by different authors. Koul and Stute (1999) considered the case $X_t = Y_{t-1}$ (Markovian hypothesis of order one) with $\omega(X_t, x) = \mathbb{I}(X_t \leq x)$. With the same weight function, Dominguez and Lobato (2003) considered a Markovian hypothesis of higher order, also studied by Stute et al. (2006) for a general linear model.

GoF tests for a parametric time series using empirical regression processes have been proposed by Escanciano (2006a, 2007a), establishing weak convergence of the process and investigating the asymptotic properties under the null hypothesis and for Pitman alternatives, being bootstrap an option for calibration. For possibly non-stationary time series, under martingale conditions, Escanciano (2007b) proved the weak convergence of a class of empirical processes. In related works, Escanciano and Velasco (2006a) and Escanciano (2009) proposed tests for the martingale difference hypothesis.

Also in this context, model identification based on cummulative lagged conditional mean and variance, for a parametric time series model, was earlier studied by McKeague and Zhang (1994). Diagnostic tests for self-exciting threshold autoregressive models have been proposed by Koul et al. (2005) whereas the extension to a multivariate context was studied by Chabot-Hallé and Duchesne (2008).

Weak convergence of the process α_n^* is obtained by martingale difference theory, which is also useful for tests in the spectral domain (see Escanciano and Velasco 2006a), playing a key role for testing about the dependence structure.

For dependent data in general, and for a second-order stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ in particular, it is also worthwhile to assess hypotheses about the dependence structure, characterized by the covariance function $C(k) = \mathrm{Cov}(X_t, X_{t+k})$ or by its Fourier transform, the spectral density, f. The spectral density can be estimated nonparametrically by the periodogram

$$I(\lambda_k) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t e^{-it\lambda_k}\right|, \quad \lambda_k = \frac{2\pi k}{N},\ k = 1, \ldots, N = \left\lfloor \frac{n-1}{2} \right\rfloor$$

with λ_k the Fourier frequencies. If $\{X_t\}_{t\in\mathbb{Z}}$ admits a linear representation, the periodogram can be written as the response variable in the following regression model:

$$I(\lambda_k) = f(\lambda_k)V_k + R_k \tag{35}$$

with $\{V_k\}_{k=1}^N$, a sequence of independent standard exponential random variables, and R_k an asymptotically negligible term (see Brockwell and Davis 1991). Taking logarithms in (35), and denoting by $Y_k = \log I(\lambda_k)$, it is easy to see that

$$Y_k = m(\lambda_k) + z_k + r_k \tag{36}$$

with $m = \log f$, the log-spectral density, $\{z_k\}_{k=1}^N$ are i.i.d. random variables with Gumbel(0, 1) distribution and $r_k = \log(1 + R_k/(f(\lambda_k)V_k))$. Ignoring the negligible terms in models (35) and (36), these expressions can be interpreted as a multiplicative and an additive regression model respectively.

Based on model (35), Paparoditis (2000) proposed a GoF for a parametric model for the spectral density, following the ideas in Bickel and Rosenblatt (1973). Specifically, for testing $H_0 : f \in \mathcal{F}_\theta$ vs. $H_a : f \notin \mathcal{F}_\theta$ and based on representation (35), the author considered the test statistic

$$T_n = nh^{-1/2} \int_{-\pi}^{\pi} \left(\frac{1}{n} \sum_{k=-N}^{N} K_h(\lambda - \lambda_k) \left(\frac{I(\lambda_k)}{f_{\widehat{\theta}}(\lambda_k)} - 1 \right) \right)^2 d\lambda,$$

which has asymptotic distribution-free bias and variance,

$$h^{-1/2} \int_{-\pi}^{\pi} K^2(x)\, dx, \quad \pi^{-1} \int_{-2\pi}^{2\pi} \left(\int_{-\pi}^{\pi} K(u)K(u + x)\, du \right)^2 dx.$$

Focusing on the regression model in the log-spectral scale (36), Fan and Zhang (2004) proposed a testing method for assessing a parametric model for the log-spectral density function ($H_0 : m \in \mathcal{M}_\theta$ vs. $H_a : m \notin \mathcal{M}_\theta$) using the GLRT theory introduced in Fan et al. (2001) and discussed in Sect. 2. Ignoring r_k in model (36), the log-likelihood function is given by $\sum_{k=-N}^{N} (Y_k - m(\lambda_k) - e^{Y_k - m(\lambda_k)})$. A parametric estimator under the null hypothesis can be obtained by Whittle's log-likelihood, whereas local linear smoothing is used for the nonparametric estimator under the general model. The GLRT in this context also exhibits the Wilks phenomenon mentioned before.

Apart from these procedures which make use of smoothing methods, some other tests have been proposed based on the ideas of the integrated regression function on the ratio between the periodogram and the parametric spectral density. See Delgado et al. (2005), Hidalgo and Kreiss (2006), and Delgado and Velasco (2010).

There are also quite recent contributions in this topic, such as the proposal of a GoF tests based on the supremum norm in the time domain (see Hidalgo 2008) or a testing procedure for nonstationary models (see Gao et al. 2009). Confidence bands for the spectral density have been developed by Neumann and Paparoditis (2008a). GoF for multivariate covariance structures have been studied by Eichler (2008), Dette and Paparoditis (2009), and Dette and Hildebrandt (2012). A general Markovian hypothesis has been tested by Neumann and Paparoditis (2008b). There are also some further works on GoF tests in the spectral domain as Sergides and Paparoditis (2007, 2009), Paparoditis (2009, 2010).

5.2 Testing in spatial and spatio-temporal models

Literature on GoF tests with spatial or spatio-temporal dependent data, inspired by GoF tests from time series models, is quite recent. In spatial statistics (see Cressie

1993), the dependence structure of second-order stationary processes can be modeled by the covariogram (or by the variogram, in an intrinsic stationary context). Denoting the spatial process by $Z(s)$, with $s \in \mathcal{D} \subset \mathbb{R}^d$ (and $d = 2$ for spatial data), second-order stationarity implies that $\mathbb{E}(Z(s)) = \mu$, $\forall s \in D$ and $\mathrm{Cov}(Z(s_1), Z(s_2)) = C(s_1 - s_2)$, $\forall s_1, s_2 \in D$. Under intrinsic stationarity, the dependence structure can be described by the variogram function $2\gamma(s_1 - s_2) = \mathrm{Var}(Z(s_1) - Z(s_2))$, being the variogram and the covariogram related for a second-order stationary process: $2\gamma(s) = 2(C(0) - C(s))$. Determining the dependence is crucial for spatial and spatio-temporal prediction.

Estimation of the covariogram and the variogram function, given a realization of a spatial process $\{Z(s_1), \dots, Z(s_n)\}$, is a well-studied problem in the spatial statistical literature, but this is not the case for testing problems. Diblasi and Bowman (2001) proposed a test for independence for the spatial case, which in terms of the variogram (or more precisely, in terms of the semivariogram γ) can be written as $H_0 : \gamma = \sigma^2$ vs. $H_a : \gamma \neq \sigma^2$. The test statistic is based on the comparison of smooth estimators of the semivariogram, similar to an F-test. The idea was extended by Diblasi and Maglione (2004) for testing a parametric family for the variogram.

For second-order stationary spatial processes, similar to time series, tests about the dependence structure can be formulated from the spectral domain. The spatial spectral density (bidimensional Fourier transform of the spatial covariance) can be estimated nonparametrically by the spatial periodogram, with suitable modifications to guarantee consistency (see Guyon 1982). Adaptations of the test proposed by Paparoditis (2000) and the GLRT test can be found in Crujeiras et al. (2010a), where the hypothesis to test is $H_0 : f \in \mathcal{F}_\theta$, f being the spatial spectral density. Hidalgo (2009) presented an alternative method based on empirical processes associated with the spatial periodogram. In these works, bootstrap calibration procedures are also detailed.

One of the simplifying assumptions in spatio-temporal processes is separability. Separability implies that the covariance structure can be factorized as a product of a spatial and a temporal covariance functions. That is, for a spatio-temporal process $\{Z(s,t), s \in D, t \in T\}$, the covariance $\mathrm{Cov}(Z(s+u, t+v) - Z(s,t)) = C(u,v)$ can be written as $C(u,v) = C_S(u)C_T(v)$, where C_S and C_T are spatial and temporal covariance functions, respectively. It is straightforward to check that, under separability, the log-spectral density is the sum of a spatial log-spectral density $\log f_S$ plus a temporal log-spectral density $\log f_T$. Hence, testing for separability can be seen as testing for additivity in the log-spectral domain. Crujeiras et al. (2010b) propose testing procedures for separability adapting additivity tests described in Sect. 5.

5.3 Testing in continuous time models

Consider a continuous time diffusion process, commonly used to model the dynamics of interest rates or stock prices exchanges in finance:

$$dX_t = m(X_t)\,dt + \sigma(X_t)\,dW_t, \quad t \in \mathcal{T} \subset \mathbb{R}^+ \tag{37}$$

where $\{X_t\}_{t\in\mathcal{T}}$ is a continuous time process, m is the drift function and σ is the diffusion function, with $\{W_t\}_{t\in\mathcal{T}}$ a standard Brownian motion. It may be of interest to test a null hypothesis such as $H_0 : dX_t = m_\theta(X_t)\,dt + \sigma_\theta(X_t)\,dW_t$ with $\theta \in \Theta$. As a special case in the financial context, $X_t = r_t$ may be an interest rate process, which has motivated a significant amount of papers related to the study of (37).

A reasonable strategy consists in discretizing Eq. (37) for $\{X_t\} = \{r_t : t \in \mathcal{T}\}$, observed at time moments $t_i = i\Delta$, for $i = 0, 1, \dots, n$ (with spacing $\Delta > 0$). The discretized version of (37) is then given by

$$Y_{t_i} = r_{t_{i+1}} - r_{t_i} = m(r_{t_i})\Delta + \sigma(r_{t_i})(W_{t_{i+1}} - W_{t_i}),$$

for $i = 0, 1, \dots, n-1$, which can be rewritten as a regression model as

$$\Delta^{-1} Y_{t_i} = m(r_{t_i}) + \sigma(r_{t_i}) \frac{\varepsilon_{t_i}}{\sqrt{\Delta}}, \tag{38}$$

with $\varepsilon_{t_i} \sim \mathcal{N}(0,1)$ independent of r_{t_i}. Hence, taking as initial random sample $\{(r_{t_i}, \Delta^{-1}Y_{t_i})\}_{i=0}^{n-1}$, the previous GoF testing techniques can be adapted for the drift function: smoothed-based tests, tests using empirical regression process, GLRT, ... always testing a null hypothesis over a continuous time model by a discretized approximation.

There is a large collection of references on this topic, which could be classified attending to the target function to test. For instance, Aït-Sahalia (1996) proposed testing parametric models for the marginal density of X_t rates by comparing the implied density with a nonparametric estimator, without considering discrete approximations to the process, whereas Corradi and Swanson (2005) discussed GoF tests by comparing cumulative distribution functions instead of densities. Tests for marginal density functions have been also studied by Gao and King (2004) and Hong and Li (2005), involving kernel estimators, as well as Arapis and Gao (2006) and the Bickel and Rosenblatt test was adapted by Lee (2006) for diffusion processes.

GoF tests for the drift coefficient were studied by Negri and Nishiyama (2009), using empirical processes and by Negri and Nishiyama (2010), based on continuous observations. Kutoyants (2010) considered the problem of testing a simple null hypothesis for the drift coefficient, based on a Cramér–von Mises test. Gao and Casas (2008) proposed tests for the specification of the drift and the volatility functions of a semiparametric diffusion process, under a discrete approximation, using nonparametric estimation methods.

With respect to the variance, specification tests have been analyzed by Corradi and White (1999), Dette and von Lieres und Wilkau (2003), and Li (2007) without requiring a complete knowledge of the functional form of the drift model. Dette et al. (2006) and Dette and Podolskij (2008) assessed the parametric form based on estimations and stochastic processes of the integrated volatility.

Finally, joint model specification problems have been also analyzed. A test for a parametric model specification, based on kernel estimation and using the ideas of empirical likelihood, was proposed by Chen et al. (2008). Kristensen (2011) introduced misspecification tests for semiparametric and fully parametric univariate diffusion models. GLRT methodology has been also adapted to diffusion process by Fan and Zhang (2003) and Fan et al. (2003). Martingale theory has been used by Masuda et al. (2010) and Song (2011) for constructing GoF tests for diffusion models. Aït-Sahalia et al. (2010) proposed procedures to test a Markov hypothesis in a mixing stationary process. Lee and Wee (2008) studied the residual empirical processes for model (37) and Monsalve-Cobis et al. (2011) introduce a bootstrap GoF tests based on empirical processes for testing parametric hypothesis on the drift and the volatility functions.

6 Goodness-of-fit tests for regression models with complex data

Usually, the sample at hand does not provide complete information about the underlying population, posing some problems for the application of GoF tests. Apart from the dependent data case, data may also be censored and/or truncated, some data may be missing, present some error measurement or even length-bias.

6.1 Censored and/or truncated data

The main motivation for the analysis of censored and/or truncated data can be found in survival and reliability analysis, and more recently, also in econometrics. Truncation appears when the individual lifetime finishes before the study follow up (left truncation, LT) whereas censoring happens when the individual lifetime is not observed until its end due to a previous censoring event (right censoring, RC).

In this general context, information for each individual is given by the value of (X, T, Z, δ), where X is a vector of covariates, T denotes the truncation time, $Z = \min\{Y, C\}$ is the observed value where Y is the lifetime of the individual and C is the censoring time and $\delta = \mathbb{I}(Y \leq C)$. The vector is completely observed if $T \leq Z$.

Some GoF tests for different models based on an observed sample denoted by $\{(X_i, T_i, Z_i, \delta_i)\}_{i=1}^n$ were proposed by Cao and González-Manteiga (2008) in the following scenarios:

– Polynomial regression model:

$$H_{0\mathrm{PR}} : \Psi\big(F(\cdot|X = x)\big) = A^t(X)\theta \tag{39}$$

with $A : \mathbb{R}^p \to \mathbb{R}^q$, $\theta \in \mathbb{R}^q$ and $\Psi(Q) = \int_0^1 Q^{-1}(s)J(s)\,ds$ for any distribution Q with quantile function $Q^{-1}(s) = \inf\{u; Q(u) \geq s, \}$, for $s \in [0, 1]$. J is a non-negative function such that $\int_0^1 J(s)\,ds = 1$ and $F(\cdot|X = x)$ denotes the conditional distribution of $Y|X = x$. If J is taken as the uniform density, then $\Psi(F(\cdot|X = x)) = A^t(X)\theta = \mathbb{E}(Y|X = x)$. For $p = 1$ and $A(x) = (1, x, \ldots, x^{q-1})$, the polynomial regression model is obtained.

– Proportional hazard model:

$$H_{0\mathrm{PH}} : \lambda(t|X = x) = \lambda_0(t)\exp\big(A^t(X)\theta\big), \tag{40}$$

where $\lambda(\cdot|X = x)$ is the conditional hazard rate of $Y|X = x$ and λ_0 is a baseline function. The model in $H_{0\mathrm{PH}}$ is the well-known Cox regression model, introduced by Cox (1972).

– Additive risk models:

$$H_{0\mathrm{AR}} : \lambda(t|X = x) = \lambda_0(t) + A^t(X)\theta. \tag{41}$$

– Proportional odds model:

$$H_{0\mathrm{PO}} : \operatorname{logit}\big(1 - \exp\big(-\Lambda(t|X = x)\big)\big) = \log\left(\frac{\mathbb{P}(Y \le t|X = x)}{\mathbb{P}(Y > t|X = x)}\right) \tag{42}$$

$$= \alpha_0(t) + A^t(X)\theta, \tag{43}$$

with $\operatorname{logit}(u) = \log(u/(1-u))$, α_0 an increasing function and $\Lambda(\cdot|X = x)$ the cumulative hazard rate function.

Estimation of the previous models have been studied by Grigoletto and Akritas (1999). For instance, considering $A(x) = (1, x, \dots, x^{q-1})$, a suitable statistic for testing problems (39)–(43) is

$$T_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}\big(\widehat{\Omega}_r - \big(\theta_0 + \theta_1 X_t + \cdots + \theta_p X_t^{q-1}\big)\big)^2,$$

with $\widehat{\Omega}_r$ an estimator of $\Omega_r = \Omega(X_r)$, an adequate transform of $\Lambda(\cdot|X)$ for obtaining a regression model. Specifically, for (39), this Ω transform is $\Omega(x) = \Psi(F(\cdot|X = x))$; for (40) is $\Omega(x) = \int_0^\infty \log \Lambda(s|X = x)\, d\omega(s)$ with ω a nonnegative weight function such that $\int_0^1 d\omega(s) = 1$; for (41), $\Omega(x) = \int_0^\infty \Lambda(s|X = x)\, d\widetilde{\omega}(s)$ with $\widetilde{\omega}(s) = \omega(s)/\int_0^\infty u\, d\omega(u)$; finally, for (43), $\Omega(x) = \int_0^\infty \operatorname{logit}(1 - \exp(-\Lambda(s|X = x)))\, d\omega(s)$.

The key issue is the nonparametric estimation of $\Omega(x)$ via the estimation of $F(\cdot|X = x)$ under censoring and/or truncation. In Iglesias-Pérez and González-Manteiga (1999), a generic estimator for this conditional distribution function was established. Based on an i.i.d. representation of $\hat{F}(\cdot|X = x)$, Cao and González-Manteiga (2008) derive the limit behavior for $nh^{1/2}T_n$, and calibration is performed by Monte Carlo methods. For the complete data case, $\mathbb{E}(Y|X = x) = \Psi(F(\cdot|X = x))$ resulting the discretized version of (5), T_{1n}^D for polynomial regression as a particular case. A generalization of hypothesis (39) is $H_0 : \Psi(F(u|X = x)) = A_u^t(x)\theta$ with time u, was analyzed in Teodorescu and Van Keilegom (2010) and Teodorescu et al. (2010).

All these proposals use a generalized Kaplan–Meier estimator, following Iglesias-Pérez and González-Manteiga (1999), under the assumption of conditional independence between Y and C given X. In the particular case of no truncation and assuming that Y and C are independent and $\mathbb{P}(\delta = 1|X, Y) = \mathbb{P}(\delta = 1|Y)$ (see Stute 1993 for some insight in this assumption), we have $\mathbb{E}(\frac{\delta}{1-G(Z^-)}\Phi(X, Z)|X) = \mathbb{E}(\Phi(X, Z)|X)$, for any functional Φ, with G the distribution of the censoring variable. This allows for an adaptation to the GoF for regression models theory with $\Phi(X, Z) = Z$, considering $\widehat{Y}_i^* = \delta_i Z_i/(1 - \widehat{G}(Z_i))$ the new response values and $\widehat{G}$ the Kaplan–Meier estimator of $G(t) = \mathbb{P}(C \le t)$. In Lopez and Patilea (2009), the test statistic (7) was adapted to this context. That is, under a parametric null hypothesis $H_0 : m \in \mathcal{M}_\theta$, the residual $\widehat{\varepsilon}_{0i} = Y_i - m_{\widehat{\theta}}(X_i)$ is replaced by $\widehat{\varepsilon}_{0i}^* = \widehat{Y}_i^* - m_{\widehat{\theta}}(X_i)$, with $\widehat{\theta} = \arg\min_\theta \sum_{i=1}^n (\widehat{Y}_i^* - m_\theta(X_i))^2$, a least squares estimator as proposed by Koul et al. (1981). This reasoning can be extended to other testing procedures, and also adapted for more complex hypothesis such as additivity (see Debbarh and Viallon 2008). Another alternative is to change the estimation of θ and consider

$$\widetilde{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \big(Z_i - m_\theta(X_i)\big)^2 W_{in} = \arg\min_{\theta} \int \big(y - m_\theta(x)\big)\, d\widehat{F}_n(x, y),$$

$\widehat{F}_n$ being the generalized Kaplan–Meier estimator with covariates (see Stute 1993, 1996, 1999; Stute et al. 2000), and $W_{in} = \frac{\delta_i}{n(1-\widehat{G}(Z_i^-))}$, the Kaplan–Meier weights over the response variable Y. The residuals are now given by $\widehat{\varepsilon}^*_{0i} = \frac{\delta_i}{1-\widehat{G}(Z_i^-)}(Z_i - m_{\widehat{\theta}}(X_i))$, for $i = 1, \ldots, n$.

Tests based on empirical regression processes can also be adapted, estimating the integrated regression function $\mathbb{E}(Y\mathbb{I}(X \leq x))$ by $\int z\mathbb{I}(u \leq x)\, d\widehat{F}_n(u, z)$ (see Stute et al. 2000 and Sánchez-Sellero et al. 2005).

Under the assumption of conditional independence between Y and C given X, there is another alternative to GoF testing, by generating a random variable that estimates $\Upsilon(X, Y)$ properly, being this function $\Upsilon(x, y) = y$ or $\Upsilon(x, y) = (y - m_{\theta_0}(x))^2$, for instance. Thus, $\Upsilon^*(X, Z, \delta) = \Upsilon(X, Z)\delta + \mathbb{E}(\Upsilon(X, Z)|Y > C, X)(1 - \delta)$, and in this case $\mathbb{E}(\Upsilon^*(X, Z, \delta)|X) = \mathbb{E}(\Upsilon(X, Y|X))$. For generating such a random variable, Υ^* is estimated by the conditional Kaplan–Meier under censoring (see Beran 1981 for its introduction and González-Manteiga and Cadarso-Suárez 1994). In González-Manteiga et al. (2007), using the generalized integrated regression $\mathbb{E}(\Upsilon(X, Z)\mathbb{I}(X \leq x))$ with artificial sample $\{(X_i, \widehat{Y}_i^*)\}_{i=1}^n$, some GoF tests based on empirical processes for the regression function were proposed, as well as some other tests for the conditional variance.

The methodology based on the empirical distribution of the residuals, with the conditional independence assumption, can also be extended to this context replacing the residuals obtained with complete data by the following ones:

$$\left\{\frac{Z_i - m_{\widehat{\theta}}(X_i)}{\widehat{\sigma}(X_i)}, \delta_i\right\}_{i=1}^n, \quad \text{and} \quad \left\{\frac{Z_i - \widehat{m}(X_i)}{\widehat{\sigma}(X_i)}, \delta_i\right\}_{i=1}^n.$$

With these residuals, the error distribution was obtained adapting the Kaplan–Meier estimator (see Pardo-Fernández et al. 2007a). See also Dette and Heuchenne (2012) for a test on the variance function in this setting.

6.2 Missing data

Given a regression model in location-scale form, $Y = m(X) + \sigma(X)\varepsilon$, there may be missing data (missing at random, along this section) in the sample, possibly since the response variable Y cannot be observed. Hence, each data is (X_i, Y_i) if Y is observed and $(X_i, \cdot)$ if the response value is not recorded. In order to model this scenario, a new variable δ is introduced, such that $\delta_i = 1$ if Y_i is observed and $\delta_i = 0$ if not. The *missing at random* assumption states that $\mathbb{P}(\delta = 1|Y, X) = \mathbb{P}(\delta = 1|X)$ for $X \in \mathbb{R}^p$. In González-Manteiga and Pérez-González (2006), the test statistic (5) was adapted to this context comparing which is a best option: (a) using the complete information (observed data) through the test statistic

$$n|H|^{1/4}T_{1n}^{M1} = n|H|^{1/4}\int \big(m_{nH}(x) - m_{\widehat{\theta}}(x)\big)^2\omega(x)\,dx,$$

with m_{nH} the (multidimensional) Nadaraya–Watson or the local-linear estimator with smoothing matrix H and $\widehat{\theta}$ a $\sqrt{n}$-consistent estimator with the complete sample, or (b) using imputed values for the unobserved data via the statistic

$$n|H|^{1/4}T_{1n}^{M2} = n|H|^{1/4}\int \big(m^I_{nH\widetilde{H}}(x) - m_{\widehat{\theta}}(x)\big)^2\omega(x)\,dx,$$

with $m^I_{nH\widetilde{H}}$ the smooth estimator with the imputed sample $\{(X_i, \delta_i Y_i + (1 - \delta_i)m_{n\widetilde{H}}(X_i))\}_{i=1}^n$, where a smoothing matrix $\widetilde{H}$ is used for imputation. For the particular case $p = 1$ and complete data, $n|H|^{1/4}T_{1n}^{M1} = n|H|^{1/4}T_{1n}^{M2} = nh^{1/2}T_{1n}$, obtaining the test in Alcalá et al. (1999). The choice between both options (a) and (b) is done based on the ratio $|\widetilde{H}|^{1/2}/|H|^{1/2}$, summarizing if the imputation is convenient. If the ratio tends to infinity as the sample size increases, this indicates that there is oversmoothing in the imputation, leading to large bias, so imputation is not reasonable. On the other hand, if the ratio tends to zero, both tests are equivalent, and the benefit of imputation is achieved when the ratio tends to a positive finite value. Recently, Li (2012) suggested a test for a similar situation but with imputation under the null hypothesis, that is, considering the sample $\{(X_i, \delta_i Y_i + (1 - \delta_i)m_{\widehat{\theta}}(X_i))\}_{i=1}^n$, avoiding a discussion about the pilot smoothing matrix $\widetilde{H}$.

Tests based on empirical processes can also be used with missing data. For instance, the test proposed by Stute (1997) can be used replacing in the empirical process $\overline{\overline{\alpha_n}}$ the response Y_i by the imputed values under the null hypothesis $(\delta_i Y_i + (1 - \delta_i)m_{\widehat{\theta}}(X_i))$ (see Sun and Wang 2009 for the linear model $m_\theta(X) = A^t(X)\theta$, and also Sun et al. 2009 for the extension to partially linear models).

A different problem is the absence of data in the covariates, but literature on this topic is much more scarce. We should mention the work by Zhu et al. (2009) where the methodology based on empirical regression processes was applied to tackle this problem.

6.3 Data with measurement error

Another interesting situation, with quite recent contributions, is GoF for regression when observations are given with some measurement error. Specifically, consider the following regression model $Y = m(X) + \varepsilon$, where the covariate X is not observable and $\mathbb{E}(\varepsilon) = 0$. In this setting, what the practitioner gets is a sample $\{(Z_i, Y_i)\}_{i=1}^n$, with each Z_i an observation from a random variable Z such that $X = Z + \eta$, with $\mathbb{E}(\eta) = 0$. Usually, ε, η and Z are assumed to be mutually independent.

The problem of estimating a parametric model for the regression function (known as Berkson's parametric model (Berkson 1950)) has been widely studied. In order to test a parametric model, that is, testing $H_0 : m \in \mathcal{M}_\theta$, some of the previous test statistics can be adapted. First, denote by f_ε, f_X, f_η and f_Z the density functions of ε, X, η and Z have densities, respectively, and assume that f_η is known.

Then, it is possible to obtain generalized tests. Assume first that f_η is known. Then $f_X(x) = \int f_Z(z) f_\eta(x-z)\,dz$ and a kernel estimator of this density is given by $\hat{f}_X(x) = n^{-1}\sum_{i=1}^n \overline{K}_h(x, Z_i)$ with $\overline{K}_h(x,z) = \int K_h(z-y) f_\eta(x-y)\,dy$.

Define the following functions: $\Psi(z) = \mathbb{E}(m(X)|Z=z)$ and $J(x) = \mathbb{E}(\Psi(Z)|X=x)$. A natural kernel estimator for the regression function is given by

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^n \overline{K}_h(x, Z_i) Y_i}{\sum_{i=1}^n \overline{K}_h(x, Z_i)}.$$

However, this estimator is consistent for $J(x)$ but not for $m(x)$. In order to obtain a test statistic for H_0 it should be noticed that, if this hypothesis holds, then $\Psi(z) = \Psi_\theta(z) = \mathbb{E}(m_\theta(X)|Z=z)$ and $J(x) = J_\theta(x) = \mathbb{E}(\Psi_\theta(Z)|X=x)$. Hence, an adaptation of the test (5) can be constructed as

$$T_{1n}^{\mathrm{ME}} = nh^{p/2}\int \left(\frac{1}{n\hat{f}_X(x)}\sum_{i=1}^n \overline{K}_h(x, Z_i)\big(Y_i - \Psi_{\widehat{\theta}}(Z_i)\big)\right)^2 \omega(x)\,dx,$$

where $\hat{\theta}$ is a consistent estimator for θ under H_0, $\hat{f}_X$ is the kernel density estimator for f_X and $\overline{K}_h(x, Z_i)$ has been introduced above. See details in Song (2008) and Koul and Song (2009).

From another point of view, and bearing in mind that $\Psi(z) = \mathbb{E}(Y|Z=z)$, one may consider the regression model, $Y = \Psi(Z) + \zeta$, with $\mathbb{E}(\zeta|Z) = 0$. In this setting, testing the null hypothesis $H_0 : m \in \mathcal{M}_\theta$ is equivalent to test $H_0 : \Psi \in \{\Psi_\theta\}_{\theta\in\Theta}$. For constructing a test statistic, a nonparametric estimator for Ψ is required and it is given by

$$\widehat{\Psi}_h(z) = \frac{\sum_{i=1}^n K_h(z-Z_i) Y_i}{n\hat{f}_{Z\tilde{h}}(z)}, \quad \text{with } \hat{f}_{Z\tilde{h}}(z) = \frac{1}{n}\sum_{i=1}^n K_{\tilde{h}}^*(z-Z_i),$$

with $K_{\tilde{h}}^*$ a kernel with bandwidth $\tilde{h}$, and the test statistic is

$$T_n = nh^{p/2}\int \left(\frac{1}{n\hat{f}_{Z\tilde{h}}(z)}\sum_{i=1}^n K_h(z-Z_i)\big(Y_i - \Psi_{\widehat{\theta}}(Z_i)\big)\right)^2 \omega(z)\,dz, \tag{44}$$

which extends the test in Koul and Ni (2004), and for $\tilde{h} = h$, the test in Härdle and Mammen (1993). Asymptotic behavior of the test can be seen in Koul and Song (2009).

An alternative way for building test based on smoothing methods is using deconvolution, based on characteristic functions: let f be a density in $\mathbb{R}^p$ with characteristic function Φ_f and consider

$$f_h(x) = \frac{1}{(2\pi)^p}\int_{\mathbb{R}^p} e^{-itx}\frac{\Phi_f(t)}{\Phi_K(t/h)}\,dt,$$

where K is a kernel function in $\mathbb{R}^p$. Again, recall that under the parametric null hypothesis, the function $\Psi(z) = \Psi_\theta(z)$ has also a parametric form. In the context described at the beginning of this section, a deconvolution estimator for f_X is given by

$$\hat{f}_{Xh}(x) = \frac{1}{nh^p} \sum_{i=1}^{n} f_h\left(\frac{x - Z_i}{h}\right).$$

Noticing that

$$\Psi(z) = \frac{\int m(x) f_X(x) f_\eta(z - x)\, dx}{\int f_X(x) f_\eta(z - x)\, dx},$$

an estimator for Ψ under the null hypothesis (assuming that f_η is known), is given by

$$\widehat{\Psi}_{\widehat{\theta}}(z) = \frac{\int m_{\widehat{\theta}}(x) \widehat{f}_{Xh}(x) f_\eta(z - x)\, dx}{\int \widehat{f}_{Xh}(x) f_\eta(z - x)\, dx},$$

and a test statistic similar to T_n in (44) can be constructed. Splitting sample methods for designing test statistics are used by Song (2008), overcoming the slow convergence rate of deconovolution estimators.

All the tests in this section are based on smoothing methods, but it is also possible to extend those tests dealing with empirical regression processes, which may be suitably adapted by taking

$$\overline{\overline{\widetilde{\alpha_n}}}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(Y_i - \Psi_{\widehat{\theta}}(Z_i)\big) \mathbb{I}(Z_i \leq x),$$

or rescaled versions of this process (see Koul and Song 2008). Recently, Koul and Song (2010) proposed a test for the partial linear model in this context. Some additional recent references are Hall and Ma (2007), Carroll et al. (2011), and Ma et al. (2011).

6.4 Length-biased data

In some situations, the sample from (X, Y) associated to the regression model $Y = m(X) + \varepsilon$ is not i.i.d. and what is available is just a biased version from the previous variable: (X^ω, Y^ω) with distribution

$$dF^\omega(x, y) = f^\omega(x, y)\, dx\, dy = \frac{\omega(x, y) f(x, y)}{\mu_y}\, dx\, dy$$

with F and f the distribution and density functions of (X, Y), respectively, and μ_Y the marginal expected value of Y. Selection bias is introduced via ω, whereas F^ω and f^ω are the distribution and density of the biased population. We have

$$\mathbb{E}^\omega(Y|X = x) = m(x)\left(1 + \frac{\mathrm{Cov}(Y, \omega(X, Y)|X = x)}{m(x)\mathbb{E}(\omega(X, Y)|X = x)}\right)$$

and estimation procedures relying on the data sample do not lead to the estimation of the regression function, which makes it difficult to test a null hypothesis such as the parametric one $H_0 : m \in \mathcal{M}_\theta$ based on the biased sample.

The bias sampling problem appears in several applications (see Cox 1969) and estimation methods should account for this fact. For instance, for the particular case of length-biased data, $\omega(x, y) = y$, the least squares estimation of θ is given by

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \frac{1}{Y_i} \big(Y_i - m_\theta(X_i)\big)^2,$$

and the nonparametric local-linear estimator is $\widehat{m}_{nh}(x) = \widehat{\beta}_0(x)$ such that

$$\big(\widehat{\beta}_0(x), \widehat{\beta}_1(x)\big) = \arg\min_{(\beta_0,\beta_1)} \sum_{i=1}^{n} \frac{1}{Y_i} \big(Y_i - \beta_0 - \beta_1(X_i - x)\big)^2 K_h(x - X_i).$$

The test in (5) can be easily extended for length-biased data (see Ojeda et al. 2008). The methodology based on empirical regression processes can also be adapted by setting

$$\overline{\overline{\alpha_n}}^{\omega}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{Y_i} \big(Y_i - m_{\widehat{\theta}}(X_i)\big) \mathbb{I}(X_i \le x)$$

as studied in Ojeda et al. (2011). The use of the empirical distribution of the residuals for testing in this context has been analyzed by Ojeda and Van Keilegom (2009).

A common property of the tests for complex data situations is that convergence rates are similar to those ones for complete data presented in Sect. 1, and a similar discussion can be done about power comparison. Obviously, bootstrap calibration can be suitably adapted accounting for the complexity nature under the null hypothesis. In the references along this section, there are several alternatives for these adjustments.

7 Some related tests: comparison of regression curves and applications

The comparison of two or more groups of variables is one of the principal problems in statistical inference. Checking means, medians or other characteristics of the variable of interest across groups enables this comparison. When the variable of interest Y is accompanied by a regression covariable X a more ambitious objective is to compare the regression functions $m_l(x) = \mathbb{E}(Y_l | X_l = x)$, with $l = 1, \ldots, L$, labeling the groups.

For parametric models in the different groups, $m_l = m_{\theta_l}$, with $\theta_l \in \Theta \subset \mathbb{R}^q$, one may perform a classical covariance analysis. However, when no parametric assumptions are made and $\{m_l\}_{l=1}^{L}$, with $m_l \in \mathcal{M}$ and $\mathcal{M}$ a functional space satisfying some regularity conditions, the problem is more complicated.

For the fixed design case, the regression model can be compactly written as

$$Y_{li} = m_l(t_{li}) + \sigma_l(t_{li})\varepsilon_{li}, \quad l = 1, \ldots, L,\ i = 1, \ldots, n_l \tag{45}$$

where ε_{li} are zero mean i.i.d. random variables, m_l and σ_l^2 are the regression and variance function in the lth group and $t_{li} \in [0, 1]$, without loss of generality and $N = \sum_{l=1}^{L} n_l$. The testing problem $H_0 : m_1 = \cdots = m_L$ vs. $H_1 : \exists(k, l)$ such that

$m_k \neq m_l$, can be approached from different perspectives, specifically, using smoothing methods, based on empirical regression processes or considering the empirical distribution of the residuals.

An L^2 distance may be used to compare the regression curves in the groups. The discrepancy $D = \sum_{k<l} \int_0^1 (m_k(t) - m_l(t))^2 \, dt$ is null iff the null hypothesis of equality holds and a test statistic which estimates D is

$$T_N = \sum_{k<l} \int_0^1 \big(\widehat{m}_k(t) - \widehat{m}_l(t)\big)^2 \omega_{kl}(t) \, dt,$$

where $\{\omega_{kl}\}_{k<l}$ are weight functions and $\widehat{m}_l(t) = \sum_{i=1}^{n_l} W_{li}(t) Y_{li}$ is the nonparametric kernel estimator of the regression function in each group $l = 1, \ldots, L$.

A general result on the previous test statistic distribution is derived in Vilar-Fernández and González-Manteiga (2004) for model (45), considering a correlation structure in the error. Specifically, for $L = 2$, we have

$$nh^{1/2}\left(T_N - \frac{1}{nh} \int \omega(t) \, dt \, \Gamma_\Delta \int K^2(t) \, dt\right) \xrightarrow{d} \mathcal{N}\big(0, \sigma_T^2\big),$$

with $\Gamma_\Delta = \sum_{k=-\infty}^{\infty} C_\Delta(k)$ and $\sigma_T^2 = 2\Gamma_\Delta^2 \int (K * K)^2(t) \, dt \int \omega^2(t) \, dt$, with C_Δ denoting the covariance of the error differences between two groups, with $n_1 = n_2 = n$ and $t_{li} = i/n_l$ for $l = 1, 2$.

Similar asymptotic results can be obtained for other tests, for instance, considering the variance difference test in (8), with corresponding test statistic

$$T_{3N} = \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \big(Y_{li} - \widehat{m}(t_{li})\big)^2 - \frac{1}{N} \sum_{l=1}^{L} n_l \widehat{\sigma}_l^2,$$

where $\widehat{\sigma}_l^2 = n_l^{-1} \sum_{i=1}^{n_l} (Y_{li} - \widehat{m}_l(t_{li}))^2$. An ANOVA-type test statistic such as

$$T_N^{\text{ANOVA}} = \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{n_l} \big(\widehat{m}(t_{li}) - \widehat{m}_l(t_{li})\big)^2$$

is also an alternative. Asymptotic results can also be derived studying the behavior of U-statistics with kernels changing with the sample size.

Suppose that there is a common covariate X for two groups (with $n_1 = n_2 = n$), that is, the samples are observed from (X, Y^1) and (X, Y^2), and the goal is to test if

$$H_0 : \mathbb{E}\big(Y^1 | X = x\big) = \mathbb{E}\big(Y^2 | X = x\big), \quad \forall x \in I \subset \mathbb{R},$$

I being any subinterval in $\mathbb{R}$. This is equivalent to test $H_0 : \mathbb{E}(Y^1 - Y^2 | X = x) = 0$, for all $x \in I$. The empirical regression process for this case can be built as

$$\alpha_N(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}(X_i \leq x)\big(Y_i^1 - Y_i^2\big),$$

which has an asymptotic Gaussian limit process. The different class of tests can be obtained as continuous functionals of this process. See Delgado (1993) or, more recently, Ferreira and Stute (2004) for dependent data.

However, the covariates in each group can be different and random, so a more general test can be built on the following empirical regression process:

$$\overline{\alpha_N}(x) = \frac{1}{N}\sum_{i=1}^{n_1} f_{1i}\mathbb{I}(X_{1i} \le x) - \frac{1}{N}\sum_{i=1}^{n_2} f_{2i}\mathbb{I}(X_{2i} \le x), \quad \text{with}$$

$$f_{li} = \frac{N}{n_l}\frac{Y_{li} - \widehat{m}(X_{li})}{\hat{f}_l(X_{li})},$$

for $l = 1, 2$ and $i = 1, \dots, n_l$, $\hat{f}_l$ being the corresponding kernel density estimator for the covariate (see Neumeyer and Dette 2003).

In more recent papers, there are more complex testing procedures, for instance, comparing autoregressive time series models. Consider $\{X_t\}$, $\{Y_t\}$ two location-scale time series:

$$X_t = m_1(X_{t-1}, \dots, X_{t-k}) + \sigma_1(X_{t-1}, \dots, X_{t-k})\varepsilon_t, \quad t = 1, \dots, n_1,$$

$$Y_t = m_2(Y_{t-1}, \dots, Y_{t-k}) + \sigma_2(Y_{t-1}, \dots, Y_{t-k})\eta_t, \quad t = 1, \dots, n_2,$$

with ε_t and η_t zero-mean innovations. The testing problem $H_0 : m_1 = m_2$ is studied in Dette and Weissbach (2009).

Finally, the empirical distribution of the residuals can also be used for comparing regression curves. Consider again a location-scale regression model

$$Y_l = m_l(X_l) + \sigma_l(X_l)\varepsilon_l, \quad l = 1, \dots, L$$

with a random sample $\{(X_{li}, Y_{li})\}_{i=1,\dots,n_l}$, for $l = 1, \dots, L$. Testing the equality of the regression curves in the L groups can be done by comparing the empirical distribution of the residuals

$$\widehat{F}_{\varepsilon_l}(y) = \frac{1}{n_l}\sum_{i=1}^{n_l}\mathbb{I}\left(\frac{Y_{li} - \widehat{m}_l(X_{li})}{\widehat{\sigma}_l(X_{li})} \le y\right)$$

with the corresponding estimators under the null hypothesis

$$\widehat{F}_{\varepsilon_{l0}}(y) = \frac{1}{n_l}\sum_{i=1}^{n_l}\mathbb{I}\left(\frac{Y_{li} - \widehat{m}(X_{li})}{\widehat{\sigma}_l(X_{li})} \le y\right),$$

with $\widehat{m}_l, \widehat{\sigma}_l$ the Nadaraya–Watson estimators for the regression and variance functions in each group and $\widehat{m}$ the estimator under $H_0 : m_1 = \cdots = m_L$, with the whole sample. Based on the empirical process $\alpha_l(y) = n_l^{1/2}(\widehat{F}_{\varepsilon_l}(y) - \widehat{F}_{\varepsilon_{l0}}(y))$, $l = 1, \dots, L$, the Kolmogorov–Smirnov and Cramér–von Mises tests can be constructed as

$$T_{\mathrm{NKS}} = \sum_{l=1}^{L}\sup_y\left|\alpha_l(y)\right|, \qquad T_{\mathrm{NCM}} = \sum_{l=1}^{L}\int \alpha_l^2(y)\, d\widehat{F}_{\varepsilon_{l0}}(y).$$

See Pardo-Fernández et al. (2007b) for details on these tests. See also Pardo-Fernández (2007) for tests on the equality of the error distributions, extending the work by Mora (2005) for linear m and Pardo-Fernández and Van Keilegom (2006) for the equality of regression curves with censored response. The use of the SiZer exploratory tool is illustrated in Park and Kang (2008).

From any of the three previous testing approaches (based on smoothing methods, on empirical processes or on the empirical distribution of the residuals), a resampling calibration procedure is required. In Vilar-Fernández et al. (2007), a comparative review of different bootstrap methods for fixed design and correlated errors is provided.

The problem of comparing nonparametric regression curves has been previously considered by Hall and Hart (1990), who proposed a bootstrap test and by Härdle and Marron (1990). A test for the equality of two regression curves has been proposed by King et al. (1991) based on the difference between linear but nonparametric estimators. Srihera and Stute (2010) based their approach on a weighted comparison of nonparametric estimators. Nonparametric analysis of covariance has been initially performed by Young and Bowman (1995). Dette and Neumeyer (2001) discussed different procedures for testing the equality of a collection of regression curves. Munk and Dette (1998) develop exact and asymptotic theory for comparing regression functions whereas Munk et al. (2007) extended the results for inhomogeneous and heteroskedastic errors. See also Neumeyer and Sperlich (2006) for an example where calibration must be done via subsampling. The comparison of regression curves using quasi-residuals was studied by Kulasekera (1995), who also considered the problem of smoothing parameter selection for maximizing the power of the test (see Kulasekera and Wang 1997, 1998). The problem of unbalanced groups has been considered by Munk and Dette (1998) and by Lavergne (2001). Munk and Dette (1998) and Hall et al. (1997) have been also concerned with the implications of different regressors designs and variation among covariates. More recently, Lin and Kulasekera (2010) addressed the comparison of nonparametric regression curves for single index models.

The comparison of regression curves can be useful in other contexts such as time series (see Dette and Paparoditis 2009) or spatio-temporal models, allowing for a comparison of correlation structures from the spectral domain. Based on the periodogram representation (36) Crujeiras et al. (2007, 2008) proceeded to the comparison of spatial spectral density by means of an L^2 distance.

8 Some recent directions

Just to finish this survey, some advances in random effects models, quantile, and functional regression will be presented. For each scenario, some of the current challenges in the development of GoF tests will be briefly discussed.

8.1 Testing regression models with random effects

Consider a regression model

$$\mathcal{H}\big(\mathbb{E}(Y_{li}|u_l, T_{li}, X_{li})\big) = m(T_{li}) + X_{li}^t\beta + Z_{li}^t u_l, \quad \text{where } l = 1, \ldots, L \tag{46}$$

denotes the number of sample subgroups, n_l is the sample size in each of them ($N = \sum_{l=1}^{L} n_l$) and $\{(Y_{li}, X_{li}, Z_{li}, T_{li})\}_{i=1}^{n_l}$ are the corresponding observations in each group, with dimension $(1 + p + r + q)$, X_{li} and Z_{li} being observable with $Z_{li} \subseteq (1, X_{li})^t$. In (46), $\{u_l\}_{l=1}^{L}$ represents the (independent) random effects, Y_{li} ($i = 1, \ldots, n_l$) are conditionally independent given (u_l, T_{li}, X_{li}) and Σ_u is the covariance matrix for the random effects. Finally, $\mathcal{H}$ denotes the link function, which is usually unknown.

For instance, Pan and Lin (2005) considered the case $m(T) = 0$ and test the hypothesis (46) with empirical regression processes. In Henderson et al. (2008), testing of the randomness of u_d with kernel smoothing, for $Z_{li} = 1$, was considered. See also Lombardía and Sperlich (2008) for tests about m extending the works by Härdle et al. (1998) and Müller (2001) to the random effects context. Sperlich and Lombardía (2010) adapted the previous study to random effects in small areas, with local polynomial smoothing. Based on the empirical distribution of the residuals, see Sánchez et al. (2009) and Meintanis and Portnoy (2011) for a test on the random effect of u_l based on the characteristic function of the residuals.

Calibration of the test statistics distribution for random effects model poses some challenges, given that in the resampling process, not only the error distribution must be taken into account, but also the random effect involved in the model.

8.2 Testing about quantile functions

Until now, this review has considered a regression function as $m(x) = \mathbb{E}(Y|X = x)$ with $m(x) = \hat{a}$ such that $\hat{a} = \arg\min_a \mathbb{E}((Y - a)^2|X = x)$. However, the interest may be focused on the quantile function

$$Q_p(x) = \inf\big\{y|F(y|X = x) \geq p\big\}, \quad p \in (0, 1)$$

which results from minimizing $\mathbb{E}(\rho_p(y - a) - \rho_p(y)|X = x)$, with $\rho_p(\varepsilon) = p\varepsilon^+ + (1 - p)\varepsilon^-$, ε^+ and ε^- being the positive and negative parts of ε, respectively (see Koenker and Basset 1978). Under continuity of the conditional distribution, it can be seen that $\mathbb{E}(\mathbb{I}(Y \leq Q_p(X))|X = x) = p$, motivating the development of GoF tests for the conditional p-quantile function. Setting the model under the null hypothesis as

$$H_0 : Y_i = Q_p(X_i) + u_i = m(X_i, \theta_0) + u_i, \quad i = 1, \ldots, n \tag{47}$$

for a random sample $\{(X_i, Y_i)\}_{i=1}^{n}$ with $\theta_0 \in \mathbb{R}^q$ and with the conditional quantile of u_i given x_i as $Q_p(x_i) = m_{\theta_0}(x_i)$, it can be proved that (47) can be equivalently written as $H_0 : \mathbb{E}(\mathbb{I}(Y_i \leq m_{\theta_0}(X_i))|X_i) = p$, for $i = 1, \ldots, n$. Hence, taking $\widehat{\varepsilon}_{i0} = \mathbb{I}(Y_i \leq m_{\widehat{\theta}}(X_i)) - p$, Zheng (1998) extended the test statistic (7) to this context. Quite recently, Wang (2008) adapted the procedure for censored response. An adaptative optimal rate for $p = 0.5$ was given by Horowitz and Spokoiny (2002).

He and Zhu (2003) consider the alternative based on empirical processes, taking

$$\alpha_n^p(t) = n^{-1/2} \sum_{j=1}^{n} \Psi(\varepsilon_i) X_j \mathbb{I}(X_j \leq t),$$

as an empirical process generator for linearity tests on $Q_p(x)$, where $\Psi(r) = p\mathbb{I}(r > 0) + (p-1)\mathbb{I}(r < 0)$ and $\varepsilon_i = (Y_i - \widehat{\theta}^t X_i)$. More recently, based on the proposal in Härdle and Mammen (1993) but applying Jackniffe bias reduction methods on the local kernel estimator of the quantile function, Zhou (2010) extended the theory of this type of tests to nonstationary time series. See also Escanciano and Velasco (2010) for generalizations of the empirical process approach to dependent data for dynamic quantile functions. Finally, for comparisons between quantile curves, see Sun (2006) and Dette et al. (2011).

There is a quite extensive literature in quantile regression estimation, and more contributions are expected briefly. There is still advantage to take from the extension and adaptation of quantile regression to other settings, such as those dealing with incomplete information or complex data.

8.3 Testing with functional data

Functional data analysis (FDA) has deserved an increasing attention in this last decade, specially motivated by the practical needs derived from dealing with high-dimensional data (see Ramsay and Silverman 2005 and Ferraty and Vieu 2006). Among the techniques, both exploratory and inferential, gathered under the FDA framework, regression models with functional regressors and (possibly) functional response have been also introduced. Consider the functional regression model

$$Y(t) = m\big(t, \eta(Z)\big) + \varepsilon(t),$$

where Y is a function response variable with $t \in \mathcal{T}$ an interval and η is a function of the covariate Z that may also be functional, and

$$\mathbb{E}\big(Y(t)\big) = \mathbb{E}\big(m\big(t, \eta(Z)\big)\big) = \mu_Y(t), \quad \text{with} \quad \mathbb{E}\big(\varepsilon(t)\big) = 0.$$

Although there are several works on the estimation of this model (see Ferraty and Romain 2010 for a recent handbook), the literature concerning GoF tests is scarce. Cuesta-Albertos et al. (2007) obtained some results on GoF for the distribution of functional variables and Chiou and Muller (2007) developed some tests based on functional residuals, $R(t) = Y(t) - \widehat{Y}(t)$, $t \in \mathcal{T}$, with the linear functional model (with functional response) as a particular case:

$$Y(t) = \mu_Y(t) + \int_{\mathcal{J}} \beta(s,t)\big(Z(s) - \mu_Z(s)\big)\,ds,$$

$\mathcal{J}$ being the set of indices characterizing the functional data Z.

In Bücher et al. (2011), the authors state some tests for a fixed design model with covariates in the unit interval, extending the tests developed by Dette (1999) for scalar response.

Following some of the ideas of Ferraty and Vieu (2006), considering pseudometrics for measuring distances between functional data, the test proposed by Härdle and Mammen (1993) was adapted by Delsol et al. (2011a, 2011b) for the scalar response model $Y = m(Z) + \varepsilon$ (Z being a functional covariate), taking as test statistic

$$T_n = \int \left(\sum_{i=1}^{n} \big(Y_i - \widehat{m}_{H_0}(Z_i)\big) K\left(\frac{d(z, Z_i)}{h}\right) \right)^2 \omega(z)\, dP_Z(z),$$

d being a pseudometric, K a kernel and P_Z the probability measure over the functional space of the covariate.

Some of the challenges in testing for FDA is the extension of empirical processes based GoF tests given that, in this setting, empirical processes will be indexed on an infinite-dimensional space.

This brief section is just an example of some of the ongoing GoF tests research in different statistical areas. It can be clearly seen how the developments on GoF methods for regression have been adapted to other settings, deserving also attention in the analysis of other characteristic curves, such as ROC curves, or with other data scenarios, such as directional data. In addition, some advances on statistical inference for GoF have been also made in order to obtain robustified tests. Although a complete and detailed review of all the contributions on GoF literature, specially the most recent ones, is beyond our means, we hope that this review will provide some insights on GoF tests for regression and their adaptability to other contexts.

Acknowledgements This research has been supported by Project MTM2008-03010 from the Spanish Ministry of Science and Innovation, and by the IAP network StUDyS (Developing crucial Statistical methods for Understanding major complex Dynamic Systems in natural, biomedical, and social sciences), from Belgian Science Policy. The authors also thank the comments and suggestions from the Associate Editor and an anonymous referee.

References

Ahmad IA, Cerrito PB (1993) Goodness-of-fit tests based on the L_2-norm of multivariate probability density functions. J Nonparametr Stat 2:169–181

Aït-Sahalia Y (1996) Testing continuous time models of the spot interest rate. Rev Financ Stud 9:385–426

Aït-Sahalia Y, Fan J, Jiang J (2010) Nonparametric test of the Markov hypothesis in continuous–time models. Ann Stat 38:3129–3163

Akritas M, Van Keilegom I (2001) Nonparametric estimation of the residual distribution. Scand J Stat 28:549–567

Albers CJ, Schaafsma W (2008) Goodness of fit testing using a specific density estimate. Stat Decis 26:3–23

Alcalá T, Cristóbal JA, González-Manteiga W (1999) Goodness of fit test for linear models based on local polynomials. Stat Probab Lett 42:39–46

Aneiros-Pérez G, González-Manteiga W, Vieu P (2004) Estimation and testing in a partial linear regression model under long-memory dependence. Bernoulli 10:49–78

Arapis M, Gao J (2006) Empirical comparisons in short–term interest rate models using nonparametric methods. J Financ Econom 4:310–345

Azzalini A, Bowman AW (1993) On the use of nonparametric regression for checking linear relationships. J R Stat Soc B 55:549–557

Azzalini A, Bowman AW, Härdle W (1989) On the use of nonparametric regression models. Biometrika 76:1–11

Bachmann D, Dette H (2005) A note on the Bickel-Rosenblatt test in autoregressive time series. Stat Probab Lett 74:221–234

Beran R (1981) Nonparametric regression with randomly censored survival data. Technical report, University of California, Berkeley

Berkson J (1950) Are these two regressions? J Am Stat Assoc 45:164–180
Bickel PJ, Rosenblatt M (1973) On some global measures of the deviations of density function estimates. Ann Stat 1:1071–1095
Biederman S, Dette H (2000) Testing linearity of regression models with dependent errors by kernel-based methods. Test 9:417–438
Biederman S, Dette H (2001) Optimal designs for testing the functional form of a regression via nonparametric estimation techniques. Stat Probab Lett 52:215–224
Bierens HJ (1982) Consistent model especification tests. J Econom 20:105–134
Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer, Berlin
Bücher A, Dette H (2010) Some comments on goodness-of-fit tests for the parametric form of the copula based on L^2-distances. J Multivar Anal 101:749–763
Bücher A, Dette H, Wieczorek G (2011) Testing model assumptions in functional regression models. J Multivar Anal 102:1472–1488
Cao R, González-Manteiga W (1993) Bootstrap methods in regression smoothing. J Nonparametr Stat 2:379–388
Cao R, González-Manteiga W (2008) Goodness-of-fit tests for conditional models under censoring and truncation. J Econom 143:166–190
Cao R, Lugosi G (2005) Goodness-of-fit tests based on the kernel density estimate. Scand J Stat 32:599–616
Carroll RJ, Ruppert D (1988) Transformation and weighting in regression. Chapman and Hall, New York
Carroll RJ, Delaigle A, Hall P (2011) Testing and estimating shape-constrained nonparametric density and regression in the presence of measurement error. J Am Stat Assoc 106:191–202
Chabot-Hallé D, Duchesne P (2008) Diagnostic checking of multivariate nonlinear time series models with martingale difference errors. Stat Probab Lett 78:997–1005
Chebana F (2004) On the optimization of the weighted Bickel–Rosenblatt test. Stat Probab Lett 68:333–345
Chebana F (2006) Functional asymptotic normality of the L_2-deviation of the kernel density estimation indexed by classes of weight functions. J Stat Plan Inference 136:2470–2505
Chen SX, Cui H (2003) An extended empirical likelihood for generalized linear models. Stat Sin 13:69–81
Chen SX, Gao J (2007) An adaptive empirical likelihood test for parametric time series regression models. J Econom 141:950–972
Chen SX, Van Keilegom I (2009a) A goodness of fit test for parametric and semiparametric models in multiresponse regression. Bernoulli 15:955–976
Chen SX, Van Keilegom I (2009b) A review on empirical likelihood methods for regression. Test 18:415–447
Chen SX, Härdle W, Li M (2003) An empirical likelihood goodness of fit test for time series. J R Stat Soc B 65:663–678
Chen SX, Gao J, Tang CY (2008) A test for model specification of difusion processes. Ann Stat 36:167–198
Chiou JM, Muller HG (2007) Diagnostics for functional regression via residual processes. Comput Stat Data Anal 51:4849–4863
Corradi V, Swanson N (2005) Bootstrap specification tests for diffusion processes. J Econom 124:117–148
Corradi V, White H (1999) Specification tests for the variance of a difussion. J Time Ser Anal 20:253–270
Cox DR (1969) Some sampling problems in technology. In: New developments in survey sampling. Springer, Berlin, pp 506–527
Cox DR (1972) Regression models and life tables. J R Stat Soc B 34:187–200
Cressie N (1993) Statistics for spatial data. Wiley, New York
Crujeiras RM, Fernández-Casal R, González-Manteiga W (2007) Comparing spatial dependence structures using spectral density estimators. Environmetrics 18:793–808
Crujeiras RM, Fernández-Casal R, González-Manteiga W (2008) An l2-test for comparing spatial spectral densities. Stat Probab Lett 78:2543–2551
Crujeiras RM, Fernández-Casal R, González-Manteiga W (2010a) Goodness-of-fit tests for the spatial spectral density. Stoch Environ Res Risk Assess 24:67–78
Crujeiras RM, Fernández-Casal R, González-Manteiga W (2010b) Nonparametric test for separability of spatio–temporal processes. Environmetrics 21:382–399
Cuesta-Albertos JA, del Barrio E, Fraiman R, Matrán C (2007) The random projection method in goodness of fit for functional data. Comput Stat Data Anal 51:4814–4831
de Jong P (1987) A central limit theorem for generalized quadratic forms. Probab Theory Relat Fields 75:261–277

Debbarh M, Viallon V (2008) Testing additivity in nonparametric regression under random censorship. Stat Probab Lett 78:2584–2591

Delgado M (1993) Testing the equality of nonparametric regression curves. Stat Probab Lett 17:199–204

Delgado M, González-Manteiga W (2001) Significance testing in nonparametric regression based on the bootstrap. Ann Stat 29:1469–1507

Delgado M, Velasco C (2010) Distribution–free tests for time series models specification. J Econom 108:25–42

Delgado M, Hidalgo J, Velasco C (2005) Distribution free goodness-of-fit tests for linear processes. Ann Stat 33:2568–2609

Delsol L, Ferraty F, Vieu P (2011a) Structural test in regression on functional variables. J Multivar Anal 102:422–447

Delsol L, Ferraty F, Vieu P (2011b) Focusing on structural assumptions in regression on functional variable. In: Recent advances in functional data analysis and related topics. Springer, Berlin, pp 77–82

Derbort S, Dette H, Munk A (2002) A test for additivity in nonparametric regression. Ann Inst Stat Math 54:60–82

Dette H (1999) A consistent test for the functional form of a regression based on a difference of variance estimators. Ann Stat 27:1012–1040

Dette H (2002) A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. J Stat Plan Inference 103:311–329

Dette H, Heuchenne C (2012) Scale checks in censored regression. Scand J Stat 39:323–339

Dette H, Hildebrandt T (2012) A note on testing hypotheses for stationary processes in the frequency domain. J Multivar Anal 104:101–114

Dette H, Marchlewski M (2008) A test for the parametric form of the variance function in a partial linear regression model. J Stat Plan Inference 138:3005–3021

Dette H, Munk A (1998) Testing heterocedasticity in nonparametric regression. J R Stat Soc B 60:693–708

Dette H, Neumeyer N (2001) Nonparametric analysis of covariance. Ann Stat 29:1361–1400

Dette H, Paparoditis W (2009) Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities. J R Stat Soc B 71:831–857

Dette H, Podolskij M (2008) Testing the parametric form of the volatility in continuous time diffusion models—a stochastic process approach. J Econom 143:56–73

Dette H, von Lieres und Wilkau C (2001) Testing additivity by kernel methods—what is a reasonable test? Bernoulli 7:669–697

Dette H, von Lieres und Wilkau C (2003) On a test for a parametric form of volatility in continuous time financial models. Finance Stoch 7:363–384

Dette H, Weissbach R (2009) A bootstrap test for the comparison of nonlinear time series. Comput Stat Data Anal 53:1339–1349

Dette H, von Lieres und Wilkau C, Sperlich S (2005) A comparison of different nonparametric methods for inference on additive models. J Nonparametr Stat 17:57–81

Dette H, Podolskij M, Vetter M (2006) Estimation of interpreted volatility in continuous–time financial models with applications to goodness-of-fit. Scand J Stat 33:259–278

Dette H, Neumeyer N, Van Keilegom I (2007) A new test for the parametric form of the variance function in nonparametric regression. J R Stat Soc B 69:903–971

Dette H, Wagener J, Volgushev S (2011) Comparing conditional quantile curves. Scand J Stat 38:63–88

Diblasi AM, Bowman AW (1997) Testing for constant variance in a linear model. Stat Probab Lett 33:95–103

Diblasi AM, Bowman AW (2001) On the use of the variogram in checking for independence in spatial data. Biometrics 57:211–218

Diblasi AM, Maglione D (2004) Exploring a valid model for the variogram of an isotropic spatial process. Stoch Environ Res Risk Assess 18:366–376

Diebolt J (1995) A nonparametric test for the regression function: asymptotic theory. J Stat Plan Inference 44:1–17

Diebolt J, Zuber J (1999) Goodness-of-fit tests for nonlinear heteroscedastic regression models. Stat Probab Lett 42:53–60

Diebolt J, Zuber J (2001) On testing goodness-of-fit of nonlinear heteroscedastic regression models. Commun Stat, Simul Comput 30:195–216

Dominguez MA, Lobato IN (2003) Testing the martingale difference hypothesis. Econom Rev 22:351–377

Durbin J (1973) Weak convergence of the sample distribution function when parameters are estimated. Ann Stat 1:279–290

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann Stat 7:1–26
Eichler M (2008) Testing nonparametric and semiparametric hypothesis in vector stationary processes. J Multivar Anal 99:968–1009
Escanciano JC (2004) Contrastes de especificación en modelos econométricos de series temporales. PhD thesis, University Carlos III
Escanciano JC (2006a) Goodness of fit tests for linear and nonlinear time series models. J Am Stat Assoc 101:531–541
Escanciano JC (2006b) A consistent diagnostic test for regression models using projections. Econom Theory 22:1030–1051
Escanciano JC (2007a) Model checks using residual marked empirical processes. Stat Sin 17:115–138
Escanciano JC (2007b) Weak convergence of non-stationarity multivariate marked processes with applications to martingale testing. J Multivar Anal 98:1321–1336
Escanciano JC (2009) On the lack of power of omnibus specification tests. Econom Theory 25:162–194
Escanciano JC, Song K (2009) Testing single–index restrictions with a focus on average derivatives. J Econom 156:377–391
Escanciano JC, Velasco C (2006a) Generalized spectral tests for the martingale difference hypothesis. J Econom 134:151–185
Escanciano JC, Velasco C (2006b) Testing the martingale difference using integrated regression functions. Comput Stat Data Anal 51:2278–2294
Escanciano JC, Velasco C (2010) Specification tests of parametric dynamic conditional quantiles. J Econom 159:209–221
Eubank RL, Hart J (1992) Testing goodness-of-fit in regression via ordered selection criteria. Ann Stat 20:1412–1425
Eubank RL, Hart J (1993) Commonality of Cusum, von Neumann and smoothing-based goodness-of-fit tests. Biometrika 80:89–98
Eubank RL, LaRiccia V (1993) Testing for no effect in nonparametric regression. J Stat Plan Inference 36:1–14
Eubank RL, Hart J, Simpson DP, Stefanski L (1995) Testing for additivity in nonparametric regression. Ann Stat 23:1896–1920
Eubank RL, Ching-Shang L, Wang S (2005) Testing lack of fit of parametric regression models using nonparametric regression techniques. Stat Sin 15:135–152
Fan J, Gijbels I (1996) Local polynomial modelling and its applications, Monographs on statistics and applied probability. Chapman & Hall, London
Fan J, Jiang J (2005) Nonparametric inference for additive models. J Am Stat Assoc 100:890–907
Fan J, Jiang J (2007) Nonparametric inference with generalized likelihood ratio tests. Test 16:409–444
Fan J, Zhang C (2003) A reexamination of diffusion estimators with applications to financial model validation. J Am Stat Assoc 98:118–134
Fan J, Zhang W (2004) Generalized likelihood ratio tests for spectral density. Biometrika 89:195–209
Fan J, Zhang C, Zhang J (2001) Generalised likelihood ratio statistics and Wilks phenomenon. Ann Stat 29:153–193
Fan J, Jiang J, Zhang C, Zhou Z (2003) Time–dependent diffusion models for term structure dynamics. Stat Sin 13:965–992
Fan Y (1994) Testing the goodness of fit of a parametric density function by the kernel method. Econom Theory 10:316–356
Fan Y (1998) Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. Econom Theory 14:604–621
Fan Y, Li Q (1996) Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64:865–890
Fan Y, Linton O (2003) Some higher–order theory for a consistent non-parametric model specification test. J Stat Plan Inference 109:125–154
Fermanian JD (2005) Goodness-of-fit tests for copulas. J Multivar Anal 95:119–152
Ferraty F, Romain Y (2010) The Oxford handbook on functional data analysis. Oxford University Press, Oxford
Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer, New York
Ferreira E, Stute W (2004) Testing for differences between conditional means in a time series context. J Am Stat Assoc 99:169–174
Franke J, Kreiss JP, Mammen E (2002) Bootstrap of kernel smoothing in nonlinear time series. Bernoulli 8:1–38

Gao J (2007) Nonlinear time series. Semiparametric and nonparametric methods. Chapman and Hall, London
Gao J, Casas I (2008) Specification testing in continuous–time diffusion models. Theory and practice. J Econom 147:131–140
Gao J, Gijbels I (2008) Bandwidth selection in nonparametric kernel testing. J Am Stat Assoc 103:1584–1594
Gao J, King M (2004) Adaptative testing in continuous–time diffussion models. Econom Theory 20:844–882
Gao J, King Z, Lu M, Tjøstheim D (2009) Specification testing in nonlinear and nonstationary time series autoregression. Ann Stat 37:3893–3928
Gasser T, Müller HG (1979) Kernel estimation of regression functions. In: Smoothing techniques for curve estimation. Lecture notes in mathematics, vol 757. Springer, Berlin
Gijbels I, Rousson V (2001) A nonparametric least-squares test for checking a polynomial relationship. Stat Probab Lett 51:253–261
Giné E, Mason DM (2004) The law of the iterated logarithm for the integrated squared deviation of a kernel density estimator. Bernoulli 4:721–752
González-Manteiga W, Aneiros-Pérez G (2003) Testing in partial linear regression models with dependent errors. J Nonparametr Stat 15:93–111
González-Manteiga W, Cadarso-Suárez C (1994) Asymptotic properties of a generalized Kaplan–Meier estimator with some applications. J Nonparametr Stat 4:65–78
González-Manteiga W, Cao R (1993) Testing the hypothesis of a general linear model using nonparametric regression estimation. Test 2:161–188
González-Manteiga W, Pérez-González A (2006) Goodness-of-fit tests for linear regression models with missing response data. Can J Stat 34:149–170
González-Manteiga W, Vilar-Fernández J (1995) Testing linear regression models using non-parametric regression estimators when errors are non-independent. Comput Stat Data Anal 20:521–541
González-Manteiga W, Quintela-del Río A, Vieu P (2002) A note on variable selection in nonparametric regression with dependent data. Stat Probab Lett 57:259–268
González-Manteiga W, Heuchenne C, Sánchez-Sellero C (2007) Parametric conditional mean and variance testing with censored data. In: Recent advances in applied stochastic models and data analysis. World Scientific, Singapore
Gouriéroux C, Tenreiro C (2001) Local power properties of kernel based goodness of fit tests. J Multivar Anal 78:161–190
Gozalo PL, Linton O (2001) Testing additivity in generalized nonparametric regression models with estimated parameters. J Econom 104:1–48
Grigoletto M, Akritas MG (1999) Analysis of covariance with incomplete data via semiparametric model transformations. Biometrics 55:1177–1187
Gu J, Li D, Liu D (2007) Bootstrap nonparametric significance test. J Nonparametr Stat 19:215–230
Guerre E, Lavergne P (2002) Optimal minimax rates for nonparametric specification testing in regression models. Econom Theory 18:1139–1171
Guerre E, Lavergne P (2005) Data-driven rate optimal specification testing in regression models. Ann Stat 33:840–870
Guyon X (1982) Parameter estimation for a stationary process on a d-dimensional lattice. Biometrika 69:95–105
Hall P, Hart JD (1990) Bootstrap test for difference between means in nonparametric regression. J Am Stat Assoc 85:1039–1049
Hall P, Ma Y (2007) Testing the suitability of polynomial models in error-in-variables problems. Ann Stat 35:2620–2638
Hall P, Yatchew A (2005) Unified approach to testing functional hypotheses in semiparametric contexts. J Econom 127:225–252
Hall P, Huber C, Speckman PL (1997) Covariate-matched one-sided tests for the difference between functional means. J Am Stat Assoc 92:1074–1083
Härdle W, Mammen E (1993) Comparing nonparametric versus parametric regression fits. Ann Stat 21:1926–1947
Härdle W, Marron JS (1990) Semiparametric comparison of regression curves. Ann Stat 18:63–89
Härdle W, Mammen E, Müller M (1998) Testing parametric versus semiparametric modeling in generalized linear models. J Am Stat Assoc 93:1461–1474
Härdle W, Sperlich S, Spokoiny V (2001) Structural test in additive regression. J Am Stat Assoc 96:1333–1347

Hart J (1997) Nonparametric smoothing and lack-of-fit tests. Springer, Berlin
Hart J, Wehrly JE (1992) Kernel regression when the boundary region is large, with application to testing the adequacy of polynomial models. J Am Stat Assoc 87:1018–1024
He X, Zhu LX (2003) A lack of fit test for quantile regression. J Am Stat Assoc 98:1013–1022
Henderson PJ, Carroll RJ, Li Q (2008) Nonparametric estimation and testing of fixed effects panel data models. J Econom 144:257–275
Heuchenne C, Van Keilegom I (2010) Goodness of fit tests for the error distribution in nonparametric regression. Comput Stat Data Anal 54:1942–1951
Hidalgo J (2008) Specification testing for regression models with dependent data. J Econom 143:143–165
Hidalgo J (2009) Goodness of fit for lattice processes. J Econom 151:113–128
Hidalgo J, Kreiss JP (2006) Bootstrap specification tests for linear covariance stationary processes. J Econom 133:807–839
Hjellvik V, Tjøstheim D (1995) Nonparametric tests of linearity for time series. Biometrika 82:351–368
Hjellvik V, Tjøstheim D (1996) Nonparametric statistics for testing of linearity and serial independence. J Nonparametr Stat 6:221–251
Hjellvik V, Yao Q, Tjøstheim D (1998) Linearity testing using local polynomial approximation. J Stat Plan Inference 68:295–321
Hjort NL, McKeague IW, Van Keilegom I (2009) Extending the scope of empirical likelihood. Ann Stat 37:1079–1111
Hong Y, Li H (2005) Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Rev Financ Stud 18:37–84
Horowitz J, Spokoiny V (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69:599–631
Horowitz J, Spokoiny V (2002) An adaptive, rate-optimal test of linearity for median regression models. J Am Stat Assoc 97:822–835
Hsiao C, Li Q, Racine JS (2007) A consistent model specification test with mixed discrete and continuous data. J Econom 140:802–826
Huang L, Chen J (2008) Analysis of variance, coefficient of determination and f-test for local polynomial regression. Ann Stat 36:2085–2109
Huang L, Davidson P (2010) Analysis of variance and f-tests for partial linear models with applications to environmental health data. J Am Stat Assoc 105:991–1004
Huskova M, Meintanis S (2007) Omnibus tests for the error distribution in linear regression models. Statistics 41:363–376
Huskova M, Meintanis S (2009) Goodness-of-fit tests for parametric regression models based on empirical characteristic functions. Kybernetika 45:960–971
Huskova M, Meintanis S (2010) Test for the error distribution in nonparametric possibly heterocedastic regression models. Test 19:92–112
Iglesias-Pérez MC, González-Manteiga W (1999) Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. J Nonparametr Stat 10:213–244
Ingster YI (1982) Minimax nonparametric detection of signals in white Gaussian noise. Probl Inf Transm 18:130–140
Ingster YI (1993a) Asymptotically minimax hypothesis testing for nonparametric alternatives, I. Math Methods Stat 2:85–114
Ingster YI (1993b) Asymptotically minimax hypothesis testing for nonparametric alternatives, II. Math Methods Stat 2:171–189
Ingster YI (1993c) Asymptotically minimax hypothesis testing for nonparametric alternatives, III. Math Methods Stat 2:249–268
Jiménez-Gamero MD, García JM, Pino-Mejías R (2005) Testing goodness of fit for the distribution of errors in multivariate linear models. J Multivar Anal 95:301–322
Khmadladze EV, Koul HL (2004) Martingale transforms goodness-of-fit tests in regression models. Ann Stat 37:995–1034
Khmadladze EV, Koul HL (2009) Goodness of fit problem for errors in nonparametric regression distribution free approach. Ann Stat 37:3165–3185
King E, Hart J, Wehrly TE (1991) Testing the equality of two regression curves using linear smoothers. Stat Probab Lett 12:239–247
Kitamura Y, Tripathi G, Ahn H (2004) Empirical likelihood-based inference in conditional moment restriction models. Econometrica 72:1667–1714
Koenker P, Basset G (1978) Regression quantiles. Econometrica 46:33–50

Koul H, Susarla V, Van Ryzin J (1981) Regresssion analysis with randomly right–censored data. Ann Stat 9:1276–1288
Koul HL, Ni P (2004) Minimum distance regression model checking. J Stat Plan Inference 119:109–141
Koul HL, Sakhanenko L (2005) Goodness of fit testing in regression. A finite sample comparison of bootstrap methodology and Khamaladze transformation. Stat Probab Lett 74:290–302
Koul HL, Song W (2008) Regression model checking with Berkson measurement errors. J Stat Plan Inference 138:1615–1628
Koul HL, Song W (2009) Minimum distance regression model checking with Berkson measurement errors. Ann Stat 37:132–156
Koul HL, Song W (2010) Model checking in partial linear regression models with Berkson measurement errors. Stat Sin 20:1551–1579
Koul HL, Stute W (1998) Lack of fit tests in regression with non-random design. Appl Stat Sci 3:53–69
Koul HL, Stute W (1999) Nonparametric model checks for time series. Ann Stat 27:204–236
Koul HL, Stute W, Li F (2005) Model diagnosis for setar time series. Stat Sin 15:795–817
Kozek AS (1991) A nonparametric test of fit of a parametric model. J Multivar Anal 37:66–75
Kreiss JP, Neumann MH, Yao Q (2008) Bootstrap tests for simple structures in nonparametric time series regression. Stat Interface 1:367–380
Kristensen D (2011) Semi-nonparametric estimation and misspecification testing of diffusion models. J Econom 164:382–403
Kulasekera KB (1995) Comparison of regression curves using quasi-residuals. J Am Stat Assoc 90:1085–1093
Kulasekera KB, Wang J (1997) Smoothing parameter selection for power optimality in testing of regression curves. J Am Stat Assoc 92:500–511
Kulasekera KB, Wang J (1998) Bandwidth selection for power optimality in a test of equality of regression curves. Stat Probab Lett 37:287–293
Kutoyants YA (2010) On the goodness-of-fit testing for ergodic diffusion processes. J Nonparametr Stat 22:529–543
Lavergne P (2001) An equality test across nonparametric regressions. Studies in estimation and testing. J Econom 103:307–344
Lavergne P, Patilea V (2008) Breaking the curse of dimensionality in nonparametric testing. J Econom 143:103–122
Lee S (2006) The Bickel–Rosenblatt test for diffusion processes. Stat Probab Lett 76:1494–1502
Lee S, Na S (2002) On the Bickel-Rosenblatt test for first-order autoregressive models. Stat Probab Lett 56:23–25
Lee S, Wee IS (2008) Residual empirical process for difussion processes. J Korean Math Soc 45:683–693
Li CS (2005) Using local linear kernel smoothers to test the lack of fit of nonlinear regression models. Stat Methodol 2:267–284
Li F (2007) Testing the parametric specification of the diffusion function in a difussion process. Econom Theory 23:221–250
Li F, Tkacz G (2006) A consistent bootstrap test for conditional density functions with time-series data. J Econom 133:863–886
Li Q, Wang S (1998) A simple consistent bootstrap test for a parametric regression functional form. J Econom 87:145–165
Li X (2012) Lack of fit testing of a regression model with response missing at random. J Stat Plan Inference 142:155–170
Liang HY, Jing BY (2007) The LIL for the Bickel-Rosenblatt test statistic. J Stat Plan Inference 137:1829–1837
Liang HY, Liu X, Li R, Tsai C (2010) Estimation and testing for partially linear single index model. Ann Stat 38:3811–3836
Liero H (2003) Testing homocedasticity in nonparametric regression. J Nonparametr Stat 15:31–51
Liero H, Läuter H, Konakov V (1998) Nonparametric versus parametric goodness of fit. Statistics 31:115–149
Lin W, Kulasekera KB (2010) Testing the equality of linear single-index models. J Multivar Anal 101:1156–1167
Liu R (1988) Bootstrap procedures under some non-i.i.d. models. Ann Stat 16:1696–1708
Liu Z, Stengos T, Li Q (2000) Nonparametric model check based on local polynomial fitting. Stat Probab Lett 48:327–334
Lombardía MJ, Sperlich S (2008) Semiparametric inference in generalized mixed effect models. J R Stat Soc B 70:913–930

Lopez O, Patilea V (2009) Nonparametric lack-of-fit tests for parametric mean–regression models with censored data. J Multivar Anal 100:210–230
Ma Y, Hart JD, Janicki R, Carroll RJ (2011) Local and omnibus goodness-of-fit tests in classical measurement error models. J R Stat Soc B 73:81–98
Maity A, Carroll RJ, Mammen E, Chatterjee W (2009) Testing in semiparametric models with interaction, with applications to gene−environment interactions. J R Stat Soc B 71:75–96
Masuda H, Negri I, Nishiyama Y (2010) Goodness-of-fit test for ergodic diffusions by discrete–time observation: an innovation martingale approach. J Nonparametr Stat 23:237–254
McKeague IW, Zhang MJ (1994) Identification of nonlinear time series from first order cummulative characteristics. Ann Stat 22:495–514
Meintanis SG, Portnoy S (2011) Specification tests in mixed effects models. J Stat Plan Inference 141:2545–2555
Miles D, Mora J (2002) On the performance of nonparametric specification test in regression models. Comput Stat Data Anal 42:477–490
Monsalve-Cobis A, González-Manteiga W, Febrero-Bande M (2011) Goodness-of-fit tests for interest rate models: an approach based on empirical processes. Comput Stat Data Anal 55:3073–3092
Mora J (2005) Comparing distribution functions of errors in linear models: a nonparametric approach. Stat Probab Lett 73:425–432
Mora J, Pérez-Alonso A (2009) Specification tests for the distribution of errors in nonparametric regression: a martingale approach. J Nonparametr Stat 21:441–452
Müller HG (1992) Goodness-of-fit diagnostics for regression models. Scand J Stat 19:157–172
Müller M (2001) Estimation and testing in generalized partial linear models—a comparative study. Stat Comput 11:299–399
Müller UU, Shick A, Welfemeyer W (2009) Estimating the error distribution function in nonparametric regression with multivariate covariates. Stat Probab Lett 79:957–964
Munk A, Dette H (1998) Nonparametric comparison of several regression functions: exact and asymptotic theory. Ann Stat 6:2339–2368
Munk A, Neumeyer N, Scholz A (2007) Non-parametric analysis of covariance. The case of inhomogeneous and heteroscedastic noise. Scand J Stat 34:511–534
Nadaraya EA (1964) On estimating regression. Theory Probab Appl 10:186–196
Negri I, Nishiyama Y (2009) Goodness-of-fit test for ergodic diffusion process. Ann Inst Math Stat 61:167–198
Negri I, Nishiyama Y (2010) Goodness-of-fit test for ergodic diffusion process by tick time sample scheme. Statistical inference for stochastic processes, vol 13, pp 81–95
Neumann MH, Paparoditis E (2000) On bootstrapping l_2-statistics in density testing. Stat Probab Lett 50:137–147
Neumann MH, Paparoditis E (2008a) Simultaneous confidence bands in spectral density estimation. Biometrika 95:381–397
Neumann MH, Paparoditis E (2008b) Goodness-of-fit tests for Markovian time series models: central limit theory and bootstrap approximations. Bernoulli 14:14–46
Neumeyer N (2009) Smooth residual bootstrap for empirical processes of nonparametric regression residuals. Scand J Stat 36:204–228
Neumeyer N, Dette H (2003) Nonparametric comparison of regression curves: an empirical process approach. Ann Stat 31:880–920
Neumeyer N, Sperlich S (2006) Comparison of separable components in different samples. Scand J Stat 33:444–501
Neumeyer N, Van Keilegom I (2010) Estimating the error distribution in nonparametric multiple regression with applications to model testing. J Multivar Anal 101:1067–1078
Ojeda JL, Van Keilegom I (2009) Goodness-of-fit tests for parametric regression with selection biased data. J Stat Plan Inference 139:2836–2850
Ojeda JL, Cristóbal JA, Alcalá JT (2008) A bootstrap approach to model checking for linear models under length-biased data. Ann Inst Math Stat 60:519–543
Ojeda JL, González-Manteiga W, Cristóbal JA (2011) A bootstrap based model checking for selection-biased data. Technical report, University of Santiago de Compostela
Owen A (2001) Empirical likelihood. Chapman & Hall, New York
Pan Z, Lin DY (2005) Goodness-of-fit methods for generalized linear mixed models. Biometrics 61:1000–1009
Paparoditis E (2000) Spectral density based goodness-of-fit tests for time series models. Scand J Stat 27:143–176

Paparoditis E (2009) Testing temporal constancy of the spectral structure of a time series. Bernoulli 15:1190–1221
Paparoditis E (2010) Validating stationary assumptions in time series analysis by rolling local periodograms. J Am Stat Assoc 105:839–851
Pardo-Fernández JC (2007) Comparison of error distributions in nonparametric regression. Stat Probab Lett 77:350–356
Pardo-Fernández JC, Van Keilegom I (2006) Comparison of regression curves with censored responses. Scand J Stat 33:409–434
Pardo-Fernández JC, Van Keilegom I, González-Manteiga W (2007a) Goodness-of-fit tests for parametric models in censored regression. Can J Stat 35:249–264
Pardo-Fernández JC, Van Keilegom I, González-Manteiga W (2007b) Testing for the equality of k regression curves. Stat Sin 17:1115–1137
Park C, Kang K (2008) Sizer analysis for the comparison of regression curves. Comput Stat Data Anal 52:3954–3970
Parzen E (1962) On estimation of a probability density function and mode. Ann Math Stat 33:1065–1076
Priestley MB, Chao MT (1972) Non-parametric function fitting. J R Stat Soc B 34:385–392
Ramil-Novo LA, González-Manteiga W (1998) χ^2 goodness-of-fit tests for polynomial regression. Commun Stat, Simul Comput 27:229–258
Ramil-Novo LA, González-Manteiga W (2000) F tests and regression analysis of variance based on smoothing splines estimators. Stat Sin 10:819–837
Ramsay J, Silverman B (2005) Functional data analysis. Springer, New York
Raz J (1990) Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach. J Am Stat Assoc 85:132–138
Robinson PM (1988) Root-N-consistent semiparametric regression. Econometrica 56:931–944
Roca-Pardiñas J, Sperlich S (2007) Testing the link when the index is semiparametric. a comparative study. Comput Stat Data Anal 51:6365–6581
Roca-Pardiñas J, Cadarso-Suárez C, González-Manteiga W (2005) Testing for interactions in generalized additive models: application to SO_2 pollution data. Stat Comput 15:289–299
Rodríguez-Campos C, González-Manteiga W, Cao R (1998) Testing the hypothesis of a generalized linear regression model using nonparametric regression estimation. J Stat Plan Inference 67:99–122
Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. Ann Math Stat 27:832–837
Rosenblatt M (1991) Stochastic curve estimation. Institute of Mathematical Statistics, Hayward
Samarakoon N, Song W (2010) Minimum distance conditional variance function checking in heteroscedastic regression models. J Multivar Anal 102:579–600
Samarov A (1993) Exploring regression structure using nonparametric functional estimation. J Am Stat Assoc 88:836–847
Sánchez BN, Houseman EA, Ryan LM (2009) Residual-based diagnostics for structural equation models. Biometrics 65:104–115
Sánchez-Sellero C, González-Manteiga W, Van Keilegom I (2005) Uniform representation of product–limit integrals with applications. Scand J Stat 32:563–581
Seber GAF (1977) Linear regression analysis. Wiley, New York
Seber GAF, Wild CG (1989) Nonlinear regression. Wiley, New York
Sergides M, Paparoditis E (2007) Bootstrapping the local periodogram of locally stationary processes. J Time Ser Anal 29:264–279
Sergides M, Paparoditis E (2009) Frequency domain tests of semiparametric hypotheses for locally stationary processes. Scand J Stat 36:800–821
Song K (2010) Testing semiparametric conditional moment restrictions using conditional martingale transforms. J Econom 154:74–84
Song W (2008) Model checking in errors-in-variables regression. J Multivar Anal 99:2406–2443
Song Z (2011) A martingale approach for testing diffusion models based on infinitesimal operator. J Econom 162:189–212
Speckman P (1988) Kernel smoothing in partial linear models. J R Stat Soc B 50:413–436
Sperlich S, Lombardía MJ (2010) Local polynomical inference for small area statistics: estimation, validation and prediction. J Nonparametr Stat 22:633–648
Sperlich S, Linton O, Härdle W (1999) Integration and backfitting methods in additive models: finite sample properties and comparison. Test 8:419–458
Sperlich S, Tjøstheim D, Yang L (2002) Nonparametric estimation and testing of interaction in additive models. Econom Theory 18:197–251

Spokoiny V (1996) Adaptive hypothesis testing using wavelets. Ann Stat 24:2477–2498
Spokoiny V (2001) Data driven testing the fit of linear models. Math Methods Stat 10:465–497
Srihera R, Stute W (2010) Nonparametric comparison of regression functions. J Multivar Anal 101:2039–2059
Staniswalis JG, Severini TA (1991) Diagnostics for assessing regression models. J Am Stat Assoc 86:684–692
Stute W (1993) Consistent estimation under random censorship when covariables are present. J Multivar Anal 45:89–103
Stute W (1996) Distributional convergence under random censorship when covariables are present. Scand J Stat 23:461–471
Stute W (1997) Nonparametric model checks for regression. Ann Stat 25:613–641
Stute W (1999) Nonlinear censored regression. Stat Sin 25:613–641
Stute W, González-Manteiga W (1996) Nn goodness-of-fit tests for linear models. J Stat Plan Inference 53:75–92
Stute W, Zhu L (2005a) Nonparametric checks for single–index models. Ann Stat 33:1048–1083
Stute W, Zhu L (2005b) Model checks for generalized linear models. Scand J Stat 29:535–545
Stute W, González-Manteiga W, Presedo-Quindimil MA (1993) Boostrap based goodness-of-fit tests. Metrika 40:243–256
Stute W, González-Manteiga W, Presedo-Quindimil M (1998a) Bootstrap approximations in model checks for regression. J Am Stat Assoc 93:141–149
Stute W, Thies S, Zhu LX (1998b) Model checks for regression: an innovation process approach. Ann Stat 26:1916–1934
Stute W, González-Manteiga W, Sánchez-Sellero C (2000) Nonparametric model checks in censored regression. Commun Stat, Theory Methods 29:1611–1629
Stute W, Presendo-Quindimil M, González-Manteiga W, Koul HL (2006) Model checks for higher order time series. Stat Probab Lett 76:1385–1396
Stute W, Xu WL, Zhu X (2008) Model diagnosis for parametric regression in high–dimensional spaces. Biometrika 95:451–467
Su JQ, Wei LJ (1991) A lack of fit test for the mean function in a generalized linear model. J Am Stat Assoc 86:420–426
Sun Y (2006) A consistent nonparametric equality test of conditional quantile functions. Econom Theory 22:614–632
Sun Z, Wang Q (2009) Checking the adequacy of a general linear model with responses missing at random. J Stat Plan Inference 139:3588–3604
Sun Z, Wang Q, Dai P (2009) Model checking for partially linear models with missing responses at random. J Multivar Anal 100:636–651
Tenreiro C (2007) On the asymptotic behaviour of location-scale invariant Bickel-Rosenblatt tests. J Stat Plan Inference 137:103–116. Erratum, no 139:2115
Tenreiro C (2009) On the choice of the smoothing parameter for the bhep goodness-of-fit test. Comput Stat Data Anal 53:1038–1053
Teodorescu B, Van Keilegom I (2010) A goodness-of-fit test for generalized conditional linear models under left truncation and right censoring. J Nonparametr Stat 22:547–566
Teodorescu B, Van Keilegom I, Cao R (2010) Generalized conditional linear models under left truncation and right censoring. Ann Inst Math Stat 62:465–485
Tripathi G, Kitamura Y (2003) Testing conditional moment restrictions. Ann Stat 31:2059–2095
Van Keilegom I, González-Manteiga W, Sánchez-Sellero C (2008a) Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. Test 17:401–415
Van Keilegom I, Sánchez-Sellero C, González-Manteiga W (2008b) Empirical likelihood based testing for regression. Electron J Stat 2:581–604
Vilar-Fernández JM, González-Manteiga W (1996) Bootstrap test of goodness of fit to a linear model when errors are correlated. Commun Stat, Theory Methods 25:2925–2953
Vilar-Fernández JM, González-Manteiga W (2000) Resampling for checking linear regression models via non-parametric regression estimation. Comput Stat Data Anal 35:211–231
Vilar-Fernández JM, González-Manteiga W (2004) Nonparametric comparison of curves with dependent errors. Statistics 38:81–99
Vilar-Fernández JM, Vilar-Fernández JA, González-Manteiga W (2007) Bootstrap tests for nonparametric comparison of regression curves with dependent errors. Test 16:123–144
Wang L (2008) Nonparametric test for checking lack of fit of the quantile regression model under random censoring. Can J Stat 36:321–336

Watson GS (1964) Smooth regression analysis. Sankhyā Ser A 26:359–372
Wong H, Liu F, Chen M, Cheung IW (2009) Empirical likelihood based diagnostics for heteroscedasticity in partial linear models. Comput Stat Data Anal 53:3466–3477
Wooldridge JM (1992) A test for functional form against nonparametric alternatives. Econom Theory 4:935–955
Wu CFJ (1986) Jackknife, bootstrap and other resampling methods in regression analysis. Ann Stat 14:1261–1350
Xia Y (2009) Model checking in regression via dimension reduction. Biometrica 96:133–148
Xia Y, Li WK, Tong H, Zhang D (2004) A goodness-of-fit test for single-index models. Stat Sin 14:1–39
You J, Chen G (2005) Testing heteroscedasticity in partially linear regression models. Stat Probab Lett 73:61–70
Young S, Bowman AW (1995) Non–parametric analysis of covariance. Biometrics 51:920–931
Zhang C (2003) Calibrating the degrees of freedom for automatic data smoothing and effective curve checking. J Am Stat Assoc 98:609–629
Zhang C (2004) Assessing the equivalence of nonparametric regression tests based on spline and local polynomial smoothers. J Stat Plan Inference 126:73–95
Zhang C, Dette H (2004) A power comparison between nonparametric regression tests. Stat Probab Lett 66:289–301
Zheng JX (1996) A consistent test of functional form via nonparametric estimation techniques. J Econom 75:263–289
Zheng JX (1998) A consistent nonparametric test of parametric regression models under conditional quantile restrictions. Econom Theory 14:123–138
Zhou Z (2010) Nonparametric inference of quantile curves for nonstationary time series. Ann Stat 38:2187–2217
Zhu H, Ibrahim JG, Shi X (2009) Diagnostic measures for generalized linear models with missing covariates. Scand J Stat 36:686–712
Zhu L (2005) Nonparametric Monte Carlo tests and their applications. Lecture notes in statistics, vol 182. Springer, Berlin
Zhu L, Ng KW (2003) Checking the adequacy of a partial linear model. Stat Sin 13:763–781
Zhu L, Fujikoshi Y, Naito K (2001) Heteroscedasticity checks for regression models. Sci China 44:1236–1252