Validation tests for semiparametric models

Simos.G. Meintanis[1,2], Jochen Einbeck[3]

[1] *Department of Economics, National and Kapodistrian University of Athens*
8 Pesmazoglou Street, 105 59 Athens, Greece,
simosmei@econ.uoa.gr

[2] *Unit for Business Mathematics & Informatics, North–West University, Potchefstroom, South Africa*[1]

[3] *Department of Mathematical Sciences, University of Durham,*
South Rd, Durham DH1 3LE, United Kingdom,
jochen.einbeck@durham.ac.uk,
Tel +44 191 3343125, Fax +44 191 3343051 **(corresponding author)**

Abstract. Tests are proposed for validation of the hypothesis that a partial linear regression model adequately describes the structure of a given data set. The test statistics are formulated following the approach of Fourier–type conditional expectations first suggested by Bierens (1982). The proposed procedures are computationally convenient, and under fairly mild conditions lead to consistent tests. Corresponding bootstrap versions are compared with alternative procedures for a wide selection of different estimators of the underlying partial linear model.

Keywords. Semi–linear model; Goodness–of–fit test; Empirical characteristic function, Bootstrap test.

MSC2010 codes. G2G08, G2G09, G2G10.

[1] On sabbatical leave from the University of Athens

1 Introduction

Let $\boldsymbol{x} = (x_1, \ldots, x_p)'$ and $\boldsymbol{z} = (z_1, \ldots, z_q)'$ be covariate vectors, and consider the partial linear model

$$(1.1) \quad y = \boldsymbol{x}'\boldsymbol{\beta} + g(\boldsymbol{z}) + \varepsilon,$$

where both $\boldsymbol{\beta}$ and $g(\cdot)$ are unknown. The (potentially heteroscedastic) error $\varepsilon = \varepsilon(\boldsymbol{x}, \boldsymbol{z})$ is assumed to follow an unknown distribution $F(\cdot|\boldsymbol{x}, \boldsymbol{z})$, with $\mathbb{E}(\varepsilon|\boldsymbol{x}, \boldsymbol{z}) = 0$ and $\mathbb{E}(\varepsilon^2|\boldsymbol{x}, \boldsymbol{z}) < \infty$. Important subcases nested in model (1.1) are: (i) the linear model for $q = 0$ and (ii) the nonparametric regression model for $p = 0$.

On the basis of the data $\{y_j, \boldsymbol{x}_j, \boldsymbol{z}_j\} \in \mathbb{R}^{1+p+q}$, $j = 1, \ldots, n$, we wish to test the null hypothesis

$$(1.2) \quad \mathcal{H}_0 : \mathbb{E}(y|\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{x}'\boldsymbol{\beta} + g(\boldsymbol{z}),$$

for some $\boldsymbol{\beta} \in \mathbb{R}^p$, and some function $g : \mathbb{R}^q \mapsto \mathbb{R}$, against general alternatives. In this general context there exist very few works on the specification problem of testing $\mathcal{H}_0$. For instance, Yatchew (1992) proposed a test based on the sum of squared residuals from a least squares fit, which also allows for heteroscedasticity but requires sample splitting. Avoiding the latter, Fan & Li (1996) suggested a test of the null hypothesis in (1.2) based on the limiting normal distribution of degenerate U-statistics, while Zhu & Ng (2003) considered a resampling approach, where the test statistic is constructed through a residual-marked cusum process.

Recently, an alternative route to the construction of hypothesis tests, employing the empirical characteristic function of the residuals, has been successfully pursued in related settings. These include tests for the shape of the error distribution in linear (Hušková & Meintanis, 2007), nonparametric (Hušková & Meintanis, 2010), and semi-parametric models (Meintanis & Einbeck, 2012). The abovementioned procedures are conceptually simple, easy to implement, and have been found to outperform, or at least be highly competitive to more traditional methods which employ, for instance, the empirical distribution function of the residuals (rather than their characteristic function).

In order to make use of this Fourier–type approach for testing (1.2), we formulate our procedure by using the characterization of Bierens (1982): for real y and given a function $w(\boldsymbol{v})$ of a k–vector $\boldsymbol{v}$, the equation $\mathbb{E}(y|\boldsymbol{v}) = w(\boldsymbol{v})$ holds if and only if $\mathbb{E}[\{y - w(\boldsymbol{v})\}e^{i\boldsymbol{t}'\boldsymbol{v}}] = 0$, for all $\boldsymbol{t} \in \mathbb{R}^k$. By adapting this result to the current set–up, with $\boldsymbol{v} = (\boldsymbol{x}', \boldsymbol{z}')'$ and $w(\boldsymbol{v}) = \boldsymbol{x}'\boldsymbol{\beta} + g(\boldsymbol{z})$, we suggest to test the null hypothesis $\mathcal{H}_0$ by checking whether

$$\text{(1.3)} \quad \mathbb{E}[\{y - \boldsymbol{x}'\boldsymbol{\beta} - g(\boldsymbol{z})\}e^{i\boldsymbol{t}'\boldsymbol{v}}] = 0, \ \forall \ \boldsymbol{t} \in \mathbb{R}^{p+q}.$$

Given suitable estimators $(\widehat{\boldsymbol{\beta}_n}, \widehat{g_n}(\cdot))$ of $(\boldsymbol{\beta}, g(\cdot))$, it is straightforward to estimate the conditional expectation figuring in the left–hand side of (1.3) by

$$E_n(\boldsymbol{t}) = \frac{1}{n}\sum_{j=1}^{n} \widehat{\varepsilon}_j e^{i\boldsymbol{t}'\boldsymbol{v}_j},$$

where

$$\text{(1.4)} \quad \widehat{\varepsilon}_j = y_j - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}_n} - \widehat{g_n}(\boldsymbol{z}_j), \ j = 1, ..., n.$$

Consequently an omnibus procedure for specification testing is to reject the null hypothesis $\mathcal{H}_0$ for large values of the test statistic

$$\text{(1.5)} \quad T_{n,W} = \int_{\mathbb{R}^{p+q}} |E_n(\boldsymbol{t})|^2 \, W(\boldsymbol{t}) d\boldsymbol{t},$$

where $W(\cdot)$ is an appropriate weight function which we discuss in detail in the following section.

The rest of the paper is organized as follows. In Section 2 we discuss the computation of the test statistics, the consistency of the test, and an approximation of the test statistic. Bootstrap versions of the tests are introduced in Section 3, while we consider the important task of estimation in Section 4. An extensive Monte Carlo study is provided in Section 5. Section 6 contributes a discussion including a real data example. Finally, in the Appendix the consistency of the proposed test under general alternatives is proved.

To fix notation used throughout the manuscript, denote $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$, $\boldsymbol{y} = (y_1, \ldots, y_n)'$, and $\boldsymbol{g} = (g(\boldsymbol{z}_1), \ldots, g(\boldsymbol{z}_n))'$. Also, in what follows bold small symbols generally denote vectors, and bold capital symbols denote matrices.

2 Computations and asymptotics

2.1 Computation of test statistic

In order to compute the test statistic $T_{n,W}$ in (1.5), first notice that

$$|E_n(\boldsymbol{t})|^2 = \frac{1}{n^2} \sum_{j,k=1}^{n} \widehat{\varepsilon}_j \widehat{\varepsilon}_k \cos\{\boldsymbol{t}'(\boldsymbol{v}_j - \boldsymbol{v}_k)\}.$$

Hence, one may write the aforementioned test statistic as

$$T_{n,W} = \frac{1}{n^2} \sum_{j,k=1}^{n} \widehat{\varepsilon}_j \widehat{\varepsilon}_k I_W(\boldsymbol{v}_j - \boldsymbol{v}_k), \tag{2.1}$$

where

$$I_W(\boldsymbol{u}) = \int_{\mathbb{R}^d} \cos(\boldsymbol{t}'\boldsymbol{u}) W(\boldsymbol{t}) d\boldsymbol{t}$$

with $d = p + q$ denoting the total number of covariates. Thus the computation of $T_{n,W}$ boils down to the computation of $I_W(\cdot)$, and consequently tests which are free of numerical integration correspond to suitable weight functions $W(\cdot)$ which render the integral $I_W(\cdot)$ in a convenient closed–form expression.

In order to investigate the spectrum of possible weight functions, suppose that $W(\cdot)$ may be written as a product

$$W(\boldsymbol{t}) = \prod_{\ell=1}^{d} w(t_\ell), \tag{2.2}$$

where $\boldsymbol{t} = (t_1, ..., t_d)'$ and the univariate function w satisfies $w(t) = w(-t),\ t \in \mathbb{R}$. Then by straightforward algebra we have $I_W(\boldsymbol{u}) = \prod_{\ell=1}^{d} I_w(u_\ell)$ where $\boldsymbol{u} = (u_1, ..., u_d)'$ and $I_w(z) = \int_{-\infty}^{\infty} \cos(tz) w(t) dt,\ z \in \mathbb{R}$. Hence the test statistic in (2.1) becomes dependent solely on w and takes the form

$$T_{n,w} = \frac{1}{n^2} \sum_{j,k=1}^{n} \widehat{\varepsilon}_j \widehat{\varepsilon}_k \left\{ \prod_{\ell=1}^{d} I_w(v_{j\ell} - v_{k\ell}) \right\}, \tag{2.3}$$

where $v_{j\ell}, \ell = 1, ..., d$, denotes the ℓ^{th} coordinate of $\boldsymbol{v}_j,\ j = 1, ..., n$. Expression (2.3) provides considerable flexibility in choosing the weight function. In principle, the product decomposition in (2.2) allows for any even function $w(t)$ which renders the integral $I_w(u)$ finite to stand as a candidate. Standard choices are exponential functions of the type $w(t) = e^{-a|t|^m}$, $a > 0$, $m = 1, 2$, uniform–type weight functions $w(t) = 1, |t| \le a$, and $w(t) = 0, |t| > a$, or trigonometric oscillations such as $w(t) = \sin(at)/(at)$, $t \in (-\pi/a, \pi/a)$. Some particularly important weight functions and associated test statistics are discussed below.

Normal weight functions. Our first choice for the weight function comes from the d–variate normal distribution with i.i.d. components each with mean zero and variance a. The characteristic function of this distribution yields

$$\int_{\mathbb{R}^d} \cos(\boldsymbol{t}'\boldsymbol{u}) e^{-a||\boldsymbol{t}||^2} d\boldsymbol{t} = \left(\frac{\pi}{a}\right)^{d/2} e^{-||\boldsymbol{u}||^2/4a}, \ a > 0.$$

From this integral and by letting $W(\boldsymbol{t}) = e^{-a||\boldsymbol{t}||^2}$ in (2.1) we have

$$(2.4) \quad T_{n,W} = \left(\frac{\pi}{a}\right)^{d/2} \frac{1}{n^2} \sum_{j,k=1}^{n} \widehat{\varepsilon}_j \widehat{\varepsilon}_k e^{-||\boldsymbol{v}_j - \boldsymbol{v}_k||^2/4a} := T_a$$

We observe that T_a, apart from the irrelevant constant $(\pi/a)^{d/2}$, is the weighted mean of the cross product $\widehat{\varepsilon}_j \widehat{\varepsilon}_k$ of the residuals. In turn, the weight that each term $\widehat{\varepsilon}_j \widehat{\varepsilon}_k$ receives in the sum is determined by the squared distance between the corresponding regressor vectors $\boldsymbol{v}_j$ and $\boldsymbol{v}_k$, this distance being exponentially transformed to $e^{-||\boldsymbol{v}_j - \boldsymbol{v}_k||^2/4a}$.

Notice that the weight function $e^{-a||\boldsymbol{t}||^2}$ is a particular case of the product decomposition in (2.2). Specifically we could alternatively arrive at (2.4) by using (2.2) with $w(t) = e^{-at^2}$, (2.3) and

$$\int_{-\infty}^{\infty} \cos(tz) e^{-at^2} dt = \sqrt{\frac{\pi}{a}} e^{-z^2/4a}.$$

Laplace-like weight functions. Further, we consider the univariate weight function $w(t) = e^{-a|t|}$ in (2.2). Using that

$$\int_{-\infty}^{\infty} \cos(tz) e^{-a|t|} \, dt = \frac{2a}{a^2 + z^2}$$

one obtains via (2.3),

$$(2.5) \quad T_{n,w} = \frac{(2a)^d}{n^2} \sum_{j,k=1}^{n} \hat{\varepsilon}_j \hat{\varepsilon}_k \left\{ \prod_{\ell=1}^{d} \frac{1}{a^2 + (v_{j\ell} - v_{k\ell})^2} \right\} := \mathcal{T}_a.$$

Note that this result is *not* obtained when using the multivariate weight $e^{-a||\boldsymbol{t}||}$ in (2.1), since $\prod_{\ell=1}^{d} e^{-a|t_\ell|} = e^{-a||\boldsymbol{t}||_1}$, where $||\cdot||_1$ is the L_1 "city-block" (but not the Euclidean!) norm.

2.2 Behavior for $n \longrightarrow \infty$

In this subsection we investigate the stochastic behavior of the test statistic $T_{n,W}$, as $n \to \infty$. Under fairly mild conditions, we have the following result.

Theorem 1 *Under the assumptions (A1) to (A4),*

$$\text{(2.6)} \qquad \lim_{n \longrightarrow \infty} T_{n,W} = \int_{\mathbb{R}^{p+q}} |E(\boldsymbol{t})|^2 \, W(\boldsymbol{t}) d\boldsymbol{t} := T_W,$$

almost surely, where $E(\boldsymbol{t}) = \mathbb{E}[\varepsilon e^{i\boldsymbol{t}'\boldsymbol{v}}]$.

The assumptions (A1) to (A4) as well as the proof are provided in the Appendix. In view of (1.3), the limit statistic T_W in (2.6) is equal to zero under the null hypothesis $\mathcal{H}_0$. On the other hand, suppose $\mathcal{H}_0$ is not true. Then there exists some $\boldsymbol{t}_0 \in \mathbb{R}^{p+q}$ such that $\mathbb{E}[\{y - \boldsymbol{x}'\boldsymbol{\beta} - g(\boldsymbol{z})\}e^{i\boldsymbol{t}_0'\boldsymbol{v}}] \neq 0$, and due to continuity, the last relation holds not just at $\boldsymbol{t}_0$, but in a δ–neighborhood $\mathcal{N}_0(\delta) := \{\boldsymbol{t} \in \mathbb{R}^{p+q} : \|\boldsymbol{t} - \boldsymbol{t}_0\| < \delta\}$. Consequently we have

$$T_W = \int_{\mathbb{R}^{p+q}} |E(\boldsymbol{t})|^2 \, W(\boldsymbol{t}) d\boldsymbol{t} \geq \int_{\mathcal{N}_0(\delta)} |E(\boldsymbol{t})|^2 \, W(\boldsymbol{t}) d\boldsymbol{t} > 0,$$

which implies the consistency of the test which rejects the null hypothesis $\mathcal{H}_0$ for large values of the test statistic $T_{n,W}$.

2.3 Behavior for $a \longrightarrow \infty$

We now consider the limit value of the statistics T_a and $\mathcal{T}_a$, as $a \to \infty$. Clearly such an approximation is equivalent to replacing of the corresponding weight function by a Dirac type function. To this end expand the exponential function, and notice that after some further algebra we may write

$$\text{(2.7)} \qquad \pi^{d/2} \frac{1}{n^2} \sum_{j,k=1}^{n} \widehat{\varepsilon}_j \widehat{\varepsilon}_k - a^{d/2} T_a = \frac{\pi^{d/2}}{4a} \frac{1}{n^2} \sum_{j,k=1}^{n} \widehat{\varepsilon}_j \widehat{\varepsilon}_k ||\boldsymbol{v}_j - \boldsymbol{v}_k||^2 + \mathrm{o}\left(\frac{1}{a}\right), a \to \infty,$$

for the test statistic in (2.4). From (2.7) it clearly follows that,

$$\text{(2.8)} \qquad \lim_{a \longrightarrow \infty} \left(\frac{a}{\pi}\right)^{d/2} T_a = \frac{1}{n^2} \sum_{j,k=1}^{n} \widehat{\varepsilon}_j \widehat{\varepsilon}_k := T_\infty$$

so that one may consider using the right hand side of this expression as an approximate, limit test statistic. A further scaling in (2.7) yields the limit

$$\lim_{a\to\infty} a\left(\frac{1}{n^2}\sum_{j,k=1}^{n}\widehat{\varepsilon}_j\widehat{\varepsilon}_k-\left(\frac{a}{\pi}\right)^{d/2}T_a\right)=\frac{1}{4}\frac{1}{n^2}\sum_{j,k=1}^{n}\widehat{\varepsilon}_j\widehat{\varepsilon}_k||\boldsymbol{v}_j-\boldsymbol{v}_k||^2. \tag{2.9}$$

Consequently we conclude that the test statistic T_a when scaled and centered properly, attains a limit value as $a \to \infty$, which coincides with the weighted mean of the cross product of the residuals with weights equal to the squared distance between the corresponding covariate–vectors. This quantity is rather of theoretical than practical interest — it does not perform well as a test statistic, since, though resembling (2.4) at first glance, the residuals are now associated with weights which increase, rather than decrease, with distances $||\boldsymbol{v}_j-\boldsymbol{v}_k||^2$. The test statistic (2.8) can be seen as a middle ground between these two extremes, which does not make use of the distances at all.

Analogously, it follows from (2.5) that

$$\lim_{a\to\infty}\left(\frac{a}{2}\right)^{d}\mathcal{T}_a=\frac{1}{n^2}\sum_{j,k=1}^{n}\widehat{\varepsilon}_j\widehat{\varepsilon}_k,$$

which along with (2.8) shows that, in first order approximation, the test statistics based on Gaussian and Laplacian weights are equivalent.

3 Bootstrap procedures

The limit null distribution of the test statistic $T_{n,W}$ is highly non–trivial and depends on unknown quantities; see Bierens (1990). Therefore it is not practical to use this distribution for calculating critical points and actually performing the test. However there exists a suitable resampling procedure which is specifically tailored for the testing problem (1.2) and circumvents this drawback. The consistency of this resampling scheme has been shown by Zhu (2005, §5), Delgado & González–Manteiga (2001) and Härdle et al. (1998), with different testing procedures for the partial linear model hypothesis (1.2). In practical terms the resampling is implemented as follows: Conditionally on the observations $\{y_j,\boldsymbol{x}_j,\boldsymbol{z}_j\}_{j=1}^n$, fit the model according to $\mathcal{H}_0$, yielding residuals $\{\widehat{\varepsilon}_j\}_{j=1}^n$, and calculate the corresponding value, say T, for the test statistic. Then,

(i) Generate independent random variables $\{e_j\}_{j=1}^n$ with mean zero and unit variance, and let $\varepsilon_j^* = e_j \widehat{\varepsilon}_j, \ j = 1, ...n$.

(ii) Compute the bootstrap responses $y_j^* = \boldsymbol{x}_j' \widehat{\boldsymbol{\beta}_n} + \widehat{g}_n(\boldsymbol{z}_j) + \varepsilon_j^*, \ j = 1, ..., n$.

(iii) On the basis of the observations $\{y_j^*, \boldsymbol{x}_j, \boldsymbol{z}_j\}_{j=1}^n$ refit the model in (1.1), compute the residuals and the corresponding value of the test statistic, say T^*.

(iv) Repeat steps (i) to (iii) a number of times, say B, and obtain $\{T_b^*\}_{b=1}^{\mathrm{B}}$.

(v) Calculate the p–value as $\hat{p} = k/(\mathrm{B}+1)$ where $k := \sum_{b=1}^{\mathrm{B}} \mathbf{1}_{\{T_b^* \geq T\}}$ denotes the number of times that $T_b^*, \ b = 1, ..., \mathrm{B}$, was greater than or equal to T.

(vi) Reject the null hypothesis $\mathcal{H}_0$ if $\hat{p} \leq \alpha$, where α denotes the designated level of significance.

Note that this bootstrap method is a version of the wild bootstrap which is employed when heteroscedasticity is present, and may also be found under the name 'external bootstrap'; see for instance Delgado & Fiteni (2002).

In order to assess the sensitivity of the proposed test to the choice of the weight parameter a, we have carried out a small simulation study resembling the design of that one in Meintanis & Einbeck (2012). We consider two different data-generating mechanisms, namely

$$\text{(M0)} \quad y = x + \sin(2\pi z) + 0.5\epsilon$$

and

$$\text{(M1)} \quad y = x^2 + 0.5\epsilon$$

with $\epsilon \sim N(0,1)$, $x \sim N(1/2, 1/2)$, and $z \sim U[0,1]$. As the null hypothesis we use

$$\mathcal{H}_0 : y = \beta x + g(z) + \varepsilon,$$

i.e. for model (M0) $\mathcal{H}_0$ is true and for model (M1) $\mathcal{H}_0$ is false. The sample size of the simulated data sets is $n = 100$, of which we produce $B = 200$ bootstrap replicates each, with Gaussian e_j in step (i). We use the test statistic T_a in (2.4), experimenting with a wide range of values for the weight parameter a; specifically these are: $a =$

$0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100, 10^3, 10^4,\ 10^5, 10^6,\ 10^7$, and 10^8. Further we consider the asymptotic version T_∞ in (2.8). The model is fitted using R function `gam` in R package **gam**, using smoothing splines and the default bandwidth (corresponding to 5 degrees of freedom including the intercept) at all occasions. The null hypothesis is rejected if the bootstrapped $p-$value falls below the target significance level, which we take to be either $\alpha = 0.05$ or $\alpha = 0.10$. The entire bootstrap routine is repeated 2000 times.

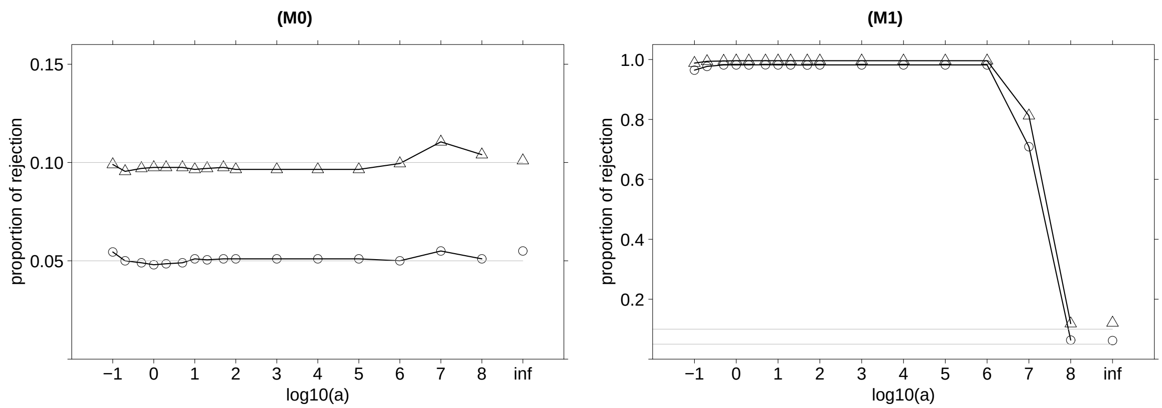
The resulting proportions of rejection under (M0) and (M1) are depicted in Figure 1. We observe that (i) when $\mathcal{H}_0$ is true, the tests attain throughout a rejection rate which is very close to the target significance level; (ii) the performance of the tests is largely independent of the weight parameter in a wide range of values from $a = 0.1$ to about $a = 100000$; (iii) however, a cannot be arbitrarily increased, as from ca. $a \approx 10^7$ on the test power falls dramatically, rendering the asymptotic version infeasible in practice. (iv) Otherwise, the test powers under (M1) achieve excellent values which are very close to 1.

We also experimented with a range of possible values for the number of bootstrap replicates, B. Increasing the number beyond 200 led to similar results to those obtained for $B = 200$, at large computational cost. Decreasing the number B led naturally to an increased variability of the observed powers, though we still obtained sensible simulation results for values as small as $B = 50$. Overall, the setting $B = 200$ appears to be a good compromise between computational efficiency and test accuracy.

4 Estimation

It remains the important task of how to find suitable estimators $\widehat{\boldsymbol{\beta}_n}$ and $\widehat{g_n}(\cdot)$ for the partial linear model (1.1). Concerning the nonparametric term, it is sufficient for our purposes to estimate g only at the location of the predictors $\boldsymbol{z}_j, j = 1, \ldots, n$. Hence, we seek estimators of type $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{g}}) \in \mathbb{R}^{p+n}$, with $\hat{\boldsymbol{g}} = (\widehat{g_n}(\boldsymbol{z}_1), \ldots, \widehat{g_n}(\boldsymbol{z}_n))'$, where the dependence on n is now suppressed for notational ease. There exists quite large number of different estimators proposed in the literature, many of which make use of

Figure 1: Proportion of rejection of $\mathcal{H}_0$ when (M0) [left] and (M1) [right] is true, in dependence of a. The two rightmost symbols refer to the limiting statistic (2.8). The bottom line (○) and the top line (△) correspond to the target significance levels $\alpha = 0.05$ and $\alpha = 0.10$, respectively.



the following simple idea. Noting that

$$\mathbb{E}(y|\boldsymbol{z}) = \mathbb{E}(\boldsymbol{x}|\boldsymbol{z})'\boldsymbol{\beta} + g(\boldsymbol{z}), \tag{4.1}$$

one finds by subtraction of (4.1) from (1.1)

$$\tilde{y} = \tilde{\boldsymbol{x}}'\boldsymbol{\beta} + \epsilon, \tag{4.2}$$

where $\tilde{y} = y - \mathbb{E}(y|\boldsymbol{z})$ and $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \mathbb{E}(\boldsymbol{x}|\boldsymbol{z})$. Now the vector $\boldsymbol{\beta}$ can be consistently estimated by LS if equation (4.2) is viewed as a linear regression of $\tilde{y}_j$ on $\tilde{\boldsymbol{x}}_j,\ j = 1, ..., n$. The unknown expectations in $\tilde{y}$ and $\tilde{\boldsymbol{x}}$ are estimated as

$$\widehat{\mathbb{E}(\tilde{y}|\boldsymbol{z})} = \sum_{j=1}^{n} w_j(\boldsymbol{z}) y_j \tag{4.3}$$

and

$$\widehat{\mathbb{E}(\tilde{\boldsymbol{x}}|\boldsymbol{z})} = \sum_{j=1}^{n} w_j(\boldsymbol{z}) \boldsymbol{x}_j \tag{4.4}$$

respectively, where $w_j(\boldsymbol{z}),\ j = 1, \ldots, n$ are weight functions corresponding to a suitable linear smoother. The most common choice, suggested by Robinson (1988), is to use a simple kernel-weighted estimator, which, in the case $q = 1$, can be written as

$$(4.5) \quad w_j(z) = \frac{K((z_j - z)/h)}{\sum_{i=1}^{n} K((z_i - z)/h)}.$$

We will use this estimator of $(\boldsymbol{\beta}, \boldsymbol{g})$ in Section 5 and denote it by (I). Alternatively, it has been proposed to use smoothers based on piecewise polynomials (Chen, 1988), or least squares tensor product splines (Schick, 1996). A variant, (II), of Robinson's method, which has arisen mainly out of the necessity to cross-validate the estimator in order to select a suitable bandwidth h for (4.5), is to replace, for estimation at $\boldsymbol{z} = \boldsymbol{z}_\ell$, all sums $\sum_{j=1}^{n}$ in (4.3), (4.4), and (4.5) by their leave-one-out versions, $\sum_{j \neq \ell}$ (Zhu & Ng, 2003). Furthermore, Bianco & Boente (2004) suggested to replace the (non–robust) weighted expectations in (4.3) and (4.4) by (robust) medians weighted via (4.5), and additionally to replace the LS estimator of $\boldsymbol{\beta}$ by a robust L1 estimator. We denote this estimator, which is not a linear smoother any more, by (III).

We return to the 'classical' estimator (4.3, 4.4) but reformulate it a little. For any choice of weights, define the *weight diagrams* as

$$\boldsymbol{s}_\ell' = (w_1(\boldsymbol{z}_\ell), \ldots, w_n(\boldsymbol{z}_\ell)), \quad \ell = 1, \ldots, n.$$

These can be thought of as the rows of a smoother matrix, $\boldsymbol{S}$, so that, *if one had* $p = 0$ *in (1.1)*, the fitted response would be given by $\hat{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{y}$. In this notation, one can show that the approach via (4.3, 4.4) is exactly equivalent to solving

$$(4.6) \quad \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{K})\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{K})\boldsymbol{y}$$

$$(4.7) \quad \hat{\boldsymbol{g}} = \boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$$

where $\boldsymbol{K} = \boldsymbol{S} + \boldsymbol{S}' + \boldsymbol{S}'\boldsymbol{S}$ (Speckman, 1988). A related class of estimators, (IV), is obtained by setting $\boldsymbol{K} = \boldsymbol{S}$. For symmetric matrices $\boldsymbol{S}$, this new class of estimators can be shown to correspond to solutions of a suitable penalized least squares problem (Green, Jennison, & Seheult, 1985). For kernel weights (4.5) as used in this paper, $\boldsymbol{S}$ is non-symmetric.

In the simulation study in Section 5, we will furthermore consider the semiparametric regression estimator, (V), produced by R function `gam` in R package **gam** (Hastie, 2010), using the identity link function and the default smoother (smoothing splines).

Table 1: Simulation design according to Zhu & Ng (2003). The table is arranged to match the graphical layout in Figures 3 to 7.

		ϵ	
		N(0,1)	U[-0.5,0.5]
x	U[-0.5,0.5]	Uni-Nor	Uni-Uni
	N(0,1)	Nor-Nor	Nor-Uni

This corresponds effectively to using equations (4.6) and (4.7), but with $\boldsymbol{S}$ corresponding to the cubic spline smoother matrix as provided explicitly in Fahrmeir & Tutz (2001), page 181.

All estimators (I) - (V) fulfil the assumptions (A1)-(A4) of Theorem 1, as detailed at the end of the Appendix.

5 Simulation study

Next, we carry out an extensive simulation study, using the same setup as Zhu & Ng (2003). Specifically, data sets of size $n = 100$ are simulated from the model

$$y = \beta x + bx^2 + (z^2 - 1/3) + \sqrt{12}(z - 1/2)\epsilon, \tag{5.1}$$

where all of x, z and ϵ are random variables. This is a model of type (1.1) with $\varepsilon = \varepsilon(x, z) = \sqrt{12}(z - 1/2)\epsilon$. The null hypothesis corresponds to the setting $b = 0$. Zhu & Ng found that the performance of their test, in terms of the test power, depends strongly on the type of the distribution of x and ϵ. Therefore, we follow their simulation design and use different combinations of the distributions of the variables in our study. These four combinations, along with the labels that we are going to use for them, are displayed in Table 1. For the random variable z, we use throughout a uniform distribution on [0,1].

We investigated and compared estimation methods (I)-(V) presented in Section 4, where the biweight kernel $K(u) = \frac{15}{16}(1-u^2)\mathbf{1}_{\{-1\leq u\leq 1\}}$ was used in (I)-(IV), and splines where used for estimator (V). For the smoothing parameters, we used for (I) to (IV) the values of h as found in Zhu & Ng via generalized cross-validation, namely $h = 0.30$

and $h = 0.57$ for the cases with uniform and normal error, respectively. Clearly, these bandwidths have only been found to be optimal for smoother (II), while the optimal bandwidths for (I), (III), and (IV) are likely to be different. However, there is no objective way of imposing an equal degree of smoothness onto all smoothers, since the trace of the smoother matrix is a misleading quantity for a leave–one–out estimator, and (III) does not possess a smoother matrix at all. Hence, working with equal smoothing parameters seems a fair compromise. As a contrast, for the `gam` function in (V), we used the default smoothing parameter corresponding to $\mathrm{tr}(\boldsymbol{S}) = 5$ (including the intercept). For comparison, the trace of the smoother matrix $\boldsymbol{S}$ used in the Nadaraya-Watson smoothers in (I) and (IV), is 3.5. Hence, while (I)-(IV) should be compared between themselves (all using kernels with the same bandwidth), the estimator (V) serves as an external benchmark (different smoother and smoothness).

We fitted model (5.1) using using B = 200 bootstrap replicates for each simulated data set, and repeated the procedure 3000 times (these numbers were chosen so that comparability with the results reported in Zhu & Ng (2003) is warranted). Gaussian random variables e_j were used in bootstrap step (i), and the test statistics T_a based on normal weights (2.4) was used in steps (iii) and (iv). For comparative purposes, we also applied Fan and Huang's (2001) Adaptive-Neyman test (Adj-FH), with normalized test statistics

$$T_{AN} = \sqrt{2 \log\log n}\, T_{AN}^* - \{2 \log\log n + \frac{1}{2} \log\log\log n - \frac{1}{2} \log(4\pi)\}, \tag{5.2}$$

where

$$T_{AN}^* = \max_{1 \le m \le n} \frac{1}{\sqrt{m \hat{\sigma}_2^2}} \sum_{j=1}^{m} (\hat{\varepsilon}_j^{*2} - \hat{\sigma}_1^2)$$

and $\hat{\varepsilon}_j^{*2}$ are discrete Fourier transforms of the original residuals, $\hat{\varepsilon}_j$. Formulae for the estimation of $\sigma_1^2 = \mathrm{Var}(\varepsilon)$ and $\sigma_2^2 = \mathrm{Var}(\varepsilon^2)$ are provided in Fan & Huang (2001).

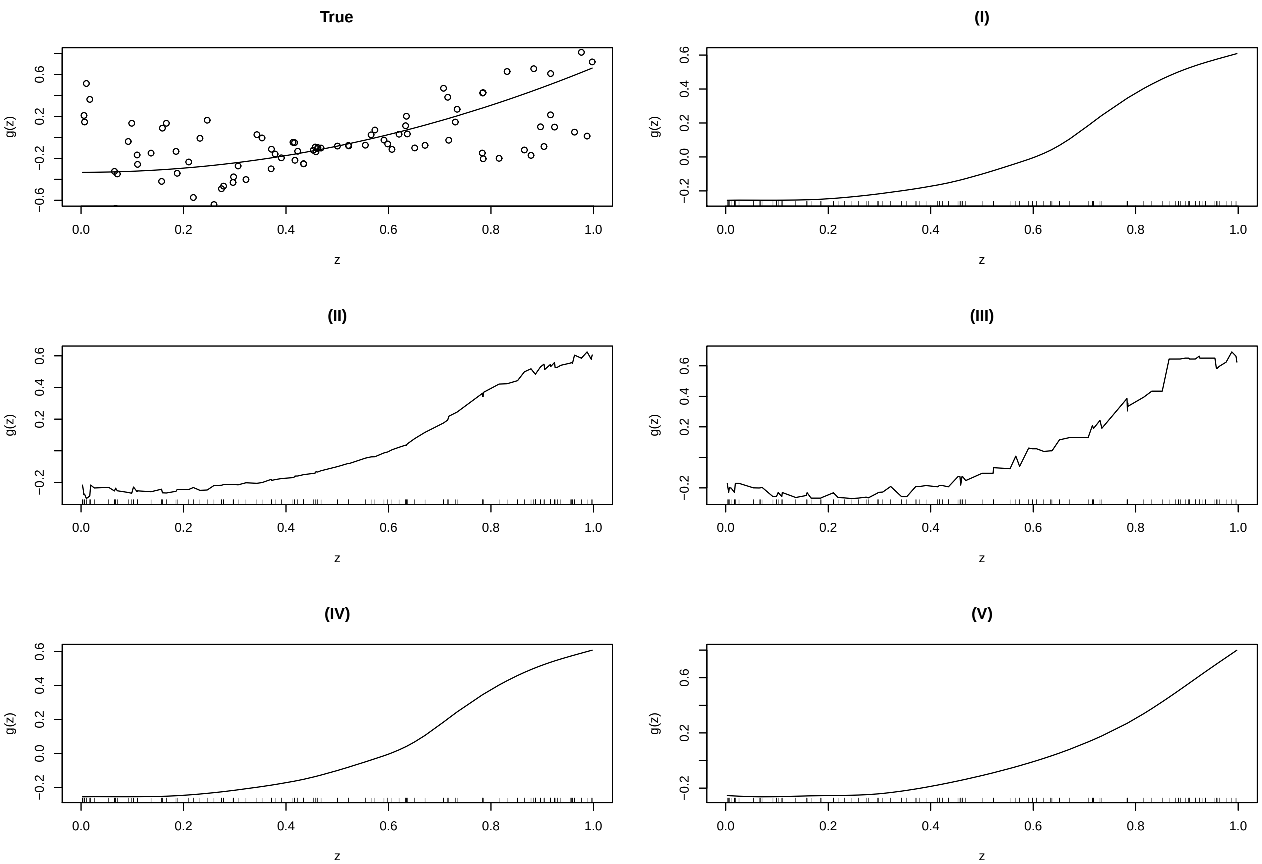
It is important to note that the Adj-FH test was developed for tests under a parametric null hypothesis, while the the proper test in the context of the null hypothesis $\mathcal{H}_0$ would be that by Fan & Li (1996). However, apart from being easier to implement, the Adj-FH test also compares well with the test of Fan & Li (1996) as shown in the simulation results reported by Zhu & Ng. An issue that will arise for either test is that the power depends strongly on the variance estimates, which become poor for large b. Zhu & Ng addressed this issue by resorting to the variance estimated under $\mathcal{H}_0$. We opted for using the true, known variances, which, for the simulation problem at hand, can be straightforwardly worked out. Exploiting the independence of z and ϵ, one finds $\text{Var}[(z-\frac{1}{2})\epsilon] = \text{Var}(z)E(\epsilon^2)$, and hence $\sigma_1^2 = 12\text{Var}(z)E(\epsilon^2) = 1/12$ for uniform error and $\sigma_1^2 = 1$ for normal error. An analogous computation gives $\sigma_2^2 = 7/450$ and $\sigma_2^2 = 22/5$, for uniform error and for normal error, respectively. According to Fan & Lin (1998), the test statistic in (5.2) satisfies:

$$(5.3) \quad P(T_{AN} < x) \longrightarrow \exp(-\exp(-x)), \ n \longrightarrow \infty.$$

The authors also provide a tabulated list of upper α quantiles. For $\alpha = 0.05$ and $n = 100$ as considered here, Fan & Lin's table gives the critical value 3.90, which was used in the simulation below. Since our null hypothesis is semiparametric, the theoretical distribution and quantiles do not transfer one–to–one to our setup. Zhu & Ng (2003) dealt with this problem by "some constant adjustment so as to maintain the significance level". We do not follow this route, but provide the observed powers at face value, hence allowing insights into the magnitude of the bias in the significance level under the different scenarios.

To give an initial impression of the nature of this heteroscedastic nonparametric estimation problem, Figure 2 (top left) gives the true function $g(z) = z^2 - 1/3$, along with partial errors $y_i - \beta x_i$ from a randomly chosen simulation loop, in the special case $\beta = 1, b = 0$. The subsequent figures give the nonparametric estimates $\hat{\boldsymbol{g}} = (\widehat{g_n}(z_1), \ldots, \widehat{g_n}(z_n))'$. As expected, the median–based estimate (III) is less smooth than the others, but notably, also the leave–one–out estimator (II) demonstrates some considerable deviation from what one would expect to be a smooth curve, especially in the boundary region. This is remarkable as the bandwidth parameters were optimized for this very estimator! We see that leaving out the "most informative observation" at each point impacts on the fitted curve considerably. Therefore, we do not think this estimator should be used generally for the actual *estimation* of a nonparametric term, unless this happens explicitly and intentionally within a cross–validation routine for the purpose of bandwidth selection. We will see that the smoothness of the nonparametric estimate $\hat{\boldsymbol{g}}$ impacts on the Adj-FH test (at least, in the form that we have used

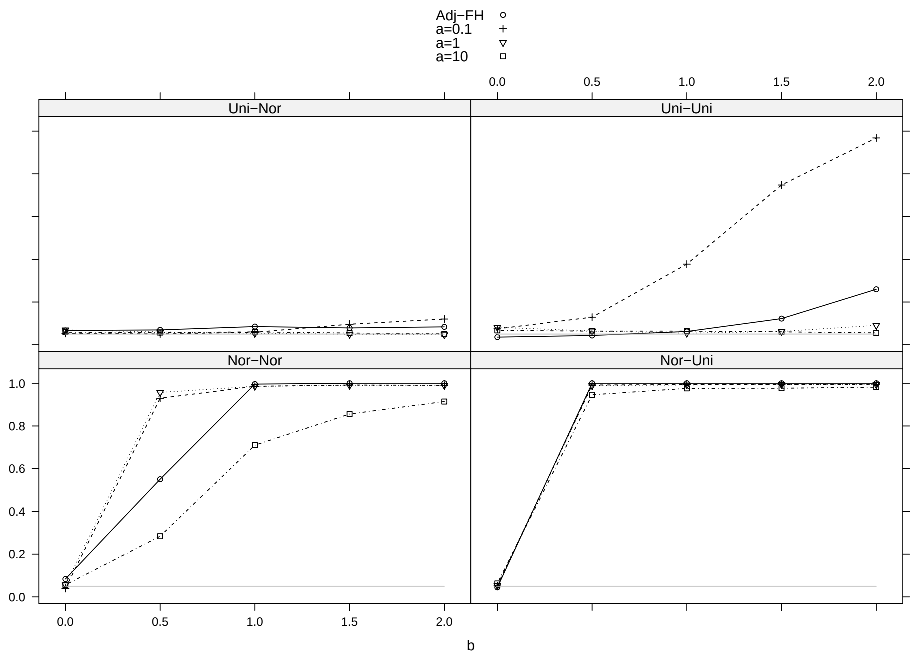
Figure 2: Top left: plot of true function $g(z) = z^2 - 1/3$, along with partial residuals $y_i - \beta x_i$ for an exemplary data set simulated from (5.1) with $\beta = 1$ and $b = 0$; other panels: nonparametric estimates $\hat{\boldsymbol{g}}$ using (I) to (V).



it), while it impacts less strongly on the test proposed herein.

The test powers (in terms of percentages of rejections of $\mathcal{H}_0 : b = 0$, over all 3000 replicates) are provided, for estimation routines (I) to (V), in graphical form in Figures 3 to 7. The target significance level is set at $\alpha = 0.05$, i.e. we reject $\mathcal{H}_0$ if the bootstrapped p–value is less than or equal 0.05. Each panel shows the percentage of rejection of $\mathcal{H}_0$ as a function of the slope parameter b. In terms of the weight parameter a, we use the values $a = 0.1$ (dashed), 1 (dotted), and 10 (dashed–dotted). While the former two settings worked generally well, the latter setting performed suboptimal, but, bearing in mind the results from Section 2.1, we included it for comparative purposes.

Figure 3: Percentage of rejection of $\mathcal{H}_0$ as a function of b, for estimator (I).



The solid line in each plot corresponds to the Adj-FH test.

Important observations to draw from these figures are: (a) For all employed estimators, and for all considered values of the weight parameter a, the target significance level is closely met under $\mathcal{H}_0$ (i.e., b=0). (b) The test power rises very quickly to 1 for Nor-Nor and Nor-Uni design, moderately quickly for Uni-Uni design, and performs very poorly for Uni-Nor design. This is in line with the results from Zhu & Ng (2003). (c) Of the three values of a considered, the best performance was observed for $a = 0.1$, closely followed by $a = 1$. The largest differences between different values of a were observed for Uni-Uni and Nor-Nor design, while the differences are negligible for Uni-Nor (here all settings perform poorly) or Nor-Uni (here all settings perform well) design. (d) Comparing Figure 4 for the leave–one–out estimator with Figure 1 in Zhu & Ng (2003), we see that our estimator, for $a = 0.1$ and Uni-Uni design, performs comparable to their CV_n test, while it performs considerably better under the Nor-Nor and Nor-Uni scenarios. Our test appears to perform even worse than CV_n under the difficult Uni-Nor scenario. (e) For normally distributed x, the Adj-FH test produced results which were not dissimilar to the test proposed herein, apart from lower powers

Figure 4: Percentage of rejection of $\mathcal{H}_0$ as a function of b, for estimator (II).

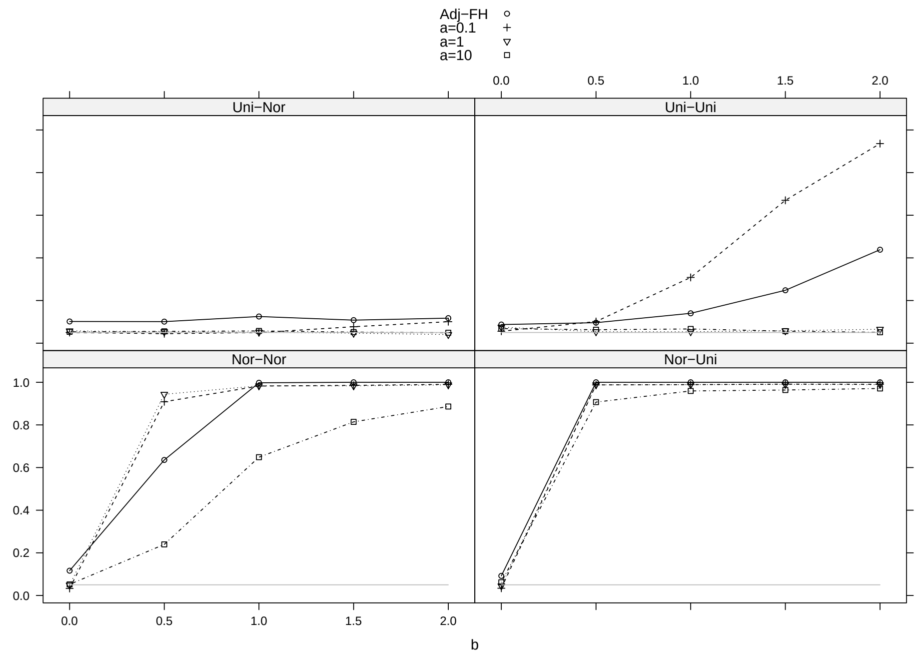


Figure 5: Percentage of rejection of $\mathcal{H}_0$ as a function of b, for estimator (III).

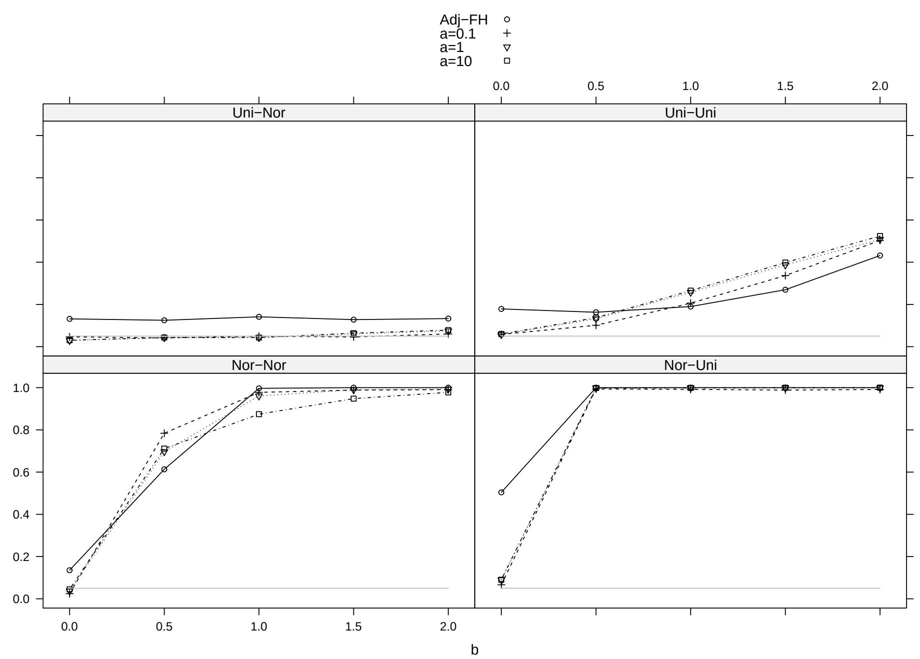
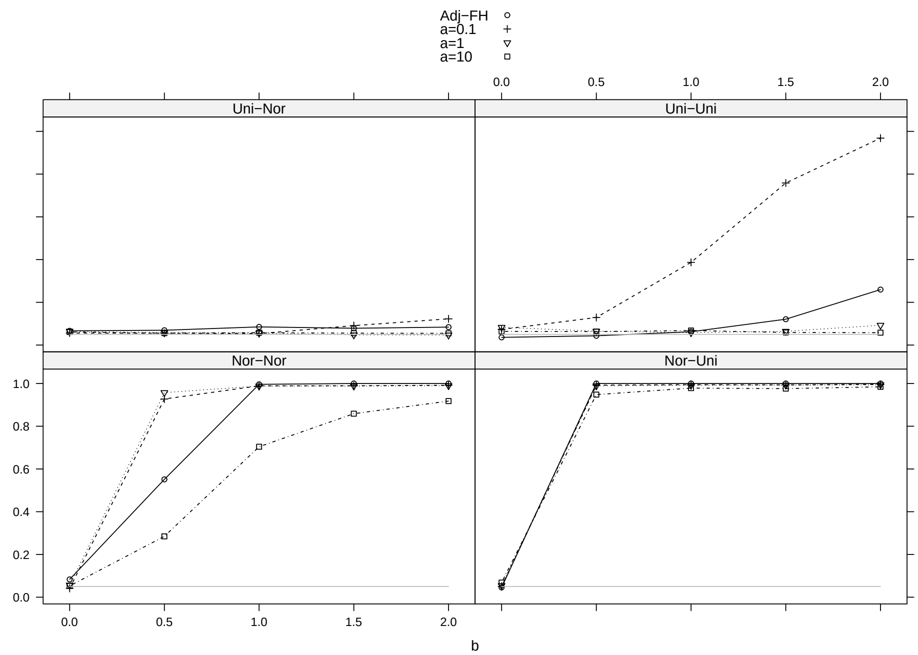


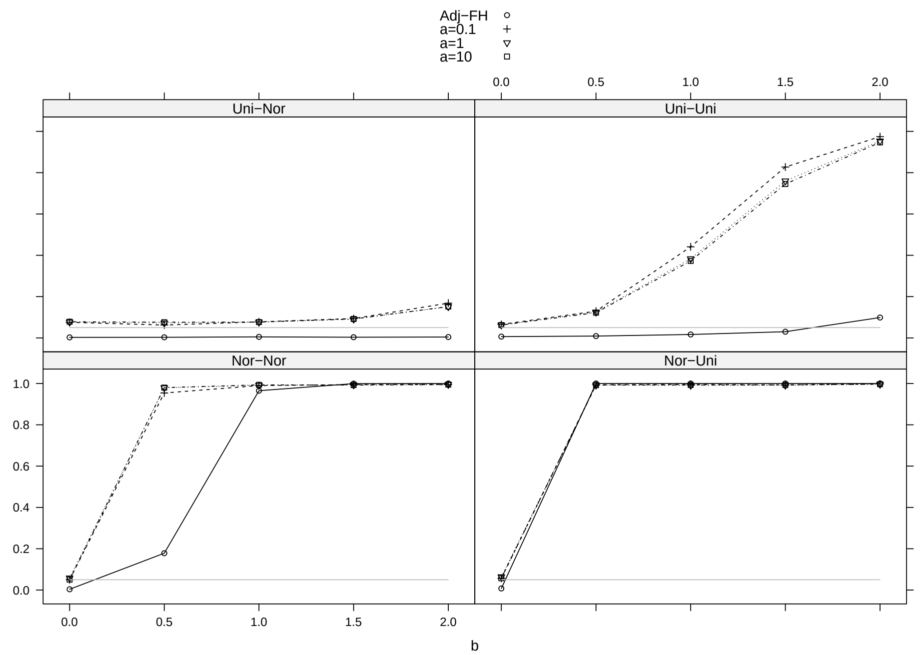
Figure 6: Percentage of rejection of $\mathcal{H}_0$ as a function of b, for estimator (IV).



for small values of b. The Adj-FH test is not suitable for uniformly distributed x, as also found by Zhu & Ng. While the target significance level was slightly undershot for estimators (I), (IV), and (V) under some design configurations, it was overshot for estimator (II) and especially strongly for estimator (III). The reason for the latter behavior is apparently found in the non–smoothness of these estimators as evidenced in Figure 2: the more the estimated function $\hat{\boldsymbol{g}}$ deviates from a smooth function, the more heavily the assumption of a parametric model (underlying the Adj-FH test) is violated. It is in these cases where Zhu & Ng's ad–hoc adjustment to maintain the significance level may be beneficial, though it is reassuring to see that in the majority of considered scenarios the Adj-FH test performs reasonably without this course of action.

A bootstrapped version of the Adj-FH test was also attempted (using again $B = 200$, and the same wild residuals as in Section 3). This led to unsatisfactory results, with unacceptably low powers overall, which we do not report in this paper. The precise reason for the failure of the wild bootstrap in this context is unclear to us. It is probably due to the type of asymptotic distribution attained by the test statistic

Figure 7: Percentage of rejection of $\mathcal{H}_0$ as a function of b, for estimator (V).



(5.2). Specifically as it may be seen from (5.3), this distribution belongs to the family of extreme value distributions and these distributions are known to often lead to failure of the bootstrap; see for instance Chernick (1999), §9.3.

6 Discussion

Some final words on the choice of the weight parameter a are in order. Though the results in the small simulated example in Section 3 indicate that the test performs quite insensitively to the choice of a, it is still impossible to pinpoint some specific value of a which would serve for all data sets, the reason being that the impact of a certain value of a onto the test statistic, such as (2.4), depends on the scale of the predictors $\boldsymbol{v} = (\boldsymbol{x}', \boldsymbol{z}')'$. However, the reassuring message to be taken from the abovementioned simulation is that, for the normal weight function, we do not need to be concerned about fixing a *precise* value of a, it is merely important to identify its right magnitude. It should be added at this occasion that this simulation was repeated using Laplace–weights (2.5). It turned out that this weight function reacts far more temperamentally to the choice of a, and powers of similar strength as for the normal weights were only achieved in the small window $0.5 \leq a \leq 1$.

For the normal weight function, a simple rule of thumb allowing to identify the right magnitude of a (without needing to fit a pilot model) is to set

$$\hat{a} = \frac{1}{n^2} \sum_{j,k=1}^{n} ||\boldsymbol{v}_j - \boldsymbol{v}_k||^2.$$

This rule delivers a value $\hat{a} \approx 0.7$ for the simulations in Section 3 and of $\hat{a} \approx 0.35$ for the simulations in Section 5. As far as the second simulation is concerned, this falls just well in between $a = 0.1$ and $a = 1$, both of which performed well in the simulation study.

To highlight the importance of this issue, let us consider briefly the Australian onions data set investigated in detail in Meintanis and Einbeck (2012), featuring 84 observations on the areal `density` of plants (plants/m^2), a `location` indicator variable, and the response log(`yield`). The obvious candidate models include (A) a linear model, and (B) a semi–linear model, with a linear term for `location` and a non–parametric term for `density`. The rule of thumb above gives, for this data set, $\hat{a} = 3400$! Application of the proposed test routine using test statistic (2.4) with weight parameter $a = 3400$, estimator (V), and 200 bootstrap replicates, yields a $p-$value of 0.00 for model (A) and 0.195 for model (B), giving, as expected, clear evidence that the semi-parametric model is favorable. However, an unreflected use of, say, $a = 1$, would lead to a loss of the test power, yielding $p \geq 0.15$ for both models.

In closing we wish to note that the notion of *validation* advocated here refers to the appropriateness of the partial linear model as a model for a given data set *and not* to that of *significance* of parameters, or to the notion of *goodness–of–fit* with respect to the error distribution. (This latter aspect was investigated in the context of partial linear models by Meintanis & Einbeck, 2012). In fact, diagnostic procedures for the problem discussed herein should precede any kind of investigative work with respect to the error–distribution, since it is the greater picture of the model that comes first and then come certain specifications of the ingredients of the model (be it parameter significance or goodness-of-fit for the error distribution). It consequently becomes clear that the current work is nicely complemented by that in Meintanis & Einbeck (2012) since the two procedures can be used in conjunction if one wishes to validate a partial linear model with a specific structure for the error distribution.

Acknowledgments

The collaboration between the two authors was supported by LMS Scheme 4 Grant Ref 4902. Simos Meintanis acknowledges support by Special Account Research Grant 11105 of the Department of Economics, National and Kapodistrian University of Athens.

Appendix

Assumptions for Theorem 1.

(A1) The weight function satisfies $W(\boldsymbol{t}) > 0$, except for a set of measure zero, and $\int_{\mathbb{R}^{p+q}} W(\boldsymbol{t}) d\boldsymbol{t} < \infty$.

(A2) The estimator of $\boldsymbol{\beta}$ satisfies $\widehat{\boldsymbol{\beta}_n} \longrightarrow \boldsymbol{\beta}$, almost surely, as $n \to \infty$.

(A3) The estimator of $g(\boldsymbol{z})$ satisfies $\max_{\boldsymbol{z}} |\widehat{g_n}(\boldsymbol{z}) - g(\boldsymbol{z})| \longrightarrow 0$, almost surely.

(A4) The regressor vector satisfies $\mathbb{E}\|\boldsymbol{x}\|_1 < \infty$.

Proof of Theorem 1. Let $\widetilde{E}_n(t) = n^{-1} \sum_{j=1}^n \varepsilon_j e^{i\boldsymbol{t}'\boldsymbol{v}_j}$. Then by straightforward algebra we have

$$(6.1) \quad \left|\widetilde{E}_n(\boldsymbol{t}) - E_n(\boldsymbol{t})\right| \le \frac{1}{n}\sum_{j=1}^n \left|\boldsymbol{x}_j'(\widehat{\boldsymbol{\beta}_n} - \boldsymbol{\beta})\right| + \frac{1}{n}\sum_{j=1}^n \max_{\boldsymbol{z}} |\widehat{g_n}(\boldsymbol{z}) - g(\boldsymbol{z})| \longrightarrow 0,$$

almost surely, as $n \to \infty$. Also by the Law of Large Numbers $\widetilde{E}_n(\boldsymbol{t}) \longrightarrow E(\boldsymbol{t})$, almost surely, as $n \to \infty$, which in conjunction with (6.1) and by invoking Fatou's Lemma yields

$$(6.2) \quad \liminf_{n\to\infty} T_{n,W} \ge \int_{\mathbb{R}^{p+q}} |E(\boldsymbol{t})|^2 W(\boldsymbol{t}) d\boldsymbol{t},$$

almost surely. By following analogous arguments to those in (6.1) it follows that $n^{-1}\sum_{j=1}^n |\varepsilon_j| - n^{-1}\sum_{j=1}^n |\widehat{\varepsilon}_j| \longrightarrow 0$, almost surely, and since $|E_n(\boldsymbol{t})|^2 \le \left(n^{-1}\sum_{j=1}^n |\widehat{\varepsilon}_j|\right)^2$, a variation of Fatou's Lemma (see Stroock, 1999, §3.3) yields

$$(6.3) \quad \limsup_{n\to\infty} T_{n,W} \le \int_{\mathbb{R}^{p+q}} |E(\boldsymbol{t})|^2 W(\boldsymbol{t}) d\boldsymbol{t},$$

almost surely. Finally (6.2) and (6.3) imply (2.6) which concludes the proof ■

Remark. Assumptions (A2) and (A3) are fairly mild (in that they only imply consistency and not asymptotic normality), and hold true for most estimators found in the literature; see Schick (1996), Härdle et al. (2000), Gannaz (2007), and Wong et al. (2009) for general references. More specifically, we note that the estimators (I), (II) and (III) considered herein all satisfy assumption (A2) according to Robinson (1988, p. 939) or Speckman (1988, Theorem 2), Zhu & Ng (2003, p. 766), and Bianco & Boente (2004, Theorem 1), respectively. Likewise, for assumption (A3) the reader is referred to Speckman (1988, Theorem 3), Zhu (2005, Theorem 5.2.1) or Zhu &Ng (2003, p. 766), and Bianco & Boente (2004, p. 236), respectively. For estimators (IV) and (V), the corresponding results are found in Speckman (1998), Theorem 1 (for A2) and Theorem 3 (for A3).

References

Bianco, A. and Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *J. Statist. Plann. Infer.*, *122*, 229–252.

Bierens, H.J. (1982). Consistent model specification tests. *J. Econometr.*, *20*, 105–134.

Bierens, H.J. (1990). A consistent conditional moment test of functional form. *Econometrica*, *58*, 1443–1458.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.*, *16*, 136–146.

Chernick, M.R. *Bootstrap Methods. A Practitioner's Guide.* Wiley, 1999.

Delgado, M.A. and Fiteni, I. (2002). External bootstrap tests for parameter stability. *J. Econometr.*, *109*, 275–303.

Delgado, M.A. and González–Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.*, *29*, 1469–1507.

Fahrmeir, L., and Tutz, G. (2001). *Multivariate Statistical Modelling based on Generalized Linear Models.* Springer, Berlin.

Fan, J. and Huang, L. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.*, *96*, 310–320.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica*, *64*, 865–890.

Fan, Y. and Lin, S.-K. (1998). Test of significance when data are curves. *J. Amer. Statist. Assoc.*, *93*, 1007–1021.

Gannaz, I. (2007). Robust estimation and wavelet thresholding in partially linear models. *Statist. & Comput.*, *17*, 293–310.

Green, P., Jennison, C., and Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *J. Roy. Statist. Soc. B*, *47*, 299–315.

Härdle, W., Mammen, E., and Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *J. Amer. Statist. Assoc.*, *93*, 1461–1474.

Härdle, W., Liang, H. and Gao, J. *Partially Linear Models*. Physica–Verlag, 2000.

Hastie, T. *gam: Generalized additive models.* (2010). R package version 1.03, `http://CRAN.R-project.org/package=gam`.

Hušková, M. and Meintanis, S.G. (2007). Omnibus tests for the error distribution in the linear regression model. *Statistics*, *41*, 363–376.

Hušková, M. and Meintanis, S.G. (2010). Tests for the error distribution in nonparametric possibly heteroscedastic regression models. *Test*, *19*, 92–112.

Meintanis, S. G. and Einbeck, J. (2012). Goodness-of-fit tests in semi–linear models. *Statistics and Computing*, *22*, 967-979.

Robinson, M. P. (1988). Root-N-consistent semiparametric regression. *Econometrica*, *56*, 931–954.

Schick, A. (1996). Root–n consistent estimation in partly linear regression models. *Statistics & Probability Letters*, 28, 353–358.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, *50*, 413–436.

Stroock, D.W. *A Concise Introduction to the Theory of Integration*. Birkhäuser, 1999.

Wong, H., Liu, F., Chen, M. and Ip, W.C. (2009). Empirical likelihood based diagnostics for heteroscedasticity in partial linear models. *Comput. Statist. Dat. Anal.*, *53*, 3466–3477.

Yatchew, A.J. (1992). Nonparametric regression tests based on least squares. *Econometr. Theory*, *8*, 435–451.

Zhu, L.X. and Ng, K.W. (2003). Checking the adequacy of a partial linear model. *Statist. Sinica*, *13*, 763–781.

Zhu, L. (2005). Checking the adequacy of a partially linear model. *Lecture Notes in Statistics*, *182*, 123–139.