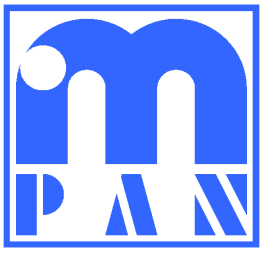
INSTITUTE OF MATHEMATICS
of the
Polish Academy of Sciences

ul. Śniadeckich 8, P.O.B. 21, 00-956 Warszawa 10, Poland http://www.impan.gov.pl



IM PAN Preprint 665 (2006)

Tadeusz Inglot and Teresa Ledwina

Data driven score tests of fit for a semiparametric homoscedastic linear regression model

Published as manuscript

Received 3 January 2006

DATA DRIVEN SCORE TESTS OF FIT FOR A SEMIPARAMETRIC HOMOSCEDASTIC LINEAR REGRESSION MODEL

Tadeusz Inglot

Polish Academy of Sciences and Wrocław University of Technology

Teresa Ledwina

Polish Academy of Sciences

Summary. We propose new tests for testing the validity of a semiparametric random-design linear regression model. The construction consists of several steps. First, we follow the classical idea of overfitting and replace the basic problem by a series of auxiliary subproblems. Next, to test whether extra terms are significant we construct a counterpart of classic score statistic. In passing, a handy way of deriving the efficient score is proposed and developed. Finally, we combine the solution with smoothing methods providing guidelines to choose the right subproblem. This leads to data driven score tests for the initial testing problem. We show that under the null model our construction is asymptotically distribution free and illustrate this result by a small simulation study. We also compare the finite sample performace of our tests with the the recent solution introduced by Guerre and Lavergne (2005), as well as to Cramér-von Mises type construction. The simulation experiment indicates the very good performance of the proposed tests.

Key words and phrases. Cramér-von Mises test, efficient score, hypothesis testing, data driven test, linear regression, selection rule, semiparametric inference, smoothing methods.

1. Introduction. The problem of verifying the linear structure of a regression function is central in applied statistics. Therefore, it is not surprising that there is an extensive literature on several possible solutions under a variety of different restrictions. Some of the solutions are briefly discussed in Section 2, below. For further references, mostly focused on the fixed design set-up, see Hart (1997).

The purpose of this paper is to propose and investigate some new tests of fit for the following problem. Let $Z = (X, Y)$ denote a random vector in $I \times R$, $I = [0, 1]$. We would like to verify the null hypothesis H_0 asserting

$$Y = \beta[v(X)]^T + \epsilon, \tag{1.1}$$

where X and ϵ are independent, $E\epsilon = 0$, $E\epsilon^2 < \infty$, $\beta \in R^q$ is a vector of unknown real valued parameters while $v(x) = (v_1(x), ..., v_q(x))$ is a vector of known functions. The symbol ^{T} denotes transposition. All vectors are considered as row vectors.

We follow the classical idea of overfitting and reducing the verification of (1.1) to testing whether extra terms are significant. More precisely, given a fixed k, we embed our null model (1.1) into the following auxiliary model

$$\texttt{M(k)} \qquad Y = \theta[u(X)]^T + \beta[v(X)]^T + \epsilon, \tag{1.2}$$

which satisfies the following assumptions

$u(x) = (u_1(x), ..., u_k(x))$, $v(x) = (v_1(x), ..., v_q(x))$, $x \in I$, and the measurable functions $u_1, ..., u_k, v_1, ..., v_q$ are bounded and linearly independent;

$\theta \in R^k$, $\beta \in R^q$ are unknown parameters;

$< \mathbf{M1} >$ X has an unknown density g with respect to the Lebesgue measure λ supported on I;

ϵ has an unknown density f with respect to the Lebesgue measure λ on R. The density f satisfies $E_f\epsilon = 0$, $\tau = E_f\epsilon^2$ and $0 < \tau < \infty$;

X and ϵ are independent.

At the first step we construct appropriate score test statistic, for the fixed k, for testing $H_0(k) : \theta = 0$ against $\theta \neq 0$ in $\texttt{M(k)}$ satisfying $< M1 >$ and some further regularity conditions $< M2 >$ and $< M3 >$. An efficient score vector along with its appropriate estimator play the central role in this construction. Section 7 briefly presents our approach to a derivation of efficient scores. This section may be of independent interest. The next step consists in incorporating into this statistic a score - based selection rule for determining the dimension k. The both steps are presented in detail in Section 3. This section is preceded by Section 2 containing motivation for the proposed construction, related discussion and some references to existing solutions of the considered problem. Section 4 presents the results of simulation study. Section 5 contains a proof of the crucial result on the asymptotic behaviour of the estimate of the efficient score vector under the null model (1.1). In Section 6 we discuss various aspects of our general assumptions. Finally in the Appendix we check the assumptions related to our implementation of the test in Section 4.

2. Motivation of the approach. The first rigorous approach to defining and constructing tests which are asymptotically optimal was by Neyman (1937). Roughly speaking, the

paper introduced an asymptotically locally most powerful test of fit to a completely specified null distribution. The resulting solution was called the smooth test and can be seen to be a standard score statistic [under the set-up considered by Neyman]. Note that this score statistic is simply the Euclidean norm of the score vector. In 1959 Neyman successfully extended this idea to cover the case of testing a parametric hypothesis in the case where some Euclidean nuisance parameters are present [cf. also Neyman (1954) and Le Cam (1956) for some preliminary results and their improvements]. The key elements of Neyman's asymptotically locally optimal solutions (1954, 1959) were residual scores calculated as the residuals from projections [derived under the null hypothesis] of scores for the parameters of interest onto scores for the nuisance parameters. Nowadays the residuals are called efficient scores. Neyman's resulting statistic is some norm of the efficient score vector.

In the thirties other goodness of fit statistics for a completely specified null distribution were introduced. Cramér-von Mises and Kolmogorov-Smirnov proposals are prominent examples. In contrast to Neyman's solution, these statistics were based mainly on intuition, as being measures of distance between theoretical and empirical distributions. Goodness of fit testing was dominated by solutions of this kind for decades. This remark applies also to goodness of fit tests for semiparametric regression in the case X is random. In particular, Stute (1997) and Stute et al. (1998a,b) developed some Cramér-von Mises type tests. Some simplified variants of such statistics were proposed by Diebolt and Zuber (2000). Kozek (1991), Härdle and Mammen (1993) and many others proposed defining the distance between parametric and nonparametric estimators. Horowitz and Spokoiny (2001) refined this approach by using data driven choice of a smoothing parameter. Recently, an alternative construction based on nonparametric smoothing methods and penalization was introduced by Guerre and Lavergne (2005). Roughly speaking, these solutions rely on the not entirely justified belief that good estimators produce sensitive tests. The papers by Cox et al. (1988), Azzalini and Bowman (1993), Aerts et al. (2000) and Fan and Huang (2001) were exceptions to the mainstream. In these articles the starting point for test construction were some ideas related to testing theory. These four papers deal with the case of fixed design. The study of Dette (2000) extended the solution of Azzalini and Bowman (1993) to the case of random design.

Returning to Neyman's approach, it should be noted that smooth tests rose little interest for many years, while nowadays Neyman's 1937 paper is considered to be ingenious [cf. Le Cam and Lehmann (1975), p. ix]. Renewed interest in this solution and its 1959 extension, related to goodness of fit problems, was observed after the paper by Thomas and Pierce (1979) and accelerated by the book of Rayner and Best (1989). It should also be noted that the theory and applications elaborated there concerned goodness of fit testing in the case where some Euclidean niusance parameters are present. The resulting solutions were also called smooth tests. A justification for the name was provided by Javitz (1975), who showed that Neyman's tests are simply efficient score tests for some natural parametric family.

However, it was increasingly clear that the practical application of smooth tests to goodness of fit problems should be accompanied by careful selection of the number of components in the test statistic. In the case of a fully specified null distribution, solutions of this kind were proposed by Eubank et al. (1993), Ledwina (1994), Fan (1996), Aerts et al. (2000), to mention few. In particular, the construction introduced in Ledwina (1994) is closely related to the original idea of Neyman (1937), as it provides asymptotically locally most powerful test for a large class of nonparametric alternatives [for some evidence see e.g. Inglot and Ledwina (1996) and Inglot and Ledwina (2001a)]. The solution relies on using Neyman's smooth test with the number of components defined by Schwarz selection rule. The case of testing goodness of fit when some Euclidean nuisance parameters are present was solved also in a similar way [cf. Inglot et al. (1997) and Inglot and Ledwina (2001b)]. The aim of this article is to apply a

suitable counterpart of Neyman's solution, along with the data driven choice of the number of components incorporated in order to construct test of fit for the model (1.1).

It should be said that the last few decades have been a period of vigorous developement of semiparametric estimation theory. Efficient scores also play central role in it. An important idea of applying results derived in semiparametric estimation, in order to construct some score tests in the case where functional nuisance parameters are present, has been considered in Choi (1989) and Choi et al. (1996). See also Bickel et al. (1998) for an alternative approach. In order to link our solution more clearly to these important contributions, we shall use the name score test instead of smooth test. Moreover, note that the name score test is an abbreviation of a more suitable name: efficient score test. Finally, let us recall that the primary importance of efficient score tests lie in the fact that, under the null model, the influence of the nuisance parameters on the null distribution is asymptotically negligible. The second advantage of efficient score tests is that they are locally optimal solutions.

In the context presented above, it is quite obvious that the paper by Choi et al. (1996) stimulated us. On the other hand, it seems to be a difficult task to follow the outline and suggestions sketched in Section 7 of that paper to someone not experienced in the details and particular concerns of techniques of efficient estimation. The guidelines given in that paper are very rough and a lot of work is needed to adapt them to a working solution in some particular application. For some further discussion on this point see Remark 4 of Section 5. Anyway, the idea turns out to be worthy of this effort. To extract, among other things, the essence of the technicalities which are needed in constructing a test, we decided to rederive some results on efficient scores stated in the literature and to present a minimal set of readable assumptions under which these results are valid in our set-up. In particular, by embedding the underlying probability model into a related abstract setting, we manage to clearly separate purely analytical work, such as differentiation and projections, from probabilistic arguments. We comment on this approach in Section 6. It seems that this may be of independent interest. Moreover, we propose an estimator of the efficient score vector and provide a detailed proof that its limiting null distribution is independent of the nuisance parameters. In this proof we used some well established ideas, as well as a very useful recent result of Schick (2001).

Having constructed an appropriate score statistic, we define a score-based selection rule, which mimics the Schwarz criterion in the application considered. We also propose a refinement of this selection rule, which combines the advantages of the Schwarz and Akaike criteria. This two ingredients, the score statistic and the selection rule for the number of components in the score statistic, lead to the final solution - a data driven score test, which we present in Section 3. The simulation results presented in Section 4 show that these data driven constructions possess two fundamental advantages of efficient score statistics. Namely, for moderate sample sizes the critical values are stable for a variety of nuisance parameters, while empirical powers are high, considerably dominating those of the best existing solutions in the area.

Though the present paper concentrates on one particular problem, it is obvious that similar approach can be adopted and developed for many others semiparametric and nonparametric testing problems.

3. Data driven score tests.

Before we introduce the test statistics, we present a series of auxiliary constructions and results.

3.1. *Efficient score vector for testing $\theta = 0$ in* `M(k)`.

As mentioned in Section 2, we rederived some existing results on score vectors in the model `M(k)` and derived an efficient score vector for testing (1.1). The calculations for (1.2), as well as in the more general heteroscedastic case, are given in Inglot and Ledwina (2003a). A general result for score vectors in some large

class of regression models is given in Schick (1997).

In the case under consideration, in addition to the basic model assumptions $< M1 >$ we need the following ones

$< \mathbf{M2} >$ $\quad f'(y)$ exists for all $y \in R$ and $J = J(f) = \int_R \frac{[f'(y)]^2}{f(y)} \lambda(dy) < \infty$,

$< \mathbf{M3} >$ $\quad g > 0 \quad \lambda$ - a.e.

Under these three assumptions the efficient score vector for testing $H_0(k) : \theta = 0$ in `M(k)` is of the form

$$\ell^*(z) = -\left[\frac{f'}{f}\left(y - v(x)\beta^T\right)\right][\tilde{u}(x) - \tilde{v}(x)\mathbf{V}^{-1}\mathbf{M}] + \\ + \frac{1}{\tau}[y - v(x)\beta^T][m_1 - m_2\mathbf{V}^{-1}\mathbf{M}], \tag{3.1}$$

where

$$m_1 = E_g u(X),\ m_2 = E_g v(X),\ m = (m_1, m_2), \\ \tilde{w}(x) = (\tilde{u}(x), \tilde{v}(x)),\ \tilde{u}(x) = u(x) - m_1,\ \tilde{v}(x) = v(x) - m_2, \tag{3.2}$$

while $\mathbf{M}$ and $\mathbf{V}$ are blocks in

$$\mathbf{W} = \begin{pmatrix} \mathbf{U} & \mathbf{M}^T \\ \mathbf{M} & \mathbf{V} \end{pmatrix} = \frac{1}{4}\left\{JE_g[\tilde{w}(X)]^T[\tilde{w}(X)] + \frac{1}{\tau}m^T m\right\}. \tag{3.3}$$

Note that, due to $< M3 >$, $\mathbf{W}$ is positive definite [cf. Remark C.13 in Inglot and Ledwina (2003a)].

3.2. *Efficient score statistic and a general result.* We introduce the additional notation

$$\vartheta = (\sqrt{g}, \sqrt{f}),\ \ \eta = (\beta, \vartheta)\ \text{ and }\ \ell^*(z;\eta) = \ell^*(z).$$

Moreover, let P_η^n denote the joint distribution of $Z_1, ..., Z_n$ under the null model (1.1).

Finally set

$$\mathbf{W}^{11} = (\mathbf{U} - \mathbf{M}^T\mathbf{V}^{-1}\mathbf{M})^{-1},\ \ \mathbf{L} = \frac{1}{4}\mathbf{W}^{11} \tag{3.4}$$

and define

$$W_k(\eta) = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \ell^*(Z_i;\eta)\right]\mathbf{L}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \ell^*(Z_i;\eta)\right]^T.$$

From $< M1 > - < M3 >$, Corollaries C.16, C.18 and Remark C.13 of Inglot and Ledwina (2003a), e.g., under the null hypothesis $H_0(k)$, $\mathbf{L}$ is positive definite and it holds that

$$E_\eta \ell^*(Z;\eta) = 0,\ \ \left\{E_\eta[\ell^*(Z;\eta)]^T[\ell^*(Z;\eta)]\right\}^{-1} = \mathbf{L},\ \ W_k(\eta) \xrightarrow{\mathcal{D}} \chi^2_k, \tag{3.5}$$

where χ^2_k denotes a random variable from the central chi-square distribution with k degrees of freedom.

Define

$$W_k(\hat{\eta}) = \left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\ell}^*(Z_i;\hat{\eta})\right]\hat{\mathbf{L}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{\ell}^*(Z_i;\hat{\eta})\right]^T, \tag{3.6}$$

where $\hat{\ell}^*(\bullet;\hat{\eta})$ is an estimator of $\ell^*(\bullet;\eta)$, while $\hat{\mathbf{L}}$ is an estimator of $\mathbf{L}$.

Finally, let $||\bullet||$ denote the Euclidean norm of a given vector, while the symbol $\bigwedge_{\bullet}$ stands for the statement: for every $\bullet$. The relation (3.5) and a simple argument yield the following result.

PROPOSITION 1. *Assume the null hypothesis $H_0(k): \theta = 0$ is true and the assumptions $< M1 >, < M2 >$ and $< M3 >$ are fulfilled. Suppose that $\hat{\mathbf{L}}$ is a consistent estimator of $\mathbf{L}$ and the estimator $\hat{\ell}^*(\bullet; \hat{\eta})$ satisfies the following condition*

$$\bigwedge_{\delta>0} \quad P_{\eta}^{n}\left(\frac{1}{\sqrt{n}}\left|\left|\sum_{i=1}^{n}[\hat{\ell}^*(Z_i;\hat{\eta}) - \ell^*(Z_i;\eta)]\right|\right| \geq \delta\right) \to 0 \quad as \quad n \to \infty. \tag{3.7}$$

Then for the test statistic $W_k(\hat{\eta})$ defined in (3.6) it holds that

$$W_k(\hat{\eta}) \xrightarrow{\mathcal{D}} \chi_k^2, \quad \text{as} \quad n \to \infty.$$

REMARK 1. $W_k(\hat{\eta})$ is an efficient score statistic for testing $H_0(k)$ in `M(k)`. As said before, we shall abbreviate this name to score statistic. Choi et al. (1996) used the name efficient test statistic for such a construction.

3.3. *Some class of estimators $\hat{\ell}^*$ of ℓ^* satisfying (3.7).* We follow some well established ideas. On one hand, our construction is obviously linked to the approach of Bickel (1982), Example 3. On the other hand, our solution incorporates the very useful contribution of Schick (1986) showing that using only a small fraction of the sample to estimate the score function, as Bickel (1982) did, can be avoided.

Suppose $Z_1, ..., Z_n$ are i.i.d. vectors obeying (1.2). Note that, as usual in score test theory, all considerations below are done under the assumption $\theta = 0$.

Take $\zeta = \left[\frac{n}{2}\right]$ and divide $Z_1, ..., Z_n$ into two parts $Z_1, ..., Z_\zeta$ and $Z_{\zeta+1}, ..., Z_n$. In order to clearly show an important feature of our construction, we shall, for a moment, display in formulas the expectation m as if it were the next nuisance parameter. Additionally set $< 1 >= \{1, ..., \zeta\}$, $< 2 >= \{\zeta + 1, ..., n\}$. The superscript (j), $j = 1, 2$, appearing below, indicates from which part of the sample we estimate the related quantity.

The basic structure of $\hat{\ell}^*$ at the observed points $Z_1, ..., Z_n$ is as follows

$$\hat{\ell}^*(Z_i;\hat{\eta}) = \ell^*\left(Z_i; \hat{\beta}_*^{(2)}, \hat{g}^{(2)}, \hat{f}^{(2)}, \hat{m}^{(1)}\right), \quad \text{if } i \in < 1 >$$

and (3.8)

$$\hat{\ell}^*(Z_i;\hat{\eta}) = \ell^*\left(Z_i; \hat{\beta}_*^{(1)}, \hat{g}^{(1)}, \hat{f}^{(1)}, \hat{m}^{(2)}\right), \quad \text{if } i \in < 2 >,$$

where

$$\hat{m}_1^{(1)} = \frac{1}{\zeta}\sum_{i\in<1>} u(X_i), \quad \hat{m}_2^{(1)} = \frac{1}{\zeta}\sum_{i\in<1>} v(X_i),$$

$$\hat{m}_1^{(2)} = \frac{1}{n-\zeta}\sum_{i\in<2>} u(X_i), \quad \hat{m}_2^{(2)} = \frac{1}{n-\zeta}\sum_{i\in<2>} v(X_i),$$

$$\tilde{u}^{(j)}(\bullet) = u(\bullet) - \hat{m}_1^{(j)}, \quad \tilde{v}^{(j)}(\bullet) = v(\bullet) - \hat{m}_2^{(j)}, \quad j = 1, 2,$$

while $\hat{\beta}_*^{(j)}$ is a discretized version of a $\sqrt{n}$ - consistent estimator $\hat{\beta}^{(j)}$ of β, based on the jth part of the sample.

The specific form of $\hat{m}^{(j)}$, together with the fact that in the construction of $\hat{\ell}^*$ only the estimators $\hat{m}^{(j)}$ are matched to Z_i with i from $< j >$ guarantee that the important property (5.7) holds [cf. Section 5]. Moreover, the requirements for $\sqrt{n}$ - consistency of an estimator for β and the specific form of $\hat{m}^{(j)}$ are the strongest requirements on estimators we imposed in the construction. When estimating other quantities there is a lot of freedom, as seen from Theorem 1, below.

To write the form of the estimators $\hat{\ell}^*(Z_i;\hat{\eta}),\ i \in< j >,\ j = 1,2$, explicitly denote by $\hat{\mathbf{V}}^{(j)}, \hat{\mathbf{M}}^{(j)}, \hat{\tau}^{(j)}, \hat{\tau}^{(j)} > 0$ – a.e. and $\widehat{[f'/f]}^{(j)}$ the related estimators of the appropriate quantities. Note that having these estimators, we do not need to estimate the density g itself. We also introduce auxiliary functions $\mathcal{L}_j^*,\ j = 1,2$, defined as follows:

$$\mathcal{L}_1^*(z;\beta) = -\widehat{[f'/f]}^{(2)}\left(y - v(x)\beta^T\right)\left[\tilde{u}^{(1)}(x) - \tilde{v}^{(1)}(x)[\hat{\mathbf{V}}^{(2)}]^{-1}\hat{\mathbf{M}}^{(2)}\right] + \\ + \frac{1}{\hat{\tau}^{(2)}}\left[y - v(x)\beta^T\right]\left[\hat{m}_1^{(1)} - \hat{m}_2^{(1)}[\hat{\mathbf{V}}^{(2)}]^{-1}\hat{\mathbf{M}}^{(2)}\right], \tag{3.9}$$

$$\mathcal{L}_2^*(z;\beta) = -\widehat{[f'/f]}^{(1)}\left(y - v(x)\beta^T\right)\left[\tilde{u}^{(2)}(x) - \tilde{v}^{(2)}(x)[\hat{\mathbf{V}}^{(1)}]^{-1}\hat{\mathbf{M}}^{(1)}\right] + \\ + \frac{1}{\hat{\tau}^{(1)}}\left[y - v(x)\beta^T\right]\left[\hat{m}_1^{(2)} - \hat{m}_2^{(2)}[\hat{\mathbf{V}}^{(1)}]^{-1}\hat{\mathbf{M}}^{(1)}\right].$$

Finally set

$$\hat{\ell}^*(Z_i;\hat{\eta}) = \mathcal{L}_1^*(Z_i;\hat{\beta}_*^{(2)}) \quad \text{for } i \in< 1 >, \quad \hat{\ell}^*(Z_i;\hat{\eta}) = \mathcal{L}_2^*(Z_i;\hat{\beta}_*^{(1)}) \quad \text{for } i \in< 2 > . \tag{3.10}$$

THEOREM 1. *Suppose that under the null distribution P_η^n for $j = 1,2$ the following hold : $\hat{\beta}^{(j)}$ are $\sqrt{n}$ - consistent estimators of β, while $\hat{\tau}^{(j)}, \hat{\mathbf{V}}^{(j)}$ and $\hat{\mathbf{M}}^{(j)}$ are consistent estimators of $\tau, \mathbf{V}$ and $\mathbf{M}$, respectively. Moreover, assume that the estimators $\widehat{[f'/f]}^{(j)}$, $j = 1,2$, of f'/f are consistent in the L_2 norm, i.e.*

$$\bigwedge_{\delta>0} \quad P_\eta^n\left(\int_R\left(\widehat{[f'/f]}^{(j)}(y) - [f'/f](y)\right)^2 f(y)\lambda(dy) > \delta\right) \to 0 \quad \text{as} \quad n \to \infty. \tag{3.11}$$

Then the estimator $\hat{\ell}^*$ of ℓ defined in (3.10) satisfies the condition (3.7) of Proposition 1.

REMARK 2. Theorem 1 shows that there is a lot of flexibility in chosing estimators defining $\hat{\ell}^*$. In Section 4.1 we propose a particular choice, while in the Appendix we check that the selected estimators fulfil the above requirements. The flexibility in choosing estimators is an attractive feature of the approch. It permits refinements of our relatively simple implementation by using robust estimators of β and τ [cf. Maronna and Yohai (1981) for a related approach] or more sophisticated estimators of f'/f and J, which we need to estimate $\mathbf{L}$, cf. (3.3) and (3.4), [see e.g. Koul and Susarla (1983) and Csörgő and Révész (1986) for such solutions for a fixed design and classical score function estimation, respectively, as well as Faraway (1992) and Jin (1992) in random design regression models] etc. Presumably, by incresing the complication of calculations a sample splitting scheme could be avoided, as done in a series of estimation problems by van der Vaart (1988) and Schick (1993, 1994, 1997). On the other hand, Schick's sample splitting has some other advantages, apart from being relatively simple. For evidence

see Klaassen (2001). Also, after incorporating more complicated notation and considerations, a more complex system of functions u_i's, such as piecewise polynomials, splines etc., could be included for modelling the alternatives `M(k)`. Since our primary and ultimate goal was, however, to understand the basic features of the approach and to propose simple working solutions, we have not considered these possibilities.

3.4. *Determining k in $W_k(\hat{\eta})$ by some score-based selection rules.* We now consider a nested family of auxiliary models `M(k)`, $k = 1, ..., d$, where d is fixed but otherwise arbitrary. Following the construction proposed in Ledwina (1994), as e.g. in Kallenberg and Ledwina (1997a) we define score-based selection rule $S1$ as follows

$$S1 = \min\{1 \le k \le d : W_k(\hat{\eta}) - k \log n \ge W_s(\hat{\eta}) - s \log n), \;\; s = 1, ..., d\}.$$

The rule $S1$ mimics the Schwarz BIC criterion. Since the penalty $s \log n$ is relatively heavy, $S1$ is well suited to detect low dimensional models `M(k)`. In contrast, the rule

$$A1 = \min\{1 \le k \le d : W_k(\hat{\eta}) - 2k \ge W_s(\hat{\eta}) - 2s, \;\; s = 1, ..., d\},$$

imitating the Akaike AIC criterion, is expected to work well when high dimensional disturbances `M(k)` of the null model `M(0)`: $Y = \beta[v(X)]^T + \epsilon$ are present. Based on our experience and some previous articles, the following "intermediate" solution was proposed and discussed in Inglot and Ledwina (2005). Use $A1$ when the distribution of the data at hand is very distinct from the null model and $S1$ otherwise. To provide a threshold defining which rule should be applied, we propose looking at the magnitude of the estimated standardized components of the efficient score vector. More precisely, in the present set-up, under the assumptions and notation of Proposition 1, set

$$(\mathcal{Y}_1, ...\mathcal{Y}_k) = \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\ell}^*(Z_i; \hat{\eta})\right] \hat{\mathbf{L}}^{1/2}.$$

Then, obviously, $W_k(\hat{\eta}) = ||(\mathcal{Y}_1, ..., \mathcal{Y}_k)||^2$. Following the discussion presented in Inglot and Ledwina (2005), we propose using the following penalty in this problem

$$\pi(s, n, c) = \begin{cases} s \log n, & \text{if } \max_{1 \le s \le d} |\mathcal{Y}_s| \le \sqrt{c \log n} \\ 2s, & \text{if } \max_{1 \le s \le d} |\mathcal{Y}_s| > \sqrt{c \log n}, \end{cases} \tag{3.12}$$

where c is some fixed positive number. This strategy leads to the following refined selection rule

$$T1 = \min\{1 \le k \le d : W_k(\hat{\eta}) - \pi(k, n, c) \ge W_s(\hat{\eta}) - \pi(s, n, c), \;\; s = 1, ..., d\}.$$

It is evident that small c's result in $T1$ being in practice equivalent to $A1$, while large c's lead to $T1$ being very similar to $S1$. "Moderate" values of c give a meaningful "switching effect".

For $n \ge 8$, $S1 \le T1 \le A1$. Moreover, since under the null model $(\mathcal{Y}_1, ..., \mathcal{Y}_k) \xrightarrow{\mathcal{D}} N(0, \mathbf{I}_k)$, then $P_\eta^n(T1 \ne S1) \to 0$ as $n \to \infty$. On the other hand, under H_0, for any $s \in \{2, ..., d\}$, $P_\eta^n(S1 = s) \le P_\eta^n(W_s(\hat{\eta}) \ge (s-1) \log n)$. Hence, Proposition 1 yields

PROPOSITION 2. *Under the null hypothesis $H_0 : Y = \beta[v(X)]^T + \epsilon$, the assumptions of Proposition 1 and $n \to \infty$, it holds that*

$$P_\eta^n(S1 > 1) \to 0, \quad W_{S1}(\hat{\eta}) \xrightarrow{\mathcal{D}} \chi_1^2,$$

and

$$P_\eta^n(T1 > 1) \to 0, \quad W_{T1}(\hat{\eta}) \xrightarrow{\mathcal{D}} \chi_1^2.$$

REMARK 3. We shall call $W_{S1}(\hat{\eta})$ and $W_{T1}(\hat{\eta})$ data driven score statistics for testing the validity of (1.1). We shall show in Section 4.4 that W_{T1} considerably extends the range of sensitivity of W_{S1}. Obviously, more general selection rules could be considered and incorporated into constructing data driven score statistics. However, as emphasized in Remark 2, our primary goal is to propose a practical solution. Therefore, we reduce the technical scope of the paper to its minimum. Finally, note that in our implementation we used numerical algorithm, based on the Schur decomposition, to calculate $\hat{\mathbf{L}}^{1/2}$.

4. Simulation study.

The aim of the simulations was to investigate the behaviour of W_{S1} and W_{T1} under H_0, as well as to compare the empirical powers of our tests to those of the recent solution in the field, which was proposed by Guerre and Lavergne (2005). We shall also present the empirical powers of the related Cramér-von Mises test, which represents a classical type of construction, still very often applied for testing goodness of fit. Practical implementation of W_{S1} and W_{T1} requires some specification of the estimators appearing in (3.6) and (3.9). So, we shall first discuss this point.

4.1. *Specification of estimators.* We define W_{S1} and W_{T1} in the following way. The sample splitting scheme and estimators $m_i^{(j)}$, $i,j=1,2$, were applied according to the description in Section 3.3. The remaining parameters were estimated on the basis of the jth part of the sample, $j=1,2$, as follows. The components of $\hat{\beta}^{(j)}$ were ordinary least square estimators. The discretization was neglected in the simulations. $\hat{\tau}^{(j)}$ was the adjusted Rice (1984) estimator [cf. Section A1]. We estimated f'/f by $[\widetilde{f}^{(j)}]'/\widetilde{f}^{(j)}$, where $\widetilde{f}^{(j)}$ is the kernel estimator of f defined as follows

$$\widetilde{f}^{(1)}(y)=\gamma_\zeta+\frac{1}{\zeta\hat{\alpha}_\zeta^{(1)}}\sum_{i\in<1>}K\left(\frac{y-\hat{\epsilon}_i^{(1)}}{\hat{\alpha}_\zeta^{(1)}}\right),\quad \widetilde{f}^{(2)}(y)=\gamma_\zeta+\frac{1}{\zeta\hat{\alpha}_\zeta^{(2)}}\sum_{i\in<2>}K\left(\frac{y-\hat{\epsilon}_i^{(2)}}{\hat{\alpha}_\zeta^{(2)}}\right),$$

where K is the standard Gaussian kernel, while $\gamma_\zeta=0.0001$, $\hat{\alpha}_\zeta^{(j)}=(0.9)[\hat{\tau}^{(j)}]^{1/2}\zeta^{-1/7}$, $\hat{\epsilon}_i^{(j)}=Y_i-v(X_i)[\hat{\beta}^{(j)}]^T$. To have some flexibility, we used a simple random bandwidth. Our choice was inspired by one of the simplest solutions in density estimation [cf. Silverman (1986), p. 45] and the result of Mammen and Park (1997), p. 338, on optimal bandwidth rate in shift models.

We estimated J in the first part of the sample by

$$\hat{J}^{(1)}=\frac{1}{\zeta}\sum_{c\in<1>}\left[\frac{\widetilde{f}'^{(2)}}{\widetilde{f}^{(2)}}\left(\hat{\epsilon}_c^{(2)}\right)\right]^2$$

and in the second part of the sample by the analogous expression. We estimated $\mathbf{W}$ by $\hat{\mathbf{W}}=(\hat{\mathbf{W}}^{(1)}+\hat{\mathbf{W}}^{(2)})/2$, where

$$\hat{\mathbf{W}}^{(1)}=\frac{1}{4}\left\{\hat{J}^{(1)}\frac{1}{\zeta}\sum_{i\in<1>}[\tilde{w}^{(1)}(X_i)]^T[\tilde{w}^{(1)}(Xi)]+\frac{1}{\hat{\tau}^{(1)}}[\hat{m}^{(1)}]^T\hat{m}^{(1)}\right\}$$

and $\hat{\mathbf{W}}^{(2)}$ is defined analogously. Finally, we considered the natural estimator $\hat{\mathbf{L}}$ of $\mathbf{L}$ given by $\hat{\mathbf{L}}=\frac{1}{4}(\hat{\mathbf{U}}-\hat{\mathbf{M}}^T\hat{\mathbf{V}}^{-1}\hat{\mathbf{M}})^{-1}$, where $\hat{\mathbf{M}}$, $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are blocks of $\hat{\mathbf{W}}$. Following our earlier considerations [cf. the discussion in Inglot and Ledwina (2005)], we took $c=2.4$ in (3.12). The

choice of u_i's and d is given in Section 4.2.

4.2. *Models used in the simulations.* The scheme of our study matches those used in the papers by Baraud et al. (2003), Diebolt and Zuber (2000), Guerre and Lavergne (2003, 2005), as well as Horowitz and Spokoiny (2001). We considerd the problem of testing

$$H_0 : \ Y = 1 + 2X + \epsilon.$$

To construct W_{S1} and W_{T1} we considered $d = 10$ auxiliary models `M(k)`, which we defined pertain to $u_i(x) = \cos([i+1]\pi x),\ i = 1, ..., 10$. Obviously, the user chooses of the supplementing system $\{u_i\},\ \ i \geq 1$. We find this to be an advantage of the procedure. A statistician can use information on the phenomenon under investigation to build apropriate and convenient class of alternative models. We decided to consider the cosine system in view of the competitors we shall investigate. The statistic of Guerre and Lavergne is based on equispaced partitions, while the Cramér-von Mises test is tightly linked to cosine functions. Therefore, such a choice gives conditions for fair comparison.

We consider ϵ obeying one of three laws: Gaussian with 0 mean and standard deviation σ [$G(\sigma)$ for short], Laplace with 0 mean and standard deviation $\sqrt{2}/\varphi$ [$L(\varphi)$ for short] and normal mixture $(0.7)\phi(x - \mu/(0.7)) + (0.3)\phi(x + \mu/(0.3))$ [denoted $NM(\mu)$ in what follows], where ϕ is the $N(0, 1)$ density function.

X was assumed to be independent of ϵ and from a beta distribution on [0,1]. Since changing, to some reasonable extent, the parameters of the beta distribution had no essential influence on the general picture, we restricted the presentation of results to the case where X is uniformly distributed.

The alternatives were defined by disturbing the pattern $1 + 2x$ [with each type of error: $G(\sigma),\ \ L(\varphi),\ \ NM(\mu)$] by one of the functions $r_l(x),\ \ l = 1, ..., 6$, where

$$r_1(x) = c \times \cos(o\pi x),\ \ c \in R,\ \ o = 2, 3, ...$$

$$r_2(x) = c \times L_s(x),\ \ c \in R,\ \ s = 2, 3, ...\ \ L_s - \text{ sth normalized Legendre polynomial on } [0, 1],$$

$$r_3(x) = c \times \frac{1}{t}\phi\left(\frac{x - 0.5}{t}\right),\ \ c \in R,\ \ t \in R_+,\ \ \phi - \text{the } N(0, 1) \text{ density function},$$

$$r_4(x) = c \times (x - a)\mathbf{1}_{[a,1]}(x),\ \ c \in R,\ \ a \in (0, 1),$$

$$r_5(x) = c \times \text{arctg}[b(2x - 1)],\ \ c \in R,\ \ b \in (0, \infty),$$

$$r_6(x) = c \times \max\{\min\{(2x - 1)/(1 - 2a), 1\}, -1\},\ \ c \in R,\ \ a \in (0, 1/2).$$

The disturbance r_1 was considerd by Guerre and Lavergne (2003, 2005), r_2 by Diebolt and Zuber (2000), r_3 was used by Horowitz and Spokoiny (2001), while r_4 by Baraud et al. (2003). We added the functions r_5 and r_6 to the above list of disturbances to cover some cases of three-phase regression.

4.3. *Empirical behaviour of test statistics under H_0.* All simulation experiments presented in the paper were done for the same sample size $n = 300$ and $N = 10000$ Monte Carlo runs. Throughout we considered tests at the significance level $\alpha = 0.05$.

Let us start our discussion with some remarks on the behaviour of W_{S1} and W_{T1} under H_0. The asymptotic critical value of W_{S1} and W_{T1} is 3.841. In order to illustrate how the asymptotic theory works in the case of our implementation, Table 1 presents simulated critical values of W_{S1} and W_{T1} under different error distributions.

TABLE 1

Simulated critical values of W_{S1}, W_{T1} and CvM under the null model $Y = 1 + 2X + \epsilon$ with X uniform on [0,1] and different errors. Sample size $n = 300$. 5% significance level, $N = 10000$ MC runs.

Error distribution	**Parameter**	**Variance**	**Critical values**		
			W_{S1}	W_{T1}	**CvM**
$G(\sigma)$	0.25	0.063	5.91	6.11	27.88
	0.50	0.250	5.63	5.92	7.00
	0.75	0.563	5.83	6.04	3.22
	1.00	1.000	5.79	6.02	1.73
$L(\varphi)$	4.00	0.125	5.29	5.57	15.72
	2.00	0.500	5.27	5.50	3.86
	1.00	2.000	5.75	5.93	0.94
	0.50	8.000	5.61	5.82	0.23
$NM(\mu)$	0.20	1.191	5.94	6.08	1.52
	0.40	1.762	5.67	6.00	0.97
	0.60	2.714	5.81	6.05	0.63
	0.80	4.048	5.66	5.85	0.43

As seen, an evident feature of the new procedures is that the critical values are very stable to changes of the error distributions and their parameters. Since the penalty in the selection rule $T1$ is slightly smaller, the respective critical values of the test W_{T1} are slightly larger. We would like to emphasize that critical values are also very stable with respect to the choice of d. Any choice of $d \geq 4$ gives practically the same simulated critical value. This follows from the fact that, under the null model and $n = 300$, in all the cases we investigated, the proportion of cases with $\{S1 = 1\}$ and $\{T1 = 1\}$ is in [0.97, 0.98] and [0.96, 0.97], respectively, and the remaining mass is mostly concentrated on dimensions 2 - 3. On the other hand, the simulated critical values are larger than the limiting values. This is a characteristic phenomenon for data driven tests, which was discussed in detail in some earlier papers. We would like to recall the basic reason for this phenomenon. As said above, in some small percentage of cases the selection rules take values greater than 1, which makes the test statistic stochastically larger than the limiting χ^2_1 random variable. For some classical testing problems we developed nicely working approximations, which can be used to estimate p - values. For some evidence see e.g. Kallenberg and Ledwina (1995, 1997b). In the present set-up, to provide a practical way of generating critical values, one can apply the residual bootstrap, described e.g. on pp. 142 - 143 of Stute et al. (1998b). We implemented this procedure in our simulation study and found that it works well. We conducted several trials with $B = 400$ bootstrap replications and $N = 10000$ Monte Carlo [MC] runs, observing that the resulting critical values are, on average throughout the $N = 10000$ repetitions, very stable, slightly overestimating the "actual" critical values given in Table 1. It was verified that this overestimation has only a negligible effect on the empirical powers. More precisely, we compared the resulting powers with those obtained by simulations when the "actual" critical values, as well as the average of the twelve "actual" critical values, 5.68 for W_{S1} and 5.91 for W_{T1}, are used. The conclussion was that, in many cases there is no difference, in some cases the powers differ by 1% - 3%. Therefore, we present simulated powers for W_{S1} and W_{T1} in the case the averaged critical values 5.68 and 5.91 are used. This also allows us to some further comparisons with simulation results presented in Inglot and Ledwina (2003b). A formal proof of the consistency of the residual bootstrap in our implementation is beyond the scope of the present paper. Finally, we would like to emphasize that the stability

of critical values of the data driven tests with respect to the choice of d allows us to choose practically arbitrary $d \geq 4$. Enlarging d does not spoil empirical powers achieved for choices of smaller d's. Therefore, reasonable choice of d only depends on two factors: how complicated alternatives one likes to detect and how much time consuming calculations are reasonable in this context.

In our implementation of Guerre and Lavergne's (2003, 2005) solution we took binwidths in $\{2^{-2}, 2^{-3}, ..., 2^{-7}\}$, $c = 1.5$, $J_n = 6$ and used the adjusted Rice estimator for the variance of errors [cf. p. 17 in Guerre and Lavergne (2003)]. They noticed that the normal approximation to the distribution of their statistic under the null hypothesis may not be very accurate for finite samples. Indeed, we did some experiments and observed, under various error distributions, such extreme simulated critical values as e.g. 0.11 and 40.50. Therefore, in our simulation study we followed their prescription and applied the wild bootstrap with the two-point distribution for w_i given on p. 17 of Guerre and Lavergne (2003). We did $B = 400$ bootstrap replications and $N = 10000$ MC runs in each experiment. For simplicity, we shall denote the test statistic introduced by Guerre and Lavergne (2003, 2005) by $\hat{T}$.

To complete the picture, we also investigated the empirical behaviour of a transformed Cramér- von Mises statistic, which was developed in Stute et al. (1998a). We shall denote this statistic by CvM. This transformation was introduced by Khmaladze (1981) to remove the influence on nuisance parameters on the null distribution. The simulations reported in Table 1

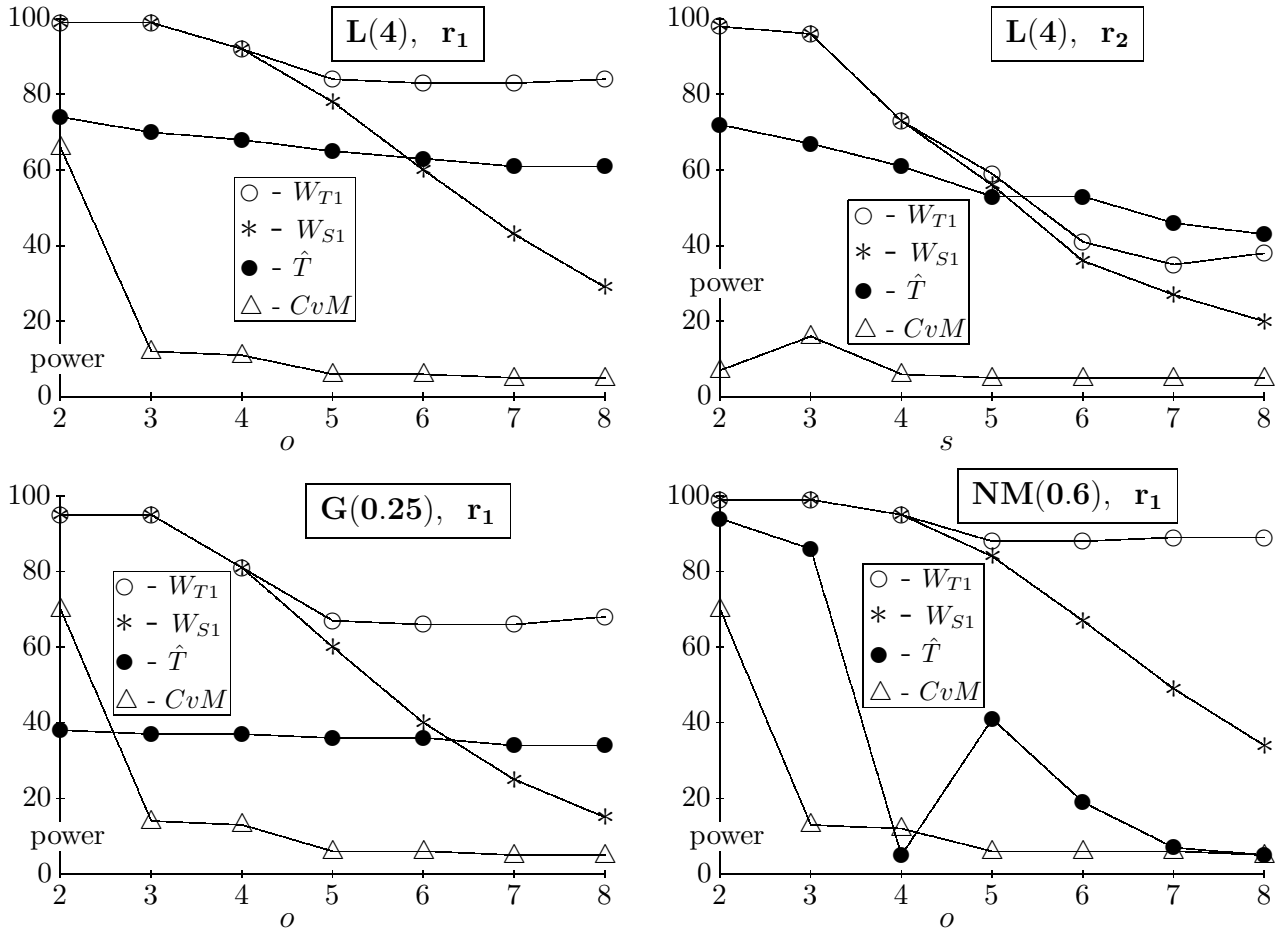


FIG. 1. *Simulated powers of tests based on W_{T1}, W_{S1}, $\hat{T}$ and CvM under the alternatives $Y = 1 + 2X + r_l(X) + \epsilon$, $l = 1, 2$, X uniform on [0,1] and different errors. Signal/noise 0.25. 5% nominal level, n=300, N=10000 MC runs.*

show that the simulated critical values of CvM are highly unstable. A similar observation was made earlier in Diebolt and Zuber (2000) and can be inferred from Koenker and Xiao (2001). As demonstrated by Stute et al. (1998b), the wild bootstrap and residual bootstrap are much better suited to provide distributional feasibility of untransformed Cramér- von Mises statistic for model check in homoscedastic regression.

4.4. *Empirical powers.* As concluded in Section 4.3, in order to simulate powers of W_{S1} and W_{T1}, $\hat{T}$, as well as CvM, we used the averaged, bootstrap and "actual" critical values, respectively.

A representative selection of simulation results is presented in Figures 1 and 2. In Figure 1 we present results of experiments which serve to understand the behaviour of test statistics when alternatives are oscillating, i.e. r_1 and r_2, given in Section 4.2, are taken into account. In all four cases considered there, the ratio signal/noise $= c||r_l||_2/\sqrt{\mathrm{Var}\epsilon}$, where $||\bullet||_2$ denotes the $L_2[0,1]$ norm, equals 0.25. Figure 2 exhibits the behaviour of tests under more "smooth" deviations i.e. the disturbances r_3 - r_6.

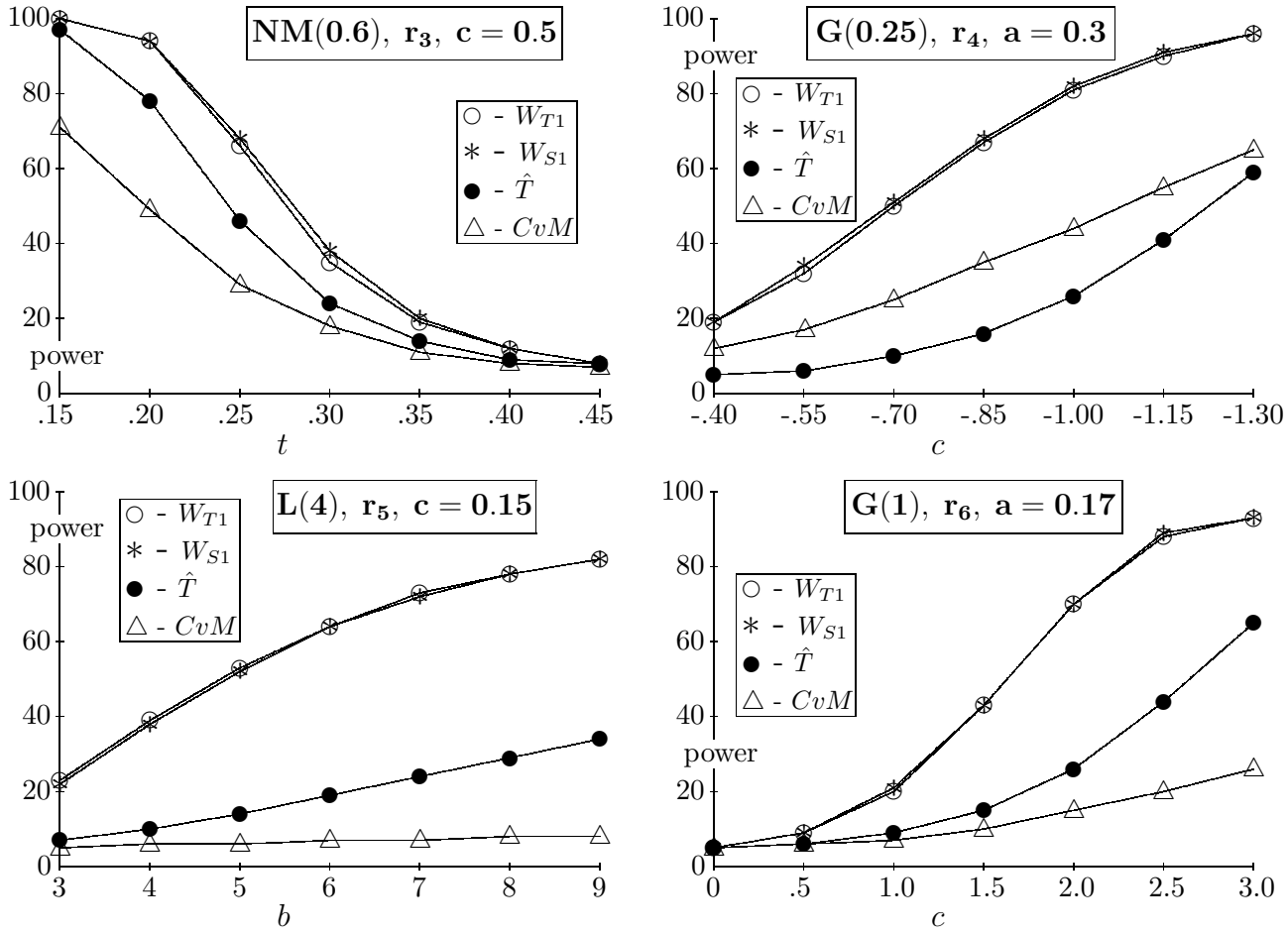


FIG 2. *Simulated powers of tests based on W_{T1}, W_{S1}, $\hat{T}$ and CvM under the alternatives $Y = 1 + 2X + r_l(X) + \epsilon$, $l = 3, 4, 5, 6$, X uniform on [0,1] and different errors. 5% nominal level, $n = 300$, $N = 10000$ MC runs.*

The simulation results confirm what might have been expected from our earlier experience.

As characteristic to data-driven Neyman tests with a Schwarz penalty, W_{S1} is powerful for "smooth" deviations from linear regression, while the minimax data-driven chi-square-type statistic $\hat{T}$ of Guerre and Lavergne (2005) is more powerful for some extreme deviations, [such

as highly oscillating alternatives, in particular]. However, Figure 1 [see the case $NM(0.6)$, r_1] shows that under difficult conditions [large variance, asymmetrical bimodal error density] the $\hat{T}$ test has some difficulty in detecting high frequency oscillations. Under larger c and/or smaller μ this drawback disappears. We also observed that $\hat{T}$ loses a lot of its power when the variance of the model error is small [cf. Figure 1, $G(0.25), r_1$] or the model is very close to the null model [cf. Figure 2]. Some more simulations on W_{S1} and $\hat{T}$ can be found in Inglot and Ledwina (2003b). Note also that in Guerre and Lavergne (2003) it is shown that $\hat{T}$ compares favourably with the solution of Horovitz and Spokoiny (2001). The refined selection rule $T1$ works very well. In comparison to W_{S1}, one observes only a slight decrease in power of W_{T1} under low dimensional deviations and, simultaneously, very substantial gain in power under high dimensional alternatives.

In all the cases considered, except $L(4)$ and r_2 with $o = 6,\ 7,\ 8$, the power of W_{T1} is larger than that of $\hat{T}$ and in many cases powers dramatically differ in favour of W_{T1}. At first glance, such big differences in behaviour of $\hat{T}$, on one hand, and W_{S1} and W_{T1}, on the other hand, might be considered as somewhat surprising. One could argue, loosely speaking, that the structure of the three solutions is similar: one has a quadratic statistic with the dimension fitted by the associated criterion. However, the quadratic statistic $W_k(\hat{\eta})$, which we derived, exploits the information contained in the data and the structure of the alternative models more efficiently. In contrast, $\hat{T}$ compares, in some way, a parametric and nonparametric fit. Moreover, other ingredients built into the constructions [i.e. the system of functions and the penalties] play some role. In this respect, the situation is to a large extent similar to a corresponding one in which data driven chi-square tests and data driven smooth tests for uniformity are compared. Therefore, we refer the reader to related discussions in Inglot et al. (1994), Bogdan (1995), Inglot and Janic-Wróblewska (2003) and references therein.

The behaviour of CvM is unsatisfactory. Obviously, the poor power of CvM test is not due to the transformation, but follows from the nature of such a statistic. It is known that, under resonable sample sizes, the Cramér-von Mises test is only capable of detecting very smooth deviations from the null model. Various aspects of this drawback are discussed e.g. in Inglot and Ledwina (2001a, 2004).

In view of the above, it seems that W_{T1} can be recommended as a stable and powerful tool for checking validity of (1.1).

5. Proof of Theorem 1. Obviously, it is enough to show (3.7) for $i \in < j >,\ \ j = 1, 2$. Therefore, we shall restrict attention to the case $j = 1$ and prove that

$$\bigwedge_{\delta>0} P_\eta^n \left(\frac{1}{\sqrt{n}} \left\| \sum_{i\in<1>} [\hat{\ell}^*(Z_i;\hat{\eta}) - \ell^*(Z_i;\eta)] \right\| \geq \delta \right) \to 0 \ \text{ as } \ n \to \infty. \tag{5.1}$$

To facilitate reading, recall that $\vartheta = (\sqrt{g}, \sqrt{f})$ and $\eta = (\beta, \vartheta)$ and concisely denote the estimate of ϑ by $\hat{\vartheta}$. We also introduce the class $\mathcal{B}(\beta)$ of deterministic sequences $\{b_n\},\ \ b_n \in R^q$, such that $\sqrt{n}(b_n - \beta)$ stays bounded.

The proof consists of four basic steps.

- The discretization allows us to replace checking (5.1) by proving that for any $\{b_n\} \in \mathcal{B}(\beta)$ it holds that

$$\bigwedge_{\delta>0} P_\eta^n \left(\frac{1}{\sqrt{n}} \left\| \sum_{i\in<1>} [\hat{\ell}^*(Z_i;b_n,\hat{\vartheta}) - \ell^*(Z_i;\beta,\vartheta)] \right\| \geq \delta \right) \to 0 \ \text{ as } \ n \to \infty. \tag{5.2}$$

- From Schick's (2001) results we infer that for any $\{b_n\} \in \mathcal{B}(\beta)$

$$\bigwedge_{\delta>0} P_\eta^n \left(\frac{1}{\sqrt{n}} \left|\left| \sum_{i\in<1>} [\ell^*(Z_i;\beta,\vartheta) - \ell^*(Z_i;b_n,\vartheta)] \right|\right| \geq \delta \right) \to 0 \text{ as } n \to \infty, \tag{5.3}$$

and therefore, to prove (5.2), it is enough to show that for any $\{b_n\} \in \mathcal{B}(\beta)$ and $\eta = (\beta, \vartheta)$

$$\bigwedge_{\delta>0} P_{(\beta,\vartheta)}^n \left(\frac{1}{\sqrt{n}} \left|\left| \sum_{i\in<1>} [\hat{\ell}^*(Z_i;b_n,\hat{\vartheta}) - \ell^*(Z_i;b_n,\vartheta)] \right|\right| \geq \delta \right) \to 0 \text{ as } n \to \infty. \tag{5.4}$$

- The contiguity of $\{P_{(\beta,\vartheta)}^n\}$ and $\{P_{(b_n,\vartheta)}^n\}$, where $\{b_n\} \in \mathcal{B}(\beta)$, allows us to replace (5.4) by

$$\bigwedge_{\delta>0} P_{(b_n,\vartheta)}^n \left(\frac{1}{\sqrt{n}} \left|\left| \sum_{i\in<1>} [\hat{\ell}^*(Z_i;b_n,\hat{\vartheta}) - \ell^*(Z_i;b_n,\vartheta)] \right|\right| \geq \delta \right) \to 0 \text{ as } n \to \infty. \tag{5.5}$$

- Checking (5.5) is simplified by introducing some conditioning related to the sample splitting scheme. Under this conditioning, the structure of the model and the choice of estimators are exploited. In particular, the structure of the model with the shift $v(X_i)b_n^T$, as well as the choice of $\hat{m}_i^{(j)}$, $i, j = 1, 2$ are essential to get the final result.

Some details are given below.

We start with some brief comments on the discretization. Suppose R^q is covered by cubes with edges of length $2n_0/\sqrt{n}$, where n_0 is a given natural number. The discretized version $\hat{\beta}_*^{(2)}$ of $\hat{\beta}^{(2)}$ is defined as the center of the cube into which $\hat{\beta}^{(2)}$ has fallen [with some additional rule for the boundaries of cubes]. The crucial property of the discretized estimator $\hat{\beta}_*^{(2)}$ is the following one: given $\gamma > 0$, there exists M_γ such that for the set $B_n = \{\sqrt{n}||\hat{\beta}_*^{(2)} - \beta|| < M_\gamma\}$ it follows that $P_\eta^n(B_n) > 1 - \gamma$ and on the set B_n the estimator $\hat{\beta}_*^{(2)}$ takes only a finite number of values, which depend on M_γ solely. The discretization trick was introduced by Le Cam (1956). For an insightful exposition see Bickel et al. (1993), p. 44, or Kreiss (1987), p. 120. The application is immediate and therefore we skip the details.

To get (5.3), it is enough to show that

$$H_n(\beta) = \frac{1}{\sqrt{\zeta}} \sum_{i=1}^{\zeta} \ell^*(Z_i;\beta,\vartheta)$$

with $\beta \in R^q$ and the other parameters fixed, but otherwise arbitrary, is asymptotically differentiable at β with the matrix $\mathbf{D}_\beta = 0$. Indeed, the definition of asymptotic differentiability [cf. Schick (2001), p. 15] and the definition $\zeta = [n/2]$ immediately yield that for

$$R_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{\zeta} [\ell^*(Z_i;\beta,\vartheta) - \ell^*(Z_i;b_n,\vartheta)]$$

it follows that $P_\eta^n(||R_n|| \geq \delta) \to 0$ as $n \to \infty$, which is the desired result.

To check the asymptotic differentiability of $H_n(\beta)$ we shall apply Theorem 2.3 from Schick (2001). First note that from the results of Section 7 of this paper, it follows that the null model density $g(x)f(y - v(x)\beta^T)$ has Hellinger derivative [κ_β in Schick's notation] of the form

$-v(x)\frac{f'}{f}(y - v(x)\beta^T)$ [cf. (7.5)]. As $\ell^*(z;\beta,\vartheta)$, by definition, is [under P_η] orthogonal to the scores for the nuisance parameters, we immediately obtain

$$\mathbf{D}_\beta = \int_R \int_I [v(x)]^T \ell^*(x,y,\beta;\vartheta) \left[\frac{f'}{f}(y - v(x)\beta^T)\right] g(x) f(y - v(x)\beta^T) dxdy = 0.$$

Here and throughout the remaining part of the proof we abbreviate $\lambda(dx)$ and $\lambda(dy)$ to dx, dy.

Therefore, it remains to show that the assumption (2.1) of Theorem 2.3 is fulfilled. In the problem considered, (2.1) reads as

$$\lim_{\tilde\beta \to \beta} r_n(\tilde\beta, \beta) = 0,$$

where

$$r_n(\tilde\beta,\beta) = \int_R \int_I \left|\left| \ell^*(x,y;\tilde\beta,\vartheta)\sqrt{f(y - v(x)\tilde\beta^T)g(x)} - \ell^*(x,y;\beta,\vartheta)\sqrt{f(y - v(x)\beta^T)g(x)} \right|\right|^2 dxdy.$$

The definition of ℓ^* [cf. (3.1)] and a change of variables in the integral yield the bound

$$r_n(\tilde\beta,\beta) \le 4\int_R \int_I \left[\frac{f'}{\sqrt{f}}\left(y - v(x)(\tilde\beta - \beta)^T\right) - \frac{f'}{\sqrt{f}}(y)\right]^2 ||\tilde u(x) - \tilde v(x)\mathbf{V}^{-1}\mathbf{M}||^2 g(x)dxdy +$$

$$4\int_R \int_I \frac{1}{\tau^2}\left[\left(y - v(x)(\tilde\beta - \beta)^T\right)\sqrt{f(y - v(x)(\tilde\beta - \beta)^T)} - y\sqrt{f(y)}\right]^2 ||m_1 - m_2\mathbf{V}^{-1}\mathbf{M}||^2 g(x)dxdy.$$

It follows from $< M1 >$ and $< M2 >$ that the functions $f'/\sqrt{f}$ and $y\sqrt{f(y)}$ are from $L_2(R,\lambda)$. Obviously, $\sqrt{g(x)} \in L_2(I,\lambda)$. Therefore, setting $t = ||\tilde\beta - \beta||$ and $\varphi_t(x) = v(x)(\tilde\beta - \beta)^T/||\tilde\beta - \beta||$ we see that (iii) of Section 6 of this paper is satisfied. This concludes the proof of (5.3).

As mentioned above, from Section 7 it follows that the model density $p(z;\beta,\vartheta)$ is Hellinger differentiable at β [cf. Sec. 2 of Schick (2001) for the terminology]. Therefore, by Lemma 2.3 in Schick (1997), the sequences of product measures $\{P^n_{(\beta,\vartheta)}\}$ and $\{P^n_{(b_n,\vartheta)}\}$, where $\{b_n\} \in \mathcal{B}(\beta)$, are indeed mutually contiguous. This implies that the proof reduces to proving (5.5).

Let us now rewrite (5.5) in a more convenient form. For this purpose let us set

$$T_{n,s} = \frac{1}{\sqrt{n}} \sum_{i \in <1>} \left[\hat\ell^*_s(Z_i; b_n, \hat\vartheta) - \ell^*_s(Z_i; b_n, \vartheta)\right], \tag{5.6}$$

where the symbol ν_s denotes the sth component of a k - dimensional vector. Using this notation, (5.5) reads as

$$T_{n,s} = o_{P^n_{(b_n,\vartheta)}}(1), \quad \text{for each} \quad s = 1,...,k \quad \text{and} \quad \{b_n\} \in \mathcal{B}(\beta). \tag{5.7}$$

To check (5.7) we shall apply the following result, which is proved for completeness in Appendix B.

PROPOSITION 3. *Suppose for each $n \ge 1$, T_n is a random variable defined on a probability space $(\mathcal{T}_n, \mathcal{B}_n, P_n)$, $E_{P_n}|T_n| < \infty$, $n \ge 1$. Let $\mathcal{F}_n$ be a sub-σ-field of $\mathcal{B}_n$. If $E\left(|T_n|\big|\mathcal{F}_n\right) \overset{P_n}{\to} 0$ then $P_n(|T_n| > \delta) \to 0$ for every $\delta > 0$.*

This proposition shall be applied to each $T_{n,s}$, $s = 1, ..., k$. We shall take $\mathcal{T}_n = R^{2n}$, $\mathcal{B}_n$ - the Borel σ-field in R^{2n}, $P_n = P^n_{(b_n,\vartheta)}$ and $\mathcal{F}_n = \sigma(X_1, ..., X_n, Y_{\zeta+1}, ..., Y_n)$.

We shall first prove that (everywhere)

$$E\left(T_{n,s}\,|\mathcal{F}_n\right) = 0, \quad s = 1, ..., k. \tag{5.8}$$

Since the conditional density of $(Y_1, ..., Y_\zeta)$ under $\mathcal{F}_n$ is of the form $\prod_{l=1}^{\zeta} f(y_l - v(X_l)b_n^T)$, we have for $T_n = (T_{n,1}, ..., T_{n,k})$

$$E\left(T_n\,|\mathcal{F}_n\right) = \frac{1}{\sqrt{n}} \sum_{i\in<1>} \int_R \left[\hat{\ell}^*(X_i, y; b_n, \hat{\vartheta}) - \ell^*(X_i, y; b_n, \vartheta)\right] f(y - v(X_i)b_n^T)dy. \tag{5.9}$$

Moreover, from $\int_R f'(y)dy = \int_R yf(y)dy = 0$ we get

$$\int_R \ell^*(X_i, y; b_n, \vartheta) f(y - v(X_i)b_n^T)dy = 0.$$

This, a change of variables in the integral (5.9) and another application of $\int_R yf(y)dy = 0$ yield

$$\sqrt{n}E\left(T_n\,|\mathcal{F}_n\right) =$$

$$\left[\int_R \left\{-\widehat{[f'/f]}^{(2)}(y)\right\} f(y)dy\right] \times \left[\sum_{i\in<1>} u(X_i) - \zeta\hat{m}_1^{(1)} - \left\{\sum_{i\in<1>} v(X_i) - \zeta\hat{m}_2^{(1)}\right\} [\hat{\mathbf{V}}^{(2)}]^{-1}\hat{\mathbf{M}}^{(2)}\right].$$

Since, however, $\hat{m}_1^{(1)} = \frac{1}{\zeta}\sum_{i\in<1>} u(X_i)$ and $\hat{m}_2^{(1)} = \frac{1}{\zeta}\sum_{i\in<1>} v(X_i)$ we infer that $E(T_n|\mathcal{F}_n) = 0$ everywhere. This proves (5.8).

Therefore, $E(T_{n,s}^2|\mathcal{F}_n) = \mathrm{Var}(T_{n,s}|\mathcal{F}_n)$ and, by Proposition 3, to get (5.7) it is enough to check that

$$\bigwedge_{\delta>0} \quad P^n_{(b_n,\vartheta)}\left(\mathrm{Var}\left(T_{n,s}\,|\mathcal{F}_n\right) > \delta\right) \to 0 \ \text{ as } n \to \infty. \tag{5.10}$$

However, notice again that, as previously, under $P^n_{(b_n,\vartheta)}$ the conditional density of $(Y_1, ..., Y_\zeta)$ with respect to $\mathcal{F}_n$ is of the form $\prod_{l=1}^{\zeta} f(y_l - v(X_l)b_n^T)$. Hence, from (3.8), the conditional variance of $T_{n,s}$ equals

$$\frac{1}{n}\sum_{i\in<1>} \int_R \left[\hat{\ell}_s^*(X_i, y; b_n, \hat{\vartheta}) - \ell_s^*(X_i, y; b_n, \vartheta)\right]^2 f(y - v(X_i)b_n^T)dy.$$

In consequence, after changing the variables in the above integral, (5.10) reads as

$$\bigwedge_{\delta>0} \quad P^n_{(b_n,\vartheta)}\left(\frac{1}{n}\sum_{i\in<1>} \int_R \left[\hat{\ell}_s^*(X_i, y; 0, \hat{\vartheta}) - \ell_s^*(X_i, y; 0, \vartheta)\right]^2 f(y)dy > \delta\right) \to 0.$$

Now, the artificial b_n is no longer useful and applying the contiguity argument again, we see that to prove (5.10) it is enough to show that

$$\bigwedge_{\delta>0} \quad P^n_{\eta}\left(\frac{1}{n}\sum_{i\in<1>} \int_R \left[\hat{\ell}_s^*(X_i, y; 0, \hat{\vartheta}) - \ell_s^*(X_i, y; 0, \vartheta)\right]^2 f(y)dy > \delta\right) \to 0. \tag{5.11}$$

The rest of the proof consists of showing that each summand appearing in (5.11) is bounded by a common [independent of i] quantity which tends to 0 in probability with respect to P^n_η.

We have

$$\hat{\ell}_s^*(X_i, y; 0, \hat{\vartheta}) - \ell_s^*(X_i, y; 0, \vartheta) =$$

$$= y\left(\frac{1}{\hat{\tau}^{(2)}} - \frac{1}{\tau}\right)(m_{1s} - \{m_2\mathbf{V}^{-1}\mathbf{M}\}_s) + \tag{5.12}$$

$$+\frac{y}{\hat{\tau}^{(2)}}\left(\hat{m}_{1s}^{(1)} - m_{1s} - \{\hat{m}_2^{(1)}[\hat{\mathbf{V}}^{(2)}]^{-1}\hat{\mathbf{M}}^{(2)}\}_s + \{m_2\mathbf{V}^{-1}\mathbf{M}\}_s\right) + \tag{5.13}$$

$$-\left[\frac{f'}{f}(y)\right]\left(\tilde{u}_s^{(1)}(X_i) - \{\tilde{v}^{(1)}(X_i)[\hat{\mathbf{V}}^{(2)}]^{-1}\hat{\mathbf{M}}^{(2)}\}_s - \tilde{u}_s(X_i) + \{\tilde{v}(X_i)\mathbf{V}^{-1}\mathbf{M}\}_s\right) + \tag{5.14}$$

$$-\left(\widehat{[f'/f]}^{(2)}(y) - [f'/f](y)\right)\left(\tilde{u}_s^{(1)}(X_i) - \{\tilde{v}^{(1)}(X_i)[\hat{\mathbf{V}}^{(2)}]^{-1}\hat{\mathbf{M}}^{(2)}\}_s\right). \tag{5.15}$$

Now we shall consider the integrals of the squared terms from (5.12) - (5.15). Denote by Π_f the distribution on R with density f with respect to λ.

As $\int_R y^2 f(y)dy = \tau \in (0, \infty)$ and $\hat{\tau}^{(2)}$ is a consistent estimator of τ, we infer that the $L_2(R, \Pi_f)$ norm of (5.12) tends to 0 in probability. In addition, exploiting the consistency of $\hat{m}_1^{(1)}$, $\hat{m}_2^{(1)}$, $\hat{\mathbf{V}}^{(2)}$, $\hat{\mathbf{M}}^{(2)}$ the same conclussion follows for (5.13).

Now, rewrite (5.14) as follows

$$-\left[\frac{f'}{f}(y)\right]\left[(\hat{m}_{1s}^{(1)} - m_{1s}) - \left\{v(X_i)\left([\hat{\mathbf{V}}^{(2)}]^{-1}\hat{\mathbf{M}}^{(2)} - \mathbf{V}^{-1}\mathbf{M}\right)\right\}_s + \right.$$

$$\left. + \left\{\hat{m}_2^{(1)}[\hat{\mathbf{V}}^{(2)}]^{-1}\hat{\mathbf{M}}^{(2)} - m_2\mathbf{V}^{-1}\mathbf{M}\right\}_s\right].$$

As $J = J(f) < \infty$ and $\sup_x ||v(x)|| < \infty$, the consistency of the involved estimators implies the required convergence to 0.

Finally, consider (5.15). Estimating $|\tilde{u}_s^{(1)}(X_i)|$ by $2\sup_x ||u(x)||$ and treating $\tilde{v}^{(1)}(X_i)$ analogously, we see that the consistency of $\hat{\mathbf{V}}^{(2)}$ and $\hat{\mathbf{M}}^{(2)}$ reduces the problem to showing that the $L_2(R, \Pi_f)$ norm of $\left(\widehat{[f'/f]}^{(2)} - f'/f\right)$ tends to 0 in P_η^n. However, this is just our assumption (3.11).

To close the proof, note that when applying Proposition 3 to the case $j = 2$, it is convenient to take $\mathcal{F}_n = \sigma(X_1, ..., X_n, Y_1, ..., Y_\zeta)$. The rest of the argument is identical. □

Remark 4. The proof of Theorem 1 shows that to get the key result $W_k(\hat{\eta}) \xrightarrow{\mathcal{D}} \chi_k^2$, using several steps, the problem can be reduced to checking the following conditions

$$(*) \qquad \sum_{i \in <j>} \int_R \hat{\ell}^*(X_i, y; b_n, \hat{\vartheta}) f(y - v(X_i)b_n^T)\lambda(dy) = 0,$$

and

$$(**) \qquad \frac{1}{n}\sum_{i \in <j>} \int_R \left|\left|\hat{\ell}^*(X_i, y; b_n, \hat{\vartheta}) - \ell^*(X_i, y; b_n, \vartheta)\right|\right|^2 f(y - v(X_i)b_n^T)\lambda(dy) = o_{P^n_{(b_n,\vartheta)}}(1),$$

for $j = 1, 2$ and every sequence $\{b_n\} \in \mathcal{B}(\beta)$. In the problem considered, the conditions $(*)$ and $(**)$ play the role of handy counterparts of (i) and (ii) in the basic proposition on p. 854 of Choi et al. (1996).

6. Discussion of the general assumptions. The assumption on the compactly supported density g was imposed for technical convenience. It could be removed when assuming

some extra conditions on the tails of g. The restriction $g > 0, \lambda$ - a.e. guarantees that the matrix $\mathbf{W}$ is positive definite. Obviously, the assumption is not necessary.

Extensions to multivariate explanatory variable seems to be rather straightforward.

The assumption that the lenght d of the list of models is independent on n substantially simplifies the considerations. From the practical point of view, fixing this number seems to be reasonable. Recall that the critical values of our tests are stable with respect to the choice of d and enlarging d does not spoil empirical powers achieved for choices of smaller d's [cf. our discussion in Section 4.3]. We would like to quote here also the general opinion phrased by Bickel and Kwon (2001), p. 948, which supports our view point : "When considering nonparametric alternatives however, as Bickel, Ritov and Stoker (2001) point out, it may be important to tailor tests to directions which a priori appear important and save some power for grossly divergent alternatives in other directions, rather than have negligible power in all directions". Introducing $d = d(n),\ d(n) \to \infty$ as $n \to \infty$, has however some aesthetical aspect. Namely, in price of some fine technical work one can get then consistency of the related data driven test for essentially any alternative. Such a program, in case of Euclidean nuisance parameters, was elaborated in details in Inglot, Kallenberg and Ledwina (1997).

7. Generalized shift operators and the efficient score vector. The degree to which efficient estimation is developed is well illustrated by the fact that nowadays many proofs and derivations are not published. For example, the efficient score vector for a complicated regression problem is introduced in Schick (1997), p. 375, as follows: "define a map". This is not very instructive, especially if e.g. one likes to do some modifications. Therefore we rederived "some maps" in Inglot and Ledwina (2003a). In the course of the work we observed that it would be useful to generalize some standard results of Hájek and Šidák (1967) and simplify some traditional calculations in this way. Therefore, we briefly comment here on our observations.

Consider the model

$\mathtt{M(k)}$ $$Y = u(X)\theta^T + v(X)\beta^T + \epsilon$$

and define

$$w(x) = (u(x), v(x)),\ \ a = (\theta, \beta),\ \ \eta = (\beta, \sqrt{g}, \sqrt{f}),\ \ \kappa = (\theta, \eta).$$

Under this notation set $p(z;\kappa)$ to be the density of $Z = (X, Y)$ under $\mathtt{M(k)}$. We have

$$p(z;\kappa) = g(x)f(y - w(x)a^T). \tag{7.1}$$

Observe that $p^{1/2}(z;\kappa)$, seen as a function of κ, is a map from $\Omega \to \mathcal{H}$, where $\Omega = \mathcal{A} \times \mathcal{B} \times \mathcal{C}$, while $\mathcal{A} = R^{k+q},\ \mathcal{B} = L_2(I, \lambda),\ \mathcal{C} = L_2(R, \lambda),\ \mathcal{H} = L_2(I \times R, \lambda \times \lambda)$. The specific structure of $p^{1/2}(z;\kappa)$ [cf. (7.1)] motivates the introduction of an abstract map

$$\Phi : \Omega \to \mathcal{H},\ \ \Phi(\omega) = \Phi(a, b, c) = \Delta_{wa^T}(bc),$$

where for an arbitrary measurable function φ on I we define $\Delta_\varphi : \mathcal{H} \to \mathcal{H}$ by

$$\Delta_\varphi h(x, y) = h(x, y - \varphi(x)),\ \ x \in I,\ y \in R,\ h \in \mathcal{H}. \tag{7.2}$$

It can be shown that Φ is Hadamard differentiable at each point $\omega = (a, b, c)$ such that c is differentiable for every $y \in R$ and $\int_R [c']^2 d\lambda < \infty$. Moreover, for any $(a_0, b_0, c_0) \in \Omega$ the following holds

$$\dot{\Phi}_{(a,b,c)}(a_0, b_0, c_0) = \Delta_{wa^T}(-bc'[wa_0^T] + b_0 c + bc_0) \tag{7.3}$$

[cf. Theorem B.11 in Inglot and Ledwina (2003a)]. The result was derived by exploiting the chain rule for Hadamard differentiability and the following basic properties of the shift operator Δ_φ.

For any arbitrary measurable φ defined on I
(i) Δ_φ is an isometry on $\mathcal{H}$,
(ii) for each $h \in \mathcal{H}$ the following holds

$$\lim_{t\to 0} ||\Delta_{t\varphi} h - h||_{\mathcal{H}} = 0.$$

Moreover,
(iii) if $\{\varphi_t,\ t \in R\}$ is a family of measurable functions on I satisfying $\lim_{t\to 0} t\varphi_t(x) = 0$ for almost all x, then for each $h \in \mathcal{H}$ it follows that

$$\lim_{t\to 0} ||\Delta_{t\varphi_t} h - h||_{\mathcal{H}} = 0.$$

Δ_φ plays a similar role to the standard location operator $\Delta_t^* f(y) = f(y-t)$ investigated in Hájek and Šidák (1967), pp. 210-212, and exploited in later articles on semiparametric estimation. For the proof of (i) - (iii) and other useful properties of Δ_φ and related scale operators see Inglot and Ledwina (2003a), Section A. Also note that some general shift operators were studied in the Appendix of Koul and Schick (1996).

Consider now the question of the differentiability of $p^{1/2}(\bullet;\kappa)$ itself. Take $b = \sqrt{g},\ c = \sqrt{f}$. Obviously, f and g satisfy $\int_I g d\lambda = \int_R f d\lambda = 1$. So, if one wants to approach $p^{1/2}(z;\kappa)$ through some, possibly completely artificial, "paths" within the space of densities, then one can disturb $b = \sqrt{g}$ by $b_n \in \mathcal{B},\ b_n \to b_0 \in \mathcal{B}$, in the following way. Take a real sequence $\{t_n\},\ t_n \to 0$, such that for large n the function $[b + t_n b_n]^2$ is a probability density [with respect to λ in our setting]. This implies that b_0 has to satisfy $\int_I b_0 \sqrt{g} d\lambda = 0$. Therefore, given $b \in \mathcal{B}$, define $\mathcal{B}_0 \subset \mathcal{B}$ by

$$\mathcal{B}_0 = \{b_0 \in \mathcal{B} :\ \int_I b_0 b d\lambda = 0\}.$$

Analogously, taking $c = \sqrt{f},\ c_n \to c_0 \in \mathcal{C},\ t_n \to 0$ such that for large n $\int_R [c + t_n c_n]^2 d\lambda = 1$ and $\int_R \iota[c + t_n c_n]^2 d\lambda = 0$, $\iota(y) = y$, [cf. the model assumptions $< M1 >$], one can easily infer that c_0 has to belong to the subspace

$$\mathcal{C}_0 = \{c_0 \in \mathcal{C} : \int_R c_0 c d\lambda = \int_R \iota c_0 c d\lambda = 0\}.$$

Set $\Omega_0 = \mathcal{A} \times \mathcal{B}_0 \times \mathcal{C}_0$. Take f and g satisfying $< M1 >$ and

$$\omega = \kappa = (a, \sqrt{g}, \sqrt{f}).$$

Moreover, consider a sequence $\{\omega_n\} \subset \Omega,\ \omega_n \to \omega_0 \in \Omega_0$ and $t_n \to 0$. In this setting (7.3) is applicable at $\omega = (a, b, c) = \kappa$ and the following holds

$$\frac{1}{t_n} \left|\left| p^{1/2}(\bullet;\kappa + t_n\omega_n) - p^{1/2}(\bullet;\kappa) - \frac{1}{2} t_n \left[\frac{\dot\Phi_\kappa(\omega_n)}{\frac{1}{2}p^{1/2}(\bullet;\kappa)}\right] p^{1/2}(\bullet;\kappa) \right|\right|_{\mathcal{H}} \to 0.$$

This relation shows that $\dot\Phi_\kappa(\bullet)/[\frac{1}{2}p^{1/2}(\bullet;\kappa)]$ is the standard form of the Hadamard derivative $\dot s_\kappa(\bullet)$, say, of $s_k(\bullet) = p^{1/2}(\bullet;\kappa)$, cf. e.g. van der Vaart (1991). So, we have

$$\dot s_\kappa(\bullet) = \frac{\dot\Phi_\kappa(\bullet)}{\frac{1}{2}p^{1/2}(\bullet;\kappa)}. \tag{7.4}$$

This, together with (7.3), implies that the operator $\dot{s}_\kappa(\bullet)$ is defined by the vector

$$\Delta_{wa^T}\left(-u\left[\frac{f'}{f}\right], -v\left[\frac{f'}{f}\right], \frac{1}{\sqrt{f}}, \frac{1}{\sqrt{g}}\right). \tag{7.5}$$

This vector is not affected by the restrictions on the set of directions Ω_0 from which one approaches the model density. However, the restricted set of directions Ω_0 plays an essential role when calculating projections of some components of (7.5) onto the subspace spanned by the remaining components of (7.5). Also note that the argument relating $\dot{\Phi}_\kappa$ to $\dot{s}_\kappa$ shows that to get the efficient score vector (3.1), it is enough to project the first k components of $\Delta_{wa^T}(-bc'w, c, b)$ onto the subspace

$$\left\{bc'[v\beta_0^T] + bc_0 + cb_0 :\ \beta_0 \in R^q,\ b_0 \in \mathcal{B}_0,\ c_0 \in \mathcal{C}_0\right\}$$

in the standard space $\mathcal{H} = L_2(I \times R, \lambda \times \lambda)$ and, at the final stage, to divide the resulting expressions by $\frac{1}{2}p^{1/2}(\bullet;\kappa)$. To calculate projections in $\mathcal{H}$, one can exploit standard results on Hilbert spaces, very nicely presented in Appendices A.2 and A.4 of Bickel et al. (1993). Some traditionally applied projections in $L_2(I \times R, P_\kappa)$ can be avoided in this way. Thus, this approach allows to extract purely analytical calculations and separate them from other derivations for which a probability space is really needed. This seems to simplify the presentation. We applied this method of derivation of an efficient score ℓ^* [cf.(3.1)] in Sections B[1] and C[1] of Inglot and Ledwina (2003a).

APPENDIX A

In this section we collect some auxiliary results that are needed to justify our implementation of the precedure. Note that we only need to check the consistency of several estimators under the null distribution P_η^n.

$\sqrt{n}$ - consistency of $\hat{\beta}$ follows by considering the normal equations and exploiting $< M1 >$. Therefore we start with

A1. *$\sqrt{n}$ - consistency of the adjusted Rice estimator of τ under the null model.* Rice defined the estimator in a nonparametric regression model with fixed design. Similarly to Guerre and Lavergne (2003, 2005), we used an adjusted version of this estimator in our simulation study. To be specific, the estimator of τ for a sample of size n is defined as follows. Let $(X_1, Y_1), ..., (X_n, Y_n)$ be i.i.d. random variables where $Y_i = \beta[v(X_i)]^T + \epsilon$ satisfies $< M1 >$. We now introduce the vector of order statistics $X_{(1)} \le ... \le X_{(n)}$, the vector of ranks $(R_1, ..., R_n)$ of $(X_1, ..., X_n)$ and the vector of antiranks $(D_1, ..., D_n)$, being, by definition, the inverse permutation of $(R_1, ..., R_n)$. We have $X_i = X_{(R_i)}$ and $X_{(i)} = X_{D_i}$ while the estimator of τ is defined by

$$\hat{\tau} = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \left(Y_{D_{i+1}} - Y_{D_i}\right)^2.$$

Note that, formally, we only need to consider the consistency of $\hat{\tau}$. However, knowing the consistency rate as well we have more flexibility when defining $\hat{\alpha}_n$ in our implementation.

LEMMA A.1. *In addition to $< M1 >$, assume that the functions $v_1(x), ..., v_q(x),\ x \in I$, satisfy the Lipschitz condition. Then*

$$\sqrt{n}(\hat{\tau} - \tau) = O_{P_\eta^n}(1).$$

PROOF. From $< M1 >$, $(D_1, ..., D_n)$ is independent of $(\epsilon_1, ..., \epsilon_n)$. This implies

$$(\epsilon_1, ..., \epsilon_n) \stackrel{\mathcal{D}}{=} (\epsilon_{D_1}, ..., \epsilon_{D_n}). \tag{A.1}$$

Set

$$E_i = \epsilon_{D_{i+1}} - \epsilon_{D_i}, \quad E = (E_1, ..., E_{n-1}), \quad V_i = v(X_{(i+1)}) - v(X_{(i)}), \quad \mathbf{V_0} = (V_1^T, ..., V_{n-1}^T).$$

Note that $\mathbf{V_0}$ is $q \times (n-1)$ random matrix and

$$\hat{\tau} = \frac{1}{2(n-1)} \left|\left|\beta^T \mathbf{V_0} + E\right|\right|^2, \tag{A.2}$$

where $|| \bullet ||$ stands for the Euclidean norm in R^{n-1}. The property (A.1) and the assumptions on the errors yield $||E|| = O_{P_\eta^n}(\sqrt{n})$. By (A.2) and the triangular inequality

$$||E|| - r_n \leq \sqrt{2(n-1)\hat{\tau}} \leq ||E|| + r_n, \text{ where } r_n = ||\beta^T \mathbf{V_0}||.$$

As $\sum_{i=1}^{n-1}[X_{(i+1)} - X_{(i)}] \leq 1$ we have $\sum_{i=1}^{n-1}[X_{(i+1)} - X_{(i)}]^2 \leq 1$ and the Lipschitz condition along with the Schwarz inequality yield $r_n = O_{P_\eta^n}(1)$. Therefore, (A.2) implies

$$\sqrt{n-1}(\hat{\tau} - \tau) = \frac{1}{2\sqrt{n-1}} \sum_{i=1}^{n-1} [E_i^2 - 2\tau] + O_{P_\eta^n}(1).$$

Hence, the conclussion follows from (A.1) and the assumptions on the ϵ_i's. □

A.2. *Estimating the score function.* In many papers on semiparametric estimation the score function f'/f is estimated simply by $\hat{f}'/\hat{f}$, where

$$\hat{f}(y) = \gamma_n + \frac{1}{na_n} \sum_{i=1}^{n} K\left(\frac{y - \hat{\epsilon}_i}{a_n}\right),$$

while $\hat{\epsilon}_i = Y_i - v(X_i)\hat{\beta}^T$, $\{\gamma_n\}$ and $\{a_n\}$ are deterministic sequences tending to 0. However, it is more effective to use a random bandwidth $\hat{\alpha}_n$ instead of a_n in the above formula. Below we show a consistency result for such modified estimator of f. As in Section A.1, to simplify the notation we formulate the result for one sample of size n.

LEMMA A.2. *Consider the regression model $Y = v(X)\beta^T + \epsilon$, where $\beta \in R^q$, while v is a vector of bounded functions on I. Assume that ϵ possesses a density f which is differentiable a.s. and the Fisher information $J = J(f)$ is finite. Let $\hat{\beta}$ be a $\sqrt{n}$ - consistent estimator of β and $\hat{\alpha}_n, \hat{\alpha}_n > 0$, a random bandwidth. For n i.i.d. (X_i, Y_i) obeying the model, set*

$$\widetilde{f}(y) = \gamma_n + \frac{1}{n\hat{\alpha}_n} \sum_{i=1}^{n} K\left(\frac{y - \hat{\epsilon}_i}{\hat{\alpha}_n}\right),$$

where $\hat{\epsilon}_i = Y_i - v(X_i)\hat{\beta}^T$, $\{\gamma_n\}$ is a deterministic sequence converging to 0, while K is a positive twice differentiable symmetric density such that $K \leq C$ and $\max\{|K'|, |K''|\} \leq CK$. Suppose in addition that for $\hat{\alpha}_n$ and an auxiliary deterministic sequence $\{\alpha_n\}$, such that $\alpha_n \to 0$ and $n\gamma_n^2\alpha_n^6 \to \infty$, we have $\sqrt{n}(\alpha_n/\hat{\alpha}_n - 1) = O_{P_\eta^n}(1)$. Then

$$\bigwedge_{\delta > 0} \quad P_\eta^n \left(\int_R \left(\frac{\tilde{f}'}{\tilde{f}}(y) - \frac{f'}{f}(y) \right)^2 f(y)\lambda(dy) > \delta \right) \to 0 \quad \text{as} \quad n \to \infty. \tag{A.3}$$

PROOF. First we introduce some auxiliary notation. Let

$$\epsilon_i = Y_i - v(X_i)\beta^T, \quad f_n(y) = \gamma_n + \frac{1}{n\alpha_n}\sum_{i=1}^{n} K\left(\frac{y-\epsilon_i}{\alpha_n}\right) \quad \text{and} \quad \bar{f}_n(y) = \gamma_n + \frac{1}{n\hat{\alpha}_n}\sum_{i=1}^{n} K\left(\frac{y-\epsilon_i}{\hat{\alpha}_n}\right).$$

From Forrester et al. (2003), Section 8, [by inserting f_n instead of their $\hat{f}_n$], since $n\gamma_n^2\alpha_n^6 \to \infty$, we infer

$$P_\eta^n\left(\int_R \left[\frac{f_n'}{f_n}(y) - \frac{f'}{f}(y)\right]^2 f(y)\lambda(dy) > \delta\right) \to 0 \text{ as } n \to \infty.$$

Therefore, to get (A.3) it is enough to show that both relations

$$P_\eta^n\left(\int_R \left[\frac{\widetilde{f}'}{\widetilde{f}}(y) - \frac{\bar{f}_n'}{\bar{f}_n}(y)\right]^2 f(y)\lambda(dy) > \delta\right) \to 0, \quad P_\eta^n\left(\int_R \left[\frac{\bar{f}_n'}{\bar{f}_n}(y) - \frac{f_n'}{f_n}(y)\right]^2 f(y)\lambda(dy) > \delta\right) \to 0 \tag{A.4}$$

hold as $n \to \infty$. To deal with the first term of (A.4), note that

$$\left|\frac{\widetilde{f}'}{\widetilde{f}} - \frac{\bar{f}_n'}{\bar{f}_n}\right| \le \left|\frac{\widetilde{f}'}{\widetilde{f}}\right|\left|\frac{\widetilde{f} - \bar{f}_n}{\bar{f}_n}\right| + \left|\frac{\widetilde{f}' - \bar{f}_n'}{\bar{f}_n}\right|.$$

Since $|K'| \le CK$, neglecting γ_n in $\widetilde{f}$, for any y we get

$$\left|\frac{\widetilde{f}'}{\widetilde{f}}(y)\right| \le \frac{C}{\hat{\alpha}_n}. \tag{A.5}$$

From our assumptions, K and K' are Lipschitz with bounding constant C^2. Moreover, as before set $\sup_x ||v(x)|| = ||v||_\infty$. Therefore, from the definition of $\widetilde{f}$ and f_n, we infer that for any y

$$|\widetilde{f}'(y) - \bar{f}_n'(y)| \le \frac{C^2}{n\hat{\alpha}_n^2}\sum_{i=1}^{n}\frac{|\hat{\epsilon}_i - \epsilon_i|}{\hat{\alpha}_n} \le \frac{C^2||v||_\infty}{\hat{\alpha}_n^3}||\hat{\beta} - \beta||.$$

In consequence

$$\left[\frac{\widetilde{f}'(y) - \bar{f}_n'(y)}{\bar{f}_n(y)}\right]^2 \le \frac{1}{\gamma_n^2}\{\widetilde{f}'(y) - \bar{f}_n'(y)\}^2 \le \left[\frac{C^2||v||_\infty\sqrt{n}||\hat{\beta} - \beta||}{\sqrt{n}\gamma_n\alpha_n^3}\right]^2\left[\frac{\alpha_n}{\hat{\alpha}_n}\right]^6. \tag{A.6}$$

Similarly, from the Lipschitz condition for K

$$\left[\frac{\widetilde{f}(y) - \bar{f}_n(y)}{\bar{f}_n(y)}\right]^2 \le \left[\frac{C^2||v||_\infty\sqrt{n}||\hat{\beta} - \beta||}{\sqrt{n}\gamma_n\alpha_n^2}\right]^2\left[\frac{\alpha_n}{\hat{\alpha}_n}\right]^6. \tag{A.7}$$

As $\hat{\beta}$ is $\sqrt{n}$ - consistent, due to the assumptions on γ_n, α_n and $\hat{\alpha}_n$, from (A.5)-(A.7), the first term in (A.4) converges to 0. Using similar arguments, the second term of (A.4) tends to 0 as well. □

COROLLARY A.3. *Choose $\alpha_n = c_0\tau^{1/2}n^{-1/7}$, where c_0 is a positive constant and γ_n satisfies $n\gamma_n^2\alpha_n^6 \to \infty$. From the $\sqrt{n}$ - consistency of $\hat{\tau}$ it follows that $\hat{\alpha}_n = c_0[\hat{\tau}]^{1/2}n^{-1/7}$ fulfils the assumptions of Lemma A.2.*

REMARK A.4. Choose $\hat{\alpha}_n = a_n$ a.s. for each n with a_n satisfying $n\gamma_n^2 a_n^6 \to \infty$. Then, obviously, $\{a_n\}$ satisfies the assumption of Lemma A.2 and we have in particular

$$\bigwedge_{\delta>0} \quad P_\eta^n \left(\int_R \left(\frac{\hat{f}'}{\hat{f}}(y) - \frac{f'}{f}(y) \right)^2 f(y)\lambda(dy) > \delta \right) \to 0 \quad \text{as} \quad n \to \infty. \tag{A.8}$$

REMARK A.5. A typical kernel K satisfying these assumptions is the logistic one. However, our experience with using this kernel in simulations was discouraging. Therefore we used a Gaussian kernel. Formally, this kernel requires some modification to satisfy $\max\{|K'|, |K''|\} \leq CK$. Obviously, such modification, which does not influence the simulation results, can be done. Therefore, we did not utilize it in our implementation. Also note that some kernel-type estimators of f'/f with random bandwidth have been applied earlier in a regression context by Koul and Susarla (1983), e.g.

A.3. *Consistency of the estimator of J.* The consistency of $\hat{J}^{(j)}$, $j = 1, 2$, is established in the following lemma.

LEMMA A.5. *Suppose that the assumptions of Lemma A.2 hold. We specify $\hat{\beta}$ here to be a discretized version of some $\sqrt{n}$ - consistent estimator. Consider the sample splitting scheme from Section 3.3. Set $\widetilde{f}^{(j)}(y)$, $j = 1, 2$, for the adjusted versions of $\widetilde{f}$ defined in Lemma A.2. Set*

$$\hat{J}^{(1)} = \frac{1}{\zeta} \sum_{c\in<1>} \left[\frac{\widetilde{f}'^{(2)}}{\widetilde{f}^{(2)}} \left(\hat{\epsilon}_c^{(2)} \right) \right]^2 \quad \textit{and} \quad \hat{J}^{(2)} = \frac{1}{n-\zeta} \sum_{c\in<2>} \left[\frac{\widetilde{f}'^{(1)}}{\widetilde{f}^{(1)}} \left(\hat{\epsilon}_c^{(1)} \right) \right]^2,$$

where $\hat{\epsilon}_c^{(j)} = Y_c - v(X_c)[\hat{\beta}^{(j)}]^T$, $c = 1, ..., \zeta$ for $j = 2$ and $c = \zeta+1, ..., n$ for $j = 1$. Then, under P_η^n, $\hat{J}^{(j)}$, $j = 1, 2$, are consistent estimators of J.

PROOF. We shall apply similar arguments to these used in Section 5 and some details shall therefore be omitted. As previously, it suffices to consider e.g. $j = 1$. Due to the discretization, we can restrict our attention to an analysis of

$$\hat{J}_n^{(1)} = \frac{1}{\zeta} \sum_{c\in<1>} \left[\frac{\widetilde{f}'^{(2)}}{\widetilde{f}^{(2)}} \left(\epsilon_{cn} \right) \right]^2,$$

where $\epsilon_{cn} = Y_c - v(X_c)b_n^T$, while $\{b_n\}$ is such that $\{\sqrt{n}(b_n - \beta)\}$ stays bounded. Recall that $\eta = (\beta, \vartheta)$. Since the sequences $\{P_{(\beta,\vartheta)}^n\}$ and $\{P_{(b_n,\vartheta)}^n\}$ are contiguous, then it is enough to prove that $P_{(b_n,\vartheta)}^n \left(|\hat{J}_n^{(1)} - J| > \delta \right) \to 0$ holds for any positive δ. We shall show this by proving that for any $\delta > 0$

$$P_{(b_n,\vartheta)}^n \left(|J_n^{(1)} - J| > \delta \right) \to 0 \quad \text{and} \quad P_{(b_n,\vartheta)}^n \left(|\hat{J}_n^{(1)} - J_n^{(1)}| > \delta \right) \to 0, \tag{A.9}$$

where

$$J_n^{(1)} = \frac{1}{\zeta} \sum_{c\in<1>} \left[\frac{f'}{f}(\epsilon_{cn}) \right]^2.$$

As, under $P_{(b_n,\vartheta)}$ the $\epsilon_{i,n}$'s have the same distribution as the Y_i's under $P_{(0,\vartheta)}$, the first term of (A.9) tends to 0 from the weak law of large numbers. This convergence also implies that $P_{(b_n,\vartheta)}^n \left(|\sqrt{J_n^{(1)}} - \sqrt{J}| > \delta \right) \to 0$. This fact allows us to replace the second part of (A.9) by

$P^n_{(b_n,\vartheta)}\left(|\sqrt{\hat{J}^{(1)}_n}-\sqrt{J^{(1)}_n}|>\delta\right)\to 0$, as
$\hat{J}^{(1)}_n - J^{(1)}_n = \left(\sqrt{\hat{J}^{(1)}_n}-\sqrt{J^{(1)}_n}\right)\left(\sqrt{\hat{J}^{(1)}_n}-\sqrt{J^{(1)}_n}+2[\sqrt{J^{(1)}_n}-\sqrt{J}]+2\sqrt{J}\right)$. Therefore, by the triangular inequality, it is enough to show that for

$$T_n=\frac{1}{\zeta}\sum_{c\in<1>}\left[\frac{\widetilde{f}'^{(2)}}{\widetilde{f}^{(2)}}(\epsilon_{cn})-\frac{f'}{f}(\epsilon_{cn})\right]^2$$

it follows that $P^n_{(b_n,\vartheta)}(T_n>\delta)\to 0$. To prove this we shall apply Proposition 3 with $P_n=P^n_{(b_n,\vartheta)}$ and $\mathcal{F}_n=\sigma(X_{\zeta+1},...,X_n,Y_{\zeta+1},...,Y_n)$. Since we have

$$E(T_n|\mathcal{F}_n)=\int_R\left[\frac{\widetilde{f}'^{(2)}}{\widetilde{f}^{(2)}}(y)-\frac{f'}{f}(y)\right]^2 f(y)dy$$

an application of Lemma A.2 concludes the proof. □

APPENDIX B

Proof of Proposition 3. Recall that for each n, T_n is a random variable defined on $(\mathcal{T}_n,\mathcal{B}_n,P_n)$. Therefore, for each n, by Jensen inequality it holds $\{E(|T_n||\mathcal{F}_n)\}^2\le E(T_n^2|\mathcal{F}_n)$ P_n - a.s. Hence, for any $\delta\in(0,1)$

$$P_n\left(\frac{E(|T_n||\mathcal{F}_n)}{1+E(|T_n||\mathcal{F}_n)}>\delta\right)=P_n\left(E(|T_n||\mathcal{F}_n)>\frac{\delta}{1-\delta}\right)\le P_n(E(T_n^2|\mathcal{F}_n)>\delta^2) \qquad (B.1)$$

and by the assumption the right hand side of (A.4) tends to 0 as $n\to\infty$.

Applying again Jensen inequality for the concave function $h(x)=x/(1+x)$, $x\in(0,\infty)$ we get

$$E(h(|T_n|)|\mathcal{F}_n)\le\frac{E(|T_n||\mathcal{F}_n)}{1+E(|T_n||\mathcal{F}_n)}\quad P_n-\text{a.e.} \qquad (B.2)$$

By the above we infer that

$$E(h(|T_n|)|\mathcal{F}_n)\to 0 \text{ in } P_n. \qquad (B.3)$$

Now observe that

$$E(h(|T_n|))=E\left[E(h(|T_n|)|\mathcal{F}_n)\right]=\int_{\mathcal{T}_n}E(h(|T_n|)|\mathcal{F}_n)dP_n=$$

$$=\int_0^1 P_n(E(h(|T_n|)|\mathcal{F}_n)>\delta)d\delta. \qquad (B.4)$$

By (B.3) and the Lebesgue Dominated Convergence Theorem the right hand side of (B.4) tends to 0 and $Eh(|T_n|)\to 0$ as well.

On the other hand, for any positive δ

$$Eh(|T_n|)\ge\frac{\delta}{1+\delta}\int_{\mathcal{T}_n}\mathbf{1}_{\{|T_n|>\delta\}}dP_n=\frac{\delta}{1+\delta}P_n(|T_n|>\delta). \qquad (B.5)$$

Thus, by the above, (B.5) yields the conclussion of Proposition 2. □

Acknowledgements. The second named author is deeply indebted to A. Schick for insightful discussion in Oberwolfach, September 2003. The discussion resulted, in particular, in sharpening the previous version of the present Theorem 1. She thanks also G.R. Ducharme and W.C.M. Kallenberg for providing copies of Choi (1989) and Neyman (1954), respectively, and to P. Lavergne and J. Mielniczuk for helpful comments. The programming work was done by A. Janic-Wróblewska under support from the KBN 5 P03A 03020 grant. Her kind co-operation is gratefully acknowledged. The computations were in part done at the Institute of Mathematics of Wrocław University of Technology. We especially thank M. Kaczmarz, W. Rutkowski and M. Wyłomański for help and advice in this regard. This research was partially supported by grant 5 P03A 03020 from KBN.

REFERENCES

Aerts, M., Claeskens, G. and Hart, J. D. (2000). Testing lack of fit in multiple regression. *Biometrika* **87** 405-424.

Azzalini, A. and Bowman, A. (1993). On use of nonparametric regression for checking linear relationships. *J. R. Statist. Soc.* B **55** 549-557.

Baraud, Y., Huet, S. and Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *Ann. Statist.* **31** 225-251.

Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647-671.

Bickel, P. J., Klaassen, C. A., Ritov, Y. and Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. John Hopkins Univ. Press, Baltimore.

Bickel, P. J., Kwon, J. (2001). Inference for semiparametric models: some questions and an answer. *Statist. Sinica* **11** 863-960.

Bickel, P. J., Ritov, Y. and Stocker, T. (1998). Testing and the method of sieves. Technical Report, University of California, Berkeley.

Bogdan, M. (1995). Data driven versions of Pearson's chi-square test for uniformity. *J. Statist. Comput. Simul.* **52** 217-237.

Choi, S. (1989). On asymtotically optimal tests. Ph.D. dissertation, Dept. Statistics, University of Rochester, Rochester.

Choi, S., Hall, W. J. and Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24** 841-861.

Cox, D., Koh, E., Wahba, G. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16** 113-119.

Csörgő, M. and Révész, P. (1986). A nearest neighbour-estimator of the score function. *Probab. Th. Rel. Fields* **71** 293-305.

Dette, H. (2000). On a nonparametric test for linear relationships. *Statist. Probab. Lett.* **46** 307-316.

Diebolt, J. and Zuber, J. (2000). On the half-sample method for goodness-of-fit in regression. Technical Report 00.1 Department of Mathematics and Statistics, La Trobe University.

Eubank, R. L., Hart, J. D. and LaRiccia, V. N. (1993). Testing goodness of fit via nonparametric function estimation techniques. *Commun. Statist. - Theory Meth.* **22** 3327-3354.

Fan, J. (1996). Tests of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674-688.

Fan, J. and Huang, L.-S. (2001). Goodness-of-fit tests or parametric regression models. *J. Amer. Statist. Assoc.* **96** 640-652.

Faraway, J. J. (1992). Smoothing in adaptive estimation. *Ann. Statist.* **20** 414-427.

Forrester, J., Hooper, W., Peng, H. and Schick, A. (2003). On the construction of efficient estimators in semiparametric models. *Statist. Decisions* **21** 109-138.

Guerre, E. and Lavergne, P. (2003). Data-driven rate-optimal specification testing in regression models. (`http://lavergne.pascal.free.fr`)

Guerre, E. and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33** 840-870.

Hájek, J. and Šidák, Z. (1967). *Theory of rank tests.* Academia, Prague.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parmetric regression fits. *Ann. Statist.* **21** 1926-1947.

Hart, J. D. (1997). *Nonparametric smoothing and lack-of-fit tests.* Springer, New York.

Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599-631.

Inglot, T. and Janic-Wróblewska, A. (2003). Data driven chi-square test for uniformity with nonequal cells. *J. Statist. Comput. Simul.* **73** 545-561.

Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (1994). Power approximations to and power comparison of smooth goodness-of-fit tests. *Scand. J. Statist.* **21** 131-145.

Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (1997). Data driven smooth tests for composite hypotheses. *Ann. Statist.* **25** 1222-1250.

Inglot, T. and Ledwina, T. (1996). Asymptotic optimality of data driven Neyman's tests for uniformity. *Ann. Statist.* **24** 1982-2019.

Inglot, T. and Ledwina, T. (2001a). Intermediate approach to comparison of some goodness-of-fit tests. *Ann. Inst. Statist. Math.* **53** 810-834.

Inglot, T. and Ledwina, T. (2001b). Asymptotic optimality of data driven smooth tests for location-scale family. *Sankhyā*, Ser. A **63** 41-71.

Inglot, T. and Ledwina, T. (2003a). Semiparametric regression: Hadamard differentiability and efficient score functions for some testing problems. Preprint 012, Institute of Mathematics, Wrocław University of Technology.

Inglot, T. and Ledwina, T. (2003b). Data driven smooth test of fit for semiparametric homoscedastic regression model. Preprint 011, Institute of Mathematics, Wrocław University of Technology.

Inglot, T. and Ledwina, T. (2004). On consistent minimax distinguishability and intermediate efficiency of Cramér-von Mises test. *J. Statist. Plann. Inference* **124** 453-474.

Inglot, T. and Ledwina, T. (2005). Towards data driven selection of a penalty function for data driven Neyman tests. *Linear Algebra Appl.*, in print.

Javitz, H. S. (1975). Generalized smooth tests of goodness of fit, independence and equality of distributions. Ph. D. dissertation, Univ. California, Berkeley.

Jin, K. (1992). Empirical smoothing parameter selection in adaptive estimation. *Ann. Statist.* **20** 1844-1874.

Kallenberg, W. C. M. and Ledwina, T. (1995). On data-driven Neyman's tests. *Probability and Mathematical Statistics* **15**, 409-426.

Kallenberg, W. C. M. and Ledwina, T. (1997a). Data driven smooth tests for composite hypotheses: Comparison of powers. *J. Statist. Comput. Simul.* **59** 101-121.

Kallenberg, W. C. M. and Ledwina, T. (1997b). Data-driven smooth tests when the hypothesis is composite. *J. Amer. Statist. Assoc.* **92** 1094-1104.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240-257.

Klaassen, C. A. J. (2001). Comments on "Inference for semiparametric models: some questions and an answer" by P. J. Bickel and J. Kwon. *Statistica Sinica* **11** 906-909.

Koenker, R. and Xiao, Z. (2001). Inference on the quantile regression process: appendices. (`http://www.econ.uiuc.edu/~roger/research/inference`)

Koul, H. L. and Schick, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *Ann. Statist.* **24** 1025-1052.

Koul, H. L. and Susarla, V. (1983). Adaptive estimation in linear regression. *Statist. Decisions* **1** 379-400.

Kozek, A. S. (1991). A nonparametric test of fit of a parametric model. *J. Multivariate Anal.* **37** 66-75.

Kreiss, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112-133.

Le Cam, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 129-156. Univ. California Press, Berkeley.

Le Cam, L. and Lehmann, E. L. (1974). J. Neyman - On the occasion of his 80th birthday. *Ann. Statist.* **2** vii-xiii.

Ledwina, T. (1994). Data driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89** 1000-1005.

Mammen, E. and Park, B. V. (1997). Optimal smoothing in adaptive location estimation. *J. Statist. Plann. Inference* **58** 333-348.

Maronna, R. A. and Yohai, V. (1981). Asymptotic behaviour of general M-estimates for regression and scale with random carriers. *Z. Wahrsch. Gebiete* **58** 7-20.

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

Neyman, J. (1954). Sur une famille de tests asymptotiques des hypothéses statistiques composées. *Trabajos de Estadistica* **5** 161-168.

Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics:* The Harald Cramér Volume (U. Grenander, ed.) 213-234. Wiley, New York.

Rayner, J. C. W. and Best, D. J. (1989). *Smooth tests of goodness of fit.* Oxford Univ. Press, New York.

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

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

Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21** 1486-1521. Correction (1995) **23** 1862-1863.

Schick, A. (1994). On efficient estimation in regression models with unknown scale functions. *Math. Meth. Statist.* **3** 171-212.

Schick, A. (1997). Efficient estimates in linear and nonlinear regression with heteroscedastic error. *J. Statist. Plann. Inference* **58** 371-387.

Schick, A. (2001). On asymptotic differentiability of averages. *Statist. Probab. Lett.* **51** 15-23.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613-641.

Stute, W., Thies, S. and Zhu, L.-X. (1998a). Model checks for regression: an innovation process approach. *Ann. Statist.* **26** 1916-1934.

Stute, W., Manteiga, W. G. and Quindimil, M. P. (1998b). Bootstrap approximation in model checks in regression. *J. Amer. Statist. Assoc.* **93** 141-149.

Thomas, D. and Pierce, D. (1979). Neyman's smooth goodness-of-fit test when the hypothesis is composite. *J. Amer. Statist. Assoc.* **74** 441-445.

van der Vaart, A. W. (1988). Estimating a real parameter in a class of semiparametric models. *Ann. Statist.* **16** 1450-1474.

van der Vaart, A. W. (1991). On differentiable functionals. *Ann. Statist.* **19** 178-204.

Tadeusz Inglot, Institute of Mathematics and Informatics, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail address: inglot@im.pwr.wroc.pl

Teresa Ledwina, Institute of Mathematics, Polish Academy of Sciences, ul. Kopernika 18, 51-617 Wrocław, Poland
e-mail address: ledwina@impan.pan.wroc.pl