PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 26, Fasc. 1 (2006), pp. 41–61

DATA-DRIVEN SCORE TESTS FOR HOMOSCEDASTIC LINEAR REGRESSION MODEL: ASYMPTOTIC RESULTS

BY

TADEUSZ INGLOT AND TERESA LEDWINA (WROCŁAW)

Abstract. We describe and investigate new tests for testing the validity of a semiparametric random-design linear regression model. The tests were introduced in Inglot and Ledwina (2006a, b). We repeat here basic steps of the constructions. The resulting statistics are closely linked to some norms of the appropriate efficient score vector and related quantities. A useful way of deriving the efficient score vector is proposed and discussed. We introduce also a large class of estimators of the efficient score vector and prove that under the null model our constructions are asymptotically distribution free. The proof adopts and exploits some ideas and results developed in the area of semiparametric estimation. We give also the limiting distribution of the test statistics under the null hypothesis. The simulation results contained in Inglot and Ledwina (2006a, b) show the very good performance of the proposed tests.

2000 AMS Mathematics Subject Classification: 62G10, 62E20, 62J05.

Key words and phrases: 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 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).

In Inglot and Ledwina (2006b) we proposed some new tests of fit for the following problem.

Let $Z = (X, Y)$ denote a random vector in $I \times R$, $I = [0, 1]$. The null hypothesis H_0 asserts

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

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

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

$$\mathfrak{M}(k)\colon\ Y = \theta[u(X)]^T + \beta[v(X)]^T + \varepsilon, \tag{1.2}$$

which satisfies the following assumptions:

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

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

$\langle M1 \rangle$ 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 $E_f\varepsilon = 0$, $\tau = E_f\varepsilon^2$ and $0 < \tau < \infty$;

X and ε are independent.

At the first step we constructed appropriate score test statistic, for the fixed k, for testing $H_0(k)\colon\ \theta = 0$ against $\theta \neq 0$ in $\mathfrak{M}(k)$ satisfying $\langle M1 \rangle$ and some further regularity conditions $\langle M2 \rangle$ and $\langle M3 \rangle$. An efficient score vector along with its appropriate estimator play the central role in this construction. Section 6 of the present paper briefly introduces our approach to a derivation of efficient scores. This section may be of independent interest. The next step was incorporating into this statistic a score-based selection rule for determining the dimension k. The both steps are repeated in Section 3. Section 3 contains also formulation of our basic asymptotic results. 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 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 5 we discuss various aspects of our general assumptions. Small sample behaviour of the introduced tests is discussed in Inglot and Ledwina (2006b).

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 of random X. 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). Koenker and Xiao (2002) as well as Khmaladze and Koul (2004) studied Kolmogorov–Smirnov type tests. Kozek (1991), Härdle and Mammen (1993), Simonoff and Tsai (1999) and many others proposed to measure a distance between parametric and nonparametric estimators. Horowitz and Spokoiny (2001) refined such an 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), Fan and Huang (2001) and Baraud et al. (2003) were exceptions to the mainstream. In these articles the starting point for test construction were some ideas related to testing theory. These five papers deal with the case of fixed design. Last two additionally assume that the errors are Gaussian. The study of Dette (2000) extended the solution of Azzalini and Bowman (1993) to the case of random design. For some comparison of few of these solutions with adequate data-driven score test see Inglot and Ledwina (2006c, 2006e).

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 (1974), p. ix). Renewed interest in this solution and its 1959 extension, related to goodness-of-fit problems, was observed after publishing 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 nuisance 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 a few. In particular, the construction introduced in Ledwina (1994) is closely related to the original idea of Neyman (1937), as it provides an 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 development 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 lies 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) was stimulating to 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 5 of Section 4 of the present paper. 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. These 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 is presented in Section 3. The simulation results given in Inglot and Ledwina (2006b) show that these data-driven constructions have 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 issue, it is obvious that a similar approach can be adopted and elaborated for many other 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, as developed in Inglot and Ledwina (2006b).

3.1. Efficient score vector for testing $\theta = 0$ in $\mathfrak{M}(k)$. A general result for score vectors in some large class of regression models is stated in Schick (1997). As mentioned in Section 2, we reproved some existing results on score vectors in the model $\mathfrak{M}(k)$ and derived an efficient score vector for testing (1.1). The calculations for $\mathfrak{M}(k)$, as well as in the more general heteroscedastic case, are given in Inglot and Ledwina (2003). In the sequel we quote some of these results.

In the case under consideration, in addition to the basic model assumptions $\langle M1 \rangle$ we needed the following ones:

$\langle M2 \rangle$ $f'(y)$ exists for all $y \in R$ and

$$J = J(f) = \int_R \frac{[f'(y)]^2}{f(y)} \lambda(dy) < \infty,$$

$\langle M3 \rangle$ $g > 0$ λ-a.e.

Under these three assumptions the efficient score vector for testing $H_0(k)$: $\theta = 0$ in $\mathfrak{M}(k)$, calculated at $z = (x, y)$, is of the form

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

where

$$m_1 = E_g u(X), \quad m_2 = E_g v(X), \quad m = (m_1, m_2),$$

$$(3.2)\quad \tilde{w}(x) = (\tilde{u}(x), \tilde{v}(x)), \quad \tilde{u}(x) = u(x) - m_1, \quad \tilde{v}(x) = v(x) - m_2,$$

while M and V are blocks in

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

Note that, due to $\langle M3\rangle$, W is positive definite (cf. Remark C.13 in Inglot and Ledwina (2003)).

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

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

Moreover, P_η^n denotes the joint distribution of $Z_1, \ldots, Z_n$ under the null model (1.1).

Finally, set

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

and define

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

From $\langle M1\rangle$–$\langle M3\rangle$, Corollaries C.16, C.18 and Remark C.13 of Inglot and Ledwina (2003), e.g., under the null hypothesis $H_0(k)$, L is positive definite and it follows that

$$(3.5)\quad E_\eta l^*(Z; \eta) = 0, \quad \{E_\eta[l^*(Z; \eta)]^T[l^*(Z; \eta)]\}^{-1} = L, \quad W_k(\eta) \xrightarrow{\mathcal{D}} \chi_k^2,$$

where χ_k^2 denotes a random variable having the central chi-square distribution with k degrees of freedom.

Consider

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

where $\hat{l}^*(\cdot; \hat{\eta})$ is an estimator of $l^*(\cdot; \eta)$, while $\hat{\boldsymbol{L}}$ is an estimator of $\boldsymbol{L}$.

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

PROPOSITION 1. *Let the null hypothesis $H_0(k)$: $\theta = 0$ be true and the assumptions $\langle M1\rangle$, $\langle M2\rangle$ and $\langle M3\rangle$ be fulfilled. Suppose that $\hat{\boldsymbol{L}}$ is a consistent estimator of $\boldsymbol{L}$ and the estimator $\hat{l}^*(\cdot; \hat{\eta})$ satisfies the following condition:*

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

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

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

Remark 1. $W_k(\hat{\eta})$ is an efficient score statistic for testing $H_0(k)$ in $\mathfrak{M}(k)$. 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 $\hat{l}^*$ of l^* satisfying (3.7). We followed some well-established ideas in the area of semiparametric estimation. 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 proposed in Bickel (1982), can be avoided.

Suppose $Z_1, \ldots, Z_n$ 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 $\zeta = [n/2]$ and divide $Z_1, \ldots, Z_n$ into two parts: $Z_1, \ldots, Z_\zeta$ and $Z_{\zeta+1}, \ldots, Z_n$. 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 $\langle 1\rangle = \{1, \ldots, \zeta\}$, $\langle 2\rangle = \{\zeta+1, \ldots, n\}$. The superscript (j), $j = 1, 2$, appearing below, indicates from which part of the sample we estimate the related quantity.

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

$$\hat{l}^*(Z_i; \hat{\eta}) = \begin{cases} l^*(Z_i; \hat{\beta}_*^{(2)}, \hat{g}^{(2)}, \hat{f}^{(2)}, \hat{m}^{(1)}) & \text{if } i \in \langle 1\rangle, \\ l^*(Z_i; \hat{\beta}_*^{(1)}, \hat{g}^{(1)}, \hat{f}^{(1)}, \hat{m}^{(2)}) & \text{if } i \in \langle 2\rangle, \end{cases} \tag{3.8}$$

where

$$\hat{m}_1^{(1)} = \frac{1}{\zeta} \sum_{i \in \langle 1\rangle} u(X_i), \qquad \hat{m}_2^{(1)} = \frac{1}{\zeta} \sum_{i \in \langle 1\rangle} v(X_i),$$

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

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

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 $\hat{l}^*$ only the estimators $\hat{m}^{(j)}$ are matched to Z_i with i from $\langle j\rangle$ guarantee that the important property (4.7) holds (cf. also the discussion of the proof in Section 4). Moreover, the requirements for $\sqrt{n}$-consistency of an estimator for β, its discretization 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 explicitly the form of the estimators $\hat{l}^*(Z_i; \hat{\eta})$, $i\in\langle j\rangle$, $j = 1, 2$, denote by $\hat{V}^{(j)}$, $\hat{M}^{(j)}$, $\hat{\tau}^{(j)}$, $\hat{\tau}^{(j)} > 0$ — a.e., and $[f'/f]^{(j)}$ 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 $\mathscr{L}_j^*$, $j = 1, 2$, defined as follows:

$$\begin{aligned}
\mathscr{L}_1^*(z;\beta) = &-[f'/f]^{(2)}\big(y - v(x)\beta^T\big)\big[\tilde{u}^{(1)}(x) - \tilde{v}^{(1)}(x)[\hat{V}^{(2)}]^{-1}\hat{M}^{(2)}\big] \\
&+\frac{1}{\hat{\tau}^{(2)}}\big[y - v(x)\beta^T\big]\big[\hat{m}_1^{(1)} - \hat{m}_2^{(1)}[\hat{V}^{(2)}]^{-1}\hat{M}^{(2)}\big], \\
\mathscr{L}_2^*(z;\beta) = &-[f'/f]^{(1)}\big(y - v(x)\beta^T\big)\big[\tilde{u}^{(2)}(x) - \tilde{v}^{(2)}(x)[\hat{V}^{(1)}]^{-1}\hat{M}^{(1)}\big] \\
&+\frac{1}{\hat{\tau}^{(1)}}\big[y - v(x)\beta^T\big]\big[\hat{m}_1^{(2)} - \hat{m}_2^{(2)}[\hat{V}^{(1)}]^{-1}\hat{M}^{(1)}\big].
\end{aligned} \tag{3.9}$$

Finally, set

$$\hat{l}^*(Z_i;\hat{\eta}) = \begin{cases} \mathscr{L}_1^*(Z_i;\hat{\beta}_*^{(2)}) & \text{for } i\in\langle 1\rangle, \\ \mathscr{L}_2^*(Z_i;\hat{\beta}_*^{(1)}) & \text{for } i\in\langle 2\rangle. \end{cases} \tag{3.10}$$

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

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

Then the estimator $\hat{l}^*$ 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 choosing estimators defining $\hat{l}^*$. In Inglot and Ledwina (2006a, b) we proposed and discussed a particular choice and in Inglot and Ledwina (2006a) we proved that under this choice the assumptions of Theorem 1 are satisfied. The flexibility in choosing estimators is an attractive feature of the approach. It permits refinements of our relatively simple implementation by using some more fine choices, which are briefly discussed in Inglot and Ledwina (2006a).

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

$$S1 = \min\{1 \leqslant k \leqslant d\colon\ W_k(\hat{\eta}) - k\log n \geqslant W_s(\hat{\eta}) - s\log n,\ s = 1, \ldots, d\}.$$

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 $\mathfrak{M}(k)$. In contrast, the rule

$$A1 = \min\{1 \leqslant k \leqslant d\colon\ W_k(\hat{\eta}) - 2k \geqslant W_s(\hat{\eta}) - 2s,\ s = 1, \ldots, d\},$$

imitating the Akaike AIC criterion, is expected to work well when high-dimensional disturbances $\mathfrak{M}(k)$ of the null model $\mathfrak{M}(0)$: $Y = \beta[v(X)]^T + \varepsilon$, are present. Based on our experience and some previous articles, the following "intermediate" solution was proposed and discussed in Inglot and Ledwina (2006d). 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

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

Then, obviously, $W_k(\hat{\eta}) = \|(\mathscr{Y}_1, \ldots, \mathscr{Y}_k)\|^2$. Following the discussion presented in Inglot and Ledwina (2006d), we propose using the following penalty in this problem:

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

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

$$T1 = \min\{1 \leqslant k \leqslant d\colon\ W_k(\hat{\eta}) - \pi(k, n, c) \geqslant W_s(\hat{\eta}) - \pi(s, n, c),\ s = 1, \ldots, d\}.$$

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 $n \geqslant 8$, $S1 \leqslant T1 \leqslant A1$. Moreover, since under the null model

$$(\mathscr{Y}_1, \ldots, \mathscr{Y}_k) \xrightarrow{\mathscr{D}} N(0, I_k),$$

we have

$$P_\eta^n(T1 \neq S1) \to 0 \quad \text{as } n \to \infty.$$

On the other hand, under H_0, for any $s \in \{2, \ldots, d\}$,

$$P_\eta^n(S1 = s) \leqslant P_\eta^n\big(W_s(\hat{\eta}) \geqslant (s-1)\log n\big).$$

Hence, Proposition 1 yields

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

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

and

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

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). 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 was to propose a practical solution. Therefore, we reduced the technical scope of the paper to the minimum.

Remark 4. The simulation results reported in Inglot and Ledwina (2006a, b) show that the critical values of new data-driven score statistics are very stable for large variety of cases. Anyway, it should be said that the simulated critical values of our solutions are usually slightly 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. Namely, 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_1^2 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, 1997). In the present set-up, to provide a practical and automatic way of generating critical values, one can apply the residual bootstrap, described e.g. on pp. 142–143 of Stute et al. (1998a). We implemented this procedure in our simulation study and found that it works well.

4. PROOF OF THEOREM 1

Obviously, it is enough to show (3.7) for $i \in \langle j \rangle$, $j = 1, 2$. Therefore, we shall restrict attention to the case $j = 1$ and prove that

$$(4.1) \qquad \bigwedge_{\delta>0} P_\eta^n\left(\frac{1}{\sqrt{n}}\left\| \sum_{i\in\langle 1\rangle} [\hat{l}^*(Z_i; \hat{\eta}) - l^*(Z_i; \eta)]\right\| \geqslant \delta\right) \to 0 \quad \text{as } n \to \infty.$$

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 $\mathscr{B}(\beta)$ 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 (4.1) by proving that for any $\{b_n\} \in \mathscr{B}(\beta)$ it follows that

$$(4.2) \qquad \bigwedge_{\delta>0} P_\eta^n\left(\frac{1}{\sqrt{n}}\left\| \sum_{i\in\langle 1\rangle} [\hat{l}^*(Z_i; b_n, \hat{\vartheta}) - l^*(Z_i; \beta, \vartheta)]\right\| \geqslant \delta\right) \to 0 \quad \text{as } n \to \infty.$$

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

$$(4.3) \qquad \bigwedge_{\delta>0} P_\eta^n\left(\frac{1}{\sqrt{n}}\left\| \sum_{i\in\langle 1\rangle} [l^*(Z_i; \beta, \vartheta) - l^*(Z_i; b_n, \vartheta)]\right\| \geqslant \delta\right) \to 0 \quad \text{as } n \to \infty,$$

and therefore, to prove (4.2), it is enough to show that for any $\{b_n\} \in \mathscr{B}(\beta)$ and $\eta = (\beta, \vartheta)$

$$(4.4) \qquad \bigwedge_{\delta>0} P_{(\beta,\vartheta)}^n\left(\frac{1}{\sqrt{n}}\left\| \sum_{i\in\langle 1\rangle} [\hat{l}^*(Z_i; b_n, \hat{\vartheta}) - l^*(Z_i; b_n, \vartheta)]\right\| \geqslant \delta\right) \to 0 \quad \text{as } n \to \infty.$$

• The contiguity of $\{P_{(\beta,\vartheta)}^n\}$ and $\{P_{(b_n,\vartheta)}^n\}$, where $\{b_n\} \in \mathscr{B}(\beta)$, allows us to replace (4.4) by

$$(4.5) \qquad \bigwedge_{\delta>0} P_{(b_n,\vartheta)}^n\left(\frac{1}{\sqrt{n}}\left\| \sum_{i\in\langle 1\rangle} [\hat{l}^*(Z_i; b_n, \hat{\vartheta}) - l^*(Z_i; b_n, \vartheta)]\right\| \geqslant \delta\right) \to 0 \quad \text{as } n \to \infty.$$

• Checking (4.5) 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 $\hat{\beta}^{(2)}$ fall (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 solely on M_γ. The discretization device 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 (4.3), we shall show that

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

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

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

it follows that $P_\eta^n(\|R_n\| \geqslant \delta) \to 0$ 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 6 of our 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. (6.5)). As $l^*(z;\, \beta,\, \vartheta)$, by definition, is (under P_η) orthogonal to the scores for the nuisance parameters, we immediately obtain

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

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

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

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

where

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

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

$$
\begin{aligned}
& r_n(\tilde{\beta}, \beta) \\
& \leqslant 2 \int_R \int_I \left[\frac{f'}{\sqrt{f}}(y - v(x)(\tilde{\beta}-\beta)^T) - \frac{f'}{\sqrt{f}}(y) \right]^2 \|\tilde{u}(x) - \tilde{v}(x) V^{-1} M\|^2 g(x)\, dx dy \\
& \quad + 2 \int_R \int_I \frac{1}{\tau^2} \left[(y - v(x)(\tilde{\beta}-\beta)^T) \sqrt{f(y - v(x)(\tilde{\beta}-\beta)^T)} - y\sqrt{f(y)} \right]^2 g(x)\, dx dy \\
& \quad \times \|m_1 - m_2 V^{-1} M\|^2.
\end{aligned}
$$

It follows from $\langle M1 \rangle$ and $\langle M2 \rangle$ 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 (4.3).

As mentioned above, from Section 6 it follows that the model density $p(z; \beta, \vartheta)$ is Hellinger differentiable at β (cf. Section 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 $\{b_n\} \in \mathscr{B}(\beta)$, are indeed mutually contiguous. This implies that the proof reduces to proving (4.5).

Let us now rewrite (4.5) in a more convenient form. For this purpose let us set

$$
T_{n,s} = \frac{1}{\sqrt{n}} \sum_{i \in \langle 1 \rangle} [\hat{l}^*_s(Z_i; b_n, \hat{\vartheta}) - l^*_s(Z_i; b_n, \vartheta)], \tag{4.6}
$$

where the symbol v_s denotes the sth component of a k-dimensional vector v. Using this notation, we read (4.5) as

$$
T_{n,s} = o_{P^n_{(b_n,\vartheta)}}(1) \quad \text{for each } s = 1, \ldots, k \text{ and } \{b_n\} \in \mathscr{B}(\beta). \tag{4.7}
$$

To check (4.7) we shall apply the following result, which can be obtained by an application of Jensen's inequality and the Lebesgue Dominated Convergence Theorem. For details see Appendix B in Inglot and Ledwina (2006a).

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

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

We shall first prove that

$$
E(T_{n,s} \,|\, \mathscr{F}_n) = 0, \quad s = 1, \ldots, k. \tag{4.8}
$$

Since the conditional density of $(Y_1, \ldots, Y_\zeta)$ under $\mathscr{F}_n$ is of the form

$$\prod_{l=1}^{\zeta} f(y_l - v(X_l) b_n^T),$$

we have for $T_n = (T_{n,1}, \ldots, T_{n,k})$

(4.9) $$E(T_n \mid \mathscr{F}_n)$$

$$= \frac{1}{\sqrt{n}} \sum_{i \in \langle 1 \rangle} \int_R [\hat{l}^*(X_i, y; b_n, \hat{\vartheta}) - l^*(X_i, y; b_n, \vartheta)] f(y - \underline{v}(X_i) b_n^T) dy.$$

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

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

This, a change of variables in the integral (4.9) and another application of $\int_R y f(y) dy = 0$ yield

$$\sqrt{n}\, E(T_n \mid \mathscr{F}_n) = \Big[\int_R \{-[f'/f]^{(2)}(y)\} f(y) dy\Big]$$

$$\times \Big[\sum_{i \in \langle 1 \rangle} u(X_i) - \zeta \hat{m}_1^{(1)} - \Big\{\sum_{i \in \langle 1 \rangle} v(X_i) - \zeta \hat{m}_2^{(1)}\Big\} [\hat{V}^{(2)}]^{-1} \hat{M}^{(2)}\Big].$$

Since, however, $\hat{m}_1^{(1)} = \zeta^{-1} \sum_{i \in \langle 1 \rangle} u(X_i)$ and $\hat{m}_2^{(1)} = \zeta^{-1} \sum_{i \in \langle 1 \rangle} v(X_i)$ we infer that $E(T_n \mid \mathscr{F}_n) = 0$. This proves (4.8).

Therefore, $E(T_{n,s}^2 \mid \mathscr{F}_n) = \mathrm{Var}(T_{n,s} \mid \mathscr{F}_n)$ and, by Proposition 3, to get (4.7) it is enough to check that

(4.10) $$\bigwedge_{\delta > 0} P^n_{(b_n, \vartheta)}(\mathrm{Var}(T_{n,s} \mid \mathscr{F}_n) > \delta) \to 0 \quad \text{as } n \to \infty.$$

However, notice again that, as previously, under $P^n_{(b_n,\vartheta)}$ the conditional density of $(Y_1, \ldots, Y_\zeta)$ with respect to $\mathscr{F}_n$ 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

$$\frac{1}{n} \sum_{i \in \langle 1 \rangle} \int_R [\hat{l}_s^*(X_i, y; b_n, \hat{\vartheta}) - l_s^*(X_i, y; b_n, \vartheta)]^2 f(y - v(X_i) b_n^T) dy.$$

In consequence, after changing the variables in the above integral, (4.10) reads as

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

Now, the artificial b_n is no longer useful and applying the contiguity argument again, we see that to prove (4.10) it is enough to show that

$$\text{(4.11)} \qquad \bigwedge_{\delta>0} P_\eta^n\left(\frac{1}{n}\sum_{i\in\langle 1\rangle}\int_R [\hat{l}_s^*(X_i, y; 0, \hat{\vartheta}) - l_s^*(X_i, y; 0, \vartheta)]^2 f(y)\,dy > \delta\right) \to 0.$$

The rest of the proof consists of showing that each summand appearing in (4.11) is bounded by a common (independent of i) quantity which tends to 0 in probability with respect to P_η^n.

We have

$$\hat{l}_s^*(X_i, y; 0, \hat{\vartheta}) - l_s^*(X_i, y; 0, \vartheta)$$

$$\text{(4.12)} \qquad = y\left(\frac{1}{\hat{\tau}^{(2)}} - \frac{1}{\tau}\right)(m_{1s} - \{m_2 V^{-1} M\}_s)$$

$$\text{(4.13)} \qquad + \frac{y}{\hat{\tau}^{(2)}}(\hat{m}_{1s}^{(1)} - m_{1s} - \{\hat{m}_2^{(1)}[\hat{V}^{(2)}]^{-1}\hat{M}^{(2)}\}_s + \{m_2 V^{-1} M\}_s)$$

$$\text{(4.14)} \qquad -\left[\frac{f'}{f}(y)\right](\tilde{u}_s^{(1)}(X_i) - \{\tilde{v}^{(1)}(X_i)[\hat{V}^{(2)}]^{-1}\hat{M}^{(2)}\}_s - \tilde{u}_s(X_i) + \{\tilde{v}(X_i) V^{-1} M\}_s)$$

$$\text{(4.15)} \qquad -([f'/f]^{(2)}(y) - [f'/f](y))(\tilde{u}_s^{(1)}(X_i) - \{\tilde{v}^{(1)}(X_i)[\hat{V}^{(2)}]^{-1}\hat{M}^{(2)}\}_s).$$

Now we shall consider the integrals of the squared terms from (4.12)–(4.15). 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 (4.12) tends to 0 in probability. In addition, exploiting the consistency of $\hat{m}_1^{(1)}$, $\hat{m}_2^{(1)}$, $\hat{V}^{(2)}$, $\hat{M}^{(2)}$ we see that the same conclusion holds for (4.13).

Now, rewrite (4.14) as follows:

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

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

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

Finally, consider (4.15). 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 $\hat{V}^{(2)}$ and $\hat{M}^{(2)}$ reduces the problem to showing that the $L_2(R, \Pi_f)$-norm of $([f'/f]^{(2)} - f'/f)$ tends to 0 in P_η^n. However, this is just our assumption (3.11).

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

Remark 5. The proof of Theorem 1 shows that to get the key result $W_k(\hat{\eta}) \xrightarrow{\mathscr{D}} \chi^2_k$, using several steps, the problem can be reduced to checking the following conditions:

$$(*) \qquad \sum_{i \in \langle j \rangle} \int_R \hat{l}^*(X_i, y; b_n, \hat{\vartheta}) f\left(y - v(X_i) b_n^T\right) \lambda(dy) = 0$$

and

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

for $j = 1, 2$ and every sequence $\{b_n\} \in \mathscr{B}(\beta)$. In the considered problem, 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).

5. DISCUSSION OF THE GENERAL ASSUMPTIONS

The assumption on the compactly supported density g was imposed for technical convenience. The restriction $g > 0$ λ-a.e. guarantees that the matrix W is positive definite. Obviously, both the assumptions are not necessary ones.

Extensions to multivariate explanatory variable seem to be rather straightforward.

The assumption that the length d of the list of models is independent of n substantially simplifies the considerations. From the practical point of view, fixing this number seems to be reasonable. Besides, the very important property of our solution is 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. 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. Possible introducing $d = d(n)$, $d(n) \to \infty$ as $n \to \infty$, in our constructions has however some aesthetical aspect. Namely, for a 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 detail in Inglot et al. (1997).

6. 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 (2003). 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

$$\mathfrak{M}(k) \qquad Y = u(X)\,\theta^T + v(X)\,\beta^T + \varepsilon$$

and define

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

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

$$p(z; \kappa) = g(x)\, f\big(y - w(x)\, a^T\big). \tag{6.1}$$

Observe that $p^{1/2}(z; \kappa)$, seen as a function of κ, is a map from $\Omega \to \mathscr{H}$, where $\Omega = \mathscr{A} \times \mathscr{B} \times \mathscr{C}$, while $\mathscr{A} = R^{k+q}$, $\mathscr{B} = L_2(I, \lambda)$, $\mathscr{C} = L_2(R, \lambda)$, and $\mathscr{H} = L_2(I \times R, \lambda \times \lambda)$. Let $\|\cdot\|_{\mathscr{H}}$ denote the related norm in $\mathscr{H}$. The specific structure of $p^{1/2}(z; \kappa)$ (cf. (6.1)) motivates the introduction of an abstract map

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

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

$$\Delta_\varphi h(x, y) = h\big(x, y - \varphi(x)\big), \qquad x \in I,\ y \in R,\ h \in \mathscr{H}. \tag{6.2}$$

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:

$$\dot{\Phi}_{(a,b,c)}(a_0, b_0, c_0) = \Delta_{wa^T}(-bc'\,[wa_0^T] + b_0\, c + bc_0) \tag{6.3}$$

(cf. Theorem B.11 in Inglot and Ledwina (2003)). 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 $\mathscr{H}$;

(ii) for each $h \in \mathscr{H}$,

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

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 $h \in \mathscr{H}$ it follows that

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

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) as well as other useful properties of Δ_φ and related scale operators see Inglot and Ledwina (2003), Section A. Note also that some general shift operators were studied in the Appendix of Koul and Schick (1996).

Consider now the question of the differentiability of $p^{1/2}(\cdot; \kappa)$ 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 $b_n \in \mathscr{B}$, $b_n \to b_0 \in \mathscr{B}$, 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 $b \in \mathscr{B}$, define $\mathscr{B}_0 \subset \mathscr{B}$ by

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

Analogously, taking $c = \sqrt{f}$, $c_n \to c_0 \in \mathscr{C}$, $t_n \to 0$ such that for large n

$$\int_R [c + t_n c_n]^2 d\lambda = 1 \quad \text{and} \quad \int_R \imath [c + t_n c_n]^2 d\lambda = 0, \quad \text{where } \imath(y) = y$$

(cf. the model assumptions $\langle M1 \rangle$), one can easily infer that c_0 has to belong to the subspace

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

Set $\Omega_0 = \mathscr{A} \times \mathscr{B}_0 \times \mathscr{C}_0$. Take f and g satisfying $\langle M1 \rangle$ 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 (6.3) is applicable at $\omega = (a, b, c) = \kappa$ and the following holds:

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

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

$$\dot{s}_\kappa(\cdot) = \frac{\dot{\Phi}_\kappa(\cdot)}{\frac{1}{2} p^{1/2}(\cdot; \kappa)}. \tag{6.4}$$

This, together with (6.3), implies that the operator $\dot{s}_\kappa(\cdot)$ is defined by the vector

$$\Delta_{waT}\left(-u\left[\frac{f'}{f}\right], -v\left[\frac{f'}{f}\right], \frac{1}{\sqrt{f}}, \frac{1}{\sqrt{g}}\right). \tag{6.5}$$

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 (6.5) onto the subspace spanned by the remaining components of (6.5). 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 $\Delta_{waT}(-bc'w, c, b)$ onto the subspace

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

in the standard space $\mathscr{H} = L_2(I \times R, \lambda \times \lambda)$ and, at the final stage, to divide the resulting expressions by $\frac{1}{2}p^{1/2}(\cdot; \kappa)$. To calculate projections in $\mathscr{H}$, 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 (2003).

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. The authors thank a referee for helpful remarks. This research was partially supported by grant 5 P03A 03020 from KBN.

REFERENCES

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

[2] A. Azzalini and A. Bowman, *On use of nonparametric regression for checking linear relationships*, J. Roy. Statist. Soc. B 55 (1993), pp. 549–557.

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

[4] P. J. Bickel, *On adaptive estimation*, Ann. Statist. 10 (1982), pp. 647–671.

[5] P. J. Bickel, C. A. Klaassen, Y. Ritov and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, John Hopkins Univ. Press, Baltimore 1993.

[6] P. J. Bickel, Y. Ritov and T. Stocker, *Testing and the method of sieves*, Technical Report, University of California, Berkeley, 1998.

[7] S. Choi, *On asymptotically optimal tests*, Ph.D. Thesis, Dept. Statistics, University of Rochester, Rochester, 1989.
[8] S. Choi, W. J. Hall and A. Schick, *Asymptotically uniformly most powerful tests in parametric and semiparametric models*, Ann. Statist. 24 (1996), pp. 841–861.
[9] D. Cox, E. Koh, G. Wahba and B. Yandell, *Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models*, Ann. Statist. 16 (1988), pp. 113–119.
[10] H. Dette, *On a nonparametric test for linear relationships*, Statist. Probab. Lett. 46 (2000), pp. 307–316.
[11] J. Diebolt and J. Zuber, *On the half-sample method for goodness-of-fit in regression*, Technical Report 00.1, Department of Mathematics and Statistics, La Trobe University, 2000.
[12] R. L. Eubank, J. D. Hart and V. N. LaRiccia, *Testing goodness of fit via nonparametric function estimation techniques*, Comm. Statist. Theory Methods 22 (1993), pp. 3327–3354.
[13] J. Fan, *Test of significance based on wavelet thresholding and Neyman's truncation*, J. Amer. Statist. Assoc. 91 (1996), pp. 674–688.
[14] J. Fan and L.-S. Huang, *Goodness-of-fit tests for parametric regression models*, J. Amer. Statist. Assoc. 96 (2001), pp. 640–652.
[15] E. Guerre and P. Lavergne, *Data-driven rate-optimal specification testing in regression models*, Ann. Statist. 33 (2005), pp. 840–870.
[16] J. Hájek and Z. Šidák, *Theory of Rank Tests*, Academia, Prague 1967.
[17] W. Härdle and E. Mammen, *Comparing nonparametric versus parametric regression fits*, Ann. Statist. 21 (1993), pp. 1926–1947.
[18] J. D. Hart, *Nonparametric Smoothing and Lack-of-fit Tests*, Springer, New York 1997.
[19] J. L. Horowitz and V. G. Spokoiny, *An adaptive, rate optimal test of a parametric mean-regression model against a nonparametric alternative*, Econometrica 69 (2001), pp. 599–631.
[20] T. Inglot, W. C. M. Kallenberg and T. Ledwina, *Data driven smooth tests for composite hypotheses*, Ann. Statist. 25 (1997), pp. 1222–1250.
[21] T. Inglot and T. Ledwina, *Asymptotic optimality of data driven Neyman's tests for uniformity*, Ann. Statist. 24 (1996), pp. 1982–2019.
[22] T. Inglot and T. Ledwina, *Intermediate approach to comparison of some goodness-of-fit tests*, Ann. Inst. Statist. Math. 53 (2001a), pp. 810–834.
[23] T. Inglot and T. Ledwina, *Asymptotic optimality of data driven smooth tests for location-scale family*, Sankhyā, Ser. A 63 (2001b), pp. 41–71.
[24] T. Inglot and T. Ledwina, *Semiparametric regression: Hadamard differentiability and efficient score functions for some testing problems*, Preprint 012, Institute of Mathematics, Wrocław University of Technology, 2003.
[25] T. Inglot and T. Ledwina, *Data driven score tests for homoscedastic linear regression model*, Preprint 665, Institute of Mathematics of the Polish Academy of Sciences, 2006a.
[26] T. Inglot and T. Ledwina, *Data driven score tests for homoscedastic linear regression model: the construction and simulations*, in: *Prague Stochastics 2006*, Proceedings, M. Hušková and M. Janžura (Eds.), matfyzpress, Prague 2006b, pp. 124–137.
[27] T. Inglot and T. Ledwina, *Intermediate efficiency of some max-type statistics*, J. Statist. Plann. Inference 136 (2006c), pp. 2918–2935.
[28] T. Inglot and T. Ledwina, *Towards data driven selection of a penalty function for data driven Neyman tests*, Linear Algebra Appl. 417 (2006d), pp. 124–133.
[29] T. Inglot and T. Ledwina, *Asymptotic optimality of new adaptive test in regression model*, Ann. Inst. H. Poincaré (2006e).
[30] H. S. Javitz, *Generalized smooth tests of goodness of fit, independence and equality of distributions*, Ph. D. Thesis, Univ. California, Berkeley, 1975.
[31] W. C. M. Kallenberg and T. Ledwina, *On data driven Neyman's tests*, Probab. Math. Statist. 15 (1995), pp. 409–426.

[32] W. C. M. Kallenberg and T. Ledwina, *Data-driven smooth tests when the hypothesis is composite*, J. Amer. Statist. Assoc. 92 (1997), pp. 1094–1104.
[33] E. V. Khmaladze and H. L. Koul, *Martingale transforms goodness-of-fit tests in regression models*, Ann. Statist. 32 (2004), pp. 995–1034.
[34] R. Koenker and Z. Xiao, *Inference on the quantile regression process*, Econometrica 70 (2002), pp. 1583–1612.
[35] H. L. Koul and A. Schick, *Adaptive estimation in a random coefficient autoregressive model*, Ann. Statist. 24 (1996), pp. 1025–1052.
[36] A. S. Kozek, *A nonparametric test of fit of a parametric model*, J. Multivariate Anal. 37 (1991), pp. 66–75.
[37] J.-P. Kreiss, *On adaptive estimation in stationary ARMA processes*, Ann. Statist. 15 (1987), pp. 112–133.
[38] L. Le Cam, *On the asymptotic theory of estimation and testing hypotheses*, Proc. Third Berkeley Symp. Math. Statist. Probab. 1, Univ. California Press, Berkeley 1956, pp. 129–156.
[39] L. Le Cam and E. L. Lehmann, *J. Neyman — On the occasion of his 80th birthday*, Ann. Statist. 2 (1974), pp. vii–xiii.
[40] T. Ledwina, *Data driven version of Neyman's smooth test of fit*, J. Amer. Statist. Assoc. 89 (1994), pp. 1000–1005.
[41] J. Neyman, *"Smooth test" for goodness of fit*, Skand. Aktuarietidskr. 20 (1937), pp. 149–199.
[42] J. Neyman, *Sur une famille de tests asymptotiques des hypothèses statistiques composées*, Trabajos de Estadistica 5 (1954), pp. 161–168.
[43] J. Neyman, *Optimal asymptotic tests of composite statistical hypotheses*, in: *Probability and Statistics: The Harald Cramér Volume*, U. Grenander (Ed.), Wiley, New York 1959, pp. 213–234.
[44] J. C. W. Rayner and D. J. Best, *Smooth Tests of Goodness of Fit*, Oxford Univ. Press, New York 1989.
[45] A. Schick, *On asymptotically efficient estimation in semiparametric models*, Ann. Statist. 14 (1986), pp. 1139–1151.
[46] A. Schick, *Efficient estimates in linear and nonlinear regression with heteroscedastic error*, J. Statist. Plann. Inference 58 (1997), pp. 371–387.
[47] A. Schick, *On asymptotic differentiability of averages*, Statist. Probab. Lett. 51 (2001), pp. 15–23.
[48] J. S. Simonoff and C. L. Tsai, *Semiparametric and additive model selection using an improved Akaike Information Criterion*, J. Comput. Graph. Statist. 8 (1999), pp. 22–40.
[49] W. Stute, *Nonparametric model checks for regression*, Ann. Statist. 25 (1997), pp. 613–641.
[50] W. Stute, W. G. Manteiga and M. P. Quindimil, *Bootstrap approximation in model checks in regression*, J. Amer. Statist. Assoc. 93 (1998a), pp. 141–149.
[51] W. Stute, S. Thies and L.-X. Zhu, *Model checks for regression: an innovation process approach*, Ann. Statist. 26 (1998b), pp. 1916–1934.
[52] D. Thomas and D. Pierce, *Neyman's smooth goodness-of-fit test when the hypothesis is composite*, J. Amer. Statist. Assoc. 74 (1979), pp. 441–445.
[53] A. W. van der Vaart, *On differentiable functionals*, Ann. Statist. 19 (1991), pp. 178–204.

Tadeusz Inglot
Institute of Mathematics and Informatics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
and
Polish Academy of Sciences
E-mail: inglot@im.pwr.wroc.pl

Teresa Ledwina (corresponding author)
Institute of Mathematics
Polish Academy of Sciences
ul. Kopernika 18
51-617 Wrocław, Poland
E-mail: ledwina@impan.pan.wroc.pl

Received on 19.3.2006