Data-Driven Smooth Tests for a Location-Scale Family Revisited

A. Janic, *Institute of Mathematics and Informatics, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland.*
Email: Alicja.Janic@pwr.wroc.pl

T. Ledwina, *Institute of Mathematics, Polish Academy of Sciences, ul. Kopernika 18, 51-617 Wrocław, Poland.*
Email: ledwina@impan.pan.wroc.pl

Abstract

A new data-driven, smooth goodness of test for a location-scale family is proposed and studied. The new test statistic is a combination of an efficient score statistic and an appropriate selection rule. Some examples are presented and by using extensive simulations the test is shown to have desirable properties.

AMS Subject Classification: 62G10, 62G20.

Key Words: Data-driven test, efficient score statistic, goodness-of-fit test, location-scale family, model selection, Schwarz's rule, test for extreme value distribution, test for normality.

---

1. Introduction

Location-scale families play an important role in modeling data. Goodness of fit tests for such families have been the subject of numerous papers. Various concepts for constructing such tests have been applied. For example, see Thomas and Pierce (1979), LaRiccia and Mason (1985), D'Agostino and Stephens (1986), Rayner and Best (1989), LaRiccia (1991), Morales et al. (1993), Stute et al. (1993), Fromont and Laurent (2006), Meintanis and Swanepoel (2007) and references therein. Inglot et al. (1997) and Janic-Wróblewska (2004 a) investigated some data-driven smooth tests. The test statistic is a score statistic based on an appropriate parametric model with the dimension of the model determined by a suitable selection rule. In these two papers attention was focused on a Schwarz-type selection rule and some simplifications. For more discussion see Sec. 3.5 of Janic-Wróblewska (2004 a).

Some other tests for testing goodness of fit to a parametric model that use model selection criteria have also been proposed. These include, in particular, tests based on a score or likelihood ratio test (LRT) statistic combined with Akaike's information criterion and Schwarz's rule applied to some nested, as well as non-nested lists, of models. For more details, see Aerts et al. (1999, 2000) and Claeskens and Hjort (2004).

The purpose of this paper is to apply and investigate a more convenient selection rule, mimicking Schwarz's criterion, and the related data-driven test. We consider a so called score-based selection rule, a special case of which was considered in Inglot et al. (1994) and Kallenberg and Ledwina (1997 a); see also the brief comment on p. 1240 of Inglot et al. (1997). Advantages of this rule are its relative simplicity, flexibility and wide applicability. Though the idea is not new, the technical and practical aspects of this test have not been fully investigated in the case of location–scale families. The aim of the paper is to fill this gap. For further remarks see Sec. 2.1. The score-based selection rule and related data-driven test, restricted to the present setting, are introduced in Sec. 2 and discussed in Sec. 4, while some encouraging simulation results are presented in Sec. 3. In Sec. 3 we consider applications of the general approach to testing in the cases of extreme value and normal distributions. These new tests appear to be powerful when compared to the Anderson-Darling and Shapiro-Wilk tests, widely recommended as the best existing solutions for the two testing problems considered. Proofs and some technical remarks are presented in Appendices A and B. Some arguments crucial to the asymptotic analysis of some data-driven smooth tests with Euclidean nuisance parameters were developed in Inglot et al. (1997) and Inglot and Ledwina (2001). In the case of location-scale families this technique was further studied and simplified in Janic-Wróblewska (2004 a). In this article we adapt the arguments used in these papers, to obtain the asymptotic results appropriate to the present setting.

2. Basic results

We start by introducing some necessary notation. Let $X_1, ..., X_n$ be independent and identically distributed random variables with density function $g(x),\ x \in R$. Set $\beta = (\beta_1, \beta_2),\ \beta \in \mathcal{B} = R \times R_+$ and

$$f(x;\beta) = \frac{1}{\beta_2} f\Big(\frac{x-\beta_1}{\beta_2}\Big), \tag{2.1}$$

where f is a density function on R. Given f, consider testing the hypothesis

$$\mathcal{H}_0 : g(x) \in \Big\{ f(x;\beta),\ \beta \in \mathcal{B} \Big\}.$$

To construct a data-driven test statistic for $\mathcal{H}_0$, we embed $f(x;\beta)$ into a larger parametric family. For this purpose set $F(x;\beta)$ to be the cdf corresponding to $f(x;\beta)$ and consider the system $\phi_0 \equiv 1, \phi_1, \phi_2, ...$ of orthonormal Legendre polynomials on [0,1]. Now, for a given natural number k, we define an auxiliary parametric family

$$g_k(x;\eta) = c_k(\theta)\exp\Big\{\sum_{j=1}^{k}\theta_j\phi_j(F(x;\beta))\Big\}f(x;\beta),$$

where $\theta = (\theta_1, ..., \theta_k) \in R^k$, $\eta = (\theta, \beta)$ and $c_k(\theta)$ is the normalizing constant. In the family $g_k(x;\eta)$ the hypothesis $\mathcal{H}_0$ is equivalent to

$$\mathcal{H}_0^*(k) : \eta = \eta_0 = (0, \beta).$$

Set

$$\ell(x;\eta) = \log g_k(x;\eta), \quad \dot{\ell}_\eta(x;\eta) = \frac{\partial}{\partial\eta}\ell(x;\eta) = \Big(\dot{\ell}_\theta(x;\eta), \dot{\ell}_\beta(x;\eta)\Big) \tag{2.2}$$

and

$$I(\eta_0) = E_{\eta_0}[\dot{\ell}_\eta(X_1;\eta_0)]^T[\dot{\ell}_\eta(X_1;\eta_0)] = \begin{pmatrix} I_{\theta\theta}(\eta_0) & I_{\theta\beta}(\eta_0) \\ I_{\beta\theta}(\eta_0) & I_{\beta\beta}(\eta_0) \end{pmatrix}, \tag{2.3}$$

where ^{T} denotes transposition. Assume that $I(\eta_0)$ is invertible. Note also that the following hold for the family $g_k(x;\eta)$

$$\dot{\ell}_\theta(x;\eta_0) = \Big(\phi_1(F(x;\beta)), ..., \phi_k(F(x;\beta))\Big) \quad \text{and} \quad \dot{\ell}_\beta(x;\eta_0) = \frac{\partial}{\partial\beta}\log f(x;\beta). \tag{2.4}$$

The efficient score vector for $\mathcal{H}_0^*(k)$ is defined by

$$\ell^*(x;\eta_0) = \dot{\ell}_\theta(x;\eta_0) - \dot{\ell}_\beta(x;\eta_0)[I_{\beta\beta}(\eta_0)]^{-1}I_{\beta\theta}(\eta_0).$$

Now define

$$I^*(\eta_0) = E_{\eta_0}[\ell^*(X_1;\eta_0)]^T[\ell^*(X_1;\eta_0)]. \tag{2.5}$$

Note that in the case of a location-scale family, $I^*(\eta_0)$ does not depend on η_0, or equivalently on β. Moreover, $I^*(\eta_0)$ is invertible, as $I(\eta_0)$ is, and the following holds

$$\Big[I^*(\eta_0)\Big]^{-1} = I + I_{\theta\beta}(\eta_0)\Big[I_{\beta\beta}(\eta_0) - I_{\beta\theta}(\eta_0)I_{\theta\beta}(\eta_0)\Big]^{-1}I_{\beta\theta}(\eta_0),$$

where I is the $k \times k$ identity matrix.

Now set $\tilde{\eta}_0 = (0, \tilde{\beta})$, where $\tilde{\beta}$ is an estimator of β. We shall also use the following abbreviated notation

$$\ell(x;\eta_0) = \ell(x;\beta), \quad \dot{\ell}_\eta(x;\eta_0) = \Big(\dot{\ell}_\theta(x;\beta), \dot{\ell}_\beta(x;\beta)\Big), \quad I^*(\eta_0) = I^*,$$

$$\ell^*(x;\eta_0) = \ell^*(x;\beta) = \Big(\ell_1^*(x;\beta), ..., \ell_k^*(x;\beta)\Big), \quad \ell^*(x;\tilde{\eta}_0) = \ell^*(x;\tilde{\beta}). \tag{2.6}$$

Using this notation, the efficient score statistic for $\mathcal{H}_0^*(k)$ is given by

$$W_k^*(\tilde{\beta}) = \Big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^*(X_i;\tilde{\eta}_0)\Big]\Big[I^*(\tilde{\eta}_0)\Big]^{-1}\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^*(X_i;\tilde{\eta}_0)\Big]^T$$

$$= \Big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^*(X_i;\tilde{\beta})\Big]\Big[I^*\Big]^{-1}\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^*(X_i;\tilde{\beta})\Big]^T.$$

The notation used here is similar to that used typically in the theory of efficient estimation and is different from that used earlier in Inglot et al. (1997) and Janic-Wróblewska (2004 a).

Efficient score statistics and tests were introduced by Neyman (1954, 1959) and further extended by Bühler and Puri (1966). Javitz (1975) studied, among other things, the case of testing $\mathcal{H}_0^*(k)$. Javitz called such a solution a generalized smooth test. Neyman (1959) called such solutions $C(\alpha)$ tests. For comments on some counterparts of $C(\alpha)$ tests in the econometrics literature see Sec. 3.3 of Bera and Bilias (2001 a). To close this brief discussion we quote the opinion of Le Cam and Lehmann (1974) on $C(\alpha)$ tests: "They can be applied to a variety of complex problems which can hardly be tackled otherwise and constitute a delightfully growing chapter of asymptotic theory."

2.1. A score-based selection rule and data-driven smooth test

The family $g_k(x;\eta)$ serves to model possible deviations from $\mathcal{H}_0$. Given k, $W_k^*(\tilde{\beta})$ is an asymptotically optimal test for $\mathcal{H}_0$ within g_k, i.e. for $\mathcal{H}_0^*(k)$, cf. Bühler and Puri (1966), Sec. 3.4 of Rayner and Best (1989), as well as Bera and Bilias (2001 b). However, the crucial question is what value of k should be used in practice. Figures 1 and 2 in Kallenberg and Ledwina (1997 b) clearly illustrate the essence of the problem. Ledwina (1994) suggested using the Schwarz selection rule to solve a similar problem in a simpler setting. For $g_k(x;\eta)$ a selection rule mimicking Schwarz's original solution was proposed and investigated in Inglot et al. (1997) and Kallenberg and Ledwina (1997 b). This solution requires extensive numerical calculations. Therefore, some simplifications were proposed, cf. ibidem. After a further decade of research, it seems that the most natural simplification, in the case of a nested list, is the following selection rule

$$S1 = S1(\tilde{\beta}) = \min\Big\{k,\ k = 1, ..., d(n) : W_k^*(\tilde{\beta}) - k\log n \geq W_j^*(\tilde{\beta}) - j\log n,\ j = 1, ..., d(n)\Big\},$$

where $d(n)$ is the number of models in the list. One motivation for $S1$ is that a good approximation of the LRT statistic under the null hypothesis and local alternatives is given by 1/2 times the efficient score statistic. For a justification of this fact see Javitz (1975), Sec. 13.

The rule $S1$ defines a data-driven smooth test or, equivalently, a data-driven efficient score test, using the test statistic

$$W_{S1}^*(\tilde{\beta}) = W_{S1(\tilde{\beta})}^*(\tilde{\beta}).$$

The rule $S1$ was introduced in Inglot et al. (1994) for the case where $\tilde{\beta}$ is the maximum likelihood estimator and later investigated using simulations in Kallenberg and Ledwina (1997 a). Aerts et al. (2000) also recommended $S1$ using maximum likelihood estimators. In such cases the efficient score vector reduces to an ordinary score vector ℓ_θ and W_k^* is corresponds to the standard Rao score statistic, which is the Lagrangian multiplier statistic, in the econometric literature. Further articles, e.g. Kallenberg and Ledwina (1999) and Inglot and Ledwina (2006 a, b), show that using efficient scores to construct data-driven tests is useful in various settings. For some general remarks on the construction of efficient score vectors we refer the reader to Bickel et. al. (2006) and Inglot and Ledwina (2006 a). Bickel et al. (2006) also present some applications of efficient score vectors in the construction of "tailor-made tests".

2.2. Asymptotic results under $\mathcal{H}_0$

To introduce the assumptions, we start with some auxiliary notation. Throughout let $||\cdot||$ be the Euclidean norm in an appropriate space. Let P_β stand for the distribution pertaining to $f(x;\beta)$. Set

$$\beta^o = (0,1), \quad f(x) = f(x;\beta^o), \quad F(x) = F(x;\beta^o), \quad \mathcal{L}(x) = \log f(x), \quad \mathcal{L}^{(l)}(x) = \frac{\partial^l}{\partial x^l}\mathcal{L}(x). \tag{2.7}$$

We impose the following restrictions.

(A1) $\int_R |x|^{4+\delta} f(x)dx < \infty$ for some $\delta > 0$;
(A2) $\mathcal{L}^{(l)},\ l = 1,2,3$, exist almost everywhere and the first derivative of f is continuous;
(A3) $\int_R (1+x^4)[\mathcal{L}^{(1)}(x)]^4 f(x)dx < \infty$;
(A4) $\int_R (1+|x|^l)\sup_{\{||\beta-\beta^o||<\eta\}} \left|\mathcal{L}^{(l)}([x-\beta_1]/\beta_2)\right| f(x)dx < \infty$ for some $\eta \in (0,1)$ and $l = 1,2,3$.
(A5) $\int_R (1+x^4)[\mathcal{L}^{(2)}(x)]^2 f(x)dx < \infty$;
(A6) The matrix $I(\eta_0)$ is nonsingular and the largest eigenvalue $\lambda(k)$ of $[I^*]^{-1}$ satisfies $\lambda(k) = O(k^v)$ for some positive v;
(A7) $\tilde\beta$ is location-scale equivariant;
(A8) There exist positive constants c_1, c_2 and n_1 such that

$$P_\beta\Big(\sqrt{n}||\tilde\beta - \beta|| \geq x\Big) \leq c_1 \exp\Big\{-c_2 x^2\Big\} \tag{2.8}$$

for $x = \rho\sqrt{\log n}$ with $\rho > 0$ and $n \geq n_1$.

Note that (A1)-(A4), together with the assumed positive definiteness of $I_{\beta\beta}(\eta_0)$, imply (R1)-(R3) in Inglot et al. (1997). (A8) is a slightly stricter version of (R4) from that paper.

Theorem 1. Assume (A1)-(A8) hold, $\mathcal{H}_0$ is true and $d(n) = o([\log n]^{1/v}[\log\log n]^{-1})$. Then

$$\lim_{n\to\infty} P_\beta\Big(S1(\tilde\beta) \geq 2\Big) = 0.$$

The proof of Theorem 1 is given in Appendix B.

Corollary 1. Assume that the conditions given in Theorem 1 hold. Then, asymptotically, $W^*_{S1}(\tilde{\beta})$ has a central chi-square distribution with 1 degree of freedom.

2.3. Consistency of the data-driven test based on $W^*_{S1}(\tilde{\beta})$

We start by introducing an auxiliary assumption and related "artificial" parameter.

(A9) If $X_1, ..., X_n$ are i.i.d. and $X_1 \sim I\!P$, then there exists $\beta = \beta(I\!P) = \Big(\beta_1(I\!P), \beta_2(I\!P)\Big)$ such that

$$||\tilde{\beta} - \beta(I\!P)|| \xrightarrow{I\!P} 0 \quad \text{as} \quad n \to \infty.$$

So, the new parameter $\beta(I\!P)$ is defined via the distribution $I\!P$ and the estimator $\tilde{\beta}$. Note that under the null model P_β, the following holds for $k \in N$: $E_{P_\beta}\Big(\ell^*_1(X_1;\beta), ..., \ell^*_k(X_1;\beta)\Big) = 0 \in R^k$, where N stands for the set of natural numbers. Therefore, we shall say that $I\!P$ is the alternative (to $P_{\beta(I\!P)}$) if there exists $k = k(I\!P) \in N$ such that $E_{I\!P}\ell^*_1(X_1;\beta(I\!P)) = ... = E_{I\!P}\ell^*_{k-1}(X_1;\beta(I\!P)) = 0$ and $E_{I\!P}\ell^*_k(X_1;\beta(I\!P)) \neq 0$.

Theorem 2. Under the assumptions (A1)-(A9), a test which rejects $\mathcal{H}_0$ for large realisations of $W^*_{S1}(\tilde{\beta})$ is consistent against any alternative $I\!P$.

Some arguments related to the proof of Theorem 2 are found in Appendix C.

2.4. Some comments

Remark 1. The assumption regarding the equivariance of $\tilde{\beta}$ is natural and was imposed to simplify the formulations of the assumptions, as well as proofs. Analogous results hold without this restriction; cf. Inglot et al. (1997).

Remark 2. To reduce the number of assumptions and simplify the notation, as in Janic-Wróblewska (2004 a, b), we restricted attention to the Legendre system when modeling alternatives. Extensions to many other systems are possible; cf. e.g. Bogdan and Ledwina (1996), as well as Inglot et al. (1997).

Remark 3. Assumptions (A1)-(A5) are easy to check for most standard f, including the extreme value and normal distributions.

Remark 4. To calculate v, which appears in (A6), one can use the results from Section 5.10 of Inglot and Ledwina (2001). It follows that $\lambda(k) = O(k^2)$ for testing exponentiality when Legendre polynomials are used. However, for other distributions the calculations involved are laborious. We briefly discuss this point in Appendix A. See also the last paragraph of Remark 5.

Remark 5. Checking (A8) requires the most technically involved calculations. However, many results in this area are already known. For selected results see e.g. Rubin and Sethuraman (1965), Inglot and Ledwina (1993), Inglot and Kallenberg (2003), Baklanov and Borisov (2003) and references therein.

In particular, in the case of moment estimators, to construct a test for the extreme value distribution, one can apply (B.3) on p. 349 in Janic-Wróblewska (2004 a), inequality (101) from Rubin and Sethuraman (1965) and their comment following it, as all moments exist. Also, the method based on probability-weighted moments, which we may use for the same problem, leads to simple estimators for which Theorem 6 of Rubin and Sethuraman (1965) is applicable.

Our test of normality, presented below, exploits the estimator of scale parameter developed in Chen and Shapiro (1995). Using results from Bai and Chen (2003), one can reduce checking that (A8) holds to investigate the probability of moderate deviations for

$$\bar{\sigma}_n = \frac{1}{n}\sum_{i=1}^{n} \Phi^{-1}\Big(\frac{1}{n}\Big) X_{n:i}, \tag{2.9}$$

where $X_{n:1} \leq ... \leq X_{n:n}$ are order statistics for n observations of a normal random variable, while Φ^{-1} is the inverse cdf of the $N(0,1)$ df. To the best of our knowledge, there is no such proof available in the existing literature. In general, studies of L-statistics with such heavy weights are particularly difficult. See e.g. Ch. 19 of Shorack and Wellner (1986) for an overview of the techniques used in analysing the asymptotic behaviour of such structures. Proving such a result is beyond the scope of this contribution.

Note also that assumption (A8) on the probability of moderate deviations of $\tilde{\beta}$ is crucial in proving that under $\mathcal{H}_0$, the selection rule chooses a one dimensional model when $d(n) \to \infty$ (Theorem 1) with probability 1. On the other hand, proving consistency against essentially any alternative (Theorem 2) is possible only if $d(n) \to \infty$. From practical point of view, one should be satisfied with consistency for a smaller class of alternatives, e.g. all distributions with vanishing Fourier coefficients for all indexes greater than some fixed, but otherwise arbitrary number, say D, then taking $d(n) = D, n = 1, 2, ...$ greatly simplifies the proofs and assumptions. In particular, the rate of growth of $\lambda(k)$ in (A6) is immaterial, while (A8) can be replaced by the $\sqrt{n}$-consistency of $\tilde{\beta}$.

Remark 6. Assumption (A9) was introduced and discussed in Inglot el al. (1994, 1997) and used in various forms in later papers.

3. Two illustrative examples

3.1. Testing for the extreme value distribution

This problem was extensively studied and discussed in Janic-Wróblewska (2004 b). For completeness we recall that in this case

$$f(x;\beta)=\frac{1}{\beta_2}\exp\Big\{\frac{x-\beta_1}{\beta_2}-\exp\Big(\frac{x-\beta_1}{\beta_2}\Big)\Big\},\quad x\in R. \tag{3.1}$$

As in the above paper, we use method of moments estimator $\tilde\beta=\tilde\beta[m]$ of $\beta=(\beta_1,\beta_2)$, given by

$$\tilde\beta[m]=\Big(\bar X+\gamma T,T\Big)\quad\text{with}\quad \bar X=\frac{1}{n}\sum_{i=1}^n X_i,\;\; T=\frac{\sqrt6}{\pi}\Big\{\frac1n\sum_{i=1}^n(X_i-\bar X)^2\Big\}^{1/2},$$

where γ is the Euler constant. Moreover, we also considered the estimator $\tilde\beta=\tilde\beta[pwm]$ derived by the method of probability-weighted moments, which was introduced by Landwehr and Matalas (1979). In the case of the extreme value distribution (3.1), it is of the form

$$\tilde\beta[pwm]=\Big(-\bar X+\gamma G,G\Big)\quad\text{with}\quad G=[n(n-1)\ln 2]^{-1}\sum_{1\le j<i\le n}\Big(X'_{n:i}-X'_{n:j}\Big),$$

where $X'_{n:1}\le ...\le X'_{n:n}$ are the order statistics derived from $-X_1,...,-X_n$ (cf. Greenwood et al. 1979, as well as Landwehr and Matalas 1979, Table 2 and (A.10), or Hosking et al. 1985).

The figures below present the empirical powers of the Anderson-Darling statistic (A^2), widely accepted to be the best existing solution in this case, and of the tests based on $W^*_{S1}(\tilde\beta)$ using the estimators described above. Inspection of Figures 1–3 in Janic-Wróblewska (2004 b) and the related comments enables many other comparisons. For brevity, we focus only on the case $n=50$. The significance level α is equal to 0.05 throughout. The simulated critical values and powers are based on 10 000 Monte Carlo runs. The simulated critical values for A^2, $W^*_{S1}(\tilde\beta[m])$ and $W^*_{S1}(\tilde\beta[pwm])$ are 0.7634, 9.9225 and 6.2686, respectively. The empirical critical values of $W^*_{S1}(\tilde\beta)$ considerably exceed the asymptotic values. This is a typical feature of such constructions. For discussion and some approximations in special cases, see Kallenberg and Ledwina (1995, 1997 b). A description of the various alternatives considered and corresponding abbreviations is provided in Table 1. The symbol Z denotes an $N(0,1)$ r.v. and φ its density. U stands for an r.v. with uniform distribution on (0,1). We omit the definitions of standard distributions, such as Weibull, Beta, Stable. Detailed descriptions may be found e.g. in Kallenberg and Ledwina (1997 a).

Table 1. Description of some of the alternatives used in simulations

Symbol	Density/Definition
LC$(p;m)$	$p\varphi(x-m)+(1-p)\varphi(x),\ x\in R$
LG$(p;q)$	$q^{-p}\{\Gamma(p)\}^{-1}\exp\{px-q^{-1}\exp(x)\},\ x\in R$
SB$(g;d)$	$X=\exp\{d^{-1}(Z-g)\}[1+\exp\{d^{-1}(Z-g)\}]^{-1}$
SC$(p;d)$	$d^{-1}p\varphi\left(d^{-1}x\right)+(1-p)\varphi(x),\ x\in R$
SU$(g;d)$	$X=\sinh\{d^{-1}(Z-g)\}$
TU(λ,μ)	$X=[U^{\lambda}-1]/\lambda-[(1-U)^{\mu}-1]/\mu$

For completeness, we also simulated powers of $W^*_{S1}(\tilde\beta)$, with $\tilde\beta$ being the maximum likelihood estimator of β calculated iteratively. The resulting powers were, in most cases, lower than when the method of moments estimator was applied. To improve the readability of figures, these results are not presented. The results of our simulations are presented in Figs. 1 and 2. The corresponding figures in Janic-Wróblewska (2004 b) contain information on the ranges of skewness and kurtosis. Note that the simulated powers of $W^*_{S1}(\tilde\beta[m])$ are comparable to the corresponding powers of the solution given in Janic-Wróblewska (2004 b). Probability-weighted moments estimates for the extreme value distribution are known to be superior in many respects to the method of moments and maximum likelihood estimators. The results presented here also confirm their appealing properties and greater overall ability to identify the underlying distribution.

Figures 1 and 2 about here

3.2. Testing for the normal distribution

The parameters of

$$f(x;\beta)=\frac{1}{\sqrt{2\pi}\beta_2}\exp\Big\{-\frac{1}{2}\Big(\frac{x-\beta_1}{\beta_2}\Big)^2\Big\}$$

were estimated by $\tilde\beta=\tilde\beta[ns]$, where

$$\tilde\beta[ns]=(\bar X,\tilde\sigma),\quad\text{with}\quad \bar X=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad \tilde\sigma=\frac{1}{n-1}\sum_{i=1}^{n-1}\frac{X_{n:i+1}-X_{n:i}}{H_{i+1}-H_i},$$

where $X_{n:1}\le\ldots\le X_{n:n}$ are the order statistics based on $X_1,\ldots,X_n$ and $H_i=\Phi^{-1}\left(\frac{i-3/8}{n+1/4}\right)$, cf. Chen and Shapiro (1995). Since $\tilde\sigma$ is based on normalized spacings, we denote the estimate of β by $\tilde\beta[ns]$.

For comparison, we used the recognized standard given by the Shapiro-Wilk statistic W.

We repeated and extended the simulations reported in Tables V–VII of Kallenberg and Ledwina (1997 a). In particular, many cases from the well known simulation study of Pearson et al. (1977) were included. This enables many further conclusions. To retain a balance between symmetric and skewed alternatives, we added few symmetric distributions not considered in the above papers. In particular, we additionally used the Laplace distribution. It has density function $(\lambda/2)\exp\{-\lambda|x|\},\ x \in R,\ \lambda \in R_+$. A selection of empirical powers is presented in Figs. 3 and 4. Again, we focus on $n = 50$ using a significance level of $\alpha = 0.05$. To obtain exactly the same conditions as in Kallenberg and Ledwina (1997 a), we assumed $d(50) = 12$. The influence of $d(n)$ on empirical significance levels and powers has been already investigated in many earlier papers. Roughly speaking, the results obtained setting $d(50) = 12$ were almost identical to the ones obtained when e.g. $d(50) = 15$. For some discussion and examples see e.g. Kallenberg and Ledwina (1997 b).

Figures 3 and 4 about here

The simulated critical value for $W^*_{S1}(\tilde{\beta}[ns])$ was 5.3549. The simulated powers of $W^*_{S1}(\tilde{\beta})$, where $\tilde{\beta}$ is the maximum likelihood estimate of β, are available in Tables V–VII of Kallenberg and Ledwina (1997 a). It can be seen that the new method of estimation leads to slightly higher powers. We also tried some M-estimates of β discussed in Sec. 6.4 and 6.5 of Huber (1981), as well as some simple robust estimators (e.g. median, median deviation). The results were not encouraging.

The estimator $\tilde{\sigma}$ is unbiased and asymptotically equivalent to the asymptotically efficient estimator $\bar{\sigma}_n$, given by (2.9). The simulated powers of $W^*_{S1}(\tilde{\beta})$ using $\tilde{\beta} = (\bar{X}, \bar{\sigma}_n)$, are comparable to those obtained using the estimator $(\bar{X}, \tilde{\sigma})$, except in the case $T(\lambda, \mu)$ when the values of the parameters are relatively small. In this case the method was very unstable.

Finally, note that in the case of the extreme value distribution the probability-weighted moments estimate of β_2 is a normalized version of Gini's mean difference. A similar estimator of β_2 in the case of the normal distribution was derived, by different means, in Downton (1966). More precisely, Downton's linear estimate of $\beta = (\beta_1, \beta_2)$ is of the form

$$\tilde{\beta}[l] = \Big(\bar{X}, \sqrt{\pi}[n(n-1)]^{-1} \sum_{1 \le j < i \le n} (X_{n:i} - X_{n:j})\Big),$$

where $X_{n:1} \le ... \le X_{n:n}$ are the order statistics for a sample from the $N(\beta_1, \beta_2)$ distribution. The empirical powers of $W^*_{S1}(\tilde{\beta}[l])$ are in most cases slightly higher than obtained when the estimator $\tilde{\beta}[ns]$ is used. However, as in the case of the estimator $(\bar{X}, \bar{\sigma}_n)$ under alternatives $T(\lambda, \mu)$, a substantial loss of power was observed when the values of parameters were relatively small.

In view of the above evidence, we recommend the most stable solution $W^*_{S1}([ns])$.

4. Discussion

As previously stated, the construction presented here is based on a general idea, which successfully can be applied to many other cases. This idea is to match efficient score statistics, or generalized smooth statistics using Javitz's terminology, with an appropriate score based selection rule. This gives us a lot of freedom to choose appropriate estimates of nuisance parameters. In particular, it enables us to extend the theory to nonparametric and semiparametric problems.

In the applications presented in this paper, the distributions considered in the alternatives are not very complex in the sense that few first terms of the Fourier expansion in the Legendre basis describe them sufficiently well. In such cases, a score-based selection rule with a Schwarz-type penalty proved to be useful. If one wishes to detect distributions with sharp peaks or high frequency oscillations, it is advisable to modify and improve this rule to be sensitive to such deviations and, at the same time, retain power in the case of low dimensional departures. For some proposals and results in this field see Inglot and Ledwina (2006 a, b, c).

Note also that data-driven smooth tests have nonstandard asymptotic properties. Roughly speaking, for a large class of local alternatives they are as efficient as if the alternatives were known. For some illustration of this we refer the reader to e.g. Inglot and Ledwina (2001).

R code for data driven smooth tests considered in this paper are available at `http://cran.r-project.org/web/packages/ddst/index.html`.

Acknowledgments. T. Ledwina warmly thanks Dr. J.M. Landwehr for making some of his papers available to her.

References

Aerts, M., Claeskens, G., Hart, J.D., 1999. Testing the fit of a parametric function. *J. Amer. Statist. Assoc.* 94, 869–879.

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

Bai, Z.D., Chen, L., 2003. Weighted W test for normality and asymptotics a revist of Chen-Shapiro test for normality. *J. Statist. Plann. Inference* 113, 485–503.

Baklanov, E.A., Borisov, I.S., 2003. Probability inequalities and limit theorems for generalized L-statistics. *Lith. Math. J.* 43, 125–140.

Bera, A.K., Bilias, Y., 2001 a. Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: an essay on historical developments and some new results. *J. Statist. Plann. Inference* 97, 9–44.

Bera, A.K., Bilias, Y., 2001 b. On some optimality properties of Fisher-Rao score function in testing and estimation. *Commun. Statist.–Theory Meth.* 30, 1533–1559.

Bickel, P.J., Ritov, Y., Stocker, T.M., 2006. Tailor-made tests for goodness of fit to semiparametric hypotheses. *Ann. Statist.* 34, 721–741.

Bogdan, M., Ledwina, T., 1996. Testing uniformity via log-spline modeling. *Statistics* 28, 131–157.

Bühler, W.J., Puri, P.S., 1966. On optimal asymptotic tests of composite hypotheses with several constraints. *Z. Wahrsch. verw. Geb.* 5, 71–88.

Chen, L., Shapiro, S.S., 1995. An alternative test for normality based on normalized spacings. *J. Statist. Comput. Simulation* 53, 269–288.

Claeskens, G., Hjort, N.L., 2004. Goodness of fit via non-parametric likelihood ratios. *Scand. J. Statist.* 31, 487–513.

D'Agostino, R.B., Stephens, M.A., 1986. *Goodness–of–Fit Techniques.* Dekker, New York.

Downton, F., 1966. Linear estimates with polynomial coefficients. *Biometrika* 53, 129–141.

Fromont, M., Laurent, B., 2006. Adaptive goodness-of-fit tests in a density model. *Ann. Statist.* 34, 680–720.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.* 15, 1049–1054.

Hosking, J.R.M., Wallis, J.R., Wood, E.F., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27, 251–261.

Huber, P.J., 1981. *Robust Statistics*. Wiley, New York.

Inglot, T., Kallenberg, W.C.M., 2003. Moderate deviations of minimum contrast estimators under contamination. *Ann. Statist.* 31, 852–879.

Inglot, T., Ledwina, T., 1993. Moderately large deviations and expansions of large deviations for some functionals of weighted empirical process. *Ann. Probab.* 21, 1691–1705.

Inglot, T., Kallenberg, W.C.M., Ledwina, T., 1994. On selection rules with an application to goodness–of–fit for composite hypotheses. Memorandum 1242, University of Twente, Faculty of Applied Mathematics.

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

Inglot, T., Ledwina, T., 2001. Asymptotic optimality of data driven smooth tests for location–scale family. *Sankhyā Ser. A* 60, 41–71.

Inglot, T., Ledwina, T., 2006 a. Data-driven score tests for homoscedastic linear regression model: asymptotic results. *Probab. Math. Statist.* 26, 41–61.

Inglot, T., Ledwina, T., 2006 b. Data-driven score tests for homoscedastic linear regression model: the construction and simulations. In *Prague Stochastics 2006. Proceedings*, (eds. M. Hušková, M. Janžura), 124–137, Matfyzpress, Prague.

Inglot, T., Ledwina, T., 2006 c. Towards data driven selection of a penalty function for data driven Neyman tests. *Linear Algebra and its Appl.* 417, 579–590.

Janic-Wróblewska, A., 2004 a. Data-driven smooth test for a location-scale family. *Statistics* 38, 337–355.

Janic-Wróblewska, A., 2004 b. Data-driven smooth test for extreme value distribution. *Statistics* 38, 413–426.

Javitz, H.S., 1975. Generalized smooth tests of goodness of fit, independence and equality of distributions. Unpublished Ph.D. thesis at University of California, Berkeley.

Kallenberg, W.C.M., Ledwina, T., 1995. On data driven Neyman's tests. *Probab. Math. Statist.* 15, 409–426.

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

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

Kallenberg, W.C.M., Ledwina, T., 1999. Data driven rank tests for independence. *J. Amer. Statist. Assoc.* 94, 285–301.

Landwehr, J.M., Matalas, N.C., 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.* 15, 1055–1064.

LaRiccia, V.L., Mason, D.M., 1985. Optimal goodness-of-fit tests for location/scale families of distributions based on the sum of squares of L-statistics. *Ann. Statist.* 13, 315-330.

LaRiccia, V.L., 1991. Smooth goodness-of-fit tests: a quantile function approach. *J. Amer. Statist. Assoc.* 86, 427–431.

Le Cam, L., Lehmann, E.L., 1974. J. Neyman - O 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.

Meintanis, S., Swanepoel, J., 2007. Bootstrap goodness-of-fit tests with estimated parameters based on empirical transforms. *Statist. Probab. Lett.* 77, 1004–1013.

Morales, D., Pardo, L., Vajda, I., 1997. Some new statistics for testing hypotheses in parametric models. *J. Multivariate Anal.* 62, 137–168.

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

Neyman, J., 1959. Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics*, (ed. U. Grenander), Harald Cramér Volume, 212–234, Wiley, New York.

Pearson, E.S., D'Agostino, R.B., Bowman, K.O., 1977. Tests for departure from normality: comparison of powers. *Biometrika* 64, 231–246.

Rayner, J.C.W., Best, D.J., 1989. *Smooth Tests of Goodness of Fit.* Oxford University Press, New York.

Rubin, H., Sethuraman, J., 1965. Probabilities of moderate deviations. *Sankhyā Ser. A* 27, 325–346.

Shorack, G.R., Wellner, J.A., 1986. *Empirical Processes with Applications to Statistics.* Wiley, New York.

Stute, W., Manteiga, W.G., Quindimil, M.P., 1993. Bootstrap based goodness-of-fit tests. *Metrika* 40, 243–256.

Thomas, D.R., Pierce, D.A., 1979. Neyman's smooth goodness–of–fit test when the hypothesis is composite. *J. Amer. Statist. Assoc.* 74, 441–445.

Appendix A. Some analytical properties of the matrices $I_{\beta\beta}(\eta_0)$ and $I_{\beta\beta}(\eta_0) - I_{\beta\theta}(\eta_0)I_{\theta\beta}(\eta_0)$

We refer here to the results from Appendix A of Janic-Wróblewska (2004 a) and Sec. 5.10 of Inglot and Ledwina (2001). Recall that in the present paper row vectors are used. From (2.3)–(2.5) and (2.7), we obtain

$$I_{\beta\theta}(0,\beta^o) = I_{\beta_0}, \;\; I_{\beta\beta}(0,\beta^o) = I_{\beta_0\beta_0}$$

using the notation from Janic-Wróblewska (2004 a). We have

$$I_{\beta\theta}(0,\beta^o) = \begin{pmatrix} m_{11} & \dots & m_{1k} \\ m_{21} & \dots & m_{2k} \end{pmatrix}, \quad \begin{array}{l} m_{1j} = -\int_R \mathcal{L}^{(1)}(x)\phi_j(F(x))f(x)dx, \\ m_{2j} = -\int_R x\mathcal{L}^{(1)}(x)\phi_j(F(x))f(x)dx. \end{array} \tag{a.1}$$

Additionally set $m_j = (m_{1j}, m_{2j}),\ j = 1, ..., k$. Assumption (A3) implies that m_{ij}, $i = 1, 2$, $j = 1, ..., k$ are well defined and finite. Using (A3) again, one obtains

$$I_{\beta\beta}(0,\beta^o) = \begin{pmatrix} \int_R [\mathcal{L}^{(1)}(x)]^2 f(x)dx & \int_R x[\mathcal{L}^{(1)}(x)]^2 f(x)dx \\ \int_R x[\mathcal{L}^{(1)}(x)]^2 f(x)dx & \int_R x^2[\mathcal{L}^{(1)}(x)]^2 f(x)dx - 1 \end{pmatrix} =$$

$$\begin{pmatrix} \sum_{j=1}^{\infty} m_{1j}^2 & \sum_{j=1}^{\infty} m_{1j}m_{2j} \\ \sum_{j=1}^{\infty} m_{1j}m_{2j} & \sum_{j=1}^{\infty} m_{2j}^2 \end{pmatrix}. \tag{a.2}$$

Note also that

$$I_{\beta\theta}(\eta_0) = \frac{1}{\beta_2} I_{\beta\theta}(0,\beta^o) \quad \text{and} \quad I_{\beta\beta}(\eta_0) = \frac{1}{\beta_2^2} I_{\beta\beta}(0,\beta^o).$$

Hence,

$$I_{\beta\beta}(\eta_0) - I_{\beta\theta}(\eta_0)I_{\theta\beta}(\eta_0) = \frac{1}{\beta_2^2} \begin{pmatrix} \sum_{j=k+1}^{\infty} m_{1j}^2 & \sum_{j=k+1}^{\infty} m_{1j}m_{2j} \\ \sum_{j=k+1}^{\infty} m_{1j}m_{2j} & \sum_{j=k+1}^{\infty} m_{2j}^2 \end{pmatrix}. \tag{a.3}$$

Obviously, $I_{\beta\beta}(\eta_0)$ is the Fisher information matrix for (2.1). We introduce the following auxiliary notation:

$$I_{\beta\beta}(\eta_0) = \frac{1}{\beta_2^2} \begin{pmatrix} e_1 & e \\ e & e_2 \end{pmatrix}, \quad I_{\beta\beta}(\eta_0) - I_{\beta\theta}(\eta_0)I_{\theta\beta}(\eta_0) = \frac{1}{\beta_2^2} \begin{pmatrix} r_1 & r \\ r & r_2 \end{pmatrix}. \tag{a.4}$$

Formula (a.3) indicates how the second matrix in (a.4) depends on k. Under the additional assumption:

(A10) There exists $\epsilon \in (0, 1]$ such that $r^2 \leq (1-\epsilon) r_1 r_2$ for all $1 \leq k < \infty$;

the result from p. 68 of Inglot and Ledwina (2001) implies that $\lambda(k)$, the largest eigenvalue of $[I^*]^{-1} = I + I_{\theta\beta}(\eta_0)\Big[I_{\beta\beta}(\eta_0) - I_{\beta\theta}(\eta_0)I_{\theta\beta}(\eta_0)\Big]^{-1} I_{\beta\theta}(\eta_0)$, is bounded above by

$$\lambda^*(k) = 2\epsilon^{-1}(e_1/r_1 + e_2/r_2).$$

For illustration, we briefly comment on this bound in the case of the extreme value distribution. For this distribution we have $e_1 = 1$, $e_2 = \pi^2/6 + (1-\gamma)^2$ and $e = 1 - \gamma$, where γ is the Euler constant. Moreover, from (a.1),

$$m_{1j} = \int_0^1 [1 + \log(1-u)]\phi_j(u)du, \quad m_{2j} = \int_0^1 [1 + \log(1-u)][\log(-\log(1-u))]\phi_j(u)du.$$

Also, recall that

$$\phi_j(u) = \sqrt{2j+1}\sum_{l=0}^{j}(-1)^{l+j}\binom{j}{l}\binom{j+l}{l}u^l.$$

From Sec. 2.1 of Janic-Wróblewska (2004 b), we obtain

$$r_1 = \sum_{j=k+1}^{\infty} m_{1j}^2 = \sum_{j=k+1}^{\infty}(2j+1)j^{-2}(j+1)^{-2} > (k+2)^{-3}.$$

This indicates that the magnitude of $\lambda(k)$ is presumably at least $O(k^3)$. Since this discussion serves only to explain the need for assumption (A6) and is not crucial for practical applications, we do not consider the technical details involved any further. See also Remark 4.

Appendix B. Proof of Theorem 1

Since $\lambda(k)$ is the largest eigenvalue of the $k \times k$ matrix $[I^*]^{-1}$, one obtains

$$W_k^*(\tilde{\beta}) \leq \lambda(k)\Big|\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell^*(X_i;\tilde{\beta})\Big|\Big|^2. \qquad (b.1)$$

To simplify the notation and to make it more similar to that used in Janic-Wróblewska (2004 a), set

$$\tilde{Y}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\ell^*(X_i;\beta) \quad \text{and} \quad b_k = a_k/\lambda(k) \quad \text{with} \quad a_k = (k-1)n^{-1}\log n. \qquad (b.2)$$

To shorten the exposition of the proof, in a few places we shall make use of relevant calculations made in the paper mentioned above. Observe that

$$\Big\{S1(\tilde{\beta}) \geq 2\Big\} \subset \bigcup_{k=2}^{d(n)}\Big\{W_k^*(\tilde{\beta}) \geq (k-1)\log n\Big\} \subset \bigcup_{k=2}^{d(n)}\Big\{||\tilde{Y}_n(\tilde{\beta})|| \geq \sqrt{b_k}\Big\}.$$

From axiom (A7), using (2.7), one obtains

$$P_\beta\Big(S1(\tilde\beta)\geq 2\Big)\leq \sum_{k=2}^{d(n)} P_{\beta^o}\Big(||\tilde Y_n(\beta^o)||\geq \sqrt{b_k}/2\Big)+\sum_{k=2}^{d(n)} P_{\beta^o}\Big(||\tilde Y_n(\tilde\beta)-\tilde Y_n(\beta^o)||\geq \sqrt{b_k}/2\Big). \qquad (b.3)$$

The relation (b.3) is a counterpart of (C.1) on p. 350 of Janic-Wróblewska (2004 a) with $b_k = a_k/\lambda(k)$ in place of a_k. Since, in typical situations, $\lambda(k)\to\infty$ as $k\to\infty$, one has to adapt the argument to these more restrictive conditions. Note also that our assumption (A8) is slightly stronger than the related condition (10) in the aforementioned paper. This allows for more flexibility in the proof. Using (A6) we set $a_k^* = c_3 k^{1-v} n^{-1}\log n$, where c_3 is an absolute constant, and consider the bound

$$P_\beta\Big(S1(\tilde\beta)\geq 2\Big)\leq \sum_{k=2}^{d(n)} P_{\beta^o}\Big(||\tilde Y_n(\beta^o)||\geq \sqrt{a_k^*}\Big)+\sum_{k=2}^{d(n)} P_{\beta^o}\Big(||\tilde Y_n(\tilde\beta)-\tilde Y_n(\beta^o)||\geq \sqrt{a_k^*}\Big). \qquad (b.4)$$

First we analyse the first term in (b.4). For this purpose we introduce

$$T_j = \frac{1}{n}\Big[\sum_{i=1}^n \ell_j^*(X_i;\beta^o)\Big]^2 + s_j - 1 \quad\text{with}\quad s_j = m_j I_{\beta\beta}^{-1}(0,\beta^o)m_j^T,$$

where m_j is defined in Appendix A. Note that $s_j\geq 0$, $Var_\beta \ell_j^*(X_1;\beta) = 1 - s_j$ and $\sum_{j=1}^\infty s_j = 2$. Moreover, one has $E_{\beta^o}T_j = 0$. Therefore, using the properties of Legendre polynomials, taking $\tilde a_k = c_4 k^{1-v}\log n - k + \sum_{j=1}^k s_j$ and arguing as in Step 1 of Janic-Wróblewska (2004 a), one obtains for some absolute constant c_5

$$P_{\beta^o}\Big(||\tilde Y_n(\beta^o)||\geq \sqrt{a_k^*}\Big)\leq c_5 k^{2v-1}/(\log n - k^v)^2.$$

Hence,

$$\sum_{k=2}^{d(n)} P_{\beta^o}\Big(||\tilde Y_n(\beta^o)||\geq \sqrt{a_k^*}\Big)\to 0 \quad\text{if}\quad d(n) = o\Big([\log n]^{1/v}\Big). \qquad (b.5)$$

Analysis of the remainder of (b.4) consists of a few steps. This part of the proof extensively exploits the closeness of $\tilde\beta$ to β, as expressed in assumption (A8). First observe that, using the Taylor expansion, the j-th component, $j = 1, ..., k$, of $\tilde Y_n(\tilde\beta) - \tilde Y_n(\beta)$ can be written as $\mathcal{Z}_j + \mathcal{R}_{1j} + \mathcal{R}_{2j}$, where

$$\mathcal{Z}_j = (\tilde\beta-\beta^o)E_{\beta^o}\Big[\frac{\partial}{\partial\beta^T}\ell_j^*(X_1;\beta)\Big|_{\beta=\beta^o}\Big],$$

$$\mathcal{R}_{1j} = (\tilde\beta-\beta^o)U_j^T,\quad U_j = (U_{1j},U_{2j}),\quad U_{tj} = \frac{1}{n}\sum_{i=1}^n\Big[\frac{\partial}{\partial\beta_t}\ell_j^*(X_i;\beta)\Big|_{\beta=\beta^o} - E_{\beta^o}\Big[\frac{\partial}{\partial\beta_t}\ell_j^*(X_1;\beta)\Big|_{\beta=\beta^o}\Big],$$

$$\mathcal{R}_{2j} = \frac{1}{2}(\tilde\beta-\beta^o)\Big[\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\beta^T\partial\beta}\ell_j^*(X_i;\beta)\Big|_{\beta=\xi_i}\Big](\tilde\beta-\beta^o)^T,$$

where, for each X_i, the point ξ_i lies between $\tilde\beta$ and β^o. Finally, set

$$\mathcal{Z} = (\mathcal{Z}_1, ..., \mathcal{Z}_k),\ \ \mathcal{R}_1 = (\mathcal{R}_{11}, ..., \mathcal{R}_{1k})\ \ \text{and}\ \ \mathcal{R}_2 = (\mathcal{R}_{21}, ..., \mathcal{R}_{2k}).$$

The triangle inequality implies that

$$P_{\beta^o}\Big(||\tilde{Y}_n(\tilde{\beta}) - Y_n(\beta^o)|| \geq \sqrt{a_k^*}\Big) \leq$$

$$P_{\beta^o}\Big(||\mathcal{Z}|| \geq \sqrt{a_k^*}/3\Big) + P_{\beta^o}\Big(||\mathcal{R}_1|| \geq \sqrt{a_k^*}/3\Big) + P_{\beta^o}\Big(||\mathcal{R}_2|| \geq \sqrt{a_k^*}/3\Big).$$

As in (C.8) of Janic-Wróblewska (2004 a), one obtains

$$\Big|E_\beta \frac{\partial}{\partial \beta_t}\ell_j^*(F(X_1;\beta))\Big|^2 \leq (1 - s_j)Var_\beta\Big[\frac{\partial}{\partial \beta_t}\log f(X_1;\beta)\Big], \quad t = 1, 2.$$

Therefore, from the definition of $\mathcal{Z}_j$ and relation (2.4) it follows that

$$\mathcal{Z}_j^2 \leq ||\tilde{\beta} - \beta^o||^2 \Big|\Big|E_{\beta^o}\Big[\frac{\partial}{\partial \beta}\ell_j^*(X_1;\beta)\Big|_{\beta=\beta^o}\Big]\Big|\Big|^2 \leq b^2||\tilde{\beta} - \beta^o||^2,$$

where

$$b^2 = \sum_{t=1}^{2} Var_{\beta^o}\Big[\frac{\partial}{\partial \beta_t}\log f(X_1;\beta)\Big|_{\beta=\beta^o}\Big] = \int_R (1 + x^2)\Big[\mathcal{L}^{(1)}(x)\Big]^2 f(x)dx - 1.$$

Under assumption (A3), b^2 exists and is finite. It follows that for some absolute constant c_6

$$P_{\beta^o}\Big(||\mathcal{Z}|| \geq \sqrt{a_k^*}/3\Big) \leq P_{\beta^o}\Big(\sqrt{n}||\tilde{\beta} - \beta^o|| \geq c_6\sqrt{k^{-v}\log n}\Big).$$

Therefore, from (A8), one may infer that

$$\sum_{k=2}^{d(n)} P_{\beta^o}\Big(||\mathcal{Z}|| \geq \sqrt{a_k^*}/3\Big) \to 0 \quad \text{if} \quad d(n) = o\Big([\log n]^{1/v}[\log\log n]^{-1}\Big). \tag{b.6}$$

From the definition of $\mathcal{R}_1$ it follows that for some positive absolute constants c_8 and c_9

$$P_{\beta^o}\Big(||\mathcal{R}_1|| \geq \sqrt{a_k^*}/3\Big) \leq P_{\beta^o}\Big(\sqrt{n}||\tilde{\beta} - \beta^o|| \geq c_7\sqrt{\log n}\Big) + P_{\beta^o}\Big(\sum_{j=1}^{k}\sum_{t=1}^{2} U_{tj}^2 \geq c_8 k^{1-v}\Big).$$

Using (A5) and (A8) and repeating an elementary argument from p. 353 of Janic-Wróblewska (2004 a), one obtains

$$\sum_{k=2}^{d(n)} P_{\beta^o}\Big(||\mathcal{R}_1|| \geq \sqrt{a_k^*}/3\Big) \to 0 \quad \text{if} \quad d(n) = o(n^w) \quad \text{with} \quad w = \min\Big\{c_2, 1/(5 + v)\Big\}. \tag{b.7}$$

To analyse the tails of $||\mathcal{R}_2||$ we introduce the event $B_n = \Big\{\sqrt{n}||\tilde{\beta} - \beta^o|| \leq \sqrt{\log n}\Big\}$ and denote its complement by B_n^c. We have

$$P_{\beta^o}\Big(||\mathcal{R}_2|| \geq \sqrt{a_k^*}/3\Big) \leq P_{\beta^o}\Big(B_n^c\Big) + P_{\beta^o}\Big(||\mathcal{R}_2|| \geq \sqrt{a_k^*}/3, B_n\Big).$$

Suppose B_n holds. From axiom (A4), one can apply (C.17) in Janic-Wróblewska (2004 a). This, together with (A8), results in

$$\sum_{k=2}^{d(n)} P_{\beta^o}\Big(||\mathcal{R}_2|| \geq \sqrt{a_k^*}/3\Big) \to 0 \quad \text{if} \quad d(n) = \min\Big\{o(n^{c_2}), o([n/\log n]^{1/(11+v)})\Big\}. \tag{b.8}$$

The proof follows from (b.5)–(b.8).

Appendix C. Proof of Theorem 2

The main concept behind this proof can be found in Sec. 4 of Inglot et al. (1994). If $I\!\!P$ is an alternative to $P_{\beta(I\!\!P)}$, then there exists a natural number $k = k(I\!\!P)$, such that

$$E_{I\!\!P}\ell^*_k(X_1;\beta(I\!\!P)) \neq 0. \qquad (c.1)$$

Let $K = K(I\!\!P)$ be the smallest number for which (c.1) holds and consider the case $K > 1$. We have $E_{I\!\!P}\ell^*_1(X_1;\beta(I\!\!P)) = ... = E_{I\!\!P}\ell^*_{K-1}(X_1;\beta(I\!\!P)) = 0$ and $E_{I\!\!P}\ell^*_K(X_1;\beta(I\!\!P)) \neq 0$. Thus the component ℓ^*_K provides a non-zero shift under $I\!\!P$ and $W^*_K(\tilde{\beta}) \xrightarrow{I\!\!P} \infty$. Therefore, the proof is complete provided that we show that $I\!\!P(S1(\tilde{\beta}) \geq K) \to 1$. For this purpose, observe that, by (b1) and (b2) of Appendix B, $W^*_k(\tilde{\beta}) \leq n\lambda(k)||\tilde{Y}_n(\tilde{\beta})||^2,\ k \in N$. In addition, given k, we have $[I^*]^{-1} = I + R(\beta)$, where I is the $k \times k$ identity matrix and

$$R(\beta) = I_{\theta\beta}(\eta_0)\Big[I_{\beta\beta}(\eta_0) - I_{\beta\theta}(\eta_0)I_{\theta\beta}(\eta_0)\Big]^{-1}I_{\beta\theta}(\eta_0).$$

The matrix $R(\beta)$ is nonnegative definite and independent of β. Hence, $W^*_k(\tilde{\beta}) \geq n||\tilde{Y}_n(\tilde{\beta})||^2,\ k \in N$. In this way the analysis of $S1(\tilde{\beta})$ under $I\!\!P$ can be reduced to a similar analysis of the rule $S2(\tilde{\beta})$ pertaining to $||\tilde{Y}_n(\tilde{\beta})||^2$, which has already been carried out in some of our earlier papers. See for example Sec. C.2 of Janic-Wróblewska (2004 a). The case $K = 1$ can be treated similarly.

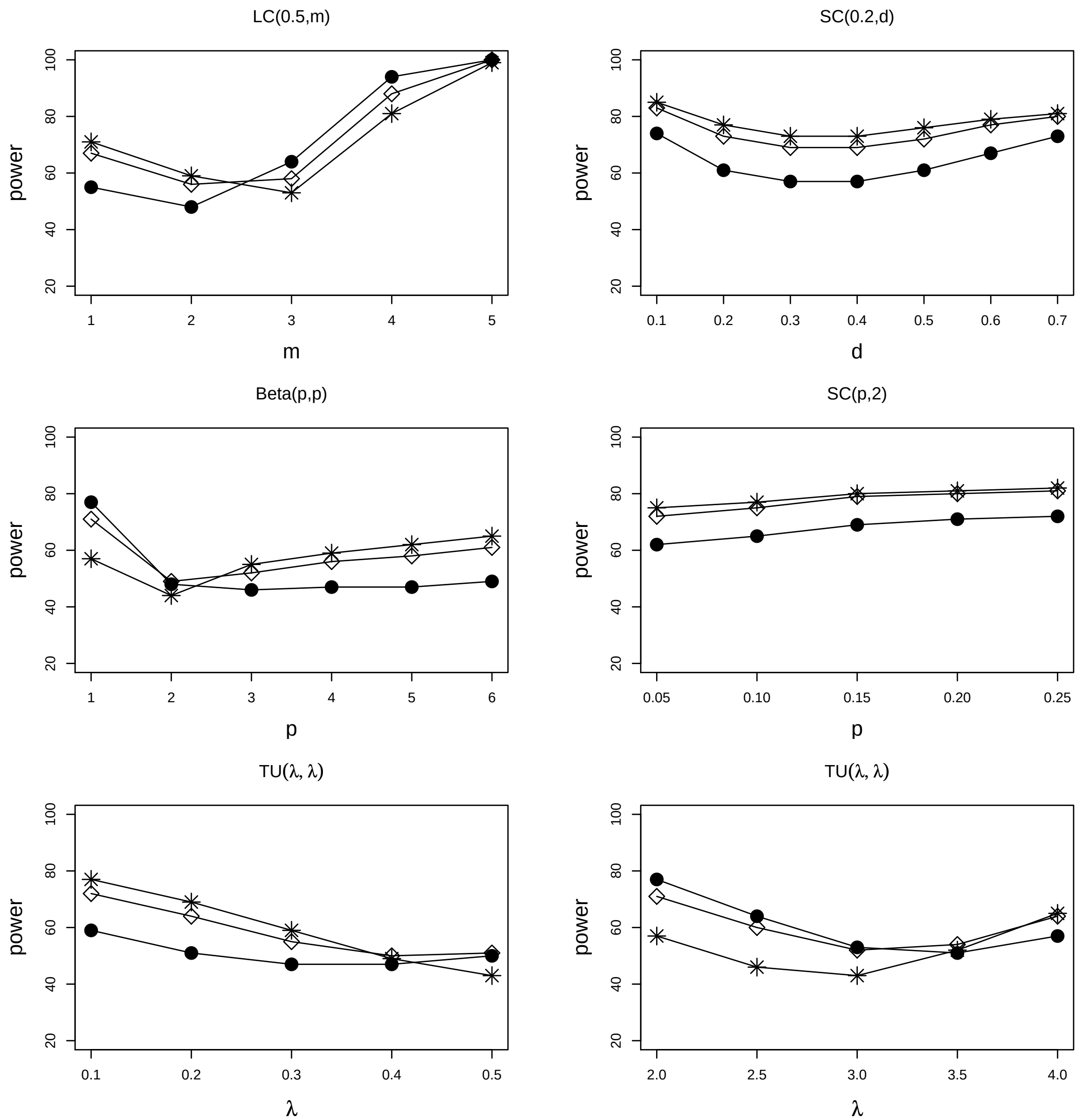

Figure 1: Testing for the extreme value distribution. Empirical powers (in %) of A^2 : $-\bullet-$, $W^*_{S1}(\tilde{\beta}[m]) : -\star-$, $W^*_{S1}(\tilde{\beta}[pwm]) : -\diamond-$. Symmetric alternatives.

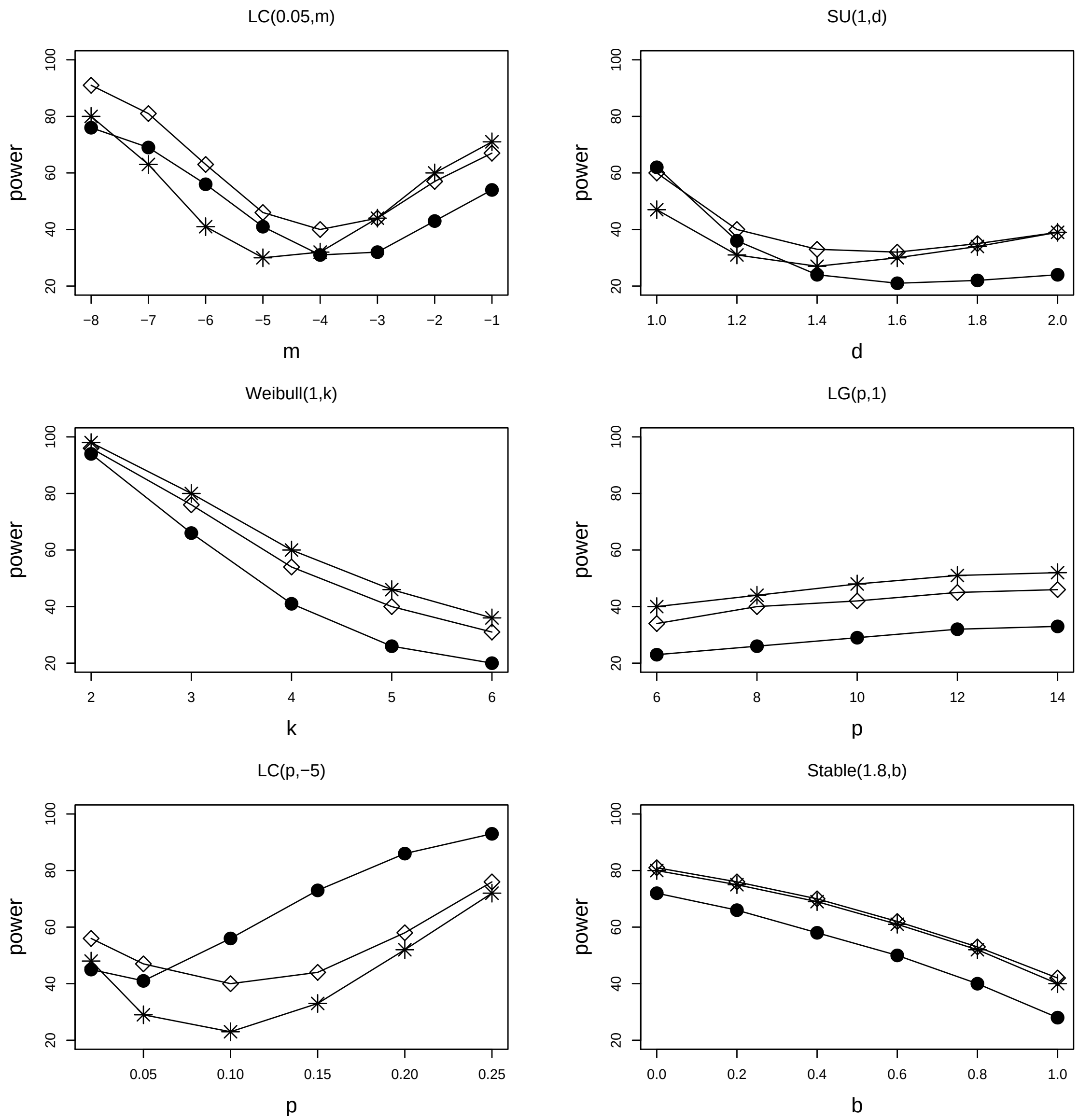

Figure 2: Testing for the extreme value distribution. Empirical powers (in %) of A^2 : $-\bullet-$, $W^*_{S1}(\tilde{\beta}[m])$: $-\star-$, $W^*_{S1}(\tilde{\beta}[pwm])$: $-\diamond-$. Skewed alternatives.

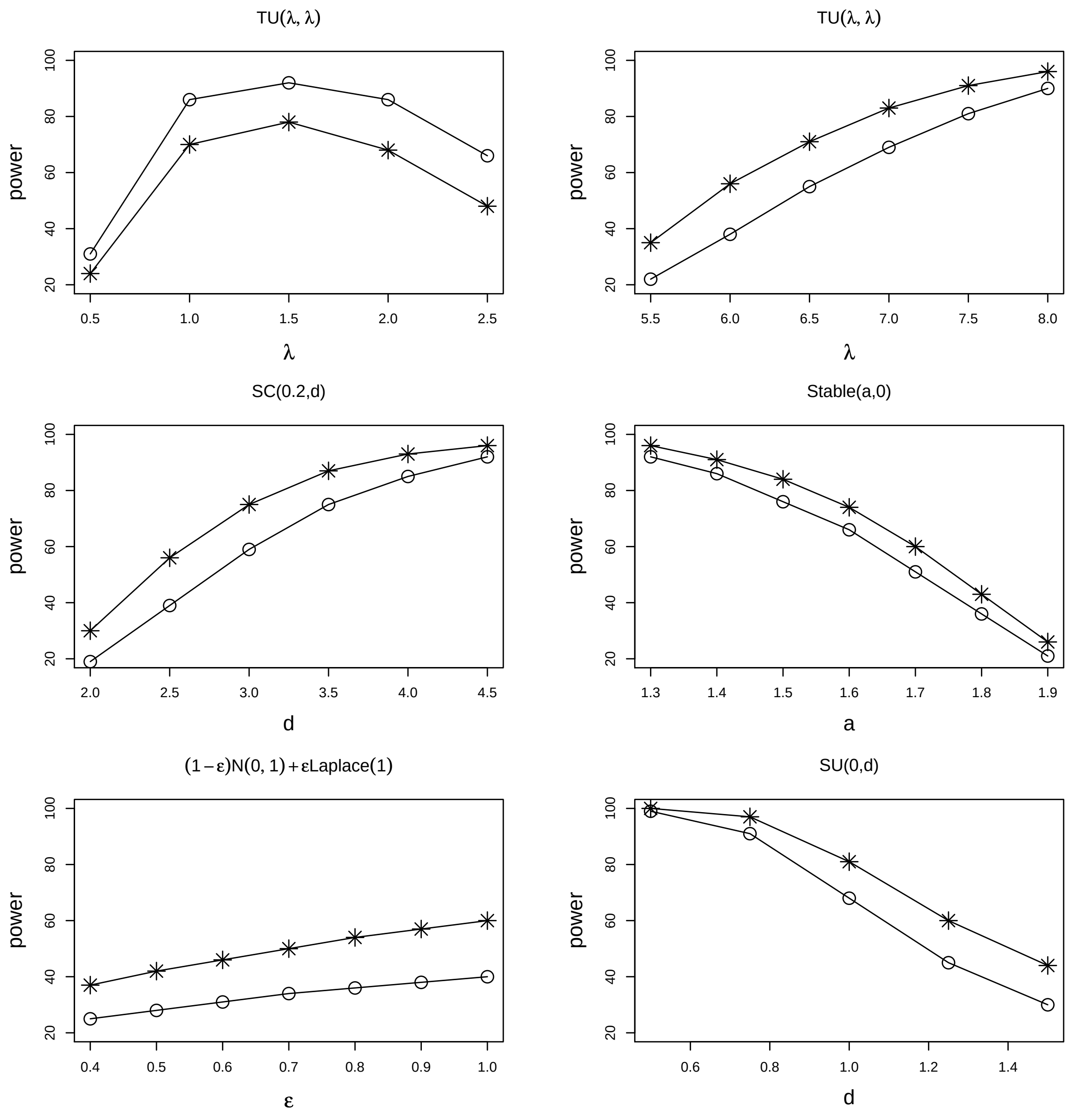

Figure 3: Testing for normality. Empirical powers (in %) of W : $-\circ-$, $W^*_{S1}(\tilde{\beta}[ns])$: $-\star-$. Symmetric alternatives.

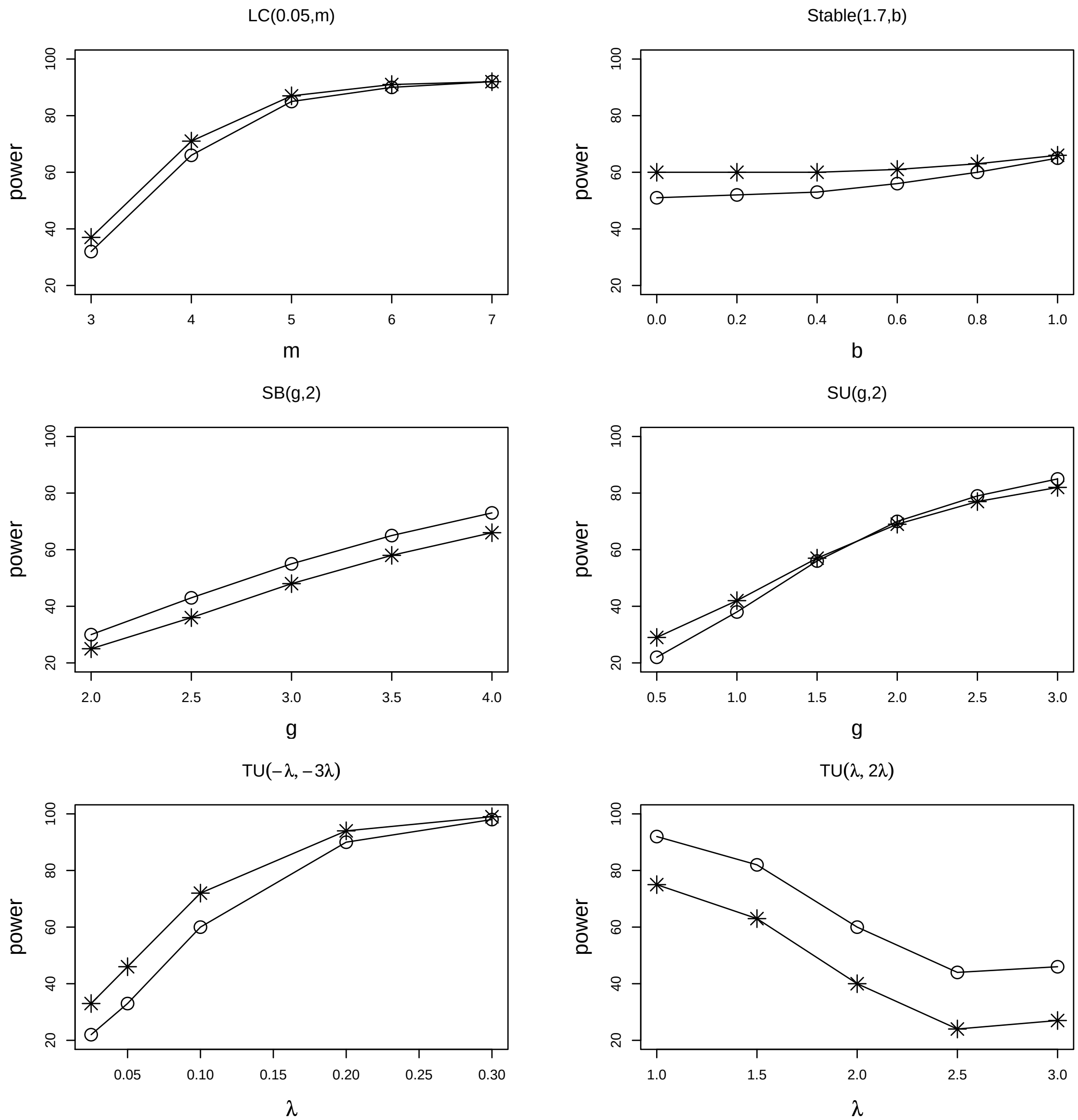

Figure 4: Testing for normality. Empirical powers (in %) of $W: -\circ-$, $W^*_{S1}(\tilde{\beta}[ns]): -\star-$. Skewed alternatives.