Faculty of Mathematical Sciences

University of Twente

University for Technical and Social Sciences

P.O. Box 217
7500 AE Enschede
The Netherlands
Phone: +31-53-4893400
Fax: +31-53-4893114
Email: memo@math.utwente.nl

Memorandum No. 1548

The penalty in data driven Neyman's tests

W.C.M. Kallenberg

October 2000

ISSN 0169-2690

The Penalty in Data Driven Neyman's Tests

Wilbert C.M. Kallenberg
Faculty of Mathematical Sciences
University of Twente
P.O. Box 217, 7500 AE Enschede
The Netherlands

Abstract Data driven Neyman's tests are based on two elements: Neyman's smooth tests in finite dimensional submodels and a selection rule to choose the "right" submodel. As selection rule usually (a modification of) Schwarz's rule is applied. In this paper we consider data driven Neyman's tests with selection rules allowing also other penalties than the one in Schwarz's rule. It is shown that the nice properties of consistency against very large classes of alternatives and the more deep result of asymptotic optimality in the sense of vanishing shortcoming continue to hold for other penalties as well, including the one corresponding to Akaike's selection rule.

Keyword and phrases: goodness-of-fit, model selection, Schwarz's criterion, Akaike's criterion, penalty, data driven test, consistency, vanishing shortcoming, intermediate efficiency, moderate deviations.

2000 Mathematics Subject Classification: 62 G 10, 62 G 20

1 Introduction

A data driven test for testing uniformity, linking Neyman's smooth tests with Schwarz's model selection procedure, was introduced by Ledwina (1994). Last years data driven tests have been developed in other contexts as well (composite goodness-of-fit problems, testing independence, the two-sample problem, testing for tail alternatives, etc.). It turns out that these tests perform very well. This is shown by simulations and by theoretical results.

Data driven tests consist of a sequence of test statistics and a selection rule to choose the "right statistic" among the collection of test statistics. Usually, the test statistics are standard test statistics (like score test statistics) for the testing problem restricted to a k-dimensional submodel. The selection rule is applied to choose the "right" dimension. The idea is that a higher dimensional, and hence more complex model, should be penalized. For a more extensive discussion see for instance Inglot and Ledwina (1996), shortly written as IL96 in the rest of the paper, and references therein.

A topic in data driven tests that is investigated so far only by a few simulations is the choice of the penalty in the selection rule. In Kallenberg and Ledwina (1997b, Section 6.1) some remarks are made about it and some simulations are presented. From the simulations performed in the context of that paper it is concluded that, "although other penalties can be considered, there is sufficient support to adopt the Schwarz rule in general". In contrast to this recommendation, recently, Janssen (2000, page 241; see also page 251) stated: "the estimator of the dimension should not be too restrictive", thus suggesting that the penalty in Schwarz's rule is too heavy for application in data driven tests (see also (2.4) and the remarks following it). It is the aim of this paper to investigate the role of the penalty more rigorously.

Theoretical support for data driven tests mostly concerns *consistency* against very large classes of alternatives. More deep results are obtained by IL96. They prove *asymptotic optimality* in the sense that the asymptotic intermediate efficiency with respect to the Neyman-Pearson test equals 1 for a large set of converging alternatives. Replacing the more complicated Schwarz's selection rule by a simplified version Inglot (1999), which is shortly written as I99 in the rest of the paper, has extended these results, thus obtaining asymptotic optimality in "almost every" direction, while classical tests are as a rule only asymptotically efficient in one direction.

A related asymptotic optimality concept is vanishing shortcoming, see Inglot et al. (2000). The shortcoming of a test is the difference between the power of the test and the power of the most powerful test for that particular alternative. It is shown in Inglot et al. (1998) that data driven tests are asymptotically optimal in the sense of vanishing shortcoming (and thus obtaining asymptotically the highest possible power) in an infinite number of orthogonal directions.

In this paper we follow this more direct approach of (asymptotic) power comparison, expressed by the concept of vanishing shortcoming with the level tending to zero. Noting the close relationship between this concept and intermediate efficiency [see Inglot et al. (2000)], the arguments in favor of intermediate efficiency as criterion for test comparison in this context, presented in Inglot and Ledwina (1998), continue to hold for vanishing shortcoming with level tending to zero as well: we get explicit quantitative results for comparison of powers and the asymptotic results, for instance when comparing data driven Neyman's tests with other goodness-of-fit tests as the Cramér-von Mises test and the Anderson-Darling test are consistent with those following from moderate finite sample size comparisons [cf. Section 5 in Inglot et al. (1998) and Sections 5.4 and 5.5 in Inglot et al. (2000) or Section 3 in Inglot and Ledwina (1998)].

Furthermore, to avoid unnecessary technicalities and to focus on the role of the penalty we investigate in this paper the goodness-of-fit testing problem of testing uniformity (which is equivalent to the simple null hypothesis case), consider contamination alternatives and use the data driven test with the simplified version of the selection rule. Moreover, the orthonormal system of the Legendre polynomials is applied, which has turned out to work very well in practice. Generalizations however, to other testing problems, to other type of alternatives, to data driven tests with other orthonormal systems or with other versions of the selection rule can be done as well.

The main result of the paper is that the theoretical support for the data driven Neyman's test with Schwarz's rule continues to hold for all other penalties, starting from the one corresponding to Akaike's criterion up to even much larger penalties than the one in Schwarz's criterion. This common behavior concerns both the consistency against very large classes of alternatives and the asymptotic optimality in the sense of vanishing shortcoming for a large set of converging alternatives in an infinite number of orthogonal directions.

From the points of view of consistency and asymptotic optimality in the sense of vanishing shortcoming, the conclusion based on the few simulations in Kallenberg and Ledwina (1997b) is supported: on the one hand other penalties can be considered also, but on the other hand the results of this paper give no reason to assume that the penalty corresponding to Schwarz's rule is too heavy. It is seen from Theorems 3.5 and 4.7 that there is no essential restriction due to the chosen penalty and hence the range of alternatives against which consistency and vanishing shortcoming holds is the same, irrespective of the chosen penalty (if it is as least as large as in Akaike's rule and not tremendously larger than in Schwarz's rule).

The paper is organized as follows. In Section 2 the data driven tests with several penalties are introduced and some notations and assumptions are presented. Section 3 is devoted to the consistency part. The more deep analysis of asymptotic optimality in the sense of vanishing shortcoming is discussed in Section 4.

2 Notation and basic assumptions

The testing problem considered in this paper is the goodness-of-fit problem based on n observations with as null hypothesis a given continuous distribution in $\mathbb{R}$. Application of the integral transformation yields that we may consider without loss of generalization i.i.d. r.v.'s $X_1, \ldots, X_n$ with values in [0, 1] and that under the null hypothesis the X_i's are uniformly distributed on [0, 1]. When the null hypothesis holds, we write P_0 for the probability measure and E_0 for the corresponding expectation.

Although in the goodness-of-fit problem all kind of alternatives are of interest, usually firstly one is focussed on a shift in the mean, then on a possible change in variance after which skewness comes in, etc. According to Neyman's original choice we therefore consider the so called Neyman's smooth test statistics with k components, given by

$$T_k = \sum_{j=1}^{k} \left\{ n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2 , \tag{2.1}$$

where ϕ_j is the j^{th} orthonormal Legendre polynomial on [0, 1]. As ϕ_j is a polynomial of degree j, T_1 is concerned with the mean, while with T_2 the variance comes in, with T_3 the skewness, etc. In this way we cover step by step more and more alternatives.

The Euclidean norm of a vector x in $\mathbb{R}^k$ is denoted by $||x||_k$ and $||x||_{jk}^2 = x_{j+1}^2 + \cdots + x_k^2$. Write $\bar{\phi} = (\bar{\phi}_1, \bar{\phi}_2, \ldots)$ with $\bar{\phi}_j = n^{-1} \sum_{i=1}^{n} \phi_j(X_i), j = 1, 2, \ldots$, where the dimension of the vector $\bar{\phi}$ follows from the context. Then we get $T_k = n||\bar{\phi}||_k^2$. Note that T_k is the score test statistic for testing uniformity in the k-dimensional submodel, given by the exponential family

$$\exp \left\{ \sum_{j=1}^{k} \theta_j \phi_j(x) - \psi_k(\theta) \right\} .$$

Here $\psi_k(\theta)$ is the normalizing constant and the null hypothesis of uniformity corresponds to $\theta_1 = \cdots = \theta_k = 0$.

For getting high power against broad classes of alternatives (and that is what one wants in a goodness-of-fit problem) the right choice of the dimension k is a major problem. Recent research [cf. Bickel and Ritov (1992), Eubank (1997), Eubank and LaRiccia (1992), Fan (1996), Kallenberg and Ledwina (1995a), Ledwina (1994)] strongly indicates that a deterministic choice, even when it is sequential as in Hušková and Sen (1985, 1986), does not give a satisfactory solution.

Looking for the right choice of k means looking for the right model. Hence, an answer can be given from the area of statistics devoted to model selection. In a series of papers [see Albers et al. (1999), Bogdan (1995, 1999), Bogdan, Bogdan and Futschik (1999), Bogdan and Ledwina (1996), I99, Inglot et al. (1997, 1998), IL96, Inglot and Ledwina (1998, 1999), Janic-Wróblewska (1999a,b), Janic-Wróblewska and Ledwina (2000), Kallenberg and Ledwina (1995a,b, 1997a,b, 1999), Kallenberg et al. (1997), Ledwina (1994)] the data driven procedure based on (modifications of) Schwarz' selection rule has been shown to be very successful.

Schwarz' rule has been introduced to select optimally (in a Bayesian sense) the dimension of a parametric model. Moreover, the rule has a nice information-theoretic interpretation in terms of minimizing the discription length of the data for a given sample size [cf. Barron and Cover (1991)]. However, these properties are not linked up with data driven tests. So, a direct and convincing motivation to apply Schwarz's rule in data driven tests is not given so far.

The original Schwarz's rule gives that $-\log L$ (with L the likelihood ratio statistic for testing $\theta_1 = \cdots = \theta_k = 0$) is penalized by $\frac{1}{2}k \log n$. In the simplified version the likelihood ratio statistic is replaced by the more simple (and asymptotically equivalent) statistic T_k in the sense that T_k plays the role of $-2\log L$.

Several other selection rules, with different penalties, appear in the area of model selection. It is the aim of this paper to investigate the influence of the penalty in the selection rule on data driven Neyman's tests. Therefore, we consider in this paper the following general selection rule

$$S = \min\{1 \leq k \leq m_n : T_k - \Delta_n k \geq T_j - \Delta_n j, 1 \leq j \leq m_n\}, \tag{2.2}$$

where m_n is a control sequence, giving the largest dimension under consideration with n observations, and where Δ_n denotes the penalty. Taking

$$\Delta_n = \log n,$$

the selection rule S equals a modified version of Schwarz's rule. Other examples are:

$$\Delta_n = 2 \text{ (AIC)}$$

$$\Delta_n = \log \log n \text{ [Hannan and Quin (1979)]}$$

$$\Delta_n = 2 \log n \text{ [Haughton, Haughton and Izenman (1990)].}$$

The null hypothesis of uniformity is rejected for large values of

$$T_S = \sum_{j=1}^{S} \left\{ n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2 \tag{2.3}$$

with S given by (2.2).

It follows from (2.2) that

$$S = 1 \Leftrightarrow \Delta_n \geq \max \left\{ \frac{T_j - T_1}{j-1} : j = 2, \ldots, m_n \right\} \tag{2.4}$$

$$S = 2 \Leftrightarrow \Delta_n < \max \left\{ \frac{T_j - T_1}{j-1} : j = 2, \ldots, m_n \right\}$$

$$\text{and } \Delta_n \geq \max \left\{ \frac{T_j - T_2}{j-2} : j = 3, \ldots, m_n \right\}$$

etc. and hence, a lower penalty implies a larger value of S. Therefore, we get a "less restrictive estimation of the dimension" [see Janssen (2000, page 241)] by taking a lower value of the penalty.

As a consequence, a lower value of the penalty implies a larger value of the test statistic T_S. On the one hand, this means that T_S gets larger values under the null hypothesis, involving a larger critical value. But on the other hand, T_S will also be larger under alternatives and it is not seen beforehand what the resulting effect is on the power of the test.

3 Consistency

In this section it is investigated for which penalties Δ_n in the selection rule the data driven test is consistent against a large class of alternatives. The idea is that under the null hypothesis T_S is bounded in probability, while under (fixed) alternatives $T_S \to \infty$.

The technical tool for getting the result under the null hypothesis is the following version of inequality (2) in Prohorov (1973), cf. also Theorem 7.7 in IL96.

Theorem 3.1 (Prohorov, 1973) *Let $Y_1, \ldots, Y_n$ be i.i.d. random vectors with values in $\mathbb{R}^k$. Let $EY_i = 0$ and let the covariance matrix of Y_i be equal to the identity matrix. Assume $||Y_1||_k \leq L$ a.e. Then, for $2k \leq y^2 \leq nL^{-2}$, we have*

$$\Pr\left(||n^{-1/2} \sum_{i=1}^{n} Y_i||_k \geq y \right) \leq \frac{C}{\Gamma(k/2)} \left(\frac{y^2}{2} \right)^{(k-1)/2} \exp\left\{ -\frac{y^2}{2}(1 - \eta_n) \right\},$$

where C is an absolute constant, while $0 \leq \eta_n < Lyn^{-1/2}$. □

The following theorem gives the boundedness in probability of T_S.

Theorem 3.2 *Assume $\Delta_n \geq 2$ and $\lim_{n\to\infty} \frac{m_n^3 \Delta_n}{n} = 0$. Then $S = O_{P_0}(1)$ and hence $T_S = O_{P_0}(1)$.*

Proof If $S \geq K$, then $T_k - T_1 \geq (k-1)\Delta_n$ for some $K \leq k \leq m_n$ and hence

$$\sum_{j=2}^{k}\left\{n^{-1/2}\sum_{i=1}^{n}\phi_j(X_i)\right\}^2 \geq (k-1)\Delta_n \text{ for some } K \leq k \leq m_n.$$

We apply Theorem 3.1 (in dimension $k-1$) with $Y_i = (\phi_2(X_i), \ldots, \phi_k(X_i))$. It is easily seen that $E_0 Y_i = 0$ and that the covariance matrix of Y_i is equal to the identity matrix. Since

$$\max_{1\leq j\leq k} \sup_{x\epsilon[0,1]} |\phi_j(x)| = (2k+1)^{1/2},$$

we get

$$||Y_1||_{k-1}^2 \leq \sum_{j=2}^{k}(2j+1) = (k-1)(k+3),$$

implying that we may take in Theorem 3.1

$$L = \sqrt{(k-1)(k+3)}. \tag{3.1}$$

In view of the conditions $\Delta_n \geq 2$ and

$$\lim_{n\to\infty} \frac{m_n^3 \Delta_n}{n} = 0,$$

we have, for all $n \geq n_1$,

$$2 \leq \Delta_n \leq \frac{n}{(m_n-1)^2(m_n+3)}.$$

Application of Theorem 3.1 now yields

$$P_0\left(\sum_{j=2}^{k}\left\{n^{-1/2}\sum_{i=1}^{n}\phi_j(X_i)\right\}^2 \geq (k-1)\Delta_n\right) \tag{3.2}$$

$$\leq \frac{C}{\Gamma((k-1)/2)}\left(\frac{(k-1)\Delta_n}{2}\right)^{(k-2)/2} \exp\left\{-\frac{(k-1)\Delta_n}{2}(1-\eta_n)\right\},$$

where C is an absolute constant while $0 \leq \eta_n \leq (k-1)\sqrt{\frac{(k+3)\Delta_n}{n}}$. Using the inequality $\log\Gamma(x) \geq -x + (x - \frac{1}{2})\log x$, the right-hand side of (3.2) is bounded above by

$$C\Delta_n^{-1/2}\exp\left[-\frac{k-1}{2}\{\Delta_n(1-\eta_n) - 1 - \log\Delta_n\}\right],$$

implying that, for all $n \geq n_1$,

$$P_0(S \geq K) \leq \sum_{k=K}^{m_n} P_0 \left(\sum_{j=2}^{k} \left\{ n^{-1/2} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2 \geq (k-1)\Delta_n \right)$$

$$\leq C\Delta_n^{-1/2} \sum_{k=K}^{m_n} \exp\left[-\frac{k-1}{2} \{\Delta_n(1-\eta_n) - 1 - \log\Delta_n\} \right].$$

In view of the assumption $\lim_{n\to\infty} \frac{m_n^3 \Delta_n}{n} = 0$ we have $\lim_{n\to\infty} \eta_n = 0$ and hence, for each $\varepsilon > 0$ there exists $K(= K_\varepsilon)$ such that, for sufficiently large $n, P_0(S \geq K) \leq \varepsilon$, which means that $S = O_{P_0}(1)$.

Noting that under the null hypothesis T_k converges to a chi-square distribution with k degrees of freedom, it now immediately follows that $T_S = O_{P_0}(1)$, as was to be proved. □

The next theorem concerns the behavior of S under alternatives. Let P denote the alternative distribution of the X_i's. Suppose that

$$E_P\phi_1(X_1) = \cdots = E_P\phi_{K-1}(X_1) = 0, \quad E_P\phi_K(X_1) \neq 0 \text{ for some } K = K(P). \tag{3.3}$$

For every alternative of interest there will be a K such that (3.3) holds. It will be assumed that

$$\liminf_{n\to\infty} m_n \geq K,$$

which is certainly the case if $\lim_{n\to\infty} m_n = \infty$, since K is fixed.

Theorem 3.3 *If* (3.3) *holds,* $\liminf_{n\to\infty} m_n \geq K$ *and* $\Delta_n = o(n)$ *as* $n \to \infty$*, then*

$$\lim_{n\to\infty} P(S \geq K) = 1.$$

Proof By the law of large numbers we have

$$\bar{\phi}_K \xrightarrow{P} E_P\phi_K(X_1) \text{ as } n \to \infty.$$

Hence, we obtain

$$T_K - K\Delta_n \geq n\bar{\phi}_K^2 - K\Delta_n \xrightarrow{P} \infty \text{ as } n \to \infty. \tag{3.4}$$

On the other hand,

$$\left(n^{-1/2} \sum_{i=1}^{n} \phi_1(X_i), \ldots, n^{-1/2} \sum_{i=1}^{n} \phi_{K-1}(X_i) \right) \xrightarrow{P} U,$$

where U is a multivariate normal distribution with expectation vector equal to 0. This implies that $T_k = O_P(1)$ for all $k = 1, \dots, K-1$ and hence, by (3.4),

$$\lim_{n\to\infty} \sum_{k=1}^{K-1} P(T_k - k\Delta_n \geq T_K - K\Delta_n) = 0.$$

Since, for $m_n \geq K$,

$$P(S < K) \leq \sum_{k=1}^{K-1} P(T_k - k\Delta_n \geq T_K - K\Delta_n),$$

the result follows. □

Theorem 3.3 is now applied to get the key result under alternatives for proving consistency.

Theorem 3.4 *If* (3.3) *holds,* $\liminf_{n\to\infty} m_n \geq K, \Delta_n = o(n)$ *and* n^{-1} *log* $m_n \to 0$ *as* $n \to \infty$*, then*

$$T_S \xrightarrow{P} \infty \text{ as } n \to \infty.$$

Proof Let $x > 0$. By Theorem 3.3 and noting that $T_S \geq T_K$ if $S \geq K$, we get

$$P(T_S \leq x) = \sum_{j=K}^{m_n} P(T_S \leq x, S = j) + o(1) \tag{3.5}$$

$$\leq m_n P(T_K \leq x) + o(1).$$

Since $E_P\phi_K(X_1) \neq 0$, Chernoff's theorem implies

$$\lim_{n\to\infty} -n^{-1}\log P\left(\left\{n^{-1/2}\sum_{i=1}^{n}\phi_K(X_i)\right\}^2 \leq x\right) > 0. \tag{3.6}$$

Using n^{-1} log $m_n \to 0$ as $n \to \infty$, combination of (3.5) and (3.6) yields the result. □

The next theorem gives the consistency of T_S against any alternative of the form (3.3), thus including essentially any alternative of interest.

Theorem 3.5 *Assume* $\Delta_n \geq 2$ *and* $\lim_{n\to\infty} \frac{m_n^3 \Delta_n}{n} = 0$*. The test based on* T_S *is consistent against any alternative of the form* (3.3)*, provided that* $\liminf_{n\to\infty} m_n \geq K$*.*

Proof Since $\lim_{n\to\infty} \frac{m_n^3 \Delta_n}{n} = 0$, we have $\Delta_n = o(n)$ and $n^{-1} \log m_n \to 0$ as $n \to \infty$ and hence the conditions of Theorems 3.2 and 3.4 are fulfilled. The result follows immediately from application of these theorems. □

It is seen from Theorem 3.5 that consistency holds against essentially any alternative of interest irrespective of the penalty, starting from the one corresponding to Akaike's criterion ($\Delta_n = 2$), up to even much larger penalties than the one in Schwarz's criterion ($\Delta_n = \log n$), provided that we take a reasonable value of m_n. (For $2 \leq \Delta_n \leq \log n$ we may take $m_n = o((n/\log n)^{1/3})$.)

Further note that the choice of the penalty gives no restriction on the class of alternatives, against which consistency is obtained.

4 Asymptotic optimality

In the previous section it was shown that under very mild conditions for almost all kind of penalties the data driven tests are consistent against essentially any alternative of interest. In this section the analysis goes a step deeper by considering the concept of vanishing shortcoming.

The shortcoming of a test is the difference between the highest obtainable power and the power of the test under consideration. More precisely, let $\beta_n^+(\alpha_n, p_n)$ be the power of the level-α_n most powerful test of P_0 (the uniform distribution) against P_n, where the dependence on P_n is denoted by its density p_n. The function $\beta_n^+(\alpha_n, p_n)$ is called the envelope power function. Similarly, we write $\beta_n(\alpha_n, p_n)$ for the power of the level-α_n data driven test. The *shortcoming* of the level-α_n data driven test is now defined as

$$R_n(\alpha_n, p_n) = \beta_n^+(\alpha_n, p_n) - \beta_n(\alpha_n, p_n).$$

By definition, the shortcoming is non-negative and asymptotic optimality in the sense of vanishing shortcoming is obtained when

$$R_n(\alpha_n, p_n) \to 0 \text{ as } n \to \infty.$$

It is shown in Inglot et al. (1998) that the data driven test with penalty $\Delta_n = \log n$, according to Schwarz's rule, is asymptotically optimal in the sense of vanishing shortcoming for a large set of alternatives. This result agrees with the asymptotic optimality of the data driven test with penalty $\Delta_n = \log n$ in the sense of intermediate efficiency, as obtained by IL96 and I99. This is not surprising in view of the close relationship between the two notions of asymptotic optimality, as shown in Inglot et al. (2000). Because of this close relationship we restrict attention here to one of these concepts: the more direct comparison of power by the concept of vanishing shortcoming.

The levels of the tests, denoted by α_n, are assumed to converge to 0. It is argued in Inglot and Ledwina (1998) that such an intermediate approach is better suited than the classical Pitman and Bahadur approaches [see also Kallenberg (1999)].

Our aim in this section is to investigate to what extend the penalty can be chosen such that in "all" directions asymptotically optimal power is obtained. To avoid too much technicalities, while still considering essentially all kind of directions, we restrict attention to contamination alternatives of the following form. Let P_n be an alternative with density

$$p_n(x) = 1 + n^{-\xi} g(x),$$

where g is bounded. Here ξ belongs to the interval $(0, 1/2)$, thus ensuring that we have intermediate alternatives between fixed alternatives ($\xi = 0$) and contiguous alternatives ($\xi = 1/2$). Equivalently, writing $g(x) = \rho a(x)$ with

$$\rho = \sqrt{\int g^2(u)du} \text{ and } a(x) = \frac{g(x)}{\sqrt{\int g^2(u)du}},$$

the alternative denisity can be expressed as

$$p_n(x) = 1 + n^{-\xi} \rho a(x).$$

Note that $\rho > 0$ and that $\int a(u)du = 0$, $\int a^2(u)du = 1$ and $\sup\{|a(u)| : 0 \leq u \leq 1\} < \infty$. Further, we assume that $a \in W_2^1$, where W_2^1 is the Sobolev space of absolutely continuous functions whose derivatives belong to $L_2[0, 1]$.

Summarizing, the set of alternatives under consideration is given by

$$\mathcal{P} = \{\{p_n(x)\} : p_n(x) = 1 + n^{-\xi} \rho a(x), \xi \in (0, \tfrac{1}{2}), \rho > 0, a \in \mathcal{A}\},$$

where

$$\mathcal{A} = \{a : [0, 1] \to \mathbb{R} : a \in W_2^1, \int a(u)du = 0, \int a^2(u)du = 1\}.$$

Note that $a \in W_2^1$ implies that a is bounded on $[0, 1]$.

Our first theorem in this section concerns the null distribution of T_S and is an extension of the result proved in Theorem 3.2. It gives the moderate deviation behavior of T_S, which is needed, because we are dealing with levels α_n tending to 0.

Theorem 4.1 *Let $\{x_n\}$ satisfy $x_n \to 0, nx_n^2 \to \infty$. Further, assume that $m_n x_n \to 0, x_n \to 0, \Delta_n \geq 2, nx_n^2 \geq \Delta_n m_n$. Then we have*

$$P_0(T_S \geq nx_n^2)$$

$$\leq \exp\left\{-\frac{nx_n^2}{2} + O(m_n nx_n^3) + O(1)\right\} \sum_{k=2}^{m_n} \left(\frac{enx_n^2}{k}\right)^{k/2}$$

$$\leq \exp\left\{-\frac{nx_n^2}{2} + \frac{m_n}{2}\log(enx_n^2) + O(m_n nx_n^3) + O(1)\right\}$$

as $n \to \infty$.

Proof We apply Theorem 3.1 with $Y_i = (\phi_1(X_i), \ldots, \phi_k(X_i)), L = \sqrt{k(k+2)}$, cf. also (3.1), and $y = \sqrt{nx_n^2}$. Since $nx_n^2 \geq \Delta_n m_n$ and $\Delta_n \geq 2$, we have $y^2 \geq 2k$ for all $k = 1, \ldots, m_n$, while $m_n x_n \to 0$ implies $y^2 \leq nL^{-2}$ for sufficiently large n. Hence, we get

$$P_0(T_S \geq nx_n^2) \leq \sum_{k=1}^{m_n} P_0(T_k \geq nx_n^2)$$

$$\leq \exp\left\{-\frac{nx_n^2}{2} + O(m_n nx_n^3) + O(1)\right\} \sum_{k=1}^{m_n} \frac{1}{\Gamma(k/2)} \left(\frac{nx_n^2}{2}\right)^{(k-1)/2}.$$

Because $\log\Gamma(x) \geq -x + (x - \frac{1}{2})\log x$, we arrive at

$$P_0(T_S \geq nx_n^2) \leq \exp\left\{-\frac{nx_n^2}{2} + O(m_n nx_n^3) + O(1)\right\} \sum_{k=1}^{m_n} \left(\frac{enx_n^2}{k}\right)^{k/2}$$

$$\leq \exp\left\{-\frac{nx_n^2}{2} + \frac{m_n}{2}\log(enx_n^2) + O(m_n nx_n^3) + O(1)\right\},$$

as was to be proved. □

The next theorem concerns the distribution of T_S under alternatives. It will be shown that T_S is asymptotically normal. Before presenting this result, we define a kind of deterministic counterpart of the selection rule S and show that under the alternatives the selection rule is (with probability tending to 1) as least as large as this quantity.

Let $A_n > 0$. Write $a_j = \int a(u)\phi_j(u)du$ and $a = (a_1, a_2, \ldots)$. Define $l_n = l_n(A_n)$ by

$$l_n = \min\{k \leq m_n : \rho^2||a||_k^2 - A_n k\Delta_n^{2\xi-1} \geq \rho^2||a||_j^2 - A_n j\Delta_n n^{2\xi-1} \text{for} j = 1, \ldots, m_n\}.$$

The number A_n is chosen in an appropriate way in the proof of the theorem. Replacing $\bar{\phi}_j = n^{-1}\sum_{i=1}^n \phi_j(X_i)$ by its expectation $E_{P_n}\phi_j(X) = \int n^{-\xi}\rho a(u)\phi_j(u)du$ in T_k and taking furthermore $A_n = 1$, gives that the selection rule S equals l_n. In this sense l_n may be viewed as a deterministic counterpart of S.

Proposition 4.2 *Assume that* $m_n \to \infty$, $\dfrac{m_n \log^{1/3} n}{n^{1/3}} \to 0$ *and* $\dfrac{m_n \log n}{n^{1-2\xi}} \to 0$. *Let* $A_n > 0$ *be such that, for some* $\varepsilon > 0$,

$$m_n = o\left(exp\left\{ \frac{(A_n^{1/2} - 1)^2 \Delta_n}{2+\varepsilon} \right\} \right) \text{ and } \frac{(A_n^{1/2} - 1)^2 \Delta_n m_n}{n} \to 0.$$

Then $P_n(S \geq l_n(A_n)) \to 1$.

Proof By definition of S we get, writing shortly l instead of $l_n(A_n)$,

$$P_n(S < l) \leq P_n(\exists j < l : T_j - \Delta_n j \geq T_l - \Delta_n l)$$

$$\leq \sum_{j=1}^{l-1} P_n\left(||\bar{\phi}||_{jl}^2 \leq (l-j)\frac{\Delta_n}{n} \right).$$

Let $\tilde{a} = n^{-\xi}\rho a$. Then $\tilde{a}_j = \int \tilde{a}(u)\phi_j(u)du = E_{P_n}\phi_j(X)$. By the triangle inequality we obtain

$$||\tilde{a}||_{jl} - ||\bar{\phi} - \tilde{a}||_{jl} \leq ||\bar{\phi}||_{jl}$$

and hence,

$$P_n(S < l) \leq \sum_{j=1}^{l-1} P_n\left(||\bar{\phi} - \tilde{a}||_{jl} \geq ||\tilde{a}||_{jl} - \left\{ (l-j)\frac{\Delta_n}{n} \right\}^{1/2} \right).$$

In view of the definition of l we have for all $1 \leq j < l$

$$||\tilde{a}||_{jl} > A_n^{1/2} \left\{ (l-j)\frac{\Delta_n}{n} \right\}^{1/2},$$

implying

$$P_n(S < l) \leq \sum_{j=1}^{l-1} P_n\left(||\bar{\phi} - \tilde{a}||_{jl} \geq (A_n^{1/2} - 1)\left\{ (l-j)\frac{\Delta_n}{n} \right\}^{1/2} \right).$$

Since we want to apply Lemma 5.2 in I99, we check the assumptions of that lemma. Our alternatives belong to $\mathcal{P}_1$, see page 494 in I99; further, (3.4) and (3.5) hold, see page 496 of I99, while (3.6) holds by the above assumption

$$\frac{m_n \text{log} n}{n^{1-2\xi}} \to 0.$$

Hence, $P_n \in \mathcal{P}_{\Phi m}$. Moreover, as stated on page 495 of I99, conditions (A1) and (A2) are fulfilled for the Legendre polynomials (with $\omega = 1/2$ and $\tau = 0$) and (A3) holds by the above assumption

$$m_n \to \infty, \frac{m_n \text{log}^{1/3} n}{n^{1/3}} \to 0.$$

Finally, note that, by assumption,

$$\frac{(A_n^{1/2}-1)^2\Delta_n m_n}{n} \to 0$$

and hence, (5.22) on page 504 of I99 holds.

Application of Lemma 5.2 in I99 yields

$$P_n(S<l) \le 2\sum_{j=1}^{l-1} \exp\left\{-\frac{(A_n^{1/2}-1)^2\Delta_n}{2}(1+o(1))\right\},$$

where $o(1)$ does not depend on j (or l). The assumption

$$m_n = o\left(\exp\left\{\frac{(A_n^{1/2}-1)^2\Delta_n}{2+\varepsilon}\right\}\right)$$

gives the result. □

Theorem 4.3 *Assume that* $m_n \to \infty, \frac{m_n \log^{1/3} n}{n^{1/3}} \to 0, \frac{m_n \log n}{n^{1/2-\xi}} \to 0, \frac{m_n}{n^\xi} \to 0, \Delta_n \ge 2$ *and* $\frac{m_n\Delta_n}{n^{1/2-\xi}} \to 0$. *Then we have*

$$\lim_{n\to\infty} P_n\left(\frac{T_S - n^{1-2\xi}\rho^2||a||_{m_n}^2}{2n^{1/2-\xi}\rho||a||_{m_n}} \le x\right) = \Phi(x). \tag{4.1}$$

If moreover, $n^{1/2-\xi}\left(||a||_{m_n}^2 - 1\right) \to 0$, *then*

$$\lim_{n\to\infty} P_n\left(\frac{T_S - n^{1-2\xi}\rho^2}{2n^{1/2-\xi}\rho} \le x\right) = \Phi(x).$$

Proof The first step in the proof is an application of Proposition 4.1 in I99. In order to do that we check the assumptions of this proposition according to Remark 5.4 on page 507 of I99. As shown in the preceding proof of Proposition 4.2 we have $P_n \in \mathcal{P}_{\Phi m}$ and the conditions (A1)-(A3) hold. Finally, (4.1) of I99 follows from our condition $m_n n^{-\xi} \to 0$ and since $||a||_{m_n} \to 1$ and $\rho > 0$, (4.4) is fulfilled provided that $k_n \ge K$, where K is the smallest integer j for which $a_j \ne 0$.

Therefore, writing again $\tilde{a} = n^{-\xi}\rho a$, we get the following result. For each sequence $K \le k_n \le m_n$ we have

$$\frac{T_{k_n} - n||\tilde{a}||_{k_n}^2 - k_n}{2\sqrt{n}||\tilde{a}||_{k_n}} = \frac{T_{k_n} - n^{1-2\xi}\rho^2||a||_{k_n}^2 - k_n}{2n^{1/2-\xi}\rho||a||_{k_n}} \xrightarrow{D} N(0,1) \text{ under } P_n.$$

To apply Proposition 4.2 we choose $A_n > 0$ such that for some $\varepsilon > 0$,

$$m_n = o\left(\exp\left\{\frac{(A_n^{1/2}-1)^2\Delta_n}{2+\varepsilon}\right\}\right) \text{ and } \frac{A_n m_n \Delta_n}{n^{1/2-\xi}} \to 0. \tag{4.2}$$

Note that

$$\frac{A_n m_n \Delta_n}{n^{1/2-\xi}} \to 0 \text{ and the condition } \frac{m_n \Delta_n}{n^{1/2-\xi}} \to 0 \text{ imply} \frac{(A_n^{1/2}-1)^2 \Delta_n m_n}{n} \to 0.$$

It is seen that under the conditions of the theorem such A_n's always can be chosen. For instance, if we take

$$(A_n^{1/2}-1)^2 \Delta_n = \log n,$$

the first requirement in (4.2) is fulfilled, because $m_n = o(n^\xi)$ and $\xi < 1/2$, and the second requirement in (4.2) holds, because

$$\frac{A_n m_n \Delta_n}{n^{1/2-\xi}} \leq \frac{2m_n \Delta_n}{n^{1/2-\xi}} + \frac{2m_n \log n}{n^{1/2-\xi}} \to 0.$$

By Proposition 4.2 we have $P_n(S < l_n) \to 0$ with $l_n = l_n(A_n)$ and A_n satisfying (4.2). Now we get

$$P_n\left(\frac{T_S - n||\tilde{a}||_{m_n}^2}{2\sqrt{n}||\tilde{a}||_{m_n}} \leq x\right) \leq P_n(S < l_n) + P_n\left(\frac{T_{l_n} - n||\tilde{a}||_{m_n}^2}{2\sqrt{n}||\tilde{a}||_{m_n}} \leq x\right)$$

and

$$P_n\left(\frac{T_S - n||\tilde{a}||_{m_n}^2}{2\sqrt{n}||\tilde{a}||_{m_n}} \leq x\right) \geq P_n\left(\frac{T_{m_n} - n||\tilde{a}||_{m_n}^2}{2\sqrt{n}||\tilde{a}||_{m_n}} \leq x\right).$$

Hence,

$$P_n\left(\frac{T_S - n||\tilde{a}||_{m_n}^2}{2\sqrt{n}||\tilde{a}||_{m_n}} \leq x\right) \to \Phi(x),$$

provided that

$$\frac{m_n}{\sqrt{n}||\tilde{a}||_{m_n}} \to 0 \text{ and } P_n\left(\frac{T_{l_n} - n||\tilde{a}||_{m_n}^2}{2\sqrt{n}||\tilde{a}||_{m_n}} \leq x\right) \to \Phi(x).$$

Noting that

$$\frac{m_n}{n^{1/2-\xi}} \to 0, \text{ since } \frac{m_n \Delta_n}{n^{1/2-\xi}} \to 0 \text{ and } \Delta_n \geq 2,$$

it follows that

$$\frac{m_n}{\sqrt{n}||\tilde{a}||_{m_n}} = \frac{m_n}{n^{1/2-\xi}\rho||a||_{m_n}} \to 0. \tag{4.3}$$

Using, cf. (4.2),

$$\frac{A_n m_n \Delta_n}{n^{1/2-\xi}} \to 0,$$

$||a||_{m_n} \to 1$ and the definition of l_n, we get

$$\frac{n||\tilde{a}||^2_{l_n m_n}}{\sqrt{n}||\tilde{a}||_{m_n}} = \frac{n^{1-2\xi}\rho^2||a||^2_{l_n m_n}}{n^{1/2-\xi}\rho||a||_{m_n}} \leq \frac{A_n(m_n - l_n)\Delta_n}{n_l^{1/2-\xi}||a||_{m_n}} \to 0.$$

As a consequence we have $n^{1/2-\xi}||a||^2_{l_n m_n} \to 0$, implying $||a||^2_{l_n m_n} \to 0$. Therefore, we get $l_n \geq K$, for sufficiently large n. Moreover, it follows that

$$\frac{||\tilde{a}||_{l_n}}{||\tilde{a}||_{m_n}} \to 1.$$

So, we obtain, cf. also (4.3),

$$\frac{T_{l_n} - n||\tilde{a}||^2_{m_n}}{2\sqrt{n}||\tilde{a}||_{m_n}}$$

$$= \frac{T_{l_n} - n||\tilde{a}||^2_{l_n} - l_n}{2\sqrt{n}||\tilde{a}||_{l_n}} \times \frac{||\tilde{a}||_{l_n}}{||\tilde{a}||_{m_n}} - \frac{n||\tilde{a}||^2_{l_n m_n}}{2\sqrt{n}||\tilde{a}||_{m_n}} + \frac{l_n}{2\sqrt{n}||\tilde{a}||_{m_n}} \xrightarrow{D} N(0,1).$$

This completes the proof of (4.1).

To prove the second statement of the theorem, note that

$$\frac{T_S - n^{1-2\xi}\rho^2}{2n^{1/2-\xi}\rho} = \frac{T_S - n^{1-2\xi}\rho^2||a||^2_{m_n}}{2n^{1/2-\xi}\rho||a||_{m_n}} \times ||a||_{m_n} + \frac{n^{1-2\xi}\rho^2\left(||a||^2_{m_n} - 1\right)}{2n^{1/2-\xi}\rho}$$

and hence the result follows by the condition

$$n^{1/2-\xi}\left(||a||^2_{m_n} - 1\right) \to 0.$$

This completes the proof of the theorem. □

Remark 4.1 Denote by W_2^r the Sobolev space of functions f on $[0,1]$ for which $f^{(r-1)}$ is absolutely continuous and whose r^{th} derivative belongs to $L_2[0,1]$. If $a \in W_2^r$ for some $r \geq 1$, then by (7.4) in Barron and Sheu (1991) we have

$$||a||^2_{m_n} - 1 = O(m_n^{-2r}).$$ □

The following theorems describe the behavior of the most powerful test against a given alternative. The standardized Neyman-Pearson statistic for testing P_0 against P_n is given by

$$W_n = (n^{1/2}\sigma_{0n})^{-1}\sum_{i=1}^{n}\{\log\, p_n(X_i) - e_{0n}\},$$

where

$$e_{0n} = E_0 \log\, p_n(X_1) \text{and } \sigma^2_{0n} = \ \text{var}_0 \log\, p_n(X_1).$$

We reject for large values of W_n. Further, we write

$$b(p_n) = \sigma_{0n}^{-1}\{E_{P_n} \log p_n(X_1) - e_{0n}\}.$$

We start with the behavior of W_n under the null hypothesis. It is an immediate consequence of a Cramér-type large deviation result obtained by Book (1976) [cf. Lemma 4.1 in Jurečkova, Kallenberg and Veraverbeke (1998)], Lemma 5.4 and Proposition 5.12 in IL96. [See also Inglot and Ledwina (1999), Theorem 3.2 (1)].

Theorem 4.4 *Let $\{z_n\}$ be a sequence satisfying $z_n \to \infty$ and $z_n = o(n^{1/2})$. Then*

$$P_0(W_n \geq z_n) = (2\pi z_n^2)^{-1/2}\exp\{-\tfrac{1}{2}z_n^2 + O(n^{-1/2}z_n^3)\}. \qquad \square$$

Application of Proposition 6.6 and Lemma 5.4 in IL96 gives the behavior of W_n under alternatives. [See also Inglot and Ledwina (1999), Theorem 3.1 (1)].

Theorem 4.5 *For any $x \in \mathbb{R}$, we have*

$$\lim_{n\to\infty} P_n(W_n - n^{1/2}b(p_n) \leq x) = \Phi(x). \qquad \square$$

The basic result on the vanishing shortcoming of the data driven tests is given in the following theorem. It concerns those α_n's for which the asymptotic power stays away from 0 and 1. Theorem 4.7 shows that in fact for all α_n's the shortcoming tends to 0.

Theorem 4.6 *Let $\{t_n\}$ be a sequence of real numbers satisfying*

$$-\infty < \liminf_{n\to\infty} t_n \leq \limsup_{n\to\infty} t_n < \infty.$$

Set

$$\alpha_n = P_0\left(\frac{T_S - n^{1-2\xi}\rho^2||a||_{m_n}^2}{2n^{1/2-\xi}\rho||a||_{m_n}} \geq t_n\right).$$

Assume that $m_n \to \infty, \dfrac{m_n \log^{1/3} n}{n^{1/3}} \to 0, \dfrac{m_n \log n}{n^{1/2-\xi}} \to 0, m_n n^{1/2-2\xi} \to 0, \dfrac{m_n}{n^\xi} \to 0, n^{1/2-\xi}\left(||a||_{m_n}^2 - 1\right) \to 0, \Delta_n \geq 2$ *and* $\dfrac{m_n \Delta_n}{n^{1/2-\xi}} \to 0$. *Then*

$$\lim_{n\to\infty} R_n(\alpha_n, p_n) = 0.$$

Proof Let $x_n^2 = n^{-2\xi}\rho^2||a||_{m_n}^2 + 2t_n n^{-1/2-\xi}\rho||a||_{m_n}$. Since $\xi < 1/2$, it is seen that $x_n = n^{-\xi}\rho(1 + o(1))$ as $n \to \infty$. Hence, the conditions of Theorem 4.1 are fulfilled, yielding

$$\log \alpha_n \leq -\frac{nx_n^2}{2} + O(m_n \log n + m_n n^{1-3\xi}). \tag{4.4}$$

Define w_n by

$$\alpha_n = P_0(W_n - n^{1/2}b(p_n) \geq w_n).$$

Since the power of the most powerful test is at least as large as the power of the test based on T_S, it follows from the definition of α_n, Theorems 4.3 and 4.5 that w_n does not tend to ∞. Direct calculation gives

$$b(p_n) = \rho n^{-\xi} + O(n^{-2\xi}).$$

Therefore, we may apply Theorem 4.4 with $z_n = \frac{1}{2}n^{1/2}b(p_n)$, yielding

$$P_0(W_n - n^{1/2}b(p_n) \geq -\tfrac{1}{2}n^{1/2}b(p_n)) = \exp\{-\tfrac{1}{8}n^{1-2\xi}\rho^2(1+o(1))\}.$$

From (4.4) we infer that

$$\log \alpha_n \leq -\tfrac{1}{2}n^{1-2\xi}\rho^2(1+o(1))$$

and hence $w_n \geq -\frac{1}{2}n^{1/2}b(p_n)$ for sufficiently large n. Therefore, we may apply Theorem 4.4 with $z_n = n^{1/2}b(p_n)+w_n$, which is of exact order $n^{1/2-\xi}$. This yields, using $w_n = O(n^{1/2}b(p_n)) = O(n^{1/2-\xi})$,

$$\begin{aligned} \log \alpha_n &= -\tfrac{1}{2}[n^{1/2}b(p_n) + w_n]^2 + O(n^{1-3\xi} + \log n) \\ &= -\tfrac{1}{2}[n^{1/2-\xi}\rho + w_n]^2 + O(n^{1-3\xi} + \log n). \end{aligned} \tag{4.5}$$

By the condition

$$n^{1/2-\xi}\left(||a||_{m_n}^2 - 1\right) \to 0,$$

we have

$$x_n^2 = n^{-2\xi}\rho^2||a||_{m_n}^2 + 2t_n n^{-1/2-\xi}\rho||a||_{m_n}$$

$$= n^{-2\xi}\rho^2 + 2t_n n^{-1/2-\xi}\rho + o(n^{-1/2-\xi})$$

and therefore, (4.4) gives

$$\log \alpha_n \leq -\tfrac{1}{2}n^{1-2\xi}\rho^2 - t_n n^{1/2-\xi}\rho + O(m_n \log n + m_n n^{1-3\xi}) + o(n^{1/2-\xi}).$$

Combination with (4.5) leads to

$$-\tfrac{1}{2}[n^{1/2-\xi}\rho + w_n]^2$$

$$\leq -\tfrac{1}{2}n^{1-2\xi}\rho^2 - t_n n^{1/2-\xi}\rho + O(m_n \log n + m_n n^{1-3\xi}) + o(n^{1/2-\xi})$$

and hence,

$$0 \leq n^{1/2-\xi}\rho w_n + \tfrac{1}{2}w_n^2 - t_n n^{1/2-\xi}\rho + O(m_n \log n + m_n n^{1-3\xi}) + o(n^{1/2-\xi}). \quad (4.6)$$

Writing

$$\tilde{w}_n = \frac{w_n}{n^{1/2-\xi}\rho},$$

(4.6) implies

$$0 \leq \tilde{w}_n + \tfrac{1}{2}\tilde{w}_n^2 + o(1).$$

Since $w_n \geq -\frac{1}{2}n^{1/2}b(p_n)$ for sufficiently large n, we get $\liminf_{n\to\infty} \tilde{w}_n \geq -\frac{1}{2}$. Using the fact that the function $w + \frac{1}{2}w^2$ is negative on $[-\frac{1}{2}, 0)$, it follows that $\liminf_{n\to\infty} \tilde{w}_n \geq 0$. On the other hand, w_n is bounded from above and hence $\limsup_{n\to\infty} \tilde{w}_n \leq 0$. Thus we get $\lim_{n\to\infty} \tilde{w}_n = 0$, or, equivalently, $w_n = o(n^{1/2-\xi})$.

Returning to (4.6), we obtain

$$0 \leq n^{1/2-\xi}\rho w_n(1 + o(1)) - t_n n^{1/2-\xi}\rho + O(m_n \log n + m_n n^{1-3\xi}) + o(n^{1/2-\xi})$$

and hence

$$0 \leq w_n(1 + o(1)) - t_n + o(1). \quad (4.7)$$

Combination of (4.7) with Theorems 4.3 and 4.5 completes the proof. □

Theorem 4.7 *Assume that* $m_n \to \infty$, $\frac{m_n \log^{1/3} n}{n^{1/3}} \to 0$, $\frac{m_n \log n}{n^{1/2-\xi}} \to 0$, $m_n n^{1/2-2\xi} \to 0$, $\frac{m_n}{n^{\xi}} \to 0$, $n^{1/2-\xi}\left(||a||^2_{m_n} - 1\right) \to 0$, $\Delta_n \geq 2$ *and* $\frac{m_n \Delta_n}{n^{1/2-\xi}} \to 0$. *Then*

$$\lim_{n\to\infty} R_n(\alpha_n, p_n) = 0.$$

Proof The proof of this theorem is completely similar to the proof of Theorem 5.2 in Inglot et al. (1998). □

It is seen from Theorem 4.7 that vanishing shortcoming holds in "all" directions, irrespective of the penalty, starting from the one corresponding to Akaike's criterion ($\Delta_n = 2$), up to even much larger penalties than the one in Schwarz's criterion ($\Delta_n = \log n$), provided that we take a reasonable value of m_n. (For $2 \leq \Delta_n \leq \log n$ and for e.g. $\xi = 1/3$ we may take $m_n = n^{\varsigma}$ with $0 < \varsigma < 1/6$, provided that $n^{1/2-\xi}\left(||a||^2_{m_n} - 1\right) \to 0$; thus we obtain vanishing shortcoming in an infinite number of orthogonal directions.)

Further, note that for $2 \leq \Delta_n \leq \log n$ the choice of the penalty gives no extra restriction on the class of alternatives, for which vanishing shortcoming is obtained.

References

Albers, W., Kallenberg, W. C. M. and Martini, F.(1999), Data driven rank tests for classes of tail alternatives. Memorandum 1504, Faculty of Mathematical Sciences, University of Twente.

Barron, A. R. and Cover, T. M. (1991). Minimum complexity density estimation. *IEEE Trans. Inform. Theory* **37** 1034-1054.

Barron, A. R. and Sheu, C. (1991). Approximation of density functions by sequences of exponentially families. *Ann. Statist.* **19** 1347-1369.

Bickel, P. J. and Ritov, Y. (1992). Testing for goodness of fit: a new approach. In *Nonparametric Statistics and Related Topics* (A. K. Md. E. Saleh, ed.) 51-57. North-Holland, Amsterdam.

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

Bogdan, M. (1999). Data driven smooth tests for bivariate normality. *J. Multivariate Anal.* **68** 26-53.

Bogdan, K., Bogdan, M. and Futschik, A. (1999). A data driven smooth test for uniformity on circle. Technical Report 65, Institute of Mathematics, Wrocław University of Technology.

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

Book, S. A. (1976). The Cramér-Feller-Petrov large deviation theorem for triangular arrays. Technical Report, Department of Mathematics, California State College, Dominguez Hills.

Eubank, R.L. (1997). Testing goodness of fit with multinomial data. *J. Amer. Statist. Assoc.* **92** 1084-1093.

Eubank, R.L. and LaRiccia, V.N. (1992). Asymptotic comparison of Cramér-von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Statist.* **20** 2071-2086.

Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 647-688.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc.* Ser. B **41** 190-195.

Haughton, D., Haughton, J. and Izenman, A. (1990). Information criteria and harmonic models in time series analysis. *J. Statist. Comput. Simul.* **35** 187-207.

Hušková, M. and Sen, P.K. (1985). On sequentially adaptive asymptotically efficient rank statistics. *Sequential Anal.* **4** 125-151.

Hušková, M. and Sen, P.K. (1986). On sequentially adaptive signed rank statistics. *Sequential Anal.* **5** 237-252.

Inglot, T. (1999). Generalized intermediate efficiency of goodness-of-fit tests. *Math. Methods Statist.* **8** 487-509.

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

Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (1998). Vanishing shortcoming of data driven Neyman's tests. In *Asymptotic Methods in Probability and Statistics*, A Volume in Honour of Miklós Csörgő, B. Szyszkowicz ed. 811-829. Elsevier Science, Amsterdam.

Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (2000). Vanishing shortcoming and asymptotic relative efficiency. *Ann. Statist.* **28** 215-238.

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

Inglot, T. and Ledwina, T. (1998). Intermediate approach to comparison of some goodness-of-fit tests, Technical Report 61, Institute of Mathematics, Wrocław University of Technology.

Inglot, T. and Ledwina, T. (1999). Vanishing shortcoming of data driven smooth tests for location-scale family, Technical Report 58, Institute of Mathematics, Wrocław University of Technology.

Janic-Wróblewska, A. (1999a). Data driven smooth test for a location-scale family. Technical Report 21, Institute of Mathematics, Wrocław University of Technology.

Janic-Wróblewska, A. (1999b). Data driven smooth tests for the extreme value distribution. Technical Report 22, Institute of Mathematics, Wrocław University of Technology.

Janic-Wróblewska, A. and Ledwina (2000). Data driven rank test for two-sample problem. *Scand. J. Statist.* **27** 281-297.

Janssen, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.* **28** 239-253.

Jurečkova, J., Kallenberg, W. C. M. and Veraverbeke, N. (1988). Moderate and Cramér-type large deviation theorems for M-estimators. *Statist. Probab. Lett.* **6** 191-199.

Kallenberg, W. C. M. (1999). Intermediate or Kallenberg efficiency. In *Encyclopedia of Statistical Sciences*, Vol. 3, (S. Kotz, ed.) 192-197. Wiley-Interscience, New York.

Kallenberg, W. C. M. and Ledwina, T. (1995a). On data driven Neyman's tests. *Probab. Math. Statist.* **15** 409-426.

Kallenberg, W. C. M. and Ledwina, T. (1995b). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.* **23** 1594-1608.

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

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

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

Kallenberg, W. C. M., Ledwina, T. and Rafajłowicz, E. (1997). Testing bivariate independence and normality. *Sankhyā* Ser. A **59** 42-59.

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

Prohorov, A. V. (1973). On sums of random vectors. *Theory Probab. Appl.* **18** 186-188.