

---

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](mailto:memo@math.utwente.nl)

---

MEMORANDUM No. 1488

Power gain by pre-testing?

W. ALBERS, P.C. BOON AND W.C.M. KALLENBERG

JUNE 1999

ISSN 0169-2690

# Power gain by pre-testing?

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

## Abstract

The aim of this paper is to study whether it is possible to gain power by pre-testing, and to give insight in when this occurs, to what extent and at which price. A pre-test procedure consists of a preliminary test which tests a particular property of a given restricted model, followed by a main test for the main hypothesis regarding the parameter of interest. After acceptance by the preliminary test, a basic main test is used in the restricted model. After rejection by the preliminary test, a more general main test is used which does not use prior information about the underlying distribution. The procedure is analyzed in the model against which the preliminary test protects. For classes of tests including the standard first-order optimal tests, a transparent expression is given for the power and size difference of the pre-test procedure compared to the power and (correct) size of the second main test. This expression is based on second-order asymptotics and gives qualitative and quantitative insight in the behavior of the procedure. It shows that substantial power gain, not merely due to size violation, is possible if the second main test really differs from the basic main test. The smaller the correlation between the two main tests, the larger the power gain.

*Keywords and phrases:* pre-test procedure, power gain, robustness of validity, second-order asymptotics.

*1991 Mathematics Subject Classification:* 62F05, 62E20

# 1 Introduction

In practice, statistical analysis frequently consists of two stages. During the first stage, some model assumptions are checked, while in the second stage the main analysis is performed. This often happens when one wants to test a null hypothesis with a given test, the validity of which relies on certain assumptions. To check these assumptions, one can carry out a preliminary test. If this test accepts, the main test is carried out as intended. But if the preliminary test rejects, a more robust alternative main test has to be used to test the main null hypothesis in a larger model. The procedure just described we shall denote in what follows as a *pre-test procedure*, and we shall be interested in studying its properties. Note that this is largely ignored in practice: usually one does not even realize that the first stage has an impact on the main test(s) from the second stage.

The idea behind such two-stage procedures is attractive in two respects. First of all, in practice people strongly prefer simple procedures and hence like to stick to the basic main test as long as possible. Only if this causes really unacceptable deviations in the size of the main test - which should be detected by the preliminary test - one shifts to the alternative main test, which may be more complicated. The second nice aspect is that in principle a higher power can be achieved if stronger assumptions can be made. Hence again, if possible one wants to apply the basic main test, because one has the idea that always using the robust second main test would imply an unnecessary loss of power in case the model assumptions for the basic main test turn out to hold. As a typical example, consider the use of a preliminary test to choose between the  $t$ -test and the sign test for testing the main hypothesis of a zero median. The main issue of this paper is to investigate in what kind of situations such power gain by the pre-test procedure indeed comes true.

However, there are some questionable aspects about such pre-test procedures. The preliminary test may fail to reject, leading to application of the main test for the smaller and invalid model, and hence to serious deviations in size with respect to the prescribed level. Furthermore, the repeated use of the same data introduces correlations which influence size and power of the combined procedure.

In literature, these drawbacks of pre-test procedures were occasionally recognized, see e.g. Saleh and Sen (1983). But general insight in the magnitude of the effects was lacking, since the available results were mainly based on numerical calculations and simulations for specific examples. Some references are Moser, Stevens, and Watts (1989) and Markowski and Markowski (1990) who consider the Behrens-Fisher problem. A preliminary  $F$ -test on equality of the variances of two normal samples is used in these papers to choose between the two-sample  $t$ -test and e.g. the more general Welch-Satterthwaite test, in testing the main hypothesis of equality of the means.

To obtain more insight, both in a qualitative and a quantitative sense, Albers,

Boon, and Kallenberg (1998b) and Albers, Boon, and Kallenberg (1997a) further analyzed the procedure for the two-sample problem. Using second-order approximations, they derived analytical results for size and power of the separate tests and of the combined procedure. To make clear what is going on, these papers also considered the more transparent one-sample case. This is analogous to the two-sample case, with  $F$ -,  $t$ - and Welch-Satterthwaite test replaced by the  $\chi^2$ -, Gauss -, and one-sample  $t$ -test, respectively.

The normal one-sample case can be considered as a special case of a more general situation where we have a parametric family of densities  $f(x; \theta, \tau)$ , with  $\theta$  the parameter of interest and  $\tau$  a nuisance parameter. A preliminary test is used to determine whether  $\tau$  may be assumed equal to some prescribed value  $\tau_0$ , implying that after acceptance of the null hypothesis  $\tau = \tau_0$ , the main hypothesis concerning  $\theta$  can be tested in the restricted family  $f(x; \theta, \tau_0)$ . Without loss of generality we take  $\tau_0 = 0$ . However, if the preliminary test rejects, the main test on  $\theta$  is performed in the complete family  $f(x; \theta, \tau)$ . For this situation, Albers, Boon, and Kallenberg (1998a) consider a general class of tests, including the standard first-order optimal ones for each of the three testing problems involved. They obtain transparent expressions for both the deviation in size and the gain in power of the pre-test procedure compared to the main test in the complete family. It is demonstrated that in general no substantial power gain can be realized without unacceptable violation of the prescribed level. The key to the explanation of this phenomenon turns out to be the correlation between the test statistic of the preliminary test on one hand, and the test statistics of the main tests on the other. Wild and uncontrollable variation in the size of the pre-test procedure can only be avoided if this correlation is really small (like in the previously studied normal case, where it in fact is zero). But then the two main tests are quite close, and their power difference will be small.

Hence in this situation the virtue of the pre-test procedure is really largely based on the aforementioned simplicity: usually no rejection by the preliminary test occurs (at least, that is what one hopes) and one can stick to the simple basic main test. Only if rejection occurs, one is forced to perform another, possibly more complicated, test.

In the above, not only  $\theta$ , but  $\tau$  as well, has been a one-dimensional parameter. However, for higher-dimensional  $\tau$ , the situation remains essentially the same: as long as optimal tests are used for each of the three components of the problem, we either have a stable size but closely related main tests with little power difference, or we get really different main tests, which opens the way to power gain, but only at the expense of unacceptable deviations in size.

Nevertheless, interesting situations involving power gain do arise, and these will be the subject of the present paper. Denote the density of our observations by  $f$ . The basic main test is performed in a restricted model  $f(x; \theta)$  and concerns testing  $H_0 : \theta = \theta_0$  against  $H_1 : \theta > \theta_0$  in this model for some given value  $\theta_0$ . Without loss of generality we take  $\theta_0 = 0$ . A short remark on notation should be

made. Writing  $f$ , an unknown density is meant, while  $f(x; \theta)$  indicates a member of a family of densities with given (and hence known) function  $f$  and unknown parameter  $\theta$ . To avoid too many different symbols we use the same notation  $f$  with different interpretation; the right one is however obvious from the context.

The restricted model  $f(x; \theta)$  may be incorrect. To check its validity a preliminary test is applied. The preliminary test picks up a particular property of the family  $f(x; \theta)$  and explores whether the data are in agreement with it. For instance, if  $f(x; \theta) = g(x - \theta)$  for some given symmetric density  $g$ , the mean and median in the family  $f(x; \theta)$  are the same and the preliminary test can be devised for testing whether the difference between mean and median equals zero. More generally, the preliminary test protects against a limited number of directions in which the density at hand may deviate from the restricted model. For simplicity, we assume that the deviations against which the preliminary test effectively protects, can be parameterized by a single parameter  $\tau$ , giving back the restricted family  $f(x; \theta)$  for  $\tau = 0$ . Hence, the preliminary test is appropriate in a model  $f(x; \theta, \tau)$  and tests  $\bar{H}_0 : \tau = 0$  against  $\bar{H}_1 : \tau \neq 0$ .

After acceptance by the preliminary test,  $\theta = 0$  is tested in the model  $f(x; \theta) = f(x; \theta, 0)$ . If the null hypothesis  $\bar{H}_0 : \tau = 0$  is rejected, the main hypothesis is tested in a nonparametric model in the following sense. If, for example,  $\theta$  can be interpreted as the median in the family  $f(x; \theta)$ , then we may also test in the nonparametric case, making no further assumptions about our density  $f$ , that the median equals zero by application of the sign test. More generally, if  $\bar{H}_0 : \tau = 0$  is rejected, we test a well defined property corresponding to  $\theta = 0$  without further assumptions about our density  $f$ . This property is often written in the form  $\int \xi f = 0$  for some given function  $\xi$ , e.g.  $\xi(x) = \text{sign}(x)$  for the case of the median. As  $f(x; \theta, \tau)$  is an extension of the restricted family  $f(x; \theta)$ , we should of course also have that  $\theta = 0$  in this family gives  $\int \xi f = 0$ , or, equivalently, that  $\int \xi(x) f(x; 0, \tau) dx = 0$  for all  $\tau$ . In principle, there is a choice which property to take, since usually  $\theta$  can be interpreted in several ways in the family  $f(x; \theta, \tau)$ . The particular choice of  $\xi$  determines the second main test. Note that in contrast to testing  $\theta = 0$  in the family  $f(x; \theta, \tau)$  considered in Albers, Boon, and Kallenberg (1998a), the second main test is often not that complicated, thus resulting in an attractive pre-test procedure from the point of view of simplicity.

Because the preliminary test protects (mainly) against alternatives of the form  $f(x; \theta, \tau)$ , we analyze the pre-test procedure in this model. The reason for possible power gain of the pre-test procedure compared to the second main test, becomes now transparent: typically the second main test provides robustness of validity in a much larger model, resulting in some loss of power in the model  $f(x; \theta, \tau)$ . (Compare again the sign test, which indeed guarantees the correct level for all  $f$ .)

It is the purpose of this paper to reveal the properties of the pre-test procedure, and in particular to clarify when power gain is obtained, while still con-

trolling the size. Although the practical application of the pre-test procedure is straightforward, its analysis is not trivial. In Theorem 3.1, our main result, an attractive expression is presented, giving qualitative and quantitative insight in the behavior of the pre-test procedure. Two quantities play a very important role: (i) the correlation between the basic main test and the preliminary test and (ii) the correlation between the basic main test and the second main test.

- *The correlation between the basic main test and the preliminary test.* The validity of the pre-test procedure, measured by its size, can only be controlled if the correlation between the basic main test and the preliminary test is small. Otherwise, the size of the pre-test procedure varies wildly and unacceptable violations of the prescribed level cannot be avoided. For an explanation of this effect we refer to Section 2 of Albers, Boon, and Kallenberg (1998a), noting that it is mainly caused by the interplay of the basic main test and the preliminary test, which are the same as in Albers, Boon, and Kallenberg (1998a). But even in case of a small correlation between the basic main test and the preliminary test non-negligible departures from the nominal level can occur. While first-order asymptotics do not reveal this feature, the use of second-order asymptotics in Theorem 3.1 makes clear what is going on.
- *The correlation between the basic main test and the second main test.* Having controlled the size of the pre-test procedure, it is indeed possible to get power gain for the pre-test procedure compared to always using the second main test. This will happen when the basic main test and the second main test are not too highly correlated. In principle, the smaller the correlation, the higher the gain in power. This qualitative statement is made more precise in Theorem 3.1, where a transparent approximation also gives the possible (asymptotic) power gain in a quantitative way. By this result the aim of the paper, which is showing whether it is possible to get power gain in pre-testing and if so, giving insight when this occurs and to what extent, is achieved.

The paper is organized as follows. In Section 2 the framework for our analysis is set. Notation is introduced, the various models are specified and the classes of statistics to be considered are given, together with some natural regularity conditions. It should be remarked that the classes of test statistics for the parametric testing problems contain all standard first-order optimal tests, like the likelihood ratio test, Rao's efficient score test or Wald's test.

In Section 3 we present an approximation for the size and power difference of the pre-test procedure with respect to the robust second main test. This approximation is based on second-order asymptotics in  $n$ , since first-order asymptotics is generally insufficient to explain deviations of the size of the procedure, as explained in Section 2. However, in our treatment of the results, we will focus

on the first order part, since possible power gain is a first-order matter and the second-order part is the same as in Albers, Boon, and Kallenberg (1998a). It turns out that for all members of the class of test statistics the same expression holds, except for one quantity coming from the basic main test.

In Section 4 the theory is exemplified by considering essentially the  $t$ -test as basic main test for the null hypothesis of a zero median, and the sign test as second main test. The accuracy of the approximations is seen from comparison with numerical results.

## 2 Notation, assumptions and preliminaries

Let  $X_1, \dots, X_n$  be i.i.d. random variables (r.v.'s) with density  $f$ . Suppose that we conjecture that this density can be parameterized as  $f(x; \theta)$ . For simplicity of presentation we take  $\theta$  one-dimensional, but an extension to higher dimensions is possible. Suppose that, according to this parameterization, we are interested in testing the main hypothesis  $H_0 : \theta = 0$  against  $H_1 : \theta > 0$ . This can be done by a test in the restricted model  $f(x; \theta)$ . But in order to determine whether this model is suitable, a preliminary test is performed. This test tests whether there are deviations from the model  $f(x; \theta)$  into a direction parameterized by a parameter  $\tau$  (also one-dimensional). Hence the preliminary test aims at testing  $\bar{H}_0 : \tau = 0$  against  $\bar{H}_1 : \tau \neq 0$  in a model  $f(x; \theta, \tau)$ . After acceptance by the preliminary test, the main hypothesis on  $\theta$  is tested by the basic main test in the restricted model  $f(x; \theta)$ . After rejection by the preliminary test, no restrictive assumptions on the density  $f$  are made and it is assumed that the hypotheses of the main testing problem can be written in the form  $H_0^* : \int \xi f = 0$  against  $H_1^* : \int \xi f > 0$  for some function  $\xi$ . Without further assumptions on the form of the distribution, it is asymptotically optimal to base the test simply on the empirical estimator of  $\int \xi f$ , given by  $n^{-1} \sum \xi(X_i)$  (van der Vaart (1998), p. 368, 385).

So, we consider three models. The smallest one is denoted by  $f(x; \theta)$ . Here, we test  $H_0 : \theta = 0$  against  $H_1 : \theta > 0$ . This model may be incorrect. This is checked by a preliminary test which is appropriate in the somewhat larger model  $f(x; \theta, \tau)$  with  $f(x; \theta, 0) = f(x; \theta)$ . Here we test  $\bar{H}_0 : \tau = 0$  against  $\bar{H}_1 : \tau \neq 0$ . Finally, if the preliminary test rejects, the main testing problem is replaced by testing  $H_0^* : \int \xi f = 0$  against  $H_1^* : \int \xi f > 0$  with the function  $\xi$  such that  $\theta = 0$  gives  $H_0^* : \int \xi(x) f(x; 0) dx = 0$ . Since,  $H_0^*$  is the extension of  $H_0$  in a larger model, including at least the model  $f(x; \theta, \tau)$ , we have also  $\int \xi(x) f(x; 0, \tau) dx = 0$  for all  $\tau$ . Because the preliminary test protects only against alternatives of the form  $f(x; \theta, \tau)$ , we analyze the pre-test procedure in this model.

Before defining the test statistics, we will introduce some further notation. We assume that regularity conditions of the same type as in Lehmann (1983), p.

429 hold. Then we can define

$$\psi_{ij}^*(x; \theta, \tau) = \frac{(\partial^{i+j}/\partial\theta^i\partial\tau^j)f(x; \theta, \tau)}{f(x; \theta, \tau)}.$$

In order to simplify notation, we will write  $\psi_{ij}^*$  or  $\psi_{ij}^*(X)$  for  $\psi_{ij}^*(X; 0, 0)$ . Expectations under the model  $f(x; \theta, \tau)$  are denoted by  $E_{\theta, \tau}$ , but will be abbreviated to  $E$  under  $(\theta, \tau) = (0, 0)$ . The Fisher information matrix reads

$$I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} = \begin{pmatrix} E(\psi_{10}^*)^2 & E\psi_{10}^*\psi_{01}^* \\ E\psi_{10}^*\psi_{01}^* & E(\psi_{01}^*)^2 \end{pmatrix}.$$

Furthermore, let

$$\psi_{ij}(x) = \psi_{ij}^*(x)I_{11}^{-i/2}I_{22}^{-j/2}.$$

Then the correlation coefficient of  $\psi_{10}(X_i)$  and  $\psi_{01}(X_i)$  under  $(\theta, \tau) = (0, 0)$  is given by

$$\rho = I_{12}(I_{11}I_{22})^{-1/2}. \quad (2.1)$$

By application of the dominated convergence theorem it follows from the regularity conditions that

$$E_{\theta, \tau}\psi_{ij}(X; \theta, \tau) = 0 \quad \text{for } i, j = 0, 1, 2, 3, \quad (i, j) \neq (0, 0). \quad (2.2)$$

Define the score functions under  $(0, 0)$

$$S = n^{-1/2} \sum_{i=1}^n \psi_{10}(X_i) \quad \text{and} \quad T = n^{-1/2} \sum_{i=1}^n \psi_{01}(X_i) \quad (2.3)$$

and let

$$\Xi = n^{-1/2} \sum_{i=1}^n \xi(X_i).$$

Our investigation will be based on asymptotics. We will consider local alternatives of the form  $(\theta_n, \tau_n)$ , with

$$\theta_n = bI_{11}^{-1/2}n^{-1/2}, \quad \tau_n = cI_{22}^{-1/2}n^{-1/2} \quad \text{with } b > 0 \text{ and } c \in \mathbb{R}, \quad (2.4)$$

where the normalization with  $I_{11}^{-1/2}$  and  $I_{22}^{-1/2}$  is done for convenience in the results. For our analysis first-order asymptotics are generally not sufficient, which is seen as follows. The test statistic for the basic main test will be based on  $S$  to first order, and will hence be asymptotically normal with expectation  $b + c\rho$  and variance 1 under  $f(x; \theta_n, \tau_n)$ . The test statistic for the second main test will be

based on  $\Xi$  to first order, and converges to a normal distribution with expectation  $bE\xi\psi_{10}$  and variance 1. In situations where  $\rho = 0$ , the size of the two main tests is thus the same to first order. Hence, to first order, the size of the pre-test procedure is equal to the nominal level if  $\rho = 0$ . Nevertheless, numerical work shows that differences in size between the pre-test procedure and the nominal level are certainly not negligible in practice. To explain these differences, we need higher order asymptotics.

Furthermore, we also expand for  $\rho$  small. First note that under  $f(x; 0, \tau_n)$ , the distribution of the test statistic for the basic main test defined in (2.5) below, converges to a normal one with expectation  $c\rho$  and variance 1. At the same time, the test statistic for the preliminary test defined in (2.6) and (2.7) below, converges to a normal r.v. with expectation  $c\sqrt{1 - \rho^2}$  and variance 1. Hence, for the preliminary test to have large power it is necessary that  $c$  is large. But then the error in the size of the basic main test, and together with it the error in the size of the total procedure, grows very large, unless  $\rho$  is small. Therefore, the procedure is unacceptable for large  $\rho$  and it is only natural to restrict attention to the case that  $\rho$  is small, which will be assumed from now on. This justifies the fact that the results given in the next section are derived using expansions not only with respect to  $n$ , but also with respect to  $\rho$ , and that the results are presented up to order  $\rho^3 + \rho n^{-1/2}$ . In this way the results are not only accurate, but also transparent, which would not have been the case if  $\rho$  were taken fixed.

Now we will define classes of tests for each of the three testing problems. The easiest one is the basic main test with which we have to test  $H_0 : \theta = 0$  in the family  $f(x; \theta, 0)$ . The statistic  $S$  corresponds to the locally most powerful test for this problem. Other standard tests for  $H_0$  are e.g. the likelihood ratio (LR) test, the test based on the maximum likelihood estimator (MLE) and Wald's test. These are first-order optimal tests, all having the same leading term, and can (up to order  $n^{-1/2}$ ) be written in the form

$$SK = S + n^{-1/2}S \left\{ n^{-1/2} \sum_{i=1}^n k(X_i) \right\} \quad (2.5)$$

for some function  $k$  with  $Ek = 0$ . Here, “ $K$ ” refers to the fact that the underlying distribution, except for  $\theta$  itself, is assumed to be known. The null hypothesis  $H_0 : \theta = 0$  is rejected in favor of  $H_1 : \theta > 0$  for large values of  $SK$ . In particular, we have the following functions  $k$  for the tests mentioned:

$$\text{LMP} \quad : k = 0$$

$$\text{LR} \quad : k = \frac{1}{2}\{\psi_{20} - (\psi_{10}^2 - 1)\} + \left(\frac{1}{3}E\psi_{10}^3 - \frac{1}{2}E\psi_{10}\psi_{20}\right)\psi_{10}$$

$$\text{MLE} \quad : k = \psi_{20} - (\psi_{10}^2 - 1) + (E\psi_{10}^3 - \frac{3}{2}E\psi_{10}\psi_{20})\psi_{10}$$

$$\text{Wald} \quad : k = \psi_{20} - (\psi_{10}^2 - 1) + \frac{1}{2}(E\psi_{10}^3 - E\psi_{10}\psi_{20})\psi_{10}.$$

Note that Rao's efficient score test coincides with the LMP test.

For testing  $\bar{H}_0 : \tau = 0$  against (two-sided) alternatives in a parametric model  $f(x; \theta, \tau)$ , the test is based on the efficient score function for  $\tau$  with nuisance parameter  $\theta$ . The corresponding test statistic has the form  $[T - \rho S](1 - \rho^2)^{-1/2}$ , and is asymptotically efficient (Pfanzagl (1980), sec.11-13). Similar to (2.5) we consider a class of first-order optimal tests, namely the class of tests based on

$$\left[ T - \rho S + n^{-1/2} \left\{ T n^{-1/2} \sum_{i=1}^n q(X_i) + S n^{-1/2} \sum_{i=1}^n r^*(X_i) \right\} \right] (1 - \rho^2)^{-1/2} \quad (2.6)$$

with functions  $q$  and  $r^*$  satisfying  $Eq = Er^* = 0$ . A further modification is needed, since up to second order the expectation and variance under  $f(x; \theta, 0)$  of this expression depend on the unknown  $\theta$ . Since we deal with a two-sided test, the small bias in expectation does not affect the size of the test up to the order  $n^{-1/2}$ , but for the variance a correction is needed. The variance of (2.6) under  $(bI_{11}^{-1/2}n^{-1/2}, 0)$  equals up to order  $n^{-1/2}$  and  $\rho^2$ , but ignoring terms of order  $\rho n^{-1/2}$ ,

$$1 + bn^{-1/2} \{ E\psi_{10}\psi_{01}^2 + 2[Eq\psi_{10} + Er^*\psi_{01}] \}.$$

Studentizing by plugging in  $S$  as an estimator of  $b$  leads to a statistic  $TU$  of the given form (2.6), with  $r^*$  replaced by

$$r = r^* - \frac{1}{2} \{ E\psi_{10}\psi_{01}^2 + 2[Eq\psi_{10} + Er^*\psi_{01}] \} \psi_{01}, \quad (2.7)$$

ignoring terms of order  $\rho n^{-1/2}$  and  $n^{-1}$ . The preliminary test rejects for large values of  $|TU|$ , where "U" refers to the fact that in this testing problem  $\theta$  is unknown.

The second main test is based on a non-parametric model. In that model, the null hypothesis is written as  $H_0^* : \int \xi f = 0$ . Since  $f(x; \theta, \tau)$  is a submodel of the large model,  $f(x; 0, \tau)$  belongs to  $H_0^*$ , implying  $E_{0,\tau}\xi = 0$  for all  $\tau$  and hence  $E\xi = E\xi\psi_{01} = E\xi\psi_{02} = 0$ . Moreover,  $f(x; \theta, 0)$  belongs to  $H_1^*$  for  $\theta > 0$ , implying  $E_{\theta,0}\xi > 0$  for  $\theta > 0$ . Hence, it is natural to assume that  $E\xi\psi_{10} \geq 0$ . This indicates that  $\xi$  and  $\psi_{10}$  are positively correlated, as they should be, since tests based on these functions are meant for the same hypothesis, written as  $H_0$  or  $H_0^*$ . Without loss of generality let also  $E\xi^2 = 1$ . Then it follows from the Cauchy-Schwarz inequality that  $E\xi\psi_{10} = E\xi(\psi_{10} - \rho\psi_{01}) \leq \sqrt{1 - \rho^2}$ . Hence, by the assumption that  $E\xi\psi_{10} \geq 0$ ,  $E\xi\psi_{10}$  can be written as  $E\xi\psi_{10} = \sqrt{(1 - \rho^2)(1 - \gamma^2)}$  for some  $\gamma$ .

In order to keep convenient clarity in the results, it will be assumed that terms of order  $\gamma n^{-1/2}$ , with  $\gamma$  from  $E\xi\psi_{10} = \sqrt{(1 - \rho^2)(1 - \gamma^2)}$ , are negligible. For  $\gamma$  small, this is of course justified. If  $\gamma$  is not small, it is also justified because then the first-order terms which depend on  $\gamma$ , will dominate. As a

consequence of this assumption, we may replace  $\xi$  by  $\psi_{10}$  in  $n^{-1/2}$ -terms with factors of the form  $E\xi g$ ,  $E\xi^2 g$ ,  $E\xi^3$ . This follows from the Cauchy-Schwarz inequality:  $E(\xi - \psi_{10})g \leq \sqrt{E(\xi - \psi_{10})^2 E g^2} = \sqrt{O(\rho^2 + \gamma^2) E g^2} = O(\rho + \gamma)$ , provided  $E g^2 < \infty$ , which we assume to hold for the  $g$ 's under consideration. Since  $E\xi\psi_{02} = 0$ , terms with  $E\psi_{10}\psi_{02}n^{-1/2}$  are of order  $\rho n^{-1/2} + \gamma n^{-1/2}$  and can also be ignored.

The negligibility of terms of order  $\rho n^{-1/2}$  and  $\gamma n^{-1/2}$  will simplify the calculation of the second-order part of the approximate power gain. By considering the power difference from the beginning on, and using the similarity of the second-order parts of the two main tests, we can avoid a lot of terms which we would have to deal with if we would calculate the powers of the two-stage procedure and of the second main test separately, and take differences afterwards.

The empirical estimator for  $\int \xi f$  equals  $n^{-1/2}\Xi$ . This can be extended to a class of test statistics based on

$$\Xi + n^{-1/2}\Xi \left\{ n^{-1/2} \sum_{i=1}^n w^*(X_i) \right\} \quad (2.8)$$

with  $Ew^* = 0$ . Under the assumptions concerning  $\xi$ , the expectation under the submodel  $f(x; 0, \tau)$  does not depend on the unknown  $\tau$  up to the considered order, but again we need to replace the unknown  $\tau$  in the variance by a  $\sqrt{n}$ -consistent estimator in order to get a well-defined test statistic. The variance under  $(0, cI_{22}^{-1/2}n^{-1/2})$  equals up to order  $n^{-1/2}$

$$1 + cn^{-1/2}\{E\xi^2\psi_{01} + 2Ew^*\psi_{01}\}$$

Studentizing leads to a statistic  $SU$  of the form (2.8) with  $w^*$  replaced by

$$w = w^* - \frac{1}{2}\{E\xi^2\psi_{01} + 2Ew^*\psi_{01}\}\psi_{01}. \quad (2.9)$$

Note that studentizing  $\Xi$  by means of its sample variance, corresponds to  $w^* = -\frac{1}{2}(\xi^2 - 1)$ . Then no further correction is needed. Also ignoring terms of order  $\rho n^{-1/2} + \gamma n^{-1/2}$ , we may replace  $\xi$  by  $\psi_{10}$  and use

$$w = w^* - \frac{1}{2}\{E\psi_{10}^2\psi_{01} + 2Ew^*\psi_{01}\}\psi_{01}. \quad (2.10)$$

The second main test rejects for large values of  $SU$ .

Now we know from (2.5), (2.6) and (2.7), and from (2.8) and (2.9) or (2.10) the form of the test statistics  $TU$ ,  $SK$  and  $SU$  for the preliminary test and the two main tests. We will investigate the power (or size) difference between the pre-test procedure and the second main test based on  $SU$ . The latter has the correct level (at least to  $O(\rho n^{-1/2} + \gamma n^{-1/2}) + o(n^{-1/2})$ ), since it does not use possibly incorrect information about the form of the distribution.

### 3 Main result

In this section we will give an approximation for the difference in size and power between the pre-test procedure and the main test based on  $SU$ . For this approximation we need a bivariate Edgeworth expansion for the simultaneous distribution of the statistics corresponding to the preliminary test and each of the main tests. From that expansion we can derive the critical values for the three testing problems. Then we have all the ingredients to state the main result in Theorem 3.1.

First, let

$$U_{1n} = \frac{SK - \mu_{1n}(b, c)}{\sigma_{1n}(b, c)}, \quad U_{2n} = \frac{SU - \mu_{2n}(b, c)}{\sigma_{2n}(b)} \quad \text{and} \quad V_n = \frac{TU - \mu_n(b, c)}{\sigma_n(c)}$$

be the standardized statistics, standardized by their expectations and variances under local alternatives. Define  $\lambda = \sqrt{1 - \gamma^2}$ . Direct calculation yields, ignoring terms of order  $\rho^4$ ,  $\rho n^{-1/2}$ ,  $\gamma n^{-1/2}$  and  $n^{-1}$

$$\begin{aligned} \mu_{1n}(b, c) &= b + c\rho + \frac{1}{2}\{b^2 E\psi_{10}\psi_{20} + 2bcE\psi_{10}\psi_{11} + 2(1 + b^2)Ek\psi_{10} \\ &\quad + 2bcEk\psi_{01}\}n^{-1/2}, \\ \sigma_{1n}(b, c) &= 1 + \frac{1}{2}\{bE\psi_{10}^3 + cE\psi_{10}^2\psi_{01} + 4bEk\psi_{10} + 2cEk\psi_{01}\}n^{-1/2}, \\ \mu_{2n}(b, c) &= b\lambda(1 - \frac{1}{2}\rho^2) + \frac{1}{2}\{b^2 E\psi_{10}\psi_{20} + 2bcE\psi_{10}\psi_{11} - bcE\psi_{10}^2\psi_{01} \\ &\quad + 2(1 + b^2)Ew\psi_{10}\}n^{-1/2}, \\ \sigma_{2n}(b) &= 1 + \frac{1}{2}b(E\psi_{10}^3 + 4Ew\psi_{10})n^{-1/2}, \\ \mu_n(b, c) &= c - \frac{1}{2}c\rho^2 + \frac{1}{2}\{c^2 E\psi_{01}\psi_{02} + 2bcE\psi_{01}\psi_{11} + b^2 E\psi_{01}\psi_{20} \\ &\quad - bcE\psi_{10}\psi_{01}^2 + 2(1 + b^2)Er\psi_{10} + 2(1 + c^2)Eq\psi_{01}\}n^{-1/2}, \\ \sigma_n(c) &= 1 + \frac{1}{2}c(E\psi_{01}^3 + 4Eq\psi_{01})n^{-1/2}. \end{aligned} \tag{3.1}$$

Except for the cancellation of  $E\psi_{10}\psi_{02}n^{-1/2}$ , the expectations and variances  $\mu_{1n}$ ,  $\sigma_{1n}$ ,  $\mu_n$  and  $\sigma_n$  for  $SK$  and  $TU$  are equal to those in Albers, Boon, and Kallenberg (1998a), since the form of these test statistics is unchanged. The only difference compared to that paper occurs in the leading term of  $\mu_{2n}$ , where  $b$  is replaced by  $b\lambda$ . The order  $n^{-1/2}$ -terms of  $\mu_{2n}$  and  $\sigma_{2n}$  remain the same (up to  $O(\rho n^{-1/2} + \gamma n^{-1/2})$ ), due to the replacement of  $\xi$  by  $\psi_{10}$ .

Under standard regularity conditions for Edgeworth expansions (cf. Albers, Boon, and Kallenberg (1997b)), we have the following expansion for the stan-

standardized statistics, which holds uniformly for  $(u, v) \in \mathbb{R}^2$

$$\begin{aligned}
P_{\theta_n, \tau_n}(U_{in} \leq u, V_n \leq v) &= \Phi(u, v; \rho_i) + n^{-1/2} \sum_{j=0}^3 c_{ij} \Phi^{(j)}(u) \Phi^{(3-j)}(v) \\
&+ O(\rho^3 + \rho n^{-1/2} + \gamma n^{-1/2}) + o(n^{-1/2}), \quad i = 1, 2.
\end{aligned} \tag{3.2}$$

Here  $\Phi(\cdot, \cdot; \rho_i)$  denotes the bivariate normal  $N(0, 0, 1, 1, \rho_i)$  distribution function, and  $\Phi^{(j)}$  the  $j$ th derivative of the standard normal distribution function  $\Phi$ . Later on,  $\Phi^{(1)}$  is also denoted by  $\varphi$ . Up to the order  $O(\rho^3 + \rho n^{-1/2} + \gamma n^{-1/2}) + o(n^{-1/2})$ , the correlations  $\rho_1$  and  $\rho_2$  between the preliminary test and the two main tests are given by

$$\begin{aligned}
\rho_1 &= n^{-1/2} r_1(b, c), \quad \rho_2 = -\rho \lambda + n^{-1/2} r_2(b, c) \\
&\text{for some functions } r_1(b, c), r_2(b, c) \text{ satisfying} \\
r_2(b, c) - r_1(b, c) &= -\frac{1}{2} b E \psi_{10}^2 \psi_{01} - b E k \psi_{01}.
\end{aligned} \tag{3.3}$$

The  $\lambda$  appears in the correlation between the second main test and the preliminary test. For the constants  $c_{ij}$  occurring in the skewness terms of order  $n^{-1/2}$  of the Edgeworth expansion, only the leading terms are needed, and we may even ignore terms of order  $O(\rho + \gamma)$ . Hence, the  $c_{ij}$  are the same as in Albers, Boon, and Kallenberg (1998a), namely

$$\begin{aligned}
c_{10} - c_{20} &= 0, \quad c_{12} - c_{22} = -\frac{1}{2} E \psi_{10}^2 \psi_{01} - E k \psi_{01}, \quad c_{11} - c_{21} = 0, \\
c_{13} &= -\frac{1}{6} E \psi_{10}^3 - E k \psi_{10}, \quad c_{23} = -\frac{1}{6} E \psi_{10}^3 - E w \psi_{10}.
\end{aligned} \tag{3.4}$$

The equality of  $c_{10}$  and  $c_{20}$  follows from the marginal distribution of  $V_n$ .

From the marginal distributions under the corresponding null hypothesis we can easily derive the critical values for the three separate testing problems. Since under the restricted model  $\sigma_{1n}(0, 0) = 1$ , we have

$$P_{0,0}(SK \leq u) = \Phi(u - \mu_{1n}(0, 0)) + n^{-1/2} c_{13} \Phi^{(3)}(u) + o(n^{-1/2})$$

with  $\mu_{1n}(0, 0)$  and  $c_{13}$  from (3.1) and (3.4). Hence, the one-sided test which rejects  $H_0$  when

$$SK > u_\alpha - c_{13} n^{-1/2} (u_\alpha^2 - 1) + \mu_{1n}(0, 0),$$

with  $u_\alpha = \Phi^{-1}(1 - \alpha)$ , has size  $\alpha + o(n^{-1/2})$  under the restricted model.

Similarly, with  $\mu_{2n}(0, c)$  and  $c_{23}$  from (3.1) and (3.4), and  $\sigma_{2n}(0) = 1$ , the test based on  $SU$  rejects  $H_0 : \theta = 0$  when

$$SU > u_\alpha - c_{23}n^{-1/2}(u_\alpha^2 - 1) + \mu_{2n}(0, c),$$

giving size  $\alpha + O(\rho n^{-1/2} + \gamma n^{-1/2}) + o(n^{-1/2})$  under  $(0, \tau_n)$ .

$TU$  is a two-sided test for  $\bar{H}_0 : \tau = 0$ . Since  $\mu_n(b, 0) = O(n^{-1/2})$  and  $\sigma_n(0) = 1$ , the test which rejects when

$$|TU| > u_{\delta/2}$$

gives size  $\delta + O(\rho n^{-1/2}) + o(n^{-1/2})$  under  $(\theta_n, 0)$ . Note that  $n^{-1/2}$ -terms which occurred in the critical values for  $SK$  and  $SU$  are not necessary here, since a shift of order  $n^{-1/2}$  of both upper and lower critical value in the same direction does not affect the size of the test.

Combining the three tests, it follows that the pre-test procedure rejects  $H_0 : \theta = 0$  against  $H_1 : \theta > 0$  when

$$SK > u_\alpha + \frac{1}{6}E\psi_{10}^3 n^{-1/2}(u_\alpha^2 - 1) + Ek\psi_{10} n^{-1/2}u_\alpha^2 \quad \text{and} \quad |TU| \leq u_{\delta/2}$$

(3.5)

or when

$$SU > u_\alpha + \frac{1}{6}E\psi_{10}^3 n^{-1/2}(u_\alpha^2 - 1) + Ew\psi_{10} n^{-1/2}u_\alpha^2 \quad \text{and} \quad |TU| > u_{\delta/2}.$$

The probability that the pre-test procedure rejects, depends on the local alternatives  $(\theta_n, \tau_n)$  parameterized by  $b$  and  $c$ , and is denoted by  $\pi^*(b, c)$ . The probability that  $SU$  rejects will be denoted by  $\tilde{\pi}(b, c)$ . The following theorem gives the main result, an expression for the approximate power gain  $\pi^*(b, c) - \tilde{\pi}(b, c)$  of the two-stage pre-test procedure compared to the second main test based on  $SU$ , which does not assume restrictive knowledge about the underlying distribution.

**Theorem 3.1** *Suppose that the regularity conditions are satisfied. Then*

$$\begin{aligned} & \pi^*(b, c) - \tilde{\pi}(b, c) \\ &= \{\Phi(u_\alpha - b\lambda) - \Phi(u_\alpha - b)\}\{\Phi(u_{\delta/2} - c) - \Phi(-u_{\delta/2} - c)\} \\ &+ h_1(c, u_{\delta/2})[\rho\varphi(u_\alpha - b) + \frac{1}{2}\rho^2\{\frac{b\lambda}{c}\varphi(u_\alpha - b\lambda) + c(u_\alpha - b)\varphi(u_\alpha - b)\} \\ &\quad + m(u_\alpha)\varphi(u_\alpha - b)n^{-1/2}] \\ &+ h_2(c, u_{\delta/2})[\rho\lambda\varphi(u_\alpha - b\lambda) + \frac{1}{2}c\rho^2\{\lambda^2(u_\alpha - b\lambda)\varphi(u_\alpha - b\lambda) \\ &\quad + \Phi(u_\alpha - b) - \Phi(u_\alpha - b\lambda)\} + m(u_\alpha)\varphi(u_\alpha - b)n^{-1/2}] \\ &+ h_3(c, u_{\delta/2})[\rho^2\lambda^2(u_\alpha - b\lambda)\varphi(u_\alpha - b\lambda)] \\ &+ O(\rho^3 + \rho n^{-1/2} + \gamma n^{-1/2}) + o(n^{-1/2}) \text{ as } \rho \rightarrow 0 \text{ and } n \rightarrow \infty, \end{aligned}$$

(3.6)

where

$$h_1(x, y) = x\{\Phi(y - x) - \Phi(-y - x)\},$$

$$h_2(x, y) = \varphi(y + x) - \varphi(y - x),$$

$$h_3(x, y) = \frac{1}{2}y\{\varphi(y + x) + \varphi(y - x)\},$$

$$m(y) = y\{\frac{1}{2}E\psi_{10}^2\psi_{01} + Ek\psi_{01}\}.$$

**Proof.** We only prove the first-order part, since the  $n^{-1/2}$ -terms and their derivation are exactly the same as in Albers, Boon, and Kallenberg (1998a), except for the fact that now also terms of order  $\gamma n^{-1/2}$  are ignored and that terms with  $E\psi_{10}\psi_{02}n^{-1/2}$  cancel. First rewrite the power gain  $\pi^* - \tilde{\pi}$  as

$$\begin{aligned} \pi^*(b, c) - \tilde{\pi}(b, c) &= P_{\theta_n, \tau_n}(U_{1n} \leq u_{1n}, V_n \leq v_n^L) - P_{\theta_n, \tau_n}(U_{2n} \leq u_{2n}, V_n \leq v_n^L) \\ &\quad - \{P_{\theta_n, \tau_n}(U_{1n} \leq u_{1n}, V_n < v_n^U) - P_{\theta_n, \tau_n}(U_{2n} \leq u_{2n}, V_n < v_n^U)\} \end{aligned}$$

with  $U_{1n}$  and  $U_{2n}$  the test statistics for the main tests based on  $SK$  and  $SU$ , standardized under local alternatives,  $u_{1n}$  and  $u_{2n}$  the standardized critical values, and  $V_n$  the standardized test statistic for the two-sided preliminary test, with corresponding lower and upper critical value  $v_n^L$  and  $v_n^U$ . These critical values equal

$$u_{1n} = u_\alpha - b - c\rho + O(n^{-1/2}),$$

$$u_{2n} = u_\alpha - b\lambda(1 - \frac{1}{2}\rho^2) + O(n^{-1/2}),$$

$$v_n^U = u_{\delta/2} - c(1 - \frac{1}{2}\rho^2) + O(n^{-1/2}),$$

$$v_n^L = -u_{\delta/2} - c(1 - \frac{1}{2}\rho^2) + O(n^{-1/2}).$$

From equation (3.2) and the rewritten form of  $\pi^* - \tilde{\pi}$ , it follows that the first-order part of the power gain is given by the difference

$$\{\Phi(u_{1n}, v_n^L; \rho_1) - \Phi(u_{2n}, v_n^L; \rho_2)\} - \{\Phi(u_{1n}, v_n^U; \rho_1) - \Phi(u_{2n}, v_n^U; \rho_2)\}.$$

Since  $\rho_1 = O(n^{-1/2})$ , we have to first order

$$\begin{aligned} &\Phi(u_{1n}, v_n^U; \rho_1) - \Phi(u_{2n}, v_n^U; \rho_2) \\ &= \{\Phi(u_{1n}) - \Phi(u_{2n})\}\Phi(v_n^U) - \rho_2\varphi(u_{2n})\varphi(v_n^U) - \frac{1}{2}\rho_2^2\varphi'(u_{2n})\varphi'(v_n^U) \\ &\quad + O(\rho_2^3 + n^{-1/2}). \end{aligned}$$

Substituting  $u_{in}$ ,  $v_n^U$  and  $\rho_2$ , expanding for  $\rho$  small, and throwing away  $O(\rho^3 + n^{-1/2})$ -terms, leads to

$$\begin{aligned} & \{\Phi(u_\alpha - b) - \Phi(u_\alpha - b\lambda)\}\Phi(u_{\delta/2} - c) \\ & + \Phi(u_{\delta/2} - c)[\{-c\rho - \frac{1}{2}c^2\rho^2(u_\alpha - b)\}\varphi(u_\alpha - b) - \frac{1}{2}b\lambda\rho^2\varphi(u_\alpha - b\lambda)] \\ & + \varphi(u_{\delta/2} - c)[\rho\lambda\varphi(u_\alpha - b\lambda) - \frac{1}{2}\rho^2\lambda^2(u_\alpha - b\lambda)(u_{\delta/2} - c)\varphi(u_\alpha - b\lambda) \\ & \quad + \frac{1}{2}c\rho^2\{\Phi(u_\alpha - b) - \Phi(u_\alpha - b\lambda)\}]. \end{aligned}$$

Taking  $u_{\delta/2}$  outside brackets and taking the difference with the expression with  $-u_{\delta/2}$  yields the result.  $\square$

Remarkably, (3.6) depends on the family of distributions and the classes of tests only through the correlations  $\rho$  and  $\lambda$ , and through the expectations  $E\psi_{10}^2\psi_{01}$  and  $E\psi_{01}$  in the function  $m$ . Hence, only the function  $k$  coming from the basic main test appears in the result, and the functions  $q$ ,  $r$  and  $w$  which define the  $n^{-1/2}$ -parts of the other tests, do not influence the final result.

The result  $\pi^* - \tilde{\pi}$  can be decomposed into two parts. Let  $\pi$  denote the power of the basic main test based on  $SK$  and  $\tilde{\pi}$  the power of the preliminary test. Then the first part can be written as the product  $(\pi - \tilde{\pi})(1 - \tilde{\pi})$ , which is the power difference of the two main tests multiplied by the acceptance probability of the preliminary test. This part is recognized in the first three lines of (3.6) together with the  $\rho^2$ -part in the second line of the term with  $h_2$ . This would have been the result if there were no correlation between the subsequent tests in the procedure. The second part consists of the remaining terms, which are due to the correlation between the preliminary test and the main tests.

First consider the size of the procedure. For the size ( $b = 0$ ) the approximation (3.6) reduces to

$$\begin{aligned} & \varphi(u_\alpha)\{ h_1(c, u_{\delta/2})[\rho + \frac{1}{2}c\rho^2u_\alpha + m(u_\alpha)n^{-1/2}] \\ & \quad + h_2(c, u_{\delta/2})[\rho\lambda + \frac{1}{2}c\rho^2\lambda^2u_\alpha + m(u_\alpha)n^{-1/2}] \\ & \quad + h_3(c, u_{\delta/2})[\rho^2\lambda^2u_\alpha] \} \end{aligned} \tag{3.7}$$

For  $\lambda = 1$ , the coefficients for  $h_1$  and  $h_2$  are equal. The behavior of the size can be deduced by analyzing the properties of the functions  $h_1$ ,  $h_2$  and  $h_3$ . Note that  $h_1$  and  $h_2$  are odd and  $h_3$  even in  $c$ , so it suffices to consider  $c \geq 0$ . Then  $h_1 \geq 0$  and  $h_2 \leq 0$ . Since  $h_1(c, 0) = h_2(c, 0) = 0$  and  $h_1 + h_2$  increases in  $u_{\delta/2}$  for  $c \geq 0$ , the sum  $h_1 + h_2 \geq 0$  for  $c \geq 0$ . Hence,  $0 \leq h_1 + h_2 \leq h_1$ . The same holds when we replace  $h_2$  by  $\lambda h_2$  or  $\lambda^2 h_2$ , since  $0 \leq \lambda \leq 1$ .

With respect to the size we can conclude the following. Direct application of the basic main test ( $\delta = 0$ ) would lead to dramatic deviations of the size, since then the deviation grows linearly in  $c$ . Application of the pre-test procedure ( $\delta > 0$ ) improves matters. Then the violation of the level is smaller than that of the basic main test, for two reasons. The first reason for this is that in the function  $h_1$ , the  $c$  is multiplied by a factor  $(1 - \bar{\pi})$  which is smaller than 1. Contrary to  $c$ , the function  $h_1$  reaches a maximum for some  $c \geq 0$ , and then decreases, reflecting that at some point the power of the preliminary test becomes large enough to detect harmful deviations from the restricted model. Secondly, the correlation term with  $h_2$  partly compensates the deviation. This compensation occurs as long as  $\rho$  is small enough in order that the  $\rho^2$ -terms in the term with  $h_3$  cannot overrule this again.

Notice however, that in situations where  $\rho = 0$ , all first-order terms cancel. Hence, the  $n^{-1/2}$ -terms are essential to explain differences between the actual size of the pre-test procedure and the prescribed level.

Now let us pay attention to the power. In case  $\lambda = 1$ , power gain is only possible at the cost of a larger size. In the present situation, where  $\lambda < 1$ , the terms with  $\Phi(u_\alpha - b\lambda) - \Phi(u_\alpha - b)$  do not cancel and lead to a possibly substantial increase of power without automatically exceeding the prescribed level.

If the correlation between the two main tests is smaller ( $\lambda$  smaller or  $\rho$  larger), indicating that the two main tests differ more, then for smaller  $\lambda$  in combination with a small  $\rho$ , a larger (possibly substantial) power gain of the pre-test procedure compared to the power of the second main test can be achieved without getting unacceptable sizes. For  $\rho = 0$  the result reduces to

$$\begin{aligned} & \{\Phi(u_\alpha - b\lambda) - \Phi(u_\alpha - b)\} \{\Phi(u_{\delta/2} - c) - \Phi(-u_{\delta/2} - c)\} \\ & + \{h_1(c, u_{\delta/2}) + h_2(c, u_{\delta/2})\} m(u_\alpha) \varphi(u_\alpha - b) n^{-1/2} \end{aligned} \tag{3.8}$$

and  $\lambda$  occurs only in the leading term. Hence, in this case the difference between the two main tests does not influence the size at all. Therefore, for  $\rho = 0$  one can first control the size, which then does not change anymore if the second test is chosen such that  $\lambda$  is smaller, and a substantial power gain can be achieved.

## 4 Numerical example: testing the median

The theory of the preceding section is now illustrated by means of the previously mentioned example, in which a preliminary test is performed to choose which of two main tests is used to test whether the median equals zero. For this example, we compare simulation results regarding power and size of the pre-test procedure with the approximation given in Theorem 3.1.

Consider the restricted model  $f(x; \theta) = \varphi(x - \theta)$ . The main testing problem in this model concerns testing  $H_0 : \theta = 0$  against  $H_1 : \theta > 0$ . From this model it

follows that  $\psi_{10}(x) = x$ . Hence, a first-order optimal test within this model is the Gauss test, corresponding to  $SK$  from (2.5) with  $k = 0$ . Its critical value equals  $u_\alpha$ , which corresponds to (3.5), since  $E\psi_{10}^3 = Ek\psi_{10} = 0$ . The null hypothesis of a zero median is rejected when  $n^{-1/2} \sum X_i > u_\alpha$ .

Note that we use the Gauss test instead of the  $t$ -test. In principle the theory can be extended with additional nuisance parameters in the restricted model to account for an unknown variance, but for this paper that is not relevant and might distract the attention from the purpose of this paper: showing that power gain is possible and in what kind of situations. For this purpose, the difference between the Gauss test and the  $t$ -test is not essential.

In case the restricted model might not hold, we need a more general test. Interpreting  $\theta$  in the restricted model as the median, we can for any continuous density  $f$  rewrite the hypothesis as  $H_0^* : \int \xi f = 0$  with  $\xi(x) = \text{sign}(x)$ . Here  $\text{sign}(x) = \pm 1$  for  $x > 0$  or  $x < 0$ , respectively, and  $\text{sign}(x) = 0$  for  $x = 0$ . The second main test becomes the sign test, based on the difference between the number of observations strictly larger than 0, and the number of observations strictly smaller than 0. Within the class of all continuous distributions, this test is well-known to be uniformly most powerful for testing  $H_0^*$  (Lehmann (1986), p. 106, 107). We take  $w^* = 0$  in (2.8). Since the variance of  $\Xi$  exactly equals 1, we do not explicitly need to studentize, and hence  $w = w^* = 0$  in  $SU$ . This is in accordance with (2.9). Again we have  $E\psi_{10}^3 = 0$  and  $Ew\psi_{10} = 0$  in (3.5). So the critical value reduces to  $u_\alpha$  again.

However, the test statistic  $\Xi$  has a lattice distribution with jumps of order  $n^{-1/2}$ . Hence, the remainder term of approximation (3.6) presented in Theorem 3.1 cannot be of order  $o(n^{-1/2})$  if the second main test rejects when  $\Xi > u_\alpha$ . But a small modification suffices to achieve the order given by the theorem. An Edgeworth expansion for the distribution of the discrete variable remains valid to order  $n^{-1/2}$  when the original distribution function is replaced by its polygonal approximant (which is continuous and equals the original distribution function at midpoints of the lattice and which equals the average of its upper and lower limits at lattice points), (see Feller (1966), p. 540). Using randomization, we can develop a test with the same size and power (to the order considered) as the test which would reject when a r.v., distributed according to the polygonal approximant of the distribution of  $\Xi$ , would exceed  $u_\alpha$ .

Let  $c_{\alpha,n} = u_\alpha n^{1/2} + 1$ . Then the nearest lattice point of the distribution of  $\sum \xi(X_i)$  below  $c_{\alpha,n}$  is the integer  $k_{\alpha,n}$  given by

$$k_{\alpha,n} = \begin{cases} 2\lfloor \frac{1}{2}c_{\alpha,n} \rfloor & \text{if } n \text{ even} \\ 2\lfloor \frac{1}{2}(c_{\alpha,n} + 1) \rfloor - 1 & \text{if } n \text{ odd.} \end{cases} \quad (4.1)$$

If  $p = \frac{1}{2}(c_{\alpha,n} - k_{\alpha,n})$  denotes the difference between the real-valued and the integer-valued ‘‘critical values’’, divided by 2 such that  $p$  lies between 0 and 1, then the

test which rejects

$$\begin{cases} \text{with probability } 1 & \text{if } \sum \xi(X_i) > k_{\alpha,n} \\ \text{with probability } 1 - p & \text{if } \sum \xi(X_i) = k_{\alpha,n}, \end{cases} \quad (4.2)$$

has the same size and power as the test based on the “continuous version” of  $\Xi$ . The approximation given by Theorem 3.1 is applied with the sign test randomized in this way.

Finally, we have to give the model  $f(x; \theta, \tau)$  and the corresponding preliminary test which has to determine whether the Gauss test or the sign test is used to test the main null hypothesis of a zero median. For this testing problem there is considerable freedom of choice. Therefore, we consider a whole class of alternative models  $f(x; \theta, \tau)$  for the deviations from the restricted model  $f(x; \theta)$ . Corresponding to this class, there is a class of preliminary tests which are (first-order) optimal for the considered model. The preliminary test aims to test  $H_0 : \tau = 0$  against  $H_1 : \tau \neq 0$ , where  $\tau$  parametrizes the deviations from the restricted model against which protection should be offered. It is natural to consider deviations of the following form

$$\begin{aligned} f(x; \theta, \tau) &= (1 - \tau)\varphi(x - \theta) + \tau g(x - \theta) \\ &= \varphi(x - \theta) \left\{ 1 + \tau \left( \frac{g}{\varphi}(x - \theta) - 1 \right) \right\} \end{aligned} \quad (4.3)$$

for some given function  $g$ . If  $g$  is a density and  $\tau$  is positive (between 0 and 1), then this can be interpreted as a contamination model.

There are two conditions which must be satisfied, as already explained in Section 2. First, it is necessary that  $E\xi\psi_{01} = 0$ , since  $f(x; 0, \tau)$  should belong to  $H_0^*$ . Secondly, the side condition of small  $\rho$  has to be respected, since otherwise the size of the procedure may deteriorate.

It is easy to construct a class of functions  $g$  for which these two conditions are satisfied. Let  $g$  be a density with corresponding distribution function  $G$ , and let  $h$  denote a symmetric density on  $(0,1)$  with distribution function  $H$ . We consider the class satisfying

$$\begin{aligned} G &= H \circ \Phi, \text{ hence } g = (h \circ \Phi)\varphi \\ &\text{with } h(t) = (2s + 1)(2t - 1)^{2s} \text{ for } 0 \leq t \leq 1, \quad s = 1, 2, \dots \end{aligned} \quad (4.4)$$

In order to achieve symmetry, the density  $h$  on  $(0,1)$  is an even power with its minimum at  $t = \frac{1}{2}$ . Then  $g$ , and hence also  $\psi_{01}^*(x) = \frac{g}{\varphi}(x) - 1$ , is even, which leads to  $E\xi\psi_{01} = 0$  and  $\rho = E\psi_{10}\psi_{01} = 0$  due to the oddness of both  $\xi$  and  $\psi_{10}$ . The boundedness of  $\psi_{01}^*$  assures that its moments do not deteriorate. Note that in order for  $f(x; 0, \tau) = \varphi(x) \{ 1 + \tau [(2s + 1)(2\Phi(x) - 1)^{2s} - 1] \}$  to be positive for all  $x$ , it is necessary and sufficient that  $-\frac{1}{2s} \leq \tau \leq 1$ .

From (4.3) and (4.4) it follows that  $\psi_{01}^*(x) = (2s + 1)(2\Phi(x) - 1)^{2s} - 1$  and  $\psi_{01}(x) = \psi_{01}^*(x)I_{22}^{-1/2}$  with  $I_{22} = \frac{(2s)^2}{4s+1}$ . The preliminary test is based on  $TU$  with  $\psi_{01}$  as given, and with  $\rho = 0$ ,  $q = 0$  and  $r^* = 0$ . Since  $E\psi_{10}\psi_{01}^2 = 0$  by oddness of  $\psi_{10}$ ,  $r = r^* = 0$  after studentization. The null hypothesis  $\bar{H}_0 : \tau = 0$  is therefore rejected when  $|n^{-1/2} \sum \psi_{01}(X_i)| > u_{\delta/2}$ .

Summarizing, first a preliminary test is performed based on  $TU$  with  $\psi_{01}(x) = \frac{\sqrt{4s+1}}{2s} \{(2s + 1)(2\Phi(x) - 1)^{2s} - 1\}$ ,  $\rho = 0$  and  $q = r = 0$ . After acceptance, the main hypothesis is tested by the Gauss test, based on  $SK$  with  $\psi_{10}(x) = x$  and  $k = 0$ . After rejection, the randomized sign test is used, based on  $SU$  with  $\xi(x) = \text{sign}(x)$  and  $w = 0$ . The pre-test procedure (with these tests) is analyzed in the model given by (4.3) and (4.4).

Now we present numerical results for the case  $s = 1$ , and  $\alpha = \delta = 0.05$ . For  $n = 25, 50, 75$  and  $100$ , and for different values of  $b$  and  $c$ , we performed 100 000 simulations. This implies that the standard deviations for the simulated power differences  $\pi^* - \tilde{\pi}$  are at most  $(100\,000)^{-1/2} = 0.0032$ . If  $\pi^*$  or  $\tilde{\pi}$  is at most  $0.05$ , then this reduces to  $0.0019$ , if they are both smaller than  $0.05$ , then the standard deviation of the difference is not more than  $0.001$ . Observations from the given density were generated using the acceptance-rejection method. As a majorizing function we used a multiple of the standard normal density, since  $f(x; 0, \tau) \leq \varphi(x)\{1 + |\tau|2s\}$  for all  $x$ . The approximate size difference and power gain are given by Theorem 3.1 with  $E\psi_{10}^2\psi_{01} = \frac{\sqrt{15}}{\pi}$ .

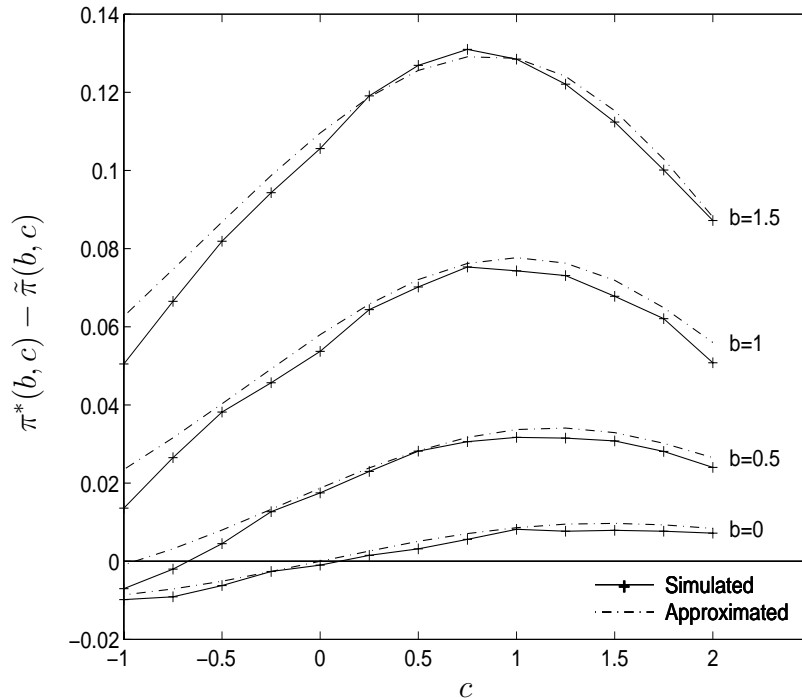
Figure 4.1 shows simulated and approximate values for the power or size difference  $\pi^* - \tilde{\pi}$  as a function of  $c$  for  $n = 50$  and for different values of  $b$ . For the other values of  $n$ , the pattern was not really different, except for the contribution of  $n^{-1/2}$ -terms. These are important for the size deviations, but are relatively small if power gain is considered.

The size difference is always smaller than  $0.014$ , and decreases for larger  $n$ , equalling at most  $0.008$  for  $n = 100$  and  $c \approx 1.5$ . (Note that this is much better than the size difference of the Gauss test, which is monotone and equals more than  $0.02$  for  $c = 2$  and  $n = 100$ , almost  $0.03$  for  $n = 50$ , and about  $0.04$  for  $n = 25$ .) The maximal error between the simulated and approximated value of the size varies between  $0.001$  for  $n = 100$ , and  $0.002$  for smaller  $n$ .

For the power gain, the maximal error of the approximate compared to the simulated value varies for different  $n$  from  $0.003$  to  $0.01$  for  $c > 0$ , and from  $0.01$  to  $0.02$  for  $c < 0$ .

The main result is the substantial power gain which is achieved: for all  $n$  its maximal value for  $b = 1.5$  equals about  $0.13$ , while the simulated power  $\tilde{\pi}$  was only  $0.24$  for  $n = 25$  and  $0.29$  for  $n = 100$ . For  $c = 0$ , where there is no size deviation, the power gain equals for all  $n$  more than  $0.10$ , compared to a power  $\tilde{\pi}$  of less than  $0.33$  for the sign test. So indeed, it is possible to gain much power without getting unacceptable size deviations.

Figure 4.1 Power gain (or size difference) for several values of  $b$ , as a function of the deviation from the restricted model for  $s = 1$ ,  $\alpha = \delta = 0.05$  and  $n = 50$ .



## References

- Albers, W., P. C. Boon, and W. C. M. Kallenberg (1997a). The asymptotic behavior of tests for normal means based on a variance pre-test. Technical Report 1403, University of Twente.
- Albers, W., P. C. Boon, and W. C. M. Kallenberg (1997b). Size and power of pre-test procedures. Technical Report 1421, University of Twente.
- Albers, W., P. C. Boon, and W. C. M. Kallenberg (1998a). Size and power of pre-test procedures. Submitted for publication.
- Albers, W., P. C. Boon, and W. C. M. Kallenberg (1998b). Testing equality of two normal means using a variance pre-test. *Statistics & Probability Letters* 38, 221–227.
- Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, Volume II. New York: Wiley.
- Lehmann, E. L. (1983). *Theory of Point Estimation*. New York: Wiley.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses* (2nd ed.). New York: Wiley.
- Markowski, C. A. and E. P. Markowski (1990). Conditions for the effectiveness of a preliminary test of variance. *The American Statistician* 44 (4), 322–326.

- Moser, B. K., G. R. Stevens, and C. L. Watts (1989). The two-sample t test versus Satterthwaite's approximate F test. *Commun. Statist. -Theor. Meth.* 18(11), 3963–3975.
- Pfanzagl, J. (1980). Asymptotic expansions in parametric statistical theory. In: P. R. Krishnaiah (ed.), *Developments in Statistics, vol. 3*, pp. 1–97. New York: Academic Press.
- Saleh, A. K. M. E. and P. K. Sen (1983). Asymptotic properties of tests of hypothesis following a preliminary test. *Statistics & Decisions* 1, 455–477.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.